Monitoring Social Media: Summarization, Classification and Recommendation

Zhaochun Ren

Monitoring Social Media: Summarization, Classification and Recommendation

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
Prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in
de Agnietenkapel
op donderdag 6 oktober 2016, te 10:00 uur

door

Zhaochun Ren

geboren te Shandong, China

Promotiecommissie

Promotor:		
	Prof. dr. M. de Rijke	Universiteit van Amsterdam
Co-promotor:		
	Dr. E. Kanoulas	Universiteit van Amsterdam
Overige leden:		
	Prof. dr. A. van den Bosch	Radboud University
	Prof. dr. J. Ma	Shandong University
	Dr. M. Marx	Universiteit van Amsterdam
	Dr. C. Monz	Universiteit van Amsterdam
	Prof. dr. M. Worring	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

SIKS Dissertation Series No. 2016-35
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

The printing of this thesis was supported by the Co van Ledden Hulsebosch Centrum, Amsterdam Center for Forensic Science and Medicine.

The research was supported by the Netherlands Organization for Scientific Research (NWO) under project number 727.011.005.

Cover by Xiaoxiao Meng, David Graus
Printed by Off Page, Amsterdam

ISBN: 978-94-6182-721-0

Acknowledgements

I started my doctoral studies at the University of Amsterdam in 2012, when I joined the Information and Language Processing Systems (ILPS) group. It has been a long journey to complete this thesis, which has been a truly life-changing experience for me. This thesis would not have been possible to come about without the support and guidance that I received from many people.

First and for most, I would like to express my sincere gratitude to my supervisor Prof. Maarten de Rijke for his supervision of my doctoral research during the past four years. Maarten taught me how to tackle research questions and express ideas. During the past four years, he gave me enormous advice, patience and support. His fabulous talent and humor have helped me overcome lots of serious research challenges.

I would like to thank my co-advisor Dr. Evangelos Kanoulas for his brilliance and motivation during our discussions. I am grateful to Prof. James Allan from the University of Massachusetts Amherst, Prof. Kathleen McKeown from Columbia University, Prof. Douglas Oard from the University of Maryland, and Prof. Gerhard Weikum from the Max-Planck-Institut für Informatik for their support during my visits.

I appreciate the financial support from the Netherlands Organisation for Scientific Research (NWO) that funded the research presented in this dissertation. I thank the Dutch Research School for Information and Knowledge Systems (SIKS) and the Co van Ledden Hulsebosch Centrum (CLHC) for their additional support.

I'm very honored to have Prof. Antal van den Bosch, Prof. Jun Ma, Dr. Maarten Marx, Dr. Christof Monz, and Prof. Marcel Worring as my committee members.

I would like to especially thank all my co-authors: Edgar, Hendrike, Hongya, Lora, Oana, Piji, Shangsong, Shuaiqiang, Willemijn and Yukun, for enthusiastic support and help during our research collaborations. I'm also very grateful to several former colleagues in ILPS: Abdo, Jiyin, Katja, Manos, Marc and Wouter, for sharing their research experience when I just started my PhD.

I want to thank all the other people in and around ILPS group. I have been lucky to work with brilliant colleagues: Adith, Aldo, Alexy, Aleksandr, Anne, Arianna, Artem, Bob, Christophe, Chuan, Cristina, Daan, Damien, David, David, Evan, Evgeny, Fei, Hamid, Harrie, Hendrik, Hendrike, Hosein, Isaac, Ilya, Jyothi, Katja, Katya, Ke, Marlies, Marzieh, Masrour, Mostafa, Nikos, Richard, Ridho, Shangsong, Tobias, Tom, and Xinyi. Thank you for discussions, reading groups, coffee breaks and countless evenings at Oerknal and De Polder. Thank you Christophe, David, Katja and Richard for sharing C3.258B. Adith, Aldo, Chuan, Evgeny, Hamid, Hosein, Marc, Richard, and Simon, we shared good times on the football courts. I would like to thank many other researchers in the Faculty of Science: Amir, Gang, Guangliang, Hao, Huan, Hui, Jiajia, Jun, Junchao, Masoud, Muhe, Ninghang, Que, Ran, Shuai, Songyu, Wei, Xiaolong, Xing, Yang, Yang, Yuan, Zijian, Zhenyang, Zhongcheng and Zhongyu, for their help as friends. I also owe my sincere gratitude to Petra and Caroline for helping me take care of countless practical details. I'm also thankful to Dr. Hans Henseler for his excellent support and management of our project meetings.

I have met and talked with many wonderful information retrieval researchers during conferences. I highly admire their work. Aixin, Chao, Chenliang, Damiano, Dawei, Jiepu, Jiyun, Laura, Liqiang, Liu, Ning, Shiri, Sicong, Weize, Xia, Xiangnan, Xirong, Yadong, Yulu, Yuxiao and Zhiyong, thank you for your discussions; your suggestions and feedback are quite valuable to me.

Many friends have helped me during my doctoral studies. It has been five years since I came to Europe from China. I thank all my friends in Amsterdam, Luxembourg and Saarbrücken for sharing life with me during this period. I'm thankful to Elsa, Fang, Hao, Lin, Liyan, Tony, Yuan, Yue, Yusi, Zhenzhen, Zhida and Zhiguang for creating such wonderful memories in Amsterdam. I'm thankful to Dalin, Jinghua, Marcela, Mike, Paul, Ran, Xin, Xuecan, Yang and Yiwen for our happy time in Luxembourg. He, Lizhen, Ran, Weijia, Yafang and Yu, thank you for having dinners together in Saarbrücken. I would also like to thank my good friends in China, the United States and Australia: Chaoran, Delei, Demin, Feng, Feng, Kai, Kang, Kun, Meng, Qiang, Shan, Shuai, Xiaoming, Zhen and Zhenyu, for their support and help.

Last, but not least, I would like to thank my parents, my grandparents and my cousins, for always supporting me spiritually throughout my studies. Special thanks go to my wife, Xiaoxiao, for her understanding, encouragement and love.

Contents

1 Introduction

With the rise of web 2.0, hundreds of millions of people are spending countless hours on social media. Defined as a group of Internet-based applications [105], social media, such as microblogs, community question-answering and web forums, provides information platforms to let people create, share, or exchange information, interests and their own viewpoints. Using social media, people can be connected anywhere and anytime, which also provides online channels to let people interact with each other. Social media has been changing our world, not only because of its timeliness and interactivity, but it also provides an ideal opportunity to observe human behavior through a new lens [265]. In recent years, *social media mining* [231, 265] has been proposed to investigate massive volumes of social media data that are being produced. Recent work on social media mining has used social media data to understand, analyze, represent and extract a range of actionable patterns [265]. Specifically, by mining social media data, we can extract bursty and salient topics [57, 160, 208], find people and groups [4], detect emergencies [57, 96, 202], and predict user behavior [45, 53, 58, 260].

A key characteristic of social media mining is the ambition to monitor the content of social media [231, 265], i.e., text from social media platforms, social relations among users, and changes in social media data over time. Monitoring text has been studied for quite a long time; indeed, it is a fundamental task in text mining [3]. Previous research on text mining has applied multiple methodologies to help people and machines understand text, e.g., document summarization [63, 165, 245] and text classification [86, 203]. Even though text understanding has become a well studied research problem, understanding social media documents remains a challenge. Social media documents are usually represented as part of a stream of documents, i.e., *social text streams* [192]. Social text streams come in various kinds, e.g., tweets from microblogs, emails from mailing lists, threads from web forums, updates from social media platforms, etc. But invariably, social media documents tend to be short, sparse, and more sensitive to the change of time than traditional news or web documents. In addition, language patterns in social text streams change with time, which leads to *topic drift* (the phenomenon that topics change over time), a serious challenge to understanding social media documents. Therefore, most existing text mining methods cannot be directly applied to understand social media data.

To understand social media text, recent work has explored various directions. Several methods aim at discovering latent patterns, e.g., topics, sentiments and viewpoints, from social media documents. Discovering topics from text has been at the core of topic detection and tracking (TDT) [10]. In recent years, topic modeling has been applied to detect and track topics from social media [65, 119, 157, 186, 273]. Focusing on understanding people's opinions from a document, sentiment analysis is another important task in understanding social media [157, 174]. Based on extracting latent patterns from social media documents, in recent years, summarization, classification and recommendation have been successfully applied to help people understand social media text. Unlike methods for generic text summarization, methods for social media summarization, such as tweets summarization [41], community question-answering summarization [229] and web forum summarization [189], need to tackle the shortness, timeliness and complicated social relations in social media. Research carried out in the area of social media mining has applied opinion summarization to understand opinions and viewpoints by summarizing opinionated documents into structured or semi-structured summaries [74, 75, 92, 108, 122]. Time-aware classification of social text streams [169] is attracting more and more attention recently. Unlike text classification for other kinds of documents, time-aware classification of social text streams has to deal with topic drift [56, 57, 169, 192]. Finally, with the development of social media, trusted social relations on many platforms, such as Yelp and TripAdvisor, have been shown to be effective in enhancing the performance of discovery and recommendation [43]; moreover, user comments from e-commerce platforms can improve the rating prediction and the interpretability of recommended results [137].

In this dissertation, we continue previous research on understanding social media documents along three lines: *summarization*, *classification* and *recommendation*. Our first line of work is the summarization of social media documents. Considering the task of time-aware tweets summarization, we first focus on the problem of selecting meaningful tweets given a user's interests and propose a dynamic latent factor model. Thereafter, given a set of opinionated documents, we address the task of summarizing contrastive themes by selecting meaningful sentences to represent contrastive themes in those documents. A viewpoint is a triple consisting of an entity, a topic related to this entity and sentiment towards this topic. In this thesis, we also propose the task of multi-viewpoint summarization of multilingual social text streams, by monitoring viewpoints for a running topic and selecting a small set of informative documents. Our second line of work concerns hierarchical multi-label classification. Hierarchical multi-label classification assigns a document to multiple hierarchical labels. Here, we focus on hierarchical multi-label classification of social text streams, in which we propose a structured learning framework to classify a short text from a social text stream to multiple classes from a predefined hierarchy. Based on a viewpoint extraction model that we propose as part of a multi-viewpoint summarization task, our third line of work applies a latent factor model for predicting item ratings that uses user opinions and social relations to generate explanations.

1.1 Research Outline and Questions

The broad question that motivates the research underlying this thesis is: *How can we understand social media documents?* Individual components for solving this problem already exist (see Chapter 2 for an overview), but other aspects, such as personalized time-aware tweets summarization, contrastive themes summarization, multi-viewpoint summarization, hierarchical multi-label classification and explainable recommendation have not yet been sufficiently investigated. This thesis aims to advance the state-of-the-art on all of those aspects and contribute new solutions to the field of social media monitoring. The work in this thesis focuses on developing methods for addressing the challenges raised in three general research themes described above: summarization, classification and recommendation of social media text.

For summarizing social media documents, in Chapter 3 we start out with our study by employing summarization approaches for selecting meaningful tweets given a user's personal interests, as previous work has found that text summarization is effective to help people understand an event or a topic on social media [41, 170, 208, 251]. Twitter has amassed over half a billion users, who produce ("tweet") over 300 million tweets per day. Twitter users can subscribe to updates from other users by following them, essentially forming a unidirectional friend relationship. Moreover, tweets can be "retweeted," basically copying a tweet posted by another user to one's own timeline. From an information retrieval point of view, the sheer volume of users and tweets presents interesting challenges. On the one hand, interesting, relevant, or meaningful tweets can easily be missed due to a large number of followed users. On the other hand, users may miss interesting tweets when none of the users they follow retweet an interesting piece of information. Tweets summarization aims at addressing this dual problem. However, how to adapt tweets summarization to a specific user is still a topic of ongoing research [179]. Moreover, previous work on tweets summarization neglects to explicitly model the temporal nature of the microblogging environment. Therefore, our research question in this first study is:

RQ1: How can we adapt tweets summarization to a specific user based on a user's history and collaborative social influences? Is it possible to explicitly model the temporal nature of a microblogging environment in personalized tweets summarization?

Multi-document summarization has become a well-studied research problem for helping people understand a set of documents. However, the web now holds a large number of opinionated documents, especially in opinion pieces, microblogs, question answering platforms and web forum threads. The growth in volume of such opinionated documents motivates the development of methods to facilitate the understanding of subjective viewpoints present in sets of documents. Given a set of opinionated documents, we define a *theme* to be a specific set of topics with an explicit sentiment opinion. Given a set of specific topics, two themes are *contrastive* if they are relevant to those topics, but opposing in terms of sentiment. The phenomenon of contrastive themes is widespread in opinionated web documents [59].

In Chapter 4, we focus on *contrastive summarization* [107, 176] of multiple themes. The task is similar to *opinion summarization*, in which opinionated documents are summarized into structured or semi-structured summaries [74, 75, 92, 108]. However, most existing opinion summarization strategies are not adequate for summarizing contrastive themes from a set of unstructured documents. To our knowledge, the most similar task in the literature is the *contrastive viewpoint summarization* task [176], where one extracts contrastive but relevant sentences to reflect contrastive topic aspects that are derived from a latent topic-aspect model [175]. However, previously proposed methods for *contrastive viewpoint summarization* neglect to explicitly model the number of topics and the relations among topics in contrastive topic modeling—these are two key features in contrastive theme modeling. The specific contrastive summarization task that we address is *contrastive theme summarization of multiple opinionated documents*. In our case, the output consists of contrastive sentence pairs that highlight every contrastive theme in the given documents. Regarding these two key features in contrastive theme modeling, we address the following question:

RQ2: How can we optimize the number of topics in contrastive theme summarization of multiple opinionated documents? How can we model the relations among topics in contrastive topic modeling? Can we find an approach to compress the themes into a diverse and salient subsets of themes?

In answering this question, we find that the definition of *viewpoint* in previous work [175, 176] neglects the importance of *entities* [158] in viewpoint modeling. Focused on an entity, in Chapter 5 we redefine a *viewpoint* to refer to a topic with a specific sentiment label. As an example, consider the entity "Japan" within the topic "#Whale hunting," with a negative sentiment. With the development of social media, we have witnessed a growth in the number of social media posts that expressing dynamically changing viewpoints in different languages around the same topic [178]. Unlike viewpoints in stationary documents, time-aware viewpoints of social text streams are dynamic, volatile and cross-linguistic [65]. Hence, the task we address is *time-aware multi-viewpoint summarization of multilingual social text streams*: we extract a set of informative social text documents to highlight the generation, propagation and drift process of viewpoints in a given social text stream.

The growth in volume of social text streams motivates the development of methods that facilitate the understanding of those viewpoints. Their multi-lingual character is currently motivating an increasing volume of information retrieval research of multilingual social text streams, in areas as diverse as reputation polarity estimation [178] and entity-driven content exploration [236]. Recent work confirms that viewpoint summarization is an effective way of assisting users to understand viewpoints in stationary documents [74, 77, 107, 127, 138, 157, 243]. However, viewpoint summarization in the context of multilingual social text streams has not been addressed yet. Compared with viewpoint summarization in stationary documents, the task of time-aware multi-viewpoint summarization of social text streams faces four challenges: (1) the ambiguity of entities in social text streams; (2) viewpoint drift, so that a viewpoint's statistical properties change over time; (3) multi-linguality, and (4) the shortness of social text streams. Therefore, existing approaches to viewpoint summarization cannot be directly applied to time-aware viewpoint summarization of social text streams. We ask the following question:

RQ3: How can we find an approach to help detect time-aware viewpoint drift? How can we detect viewpoints from multilingual social text streams? How can we generate summaries to reflect viewpoints of multi-lingual social text streams?

After our investigation into summarizing social media documents, we turn to classifying social text streams. Short text classification has been shown to be an effective way of assisting users in understanding documents in social text streams [141, 143, 169, 268]. Straightforward text classification methods, however, are not adequate for mining documents in social streams.

For many social media applications, a document in a social text stream usually belongs to multiple labels that are organized in a hierarchy. This phenomenon is widespread in web forums, question answering platforms, and microblogs [42]. Faced with many millions of documents every day, it is impossible to manually classify social streams into multiple hierarchical classes. This motivates the *hierarchical multi-label classification* (HMC) task for social text streams: classify a document from a social text stream using multiple labels that are organized in a hierarchy. Recently, significant progress has been made on the HMC task, see, e.g., [28, 34, 40]. However, the task has not yet been examined in the setting of social text streams. Compared to HMC on stationary documents, HMC on documents in social text streams faces specific challenges: (1) Because of topic drift, a document's statistical properties change over time, which makes the classification output different at different times. (2) The shortness of documents in social text streams hinders the classification process.Therefore, in Chapter 6 we address the HMC problem for documents in social text streams and provide an answer to the following question:

RQ4: Can we find a method to classify short text streams in a hierarchical multi-label classification setting? How should we tackle the *topic drift* and *shortness* in hierarchical multi-label classification of social text streams?

In our last step towards understanding social media, we turn to the problem of explainable recommendation on e-commerce portals, with the goal of generating so-called viewpoints by jointly analyzing user's reviews and trusted social relations. Many e-commerce sites, such as Yelp and TripAdvisor, have become popular social platforms that help users discuss and select items. Traditionally, an important strategy for predicting ratings in recommender systems is based on collaborative filtering (CF), which infers a user's preference using their previous interaction history. Since CF-based methods only use (previous) numerical ratings as input, they suffer from the "cold-start" problem and from the problem of unexplainable prediction results [89, 137], a topic that has received increased attention in recent years.

Explainable recommendation has been proposed to address the "cold-start" problem and the poor interpretability of recommended results by not only predicting better rating results, but also generating item aspects that attract user attention [271]. Most existing methods on explainable recommendation apply topic models to analyze user reviews to provide descriptions along with the recommendations they produce. To improve the rating prediction for explainable recommendations, in Chapter 7, our focus is on developing methods to generate so-called viewpoints by jointly analyzing user reviews and trusted social relations. Compared to "topics" in previous explainable recommendation strategies [32, 242], viewpoints, as we discussed in previous chapters, contain more useful information that can be used to understand and predict user ratings in recommendation task. We assume that each item and user in a recommender system can be represented as a finite mixture of viewpoints. Furthermore, each user's viewpoints can be influenced by their trusted social friends. Our question in this study, then, is:

RQ5: Can we find an approach to enhance the rating prediction in explainable recommendation? Can user reviews and trusted social relations help explainable recommendation? What are factors that could affect the explainable recommendations?

We seek answers to the five questions listed in five research chapters (Chapters 3–7). We record our answers in the discussion and conclusion sections of each individual chapter and in Chapter 8 we bring our answers together to summarize our findings.

In the next sections we list the contributions that this thesis makes to the field and we give an overview of the thesis and of the origins of the material.

1.2 Main Contributions

This thesis contributes at different levels: we provide new *task scenarios*, new *models and algorithms*, and new *analyses*. Our main contributions are listed below.

Task Scenarios

Personalized time-aware tweets summarization We propose the task of personalized time-aware tweets summarization, selecting personalized meaningful tweets from a collection of tweets. Unlike traditional summarization approaches that do not cover the evolution of a specific event, we focus on the problem of selecting meaningful tweets given a split of a user's history into time periods and collaborative social influences from "social circles."

Contrastive theme summarization We address the task of summarizing contrastive themes: given a set of opinionated documents, select meaningful sentences to represent contrastive themes present in those documents. Our unsupervised learning scenario for this task has three core ingredients: contrastive theme modeling, diverse theme extraction, and contrastive theme summarization.

Time-aware multi-viewpoint summarization of multilingual social text streams We propose the task of time-aware multi-viewpoint summarization of multilingual social text streams, in which one monitors viewpoints for a running topic from multilingual social text streams and selects a small set of informative social texts. The scenario includes three core ingredients: dynamic viewpoint modeling, cross-language viewpoint alignment, and, finally, multi-viewpoint summarization.

Hierarchical multi-label classification of social text streams We present the task of hierarchical multi-label classification for streaming short texts, in which we classify a document from a social text stream using multiple labels that are organized in a hierarchy. Our scenario includes three core ingredients: short document expansion, time-aware topic modeling, and chunk-based structural classification.

Models and Algorithms

An effective approach for personalized time-aware tweets summarization We propose a time-aware user behavior model, the Tweet Propagation Model (TPM), in which we infer dynamic probabilistic distributions over interests and topics. We then explicitly consider novelty, coverage, and diversity to arrive at an iterative optimization algorithm for selecting tweets.

Non-parametric models for contrastive theme modeling We present a hierarchical non-parametric model to describe hierarchical relations among topics; this model is used to infer threads of topics as themes from a nested Chinese restaurant process. We enhance the diversity of themes by using structured determinantal point processes for selecting a set of diverse themes with high quality.

An effective approach to track dynamic viewpoints from text streams We propose a dynamic latent factor model to explicitly characterize a set of viewpoints through which entities, topics and sentiment labels during a time interval are derived jointly; we connect viewpoints in different languages by using an entity-based semantic similarity measure; and we employ an update viewpoint summarization strategy to generate a time-aware summary to reflect viewpoints.

A structured learning algorithm for hierarchical multi-label classification Based on a structural learning framework, we transform our hierarchical multi-label classification problem into a chunk-based classification problem via multiple structural classifiers.

Social collaborative viewpoint regression for explainable recommendations We propose a latent factor model, called social collaborative viewpoint regression (sCVR), for predicting item ratings that uses user opinions and social relations generate explanations. To this end we use viewpoints from both user reviews and trusted social relations. Our method includes two core ingredients: inferring viewpoints and predicting user ratings. We apply a Gibbs EM sampler to infer posterior distributions for sCVR.

Analyses

An analysis of the effectiveness of summarization methods on social media We provide a detailed analysis of the effectiveness of document summarization approaches for each summarization task in this thesis. We compare those summarization methods with our own strategies in each task, and provide an extensive discussion of the advantages and disadvantages of those methods on our datasets.

An analysis of social media summarization outcomes We identify factors that affect the performance on each of the summarization tasks that we consider. For the personalized time-aware tweets summarization task time periods and social circles matter. Our analysis provides insights in the importance and impact of these dual factors. For the contrastive theme summarization, several factors play a role in our proposed summarization method. To determine the contribution of *contrast*, *diversity* and *relevance*, we provide an analysis to show the impact of those factors in contrastive summarization. For the multi-viewpoint summarization, our analysis provides the impact of each algorithmic step, and we identify the effect of *novelty* and *coverage* in summarization.

An analysis of hierarchical multi-label classification outcomes For each step in our method for hierarchical multi-label classification of social text streams, we evaluate its effectiveness. By comparing with existing work on hierarchical multi-label classification, we analyze the overall effectiveness of our own method. We also identify several factors that impact the classification results, namely, shortness of document, topic drift and number of items, and provide an extensive analysis of the impact of those factors in hierarchical multi-label classification.

An analysis of social relations and user reviews in recommendation Compared to previous work on explainable recommendation, we identify two main differences in our method: viewpoints from user reviews and influences from trusted social relations. We evaluate each factor's impact for the performance of explainable recommendation. We discuss the explainability of recommendation by analyzing outcomes of social collaborative viewpoint regression.

1.3 Thesis Overview

This thesis is organized in eight chapters. After a background chapter, we present five research chapters containing our core contributions plus a concluding chapter:

Chapter 2—Background Here, we present the background for all subsequent chapters. We place our research in the broader context of information retrieval and text mining. After a brief outline of the field, and of social media mining in particular, we review the document summarization, text classification, recommendations and topic modeling literature.

Chapter 3—Personalized time-aware tweets summarization We focus on the problem of selecting meaningful tweets given a user's interests. We consider the task of time-aware tweets summarization, based on a user's history and collaborative social influences from "social circles." We propose a time-aware user behavior model, the Tweet Propagation Model (TPM), in which we infer dynamic probabilistic distributions over interests and topics. We then explicitly consider novelty, coverage, and diversity to arrive at an iterative optimization algorithm for selecting tweets. Experimental results validate the effectiveness of our personalized time-aware tweets summarization method based on TPM.

Chapter 4—Contrastive theme summarization We address the task of summarizing contrastive themes: given a set of opinionated documents, select meaningful sentences to represent contrastive themes present in those documents. We present a hierarchical non-parametric model to describe hierarchical relations among topics; this model is used to infer threads of topics as themes from a nested Chinese restaurant process. We enhance the diversity of themes by using structured determinantal point processes for selecting a set of diverse themes with high quality. Finally, we pair contrastive themes and employ an iterative optimization algorithm to select sentences, explicitly considering contrast, relevance, and diversity. Experiments on three datasets demonstrate the effectiveness of our method.

Chapter 5—Multi-viewpoint summarization of multilingual social text streams We focus on time-aware multi-viewpoint summarization of multilingual social text streams. We propose a dynamic latent factor model to explicitly characterize a set of viewpoints through which entities, topics and sentiment labels during a time interval are derived jointly; we connect viewpoints in different languages by using an entity-based semantic similarity measure; and we employ an update viewpoint summarization strategy to generate a time-aware summary to reflect viewpoints. Experiments conducted on a real-world dataset demonstrate the effectiveness of our proposed method for time-aware multi-viewpoint summarization of multilingual social text streams.

Chapter 6—Hierarchical multi-label classification of social text streams We focus on hierarchical multi-label classification of social text streams. We extend each short document in social text streams to a more comprehensive representation via state-of-the-art entity linking and sentence ranking strategies. From documents extended in this manner, we infer dynamic probabilistic distributions over topics by dividing topics into dynamic "global" topics and "local" topics. For the third and final phase we propose a chunk-based structural optimization strategy to classify each document into multiple classes. Extensive experiments conducted on a large real-world dataset show the effectiveness of our proposed method for hierarchical multi-label classification of social text streams.

Chapter 7—Social collaborative viewpoint regression We propose a latent variable model, called social collaborative viewpoint regression (sCVR), for predicting item ratings that uses user opinions and social relations generate explanations. To this end we use so-called viewpoints from both user reviews and trusted social relations. Our method includes two core ingredients: inferring viewpoints and predicting user ratings. We apply a Gibbs EM sampler to infer posterior distributions of sCVR. Experiments conducted on three large benchmark datasets show the effectiveness of our proposed method for predicting item ratings and for generating explanations.

Chapter 8—Conclusions We summarize our main findings and point out directions for future research.

1.4 Origins

For each research chapter we list on which publication(s) it is based, and we briefly discuss the role of the co-authors.

Chapter 3. This chapter is based on Ren, Liang, Meij, and de Rijke [190] "Personalized time-aware tweets summarization," *Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval*. ACM, 2013. The scope and the design of the algorithm and experiments were mostly due to Ren. Liang and Meij contributed to the experiment. All authors contributed to the text.

Chapter 4. This chapter is based on Ren and de Rijke [188] "Summarizing contrastive themes via hierarchical non-parametric processes." *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. ACM,

2015. The design of the algorithm and the experiments were due to by Ren. All authors contributed to the text.

Chapter 5. This chapter is based on Ren, Inel, Aroyo, and de Rijke [193] "Time-aware multi-viewpoint summarization of multilingual social text streams," *Proceedings of the 25th ACM international conference on information and knowledge management*. ACM, 2016. The scope and the design of the algorithm and experiment were mostly due to Ren. All authors contributed to the text.

Chapter 6. This chapter is based on Ren, Peetz, Liang, van Dolen, and de Rijke [192] "Hierarchical multi-label classification of social text streams," *Proceedings of the 37th international ACM SIGIR conference on research and development in information retrieval*. ACM, 2014. Van Dolen contributed to the experimental setup. The scope and design of the algorithm was mostly developed by Ren. All authors contributed to the text.

Chapter 7. This chapter is based on Ren, Liang, Li, Wang, and de Rijke [194] "Social collaborative viewpoint regression for explainable recommendations," *under review*, 2016. The scope and design of the algorithm was mostly developed by Ren. Liang and Wang contributed to the design of algorithm. All authors contributed to the text.

Work on other publications also contributed to the thesis, albeit indirectly. We mention nine papers:

- van Dijk, Graus, Ren, Henseler, and de Rijke [234], "Who is involved? Semantic search for e-discovery," *Proceedings of the 15th international conference on artificial intelligence & law*, 2015.
- Graus, Ren, de Rijke, van Dijk, Henseler, and van der Knaap [82], "Semantic search in e-discovery: An interdisciplinary approach," *ICAIL 2013 workshop on standards for using predictive coding, machine learning, and other advanced search and review methods in e-discovery*, 2013.
- Liang, Ren, and de Rijke [130] "The impact of semantic document expansion on cluster-based fusion for microblog search," *Advances in information retrieval. Proceedings of the 36th european conference on IR research.* Springer, 2014.
- Liang, Ren, and de Rijke [129] "Fusion helps diversification," *Proceedings of the 37th international ACM SIGIR conference on research and development in information retrieval.* ACM, 2014.
- Liang, Ren, and de Rijke [131] "Personalized search result diversification via structured learning," *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining.* ACM, 2014.
- Liang, Ren, Weerkamp, Meij, and de Rijke [132] "Time-aware rank aggregation for microblog search," *Proceedings of the 23rd ACM international conference on conference on information and knowledge management.* ACM, 2014.
- Ren, Ma, Wang, and Liu [189] "Summarizing web forum threads based on a latent topic propagation process," *Proceedings of the 20th ACM international conference on information and knowledge management.* ACM, 2011.

- Ren, van Dijk, Graus, van der Knaap, Henseler, and de Rijke [191] "Semantic linking and contextualization for social forensic text analysis," *Proceedings of european intelligence and security informatics conference (EISIC)*. IEEE, 2013.
- Zhao, Liang, Ren, Ma, Yilmaz, and de Rijke [274] "Explainable user clustering in short text streams," *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval.* ACM, 2016.

2 Background

In this chapter, we provide the concepts and background needed in later chapters in this thesis. We start with a brief introduction to social media in Section 2.1, in which we focus on information retrieval in social media. We study the overall task that we address in this thesis, i.e., monitoring social media, from three angles: summarization, classification, and recommendation. Thus, in Section 2.2 we detail previous work on summarization to prepare for Chapters 3–5. Specifically, Section 2.2.2 surveys background material on multi-document summarization. Because our proposed summarization strategies of social media rely on update summarization algorithms, we discuss related work on update summarization in Section 2.2.3. In Section 2.2.4 we describe related work on tweets summarization, which is the subject of Chapter 3. The contrastive theme summarization and the viewpoint summarization algorithms proposed in Chapter 4 and 5, respectively, work with opinion summarization; thus we also recall previous work for sentiment analysis in Section 2.2.5. And then, in Section 2.3, we discuss background knowledge on text classification, which is the subject of Chapter 6. Specifically, our proposed hierarchical multi-label classification of social text streams in Chapter 6 utilizes short text classification and hierarchical multi-label classification algorithms; relevant methods are described in Section 2.3.2 and Section 2.3.3, respectively. In Section 2.4, we provide background for our work on recommendation. For the task of explainable recommendation in Chapter 7, we provide background material on collaborative filtering and explainable recommendations in Section 2.4.2 and Section 2.4.3, respectively.

Finally, we detail preliminaries of machine learning methods that are used in thesis. Our proposed algorithms in Chapters 3–7 work with latent topic modeling; thus we recall methods for topic modeling in Section 2.5. Section 2.6 surveys background material on the determinantal point process, which is applied in Chapter 4. We introduce structured learning methods in Section 2.7 for our proposed chunk-based structured learning algorithm in Chapter 6.

2.1 Social Media

In this section, we describe relevant research on social media. We start with a general overview of social media and then zoom in on information retrieval for social media.

2.1.1 Overview

Social media refers to websites and applications that enable users to create and share content or participate in social networking [177]. Those websites and applications include personal blogs, microblogs, web forums, community question-answering, mailing lists, and many websites with social networking services. In day-to-day language, social media also refers to social networking sites such as Facebook, G+, and LinkedIn. An increasing number of e-commerce portals and traditional newspapers, such as Yelp,[1] TripAdvisor,[2] and the New York Times,[3] have begun to provide social media services. For example, on the New York Times website, users can share, comment, and discuss each article.

Social media has been broadly defined to include widely accessible electronic tools that enable anyone to publish, access, and propagate information. An important feature of social media is social networking. According to Maslow's hierarchy of needs [153], humans need to feel a sense of belonging and acceptance among their social communities. This primary need drives the success of social media in recent years.

According to Aichner and Jacob [8], social media can be divided into eight kinds: (1) blogs; (2) microblogs; (3) e-commerce portals; (4) multimedia sharing; (5) social networks; (6) review platforms; (7) social gaming; and (8) virtual worlds. Unlike traditional media, social media documents have unique features in many aspects:

- *Shortness*: Most social media documents are shorter than documents in traditional media, e.g., in Twitter, there is a 140 character limit to the length of a tweet [190]. Compared to long documents, traditional text mining methods usually cannot successfully be applied directly to analyze social media documents.
- *Multilinguality*: With the development of social media, people using different languages are involved in the same communication platform. E.g., during global sports events such as FIFA Worldcup 2014, people discuss the same match in multiple languages on Twitter.
- *Opinions*: Social media holds a large number of opinionated documents, especially in opinion pieces, microblogs, question answering platforms and web forum threads. Thus, understanding opinions and sentiment analysis become increasingly important for content analysis in social media.
- *Timeliness*: Social media documents are posted with specific timestamps. The dynamic nature of social media makes text in social media quite different from text in traditional, more static collections. *Topic drift* and *viewpoint drift* can be found in social text streams. Because of such phenomena, the statistical properties of social media text streams change over time.

2.1.2 Information retrieval in social media

Information retrieval (IR) is about finding material of an unstructured nature that satisfies an information need within large collections [150]. According to Baeza-Yates and Ribeiro-Neto [20], information retrieval deals with the representation, storage, organi-

[1] http://www.yelp.com
[2] http://tripadvisor.com
[3] http://www.nytimes.com

zation of, and access to information items. A lot of system-oriented early IR research, from the 1950s in which the term IR was proposed by Mooers [164] until the early 1990s, focuses on *boolean retrieval models* [104], *vector space retrieval models* [204], and *probabilistic retrieval models* [152, 197].

Specifically, Boolean retrieval models are the basic retrieval models, where the input query is represented as a Boolean expression of terms, and relevance of a document to a query is binary. To tackle the disadvantages of Boolean retrieval models, researchers proposed a second generation of retrieval models, i.e., *vector space models* [204], where the "bag of words" representation is introduced. Such models tend to neglect the dependence between adjacent terms, so that context-aware information is lost in the representation. Furthermore, weighting of terms or documents in vector space models is intuitive but not always formally justified [128]. Therefore, *probabilistic retrieval models* were proposed by Maron and Kuhns [152] and Robertson and Jones [197]. Probabilistic retrieval modeling is the use of a model that ranks documents in decreasing order of their probability of relevance to a user's information needs [51]. In probabilistic retrieval models, the probability of relevance of a document to a query is set to depend on the query and document representations. With the availability of a large number of ranking functions came the need to combine their outcomes, in the late 1980s the idea of learning to rank was introduced [72]. From the late 1990s, lots of IR research focuses on *learning to rank* [103], *language models* [182], and *text mining* [3, 90, 99, 102]: With the development of machine learning, many supervised learning methods have been applied to optimize the ranking of documents, which are called *learning to rank* models [103].

In the meantime, with the emergence of the World Wide Web in the 1990s, the field of information retrieval changed in important ways [177]. Search has to be open to everyone who can access the web. And the scale of the data used in IR has changed dramatically. In parallel, another important development occurred: since 1992, the Text REtrieval Conference (TREC) [88] has been set up to support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies.

The web gave rise to a large number of ranking methods, such as PageRank [37] and HITS [111], that exploit the special nature of the web and of web pages. Instead of overtly modeling the probability of relevance of a document to a query, *language models* [182] model the idea that a document is a good match to a query if the document model is likely to generate the query, which will in turn happen if the document contains the query words often [150]. Because of the large volume of data in current information retrieval tasks, text mining in IR has received an increase number of attention [3, 90, 99, 102]. In information retrieval, text mining [3] refers to a family of techniques oriented to the study of deriving high-quality information from texts. Early text mining tasks considered in IR include text summarization, text classification, text clustering, concept extraction, sentiment analysis, and entity modeling [90, 216, 217, 241, 258, 278].

In recent years, information retrieval has been successfully applied to social media. Information retrieval in social media needs to consider the specific features of social media documents and network structure, and adjust the formulation for their research problems. Generally, IR work on social media can be divided into the following groups:

Retrieval in social media Because of the dynamic nature of social media documents, *topic drift* happens, i.e., topic distributions change over time. Thus, in the task of rankings of documents in social media, the relevance of a social media document to a query may change over time. Recently, dynamic retrieval tasks, such as microblog search [135, 173, 218] and temporal summarization [18, 19], have been tackled as tracks within TREC. In the TREC microblog track, the task can be summarized as: at time t, participants are asked to find tweets that are relevant to a query q, and rank relevant tweets by time [218]. Since the launch of the microblog track, several strategies have been proposed for microblog retrieval, many of them using temporal information related to microblogs [13, 277]. Zhang et al. [269] apply a combination method by taking the frequency of a query term in various microblogs into account with query expansion. Luo et al. [147] apply a learning to rank method by considering meta data as block features in the microblog search.

The temporal summarization track has been proposed to develop systems for efficiently monitoring the information associated with an event over time [18, 19]. Specifically, it is aimed at developing systems that can broadcast short, relevant, and reliable sentence length updates about a developing event. Following the idea of temporal summarization, Guo et al. [85] focus on updating users about time critical news events. McCreadie et al. [156] apply a regression model to tackle the incremental summarization for events.

Information diffusion in social media Understanding the propagation of information in social media communities is another crucial topic [255]. Research about information diffusion in social media can be divided to discrete-time diffusion and continuous-time diffusion. Early research focuses on discrete-time diffusion in social communities [1, 80, 123, 255]. Adar and Adamic [1] formulated diffusion as a supervised classification problem and used support vector machines combined with rich textual features to predict the occurrence of individual links. Because choosing the best set of edges maximizing the likelihood of the data is NP-hard, Gomez Rodriguez et al. [80] propose an efficient approximate algorithm for inferring a near-optimal set of directed edges. For the continuous-time setting, several authors estimate the expected number of followups a set of nodes can trigger in a time window [48, 60, 80, 81, 172, 232, 256]. Cheng et al. [49] examine the problem of predicting the growth of retweeting behavior over social communities. Gao et al. [76] focus on retweeting dynamics and predict the future popularity of given tweets by proposing an extended reinforced Poisson process model with time mapping process. Based on the influence estimation problem, the influence maximization problem is proposed where one needs to search a set of nodes whose initial adoptions of a contagion can trigger, within a given time window, the largest expected number of follow-ups [81]. Focusing on this problem, Rodriguez and Schölkopf [198] propose an efficient approximate algorithm by exploiting a natural diminishing returns property.

Monitoring social media Monitoring social media refers to a continuous systematic observation and analysis of social media communities [66]. Because of social media features that we described at the beginning of this section, monitoring social media is a challenging problem. To tackle this problem, in recent years, more and more researchers start to apply text mining methods from IR to monitor social media documents. Many tasks can be found, including understanding content of social media [62, 169, 170, 209, 215, 224, 229, 251, 259] and predicting user behavior on social media [91, 98, 244, 250]. To help understand social media content, social media summarization, clustering, and classification have been tackled using a range of approaches. The shortness of documents hinders the effectiveness of many widely used text mining methods when working with social media. Focusing on short text processing in social media, Efron et al. [62] propose a document expansion method to extend short texts to long text. Knowledge-based semantic document expansion methods have also been proved effective in social media text processing [82, 190, 191]. Liang et al. [130] integrate semantic document expansion to increase the contribution of the clustering information in cluster-based fusion for microblog search. Using word co-occurrence patterns to replace unigram semantic units in topic learning, the biterm topic model (BTM) tackles the shortness problem in short text processing [253]. Inspired by BTM, Zhao et al. [274] propose a dynamic user behavior model for user clustering of social text streams. Opinion mining is another crucial topic in social media monitoring [115]. To analyze opinionated documents in social media, Liu et al. [140] propose a smoothed language model to combine manually labeled data and noisy labeled data. Moreover, online reputation management in social media has been tackled as an evaluation exercise activity, i.e., RepLab [14–16]. Based on the RepLab 2012 and 2013 datasets, Peetz et al. [178] automatically determine the reputation polarity of a tweet by using features based on three dimensions: the source of the tweet, the contents of the tweet and the reception of the tweet.

Another important task in monitoring social media is collective user behavior modeling [23, 98]. In recent years, this task has received an increasing amount of attention [23, 98, 250, 262]. Several approaches have been proposed for the recommendation task in social media: Yang et al. [256] address recommendation and link prediction tasks based on a joint-propagation model, FTP, between social friendship and interests. Ye et al. [260] propose a generative model to describe users' behavior, given influences from social communities, for recommendation [148, 149]. Chen et al. [45] propose a collaborative filtering method to generate personalized recommendations in Twitter through a collaborative ranking procedure.

In this dissertation, our focus relates to monitoring social media. To answer the research questions listed in Chapter 1, we use three angles: summarization, classification and recommendation. As we work on automatic text summarization of social media documents in Chapters 3–5, we provide brief overviews of multi-document summarization (Section 2.2.2), update summarization (in Section 2.2.3), tweets summarization (Section 2.2.4), and opinion summarization (Section 2.2.5). As we work on hierarchical multi-label classification in Chapter 6, we provide a brief overview of short text classification (Section 2.3.2) and hierarchical multi-label classification (Section 2.3.3). And as background for our work on explainable recommendation in Chapter 7, we provide an overview of collaborative filtering (Section 2.4.2) and explainable recommendation (Section 2.4.3). Because we utilize latent factor modeling, determinantal point processes, and structured learning for social media monitoring, we introduce the background of these methods in Sections 2.5, 2.6, and 2.7.

2.2 Automatic Text Summarization

2.2.1 Overview

A text summarization system takes one or more documents as input and attempts to produce a concise and fluent summary of the most important information in the input [165]. In the 1950s, automatic text summarization was proposed by Luhn [146] with a term frequency based strategy. With the development of the World Wide Web, billions of web documents make text summarization much more important. In recent years, numerous summarization approaches have been proposed to digest news articles [52, 134, 230], text streams [156, 252], community question-answering [229], microblogs [159], and opinionated documents [74, 93, 176].

Text summarization approches can be divided into two classes: *extractive summarization* and *abstractive summarization*. Methods for extractive summarization select keywords or sentences from candidate documents to form the summary, whereas methods for abstractive summarization apply natural language generation to build an internal semantic representation for candidate documents. In this dissertation, our research mainly focuses on extractive summarization.

Early work in text summarization focused on the *single document summarization* task where the input is only one document. As research progressed, large redundancy on the web motivated research on *multi-document summarization* where the digest is generated from multiple similar but different documents. Based on multi-document summarization, *update summarization*, *tweets summarization*, and *opinion summarization* have been proposed. As we tackle automatic text summarization tasks for social media documents in Chapter 3–5, in this section, we provide background material on multi-document summarization, update summarization, tweets summarization, and opinion summarization.

2.2.2 Multi-document summarization

Multi-document summarization (MDS) is useful since it is able to provide a brief digest of large numbers of relevant documents on the same topic [165]. Most existing work on MDS is based on the extractive format, where the target is to extract salient sentences to construct a summary. Both unsupervised and supervised based learning strategies have received lots of attention. One of the most widely used unsupervised strategies is clustering with respect to the centroid of the sentences within a given set of documents; this idea has been applied by NeATS [134] and MEAD [184]. Many other recent publications on MDS employ graph-based ranking methods [63]. Wan and Yang [241] propose a theme-cluster strategy based on conditional Markov random walks. Similar methods are also applied in [245] for a query-based MDS task. Celikyilmaz and Hakkani-Tur [39] consider the summarization task as a supervised prediction problem based on a two-step hybrid generative model, whereas the Pythy summarization system [230] learns a log-linear sentence ranking model by combining a set of semantic features. As to discriminative models, CRF-based algorithms [211] and structured SVM-based classifiers [125] have proved to be effective in extractive document summarization. Learning to rank models have also been employed to query-based MDS [210] and to topic-focused MDS [279]. In recent years, with the development of social media, multi-document summarization is also being applied to social documents, e.g., tweets, weibos, and Facebook posts [41, 61, 167, 189, 190].

2.2.3 Update summarization

Traditional document summarization is retrospective in nature. *Update summarization* [11] is becoming a popular task in MDS research [165]; for this task one follows a stream of documents over time and extracts and synthesizes novel information in a collection of documents on what is new compared to what has been summarized previously [54, 156, 167, 215]. Given a *base* collection that users have already read and another *update* collection of recent documents, the goal of update summarization is to generate an update summary by analyzing the novelty, contrast and prevalence. An intuitive solution to update summarization is to remove redundancy from the output generated by a multi-document summarizer [70]. Yan et al. [252] propose an evolutionary timeline summarization strategy based on dynamic programming. Wan [240] propose a co-ranking algorithm to optimize a trade-off strategy between novelty and relevance metrics. McCreadie et al. [156] propose a pair-wise learning to rank algorithm to produce an update summary. They also train a regression model to predict the novelty of the given documents in each time period.

2.2.4 Tweets summarization

Several publications have focused on tweets summarization: the task of selecting a list of meaningful tweets that are most representative for some topic. Most work in the literature concerns tweets as basic constituents to compose a summary. Some authors bring feature-based or graph-based summarization technologies to bear on this task [170, 209], while other methods use a term-frequency based method [224] or a strategy based on mutual reinforcement between users' influence and qualifications of tweets [251].

Recently, time-aware summarization has been studied by several authors, often in the form of timeline generation on Twitter. Chakrabarti and Punera [41] separate topic related tweets into various periods as an event evolution map, and generate an update-summarization result. Evolutionary summarization approaches segment post streams into event chains and select tweets from various chains to generate a tweet summary; Nichols et al. [167] propose an effective method to separate timelines using Twitter. To the best of our knowledge, existing work on tweets summarization focuses on the extraction of representative tweets for specific topics, without considering personalization.

Other work integrates the task of selecting tweets with other web documents: Yang et al. [259] use mutual reinforcement to train both the selection of related web documents and tweets via a single graph factor model. Zhao et al. [272] extract representative keywords from tweets based on a topic model. Tweet ranking has also attracted attention: Weng et al. [247] proposed a graph-based ranking strategy for ranking tweets based on the author-topic model.

2.2.5 Opinion summarization

In recent years, *sentiment analysis* has received a lot of attention. As a fundamental task in sentiment analysis, *opinion summarization* [92] is crucial to understand user generated content in product reviews. Opinion summarization generates structured [92, 124, 145, 157] or semi-structured summaries [75, 93, 109] given opinionated documents as input. Given opinionated documents, a structured opinion summary shows positive/negative opinion polarities. Semi-structured opinion summarization extracts sentences to describe opinion polarities. Hu and Liu [93] apply a sentence ranking approach based on the dominant sentiment according to polarity. Kim et al. [109] propose a method to extract explanatory sentences as opinion summary. Ganesan et al. [75] propose an unsupervised method to generate a concise summary to reflect opinions. Other relevant work for the contrastive summarization has been published by Lerman and McDonald [122] and Paul et al. [176]. Lerman and McDonald [122] propose an approach to extract representative contrastive descriptions from product reviews. A joint model between sentiment mining and topic modeling is applied in [176]. Opinosis [74] generates a summary from redundant data sources. Similarly, a graph-based multi-sentence compression approach has been proposed in [67]. Meng et al. [159] propose an entity-centric topic-based opinion summarization framework, which is aimed at generating summaries with respect to topics and opinions.

2.3 Text Classification

2.3.1 Overview

Given input documents and pre-defined classes, the target of *text classification* is to classify each document to one or more classes. As a traditional task in text mining and machine learning [3, 30, 71], text classification has received quite lot of attention. Distinguished by the formulation of the labeling results, text classification can be divided into binary classification, multi-class classification, and multi-label classification [71]. For traditional long documents, binary text classification and multi-class text classification, as a basic machine learning task, have already become two well-studied research problem [30, 71]. In recent years, the growth in volume of social media text drives lots of research interest on short text classification [35, 46, 261], especially for text classification of social text streams [169]. Another challenging research task in text classification is *hierarchical multi-label classification* (HMC) [112], which is to classify a document using multiple labels that are organized in a hierarchy.

In Chapter 6, we address the HMC task for social text streams. Thus in this section, we discuss a selection of influential approaches proposed in the literature, on both short text classification (in Section 2.3.2) and hierarchical multi-label classification (in Section 2.3.3).

2.3.2 Short text classification

In recent years, short text classification has received considerable attention. Most previous work in the literature addresses the sparseness challenge by extending short texts using external knowledge. Those techniques can be classified into web search-based methods and topic-based ones.

Web search-based methods handle each short text as a query to a search engine, and then improve short text classification performance using external knowledge extracted from web search engine results [35, 261]. Such approaches face efficiency and scalability challenges, which makes them ill-suited for use in our data-rich setting [46]. In a different way, several other works haves been proposed via collecting a large-scale corpus to enhance the classification performance [46, 181, 220, 266].

As to topic-based techniques, Phan et al. [181] extract topic distributions from a Wikipedia dump based on the LDA [32] model. Similarly, Chen et al. [46] propose an optimized algorithm for extracting multiple granularities of latent topics from a large-scale external training set; see [220] for a similar method.

Besides those two strategies, other methods have also been employed. E.g., Sun [222] and Nishida et al. [168] improve classification performance by compressing shorts text into entities. Zhang et al. [268] learn a short text classifier by connecting what they call the "information path," which exploits the fact that some instances of test documents are likely to share common discriminative terms with the training set. Few previous publications on short text classification consider a streaming setting; none focuses on a hierarchical multiple-label version of the short text classification problem.

2.3.3 Hierarchical multi-label classification

In machine learning, *multi-label classification* problems have received lots of attention. Discriminative ranking methods have been proposed in [207], while label-dependencies are applied to optimize the classification results by [86, 113, 180]. However, none of them can work when labels are organized hierarchically.

The *hierarchical* multi-label classification problem is to classify a given document into multiple labels that are organized as a hierarchy. Koller and Sahami [112] propose a method using Bayesian classifiers to distinguish labels; a similar approach uses a Bayesian network to infer the posterior distributions over labels after training multiple classifiers [21]. As a more direct approach to the HMC task, Rousu et al. [200] propose a large margin method, where a dynamic programming algorithm is applied to calculate the maximum structural margin for output classes. Decision-tree based optimization has also been applied to the HMC task [34, 237]. Cesa-Bianchi et al. [40] develop a classification method using hierarchical SVM, where SVM learning is applied to a node if and only if this node's parent has been labeled as positive. Bi and Kwok [28] reformulate the "tree-" and "DAG-" hierarchical multi-label classification tasks as problems of finding the best subgraph in a tree and DAG structure, by developing an approach based on kernel density estimation and the condensing sort and select algorithm.

To the best of our knowledge there is no previous work on HMC for short documents in social text streams. In Chapter 6 we present a chunk-based structural learning method for the HMC task, which is different from existing HMC approaches, and which we show to be effective for both the traditional stationary case and the streaming case.

2.4 Recommender Systems

2.4.1 Overview

Recommender systems are playing an increasingly important role in e-commerce portals. Typically, the task of recommender systems, or recommendation, is to aggregate and direct input items to appropriate users [79, 195]. Formally, given a set of users, $\mathcal{U}$, and a set of candidate items, $\mathcal{V}$, during recommendation we need to learn a function f, i.e., $f : \mathcal{U} \times \mathcal{V} \rightarrow \mathcal{R}$, where $\mathcal{R}$ indicates the ratings set between users and items. Thus, given each user $u \in \mathcal{U}$, the target of the recommendation process is to find a proper item $v \in \mathcal{V}$, so that:

$$v = \arg\max_{v' \in \mathcal{V}} f(u, v'), \tag{2.1}$$

Approaches for recommender systems can be divided into *content-based recommendation* and *collaborative filtering* (CF) [2, 214]. The task of content-based recommendation is to recommend items that are similar to the ones the user preferred in the past, whereas collaborative filtering is based on the core assumption that users who have expressed similar interests in the past will share common interests in the future [79]. Recently, significant progress has been made in collaborative filtering [22, 114, 121, 163, 206]. However, since CF-based methods only use numerical ratings as input, they suffer from a "cold-start" problem and unexplainable prediction results [89, 137], a topic that has received increased attention in recent years. Explainable recommendation has been proposed to address the "cold-start" problem and the poor interpretability of recommended results by not only predicting better rating results, but also generating item aspects that attract user attention [271]. We propose an explainable recommendation approach in Chapter 7. Thus in this section, we discuss the background knowledge about collaborative filtering (Section 2.4.2) and previous work on explainable recommendation (Section 2.4.3).

2.4.2 Collaborative filtering

In recent years, collaborative filtering based techniques have received considerable attention. Unlike content-based filtering strategies [144] that predict ratings using the analysis of user profiles, collaborative filtering [221] technologies, divided into memory-based collaborative filtering and model-based collaborative filtering, make rating predictions via user-item ratings matrices. Early collaborative filtering methods apply memory-based techniques. The most widely used memory-based collaborative filtering methods include the nearest neighbor approach [22], user-based methods [196] and item-based methods [206]. Among the model-based collaborative filtering methods, latent factor models [114] have become very popular as they show state-of-the-art performance on multiple datasets. Aimed at factorizing a rating matrix into products of a user-specific matrix and an item-specific matrix, matrix factorization based methods [114, 121, 163] are widely used. Zhang et al. [270] propose a localized matrix factorization approach to tackle the problem of data sparsity and scalability by factorizing block diagonal form matrices. Recently, ranking-oriented collaborative filtering algorithms have achieved great success: using list-wise learning to rank, Shi et al. [213] propose a reciprocal rank method, called

CliMF, to rank items. Following the memory-based collaborative filtering framework, Huang et al. [94] propose ListCF to directly predict a total order of items for each user based on similar users' probability distributions over permutations of commonly rated items.

Collaborative filtering has also been applied to social media recommendation [100, 148, 149, 254]. In recent years, collaborative filtering on Twitter has attracted an increased attention. Yang et al. [256] address recommendation and link prediction tasks based on a joint-propagation model, between social friendship and interests. Ye et al. [260] propose a generative model to describe users' behavior, given influences from social communities [148, 149]. To track social influence of users in a social network, Xu et al. [250] propose a graphical mixture model to describe user's behavior in posting tweets and analyze the topic domain for a specific proposed tweet. Chen et al. [45] propose a collaborative filtering method to generate personalized recommendations in Twitter through a collaborative ranking procedure. Similarly, Pennacchiotti et al. [179] propose a method to recommend "novel" tweets to users by following users' interests and using the tweet content. However, many of these methods ignore the dynamic nature of the problem; with the change of time, user interests may also change.

2.4.3 Explainable recommendation

The "cold-start" problem and poor interpretability are two serious issues for traditional collaborative filtering methods. To address these two issues, in recent years, more and more researchers have started to consider explainable recommendation [29, 228, 271]. Explainable recommendation is known to improve transparency, user trust, effectiveness and scrutability [228]. Vig et al. [238] propose an explainable recommendation method that uses community tags to generate explanations. Based on sentiment lexicon construction, the explicit factor models [271] and Tri-Rank [89] algorithms have been proposed. By combing content-based recommendation and collaborative filtering, Wang and Blei [242] apply topic models [32] to explainable recommendation to discover explainable latent factors in probabilistic matrix factorization. Chen et al. [43] take advantage of the social trust relations by proposing a hierarchical Bayesian model that considers social relationship by putting different priors on users.

Recent work on explainable recommendations focuses on user reviews. Diao et al. [58] propose a hybrid latent factor model integrating user reviews, topic aspects and user ratings for collaborative filtering. By using a multi-dimension tensor factorization strategy, Bhargava et al. [27] propose a recommendation approach by combining users, activities, timestamps and locations. The Hidden Factors as Topic model has been proposed to learn a topic model for items using the review text and a matrix factorization model to fit the ratings [154]. To tackle the sparsity in collaborative topic filtering, the Ratings Meet Reviews model has been proposed by adopting a mixture of Gaussians, which is assumed to have the same distribution as the topic distribution, to model ratings [137].

To the best of our knowledge, there is little previous work on explainable recommendation that jointly considers using user reviews and trusted social relations to improve the rating prediction, not alone generating viewpoints from user reviews.

2.5 Topic Modeling

Early research on topic modeling addressed the topic detection and tracking (TDT) task, where one needs to find and follow topics and events in a stream of broadcast news stories [10, 12]. With the development of social media, topic modeling for social text streams has received increased attention [9, 41, 155, 190]. Yang et al. [257] propose a large-scale topic modeling system that infers topics of tweets over an ontology of hundreds of topics in real-time. Focusing on sparsity and drift, Albakour et al. [9] propose a query expansion method to tackle real-time filtering in microblogs. To help users understand events and topics in social text streams, tweets summarization has also received attention [41, 190, 215].

Topic models have been successfully applied to topic modeling [56, 186, 190, 273]. Topic models [32, 90] are employed to reduce the high dimensionality of terms appearing in text into low-dimensional, "latent" topics. Ever since Hofmann [90] presented probabilistic latent semantic indexing (pLSI), many extensions have been proposed. The latent Dirichlet allocation (LDA, [32]) is one of the most popular topic models based upon the "bag of words" assumption. The author-topic model handles users' connections with particular documents and topics [199]. The entity-topic model detects and links an entity to a latent topic in a document [87]. However, for data with topic evolution the underlying "bag of words" representation may be insufficient. To analyze topic evolution, other models have been proposed, such as the Dynamic Topic Model [31], Dynamic Mixture Models [246] and the Topic Tracking Model [98]. Topic models have not yet been considered very frequently in the setting of Twitter. Twitter-LDA is an interesting exception; it classifies latent topics into "background" topic and "personal" topics [272], while an extension of Twitter-LDA has been proved to be effective in burst detection [57]. Topic models have been extended to sentiment analysis task successfully. For instance, Paul et al. [176] propose a topic model to distinguish topics into two contrastive categories; and Li et al. [124] propose a sentiment-dependency LDA model by considering dependency between adjacent words.

Non-parametric topic models are aimed at handling infinitely many topics; they have received much attention. For instance, to capture the relationship between latent topics, nested Chinese restaurant processes generate tree-like topical structures over documents [33]. To describe the whole life cycle of a topic, Ahmed and Xing [6] propose an infinite dynamic topic model on temporal documents. Instead of assuming that a vocabulary is known *a priori*, Zhai and Boyd-Graber [267] propose an extension of the Dirichlet process to add and delete terms over time. Non-parametric topic models have also been applied to explore personalized topics and time-aware events in social text streams [56]. Traditional non-parametric topic models do not explicitly address diversification among latent variables during clustering. To tackle this issue, Kulesza and Taskar [116, 117] propose a stochastic process named structured determinantal point process (SDPP), where diversity is explicitly considered. As an application in text mining, Gillenwater et al. [78] propose a method for topic modeling based on SDPPs. As far as we know, the determinantal point process has not been integrated with other non-parametric models yet.

Unlike existing topic models, we propose a novel topic model in Chapter 3 by jointly modeling time-aware propagation and collaborative filtering from "social circles." To the best of our knowledge, there is little previous work on summarizing contrastive themes. In Chapter 4, by optimizing the number of topics, building relations among topics and enhancing the diversity among themes, we propose a hierarchical topic modeling strategy to summarize contrastive themes in the given documents. By jointly modeling temporal topics, sentiment labels and entities in multilingual social text streams, in Chapter 5 we propose a cross-language strategy to tackle the viewpoint summarization task for multilingual social text streams. In Chapter 6 we apply a modified dynamic topic model to track topics with topic drift over time, based on both local and global topic distributions. We also focus on a combination of content-based recommendation and collaborative filtering in Chapter 7 by jointly considering topic aspects, user ratings and social trust communities in a latent topic model. Our proposed topic models in Chapters 3–7 are based on latent Dirichlet allocation (LDA, [32]). To help understand our proposed topic models, we provide the basic idea of LDA.

Figure 2.1 shows a graphical representation of LDA, where shaded and unshaded nodes indicate observed and latent variables, respectively. Among the variables related to document set in the graph, z, θ, ϕ are random variables and w is the observed variable; D, N_d and K indicate the number of variables in the model. As usual, directed arrows in a graphical model indicate the dependency between two variables; the variable ϕ depends on variable β, the variable θ depends on variable α.

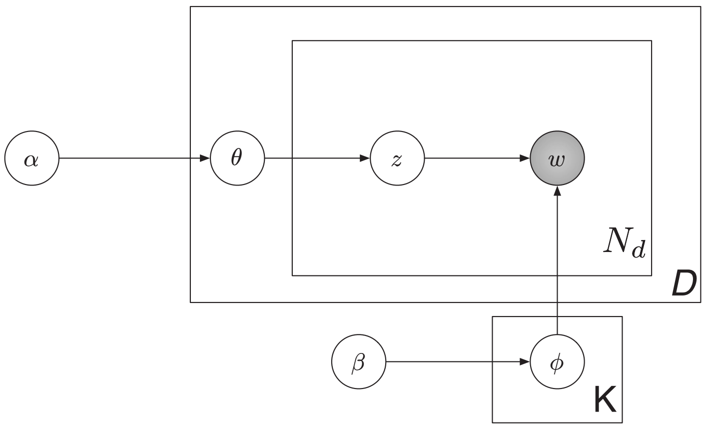


Figure 2.1: Graphical representation of latent Dirichlet allocation.

In LDA, each document is generated by choosing a distribution over topics and then each word in the document is chosen from a selected topic. The topic distributions θ for a document d are derived from a Dirichlet distribution over a hyper parameter α. Given a word $w \in d$, a topic z for word w is derived from a multinomial distribution θ over document d. We derive a probabilistic distribution ϕ over K topics from a Dirichlet distribution over hyper parameters β. The generative process for the LDA model is described in Figure 2.2.

Due to the unknown relation between ϕ and θ, the posterior topic distribution for each document d is intractable in LDA. The posterior distribution in the LDA model

1. For each topic $z, z \in [1, K]$:
   - Draw $\phi \sim Dirichlet(\beta)$;
2. For each candidate document $d \in [1, D]$:
   - Draw $\theta \sim Dirichlet(\alpha)$;
   - For each word w in d
     - Draw a topic $z \sim Multinomial(\theta)$;
     - Draw a word $w \sim Multinomial(\phi_z)$;

Figure 2.2: Generative process for latent Dirichlet allocation.

Algorithm 1: Gibbs Sampling Process for LDA

```
Input: β, α, documents D, number of iterations R, number of topics K
Output: ⟨w, z⟩, topic parameters θ and φ
Initialize values of β, α; Topic assignment for all words
r = 0;
for r < R do
    for d = 1 to D do
        for i = 1 to N_d do
            Draw ⟨w_i, z_i = j⟩ from Eq. 2.2;
            Update n_j^w and n_j^d;
        end
    end
    Calculate θ_{d,j}, φ_{w,j} from Eq. 2.3;
    r = r + 1;
end
```

can be approximated using variational inference with the expectation-maximization algorithm [32]; or an alternative inference technique uses Gibbs sampling [84]. Here we introduce Gibbs collapsed sampling [139] for inferring the posterior distributions over topics. For each iteration during our sampling process, given a word $w_i \in d$, we derive the topic z_i via the following probability:

$$p(z_i = j \mid \mathcal{W}, \mathcal{Z}_{-i}) \propto \frac{n^d_{j,-i} + \alpha}{n^d_{-i} + K\alpha} \cdot \frac{n^{w_i}_{j,-i} + \beta}{n_{j,-i} + W\beta}, \tag{2.2}$$

where $n^d_{j,-i}$ indicates the number of words in d has been assigned to topic j, excluding the current word, and n^d_{-i} indicates the number words in d, excluding the current one; $n^{w_i}_{j,-i}$ indicates the number of times that word w_i has been assigned to topic j, excluding the current word; $n_{j,-i}$ indicates the number words that have been assigned to topic j, not including the ith word in d. Algorithm 1 summarizes the Gibbs sampling inference procedure based on the equations that we have in Eq. 2.2.

During the Gibbs sampling process, we estimate the parameters of document d's topic distribution, θ_d, topic distributions over words ϕ as follows:

$$\begin{aligned} \theta_{d,j} &= \frac{n_j^d + \alpha}{\sum_{k=1}^{K} n_z^d + K\alpha} \\ \phi_{w,j} &= \frac{n_j^w + \beta}{\sum_{z=1}^{K} n_z^w + W\beta}. \end{aligned} \tag{2.3}$$

2.6 Determinantal Point Process

The second part of our contrastive summarization model in Chapter 4 is based on the determinantal point process (DPP) [116]. Here we provide a brief introduction to the DPP.

A *point process* $\mathcal{P}$ on a discrete set $\mathcal{Y} = \{y_1, y_2, \dots, y_N\}$ is a probability measure on the power set $2^{\mathcal{Y}}$ of $\mathcal{Y}$. We follow the definitions from [116]. A *determinantal point process* (DPP) $\mathcal{P}$ is a point process with a positive semidefinite matrix M indexed by the elements of $\mathcal{Y}$, such that if $\mathcal{Y} \sim \mathcal{P}$, then for each discrete set $\mathcal{A} \subseteq \mathcal{Y}$, there is $\mathcal{P}(\mathcal{A} \subseteq \mathcal{Y}) = \det(M_{\mathcal{A}})$. Here, $M_{\mathcal{A}} = [M_{i,j}]_{y_i, y_j \in \mathcal{A}}$ is the restriction of M to the entries indexed by elements of $\mathcal{A}$. Matrix M is defined as the marginal kernel, where it contains all information to compute the probability of $\mathcal{A} \subseteq \mathcal{Y}$. For the purpose of modeling data, the construction of DPP is via *L-ensemble* [36]. Using L-ensemble, we have

$$\mathcal{P}(\mathcal{Y}) = \frac{\det(L_{\mathcal{Y}})}{\sum\limits_{\mathcal{Y}' \subset \mathcal{Y}} \det(L_{\mathcal{Y}'})} = \frac{\det(L_{\mathcal{Y}})}{\det(L + I)}, \tag{2.4}$$

where I is the $N \times N$ identity matrix, L is a positive semidefinite matrix; $L_{\mathcal{Y}} = [L_{i,j}]_{y_i, y_j \in \mathcal{Y}}$ refers to the restriction of L to the entries indexed by elements of $\mathcal{Y}$, and $\det(L_{\emptyset}) = 1$. For each entry of L, we have

$$L_{ij} = q(y_i)\varphi(y_i)^{\mathrm{T}}\varphi(y_j)q(y_j), \tag{2.5}$$

where $q(y_i) \in \mathbb{R}^+$ is considered as the "quality" of an item y_i; $\varphi(y_i)^{\mathrm{T}}\varphi(y_j) \in [-1, 1]$ measures the similarity between item y_i and y_j. Here, for each $\varphi(y_i)$ we set $\varphi(y_i) \in \mathbb{R}^D$ as a normalized D-dimensional feature vector, i.e., $\|\varphi(y_i)\|_2 = 1$. Because the value of a determinant of vectors is equivalent to the volume of the polyhedron spanned by those vectors, $\mathcal{P}(\mathcal{Y})$ is proportional to the volumes spanned by $q(y_i)\varphi(y_i)$. Thus, sets with high-quality, diverse items will get the highest probability in DPP.

Building on the DPP, *structured* determinantal point processes (SDPPs) have been proposed to efficiently handle the problem containing exponentially many structures [78, 116, 117]. In the setting of SDPPs, items set $\mathcal{Y}$ contains a set of threads of length T. Thus in SDPPs, each item y_i has the form $y_i = \{y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(T)}\}$, where $y_i^{(t)}$ indicates the document at the t-th position of thread y_i. To make the normalization and sampling efficient, SDPPs assume a factorization of $q(y_i)$ and $\varphi(y_i)^{\mathrm{T}}\varphi(y_j)$ into parts, decomposing quality multiplicatively and similarity additively, as follows:

$$q(y_i) = \prod_{t=1}^{T} q(y_i^{(t)}) \quad and \quad \varphi(y_i) = \sum_{t=1}^{T} \varphi(y_i^{(t)}). \tag{2.6}$$

The quality function $q(y_i)$ has a simple log-linear model setting $q(y_i) = \exp(\lambda w(y_i))$, where λ is set as a hyperparameter that balances between quality and diversity. An efficient sampling algorithm for SDPPs has been proposed by Kulesza and Taskar [116].

Since SDPPs specifically address "diversification" and "saliency," we apply them to identify diversified and salient themes from themes sets in the contrastive theme summarization. We will detail this step in Chapter 4.

2.7 Structural SVMs

Structural SVMs have been proposed for complex classification problems in machine learning [125, 126, 205]. Generalizing the Support Vector classifier with binary output, structural SVMs generates more complicatged structured labels, such as trees, sets and strings [233, 264]. We follow the notation from [233]. Given an input instance $\mathbf{x}$, the target is to predict the structured label $\mathbf{y}$ from the output space $\mathcal{Y}$ by maximizing a discriminant $\mathcal{F} : \mathcal{X} \times \mathcal{Y} \rightarrow \Re$:

$$\mathbf{y} = f(\mathbf{x}; \mathbf{w}) = \arg\max_{\mathbf{y} \in \mathcal{Y}} \mathcal{F}\left(\mathbf{x}, \mathbf{y}; \mathbf{w}\right), \tag{2.7}$$

where the discriminant $\mathcal{F}$ measures the correlation between $(\mathbf{x}, \mathbf{y})$, and $\mathbf{w}$ indicates the weights of $\mathbf{x}$ in $\mathcal{F}$. The discriminant $\mathcal{F}$ will get its maximal value when $\mathbf{y} = f(\mathbf{x}; \mathbf{w})$, which is set as hypothesis function in structural SVMs. We assume the discriminant $\mathcal{F}$ to be linear in a joint feature space $\Psi : X \times Y \rightarrow R^K$, thus $\mathcal{F}$ can be rewritten as $\mathcal{F}(x, y; w) = \langle \mathbf{w}, \Psi(\mathbf{x}, \mathbf{y}) \rangle$. The feature mapping Ψ maps the pair $(\mathbf{x}, \mathbf{y})$ into a suitable feature space endowed with the dot product. Then the function $\mathcal{F}$ can be learned in a large-margin framework through the training set $\{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{T}$ by minimizing the objective function:

$$\min_{\zeta \geq 0} \tfrac{1}{2} \|w\|^2 + C \textstyle\sum_{i=1}^{n} \zeta_i \tag{2.8}$$

such that for all i and all $\mathbf{y} \in Y \backslash \mathbf{y}^{(i)}$:

$$w^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) - w^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}) \geq \Delta(\mathbf{y}, \mathbf{y}^{(i)}) - \zeta_i, \tag{2.9}$$

where $w^T \Psi(\mathbf{x}^{(i)}, \mathbf{y})$ indicates the hypothesis function value given $\mathbf{x}^{(i)}$ and a random $\mathbf{y}$ from $Y \backslash \mathbf{y}^{(i)}$. For each $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})$, a set of constraints (see Eq. 2.9) is added to optimize the parameters w. Note that $\mathbf{y}^{(i)}$ is the prediction that minimizes the loss function $\Delta(\mathbf{y}, \mathbf{y}^{(i)})$. The loss function equals 0 if and only if $\mathbf{y} = \mathbf{y}^{(i)}$, and it decreases when $\mathbf{y}$ and $\mathbf{y}^{(i)}$ become more similar. Given the exponential size of Y, the number of constraints in Eq. 2.9 makes the optimization challenging.

Now that we have provided the necessary background for the reminder of this thesis. Then, we move to the first research chapter, personalized time-aware tweets summarization.

3 Personalized Time-Aware Tweets Summarization

In the previous chapter we have introduced the background material for this thesis. Starting with this chapter, we begin our research and answer the research questions we listed in Chapter 1. In this chapter, we address **RQ1**, which is concerned with personalized time-aware tweets summarization.

Twitter had amassed over half a billion users as long ago as 2012, who produce ("tweet") over 300 million tweets per day.[1] Twitter users can subscribe to updates from other users by following them, essentially forming a unidirectional friend relationship. Moreover, tweets can be "retweeted," basically copying a tweet posted by another user to one's own timeline. From an information retrieval point of view, the sheer volume of users and tweets presents interesting challenges. On the one hand, interesting, relevant, or meaningful tweets can easily be missed due to a large number of followed users. On the other hand, users may miss interesting tweets when none of the users they follow retweet an interesting piece of information.

One task that is aimed at addressing this dual problem is *tweets summarization* [170]: to extract a group of representative tweets from a set of tweets. The task is similar to tweet recommendation, but tweets summarization pays more attention to the *quality* of selected results, including notions such as representativeness and diversity. So far, tweets summarization methods are typically query and user-independent. How to adapt tweets summarization to a specific user is still a topic of ongoing research [41, 45, 53, 179, 247, 260]. Current methods, whether personalized or not, also neglect to explicitly model the temporal nature of the microblogging environment; time-awareness is a key feature of Twitter in general and tweets summarization in particular. Therefore, we address the following main research question listed in Chapter 1:

RQ1: How can we adapt tweets summarization to a specific user based on a user's history and collaborative social influences? Is it possible to explicitly model the temporal nature of microblogging environment in personalized tweets summarization?

To answer this main research question, we put forward a model for personalized, time-aware tweets summarization (TaTS). We investigate three key aspects of tweets summa-

[1] `http://blog.twitter.com/2012/03/twitter-turns-six.html`.

rization: (a) novelty, preventing near-duplicate tweets to be included, (b) coverage, so as to be representative to candidate tweets, (c) diversity, covering as many aspects as possible. When working with Twitter data, several methodological challenges arise. In order to perform effective tweets summarization, we require a notion of a user's interest. Most Twitter users, however, mostly *consume* information without *producing* a lot of information. That is, they rarely post tweets of their own [179]. Hence, in order to infer a user's interest in a robust manner, we need to use other signals than just the user's tweets. To address the issue, we incorporate intuitions from the field of collaborative filtering and base our estimation of a person's interest on those of their friends on Twitter, following [45]. We assume that for each user there exist one or more "social circles," in which three or more users follow each other and form cliques. We find that people are usually connected to specific communities and assume that each user's behavior on Twitter is affected by: (a) a user's private taste, (b) a collaborative effect from social circles, and (c) a bursty component, reflecting current events.

Clearly, a user's interest can change over time. Topic modeling has proven effective for topic detection and user behavior modeling on Twitter [57, 186, 250]. As a dynamic extension of the author-topic model [199], our proposed Tweet Propagation Model (TPM) aims to track both a user's interests and any topic drift arising with the passing of time. Based on "social circles", TPM derives the user's interest from a dirichlet mixture over interests of someone who share "social circles." It does so by inferring distributions over topics and interests that change over time. Following existing topic modeling approaches for Twitter [57, 272], we extend TPM and classify the topics as (a) personal topics, (b) common topics, or (c) bursty topics. Gibbs Expectation Maximization (EM) sampling [239] is used to infer the posterior probabilities and to estimate the value of hyperparameters in our topic models. After inferring the probabilities of each tweet, we employ an iterative algorithm to optimize the tweet selection procedure, considering coverage, novelty, and diversity.

Our contributions in this chapter are as follows. (1) We propose the task of personalized time-aware tweets summarization, selecting personalized meaningful tweets from a collection of tweets. (2) We leverage a user's "collaborative influence" in order to derive the user's interests. (3) We introduce a tweet propagation model to address the potential drift in a user's interests as well as topics over time. (4) We employ a tweet selection algorithm that jointly optimizes for coverage, diversity, and novelty.

The rest of this chapter is organized as follows. Our problem formulation is detailed in Section 3.1. Our strategy for tweets summary generation, is described in Section 3.2. Section 3.3 details our experimental setup and Section 3.4 presents and discusses the experimental results and Section 3.5 concludes the chapter.

3.1 Problem Formulation

Before introducing our method for time-aware tweets summarization, we introduce our notation and key concepts. Table 3.1 lists the notation we use. Given two users u_i and u_j on Twitter, there are two main reasons for u_i and u_j to follow each other: either because they have similar interests or they have some relationship outside Twitter [250]. If two users u_i and u_j follow each other, we define them to be *friends* on Twitter. Given this

Table 3.1: Notation used in this chapter.

Symbol	Description
K	number of topics
U	number of users
V	the size of the vocabulary
T	number of time periods
$\mathcal{D}_t$	candidate tweets at time t
D_t	number of candidate tweets at time t, i.e., $\lvert\mathcal{D}_t\rvert$
u	user u on Twitter, $u \in \mathcal{U}$
$\mathcal{C}_{u,t}$	social circle for user u at t
$\mathcal{D}_{u,t}$	tweets posted by u at time t
$D_{u,t}$	number of tweets posted by u at time t, i.e., $\lvert\mathcal{D}_{u,t}\rvert$
$F_{u,t}$	number of friends of u at time t
$C_{u,t}$	number of social circles around u at time t
d_t	tweet published at time t, $d_t \in \mathcal{D}_t$
w	token/word present in some tweet, $w \in \mathcal{W}$
z_t	latent topic at t time, $z_t \in Z_t$
$c_{u,t}$	social circle around u at time t, $c_{u,t} \in \mathcal{C}_{u,t}$
$\theta_{u,t}$	distribution of u's interests over topics at time t
ϑ_t	distribution of topics within a tweet at time t
ϕ_t	distribution of words over topics at time t
Z	classification of individual topics in θ or ϑ
β, α, σ, r	hyper-parameters in TPM
N	maximum number of tweets returned
$\lambda_{u,c,t}$	weight of social circle c for user u at t

definition, we define a *social circle* around a user u to be a set of friends of u such that every pair of users in this set is in the *friend* relation. See Figure 3.1 for a schematic representation.

Similar to the author-topic model [199], we assume that each Twitter user's interests are represented by a multinomial distribution $\theta_{u,t}$, which may, however, change over time. That is, the time-aware interests of user u are represented as a multinomial distribution $\theta_{u,t}$ over topics, where each topic is represented as a probabilistic distribution over words [32]. Formally, we have $\theta_{u,t} = \{\theta_{u,t,z_1}, \ldots, \theta_{u,t,z_K}\}$, where θ_{u,t,z_i}, denotes the distribution of topic z_i for user u at time t.

We further assume that each tweet can be represented as a probabilistic distribution over topics. To cater for the phenomenon of user interests changing over time, we assume that topic distributions are dynamic and may differ between two time periods. Given a user u, we split the topic set Z_t at time t into three classes: $Z_t = Z_t^u \cup Z_t^{com} \cup Z_t^B$: there exist "private" topics Z_t^u that solely depend on the user, there are common topics Z_t^{com} that are influenced by friends from shared social circles, and there are topics from event-related, bursty sources, Z_t^B. The latter type of topic will typically transfer from initially being observed at time t into Z^{com} at some later time t'.

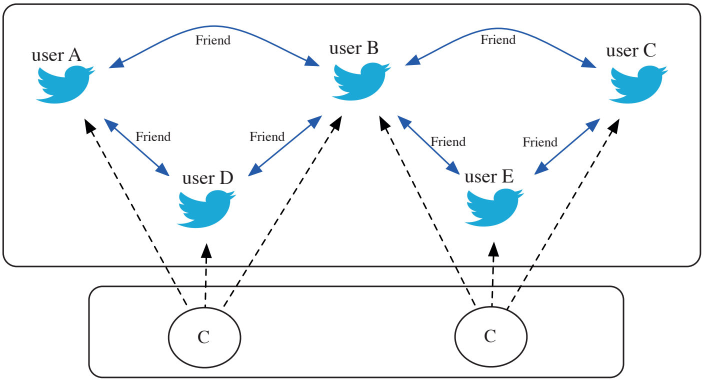


Figure 3.1: Example of social circles on Twitter: there are two social circles (indicated using the 'c') among the five users in this graph, where each pair of vertices in each social circle is connected through the "friend" relationship.

The dynamic interests of user u at time t, reflected by $\theta_{u,t}$, evolve in different ways depending on the class that a topic $z_t \in Z_t$ belongs to. For each user, $\theta_{u,t}$ is affected by the following three classes.

(a) If $z_t \in Z_t^u$ is a "private" topic, then $\theta^u_{u,t,z}$ only depends on $\theta^u_{u,t-1,z}$ at time $t-1$.
(b) If $z_t \in Z_t^{com}$ then the topic is dependent on friends in the user's social circle(s). θ^{com}_{u,t,z_t} is computed from the collaborative effect $\theta^{com}_{u_i,t-1}$ at time $t-1$ from the social circles $\{u_i \mid u_i \in \mathcal{C}_{u,t-1}\}$.
(c) If $z_t \in Z_t^B$ is a "burst" topic, $\theta^B_{u_i,z,t}$ is generated according to a distribution of "burst" words in u_i's tweets at time t.

Typically, traditional summarization does not cover the evolution of a specific event. Given a split of a user's history into time periods, the task of time-aware tweets summarization is to select the most representative tweets for each time period, covering the whole event evolution on a timeline. More precisely, given a set of tweets $\mathcal{D}$, a set of time periods T, and a maximum number of tweets per period, N, time-aware tweets summarization aims to extract multiple sets of tweets RT_t $(1 \leq t \leq T)$ from $\mathcal{D}$, where for each time period t, $RT_t = \{d_{t,x_1}, d_{t,x_2}, \ldots, d_{t,x_N}\}$ is a set of representative tweets that summarize the period. Furthermore, *personalized* time-aware tweets summarization is defined similar to time-aware tweets summarization, but in this case the tweets selected for inclusion in RT_t need to be relevant based on u's interests θ_u at time t.

3.2 Method

In this section, we detail our tweets summarization method, including the required methods for joint user-tweets topic modeling, inference and parameter estimation. As input, our method has probabilistic distributions from topic modeling. The output is the time-aware tweets summary, i.e., a selection of tweets (per period) satisfying the user's interest.

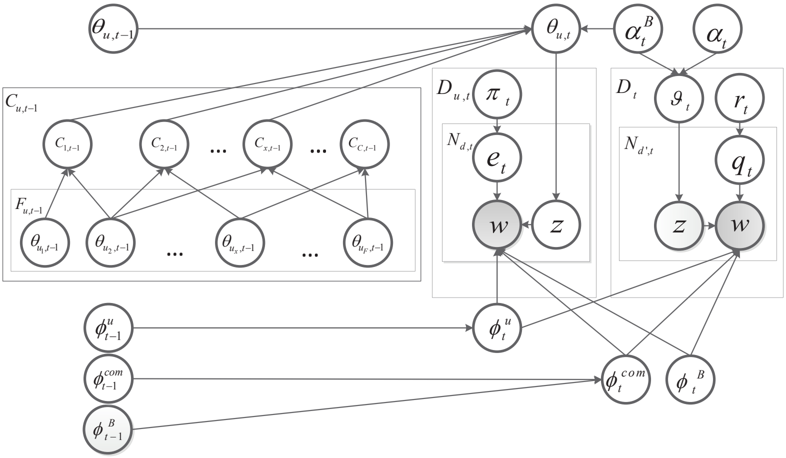


Figure 3.2: Graphical representation of TPM.

3.2.1 Topic modeling: tweets propagation

We start by proposing the tweets propagation model (TPM) to jointly track dynamic user's interests and topics. The interests of a user u are assumed to be reflected by a multinomial distribution $\theta_{u,t}$ over topics. We assume that the distribution of topics ϕ_t over words follows a dynamic propagation process with changes over time. Figure 3.2 provides a graphical overview of TPM.

In the graphical structure of TPM, we see a number of ingredients. Among the variables related to user u in the graph, z, θ and γ are random variables and w is the observed variable. In the candidate tweets part, ϑ, z and σ are random variables; $D_{u,t}$ and D_t indicate the number of variables in the model. As usual, directed arrows in a graphical model indicate the dependency between two variables; the variables $c_{u,t-1}$ depend on variables $\{\theta_{u_i,t-1} \mid u_i \in \mathcal{C}_{u,t-1}\}$. The variables ϕ_t^{com} and ϕ_t^u depend on variables $\{\phi_{t-1}^{com}, \phi_{t-1}^B\}$ and ϕ_{t-1}^u, respectively.

Now, let us give a more detailed technical account of our model. Around user u, there exist multiple social circles. For each social circle $c_{u,t}$ in time period t, there is a random parameter $\lambda_{c_{u,t}}$ indicating the importance of $c_{u,t}$ to u at t. User u's interests $\theta_{u,t}$ are composed of three parts: the personal aspect, the common topic aspect and the bursty aspect, i.e., $\theta_{u,t} = \left\{\theta_{u,t}^{com}, \theta_{u,t}^u, \theta_{u,t}^B\right\}$, where the common topics are not only influenced by the user's social circles, but also by his own previous interests. Therefore, we use a Dirichlet distribution to derive the probability of $\theta_{u,t}^{com}$ over $x_{u,t}^{com}$ as:

$$x_{u,t}^{com} = \alpha_{u,t}\theta_{u,t-1}^{com+B} + (1-\alpha_{u,t}) \sum_{c_i \in C_{u,t-1}} \lambda_{c_i}\theta_{c_i,t-1}^{com+B}, \tag{3.1}$$

where $\theta_{u,t-1}^{com+B}$ refers to the set $\{\theta_{u,t-1}^{com}, \theta_{u,t-1}^B\}$ at period $t-1$, which reflects user u's interests for common and burst topics at time $t-1$, and $\theta_{c_i,t-1}^{com+B}$ refers to the set $\left\{\theta_{c_i,t-1}^{com}, \theta_{c_i,t-1}^B\right\}$ at period $t-1$. The hyperparameter $\alpha_{u,t}$ indicates the weight of

$\theta_{u,t-1}^{com+B}$ in Eq. 3.1 that we use to calculate $\theta_{u,t}^{com+B}$. Here, the value of $\theta_{c_i,t-1}^{com+B}$ is equal to $\frac{1}{|\mathcal{C}_{u,t}|}\sum\theta_{u_i,t-1}$, where $u_i \in \mathcal{C}_{u,t-1}$.

For private topical aspects $\theta_{u,t}^{u}$, we use a Dirichlet distribution over $x_t^u = \theta_{u,t-1}^{u}$ that is derived from values in period $t-1$. For bursty topics in period t , we only focus on those "burst" words that have a high term frequency within period t. Similar to [250], we define a keyword to be "bursty" if its frequency $n_{w,t}$ at time t is above a threshold value. We derive $\theta_{u,t}^{B}$ from a Dirichlet distribution over the hyperparameter α_t^B.

For a tweet in $\mathcal{D}_t$ that is posted during time period t, a probabilistic distribution ϑ_t over topics $Z_t = Z_t^u \cup Z_t^{com} \cup Z_t^B$ is derived from a Dirichlet distribution over the hyperparameter α_t.

For each word w in tweet $d_t, d_t \in \{\mathcal{D}_{u,t}, \mathcal{D}_t\}$ proposed during period t, we assign a specific topic z from u's interests $\theta_{u,t}$ or distribution ϑ_t for candidate documents. For topic aspects z ($z \in Z_t^{com} \cup Z_t^B \cup Z_t^u$), we introduce three kinds of multinomial distribution ϕ_t^{com}, ϕ_t^u and ϕ_t^B to reflect the probability over Z^{com}, Z^B and Z^u, respectively. Based on [98, 246], we assume that the common and personal topic propagations follow a Dirichlet distribution over the value from the previous interval's distributions, with a weighted prior $\beta_t = \{\beta_t^{com}, \beta_t^u\}$: for common topics $z \in Z_t^{com}$, we use the Dirichlet distribution to infer from $\left\{\phi_{t-1}^{com}, \phi_{t-1}^{B}\right\}$; for private topics $z \in Z_t^u$, ϕ_t^u is derived from ϕ_{t-1}^u.

This concludes the technical account of the graphical model depicted in Figure 3.2. After computing the models for period t for all users in $\mathcal{U}$, we update the edge weights for the social circles ($\lambda_{u,c_i,t}$), using related users' interests θ and current social circles. Inference for our topic modeling process will then move on to period $t+1$. The generative process for the TPM model at time interval t, $0 < t < T$, is described in Figure 3.3.

3.2.2 Inference and parameter estimation

Sampling-based methods for LDA rarely include methods for optimizing hyper-parameters. In the TPM model, since $\alpha_{u,t}$ and $\beta_{z,t}^{f_l}$ indicate the weight of the results for period $t-1$ for computations for period t, it is necessary to find an optimized process for hyper-parameters $\alpha_{u,t}$ and $\beta_{z,t}^{f_l}$ during our posterior inference. Therefore, unlike many previous dynamic topic models, to infer weighted priors we use a Gibbs EM algorithm [239] to handle the approximate posterior inference step. For user u at time interval t, we first jointly sample topic z_i and parameter q_i from the ith word in tweet d ($d \in D_{u,t}$) over other variables. So for u's tweets we obtain:

$$\begin{aligned}p(e_i = l, z_i = z \mid \mathcal{W}, e_{-i}, Z_{-i}, x_{u,t}, \sigma, \beta_t^u) \propto \frac{n_{d,l,-i}^{u,t} + \sigma}{n_{d,-i}^{u,t} + 3\sigma} \cdot \\ \frac{n_{d,z,-i}^{u,t} + x_{u,z,t}^{f_l}}{\sum\limits_{z' \in Z^{f_l}} (n_{d,z,'-i}^{u,t} + x_{u,z',t}^{f_l})} \cdot \frac{n_{w,z,-i}^{u,t} + \beta_t^{f_l}}{\sum\limits_{w' \in N_{u,t}} n_{w',z,-i}^{u,t} + N_{u,t}\beta_t^{f_l}},\end{aligned} \tag{3.2}$$

where l indicates the possible values of variable e for the ith word in tweet p, and the f_l indicate the corresponding kind of topics when $e_i = l$. For private and common topics in u, i.e., $l = 0, 1$, in Eq. 3.2, $n_{d,l,-i}^{u,t}$ indicates the number of times that words in d are

1. For each topic $z, z \in Z_t^{com} \cup Z_t^B \cup Z_t^u$:
   - Draw $\phi_t^B \sim Dirichlet(\beta_t^B)$;
   - Draw $\phi_t^{com} \sim Dirichlet(\beta_t^{com}\left\{\phi_{t-1}^{com}, \phi^B{}_{t-1}\right\})$;
   - Draw $\phi_t^u \sim Dirichlet(\beta_t^u \phi_{t-1}^u)$
2. For each candidate tweet $d_t \in \mathcal{D}_t$:
   - Draw $\vartheta_t \sim Dirichlet(\alpha_t, \alpha_t^B)$; $r_t \sim Dirichlet(\gamma_t)$;
   - For each word w in d_t
     - Draw $q \in Multinomial(r)$; $z_w \sim Multinomial(\vartheta_t)$;
       * if $q = 0$: Draw $w \sim Multinomial(\phi_{z,t}^{com})$;
       * if $q = 1$: Draw $w \sim Multinomial(\phi_{z,t}^u)$;
       * if $q = 2$: Draw $w \sim Multinomal(\phi_t^B)$;
3. For user $u, u \in \mathcal{U}$:
   - Draw $\theta_{u,t} \sim Dirichlet(\left\{x_{u,t}^u, x_{u,t}^{com}, \alpha_t^B\right\})$;
   - Draw $\pi_t \sim Dirichlet(\sigma_t)$;
   - For each word $w \in d_{u,t}$, where $d_{u,t} \in D_{u,t}$:
     - Draw $e \sim Multinomial(\pi)$; $z_{w,t} \sim Multinomial(\theta_{u,t})$
       * if $e = 0$: Draw $w \sim Multinomial(\phi_{z,t}^{com})$;
       * if $e = 1$: Draw $w \sim Multinomial(\phi_{z,t}^u)$;
       * if $e = 2$: Draw $w \sim Multinomial(\phi_t^B)$;

Figure 3.3: Generative process for the TPM model.

assigned to label l except for the ith word, whereas $n_{d,-i}^{u,t}$ indicates the sum of $n_{d,l,-i}^{u,t}$ for all values of l. Furthermore, $n_{d,z,-i}^{u,t}$ is the number of times that tweet d is assigned to topic z excluding the ith word in d, whereas $n_{w,z,-i}^{u,t}$ indicates the number of times that word w is assigned by topic z excluding the ith word. According to Figure 3.3, if $e_i = 2$, we are dealing with a "bursty" topic, so the vocabulary only refers to the set of "bursty" keywords in $\{\mathcal{D}_{u,t}, \mathcal{D}_t\}$, then $x_{u,t}^B$ in Eq. 3.2 equals to α_t^B.

For the process of sampling candidate tweets from $\mathcal{D}_t$, we have a similar procedure, as follows:

$$
\begin{aligned}
p(q_i = l, z_i = z \mid \mathcal{W}, d_{-i}, Z_{-i}, \alpha_t, \gamma, \beta_t^u) \propto \frac{n_{d,l,-i}^t + \gamma}{n_{d,-i}^t + 3\gamma} \cdot \\
\frac{n_{d,z,-i}^t + \alpha_t^{f_l}}{\sum\limits_{z' \in Z^{f_l}} n_{d,z,'-i}^t + Z^{f_l}\alpha_t^{f_l}} \cdot \frac{n_{w,z,-i}^t + \beta_t^{f_l}}{\sum\limits_{w' \in N_t} n_{w',z,-i}^{u,t} + N_{u,t}\beta_t^{f_l}}.
\end{aligned}
\tag{3.3}
$$

Meanwhile, every time after sampling for $p(e_i = l, z_i = z)$ and $p(q_i = l, z_i = z)$, we optimize $\widehat{\alpha}_{u,t}$ and $\widehat{\beta}_{z,t,t-1}^{f_l}$ by maximizing the likelihood posterior distribution

$$
p(W \mid \Phi_{t-1}, x_{u,t-1}, \alpha^B, \beta_t, \sigma, \gamma),
$$

so we get

$$
\widehat{\alpha}_{u,t} = \widehat{\alpha}_{u,t} \cdot \frac{\sum\limits_{z \in Z_t^{com}} (\theta_{u,t-1,z} - \sum\limits_{c_i \in C_{u,t-1}} \lambda\theta_{c_i,t-1}) A_{u,z,t}}{\Psi(n_{u,t}^{com} + \alpha_{u,t}) - \Psi(\alpha_{u,t})},
\tag{3.4}
$$

Algorithm 2: Gibbs EM Sampling Process during period t

Input: β_t, β^B, α^B, α_t, $X_{u,t}$, Φ^f_{t-1}, $d_t, \mathcal{U}$, $\mathcal{D}_t$ and R
Output: $\hat{\beta}^{f_l}_t$,$\widehat{\alpha}_t$,$\langle e, z\rangle$ and $\langle q, z\rangle$
Initialize β_t, β^B, α^B, α_t; *Topic assignment for all words*
for $u \in \mathcal{U}$ **do**
    $r = 0$;
    for *r¡R* **do**
        E-Step:
        for $d = 1$ *to* $D_{u,t}$ **do**
            for $i = 1$ *to* N_d **do**
                Draw $\langle e_i, z_i\rangle$ from Eq. 3.2
                Update $n^{u,t}_{e,0,i}$, $n^{u,t}_{e,z,i}$ and $n^{u,t}_{w,z,i}$
            end
        end
        for $d = 1$ *to* D_t **do**
            for $i = 1$ *to* N_d **do**
                Draw $\langle q_i, z_i\rangle$ from Eq. 3.3
                Update $n^t_{q,l,i}$, $n^t_{d,z,i}$ and $n^t_{w,z,i}$;
            end
        end
        M-Step:
        Calculate $\theta^{f_l}_{u,t}$, $\phi^{f_l}_{w,t}$, $\vartheta^{f_l}_{d,t}$, *and* λ_{u,c_i} *from Eq. 3.6, 3.8*;
        Maximize $\widehat{\alpha}^{(r)}_{u,t}$*and* $\hat{\beta}^{f_l,(r)}_{z,t}$ *from Eq. 3.4, 3.5*;
        $r = r + 1$ *and go to* **E-Step**;
    end
end

and

$$\hat{\beta}^{f_l}_{z,t} = \hat{\beta}^{f_l}_{z,t} \cdot \frac{\sum\limits_{w\in N_t} \phi^{f_l}_{t-1,w} \Big(\Psi(n^{f_l}_{w,z,t} + y^{w,z}_{t,t-1}) - \Psi(y^{w,z}_{t,t-1})\Big)}{\Psi(n^{f_l}_{z,t} + \beta_{z,t}) - \Psi(\beta_{z,t})}, \tag{3.5}$$

where $\Psi(x)$ is defined by $\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$, $A_{u,z,t}$ refers to

$$\Psi(n^{com}_{u,z,t} + x^{com}_{u,z,t}) - \Psi(x^{com}_{u,z,t}),$$

and $y^{w,z}_{t,t-1}$ is defined as $\beta^{com}_t \phi^{com+B}_{t-1}$.

Algorithm 2 summarizes the Gibbs EM sampling inference based on the equations that we have just derived. During the Gibbs EM sampling process, we estimate the parameters of user u's interests $\theta^{e=l}_{u,z,t}$, the probability of topics over candidate tweets

$\vartheta^{q=l}_{d,z,t}$, topic distributions over words $\phi^{f_l}_{w,z,t}$ and $\{\pi_{d,l,t}, r_{d,l,t}\}$ as follows:

$$
\begin{aligned}
\theta^{f_l}_{u,z,t} &= \frac{n^{u,t}_{z} + x^{f_l}_{u,z,t}}{\sum\limits_{z \in Z^{f_l}} n^{u,t}_{z'} + x^{f_l}_{u,z',t}} \\
\vartheta^{f_l}_{d,z,t} &= \frac{n_{d,z,t} + \alpha_{z,t}}{\sum\limits_{z \in Z^{f_l}} n_{d,z',t} + \alpha_{z',t}} \\
\phi^{f_l}_{w,z,t} &= \frac{n_{w,z,t} + \beta^{f_l}_{w,t}}{\sum\limits_{z \in Z^{f_l}} n_{w,z,t} + \beta^{f_l}_{w,t}} \\
\pi_{d,l,t} &= \frac{n^{u,t}_{d,l} + \sigma}{n^{u,t}_{d} + 3\sigma} \\
r_{d,l,t} &= \frac{n^{t}_{d,l} + \gamma}{n^{t}_{d} + 3\gamma}.
\end{aligned} \tag{3.6}
$$

To compute the weight $\lambda_{c_{u,t}}$, we use a Markov random walk strategy, which calculates saliency of a social circle based on "voting" from others. Since each social circle can be considered as a set of users, an interest distribution $\theta^{com+B}_{c_i,t-1}$ for each social circle c_i can be computed as $\sum_{u' \in c_i} \theta^{com+B}_{u',t-1}$. Thus we compute a $\theta^{com+B}_{u,t}$-based similarity matrix $SIM^{u,t}$ among different social circles, where each item $SIM^{u,t}_{i,j}$ is computed based on the divergence between two items:

$$
div(\theta_{c_i}, \theta_{c_j} \mid \theta_u) = \sum_{z \in Z} \left| \theta_{c_i,z} \ln \frac{\theta_{c_i,z}}{\theta_{u,z}} - \theta_{c_j,z} \ln \frac{\theta_{c_i,z}}{\theta_{u,z}} \right|, \tag{3.7}
$$

We calculate the saliency of c_i after normalizing SIM into $\widehat{SIM}$:

$$
\lambda_{u,c_i,t} = \mu \sum_{i \neq j} \widehat{sim}(\theta_{c_i,t}, \theta_{c_j,t} | \theta_{u,t}) \cdot \lambda_{u,c_i,t} + \frac{(1-\mu)}{|C_{u,t}|}. \tag{3.8}
$$

3.2.3 Time-aware summarization

After Gibbs EM sampling, for each candidate tweet d_t at time t, we have two parametric distributions ϑ_t and ϕ_t that reflect the topic-tweet distribution and the word-topic distribution, respectively. I.e., $P(z_t \mid d_t) = \theta_{z_t,d_t}$ and $P(w \mid z_t) = \phi_{z,t,w}$. For user u at time t, we now derive the distribution of interests over topics $\theta_{u,t}$, i.e., $P(z_t \mid u, t)$.

Given the distribution $\theta_{u,t}$, one intuitive way to get the most meaningful tweets is to extract the most similar tweets with $\theta_{u,t}$ from among a candidate set $\mathcal{D}_t$. However, a high-degree relevance in latent topic distributions cannot be taken as the only criterion in our tweet selection. Thus after extracting a set of relevant tweets $\mathcal{R}_t$ from $\mathcal{D}_t$, there are three key requirements for an ideal summary [125] that we need to consider in generating a tweet summary: *novelty*, the *coverage* and the *diversity*.

Novelty calculates the semantic divergence between the currently selected set $RT_{u,t}$ and the results in previous time periods $RT_{t'}$. Our intention is to make the current results as different as possible from previous results as much as possible. Therefore, we have:

$$\mathcal{L}_N(RT_t|RT_{t'}) = \sum_{p \in RT} \min_{p' \in RT_{t'}}(div(\vartheta_p, \vartheta_{p'} \mid \theta_{u,t})), \tag{3.9}$$

where the divergence $div(\vartheta_p, \vartheta_{p'} \mid \theta_{u,t})$ between ϑ_p and $\vartheta_{p'}$ are calculated based on Eq. 3.7.

Furthermore, a tweet summary should contain important aspects from all related tweets and minimize the information loss with the set of all candidate tweets. Thus, given $\theta_{u,z,t}$, the *coverage* between RT and $\mathcal{D}_t$ is calculated as follows:

$$\mathcal{L}_C(RT \mid \mathcal{D}_t) = \sum_{d \in RT} e^{-\min_z \sum\limits_{d' \in \mathcal{D}_t} div(\vartheta_{d,z}, \vartheta_{d',z} | \theta_{u,z,t})}, \tag{3.10}$$

where the divergence $div(\vartheta_{d,z}, \vartheta_{d',z}|\theta_{u,z,t})$ is calculated as follows:

$$div(\vartheta_{d,z}, \vartheta_{d',z} \mid \theta_{u,z,t}) = \left| \vartheta_{d,z} \ln \frac{\vartheta_{d,z}}{\theta_{u,z,t}} - \vartheta_{d',z} \ln \frac{\vartheta_{d,z}}{\theta_{u,z,t}} \right|, \tag{3.11}$$

Diversity calculates the information divergence among all tweets within the current candidate result set. Ideally, the tweet summary results have the largest possible difference in topic distributions with each other. The equation is as follows:

$$\mathcal{L}_D(RT) = \sum_{w,w' \in RT} \max_z div(\phi_{w,z,t}, \phi_{w',z,t} \mid \prod_{d \in \mathcal{D}_t} \vartheta_{d,z}). \tag{3.12}$$

where we compute the divergence $div(\phi_{w,z,t}, \phi_{w',z,t} \mid \prod_{d \in \mathcal{D}_t} \vartheta_{d,z})$ in the same way as Eq. 3.11.

The exact process for generating $RT_{u,t}$ given user u is shown in Algorithm 3. Illuminated by a previous work [252], an iterative optimization algorithm is used to select the set $RT_{u,t}$. During each iteration n, we extract tweet d_x such that $d_x \in R_t \wedge \neg RT_{u,t}$ to substitute $d_y \in RT^{(n)}_{u,t}$ when the saliency gain $\mathcal{S}((RT_{u,t} - d_y) \cup d_x) - \mathcal{S}(RT_{u,t})$ gets a maximum value. The algorithm will converge when $\mathcal{S}(RT_{u,t})$ reaches its maximum value.

3.3 Experimental Setup

For our experiments we employ a Twitter dataset that includes both social relations and tweets: we crawl tweets via the Twitter streaming API,[2] which contains a random sample of around 10% of all items posted on Twitter. Timestamps in our dataset are from November 1, 2009 to December 31, 2010; the 2009 part contains 47,373,408 tweets and 562,361 users, while the numbers for 2010 are 295,145,421 and 5,828,356, respectively. Figure 3.4(a) shows the statistics of the number of tweets per user in our dataset, where

[2]https://dev.twitter.com/docs/streaming-apis.

Algorithm 3: Iterative Process for $RT_{u,t}$ Generation.

Input : $\mathcal{D}_t$, $RT_{u,t'}$, $\theta_{u,t}$, ϕ_t, N;
Output: $RT_{u,t}$;

Calculate Kullback-Leibler divergence $KL(\vartheta_{d,t}, \theta_{u,t})$;
Rank and extract relevant tweets to $\mathcal{R}_t$ by $e^{-KL(\vartheta_{d,t},\theta_{u,t})}$;
Initialize: Extract N tweets from $\mathcal{R}_t$ to $RT_{u,t}$;
repeat
  Extract $\mathcal{X}_t = \{d_x \in R_t \wedge \neg RT_{u,t}\}$;
  for $d_x \in \mathcal{X}_t, \forall d_y \in RT_{u,t}$ **do**
    Calculate $\mathcal{S}_{RT_{u,t}} = F(\mathcal{L}_C \cdot \mathcal{L}_N \cdot \mathcal{L}_D)$;
    Calculate $\Delta\mathcal{S}_{d_x,d_y} = \mathcal{S}((RT_{u,t} - d_y) \cup d_x) - \mathcal{S}(RT_{u,t})$;
  end
  Get $\left\langle \hat{d}_x, \hat{d}_y \right\rangle$ that $\left\langle \hat{d}_x, \hat{d}_y \right\rangle = \arg\max_{d_x,d_y} \Delta S_{d_x,d_y}$;
  $RT_{u,t} = (RT_{u,t} - \hat{d}_y) \cup \hat{d}_x$;
until $\forall \Delta S_{d_x,d_y} < \varepsilon$;
return $RT_{u,t}$.

we can find that most users (75.2%) in our dataset wrote fewer than 100 tweets. For crawling the social relations, we use the dataset from [118], which includes social relations for all users on Twitter until July 2009. In our experiments, we use only those tweets and users that appear in both datasets. In our experiments we assume social relations among users to remain the same over the entire time period.

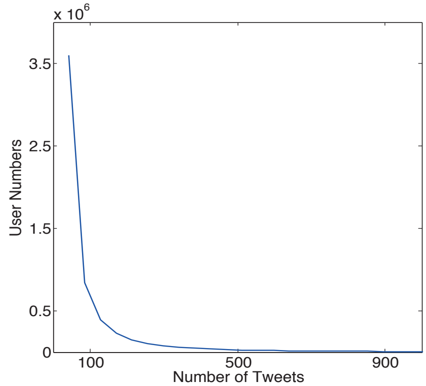
Since it is impossible to evaluate the effectiveness if a user posted nothing on Twitter, sparse postings obstruct our experimental evaluation. We therefore only consider users who posted a sufficient number of tweets for our evaluation: we collect users who post over 100 tweets in our dataset. This results in a subset containing 32,659 users. Thereafter we use social relations to build the social circles around those users. Figure 3.4(b) shows the number of tweets of these users (y-axis) versus the number of friends on the x-axis. We further remove non-English tweets through automatic language identification [38]. We remove stop words and apply Porter stemming [183].

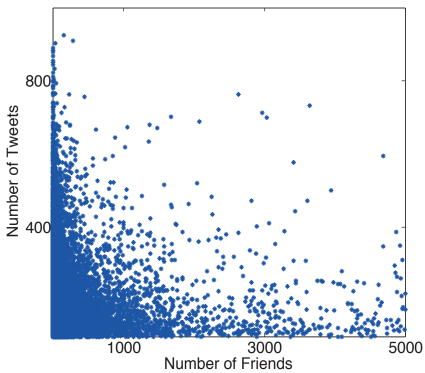
3.3.1 Data enrichment

Since each tweet is only up to 140 characters long, the amount of textual evidence to work with is very limited. To remedy this, we employ a state-of-the-art method for *linking* tweets to Wikipedia articles [158]. In particular, we employ the so-called CMNS method that uses the prior probability that Wikipedia article c is the target of a link with anchor text q within Wikipedia:

$$CMNS(c, q) = \frac{|L_{q,c}|}{\sum_{c'} |L_{q,c'}|}, \tag{3.13}$$

where $L_{q,c}$ denotes the set of all links with anchor text q and target c.





(a) (b)

Figure 3.4: Histograms of the number of users and tweets in our dataset: the left (a) indicates the number of tweets per user in our dataset where the y-axis denotes the number of tweets; while the right (b) indicates the number of tweets per user with its number of friends in Twitter, where y-axis indicates the number of tweets the user wrote and the x-axis indicates the number of friends.

After we have obtained three Wikipedia articles with the highest CMNS score, we extract the most *central* sentences from these Wikipedia articles and append them to the tweet. In particular, we apply a query-sensitive graph-based summarization method, similar to [63], to each Wikipedia article to ranking sentences, using the tweet d_t as the query. This calculates the score of each sentence via "votes" from other sentences in a document. Figure 3.5 shows four example tweets and the appended sentences. Here, the left text box in each item is a tweet and on the right we show the identified sentences from the linked Wikipedia articles.

3.3.2 Experimental setup

Following existing topic models [84], we set pre-defined values for the hyperparameters α_t and β_t in our graphical model: for the weighted parameter $\alpha_{u,t}$ and β_t, we set $50/K_t^u$ to $\alpha_{u,t}$ and 0.5 to β_t respectively. And we set $50/K_t^B$ to α^B and 0.5 to β^B respectively. For the hyperparameters γ and σ in TPM, as defined in [101], we set $\sigma_u = \gamma_{com} = 0.5$ and $\gamma_u = \sigma_{com} = 0.3$. For burst topics we set $\gamma_B = \sigma_B = 0.2$ in our experiments. The initial value of $\lambda_{u,c_i,t-1}$ for each social circle of u is set to $1/C_{u,t}$, the parameter μ is set as 0.85; and ε in Algorithm 3 is set to 0.0001. For the number of topics in our topic modeling process, the default values for Z_0^u and Z_0^{com} in our experiments are set to 100, respectively. To optimize the number of topics, we compare performance in various values and discuss it latter.

Statistical significance of observed differences between two comparisons is tested using a two-tailed paired t-test. In our experiments, statistical significance is denoted using $^\blacktriangle$ for significant differences for $\alpha = 0.01$, or $^\vartriangle$ for $\alpha = 0.05$.

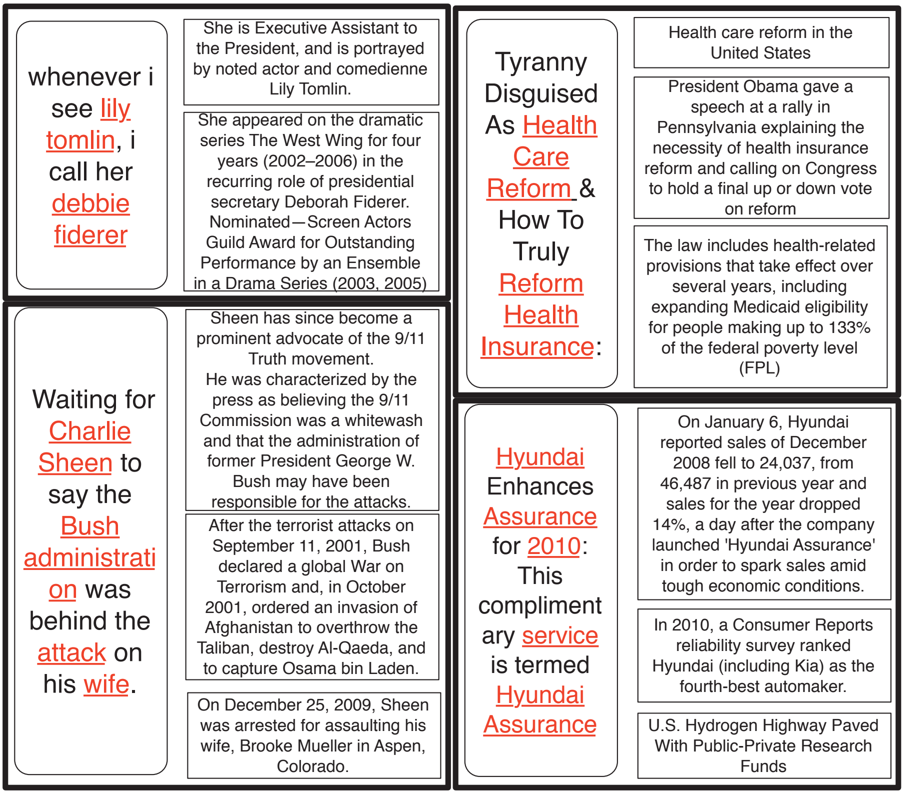


Figure 3.5: Four examples for entity linking and ranking corresponding to four individual tweets, where the textbox on left side indicates the original tweet while the textbox on the right side shows the extracted related sentences. A mixture of the tweet and extracted wiki sentences will replace the original tweet in our experiments.

3.3.3 Evaluation metrics

Evaluating the effectiveness of time-aware tweets summarization is a challenging task, especially in the absence of explicit user feedback. One possible solution is to use evidence from users themselves: we use a user's retweeted post(s) at time $t+1$ as the ground truth to evaluate performance of comparisons at time t.

We measure the quality of summaries by counting overlapping textual *units* between the generated results and the ground truth results. In our experiments, we adopt the ROUGE evaluation metrics [133], a widely-used recall-oriented metric in the task of document summarization that evaluates the overlap between a gold standard and candidate selections.[3] In our experiments, ROUGE-1 (*unigram based method*), ROUGE-2 (*bigram based method*) and ROUGE-W (*weighted longest common sequence*) are used

[3] Version 1.5.5 is used in this chapter.

as evaluation metrics.

3.3.4 Baseline comparisons

Given the TPM modeling introduced in Section 3.2.1, our contribution is twofold: (1) we introduce collaborative influence to user's interests detection; (2) we adopt time-aware propagation to infer topics. To evaluate the influence of social circles and time-aware topics, besides our overall TPM-based strategy, we also evaluate the performance of the model that only includes (1) the collaborative influence or only the (2) time-aware propagation, respectively.

We write **TPM-ALL** for the overall process as described in Section 3.2.1, which includes both the social influence modeling and time-aware topic and interests tracking. We write **TPM-SOC** for the model that only considers users' social influence (so excluding time-aware topic propagation and it does not consider if some topic is private or not). We write **TPM-TOP** for the model that uses a user's own tweets (without social circles but considering topic and interests propagation with the time).

To evaluate our proposed method in more detail, in our experiments the baselines not only include widely-used topic models, but also recent user behavior models on Twitter. For those topic models, we use the Author-Topic Model (**AT**) [199] and the Twitter-LDA [272] as baselines for topic models: (AT) focuses on various users' interests in one static corpus. Since each tweet only has one author, AT's process on Twitter coincides with the LDA modeling process on all tweets written by a specific user. As an extension of the author-topic model, Twitter-LDA **(TLDA)** classifies topics into private topics and background topic by introducing one binomial distribution. For comparison, we use one more state-of-the-art use behavior model, **UBM** [250]; here, a user's interest is tracked by a mixture graphical model that considers background knowledge, social interests and the user's own interest. The final baseline that we consider is **TF-IDF**, which uses TF-IDF to re-calculate $\mathcal{S}_{RT_{u,t}}$ in Algorithm 3. Finally, we also use **SUM-TF**, a baseline used in [41] that extract tweets by ranking *tf* scores, and **Random**, which extracts tweets randomly in each period.

For the baseline topic models, we use a similar tweet selection method as in Algorithm 3 to select tweets in each time interval. For static topic models, results at time $t, 1 \leq t \leq T$ are calculated after re-modeling for all past data before period t.

To evaluate the effectiveness of results to *personalized* aspect, we introduce several other sentence extraction procedures from the area of document summarization (without personalization) as baselines: **LexRank** and **Centroid** are two widely-used unsupervised document summarization methods, where **LexRank** [199] is a graph-based method for ranking tweet as "votes" from other tweets, and **Centroid** [185] applies the MEAD summarization method that uses statistical and structural features in tweets selection.

3.3.5 Granularities and number of topics

To test the optimal granularity of time intervals, we examine ROUGE-1 performance of TPM-ALL with different values for granularities, shown in Figure 3.6. The performance of TPM-ALL in terms of ROUGE-1 peaks when the granularity is set to 7 days.

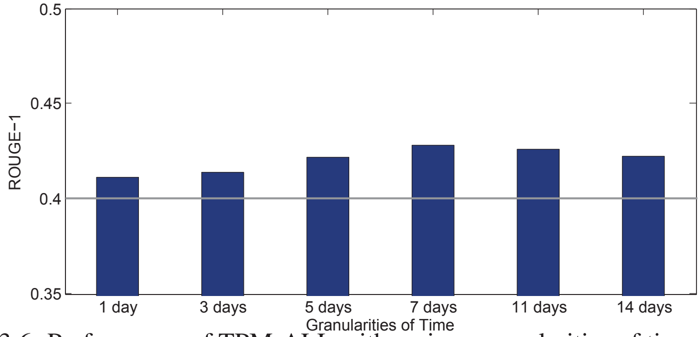


Figure 3.6: Performance of TPM-ALL with various granularities of time periods.

With fewer than 7 days, performance keeps increasing because adding more days reduces sparseness; but after 7 days, due to the increase in irrelevant and noisy tweets, the ROUGE-1 score decrease. Thus, we set the granularity to 7 days in the remainder of our experiments.

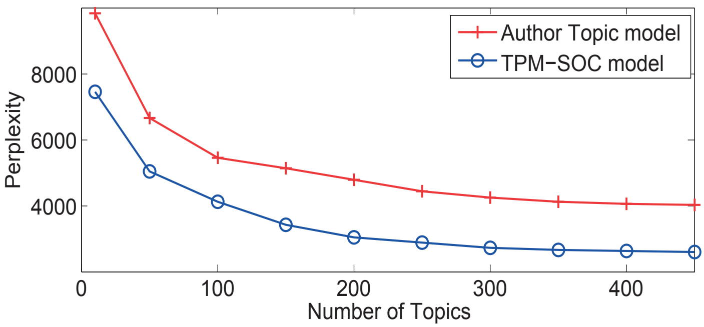


Figure 3.7: Perplexity performance with different number of topics in Author topic model and TPM-SOC model;

Optimizing the number of topics is a problem shared between all topic modeling approaches. Similar to previous work [32, 84, 250], we introduce the perplexity of a held-out test set to evaluate the performance of our topic models. The perplexity, usually used in language modeling, focuses on the inverse of the geometric mean per-word likelihood, which is calculated as follows:

$$Perplexity(W) = \exp\left[-\frac{\sum_{t\in T}\sum_{w\in\mathcal{D}_t}\log p(w)}{\sum_{t\in T} d_t}\right], \tag{3.14}$$

where $p(w)$ indicates $p(w) = p(w \mid z)p(z)$. Thus, a lower perplexity score indicates a better generalization performance [32]. Figure 3.7 shows the results of perplexity values for the author-topic model and the TPM-SOC model with differing numbers of topics on our held-out test set. After the number of topics becomes larger than 300, the perplexity of both approaches starts to flatten out. We find that TPM-SOC outperforms the author-topic model with better generalization performance. For TPM-ALL and TPM-TOP we set the number of "private" topics and "common" topics to 150, separately.

3.4 Results and Discussion

In this chapter, we divide our main research question **RQ1** into multiple research questions **RQ1.1**–**RQ1.3** that guide the remainder of the chapter:

RQ1.1 How does the TPM-based TaTS strategy perform on time-aware tweets summarization? (See §3.4.1.)

RQ1.2 How does the TPM-based TaTS strategy perform on social-aware tweets summarization? (See §3.4.2.)

RQ1.3 what is the overall performance for TPM on the task of personalized TaTS? (See §3.4.3.)

3.4.1 Time-aware comparisons

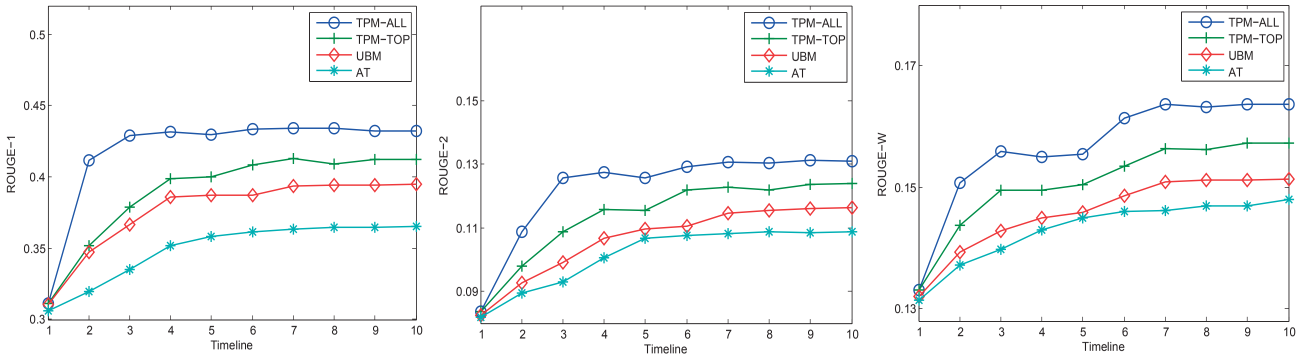
To illustrate the performance at different time periods, the evaluation results of the TPM-ALL, TPM-TOP, UBM and AT strategies at different time periods are shown in Figure 3.8, in terms of ROUGE-1, ROUGE-2 and ROUGE-W, respectively. We select 10 contiguous weeks from November 1, 2009 onwards as the test period and separate it into 10 periods.

In Figure 3.8 we observe that the AT model obtains the worst performance, while both TPM-ALL and TPM-TOP outperform all other strategies in terms of ROUGE metrics at all time intervals. This demonstrates the advantage of TPM-based strategies in time-aware comparisons. In Figure 3.8, we observe a "cold-start" phenomenon, which results from the sparseness of the context in the first time period. In that condition, TPM-ALL and TPM-TOP are nearly equivalent to the UBM and AT since there are neither social circles nor burst topics during the first time period. After that, the performance of the TPM based methods keeps increasing over time until it achieves a stable performance after $t = 3$. We find that TPM based strategies are sensitive to time-aware topic drifting. Meanwhile, we find that TPM-ALL performs better than TPM-TOP in Figure 3.8. TPM-ALL detects user's interests using social circles whereas TPM-TOP ignores them.

3.4.2 Social-aware comparisons

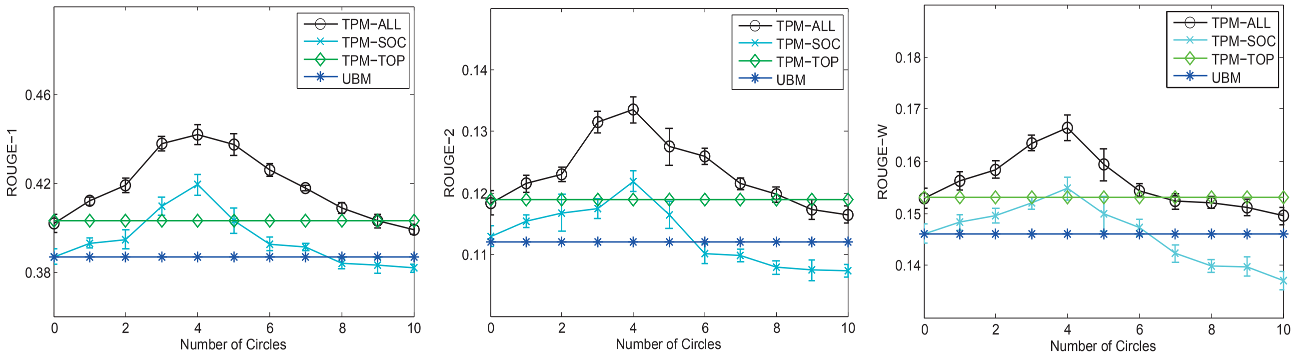
To evaluate the influence of social circles in our proposed strategy, we investigate the performance under various numbers of social circles. From our dataset, we extract users with different numbers of social circles and compare the performance of our methods on these data sets in terms of ROUGE. In Figure 3.9 we plot the values of ROUGE-1, ROUGE-2 and ROUGE-W in (a) to (c), respectively. For each figure, we compare our strategies that do consider social circles, TPM-ALL and TPM-SOC, against the TPM-TOP and UBM methods under varying number of social circles.

We observe from Figure 3.9(a) that the performance in terms of ROUGE-1 changes with the number of social circles, and the value increases and achieves a maximal value between 3 and 5 social circles. After that, the value decreases rapidly; redundant and



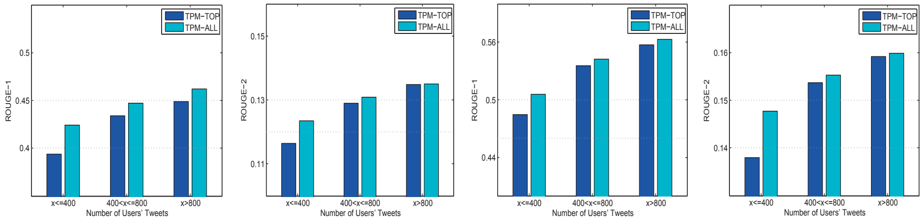
(a) ROUGE-1 (b) ROUGE-2 (c) ROUGE-W

Figure 3.8: **RQ1.1:** Time-aware performance in terms of ROUGE metrics.



(a) ROUGE-1 (b) ROUGE-2 (c) ROUGE-W

Figure 3.9: **RQ1.2:** ROUGE-1 and ROUGE-2 performance with increasing numbers of social circles.



(a) ROUGE-1, N=40 (b) ROUGE-2, N=40 (c) ROUGE-1, N=60 (d) ROUGE-2, N=60

Figure 3.10: **RQ1.2:** Performance for different kinds of users: users in our dataset are classified by their number of tweets.

Table 3.2: **RQ1.3:** Overall ROUGE Performance for All Comparisons

Metrics	TPM-ALL	TPM-TOP	TPM-SOC	UBM	TLDA	AT	TF-IDF	Centroid	Lex-R	SUM-TF	Random
Cut-off of $N = 40$ tweets per period											
ROUGE-1	**0.428** ▲	0.403 ▲	0.395 ▲	0.387	0.374	0.355	0.341	0.302	0.291	0.274	0.252
ROUGE-2	**0.125** ▲	0.119 △	0.116 △	0.114	0.112	0.102	0.095	0.081	0.077	0.079	0.037
ROUGE-W	**0.159** ▲	0.153 ▲	0.149 △	0.146	0.142	0.144	0.137	0.118	0.115	0.105	0.076
Cut-off of $N = 60$ tweets per period											
ROUGE-1	**0.513** ▲	0.497 ▲	0.482 ▲	0.461	0.457	0.423	0.411	0.362	0.369	0.329	0.281
ROUGE-2	**0.149** ▲	0.142 △	0.139 △	0.134	0.127	0.122	0.116	0.097	0.102	0.095	0.041
ROUGE-W	**0.197** ▲	0.191 ▲	0.189 ▲	0.178	0.176	0.166	0.161	0.131	0.135	0.119	0.081

irrelevant "relations" seem to enter the picture. Another plausible explanation concerns the difference of user characteristics in various social circles. Since the UBM and TPM-TOP models do not consider the social influence, their ROUGE values keep constant for different numbers of social circles. We observe a similar behavior in Figure 3.9(b) and 3.9(c) in terms of ROUGE-2 and ROUGE-W.

To evaluate the effect of collaborative filtering in TPM for various classes of users, especially for "passive" users on Twitter who rarely write a tweet, we compare the performance of different users in terms of ROUGE metrics with varying values of the number of tweets selected per period (40 or 60). We separate users into 3 classes by counting their tweets: (1) less than 400 tweets; (2) between 400 to 800; and (3) more than 800 tweets. As shown in Figure 3.10(a) and (c) that focusing on ROUGE-1, the difference between TPM-ALL and TPM-TOP is bigger for users with up to 400 tweets than for those with more than 400. This can be explained by the fact that the collaborative filtering used in TPM-ALL becomes more effective when there is a bigger data sparseness issue to overcome. In terms of ROUGE-2, similar results can be found in Figure 3.10(b) and (d).

3.4.3 Overall performance

Table 3.2 shows the average performance of our TPM-based strategies and baselines, in terms of ROUGE-1, ROUGE-2 and ROUGE-W, based on all candidate tweets in all time periods. We find that our method outperforms the baselines in every case. Except for our TPM-based strategies, UBM get the best performance than others. Since summarization baselines are not sensitive to users' interests, thus we find that Centroid, Lex-R (short for

LexRank), and SUM-TF do not perform well. Among the topic models, we found that the AT-based method yields almost the worst performance. This can be explained by the fact that the topic modeling procedure in AT does not capture topic drift and users' social circles.

We evaluated the performance of the various approaches in terms of the three ROUGE metrics for a varying number of tweets selected per period, i.e., $N = 40$ and $N = 60$. As shown in Table 3.2, TPM-ALL performs better than all baselines on all metrics. For $N = 40$, TPM-ALL achieves an increase of 10.6%, 11.6% and 8.9% over UBM in terms of ROUGE-1, ROUGE-2, and ROUGE-W respectively. For $N = 60$, TPM-ALL gives an increase of 11.2%, 11.2% and 10.1% over UBM. For the dynamic version without social influence, TPM-TOP outperforms all other baselines also, which indicates the effectiveness of detecting dynamic topics. We further compare TPM-TOP with UBM: for $N = 40$, TPM-TOP offers relative performance improvements of 4.1%, 6.25% and 4.8%, respectively, for the ROUGE-1, ROUGE-2 and ROUGE-W metrics, while the relative improvements are 7.8%, 6.7% and 7.3% on the same metrics for $N = 60$. We find that TPM-ALL outperforms the UBM baselines with a statistical significance difference at level $\alpha < 0.01$ in terms of all ROUGE metrics, whereas TPM-TOP and TPM-SOC outperforms UBM with a statistical significance difference at level $\alpha < 0.05$.

3.5 Conclusion

We have considered the task of personalized time-aware tweets summarization, based on user history and influences from "social circles." To handle the dynamic nature of topics and user interests along with the relative sparseness of individual messages, we have proposed a time-aware user behavior model. Based on probabilistic distributions from our proposed topic model, the tweets propagation model, we have introduced an iterative optimization algorithm to select tweets subject to three key criteria: novelty, coverage and diversity. In our experiments, we have provided answers to the main research question raised at the beginning of this chapter:

RQ1: How can we adapt tweets summarization to a specific user based on a user's history and collaborative social influences? Is it possible to explicitly model the temporal nature of microblogging environment in personalized tweets summarization?

To answer this question, we employ a Twitter dataset that includes both social relations and tweets. In our experiments, we illustrate the performance of our methods and baselines at different time periods. To evaluate the influence of social circles in our proposed strategy, we also investigate the performance of our methods and other baselines under various numbers of social circles. Our experiments have verified the effectiveness of our proposed method, showing significant improvements over various state-of-the-art baselines.

As to future work, we aim to employ a user-study to enhance the accuracy of interest detection, e.g., via an online evaluation. Another future direction is to take more information and features into account for our task: our current experiments ignore, e.g., URLs appearing in tweets which could enhance our entity linking setup. It will also be interesting to consider other features for modeling, such as geographic or profile information. Finally, our current model is evaluated based on fixed time intervals, which might not accurately reflect bursty topics on Twitter. Therefore, a novel graphical model that includes dynamic time bins instead of the fixed time granularities, will be another direction for future research. In the next chapter, we will turn to summarize contrastive themes for opinionated documents.

4

Contrastive Theme Summarization

In the previous chapter, we studied the task of personalized time-aware tweets summarization by considering the user history and influences from social media. In this chapter, we continue our research on summarization by addressing the task of contrastive theme summarization. In recent years multi-document summarization has become a well studied task for helping users understand a set of documents. Typically, the focus has been on relatively long, factual and grammatically correct documents [39, 95, 125, 190, 211, 241]. However, the web now holds a large number of opinionated documents, especially in social media, e.g., microblogs, question answering platforms and web forum threads. The growth in volume of such opinionated documents on the web motivates the development of methods to facilitate the understanding of subjective viewpoints present in sets of documents.

Given a set of opinionated documents, inspired by Paul et al. [176], we define a *theme* to be a specific set of topics around an explicit sentiment opinion. Given a set of specific topics, two themes are *contrastive* if they are related to the topics, but opposite in terms of sentiment. The phenomenon of contrastive themes is widespread in opinionated web documents [59]. In Figure 4.1 we show an example of three contrastive themes about the "Palestine and Israel relationship." Here, each pair of contrastive themes includes two sentences representing two relevant but opposing themes. In this chapter, our focus is on developing methods for automatically detecting and describing such contrastive themes.

The task on which we focus is *contrastive summarization* [107, 176] of multiple themes. The task is similar to *opinion summarization*, in which opinionated documents are summarized into structured or semi-structured summaries [74, 75, 92, 108]. However, most existing opinion summarization strategies are not adequate for summarizing contrastive themes from a set of unstructured documents. To our knowledge, the most similar task in the literature is the *contrastive viewpoint summarization* task [176], in which the authors extract contrastive but relevant sentences to reflect contrastive topic aspects, which are derived from a latent topic-aspect model [175]. However, their proposed method for *contrastive viewpoint summarization* neglects to explicitly model the number of topics and the relations among topics in contrastive topic modeling—these are two key features in contrastive theme modeling. The specific contrastive summarization task that we address in this chapter is *contrastive theme summarization of multiple opinionated documents*. In our case, the output consists of contrastive sentence pairs that highlight every contrastive theme in the given documents. Therefore, we address the

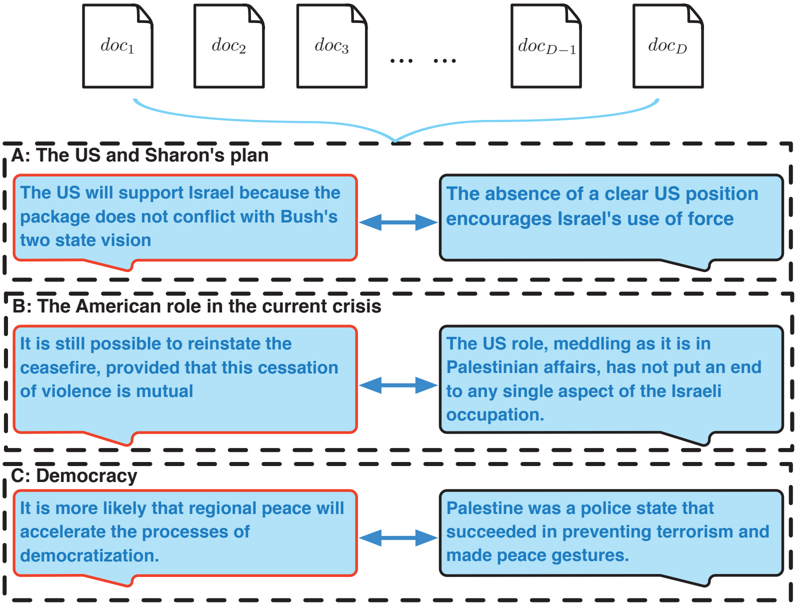


Figure 4.1: Three example contrastive themes related to "Palestine and Israel." Each contrastive theme shows a pair of opposing sentences.

following main research question listed in Chapter 1:

RQ2: How can we optimize the number of topics in contrastive theme summarization of multiple opinionated documents? How can we model the relations among topics in contrastive topic modeling? Can we find an approach to compress the themes into a diverse and salient subsets of themes?

To answer this main research question, we employ a non-parametric strategy based on the nested Chinese restaurant process (nCRP) [33]. Previous work has proved the effectiveness of non-parametric models in topic modeling [7, 187]. But none of them considers the task of contrastive theme summarization. We introduce a topic model that aims to extract contrastive themes and describe hierarchical relations among the underlying topics. Each document in our model is represented by hierarchical threads of topics, whereas a word in each document is assigned a finite mixture of topic paths. We apply collapsed Gibbs sampling to infer approximate posterior distributions of themes.

To enhance the diversity of the contrastive theme modeling, we then proceed as follows. Structured determinantal point processes (SDPPs) [116] are a novel probabilistic strategy to extract diverse and salient threads from large data collections. Given theme distributions obtained via hierarchical sentiment topic modeling, we employ SDPPs to extract a set of diverse and salient themes. Finally, based on themes extracted in the first two steps, we develop an iterative optimization algorithm to generate the final contrastive theme summary. During this process, *relevance*, *diversity* and *contrast* are considered.

Table 4.1: Notation used in this chapter.

Symbol	Description
$\mathcal{D}$	candidate documents
$\mathcal{W}$	vocabulary in corpus $\mathcal{D}$
$\mathcal{K}$	themes set in $\mathcal{D}$
$\mathcal{T}$	themes tuples from $\mathcal{K}$
d	a document, $d \in \mathcal{D}$
s_d	a sentence in document d, i.e., $s_d \in d$
w	a word present in a sentence, $w \in \mathcal{W}$
x	a sentiment label, $x \in \{neg, neu, pos\}$
o_s	sentiment distribution of sentence s
c^x	a topic path under label x
b	a topic node on a topic path
z^x	a topic level under x label
ϕ^x	topic distribution of words, under label x
$k_{c,x}$	a theme corresponding to topic path c, under label x
t	a contrastive theme tuple
θ_d	probability distribution of topic levels over d
$\mathcal{S}_t$	contrastive summary for theme tuple t

Our experimental results, obtained using three publicly available opinionated document datasets, show that contrastive themes can be successfully extracted from a given corpus of opinionated documents. Our proposed method for multiple contrastive themes summarization outperforms state-of-the-art baselines, as measured using ROUGE metrics.

To sum up, our contributions in this chapter are as follows:

- We focus on a contrastive theme summarization task to summarize contrastive themes from a set of opinionated documents.
- We apply a hierarchical non-parametric model to extract contrastive themes for opinionated texts. We tackle the diversification challenge by employing structured determinantal point processes to sample diverse themes.
- Jointly considering relevance, diversity and contrast, we apply an iterative optimization strategy to summarize contrastive themes, which is shown to be effective in our experiments.

We formulate our research problem in §4.1 and describe our approach in §4.2. Then, §4.3 details our experimental setup and §4.4 presents the experimental results. Finally, §4.5 concludes the chapter.

4.1 Problem Formulation

Before introducing our method for contrastive theme summarization, we introduce our notation and key concepts. Table 4.1 lists the notation we use in this chapter.

We have already defined the notion of *topic* in Section 2.5. Given a corpus $\mathcal{D}$, unlike

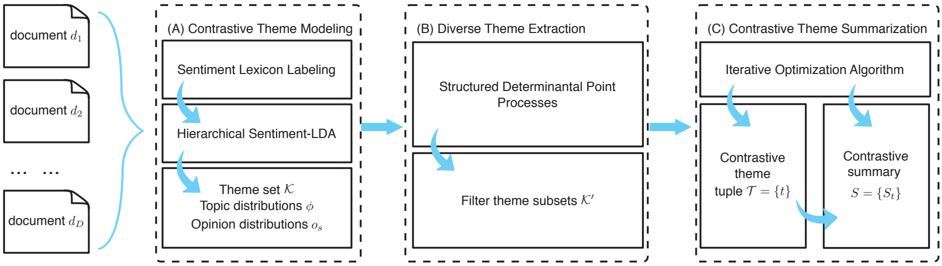


Figure 4.2: Overview of our approach to contrastive theme summarization. (A) indicates contrastive theme modeling; (B) indicates a structured determinantal point process to diversify topics; and (C) refers to the contrastive summary generation algorithm. Crooked arrows indicate the output in each step; while straight arrows indicate processing directions.

"flat" topic models [32], in this chapter we assume that each document d can be represented by multiple topics that are organized in an infinite tree-like hierarchy $c = \{(z_0, c), (z_1, c), \ldots\}$, $z_0 \prec z_1 \prec \ldots$, i.e., c indicates a path from the root topic level z_0 on the infinite tree to more specialized topics that appear at the leaves of the tree, and for each topic level z we define a topic node $b = (z, c)$ on the topic path c. Then, we define the notions of *sentiment* and *theme* in our work.

Sentiment is defined as a probability distribution over sentiment labels *positive*, *negative*, and *neutral*. A *sentiment label* x is attached with each word w. Considering the *sentiment*, we divide topics into three classes: positive topics (2), neutral topics (1) and negative topics (0).

Given all hierarchical topics and sentiment labels, we define a *theme* $k_{c,x}$ as a threaded topic path c from the root level to the leaf level for the given sentiment label x. Let $\mathcal{K}$ be the set of themes, and let $\mathcal{K}^{pos}$, $\mathcal{K}^{neg}$, $\mathcal{K}^{neu}$ indicate the set of positive, negative and neutral themes, respectively, i.e., $\mathcal{K} = \mathcal{K}^{pos} \cup \mathcal{K}^{neg} \cup \mathcal{K}^{neu}$. Furthermore, we define a *contrastive theme* to be a theme tuple $t = (c^{pos}, c^{neg}, c^{neu})$ by extracting themes fromis contained in $\mathcal{K}^{pos} \times \mathcal{K}^{neg} \times \mathcal{K}^{neu}$. Themes c^{pos}, c^{neg} and c^{neu} in each tuple t are relevant in topic but opposite in sentiment labels.

Finally, we define contrastive theme summarization. Given a set of documents $\mathcal{D} = \{d_1, d_2, ..., d_D\}$, the purpose of the contrastive theme summarization task (CTS) is to select a set of meaningful sentences $\mathcal{S}_t = \{S_{c^{pos}}, S_{c^{neg}}, S_{c^{neu}}\}$ to reflect the representative information in each possible theme tuple $t = (c^{pos}, c^{neg}, c^{neu})$.

4.2 Method

4.2.1 Overview

We provide a general overview of our method for performing contrastive theme summarization (CTS) in Figure 4.2. There are three main phases: (A) contrastive theme modeling; (B) diverse theme extraction; and (C) contrastive theme summarization. To summarize, we are given a set of documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ as input. For each document $d \in \mathcal{D}$, in phase (A) (see §4.2.2), we obtain a structured themes set $\mathcal{K}$ with a root node r, topic distributions ϕ and opinion distributions o_s.

In (B) (see §4.2.3), given the structured output themes $\mathcal{K}$, we employ a structured determinantal point process to obtain a subset $\mathcal{K}' \subseteq \mathcal{K}$ to enhance the saliency and diversity among themes.

Based on themes $\mathcal{K}'$ and their corresponding topic distributions and opinion distributions, in (C) (see §4.2.4) we generate the final contrastive theme summary $\mathcal{S}$. We develop an iterative optimization algorithm for this process: the first part in §4.2.4 is to generate the contrastive theme tuples $\mathcal{T}$, each of which includes relevant themes for a topic but contrastive in sentiment; the second part in §4.2.4 is meant to generate the final contrastive summary $\mathcal{S} = \{S_t\}$ for each theme tuple.

4.2.2 (A) Contrastive theme modeling

We start by proposing a hierarchical sentiment-LDA model to jointly extract topics and opinions from our input corpus. Unlike previous work on traditional "flat" topic models [176], our method can adaptively generate topics organized in a tree-like hierarchy.

Briefly, each document $d \in \mathcal{D}$ can be represented as a collection of sentences, whereas each sentence $s \in d$ is composed of a collection of words. By using a state-of-the-art sentiment analysis method [219], for each word w in each document d we extract its sentiment label x_w, where $x_w \in \{pos, neu, neg\}$. Generally, for document d we select three threaded topic paths $\{c^x\}$, with $x = pos$, neu, neg, each of which is generated by a nested Chinese restaurant process (nCRP) [33]. After deriving the sentiment label x, each word $w \in d$ is assigned to a specific topic level z by traversing from the root to the leave on the path c^x.

Next, we give a more detailed technical account of our model. Following the nested Chinese restaurant process [33], our topic model identifies documents with threaded topic paths generated by nCRP. Given level z, we consider each node (z, c) on a threaded topic path c as a specific topic. To select the exact topic level $z \in [1, L]$, we draw a variable θ_d from a Dirichlet distribution derived from hyperparameter m, to define a probability distribution on topic levels along the topic path c. Given a draw from a Dirichlet distribution, document d is generated by repeatedly selecting a topic level. We assume that each document $d \in \mathcal{D}$ is represented by three classes of topics: positive, negative and neutral topics.

In document d, for each sentence $s \in d$ we define a sentiment distribution o_s from a Dirichlet distribution over a hyper parameter γ. For each word $w \in \mathcal{W}$, we select three topic levels z^{pos}, z^{neg} and z^{neu} from a discrete distribution over θ_d, respectively. While the sentiment label is derived from a multinomial distribution over o_s, w is derived from a discrete distribution over topic levels $\{z^{pos}, z^{neg}, z^{neu}\}$. The generation process of our proposed model is shown in Figure 4.3.

Since exact posterior inference in hierarchical sentiment-LDA is intractable, we employ a collapsed Gibbs sampler to approximate the posterior distributions of topic level z_w for each word w and topic path c_d for each document d. In our model, two sets of variables are observed: the sentiment labels x_w for each word w, and the words set $\mathcal{W}$. Our sampling procedure is divided into two steps for each iteration: (1) sampling a topic path for each document; (2) sampling level allocation for each word.

1. For each topic level $z^x \in \mathcal{Z}^x$ in infinite tree:
   - Draw $\phi^x \sim Dirichlet(\beta^x)$;
2. For each document $d \in \mathcal{D}$:
   - Draw $c_d^x \sim nCRP(p)$;
   - Draw $\theta_d \sim Dirichlet(m)$;
   - For each sentence $s \in d$:
     - Draw opinion $o_s \sim Dirichlet(\gamma)$;
     - For each word $w \in N$:
       * Draw sentiment $x \sim Multinomial(o_s)$;
       * Draw topics $z^x \sim Discrete(\theta_d)$:
       * Draw word $w \sim Discrete(\phi_{z^x, c_d^x})$;

Figure 4.3: Generative process in hierarchical sentiment-LDA.

For the sampling procedure of thread c_d, given current other variables on document d, we have:

$$p(c_d^x \mid c_{-d}^x, z, o) \propto p(c_d^x \mid c_{-d}^x) \cdot p(W_d \mid W_{-d}, c, x, o, z) \tag{4.1}$$

where $p(c_d^x \mid c_{-d}^x)$ in (4.1) is the prior distribution implied by the nested Chinese restaurant process, whereas for each topic node (z, c_d) on path c_d, we have:

$$\begin{cases} P((z, c_d) = b_i) = \frac{n_i}{n+p-1} \\ P((z, c_d) = b_{new}) = \frac{p}{n+p-1} \end{cases} \tag{4.2}$$

where b_i indicates a node that has been taken before, b_{new} indicates a new node that has not been considered yet; n_i refers to the number of times that topic node (z, c_d) is assigned to a document. To infer $p(W_d \mid W_{-d}, c, x, o, z)$, we integrate over multinomial parameters and have:

$$p(W_d \mid W_{-d}, c, x, o, z) \propto \prod_{z=1}^{L} \frac{\Gamma(n_{-d}^{z,c} + W\beta)}{\prod_{w \in W} \Gamma(n_{w,-d}^{z,c} + \beta)} \frac{\prod_{w \in W} \Gamma(n_{w,-d}^{z,c} + n_{w,d}^{z,c} + \beta)}{\Gamma(n_{-d}^{z,c} + n_d^{z,c} + W\beta)} \prod_{s \in S_d} \frac{\prod_{x \in X} \Gamma(n_{s,x} + \gamma_x)}{\Gamma(n_s + \gamma)}, \tag{4.3}$$

where $n_{-d}^{z,c}$ indicates the number of times that documents have been assigned to topic node (z, c) leaving out document d; $n_{w,-d}^{z,c}$ denotes the number of times that word w has been assigned to the topic node (z, c) leaving out document d.

To sample topic level $z_{d,n}$ for each word w_n in document d, we find its joint probabilistic distribution of terms, sentiment labels and topics as follows:

$$p(z_{d,n}^x = \eta | z_{-(d,n)}^x, c^x, x, o, w) \propto \frac{n_{w_n,-n}^{\eta,c} + \beta}{n_{-n}^{\eta,c} + W\beta} \frac{n_d^{\eta} + m}{n_{d,-n}^{\eta} + Lm} \frac{\prod_{x \in X} \Gamma(n_{s,x} + \gamma_x)}{\Gamma(n_s + \gamma)}, \tag{4.4}$$

where $z_{-(d,n)}^x$ denotes the vectors of level allocations leaving out $z_{d,n}^x$ in document d. Further, $n_{w_n,-n}^{\eta,c}$ denotes the number of times that words have been assigned to topic node

(η, c) that are the same as word w_n; $n^{\eta}_{d,-n}$ denotes the number of times that document d have been assigned to level k leaving out word w_n.

After Gibbs sampling, we get a set of topic paths $\{c^x\}$ that can be represented as themes $\mathcal{K} = \{k_{c,x}\}$; for each word w in d, we have hybrid parametric distributions ϕ^x that reflect the topic distribution given a specific level z on path c, i.e., $P(w, x \mid c, z) = \phi^x_{z,c,w}$. For each sentence s, we have a probability distribution o_s over sentiment labels, i.e., $P(x \mid s) = o_{s,x}$.

4.2.3 (B) Diverse theme extraction

Given a set of themes $\mathcal{K} = \{k_{c,x}\}$ resulting from step (A), some further issues need to be tackled before we arrive at our desired summary. On the one hand, many themes in $\mathcal{K}$ share common topics; on the other hand, many words' topic probabilities ϕ are similar, which makes it difficult to distinguish the importance of the themes.

To address this dual problem, we employ the structured determinantal point process (SDPP) [117] to select a subset of salient and diverse themes from $\mathcal{K}$. We already have introduced the background for the determinantal point process (DPP) and the structured determinantal point process (SDPP) in Section 2.6. Following [116], here we define a structured determinantal point process $\mathcal{P}$ as a type of probability distribution over a subset of themes belonging to $\mathcal{K}$. Two main factors are considered in SDPPs: the *quality* q_i and the *similarity* ${\varphi_i}^T\varphi_j$. A subset with high quality and highly diverse themes will be assigned the highest probability $\mathcal{P}$ by the SDPPs.

Given themes $\mathcal{K}$ sampled from (A), we proceed as follows. Firstly, for each theme $k \in \mathcal{K}$ we use $q((z_i, c))$ to indicate the "quality" of topic $(z_i, c) \in k$ and we use $\varphi((z_i, c))^T\ \varphi((z_j, c')) \in [0, 1]$ to refer to a measure of similarity between two topics (z_i, c) and (z_j, c'):

$$
\begin{aligned}
q((z_i, c)) &= \sum_{w \in \mathcal{W}_H} \phi_{z_i,c,w} \\
\varphi((z_i, c))^T \varphi((z_j, c')) &= \exp\left(-\frac{\left\|\Phi_{z_i,c} - \Phi_{z_j,c'}\right\|_2^2}{2\sigma^2}\right),
\end{aligned}
\tag{4.5}
$$

where $\Phi_{z_i,c}$ indicates the vector $\{\phi_{z,c,w}\}_{w \in \mathcal{W}}$; $\left\|\Phi_{z_i,c} - \Phi_{z_j,c'}\right\|_2^2$ is the squared Euclidean distance between $\Phi_{z_i,c}$ and $\Phi_{z_j,c'}$; $\mathcal{W}_H$ indicates the top-n salient words; σ is a free parameter. Based on Eq. 2.4 and Eq. 2.5 in Chapter 2, we construct the semidefinite matrix $\mathcal{M}$ for SDPPs.

For two topic paths $c_i = \{(z_1, c_i), \ldots, (z_L, c_i)\}$ and $c_j = \{(z'_1, c_j), \ldots, (z_L, c_j)\}$, $c_i, c_j \in \mathcal{K}$, we assume a factorization of the quality $q(c)$ and similarity score $\varphi(c_i, c_j)$ into parts, decomposing quality multiplicatively and similarity additively, i.e., for topic paths c_i and c_j, $q(c_i)$ and $\varphi(c_i, c_j)$ are calculated by Eq. 2.6, respectively.

To infer the posterior results of SDPPs over themes, we adapt an efficient sampling algorithm as described in Algorithm 4. Following [116], we let $\mathcal{M} = \sum_{k=1}^{K} \lambda_k v_k v_k^T$ be an orthonormal eigen-decomposition, and let e_i be the ith standard basis K-vector. The sampling algorithm of SDPPs outputs a subset of themes, i.e., $\mathcal{K}' = \{k'_{c,x}\}$, which reflect a trade-off between high quality and high diversity.

Algorithm 4: Sampling process for SDPPs

Input : Eigenvector/values pairs $\{(v_k, \lambda_k)\}$; Themes set $\mathcal{K}$;
Output: Filtered themes set $\mathcal{K}'$ from SDPPs;
$\mathcal{J} \leftarrow 0$; $\mathcal{K}' \leftarrow 0$;
for $k \in \mathcal{K}$ **do**
  $\mathcal{J} \leftarrow \mathcal{J} \cup \{k\}$ with probability $\frac{\lambda_k}{1+\lambda_k}$;
end
$V \leftarrow \{v_k\}_{k \in \mathcal{J}}$;
while $|V| > 0$ **do**
  Select k_i from $\mathcal{K}$ with $P(k_i) = \frac{1}{|V|} \sum\limits_{v \in V} (v^T e_i)^2$;
  $\mathcal{K}' \leftarrow \mathcal{K}' \cup k_i$;
  $V \leftarrow V_\perp$ as an orthonormal basis for the subspace of V orthonormal to e_i;
end
return $\mathcal{K}'$.

4.2.4 (C) Contrastive theme summarization

In this section, we specify the sentence selection procedure for contrastive themes. Considering the diversity among topics, we only consider leaf topics in each theme $k'_{c,x} \in \mathcal{K}'$. Thus, each theme $k'_{c,x}$ can be represented by a leaf topic (z^x_L, c^x) exclusively. For simplicity, we abbreviate leaf topics sets $\{(z^x_L, c^x)\}$ as $\{c^x\}$.

Given $\{c^x\}$, we need to connect topics in various classes to a set of contrastive theme tuples of the form $t = (c^{pos}_i, c^{neg}_{ii}, c^{(neu)}_{iii})$. To assess the correlation between two topics (c^x_i) and (c^y_{ii}) in different classes, we define a *correlation* based on topic distributions $\Phi_{z,c}$ as follows:

$$1 - \frac{1}{N} \sum_{d \in \mathcal{D}} \left| \sum_{w \in d} \phi_{z_L, c^x_i, w} - \sum_{w' \in d} \phi_{z_L, c^y_{ii}, w'} \right|. \tag{4.6}$$

We sample three leaf topics from the three classes mentioned earlier (positive, negative and neutral), so that the total *correlation* values for all three topic pairs has maximal values. Next, we extract representative sentences for each contrastive theme tuple $t = (c^{pos}_i, c^{neu}_{ii}, c^{neg}_{iii})$. An intuitive way for generating the contrastive theme summary is to extract the most salient sentences as a summary. However, high-degree topical relevance cannot be taken as the only criterion for sentence selection. To extract a contrastive theme summary $\mathcal{S}_t = \{S_{c^{pos}_i}, S_{c^{neu}_{ii}}, S_{c^{neg}_{iii}}\}$ for tuple $t = (c^{pos}_i, c^{neu}_{ii}, c^{neg}_{iii})$, in addition to *relevance* we consider two more key requirements *contrast* and *diversity*. Given selected sentences $\mathcal{S}'_t$, we define a salient score $F(s_i|\mathcal{S}'_c, t)$:

$$F(s_i \mid \mathcal{S}'_t, t) = ctr(s_i|\mathcal{S}'_t, t) + div(s_i, \mathcal{S}'_t) + rel(s_i \mid t), \tag{4.7}$$

where $ctr(s_i \mid \mathcal{S}'_t, t)$ indicates the contrast between s_i and $\mathcal{S}'_t$ for t; $div(s_i, \mathcal{S}'_t)$ indicates the divergence between s_i and $\mathcal{S}'_t$; $rel(s_i \mid t)$ indicates the relevance of s_i given t.

Contrast calculates the sentiment divergence between the currently selected sentence s_i and the results of extracted sentences set $\mathcal{S}'_t$, under the given theme t. Our intention is

Algorithm 5: Iterative process for generating the summary $\mathcal{S}$.

Input : $\mathcal{T} = \{(c_i^{pos}, c_{ii}^{neg}, c_{iii}^{neu})\}$, μ,π,S,N;
Output: $\mathcal{S} = \left\{\{S_{c_i^{pos}}, S_{c_{ii}^{neg}}, S_{c_{iii}^{neu}}\}_{(t)}\right\}$;

for *each* $t = (c_i^{pos}, c_{ii}^{neg}, c_{iii}^{neu})$ **do**
  Rank and extract relevant sentences to $\mathcal{C}$ by $rel(s \mid t)$;
  Initialize: Extract $\frac{N}{|\mathcal{T}|}$ sentences from $\mathcal{C}$ to $\mathcal{S}_t$;
  repeat
    Extract $\mathcal{X} = \{s_x \in \mathcal{C} \wedge \neg \mathcal{S}_t\}$;
    for $s_x \in \mathcal{X}, \forall s_y \in \mathcal{S}_t$ **do**
      Calculate $\mathcal{L} = \sum_{s_i \in \mathcal{S}_t} F(s_i \mid \mathcal{S}_t, t)$;
      Calculate $\Delta \mathcal{L}_{s_x,s_y} = \mathcal{L}((\mathcal{S}_t - s_y) \cup s_x) - \mathcal{L}(\mathcal{S}_t)$;
    end
    Get $\langle \hat{s_x}, \hat{s_y} \rangle$ that $\langle \hat{s_x}, \hat{s_y} \rangle = \arg\max_{s_x,s_y} \Delta \mathcal{L}_{s_x,s_y}$;
    $\mathcal{S}_t = (\mathcal{S}_t - \hat{s_y}) \cup \hat{s_x}$;
  until $\forall \Delta S_{s_x,s_y} < \varepsilon$;
  $\mathcal{S} = \mathcal{S} \cup \mathcal{S}_t$;
end
return $\mathcal{S}$.

to make the current sentence as contrastive as possible from extracted sentences as much as possible. Therefore, we have:

$$ctr(s_i \mid \mathcal{S}'_t, t) = \max_{\{s \in \mathcal{S}'_t, x\}} \left|(o_{s_i,x} - o_{s,x}) \cdot (\phi^x_{z_L,c,w} - \phi^x_{z_L,c,w})\right|, \tag{4.8}$$

Diversity calculates the information divergence among all sentences within the current candidate result set. Ideally, the contrastive summary results have the largest possible difference in theme distributions with each other. The equation is as follows:

$$div(s_i \mid \mathcal{S}'_t) = \max_{s \in \mathcal{S}'_t} |rel(s_i \mid t) - rel(s \mid t)|, \tag{4.9}$$

Furthermore, a contrastive summary should contain relevant sentences for each theme t, and minimize the information loss with the set of all candidate sentences. Thus, given $\phi^x_{z_L,c,w}$, the *relevance* of sentence s_i given theme t is calculated as follows:

$$rel(s_i \mid t) = \frac{1}{N_{s_i}} \sum_x \sum_{w \in s_i} \phi^x_{z_L,c,w}, \tag{4.10}$$

Algorithm 5 shows the details of our sentence extraction procedure.

4.3 Experimental Setup

4.3.1 Research questions

We divide our main research question **RQ2** into research questions **RQ2.1–RQ2.4** that guide the remainder of the chapter.

RQ2.1 Is hierarchical sentiment-LDA effective for extracting contrastive themes from documents? (See §4.4.1.) Is hierarchical sentiment-LDA helpful for optimizing the number of topics during contrastive theme modeling? (See §4.4.2.)

RQ2.2 Is the structured determinantal point process helpful for compressing the themes into a diverse and salient subset of themes? (See §4.4.2 and §4.4.3.) What is the effect of SDPP in contrastive theme modeling? (See §4.4.3).

RQ2.3 How does our iterative optimization algorithm perform on contrastive theme summarization? Does it outperform baselines? (See §4.4.4.)

RQ2.4 What is the effect of *contrast*, *diversity* and *relevance* for contrastive theme summarization in our method? (See §4.4.5.)

4.3.2 Datasets

We employ three datasets in our experiments. Two of them have been used in previous work [175, 176], and another one is extracted from news articles of the New York Times.[1] All documents in our datasets are written in English. All three datasets include human-made summaries, which are considered as ground-truth in our experiments. As an example, Table 4.2 shows statistics of 15 themes from the three datasets that include the largest number of articles in our dataset. In total, $15,736$ articles are used in our experiments.

The first dataset ("dataset 1" in Table 4.2) consists of documents from a Gallup[2] phone survey about the 2010 U.S. healthcare bill. It contains 948 verbatim responses, collected March 4–7, 2010. Respondents indicate if they are "for" or "against" the bill, and there is a roughly even mix of the two opinions (45% for and 48% against). Each document in this dataset only includes 1–2 sentences.

Our second dataset ("dataset 2") is extracted from the Bitterlemons corpus, which is a collection of 594 opinionated blog articles about the Israel-Palestine conflict. The Bitterlemons corpus consists of the articles published on the Bitterlemons website[3] from late 2001 to early 2005. This dataset has also been applied in previous work [136, 175]. Unlike the first dataset, this dataset contains long opinionated articles with well-formed sentences. It too contains a fairly even mixture of two different perspectives: 312 articles from Israeli authors and 282 articles from Palestinian authors.

Our third dataset ("dataset 3") is a set of articles from the New York Times. The New York Times Corpus contains over 1.8 million articles written and published between January 1, 1987 and June 19, 2007. Over 650,000 articles have manually written article summaries. In our experiments, we only use *Opinion* column articles that were published during 2004–2007.

4.3.3 Baselines and comparisons

We list the methods and baselines that we consider in Table 4.3. We write HSDPP for the overall process as described in Section 4.2, which includes steps (A) contrastive theme modeling, (B) diverse theme extraction and (C) contrastive theme summarization. We

[1] http://ilps.science.uva.nl/resources/nyt_cts
[2] http://www.gallup.com/home.aspx
[3] http://www.bitterlemons.org

Table 4.2: Top 15 topics in our three datasets. Column 1 shows the name of topic; column 2 shows the number of articles included in the topic; column 3 shows the publication period of those articles, and column 4 indicates to which dataset the topic belongs.

General description	# articles	Period	Dataset
U.S. International Relations	3121	2004–2007	3
Terrorism	2709	2004–2007	3
Presidential Election of 2004	1686	2004	3
U.S. Healthcare Bill	940	2010	1
Budgets & Budgeting	852	2004–2007	3
Israel-Palestine conflict	594	2001–2005	2
Airlines & Airplanes	540	2004–2007	3
Colleges and Universities	490	2004–2007	3
Freedom and Human Rights	442	2004–2007	3
Children and Youth	424	2004–2007	3
Computers and the Internet	395	2004–2007	3
Atomic Weapons	362	2004–2005	3
Books and Literature	274	2004–2007	3
Abortion	170	2004–2007	3
Biological and Chemical Warfare	152	2004–2006	3

write HSLDA for the model that only considers steps (A) and (C), so skipping the structured determinantal point processes in (B). To evaluate the effect of *contrast*, *relevance* and *diversity*, we consider HSDPPC, the method that only considers *contrast* in contrastive theme summarization. We write HSDPPR for the method that only considers *relevance* and HSDPPD for the method that only considers *diversity* in the summarization.

To assess the contribution of our proposed methods, our baselines include recent related work. For contrastive theme modeling, we use the Topic-aspect model (TAM, [175]) and the Sentiment-topic model (Sen-TM, [124]) as baselines for topic models. Both focus on the joint process between topics and opinions. Other topic models, such as Latent dirichlet allocation (LDA) [32] and hierarchical latent dirichlet allocation (HLDA) [33], are also considered in our experiments. For the above "flat" topic models, we evaluate their performance using varying numbers of topics (10, 30 and 50 respectively). The number of topics used will be shown as a suffix to the model's name, e.g., TAM-10.

We also consider previous document summarization work as baselines: (1) A depth-first search strategy (DFS, [75]) based on our topic model. (2) The LexRank algorithm [63] that ranks sentences via a Markov random walk strategy. (3) ClusterCMRW [241] that ranks sentences via a clustering-based method. (4) Random, which extracts sentences randomly.

4.3.4 Experimental setup

Following existing models, we set pre-defined values for some parameters in our proposed method. In our proposed hierarchical sentiment-LDA model, we set m as 0.1 and

Table 4.3: Our methods and baselines used for comparison.

Acronym	Gloss	Reference
HSDPPC	HSDPP only considering *contrast* in (C) contrastive theme summarization	This chapter
HSDPPR	HSDPP only considering *relevance* in (C) contrastive theme summarization	This chapter
HSDPPD	HSDPP only considering *diversity* in (C) contrastive theme summarization	This chapter
HSLDA	Contrastive theme summarization method in (C) with HSLDA, without SDPPs	This chapter
HSDPP	Contrastive theme summarization method in (C) with HSLDA and SDPPs	This chapter
Topic models		
TAM	Topic-aspect model based contrastive summarization	[175]
Sen-TM	Sentiment LDA based contrastive summarization	[124]
LDA	LDA based document summarization	[32]
HLDA	Hierarchical LDA based document summarization	[33]
Summarization		
LexRank	LexRank algorithm for summarization	[63]
DFS	Depth-first search for sentence extraction	[75]
ClusterCMRW	Clustering-based sentence ranking strategy	[241]

γ as 0.33 as default values in our experiments.

Optimizing the number of topics is a problem shared between all topic modeling approaches. In our hierarchical sentiment-LDA model, we set the default length of L to 10, and we discuss it in our experiments. Just like other non-parametric topic models, our HSLDA model optimizes the number of themes automatically. Under the default settings in our topic modeling, we find that for the Gallup investigation data, the optimal number of topics is 23; the Bitterlemons corpus, it is 67; for the New York Times dataset, it is 282.

4.3.5 Evaluation metrics

To assess the saliency of contrastive theme modeling in our experiments, we adapt the *purity* and *accuracy* in our experiments to measure performance. To evaluate the diversity among topics we calculate the *diversity* as follows:

$$diversity = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \max \left| \phi^{x}_{z,c,w} - \phi^{x}_{z',c',w} \right| \tag{4.11}$$

We adopt the ROUGE evaluation metrics [133], a widely-used recall-oriented metric for document summarization that evaluates the overlap between a gold standard and candidate selections. We use ROUGE-1 (R-1, *unigram based method*), ROUGE-2 (R-2, *bigram based method*) and ROUGE-W (R-W, *weighted longest common sequence*) in our experiments.

Table 4.4: Part of an example topic path of hierarchical sentiment-LDA result about "College and University." Columns 2, 3 and 4 list popular positive, neutral and negative terms for each topic level, respectively.

Topic level	Positive	Neutral	Negative
1	favor, agree, accept, character paid, interest, encourage	college, university, university school, editor, year	lost, suffer, fish, wrong, ignore drawn, negative
2	education, grant, financial, benefit save, recent, lend, group	Harvard, president, summer, Lawrence university, faculty, term, elite	foreign, hard, low global, trouble lose, difficulty
3	attract, meaningful, eligible, proud essence, quarrel,qualify	summers, Boston, greek, season seamlessly, opinion, donation	short, pity, unaware, disprove disappoint, idiocy, disaster
4	practical, essay, prospect respect, piously, behoove	write, march, paragraph, analogy analogy, Princeton, english	dark, huge, hassle, poverty depression, inaction, catastrophe
5	grievance, democratic, dignity, elite interest, frippery, youthful	June, volunteer, community, Texas classmate, liberal, egger	cumbersome, inhumane, idiocy, cry mug, humble, hysteria

Statistical significance of observed differences between the performance of two runs is tested using a two-tailed paired t-test and is denoted using $^\blacktriangle$ (or $^\blacktriangledown$) for strong significance for $\alpha = 0.01$; or $^\vartriangle$ (or $^\triangledown$) for weak significance for $\alpha = 0.05$. In our experiments, significant difference are with regard to TAM and TAM-Lex for contrastive theme modeling and contrastive theme summarization, respectively.

4.4 Results and Discussion

4.4.1 Contrastive theme modeling

We start by addressing **RQ2.1** and test whether HSLDA and HSDPP are effective for the contrastive theme modeling task. First, Table 4.4 shows an example topic path of our hierarchical sentiment-LDA model. Column 1 shows the topic levels, columns 2, 3 and 4 show the 7 most representative words with positive, neutral and negative sentiment labels, respectively. For each sentiment label, we find semantic dependencies between adjacent levels.

Table 4.5 compares the *accuracy* and *purity* of our proposed methods to four baselines. We find that HSDPP and HSLDA tend to outperform the baselines. For the *Bitterlemons* and *New York Times* corpora, HSDPP exhibits the best performance both in terms of *accuracy* and *purity*. Compared to TAM, HSDPP shows a 9.5% increase in terms of *accuracy*. TAM achieves the best performance on the *Healthcare Corpus* when we set its number of topics to 10. However, the performance differences between HSDPP and TAM on this corpus are not statistically significant. This shows that our proposed contrastive topic modeling strategy is effective in contrastive topic extraction.

4.4.2 Number of themes

To start, for research question **RQ2.1**, to evaluate the effect of the length of each topic path to the performance of contrastive theme modeling, we examine the performance of HSDPP with different values of topic level L, in terms of *accuracy*. In Figure 4.4, we find that the performance of HSDPP in terms of *accuracy* peaks when the length of L equals

Table 4.5: **RQ2.1 and RQ2.2:** *Accuracy*, *purity* and *diversity* values for contrastive theme modeling. Significant differences are with respect to TAM-10 (row with shaded background). Acc. abbreviates accuracy, Pur. abbreviates purity, Div. abbreviates diversity.

	Healthcare Corpus			Bitterlemons Corpus			New York Times		
	Acc.	*Pur.*	*Div.*	*Acc.*	*Pur.*	*Div.*	*Acc.*	*Pur.*	*Div.*
LDA-10	0.336▼	0.337▼	0.156▽	0.346▼	0.350▼	0.167▽	0.321▼	0.322▼	0.172▼
LDA-30	0.313▼	0.315▼	0.134▼	0.324▼	0.332▼	0.137▼	0.317▼	0.317▼	0.144▼
LDA-50	0.294▼	0.298▼	0.115▼	0.304▼	0.309▼	0.121▼	0.295▼	0.301▼	0.134▼
TAM-10	**0.605**	0.602	0.222	0.645	0.646	0.241	0.551	0.560	0.271
TAM-30	0.532▽	0.534▽	0.194	0.623	0.626	0.224	0.564	0.564	0.242
TAM-50	0.522▽	0.525▽	0.152	0.596▽	0.596▽	0.174	0.576	0.582	0.195▼
Sen-TM-10	0.530▽	0.531	0.194	0.537▼	0.539▼	0.209	0.514	0.518	0.255
Sen-TM-30	0.484▼	0.488▼	0.184	0.492▼	0.502▼	0.163▽	0.473	0.478	0.195▼
Sen-TM-50	0.471▼	0.481▼	0.164	0.479▼	0.482▼	0.152▽	0.454▼	0.456▼	0.182▼
HLDA	0.324▼	0.326▼	0.223	0.346▼	0.342▼	0.263	0.329▼	0.330▼	0.291
HSLDA	0.591	0.598	0.225	0.658	0.660	0.269	0.573	0.578	0.292
HSDPP	0.603	**0.604**	**0.244**	**0.692**	**0.696**	**0.292**▲	**0.609**▲	**0.610**▲	**0.326**▲

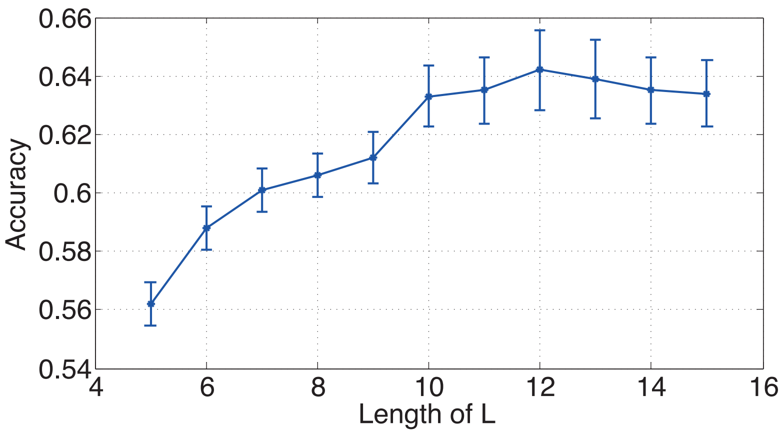


Figure 4.4: **RQ2.1:** Performance with different values of hierarchical topic level L, in terms of accuracy

12; with fewer than 12, performance keeps increasing but if the number exceeds 12, due to the redundancy of topics in contrastive summarization, performance decreases.

Unlike TAM and Sen-LDA, HSDPP and HSLDA determine the optimal number of topics automatically. In Table 4.5 we find that the results for TAM change with various number of topics. However, for HSDPP we find that it remains competitive for all three corpora while automatically determining the number of topics.

4.4.3 Effect of structured determinantal point processes

Turning to **RQ2.2**, Table 4.5 shows that performance of HSDPP and HSLDA on contrastive theme modeling in terms of *accuracy* and *purity*, for all three datasets. We find that HSDPP outperforms HSLDA in terms of both *accuracy* and *purity*. Table 4.5 also contrasts the evaluation results for HSDPP with TAM and Sen-TM in terms of diversity

Table 4.6: **RQ2.2:** Effect of structured determinantal point processes in topic modeling for the top 15 topics in our datasets. Acc. abbreviates accuracy, Div. abbreviates diversity.

	HSLDA		HSDPP	
Descriptions	Acc.	Div.	Acc.	Div.
U.S. Inter. Relations	0.532	0.294	**0.583**△	**0.312**
Terrorism	0.569	0.301	**0.621**▲	**0.341**▲
2004 Election	0.591	0.266	**0.641**▲	**0.281**
US. Healthcare	0.591	0.225	**0.603**	**0.244**
Budget	0.506	0.248	**0.551**▲	**0.299**△
Israel-Palestine	**0.658**	0.269	0.652	**0.292**
Airlines	0.602	0.325	0.602	**0.384**▲
Universities	**0.596**	0.207	0.562	**0.219**
Human Rights	0.571	0.199	**0.624**△	**0.206**△
Children	**0.712**	0.352	0.622	**0.394**▲
Internet	0.547	0.277	**0.601**▲	**0.298**
Atomic Weapons	0.614	0.292	**0.662**△	**0.306**△
Literature	0.555	0.212	**0.611**△	**0.255**△
Abortions	0.594	0.301	**0.608**	**0.322**△
Bio.&Chemi. warfare	0.596	0.275	**0.597**	**0.302**△
Overall	0.581	0.296	**0.614**△	**0.317**△

(columns 4, 7, 10). We evaluate the performance of TAM and Sen-TM by varying the number of topics. HSDPP achieves the highest diversity scores. The diversity scores for TAM and Sen-TM decrease as the number of topics increases. In Table 4.6, we see that HSDPP outperforms HSLDA for all top 15 topics in our dataset in terms of diversity. In terms of diversity, HSDPP offers a significant increase over HSLDA of up to 18.2%.

To evaluate the performance before and after structured determinantal point processes in terms of *accuracy*, Table 4.6 contrasts the evaluation results for HSDPP with those of HSLDA, which excludes structured determinantal point processes, in terms of *accuracy*. We find that HSDPP outperforms HSLDA for each topic listed in Table 4.6. In terms of *accuracy*, HSDPP offers a significant increase over HSLDA of up to 14.6%. Overall, HSDPP outperforms HSLDA with a 5.6% increase in terms of accuracy. Hence, we conclude that the structured determinantal point processes helps to enhance the performance of contrastive theme extraction.

4.4.4 Overall performance

To help us answer **RQ2.3**, Table 4.7 lists the ROUGE performance for all summarization methods. As expected, Random performs worst. Using a depth-first search-based summary method (DFS) does not perform well in our experiments. Our proposed method HSDPP significantly outperforms the baselines on two datasets, whereas on the *healthcare corpus* the LexRank-based method performs better than HSDPP, but not significantly. A manual inspection of the outcomes indicates that the contrastive summarizer in HSDPP (i.e., step (C) in Figure 4.2) is being outperformed by the LexRank summa-

Table 4.7: **RQ2.3:** ROUGE performance of all approaches to contrastive document summarization. Significant differences are with respect to TAM-Lex (row with shaded background).

	Healthcare Corpus			Bitterlemons Corpus			New York Times		
	ROUGE-1	ROUGE-2	ROUGE-W	ROUGE-1	ROUGE-2	ROUGE-W	ROUGE-1	ROUGE-2	ROUGE-W
Random	0.132▼	0.022▼	0.045▼	0.105▼	0.019▼	0.038▼	0.102▼	0.015▼	0.033▼
ClusterCMRW	0.292▼	0.071▼	0.155▼	0.263▼	0.065▼	0.106	0.252▼	0.066▼	0.098▼
DSF	0.264▼	0.064▼	0.125▼	0.235▼	0.054▼	0.091▼	0.211▼	0.047▼	0.088▼
Sen-TM-Lex	0.312▼	0.077▽	0.141	0.296▼	0.062▼	0.129	0.284▼	0.057▼	0.122
TAM-Lex	0.397	0.085	**0.147**	0.362	0.071	0.135	0.341	0.068	0.125
HSDPP	**0.398**	**0.089**	0.142	**0.404**▲	**0.082**▲	**0.159**▲	**0.393**▲	**0.082**▲	**0.149**▲

Table 4.8: **RQ2.4:** ROUGE performance of all our proposed methods in contrastive document summarization. Significant differences are with respect to the row labeled HSDPPD, with shaded background.

	Healthcare Corpus			Bitterlemons Corpus			New York Times		
	ROUGE-1	ROUGE-2	ROUGE-W	ROUGE-1	ROUGE-2	ROUGE-W	ROUGE-1	ROUGE-2	ROUGE-W
HSDPPD	0.291	0.054	0.133	0.301	0.045	0.136	0.284	0.042	0.132
HSDPPR	0.392▲	0.082▲	0.138	0.394▲	0.079▲	0.146▲	0.376▲	0.072▲	0.147▲
HSDPPC	0.362	0.078	0.136	0.319	0.059	0.136	0.308	0.067	0.141
HSDPP	**0.398**▲	**0.089**▲	**0.142**△	**0.404**▲	**0.082**▲	**0.159**▲	**0.393**▲	**0.082**▲	**0.149**▲

rizer in HSDPP-Lex on the *Healthcare* dataset because of the small vocabulary and the relative shortness of the documents in this dataset (at most two sentences per document). The summarizer in HSDPP prefers longer documents and a larger vocabulary. We can see this phenomenon on the *Bitterlemons Corpus*, which has 20–40 sentences per document, where HSDPP achieves a 10.3% (13.4%) increase over TAM-Lex in terms of ROUGE-1 (ROUGE-2), whereas the ROUGE-1 (ROUGE-2) score increases 2.2% (4.8%) over HSDPP-Lex. On the *New York Times*, HSDPP offers a significant improvement over TAM-Lex of up to 13.2% and 18.2% in terms of ROUGE-1 and ROUGE-2, respectively.

4.4.5 Contrastive summarization

Several factors play a role in our proposed summarization method, HSDPP. To determine the contribution of *contrast*, *relevance* and *diversity*, Table 4.8 shows the performance of HSDPPD, HSDPPR, and HSDPPC in terms of the ROUGE metrics. We find that HSDPP, which combines *contrast*, *relevance* and *diversity*, outperforms the other approaches on all corpora. After HSDPP, HSDPPR, which includes *relevance* during the summarization process, performs best. Thus, from Table 4.8 we conclude that *relevance* is the most important part during the summarization process.

4.5 Conclusion

We have considered the task of contrastive theme summarization of multiple opinionated documents. We have identified two main challenges: unknown number of topics and unknown relationships among topics. We have tackled these challenges by combining the nested Chinese restaurant process with contrastive theme modeling, which outputs a set of threaded topic paths as themes. To enhance the diversity of contrastive theme modeling, we have presented the structured determinantal point process to extract a subset of diverse and salient themes. Based on the probabilistic distributions of themes, we generate contrastive summaries subject to three key criteria: contrast, diversity and relevance. In our experiments, we have provided answers to the main research question raised at the beginning of this chapter:

RQ2: How can we optimize the number of topics in contrastive theme summarization of multiple opinionated documents? How can we model the relations among topics in contrastive topic modeling? Can we find an approach to compress the themes into a diverse and salient subsets of themes?

To answer this main research question, we work with three manually annotated datasets. In our experiments, we considered a number of baselines, including recent work on topic modeling and previous summarization work. Our experimental results demonstrated the effectiveness of our proposed method, finding significant improvements over state-of-the-art baselines. Contrastive theme modeling is helpful for extracting contrastive themes and optimizing the number of topics. We have also shown that structured determinantal point processes are effective for diverse theme extraction.

Although we focused mostly on news articles or news-relate articles, our methods are more broadly applicable to other settings with opinionated and conflicted content, such as comment sites or product reviews. Limitations of our work include its ignorance of word dependencies and, being based on hierarchical LDA, the documents that our methods work with should be sufficiently large.

As to future work, parallel processing methods may enhance the efficiency of our topic model on large-scale opinionated documents. Also, supervised and semi-supervised learning can be used to improve the accuracy in contrastive theme summarization. It is interesting to consider recent studies such as [129] on search result diversification for selecting salient and diverse themes. Finally, the transfer of our approach to streaming corpora should give new insights. Hence, in the next chapter, we will focus on the viewpoint summarization problem of multilingual social text streams.

5 Multi-Viewpoint Summarization of Multilingual Social Text Streams

In the previous chapter, we addressed the topic of contrastive theme summarization by using hierarchical non-parametric processes. In this chapter, we continue our research on summarization, and address the viewpoint summarization of multilingual streaming corpora. Focused on an entity [158], a *viewpoint* refers to a topic with a specific sentiment label. As an example, consider the entity "Japan" within the topic "#Whale hunting," with a negative sentiment. With the development of social media, we have witnessed a dramatic growth in the number of online documents that express dynamically changing viewpoints in different languages around the same topic [178]. Unlike viewpoints in stationary documents, time-aware viewpoints of social text streams are dynamic, volatile and cross-linguistic [65]. The task we address in this chapter is *time-aware multi-viewpoint summarization of multilingual social text streams*: we extract a set of informative social text documents to highlight the generation, propagation and drift process of viewpoints in a given social text stream. Figure 5.1 shows an example of our task's output for the topic "#FIFA WorldCup 2014."

The growth in the volume of social text streams motivates the development of methods that facilitate the understanding of those viewpoints. Their multi-lingual character is currently motivating an increasing volume of information retrieval research on multilingual social text streams, in areas as diverse as reputation polarity estimation [178] and entity-driven content exploration [236]. Recent work confirms that viewpoint summarization is an effective way of assisting users to understand viewpoints in stationary documents [74, 77, 107, 127, 138, 157, 243]—but viewpoint summarization in the context of multilingual social text streams has not been addressed yet.

The most closely related work to time-aware viewpoint summarization is the viewpoint summarization of stationary documents [176], in which a sentence ranking algorithm is used to summarize contrastive viewpoints based on a topic-aspect model [175]. Compared with viewpoint summarization in stationary documents, the task of time-aware multi-viewpoint summarization of social text streams faces four challenges: (1) the ambiguity of entities in social text streams; (2) viewpoint drift, so that a viewpoint's statistical properties change over time; (3) multi-linguality, and (4) the shortness of social text streams. Therefore, we address the following main research question listed in Chapter 1:

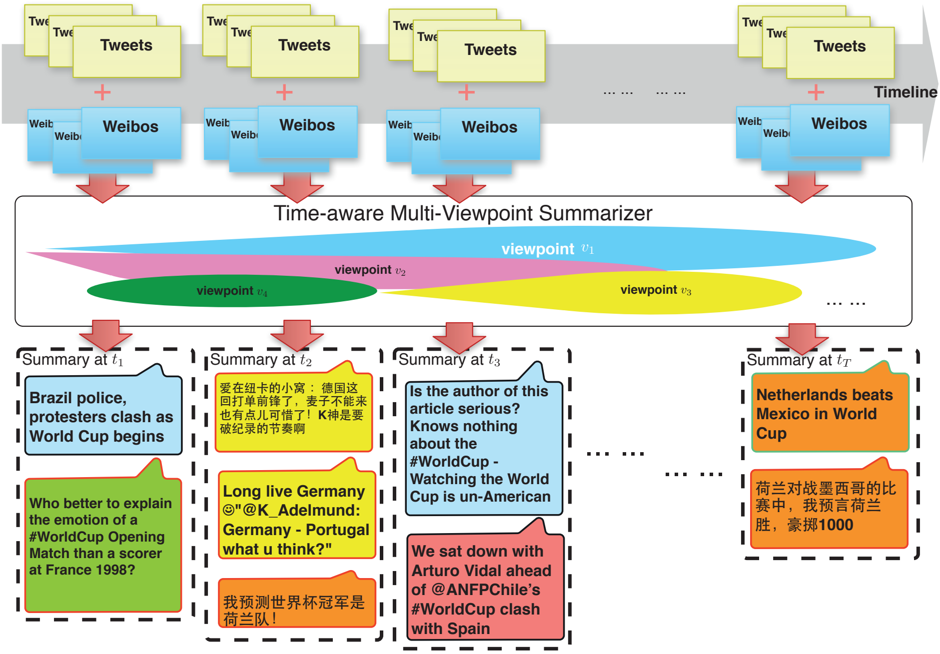


Figure 5.1: An example of time-aware multi-viewpoint summarization of multilingual social text streams about #FIFA Worldcup 2014. The timeline at the top is divided into multiple time periods. The social text stream is composed of English language tweets and Chinese language weibos, which are shown at the top as yellow and blue rectangles, respectively. The time-aware multi-viewpoint summarizer detects temporal viewpoints by analyzing social text and generating an update summary at each period to reflect salient viewpoints. The summarization results are shown as colored round rectangles.

RQ3: How can we find an approach to help detect time-aware viewpoint drift? How can we detect viewpoints from multilingual social text streams? How can we generate summaries to reflect viewpoints of multi-lingual social text streams?

We propose a method to tackle the above research question: (1) We employ a state-of-the-art entity linking method to identify candidate entities from social text; (2) We represent a viewpoint as a tuple of an entity, a topic and a sentiment label, and propose a dynamic latent factor model, called the viewpoint tweet topic model (VTTM), to discover life cycles of a viewpoint. Unlike most existing topic models, VTTM jointly tracks dynamic viewpoints and any viewpoint drift arising with the passing of time. VTTM employs Markov chains to capture the sentiment dependency between two adjacent words. At each time period, VTTM detects viewpoints by jointly generating entities, topics and sentiment labels in social text streams. Gibbs sampling is applied to approximate the posterior probability distribution. (3) Focusing on multi-linguality, we employ an entity-based viewpoint alignment method to match viewpoints in multiple languages by calculating semantic similarities between viewpoints. (4) Lastly, we present a random walk strategy to extract update summaries to reflect viewpoints.

To evaluate our proposed strategy to summarizing dynamic viewpoints in multilingual social text streams, we collect multilingual microblog posts for six well-known topics from 2014. Based on both online and offline human annotations, the evaluation of our proposed method for time-aware viewpoint summarization is shown to be effective.

To sum up, our contributions in this chapter are as follows:

- We propose the task of time-aware multi-viewpoint summarization of multilingual social text streams;
- We propose a viewpoint tweet topic model (VTTM) to track dynamic viewpoints from text streams;
- We align multilingual viewpoints by calculating semantic similarities via an entity-based viewpoint alignment method;
- We present a Markov random walk strategy to summarize viewpoints from multilingual social text streams, which is shown to be effective in experiments using a real-world dataset.

We formulate our research problem in §5.1 and describe our approach in §5.2. §5.3 details our experimental setup and §5.4 presents the experimental results. Finally, §5.5 concludes the chapter.

5.1 Problem Formulation

In this section, we introduce key concepts about time-aware multi-viewpoint summarization. First of all, Table 5.1 lists the notation we use in this chapter.

Given a social text stream $\mathcal{D}$ including T time periods, we define $\mathcal{D}_t \subset \mathcal{D}$ to be the set of documents published during the t-th period. We suppose there are two different languages used in $\mathcal{D}$; we divide $\mathcal{D}_t = \{d_1, d_2, \ldots, d_{D_t}\}$ into $\mathcal{D}_t^{(A)} \cup \mathcal{D}_t^{(B)}$, where $\mathcal{D}_t^{(A)}$ and $\mathcal{D}_t^{(B)}$ indicate the set of documents written in language A and B, respectively.

We use the same definitions of the notions of *topic* and *sentiment* in Section 2.5 and Section 4.1, respectively. Assuming K topics exist in the social text streams on which we focus, we set $z \in \{1, 2, \ldots, K\}$. Following [124], we assume that the sentiment label l_i for a word w_i depends on the sentiment label for its previous word w_{i-1} and the topic z_i simultaneously. Specifically, we set $l_i = -1$ when word w_i is "negative", whereas $l_i = 1$ when w_i is "positive." Then, we define an *entity*, denoted as e, as a rigid designator of a concept around a *topic*, e.g., "China" with "disputed islands between China and Japan". Using a state-of-the-art entity linking method [158], for each document we find an associated entity $e_d \in \mathcal{E}$.

Given a topic z, sentiment label l and entity e, we define a *viewpoint* to be a finite mixture over the *sentiment*, *entity* and *topic*, i.e., a tuple $v = \langle z, l, e\rangle$. Unlike previous work that considers viewpoints to be stationary [75, 176, 243], we assume that each viewpoint is also changing over time, which effects topics, sentiments and entities at each time interval. Thus for each viewpoint at time t, we represent it as a tuple $v = \langle z, l, e, t\rangle$. Given documents $\mathcal{D}_t$, because documents in social text streams are short, we assume that in each document $d \in \mathcal{D}_t$ only one viewpoint v_d exists. We further assume that there exist a probability distribution of viewpoints at each time period.

Table 5.1: Notation used in this chapter.

Symbol	Description
$\mathcal{D}$	all documents
$\mathcal{W}$	vocabulary of documents $\mathcal{D}$
$\mathcal{E}$	entities set in $\mathcal{D}$
$\mathcal{L}$	sentiments in $\mathcal{D}$
$\mathcal{Z}$	topics in $\mathcal{D}$
$\mathcal{V}$	viewpoints in $\mathcal{D}$
K	the number of topics, i.e., $\lvert\mathcal{Z}\rvert$
E	number of entities
$\mathcal{D}_t$	documents posted at t
$\mathcal{D}_t^{(A)}$	documents posted in language A at t
$\mathcal{N}_d$	words in document d
D_t	number of documents posted at t, i.e., $\lvert\mathcal{D}_t\rvert$
N_d	number of words in document d, i.e., $\lvert\mathcal{N}_d\rvert$
d_t	a document in $\mathcal{D}_t$ posted at t
v_d	a viewpoint in document d, $v \in \mathcal{V}$
e_d	an entity present in document d, $e \in \mathcal{E}$
w_i	the i-th word present in document, $w \in \mathcal{W}$
z_i	a topic present in word w_i, $z \in \mathcal{Z}$
l_i	a sentiment label present in word w_i
π_t	distribution of viewpoint at t
θ_t	distribution of entity over viewpoint at t
μ_t	distribution of topics over viewpoint at t
$\phi_{v,z,l,t}$	distribution of words over v, z and l at t
$\mathcal{S}_t$	time-aware multi-viewpoint summary at t

At time t, we set π_t to be a probability distribution of viewpoints at t, μ_t a probability distribution of topics over viewpoints at t, and θ_t a probability distribution of entities over viewpoints t. In social text streams, the statistical properties of viewpoints change over time. Thus we assume that the probability distribution of viewpoints π_t at time t is derived from a Dirichlet distribution over π_{t-1}. Assuming that the distribution of topics and sentiments also drifts over time, we set ϕ_t to be a probability distribution of words in topics and sentiment labels at time t, which is derived from a Dirichlet distribution over ϕ_{t-1} at the previous time $t-1$.

Finally, we define the task of time-aware multi-viewpoint summarization of multilingual social streams. Let multilingual social text streams $\mathcal{D}$ posted in T time periods be given. Then,

- at time period $t = 1$, the target of time-aware multi-viewpoint summarization of multilingual social text streams is to select a set of relevant documents as $\mathcal{S}_1$ as a summary of viewpoints $\mathcal{V}_1$;
- at a time period $t, 1 < t \leq T$, the target is to select a set of both relevant and novel documents, to summarize both the content of viewpoints $\mathcal{V}_t$ at time period t and

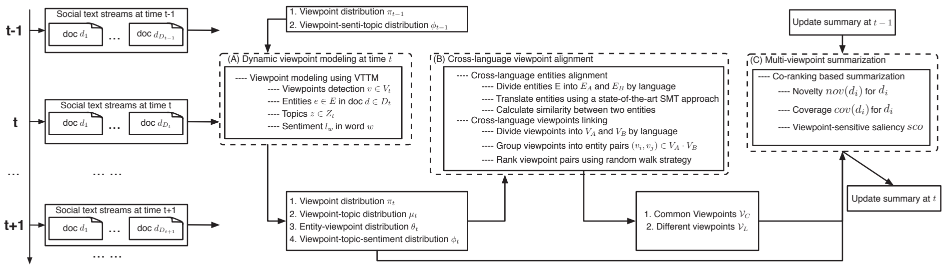


Figure 5.2: Overview of our approach to dynamic viewpoint summarization in social text streams. (A) dynamic viewpoint modeling; (B) cross-language viewpoint alignment; (C) multi-viewpoint summarization and generation of the update summary.

the difference between $\mathcal{V}_t$ and viewpoints $\mathcal{V}_{t-1}$.

5.2 Method

5.2.1 Overview

Before providing the details of our proposed method for time-aware viewpoint summarization, we first give an overview in Figure 5.2. We divide our method in 3 phases: (A) dynamic viewpoint modeling; (B) cross-language viewpoint alignment; and (C) multi-viewpoint summarization. Given a multilingual social text stream $\mathcal{D}_t = \{d_1, d_2, \ldots, d_{D_t}\}$ published at time t, in phase A we propose a dynamic viewpoint model to draw viewpoints for each document. Using a set of viewpoints $\mathcal{V}_t$ extracted from phase A, in phase B we use cross-language viewpoint alignment to link similar viewpoints in different languages by computing the similarity between two entities. Phase C then summarizes documents according to viewpoint distributions using a co-ranking based strategy. In the end we get a time-aware multi-viewpoint summary $\mathcal{S}_t$ at time t.

5.2.2 (A) Dynamic viewpoint modeling

At time period t, given documents $\mathcal{D}_t$ in two different languages, our task during phase A is to detect dynamic viewpoints from the documents in $\mathcal{D}_t$. Using an extension of dynamic topic models [31], we propose a dynamic latent factor model, the *viewpoint tweets topic model* (VTTM), that jointly models viewpoints, topics, entities and sentiment labels in $\mathcal{D}_t$ at each time interval t.

Using a state-of-the-art entity linking method for social media [158], for each document d at t, we discover entities by calculating the *COMMONNESS* value of the document. We assume that there are, in total, V viewpoints and K topics in social text steams. For each document d, there are an entity e_d and N_d words; for each word $w_i \in d$, there is a topic z_i and a sentiment label l_i. We assume that the viewpoint v_d in d is derived via a multinomial distribution over a random variable π_t that indicates a probability distribution over viewpoints at t; each topic z, each sentiment label l and each entity e in

- For each topic $z \in \mathcal{Z}$ and sentiment l at time t:
  - Draw $\phi_{z,l,t} \sim Dir(\phi_{z,l,t-1} \cdot \beta_t)$;
- For each viewpoint $v \in \mathcal{V}$:
  - Draw $\pi_{v,t} \sim Dir(\alpha \cdot \pi_{v,t-1})$;
  - Draw $\mu_{v,t} \sim Dir(\chi)$; $\theta_{v,t} \sim Dir(\delta)$
  - For each topic z, draw $\rho_{v,z} \sim Beta(\eta)$;
- For each document $d \in \mathcal{D}_t$:
  - Draw a viewpoint $v_d \sim Multi(\pi_t)$;
  - Draw an entity $e_d \sim Multi(\theta_{v_d,t})$;
  - Draw $\sigma \sim Dir(\tau)$;
  - For each word $w_i \in \mathcal{N}_d$, $0 < i < N_d$:
    * Draw a topic $z_i \sim Multi(\mu_{v_d,t})$;
    * Draw $x_i \sim Multi(\sigma)$;
    * If $x_i = 1$, draw $l_i \sim l_{i-1}$
    * If $x_i = -1$, draw $l_i \sim (-1) \cdot l_{i-1}$;
    * If $x_i = 0$, draw $l_i \sim Bern(\rho_{v_d,z_i})$;
    * Draw word $w_i \sim Multi(\phi_{z_i,l_i,t})$:

Figure 5.3: Generative process in VTTM at time period t.

document d is derived from the viewpoint v_d. The probability distribution π_t is derived from a Dirichlet mixture over the viewpoint distribution π_{t-1} at the previous period.

In VTTM we consider the sentiment dependency between two adjacent words. That is, a Markov chain is formed to represent the dependency relation between the sentiment labels of two adjacent words. Given a word w_i, the sentiment label l_i is selected depending on the previous word. The transition probability distribution is derived from the sentiment label of l_{i-1} and a transition variable x_i. The transition variable $x \in \mathcal{X}$ determines where the corresponding sentiment label comes from. If $x = 1$, then the sentiment label l_i of w_i is identical to the sentiment label l_{i-1} of word w_{i-1}; whereas if $x_i = -1$, the sentiment label l_i is opposite to l_{i-1}, which shows that the sentiment label changes from one polarity to the other. Thus, we set the transition variable $x_i = 1$ when w_i and w_{i-1} are connected by a correlative conjunction, such as "and" and "both"; and we set $x_i = -1$ when w_i and w_{i-1} are connected by an adversative conjunction, such as "but" and "whereas"; we set $x_i = 0$ for other kinds of conjunctions. The generative process of VTTM is shown in Figure 5.3.

Similar to other topic models [31, 32, 98, 242], it is intractable to derive the explicit posterior distribution of viewpoint $v_{d,t}$ at time period t. We apply a Gibbs sampling method [56] for sampling from the posterior distribution over viewpoints, entities, topics and sentiment labels. The sampling algorithm provides a method for exploring the implicit topic for each word and the particular viewpoint for each document.

At time period t, given document d, the target of our sampling is to approximate the posterior distribution $p(v_d, \vec{z}_d, \vec{l}_d, \vec{x}_d \mid \mathcal{W}, \mathcal{Z}, \mathcal{V}, \mathcal{E}, t)$, where $\vec{z}_d$, $\vec{l}_d$ and $\vec{x}_d$ indicate document d's topic vector, sentiment labels, and transition vector, respectively. Conceptually, we divide our sampling procedure into two parts. First, we sample the conditional probability of viewpoint v_d in each document $d \in \mathcal{D}_t$ given the values of inferred topics and sentiment labels, i.e., $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z})$. Second, given the current state of viewpoints, we sample the conditional probability of topic z_i with sentiment label l_i for word w_i, i.e., $P(z_i = k, l_i = l, x_i = x \mid \mathcal{X}_{-i}, \mathcal{L}_{-i}, \mathcal{Z}_{-i}, \mathcal{W}, v_d)$.

As the first step in our sampling procedure, for each document $d \in \mathcal{D}_t$, to calculate the probability of viewpoint v_d by sampling $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z})$, we have:

$$\begin{aligned} &P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}) \propto \frac{n_{v,t}^{-d} + \alpha \cdot \pi_{t-1}}{n_t^{-d} + 1} \cdot \\ &\prod_{e \in \mathcal{E}} \frac{n_{v,e,t}^{-d} + \delta}{n_{v,t}^{-d} + E\delta} \cdot \prod_{z \in \mathcal{Z}} \frac{n_{v,z,t}^{-d} + \chi_{z,t}}{n_{z,t}^{-d} + \sum_{z \in \mathcal{Z}} \chi_{z,t}} \cdot \prod_{l \in \mathcal{L}} \prod_{w \in \mathcal{N}_d} \frac{n_{z,l,v,t}^{w} + \beta_t \cdot \phi_{t-1,w}^{z,l,v}}{n_{z,l,v,t}^{-i} + \sum_{w \in \mathcal{N}} \beta_t \cdot \phi_{t-1,w}^{z,l,v}}, \end{aligned} \tag{5.1}$$

where $n_{v,t}^{-d}$ indicates the number of times that documents have been assigned to viewpoint v at t, except for document d; $n_{v,e,t}^{-d}$ indicates the number of times that entity e has been assigned to viewpoint v at t, excluding d; $n_{v,z,t}^{-d}$ indicates the number of times that topic z, at time t, has been assigned to viewpoint v, except for topic z in d; $n_{z,l,v,t}^{w}$ indicates the number of times that word w has been assigned to z, l and v jointly at t; $\phi_{t-1,w}^{z,l,v}$ is the probability of word w given v, z and l at $t-1$.

As the second step in our sampling procedure, given the viewpoint v_d sampled from document d, when $x_i \neq 0$ and $x_{i+1} \neq 0$ we sample the ith word w_i's topic z_i and sentiment label l_i using the probability in Eq. 5.2:

$$\begin{aligned} P(z_i = k, l_i = l, x_i =x \mid \mathcal{X}_{-i}, \mathcal{L}_{-i}, \mathcal{Z}_{-i}, \mathcal{W}, v_d) \propto \\ \frac{n_{v_d,k,t}^{-i} + \chi_{k,t}}{n_{v_d,t}^{-i} + \sum_{z \in \mathcal{Z}} \chi_{z,t}} \cdot \frac{n_{k,l,v_d,t}^{w_i,-i} + \beta_t \cdot \phi_{t-1,w_i}^{k,l,v_d}}{n_{k,l,v_d,t}^{-i} + \sum_{w \in \mathcal{N}} \beta_t \cdot \phi_{t-1,w}^{k,l,v_d}} \cdot \\ \frac{n_{-i,x}^{w_i} + \tau_x}{n_{-i}^{w_i} + \sum_{x \in \mathcal{X}} \tau_x} \cdot \frac{n_{-(i+1),x_{i+1}}^{w_{i+1}} + I(x_{i+1} = x_i) + \tau_x}{n_{-(i+1)}^{w_{i+1}} + 1 + \sum_{x \in \mathcal{X}} \tau_x} \end{aligned} \tag{5.2}$$

where $n_{v_d,k,t}^{-i}$ indicates the number of times that a word with viewpoint v_d has been assigned to a topic k at time period t, except for the ith word; $n_{-i,t}^{d}$ indicates the number of words in document d, except for the ith word; $n_{-i,k,l}$ indicates the number of times that a word has been assigned to topic z and sentiment l synchronously, excluding the ith word; ϕ_{t-1}^{k,l,w_i} is the probability of word w_i given z and l at $t-1$; $n_{-i,x}^{w_i}$ indicates the number of times that w_i has been assigned to x, excluding the current one; and $I(x_{i+1} = x_i)$ gets the value 1 if $x_{i+1} = x_i$, and 0 otherwise. When $x_i = 0$, w_i's sentiment label l_i is derived from a Bernoulli distribution ρ_{v_d,z_i}, thus the last part in Eq. 5.2 is replaced by a posterior distribution over η, i.e., $(n_{z,l,v,t}^{-i} + \eta_l)/(n_{v,z,t}^{-i} + \sum_{l \in \mathcal{L}} \eta_l)$.

After sampling the probability for each viewpoint v, topic z and sentiment label l, at time period t we approximate the random variable ϕ_t that indicates the probability distribution over viewpoints, topics and sentiments labels, a viewpoint distribution π_t, a topic distribution μ_t over viewpoints, and entity distribution θ_t over viewpoints, similar to Iwata et al. [98].

5.2.3 (B) Cross-language viewpoint alignment

Using VTTM, we extract viewpoints from multi-lingual social text streams. Multilinguality may make the viewpoint set $\mathcal{V}$ redundant and ambiguous. To address this, we present a cross-language viewpoint alignment strategy to connect the same viewpoint across languages. Shortness and sparseness hinder statistical machine translation in social text streams. We consider entities, i.e., concepts that can be linked to a specific Wikipedia document, as a means to connect viewpoints by comparing the similarity between two linked Wikipedia documents. We divide viewpoints $\mathcal{V}$ extracted from VTTM into $\mathcal{V}_A$ and $\mathcal{V}_B$ according to their languages L_A and L_B. Similarly, we divide entities $\mathcal{E}$ into $\mathcal{E}_A$ and $\mathcal{E}_B$ according to their languages.

Given viewpoint $v_A \in \mathcal{V}_A$, at time period t we extract the most relevant entity $e_i \in \mathcal{E}_A$ that has the highest $\theta_{v,e_i,t}$, i.e., $P(e_i \mid v, t)$. The same procedure is adapted to obtain $e_j \in \mathcal{E}_B$ for another viewpoint $v_B \in \mathcal{V}_B$. We compute the similarity between v_A and v_B by comparing the similarity between two entities e_i and e_j, shown in Eq. 5.3:

$$sim_t(v_A, v_B \mid t) = sim(e_i, e_j) \cdot \theta_{v_A,e_i,t} \cdot \theta_{v_B,e_j,t}, \tag{5.3}$$

where $sim(e_i, e_j)$ is the similarity between e_i and e_j in two languages. To compute $sim(e_i, e_j)$, we compute the similarity between two linked Wikipedia documents. Using links to English Wikipedia documents on Wikipedia pages, we translate a non-English Wikipedia document to an English Wikipedia document, i.e., a corresponding English Wikipedia document $\widehat{\mathcal{W}_{e_j}}$ for document $\mathcal{W}_{e_j}$. We use LDA [32] to represent each Wikipedia document $\mathcal{W}$ as a K-dimensional topic vector $\vec{\varphi}_{\mathcal{W}}$. Then $sim(e_i, e_j)$ is computed proportionally to the inner product of the two vectors:

$$sim(e_i, e_j) = \frac{|\vec{\varphi}_{\mathcal{W}_{e_i}} \cdot \vec{\varphi}_{\widehat{\mathcal{W}_{e_j}}}|}{|\vec{\varphi}_{\mathcal{W}_{e_i}}| \cdot |\vec{\varphi}_{\mathcal{W}_{e_j}}|}, \tag{5.4}$$

where $\vec{\varphi}_{\mathcal{W}_{e_i}}$ indicates the topic vector for entity e_i's Wikipedia document, and $\vec{\varphi}_{\widehat{\mathcal{W}_{e_j}}}$ indicates the topic vector for entity e_j's translated Wikipedia document. We sum up the similarities between v_A and v_B at all time periods to obtain the similarity between v_A and v_B: $sim(v_A, v_B) = \sum_t sim_t(v_A, v_B)$. Thus, for each viewpoint $v_A \in \mathcal{V}_A$, we find the most similar viewpoint $v_B \in \mathcal{V}_B$ to match with the highest $sim(v_A, v_B)$. By generating such viewpoint pairs, we extract a set of viewpoint pairs $\mathcal{V}_s$ from $\mathcal{V}$. To remove redundant viewpoint pairs from $\mathcal{V}_s$, we employ a random walk-based ranking strategy [64] to rank $\mathcal{V}_s$ iteratively, in which each viewpoint pair's score, sa, receives votes from other pairs. As shown in Eq. 5.5, we use the similarity between two viewpoint pairs as the transition probability from one to another:

$$tr((v_A, v_B), (v'_A, v'_B)) = \frac{|sim(v'_A, v_B) \cdot sim(v_A, v'_B)|}{|sim(v_A, v_B)| \cdot |sim(v'_A, v'_B)|}. \tag{5.5}$$

At the beginning of the iterative process, an initial score for each pair is set to $1/|\mathcal{V}_s|$, and at the c-th iteration, the score of a viewpoint pair i is computed in Eq. 5.6:

$$sa(i)^{(c)} = \mu \sum_{i \neq j} \frac{tr\,(i, j)}{\sum\limits_{j' \in \mathcal{V}_s} tr(i, j')} \cdot sa(j)^{(c-1)} + \frac{(1-\mu)}{|\mathcal{V}_s|}, \tag{5.6}$$

where $|\mathcal{V}_s|$ equals the number of viewpoint pairs; μ denotes a decay parameter that is usually set to 0.85. The iterative process will stop when it convergences. Then we extract the top $|\mathcal{V}_C|$ viewpoint pairs from the ranked list, and merge two viewpoints in a pair into a single viewpoint. Below, we write $\mathcal{V}_C$ to denote $|\mathcal{V}_C|$ common viewpoints shared by both $\mathcal{V}_A$ and $\mathcal{V}_B$, and $\mathcal{V}_L = (\mathcal{V}_A \cup \mathcal{V}_B, v) \setminus \mathcal{V}_C$ to denote viewpoints $v \notin \mathcal{V}_C$.

5.2.4 (C) Multi-viewpoint summarization

The last step of our method, after cross-language viewpoint alignment is time-aware multi-viewpoint summarization of social text streams. Following [54, 70, 156], we propose a time-aware multi-viewpoint summarization method to summarize time series viewpoints by extracting a set of documents at each time period.

Suppose a set of viewpoint summaries $\{\mathcal{S}_s\}_{s=1}^{t-1}$ has been generated and read during the previous $t-1$ time periods. Based on viewpoint pairs $\mathcal{V}_s$ and viewpoint distributions inferred via VTTM, our target is to generate an update summary $\mathcal{S}_t$ to reflect the distribution of viewpoints at time period t. Inspired by Wan [240], we employ a co-ranking based algorithm to calculate the saliency of each tweet by considering both novelty and coverage. *Novelty* concerns the semantic divergence of viewpoint probabilities between a candidate document $d_i \in \mathcal{D}_t$ and previous summaries $\{\mathcal{S}_s\}$. *Coverage* concerns the relevance of a candidate document $d_i \in \mathcal{D}_t$ to a given viewpoint. Each document d_i's total saliency score $sco(d_i)$ is composed of a novelty score $nov(d_i)$ and a coverage score $cov(d_i)$. As in co-ranking, Markov random walks are employed to optimize the ranking list iteratively. Three matrices are constructed to capture the transmission probability between two documents. Given a viewpoint $v \in \mathcal{V}_C \cup \mathcal{V}_L$, item $M^A_{i,j}$ in matrix M^A is about the similarity between two candidate documents d_i and d_j in $\mathcal{D}_t$:

$$M^A_{i,j} = \frac{\sum\limits_{e,e'} sim(e,e') \cdot \sum\limits_{z\in\mathcal{Z}} \sum\limits_{l\in\mathcal{L}} \phi^{z,l,v}_{d_i,t} \cdot \phi^{z,l,v}_{d_j,t}}{\|\Phi^v_{d_i,t}\| \cdot \|\Phi^v_{d_j,t}\|}, \tag{5.7}$$

where entity e and e' belong to $\mathcal{E}_{d_i}$ and $\mathcal{E}_{d_j}$, respectively; $\Phi^v_{d_i,t}$ is a matrix over topics and sentiment labels; each item for z, l, i.e., $\phi^{z,l,v}_{d_i,t}$ in Eq. 5.7, is calculated by averaging the value of $\phi^{z,l,v}_{t,w}$ of all words $w \in d_i$. Since the transmission matrix must be a stochastic matrix [63], we normalize M^A to $\widehat{M}^A$ by making the sum of each row equal to 1. Similarly, we use $\widehat{M}^B$ to represent the transmission matrix among summaries during the previous $t-1$ time periods; we use M^{AB} to represent the similarity between $\mathcal{D}_t$ and $\{\mathcal{S}_s\}_{s=1}^{t-1}$. We normalize M^{AB} to $\widehat{M}^{AB}$ by making the sum of each row equal to 1. The third and last matrix, W^{AB}, is about the divergence between $\mathcal{D}_t$ and $\{\mathcal{S}_s\}_{s=1}^{t-1}$; given a viewpoint v, we calculate each item $W^{AB}_{i,j}$ in $\widehat{W}^{AB}$ using Eq. 5.8:

$$W^{AB}_{i,j} = \frac{|t-s| \cdot |\pi_{v,t} - \pi_{v,s}| \cdot \|\Phi^v_{d_i,t} - \Phi^v_{d_j,t}\|}{\|\Phi^v_{d_i,t}\| \cdot \|\Phi^v_{d_j,t}\|}, \tag{5.8}$$

After row-normalization, we obtain $\widehat{W}^{AB}$ from W^{AB}. Using a co-ranking based update summarization algorithm [240], given a viewpoint v, for each iteration we use two column vectors $nov(d) = [nov(d_i)]_{i\in\mathcal{D}_t}$ and $cov(d) = [cov(d_i)]_{i\in\mathcal{D}_t}$ to denote the novelty scores and coverage scores of the documents in $\mathcal{D}_t$, respectively. In order to compute the viewpoint-biased scores of the documents, we use column vectors $\kappa_{d,v} = [\kappa_{d_i,v}]_{i \in \mathcal{D}_t}$ to reflect the relevance of the documents to the viewpoint v, where each entry in $\kappa_{d,v}$ corresponds to the conditional probability of the given viewpoint in documents, i.e., $\|\Phi^{v}_{d_i,t}\|$. Then κ is normalized to $\widehat{\kappa}$ to make the sum of all elements equal to 1. After computing the above matrices and vectors, we can compute the update scores and the coverage scores of the documents in a co-ranking process. So at the c-th iteration, the update and coverage scores of d_i are calculated as:

$$
\begin{aligned}
nov(d_i)^{(c)} = {} & \varepsilon_1 \sum_{j \in D_t}^{i \neq j} \widehat{M}^{A}_{i,j} \cdot nov(d_j)^{(c-1)} \\
& + \varepsilon_2 \sum_{j \in \{S_s\}} \hat{W}^{AB}_{i,j} \cdot nov(d_j)^{(c-1)} + \frac{(1 - \varepsilon_1 - \varepsilon_2)}{D + S} \cdot \kappa_{d_i,v},
\end{aligned}
\tag{5.9}
$$

and

$$
\begin{aligned}
cov(d_i)^{(c)} = {} & \gamma_1 \sum_{j \in D_t}^{i \neq j} \widehat{M}^{A}_{i,j} \cdot cov(d_j)^{(c-1)} \\
& + \gamma_2 \sum_{j \in \{S_s\}} \widehat{M}^{AB}_{i,j} \cdot cov(d_j)^{(c-1)} + \frac{(1 - \gamma_1 - \gamma_2)}{D + S} \cdot \kappa_{d_i,v},
\end{aligned}
\tag{5.10}
$$

where we set γ and ε as decay parameters in random walks. Initially, we set *nov*(d_i) and *cov*(d_i) as $\frac{1}{D_t}$, respectively. After each iteration c, we normalize $nov(d_i)^{(c)}$ and $cov(d_i)^{(c)}$ and calculate the *saliency* score of each document d_i as follows:

$$
sco(d_i)^{(c)} = nov(d_i)^{(c)} + cov(d_i)^{(c)} \tag{5.11}
$$

Following Eq. 5.9 and 5.10, for each given viewpoint $v \in \mathcal{V}_C \cup \mathcal{V}_L$, we rank documents in $\mathcal{D}_t$ to a ranking list $\mathcal{R}_v$, thus we apply Algorithm 6 to select documents to generate the viewpoint summary at time t. Eventually, we generate a set of summaries $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_T\}$ as the time-aware summarization result.

5.3 Experimental Setup

In §5.3.1, we divide our main research question **RQ3** into three research questions to guide our experiments; we describe our dataset in §5.3.2 and specify how data was labeled in §5.3.3; §5.3.4 details the parameters used, and §5.3.5 details our evaluation metrics; the baselines are described in §5.3.6.

5.3.1 Research questions

We divide our main research question **RQ3** into the research questions **RQ3.1–RQ3.3** that guide the remainder of the chapter.

Algorithm 6: Time-aware multi-viewpoint summarization at time period t

Input:
Viewpoints $\mathcal{V}_C$ and $\mathcal{V}_L$, ranking list $\{R_v\}_{v \in \mathcal{V}_C \cup \mathcal{V}_L}$, summaries $\{\mathcal{S}_s\}_{s=1}^{t-1}$, $\mathcal{D}_t$,
probability distributions π_t, θ_t, ϕ_t,
probability distributions $\{\pi_s\}_{s=1}^{t-1}$, $\{\theta_s\}_{s=1}^{t-1}$, $\{\phi_s\}_{s=1}^{t-1}$
Output: Multi-viewpoint summary $\mathcal{S}_t$ at t;

```
Ω ← null; T ← predefined threshold; L ← length of summary
while |Ω| < L do
    for each v do
        d_i = top document in R_v;
        R_v = R_v − d_i;
        if max_{d_j ∈ Ω} sim(d_i, d_j | v, t) < T then
            Ω = Ω + d_i;
    if |Ω| = L then
        S_t = Ω;
        Break;
```

RQ3.1 How does our viewpoint tweet topic model (VTTM) perform in time-aware viewpoint modeling? Does it help detect time-aware viewpoint drift? (See §5.4.1.)

RQ3.2 What is the performance of cross-language viewpoint alignment? Can it help detect common viewpoints from multilingual social text streams? (See §5.4.2.)

RQ3.3 How does our end-to-end time-aware multi-viewpoint summarization method (TAMVS) perform? Does it outperform the baselines? What is the effect if we only consider *novelty* or *coverage*? (See §5.4.3.)

5.3.2 Dataset

In order to assess the performance of our methods, we collect a dataset of microblogs in two languages. We define multilingual queries about six well-known topics in 2014 and crawl English and Chinese microblogs via the Twitter streaming API[1] and a Sina Weibo[2] crawler, respectively. Table 5.2 provides descriptive statistics about the dataset. The tweets and weibos are posted between January, 2014 and August, 2014.

To evaluate the effectiveness of time-aware viewpoint summarization methods in our dataset, we used a crowdsourcing platform and had workers to label the ground truth in our dataset in their native language (i.e., Chinese or English); §5.3.3 details the annotations we obtained. In total, 8,308 English tweets and 12,071 Chinese weibos were annotated.

5.3.3 Crowdsourcing labeling

We obtain our annotations using the CrowdTruth platform [97] and assess the annotations using the CrowdTruth metrics [17].

[1] `https://dev.twitter.com/docs/streaming-apis`
[2] Chinese microblogging platform, `http://www.weibo.com`.

Table 5.2: Six topics in our dataset. The first column shows the topic name. The second and third column shows the number of English tweets and Chinese weibos per topic respectively. Each item is divided into two parts: the number of documents annotated, and the number of documents for each topic.

Topic	# tweets	# weibos
1. World Economic Forum	2,000/2,000	1,978/1,978
2. Whaling hunting	566/566	1,072/1,072
3. FIFA Worldcup 2014	1,120/1,963	1,801/1,801
4. Missing MH370	3,124/6,308	4,725/4,725
5. Anti-Chinese in Vietnam 2014	825/2,001	1,095/1,095
6. Sinking of the MV Sewol	403/2,000	1,400/1,881

The *Topic* annotation task gathers relevant tweets for each topic introduced in Table 5.2, and relevant topic mentions from each given tweet. Based on the answers gathered from the crowd we construct for each topic type a set of relevant tweets and a set of relevant topic mentions. Following the CrowdTruth approach, each tweet is assigned a topic type relevance score and each topic mention a relevance score. The *Sentiment* annotation task captures the sentiment and the intensity (i.e., high, medium, low) of the tweets and their topic mentions. The crowd provides the sentiment and the intensity of each topic mention and the overall sentiment and intensity of the tweet. The *Novelty ranking* task provides a ranking of the tweets based on how much new information they bring in with regard to a given topic. As data preparation, the tweets of a given topic are sorted chronologically and split by day. The crowdsourcing task is a pair-wise comparison of the tweets by following the approach: every tweet of a particular day is compared to all the following tweets, resulting in $\frac{n(n-1)}{2}$ comparison pairs per day, where n is the total number of tweets published on that day. Given the summary of the topic, for each pair of tweets, the crowd indicates which tweet is more salient with regard to the topic. By analyzing these judgments we provide, per day, a ranked list of salient tweets.

Table 5.3 provides an overview of the annotations gathered. On each task we applied the CrowdTruth metrics [17] in order to identify and remove spam, low-quality workers and their annotations. Only the quality annotations were used as ground truth for further experiments. We validate the results by performing manual evaluation of the annotations. We extract a pool of workers, evenly distributed between low and high-quality, and annotate them in the following way: 0 for quality work and 1 for low-quality work. These scores are then used to compute the precision, recall, accuracy and F1-score, in order to confirm the CrowdTruth metrics accuracy. Overall, we obtain high scores for each of the measures (above 0.85) and across tasks, which indicates that the low-quality workers were correctly separated from quality workers.

5.3.4 Parameters

Following existing topic models [84], for the weighted parameter $\alpha_{v,t}$ and β_t, we set $\alpha_{u,t}$ to $50/V$ and β_t to 0.5. For the hyperparameters χ and δ in VTTM, we set $\chi = \delta = 0.5$.

Table 5.3: Crowdsourcing task results overview.

Task	Topic	Sentiment	Novelty ranking
Units	6,225	5,317	5,211
Jobs	92	77	82
#Total workers	6,337	6,555	5,336
#Unique workers	557	500	341
#Spam workers	1,085	1,334	1,284
#Total judgments	43,575	53,170	78,165
#Spam judgments	7,562	10,519	14,475
Total cost	\$1,136	\$1,328	\$1,444

The default number of viewpoints in VTTM is set to 20. To optimize the number of viewpoints, we compare the performance at different values (see below). In time-aware multi-viewpoint summarization we set the parameter $\varepsilon_1 = \varepsilon_2 = 0.4$ in Eq. 5.9 and $\gamma_1 = \gamma_2 = 0.4$ in Eq. 5.10; the convergence threshold in co-ranking is set to 0.0001. The length of the summary L is set to 200 words per time period.

5.3.5 Evaluation metrics

To assess VTTM, we adapt the *purity* and *accuracy* evaluation metrics, which are widely used in topic modeling and clustering experiments [176, 188]. To evaluate the performance of time-aware multi-viewpoint summarization, we adopt the ROUGE evaluation metrics: ROUGE-1 (unigram), ROUGE-2 (bigram) and ROUGE-W (weighted longest common sequence), as same as in Chapters 3 and 4.

Statistical significance of observed differences between the performance of two runs is tested using a two-tailed paired t-test and is denoted using $^\blacktriangle$ (or $^\blacktriangledown$) for strong significance for $\alpha = 0.01$; or $^\vartriangle$ (or $^\triangledown$) for weak significance for $\alpha = 0.05$.

5.3.6 Baselines and comparisons

We list the methods and baselines that we consider in Table 5.4. We divide our methods into 3 groups according to the phases A, B, and C specified in §5.2. We write VTTM for the dynamic viewpoint model we proposed in §5.2.2. In the context of **RQ3.1**, we write VTTM-S for the stationary viewpoint modeling method. We write CLVA for the LDA-based viewpoint alignment method in phase B. In the context of **RQ3.2**, we write CLVA-T for the alignment method that applies term frequency in viewpoint similarity calculation, CLVA-E for the alignment method that only checks the consistency of entities. We write TaMVS for the overall process described in §5.2, which includes dynamic viewpoint modeling, cross-language viewpoint alignment and time-aware viewpoint summarization, and TaMVS-V for the viewpoint summarization method without considering cross-language viewpoint alignment. In the context of **RQ3.3** we use TaMVSN and TaMVSC to denote variations of TaMVS that only consider *Novelty* and *Coverage*, respectively.

Table 5.4: Our methods and baselines used for comparison.

Acronym	Gloss	Reference
Dynamic viewpoint modeling		
VTTM	Dynamic viewpoint modeling in (A)	§5.2.2
VTTM-S	Stationary viewpoint modeling in (A)	§5.2.2
Cross-language viewpoint alignment		
CLVA	LDA-based strategy in (B)	§5.2.3
CLVA-T	Term similarity based strategy in (B)	§5.2.3
CLVA-E	Entity similarity based strategy in (B)	§5.2.3
Time-aware multi-viewpoint summarization		
TaMVS	Summarization strategy defined in (C)	§5.2.4
TaMVS-V	TaMVS without phase B	§5.2.4
TaMVSN	TaMVS only considering *novelty* in (C)	§5.2.4
TaMVSC	TaMVS only considering *coverage* in (C)	§5.2.4
Topic models		
Sen-TM	Sentiment LDA based contrastive summarization	[124]
TAM	Topic-aspect model based contrastive summarization	[175]
Summarization		
CoRUS	Co-Ranking update summarization	[240]
IUS	Incremental update summarization	[156]
LexRank	LexRank algorithm for summarization	[63]

No previous work has addressed the same task as we do in this chapter. However, some existing work can be considered as baselines in our experiments. To assess the contribution of VTTM in dynamic viewpoint modeling, our baselines include recent work on stationary viewpoint modeling. We use the Topic-aspect model [175, TAM] and the Sentiment-topic model [124, Sen-TM] as baselines for topic models. As baselines for summarization, we use three representative summarization algorithms, i.e., LexRank, IUS and CoRUS, as baselines: (1) the LexRank algorithm [63] ranks sentences via a Markov random walk strategy; (2) the IUS algorithm [156] generates an incremental update summary for given text streams; (3) the CoRUS algorithm [240] generates an update summary using a co-ranking strategy, but without VTTM.

5.4 Results and Discussion

We compare VTTM to baselines for viewpoint modeling in social text streams, examine the performance of CLVA for cross-language viewpoint alignment as well as the end-to-end summarization performance of TaMVS.

5.4.1 Viewpoint modeling

To begin, Table 5.5 shows four example viewpoints produced by VTTM. Column 1 shows the entities included by each viewpoint, column 2 shows topics attached with

Table 5.5: Task: dynamic viewpoint modeling. **RQ3.1:** Example viewpoints produced by VTTM. Column 1 lists the entities corresponding to the viewpoints; Column 2 list the topics in viewpoints, Columns 3, 4 and 5 list the probabilities of positive, neutral and negative labels for each topic, respectively. Column 6 shows the time interval of each viewpoint.

Entity	Topic	Positive	Neutral	Negative	Time interval
Search_for_Malaysia_Airlines_Flight_370	#Missing MH370	0.077	0.422	0.501	2014-03-27
Whaling_in_ Japan	#Whaling hunting	0.015	0.317	0.668	2014-05-05
Mexico	#World Economic Forum	0.102	0.755	0.143	2014-01-28
	#FIFA Worldcup 2014	0.241	0.262	0.497	2014-06-20
China-Japan_relations	#The World Economic Forum	0.110	0.166	0.724	2014-01-26
	#Anti-Chinese in Vietnam	0.017	0.621	0.362	2014-06-03

the entity in the viewpoint, columns 3, 4, 5 show the probability of positive, neutral and negative sentiment, respectively; column 6 shows the time period of the viewpoint. For a viewpoint about "China-Japan_relations" in Table 5.5, we find that its topic changes from "#World Economic Forum" on 2014-01-26 to "#Anti-Chinese in Vietnam" on 2014-06-03.

Next, we address **RQ3.1** and test whether VTTM is effective for the viewpoint modeling task in social text streams. Table 5.7 shows the evaluation results for viewpoint modeling in terms of purity and accuracy for English tweets and Chinese weibos. For both languages, we find that VTTM outperforms TAM for all topics in terms of purity and accuracy. VTTM achieves an increase in purity over TAM of up to 23.4%, while accuracy increases by up to 21.4%. Compared with Sen-LDA, VTTM offers an increase of up to 12.0%, whereas accuracy increases by up to 12.6%. We look at those unsuccessful results made by VTTM, and find that for 67.2% of those documents the sentiment labeling results are incorrect, whereas for 75.4% of those documents the topic prediction results are incorrect. Another aspect of **RQ3.1** concerns *viewpoint drift*, i.e., changes of statistical properties.

Figure 5.4 shows the propagation process of an example viewpoint about "FIFA World Cup 2014 Group E." The curves in Figure 5.4 plot viewpoint distributions π over time, which indicate the *viewpoint drift* between two adjacent intervals. We also find that this viewpoint's sentiment changes over time. Thus, VTTM has to respond to these drift phenomena. Table 5.6 contrasts the average performance of VTTM and VTTM-S (the stationary version of VTTM) for all periods in terms of Accuracy. For both languages, VTTM outperforms VTTM-S for each topic. We conclude that VTTM responds better to topic drift than VTTM-S, which neglects the dependency of viewpoints between two adjacent intervals.

5.4.2 Cross-language viewpoint alignment

To detect the number of common viewpoints between documents in two languages, we evaluate the ROUGE performance of TaMVS with varying numbers of common viewpoints $|\mathcal{V}_C|$. Using the same numbering of topics as in Table 5.2, Figure 5.5 shows the number of shared viewpoints $\mathcal{V}_C$ for our 6 test topics; we find that Weibo users have more

Table 5.6: Task: dynamic viewpoint modeling. **RQ3.1:** Contrasting the performance of VTTM and VTTM-S in the Chinese viewpoint modeling task.

	VTTM		VTTM-S	
Topic	pur.	acc.	pur.	acc.
World Economic Forum	**0.497**	**0.516**	0.496	0.513
Whaling hunting	**0.454**	**0.463**	0.449	0.459
FIFA Worldcup 2014	**0.472**	**0.423**	0.441	0.459
Missing MH370	**0.463**	**0.471**	0.433	0.448
Anti-Chinese in Vietnam	**0.491**	**0.511**	0.456	0.471
Sinking of the MV Sewol	**0.425**	**0.438**	0.422	0.435
Overall	**0.474**	**0.482**	0.461	0.474

Table 5.7: Task: dynamic viewpoint modeling. **RQ3.1:** Comparison of methods. Purity is abbreviated to as pur., Accuracy as acc. We use ▲ to denote statistically significant improvements of VTTM over the baseline TAM.

	English tweets						Chinese weibos					
	VTTM		TAM		Sen-LDA		VTTM		TAM		Sen-LDA	
Topic	pur.	acc.	pur.	acc.	pur.	acc.	pur.	acc.	pur.	acc.	pur.	acc.
World Economic Forum	**0.497**▲	**0.516**▲	0.401	0.415	0.419	0.425	**0.441**▲	**0.472**▲	0.352	0.371	0.391	0.407
Whaling hunting	**0.454**	**0.463**	0.432	0.435	0.451	0.462	0.493	0.505	0.442	0.458	**0.501**	**0.513**
FIFA Worldcup 2014	**0.472**△	**0.423**△	0.432	0.442	0.445	0.451	**0.541**▲	**0.561**▲	0.432	0.442	0.483	0.497
Missing MH370	**0.463**▲	**0.471**▲	0.391	0.403	0.427	0.445	**0.501**▲	**0.542**▲	0.343	0.352	0.451	0.462
Anti-Chinese in Vietnam	**0.491**▲	**0.511**▲	0.406	0.415	0.452	0.557	**0.522**▲	**0.541**▲	0.482	0.495	0.503	0.517
Sinking of the MV Sewol	**0.425**	**0.438**	0.361	0.372	0.407	0.411	**0.625**▲	**0.642**▲	0.497	0.507	0.559	0.572
Overall	**0.474**▲	**0.482**▲	0.384	0.397	0.417	0.428	**0.524**▲	**0.543**▲	0.437	0.452	0.482	0.504

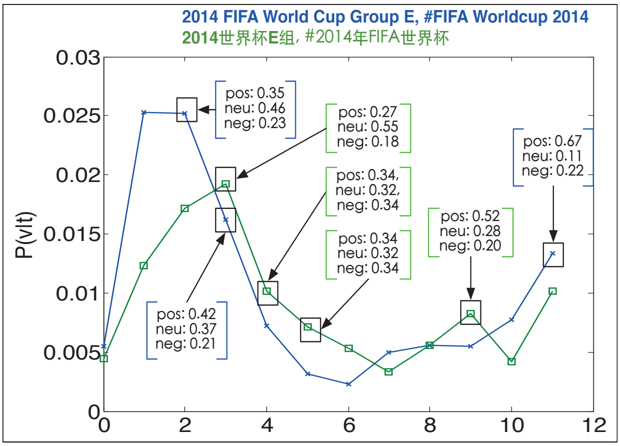


Figure 5.4: Task: dynamic viewpoint modeling. **RQ3.1:** An example viewpoint about "2014 FIFA WorldCup Group E" propagation for "#FIFA Worldcup 2014." The blue (green) text box indicates the probability distribution of English (Chinese) viewpoints' sentiment labels at a specific time interval; the blue (green) curve shows the English (Chinese) viewpoint distribution $\pi_{t,v}$ over the whole timeline.

Table 5.8: Task: cross-language viewpoint alignment. **RQ3.2:** Performance of CLVA in cross-language viewpoints alignment task, in terms of Accuracy.

Topic	CLVA	CLVA-T	CLVA-E
World Economic Forum	**0.754**	0.613	0.591
Whaling hunting	**0.737**	0.671	0.622
FIFA Worldcup 2014	**0.643**	0.588	0.521
Missing MH370	**0.727**	0.611	0.524
Anti-Chinese in Vietnam	**0.787**	0.732	0.655
Sinking of the MV Sewol	**0.854**	0.712	0.659
Overall	**0.711**	0.669	0.615

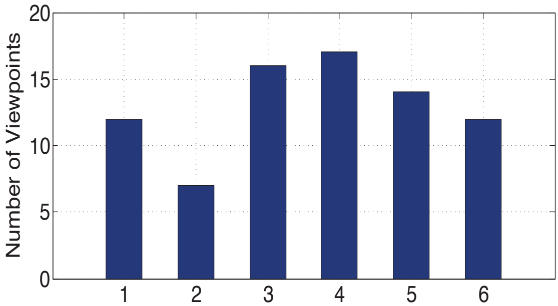


Figure 5.5: Task: cross-language viewpoint alignment. **RQ3.2:** Length of common viewpoints $\mathcal{V}_C$ in 6 topics. The numbers on the x-axis correspond to the topic numbers in Table 5.2.

common viewpoints with Twitter users on the topics “#Missing MH370” and “#FIFA Worldcup 2014” than on other topics. To test the effectiveness of our cross-language viewpoint alignment strategy in **RQ3.2**, we examine the performance of CLVA for every topic; see Table 5.8. CLVA outperforms the other two methods, CLVA-T and CLVA-E, for each topic. We find that CLVA-T outperforms CLVA-E on the cross-language viewpoint alignment task.

5.4.3 Overall performance

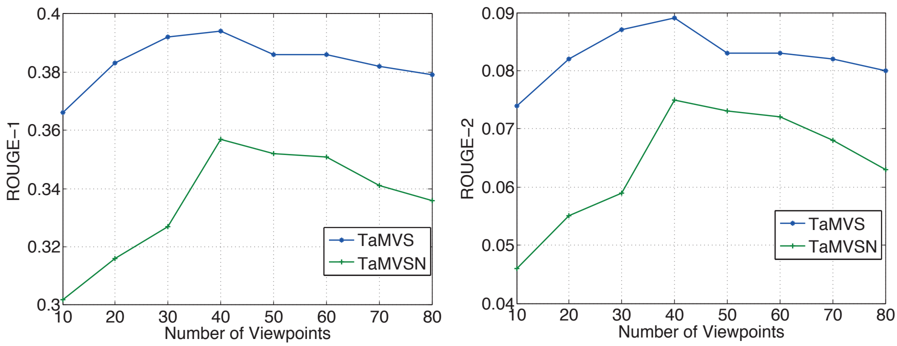
Tables 5.9 and 5.10 show the per topic time-aware multi-viewpoint summarization performance of all methods in terms of the ROUGE metrics. We begin by examining the importance of cross-language viewpoint alignment. Looking at Table 5.9, we see that TaMVS (columns 2–4) significantly outperforms TaMVS-V in which we leave out the cross-language viewpoint alignment step for each topic, and that it does so for all metrics (columns 5–7). This shows the importance of cross-language viewpoint alignment in multi-viewpoint summarization.

Turning to **RQ3.3**, to determine the contribution of novelty and coverage, we turn to Table 5.9, where columns 2–4, 8–10 and 11–13 show the performance of TaMVS, TaMVSN, and TaMVSC, respectively in terms of the ROUGE metrics. Recall that

TaMVSN only considers novelty in phase C and that TaMVSC only considers coverage in phase C. We find that TaMVS, which combines novelty and coverage, outperforms both TaMVSN and TaMVSC on all topics. After TaMVS, TaMVSN, which only includes novelty during the summarization process, performs best. Thus, from Table 5.9 we conclude that novelty is the most important part during our multi-viewpoint summarization process.

Turning to Table 5.10, we find that TaMVS outperforms the baselines on all test topics in terms of ROUGE-1, and in several cases significantly so. In terms of ROUGE-2, we see a similar picture: TaMVS outperforms the baselines, and in several cases significantly so. Meanwhile, among the baselines, LexRank gets the worst performance simply because it ignores the dynamic patterns during viewpoint modeling. And CoRUS achieves the second best performance, which indicates the importance of update summarization in our viewpoint summarization. TaMVS achieves a 3.2% and 7.5% increase over CoRUS in terms of ROUGE-1 and ROUGE-2, respectively, whereas it gives 12.1% and 37.1% increase over IUS in terms of ROUGE-1 and ROUGE-2. Compared to Sen-TM, TaMVS achieves a statistical significant improvement of up to 28.1% in terms of ROUGE-1 and 63.4% in terms of ROUGE-2. Interestingly, TaMVS performs better on test topics that have higher scores for dynamic viewpoint modeling (phase A, see Table 5.7), which underlines the importance of dynamic viewpoint modeling in time-aware multi-viewpoint summarization.

We now analyze the influence of the number of viewpoints. Figure 5.6 plots the average ROUGE performance curves for TaMVS and TaMVSN with varying numbers of



(a) ROUGE-1 (b) ROUGE-2

Figure 5.6: Task: time-aware multi-viewpoint summarization. **RQ3.3:** Performance with different numbers of viewpoints, in terms of ROUGE-1 (a) and ROUGE-2 (b).

viewpoints. We that find for both metrics and methods, the performance peaks when the number of viewpoints equals 40, i.e., higher than our default value of 20.

5.5 Conclusion and Future Work

We have considered the task of time-aware multi-viewpoint summarization of social text streams. We have identified four main challenges: ambiguous entities, viewpoint drift, multilinguality, and the shortness of social text streams. We have proposed a dynamic viewpoint modeling strategy to infer multiple viewpoints in the given multilingual social text steams, in which we jointly model topics, entities and sentiment labels. After cross-language viewpoint alignment, we apply a random walk ranking strategy to extract documents to tackle the time-aware multi-viewpoint summarization problem. In our experiments, we have provided answers to the main research question raised at the beginning of this chapter:

RQ3: Can we find an approach to help detect time-aware viewpoint drift? Can we find an approach to help detect viewpoints from multilingual social text streams? How can we generate summaries to reflect viewpoints of multi-lingual social text streams?

To answer this research question, we collect a dataset of microblogs in two languages, and we obtain our annotations using the CrowdTruth platform. We have considered some existing work as baselines in our experiments, including recent work on topic modeling and update summarization. We have demonstrated the effectiveness of our proposed method by showing a significant improvement over various baselines tested with a manually annotated dataset. Our viewpoint tweet topic model is helpful for detecting the viewpoint drift phenomenon and summarizing viewpoints over time.

Although we focused mostly on microblogs, our methods are broadly applicable to other settings with opinionated content, such as comment sites or product reviews. Limitations of our work include its ignorance of viewpoint dependencies, viewpoint diversity and, being based on LDA, its predefined number of viewpoints. As to future work, *contrastive* viewpoints in multilingual text streams are worth considering. Also, the transfer of our approach to a non-parametric extension should give new insights and an extrinsic online user evaluation would give deeper insights into the performance of our approach. A novel graphical model that includes dynamic time bins instead of the fixed time granularities, is another direction for future research. Finally, discovering new entities that are not included by Wikipedia will help our approach to explore realtime viewpoints. We have already addressed social media summarization in Chapters 3–5. In the next chapter, we change our research angle to the hierarchical multi-label classification of social text streams.

Table 5.9: Task: time-aware multi-viewpoint summarization. **RQ3.2 and RQ3.3:** ROUGE performance of all VTTM-based methods in time-aware viewpoint summarization. ROUGE-1 is abbreviated as R-1, ROUGE-2 as R-2 and ROUGE-W as R-W. Statistically significant differences are with respect to TaMVS-V.

	TaMVS			TaMVS-V			TaMVSN			TaMVSC		
Topic	R-1	R-2	R-W	R-1	R-2	R-W	R-1	R-2	R-W	R-1	R-2	R-W
World Economic Forum	**0.383**$^\blacktriangle$	**0.082**$^\blacktriangle$	**0.184**$^\blacktriangle$	0.316	0.055	0.131	0.369	0.057	0.136	0.351	0.052	0.145
Whaling hunting	**0.294**$^\blacktriangle$	**0.047**$^\blacktriangle$	**0.152**$^\blacktriangle$	0.197	0.037	0.092	0.221	0.032	0.138	0.221	0.032	0.138
FIFA Worldcup 2014	**0.436**$^\blacktriangle$	**0.094**$^\blacktriangle$	**0.202**$^\blacktriangle$	0.351	0.068	0.142	0.404	0.083	0.194	0.404	0.083	0.194
Missing MH370	**0.425**$^\blacktriangle$	**0.087**$^\triangle$	**0.232**$^\blacktriangle$	0.324	0.074	0.132	0.402	0.082	0.170	0.394	0.079	0.162
Anti-Chinese in Vietnam	**0.409**$^\blacktriangle$	**0.065**$^\blacktriangle$	**0.169**$^\blacktriangle$	0.226	0.046	0.127	0.403	0.047	0.139	0.395	0.042	0.129
Sinking of the MV Sewol	**0.373**$^\blacktriangle$	**0.064**$^\blacktriangle$	**0.171**$^\blacktriangle$	0.314	0.051	0.131	0.362	0.055	0.134	0.341	0.046	0.133
Overall	**0.387**$^\blacktriangle$	**0.085**$^\blacktriangle$	**0.188**$^\blacktriangle$	0.318	0.061	0.132	0.377	0.062	0.139	0.359	0.056	0.150

Table 5.10: Task: time-aware multi-viewpoint summarization. **RQ3.3:** Per topic performance of all methods. ROUGE-1 is abbreviated as R-1 and ROUGE-2 as R-2. We use $^\blacktriangle$ ($^\triangle$) to denote strong (weak) statistically significant improvements of **TaMVS** over CoRUS.

	TaMVS		TAM		Sen-TM		LexRank		IUS		CoRUS	
Topic	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2
World Economic Forum	**0.383**$^\triangle$	**0.082**$^\triangle$	0.295	0.047	0.298	0.051	0.321	0.072	0.347	0.066	0.372	0.077
Whaling hunting	**0.294**	**0.047**	0.237	0.038	0.224	0.032	0.221	0.032	0.273	0.042	0.292	0.044
FIFA Worldcup 2014	**0.436**$^\blacktriangle$	**0.094**$^\triangle$	0.347	0.076	0.355	0.081	0.342	0.073	0.378	0.080	0.383	0.089
Missing MH370	**0.425**$^\blacktriangle$	**0.087**$^\blacktriangle$	0.297	0.049	0.314	0.062	0.341	0.072	0.358	0.071	0.352	0.082
Anti-Chinese in Vietnam	**0.409**$^\blacktriangle$	**0.065**$^\blacktriangle$	0.222	0.039	0.239	0.042	0.238	0.040	0.269	0.046	0.253	0.052
Sinking of the MV Sewol	**0.373**$^\triangle$	**0.064**	0.284	0.044	0.295	0.049	0.251	0.040	0.327	0.071	0.369	0.062
Overall	**0.387**$^\triangle$	**0.085**	0.297	0.047	0.302	0.052	0.309	0.052	0.345	0.062	0.375	0.079

6 Hierarchical Multi-Label Classification of Social Text Streams

The previous three research chapters focused on research about social media summarization. In this chapter, we change our research angle to the hierarchical multi-label classification of social text streams. Short text classification is an effective way of assisting users in understanding documents in social text streams [141, 143, 169, 268]. Straightforward text classification methods [102, 216, 258], however, are not adequate for mining documents in social streams.

For many social media applications, a document in a social text stream usually belongs to multiple labels that are organized in a hierarchy. This phenomenon is widespread in web forums, question answering platforms, and microblogs [42]. In Figure 6.1 we show an example of several classes organized in a tree-structured hierarchy, of which several subtrees have been assigned to individual tweets. The tweet "I think the train will soon stop again because of snow ..." is annotated with multiple hierarchical labels: "Communication," "Personal experience" and "Complaint." Faced with many millions of documents every day, it is impossible to manually classify social streams into multiple hierarchical classes. This motivates the *hierarchical multi-label classification* (HMC) task for social text streams: classify a document from a social text stream using multiple labels that are organized in a hierarchy.

Recently, significant progress has been made on the HMC task, see, e.g., [28, 34, 40]. However, the task has not yet been examined in the setting of social text streams. Compared to HMC on stationary documents, HMC on documents in social text streams faces specific challenges: (1) Because of *topic drift* a document's statistical properties change over time, which makes the classification output different at different times.(2) The shortness of documents in social text streams hinders the classification process.Therefore, we ask the following research question listed in Chapter 1:

RQ4: Can we find a method to classify short text streams in a hierarchical multi-label classification setting? How should we tackle the *topic drift* and *shortness* in hierarchical multi-label classification of social text streams?

To answer the above research question, in this chapter, we address the HMC problem for documents in social text streams. We utilize structural support vector machines (SVMs)

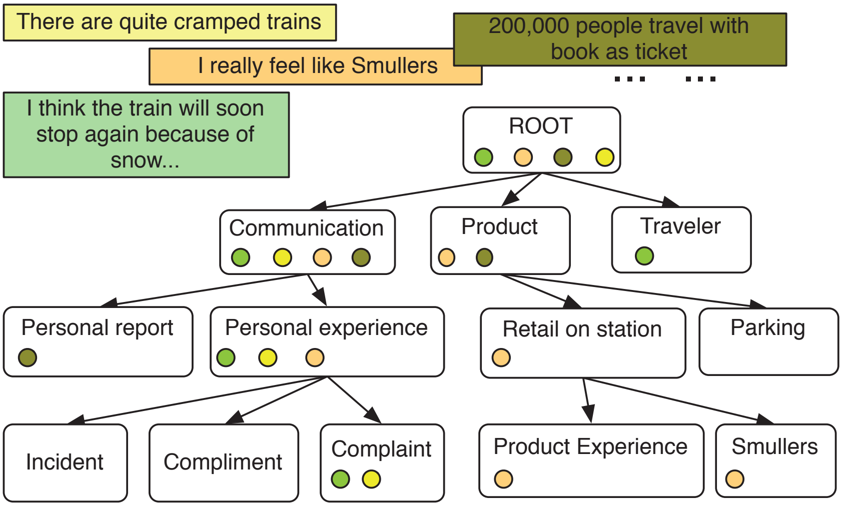


Figure 6.1: An example of predefined labels in hierarchical multi-label classification of documents in a social text stream. Documents are shown as colored rectangles, labels as rounded rectangles. Circles in the rounded rectangles indicate that the corresponding document has been assigned the label. Arrows indicate hierarchical structure between labels.

[233]. Unlike with standard SVMs, the output of structural SVMs can be a complicated structure, e.g., a document summary, images, a parse tree, or movements in video [125, 264]. In our case, the output is a 0/1 labeled string representing the hierarchical classes, where a class is included in the result if it is labeled as 1. For example, the annotation of the top left tweet in Figure 6.1 is `1100010000100`. Based on this structural learning framework, we use multiple structural classifiers to transform our HMC problem into a chunk-based classification problem. In chunk-based classification, the hierarchy of classes is divided into multiple chunks.

To address the shortness and topic drift challenges mentioned above, we proceed as follows. Previous solutions for working with short documents rely on extending short documents using a large external corpus [181]. In this chapter, we employ an alternative strategy involving both entity linking [171] and sentence ranking to collect and filter relevant information from Wikipedia. To address topic drift [9, 56, 57, 169, 223], we track dynamic statistical distributions of topics over time. Time-aware topic models, such as dynamic topic models (DTM) [31], are not new. Compared to latent Dirichlet allocation (LDA) [32], dynamic topic models are more sensitive to bursty topics. A *global* topic is a stationary latent topic extracted from the whole document set and a *local* topic is a dynamic latent topic extracted from a document set within a specific time period. To track dynamic topics, we propose an extension of DTM that extracts both global and local topics from documents in social text streams.

Previous work has used Twitter data for streaming short text classification [169]. So do we. We use a large real-world dataset of tweets related to a major public transportation system in a European country to evaluate the effectiveness of our proposed methods for hierarchical multi-label classification of documents in social text streams. The tweets were collected and annotated as part of their online reputation management campaign. As we will see, our proposed method offers statistically significant improvements over state-of-the-art methods.

Our contributions can be summarized as follows:

- We present the task of hierarchical multi-label classification for streaming short texts.
- We use document expansion to address the shortness issue in the HMC task for short documents, which enriches short texts using Wikipedia articles. We tackle the time-aware challenge by developing a new dynamic topic model that distinguishes between local topics and global topics.
- Based on a structural learning framework, we transform our hierarchical multi-label classification problem into a chunk-based classification problem via multiple structural classifiers, which is shown to be effective in our experiments using a large-scale real-world dataset.

In §6.1 we formulate our research problem. We describe our approach in §6.2; §6.3 details our experimental setup and §6.4 presents the results; §6.5 concludes the chapter.

6.1 Problem Formulation

In this section, we detail the task that we address and introduce important concepts. We begin by defining the hierarchical multi-label classification (HMC) task. We are given a class hierarchy $(C, \prec)$, where C is a set of class labels and $\prec$ is a partial order representing the parent relationship, i.e., $\forall c_i, c_j \in C$, $c_i \prec c_j$ if and only if c_i is the parent class of c_j. We write $\mathbf{x}^{(i)}$ to denote a feature vector, i.e., an element of the feature space $\mathcal{X}$, and we write $\mathbf{y}^{(i)} \in \{0,1\}^{|C|}$ for the target labeling. Let D be the set of input documents, and $|D|$ the size of D. The target of a hierarchical multi-label classifier, whether for stationary documents or for a stream of documents, is to learn a hypothesis function f: $\mathcal{X} \to \{0,1\}^C$ from training data $\{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{|D|}$ to predict a $\mathbf{y}$ when given $\mathbf{x}$. Suppose the hierarchy is a tree structure. Then, classes labeled positive by $\mathbf{y}$ must satisfy the *$\mathcal{T}$-property* [28]: if a labeled $c \in C$ is labeled positive in output $\mathbf{y}$, its parent label must also be labeled positive in $\mathbf{y}$. Given the *$\mathcal{T}$-property*, we define a *root* class r in the beginning of each C, which refers to the root vertex in HMC tree structure. Thus for each $\mathbf{y}$ in HMC, we have $\mathbf{y}^{(r)} = 1$.

Hierarchical multi-label classification for short documents in social streams (HMC-SST) learns from previous time periods and predicts an output when a new document arrives. More precisely, given a class hierarchy $(C, \prec)$ and a collection of documents seen so far, $\mathcal{X} = \{X_1, \ldots, X_{t-1}\}$, HMC-SST learns a hypothesis function f: $\mathcal{X} \to \{0,1\}^C$ that evolves over time. Thus, at time period $t, t > 1$, we are given a function f that has been trained during the past $t-1$ periods and a set of newly arriving documents X_t. For each $\mathbf{x}_t^{(i)} \in X_t$, $f(x)$ predicts $\hat{y}_t^{(i)}$ that labels each class $c \in C$ as 0 or 1. Classes in C that are labeled positive must follow the $\mathcal{T}$-property. Afterwards, f updates its parameters using X_t and their true labels $\{\mathbf{y}_t^{(i)}\}_{i=1}^{|X_t|}$.

Topic drift indicates the phenomenon that topic distributions change between adjacent time periods [73]. In streaming classification of documents [169] this problem needs to

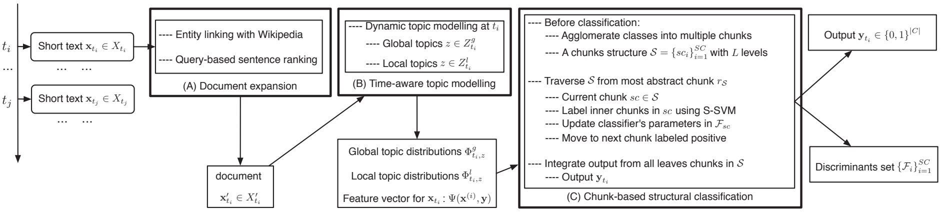


Figure 6.2: Overview of our approach to hierarchical multi-label classification of documents in social text streams. (A) indicates document expansion; (B) indicates the topic modeling process; (C) refers to chunk-based structural learning and classification.

be addressed. We assume that each document in a stream of documents is concerned with multiple topics. By dividing the timeline into time periods, we dynamically track latent topics to cater the phenomenon of *topic drift* over time. For streaming documents, global statistics such as tf-idf or topic distributions cannot reflect drift phenomena. However, local statistics derived from a specific period are usually helpful for solving this problem [31, 120, 169]. Ideally, one would find a trade-off between tracking the extreme local statistics and extreme global statistics [120]. Thus, in this chapter we address the issue of topic drift by tracking both global topics (capturing the complete corpus) and local, latent and temporally bounded, topics over time. Given a document set X_t published at time t, we split the topic set Z_t into $Z_t^g \cup Z_t^l$, with global topics Z_t^g that depend on all time periods and documents seen so far, and local topics Z_t^l derived from the previous period $t-1$ only. We then train our temporal classifier incrementally based on those global and local topic distributions.

6.2 Method

We start by providing an overview of our approach to HMC for documents in social text streams. We then detail each of our three main steps: document expansion, topic modeling and incremental structural SVM learning.

6.2.1 Overview

We provide a general overview of our scenario for performing HMC on (short) documents in social text streams in Figure 6.2. There are three main phases: (A) document expansion; (B) time-aware topic modeling; (C) chunk-based structural classification. To summarize, at time period t_i, we are given a temporally ordered short documents set $X_{t_i} = \{\mathbf{x}_{t_i}^{(1)}, \mathbf{x}_{t_i}^{(2)}, \ldots, \mathbf{x}_{t_i}^{(|X_t|)}\}$. For each short text $\mathbf{x}_{t_i} \in X_{t_i}$, in phase (A) (see §6.2.2) we expand $\mathbf{x}_{t_i}$ through entity linking and query-based sentence ranking; we obtain $\mathbf{x}'_{t_i}$ from $\mathbf{x}_{t_i}$ by extracting relevant sentences from related Wikipedia articles.

Next, in phase (B) (see §6.2.3), we extract dynamic topics Φ_{t_i}; building on an extended DTM model, we extract both global and local topical distributions for $\mathbf{x}'_{t_i}$; then, a feature vector for $\mathbf{x}'_{t_i}$ is generated as $\Psi(\mathbf{x}'^{(i)}, \mathbf{y})$.

Based on the extracted features, we train an incremental chunk-based structural learning framework in (C) in §6.2.4. We introduce multiple structural classifiers to the optimization problem by transferring the set of classes C to another representation using multiple chunks $\mathcal{S}$. Traversing from the most abstract chunk $r_{\mathcal{S}} \in \mathcal{S}$, we define each chunk $s \in \mathcal{S}$ to be a set of chunks or classes. Leaves in $\mathcal{S}$ only include classes. For each chunk $sc \in \mathcal{S}$, we employ a discriminant to address the optimization problem over parameters $\mathcal{F}_{sc}$, where sc's child chunk/class will not be addressed unless it is labeled positive during our prediction. Accordingly, multiple discriminants are applied to predict labels given $\mathbf{x}_{t_i}$ and update their parameters based on true labels $\mathbf{y}_{t_i}$.

6.2.2 (A) Document expansion

To address the challenge offered by short documents, we propose a document expansion method that consists of two parts: entity linking and query-based sentence ranking and extraction.

Entity linking

Given a short document $\mathbf{x}_t$ at time t, the target of entity linking is to identify the entity e from a knowledge base E that is the most likely referent of $\mathbf{x}_t$. For each $\mathbf{x}_t$, a *link candidate* $e_i \in E$ links an *anchor* a in $\mathbf{x}_t$ to a target w, where an anchor is a word n-gram tokens in a document and each w is a Wikipedia article. A target is identified by its unique title in Wikipedia.

As the first step of our entity linking, we aim to identify as many *link candidates* as possible. We perform lexical matching of each n-gram anchor a of document d_t with the target texts found in Wikipedia, resulting in a set of *link candidates* E for each document d_t. As the second step, we employ the commonness (*CMNS*) method from [158] and rank link candidates E by considering the prior probability that anchor text a links to Wikipedia article w:

$$CMNS(a, w) = \frac{|E_{a,w}|}{\sum_{w' \in W} |E_{a,w'}|}, \tag{6.1}$$

where $E_{a,w}$ is the set of all links with anchor text a and target w. The intuition is that link candidates with anchors that always link to the same target are more likely to be a correct representation. In the third step, we utilize a learning to rerank strategy to enhance the precision of correct link candidates. We extract a set of 29 features proposed in [158, 171], and use a decision tree-based approach to rerank the link candidates.

Query-based sentence ranking

Given the *link candidates* list, we extract the most central sentences from the top three most likely Wikipedia articles. As in LexRank [63], Markov random walks are employed to optimize the ranking list iteratively, where each sentence's score is voted from other sentences. First, we build the similarity matrix M, where each item in M indicates the similarity between two sentences given $\mathbf{x}_t$ as a query. Given two sentences s_i and s_j, we have:

$$M_{i,j} = sim(s_i, s_j|\mathbf{x}_t) / \sum_{j' \in |S|} sim(s_i, s_{j'}|\mathbf{x}_t), \tag{6.2}$$

At the beginning of the iterative process, an initial score for each sentence is set as $1/|S|$, and at the t-th iteration, the score of s_i is calculated as follows:

$$score(s_i)^{(t)} = (1-\lambda) \sum_{i \neq j} M_{i,j} \cdot score(s_j)^{(t-1)} + \lambda \frac{1}{|S|}, \tag{6.3}$$

where $|S|$ equals the number of sentences in Wikipedia documents that have been linked to the anchor text a in Eq. 6.1 and the damping factor $\lambda = 0.15$. Then the transition matrix $\widetilde{M}$ equals to:

$$\widetilde{M} = (1-\lambda)M + \bar{e}\bar{e}^T \lambda / |S|, \tag{6.4}$$

where $\bar{e}$ is a column vector with all items equal to 1. The iterative process will stop when it convergences. Since $\widetilde{M}$ is a column stochastic matrix, it can be proven that the value of *score* converges [241], and a value of *score* can be derived from the principle eigenvector of $\widetilde{M}$. We extract the top $\mathcal{E}_{\mathbf{x}_t}$ sentences from the ranked list, and extend $\mathbf{x}_t$ to $\mathbf{x}'_t$ by including those $\mathcal{E}_{\mathbf{x}_t}$ sentences in $\mathbf{x}_t$.

6.2.3 (B) Time-aware topic modeling

Topic drift makes tracking the change of topic distributions crucial for HMC of social text streams. We assume that each document in a social text stream can be represented as a probabilistic distribution over topics, where each topic is represented as a probabilistic distribution over words. The topics are not necessarily assumed to be stationary. We employ a dynamic extension of the LDA model to track latent dynamic topics. Compared to previous work on dynamic topic models [31], our method is based on the conjugate prior between Dirichlet distribution and Multinomial distribution. To keep both stationary statistics and temporary statistics, we present a trade-off strategy between stationary topic tracking and dynamic topic tracking, where topic distributions evolve over time.

Figure 6.3 shows our graphical model representation, where shaded and unshaded nodes indicate observed and latent variables, respectively. Among the variables related to document set X_t in the graph, z, θ, r are random variables and w is the observed variable; $|X_{t-1}|$, $|X_t|$ and $|X_{t+1}|$ indicate the number of variables in the model. As usual, directed arrows in a graphical model indicate the dependency between two variables; the variables ϕ_t^l depend on variables ϕ_{t-1}^l.

The topic distributions $\theta_{\mathbf{x}_t}$ for a document $\mathbf{x}_t \in X_t$ are derived from a Dirichlet distribution over hyper parameter α. Given a word $w_i \in \mathbf{x}_t$, a topic z_{w_i} for word w_i is derived from a multinomial distribution $\theta_{\mathbf{x}_t}$ over document $\mathbf{x}_t$. We derive a probabilistic distribution ϕ_t over topics $Z_t = Z_t^g \cup Z_t^l$ from a Dirichlet distribution over hyper parameters b_t: if topic $z \in Z^l$, then $b_t = \beta_t^l \cdot \phi_{w_i,t-1}$, otherwise $b_t = \beta^g$. The generative process for our topic model at time $t > 1$, is described in Figure 6.4.

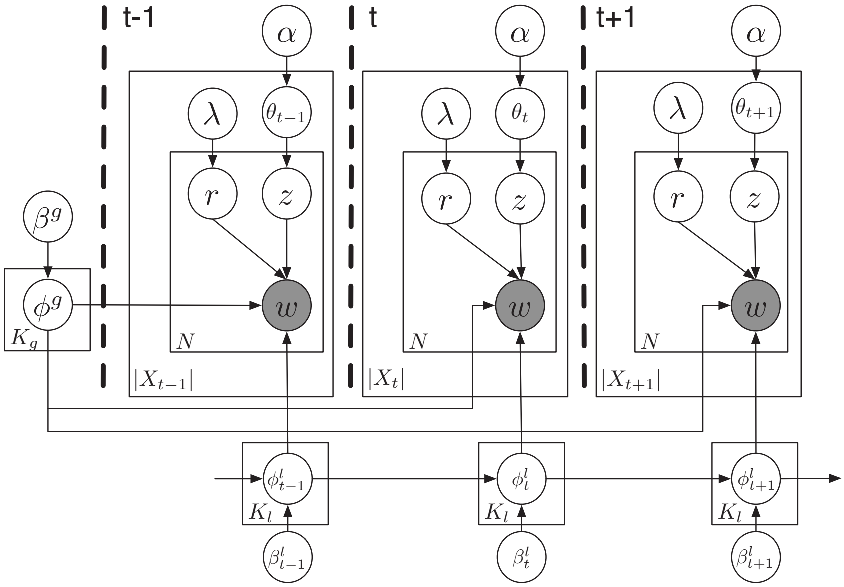


Figure 6.3: Graphical representation of topical modelling, where $t-1$, t and $t+1$ indicate three time periods.

Due to the unknown relation between ϕ_t and θ_t, the posterior distribution for each short text $\mathbf{x}_t$ is intractable. We apply Gibbs collapsed sampling [139] to infer the posterior distributions over both, global and local topics. For each iteration during our sampling process, we derive the topic z via the following probability:

$$p(r_i = m, z_i = z|\mathcal{W}, Z_{-i}, \alpha, b_t) \propto \frac{n^t_{d,m,-i} + \lambda}{n^t_{d,-i} + 2\lambda} \cdot \frac{n^t_{d,z,-i} + \alpha}{\sum\limits_{z' \in Z^m} (n^t_{d,z,'-i} + \alpha)} \cdot \frac{n^t_{w,z,-i} + b^m_{w,z,t}}{\sum\limits_{w' \in N_{u,t}} n^t_{w',z,-i} + N_t b^m_{w,z,t}}, \quad (6.5)$$

where m indicates the possible values of variable r for the ith word in document d_t, and the value m indicates the corresponding kind of topics when $r_i = m$. We set $b_{w,z,t} = \beta^l_t \cdot \phi_{w,z,t-1}$ when $r_i = 1$, and $b_{w,z,t} = \beta^g$ when $r_i = 0$. After sampling the probability for each topic z, we infer the posterior distributions for random variable $\phi_{w,z,t}$, which are shown as follows:

$$\begin{aligned} \phi^{r=0}_{w,z,t} &= \frac{n_{w,z,t} + \beta^g}{\sum\limits_{z \in Z^m} n_{w,z,t} + \beta^g} \\ \phi^{r=1}_{w,z,t} &= \frac{n_{w,z,t} + \beta^l_t \cdot \phi_{w,z,t-1}}{\sum\limits_{z \in Z^m} n_{w,z,t} + \beta^l_t \cdot \phi_{w,z,t-1}} \end{aligned} \quad (6.6)$$

1. For each topic $z, z \in Z_t^l \cup Z_t^g$:
    - Draw $\phi^g \sim Dirichlet(\beta^g)$;
    - Draw $\phi_t^l \sim Dirichlet(\beta_t^l \cdot \phi_{t-1}^l)$;
2. For each candidate short text $\mathbf{x}_t \in X_t$:
    - Draw $\theta_t \sim Dirichlet(\alpha_t)$;
    - For each word w in d_t
        - Draw $r \sim Bernoulli(\lambda)$;
        - Draw $z_w \sim Multinomial(\theta_t)$;
            * if $r = 0$: Draw $w \sim Multinomial(\phi_z^g)$;
            * if $r = 1$: Draw $w \sim Multinomial(\phi_{z,t}^l)$;

Figure 6.4: Generative process for the topic model.

6.2.4 (C) Chunk-based structural classification

Some class labels, specifically for some leaves of the hierarchy, only have very few positive instances. This skewness is a common problem in hierarchical multi-label classification [28]. To handle skewness, we introduce a multi-layer chunk structure to replace the original class tree. We generate this chunk structure by employing a continuous agglomerative clustering approach to merge multiple classes/chunks to a more abstract chunk that contains a predefined number of items. Merging classes, considered as leave nodes in the final chunk structure, our clustering strategy continues until what we call the *root chunk*, the most abstract chunk, has been generated. Following this process, we agglomerate the set of classes C into another set of chunks $\mathcal{S}$, each of which, denoted as sc, includes s items. During this continuous agglomerative clustering process from classes C to the *root chunk*, we define *successive* relations among chunks in $\mathcal{S}$. Each chunk sc's successive chunks/classes in $\mathcal{S}$ are chunks/classes that exist as items in sc, i.e., chunk sc is a successive chunk of chunk sc^{pa} if and only if there exist a vertex in sc^{pa} corresponding to chunk sc.

Thus we think of $\mathcal{S}$ as a tree structure. From the most abstract chunk $r_{\mathcal{S}} \in \mathcal{S}$ that is not included in any other chunk, each layer l of $\mathcal{S}$ is the set of child nodes in those chunks that exist in l's last layer. The leaves of $\mathcal{S}$ indicate classes. Then, a structural SVM classifier $\mathcal{F}_{sc}$ for chunk sc includes L_{sc} chunks, and its output space $\mathcal{Y}_{sc}$ refers to a set of binary labels $\{0,1\}^{L_{sc}}$ over chunks.

At each time period t, we divide the HMC for documents in social text streams into a learning process and a inference process, which we detail below.

Learning with structural SVMs

For the learning process, we train multiple structural SVM classifiers from $\mathcal{S}$'s root chunk $r_{\mathcal{S}}$ to the bottom, where the $\mathcal{T}$-property must be followed by each chunk $sc \in \mathcal{S}$. After generating the chunk structure $\mathcal{S}$, we suppose $\mathcal{S}$ has SC chunks with L levels. At time t, we are given a set of training instances $\mathcal{T}_t = \{(\mathbf{x}_t^{(1)}, \mathbf{y}_t^{(1)}), (\mathbf{x}_t^{(2)}, \mathbf{y}_t^{(2)}), \ldots, (\mathbf{x}_t^{(|X_t|)}, \mathbf{y}_t^{(|X_t|)})\}$, and our target is to update parameters of multiple structural SVM classifiers during the learning process. Thus $\mathbf{y}_t^{(i)}$ in $(\mathbf{x}_t^{(i)}, \mathbf{y}_t^{(i)})$ is divided and extended into SC parts $\bigcup_{sc \in S}\{\mathbf{y}_{t,sc}^{(i)}\}$, where $\mathbf{y}_{t,sc}^{(i)}$ indicates the output vector in chunk sc. The structural classifier $\mathcal{F}_{sc}$ for chunk $sc \in \mathcal{S}, sc \neq r_c$, learns and updates its parameters after its parent chunk $p(sc)$ has received a positive label on the item corresponding to sc. For each chunk $sc \in \mathcal{S}$, we utilize the following structural SVM formulation to learn a weight vector **w**, shown in Eq. 6.7:

$$\min_{\zeta \geq 0} \frac{1}{2} \|\mathbf{w}_{t,sc}\|^2 + C \sum_{i=1}^{n} \zeta_i \tag{6.7}$$

subject to:

1. $\forall \mathbf{y}_{t,sc} \in \mathcal{Y}_{sc} \backslash \mathbf{y}_{t,sc}^{(i)}$;
2. $\forall c \in \mathbf{c}_{\mathbf{y}_{t,sc}}, p(c) \in \mathbf{c}_{\mathbf{y}_{t,sc}}$;
3. $w^T \Psi(\mathbf{x}_t^{(i)}, \mathbf{y}_{t,sc}^{(i)}) - w^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}_{t,sc}) \geq \Delta(\mathbf{y}, \mathbf{y}_{t,sc}^{(i)}) - \zeta_i$;

where $\mathbf{c}_{\mathbf{y}_{t,sc}}$ are positive chunks labeled by $\mathbf{y}_{t,sc}^{(i)}$, and $\Psi(\mathbf{x}_t^{(i)}, \mathbf{y}_{t,sc})$ indicates the feature representation for $\mathbf{x}_t^{(i)}, \mathbf{y}_{t,sc}^{(i)}$.

Traditional SVMs only consider zero-one loss as a constraint during learning. This is inappropriate for complicated classification problems such as hierarchical multi-label classification. We define a loss function between two structured labels $\mathbf{y}$ and $\mathbf{y}_i$ based on their similarity as $\Delta(\mathbf{y}_{sc}, \mathbf{y}_{i,sc}) = 1 - sim(\mathbf{y}_{sc}, \mathbf{y}_{i,sc})$. Here, $sim(\mathbf{y}_{sc}, \mathbf{y}_{i,sc})$ indicates the structural similarity between two different subsets of sc's child sets $\mathbf{c}_{\mathbf{y}}$ and $\mathbf{c}_{\mathbf{y}^{(i)}}$. We compute the similarity between $\mathbf{y}_{t,sc}$ and $\mathbf{y}_{t,sc}^{(i)}$ by comparing the overlap of nodes in these two tree structures, as follows:

$$sim(\mathbf{y}_{t,sc}^{(i)}, \mathbf{y}_{t,sc}) = \frac{\sum\limits_{n \in \mathbf{c}_{\mathbf{y}^{(i)}}, n' \in \mathbf{c}_{\mathbf{y}}} w_{n,n'} \cdot |(n \cap n')|}{\sum\limits_{n \in \mathbf{c}_{\mathbf{y}^{(i)}}, n' \in \mathbf{c}_{\mathbf{y}}} w_{n,n'} \cdot |(n \cup n')|}, \tag{6.8}$$

where we set $w_{n,n'}$ to be the weight between two chunks n and n', each of which is included in $\mathbf{c}_{\mathbf{y}^{(i)}}$ and $\mathbf{c}_{\mathbf{y}}$ respectively. Since it is intractable to compare two chunks that are not at the same level in $\mathcal{S}$, here we set $w_{n,n'}$ to be:

$$w_{n,n'} = \begin{cases} 1/h_n & h_n = h_{n'} \\ 0 & else \end{cases} \tag{6.9}$$

To optimize Eq. 6.7, we adjust the cutting plane algorithm [69, 264] to maintain the $\mathcal{T}$-property. In general, the cutting plane algorithm iteratively adds constraints until the problem is solved by a desired tolerance ε. It starts with an empty set $\mathbf{y}_i$, for $i = 1, 2, \ldots, n$, and iteratively looks for the most violated constraint for $(\mathbf{x}_t^{(i)}, \mathbf{y}_{t,sc}^{(i)})$. Algorithm 7 shows that to maintain the $\mathcal{T}$-property, we adjust the set of positive chunks in $\hat{y}$ iteratively. The parameter $\mathbf{w}_{t,sc}$ is updated with respect to the combined working set $\bigcup_i \{\mathbf{y}_i\}$.

Making predictions

The feature representation for $\Psi(\mathbf{x}_t^{(i)}, \mathbf{y}_{t,sc})$ must enable meaningful discrimination between high quality and low quality predictions [264]. Our topic model generates a set

Algorithm 7: Cutting Plane Optimization for Eq. 6.7

Input: $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{y}^{(2)}), ..., (\mathbf{x}^{(t)}, \mathbf{y}^{(t)}), C, \zeta$
$\mathbf{y}_i = \emptyset$;
repeat
  for $i = 1, 2, ... , n$ **do**
    $\omega \equiv w^T \Psi(x^{(i)}, y^{(i)}) - w^T \Psi(x^{(i)}, y)$;
    $H(y; w) \equiv \Delta(y^{(i)}, y) + \omega$;
    compute $\hat{y} = \arg\max_{\mathbf{y} \in Y} H(\mathbf{y}; \mathbf{w})$;
    repeat
      for *leaves node* $n \in sc$ **do**
        if $p(n) \notin \boldsymbol{c}_{\hat{y}}$ **then**
          $\hat{y}+ = \hat{y} \cup p(n)$;
          $\hat{y}- = \hat{y} - n$;
          $\hat{y} = \arg\max_y (H(\hat{y}+; w), H(\hat{y}-; w))$
        end
      end
    until $\hat{y} \in Y$ *hold* $\mathcal{T}$*-property*;
    if $H(\hat{y}; w) > \zeta_i + \varepsilon$ **then**
      $\mathbf{w} \leftarrow$ optimize Eq. 6.7 over $\bigcup_i \{\mathbf{y}_i\}$
    end
  end
until *no working set has changed during iteration*;

of topical distributions, Φ_t, where each item $\phi(w|z,t) \in \Phi_t$ is a conditional distribution $P(w|z,t)$ over words w given topic z. Assuming that each document's saliency is summed up by votes from all words in the document, we then define $\Psi(\mathbf{x}, \mathbf{y})$ as follows:

$$\Psi(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \frac{1}{N_{\mathbf{x}}} \sum\limits_{w \in \mathbf{x}} \phi(w|z_1, t) \cdot \frac{1}{N_{\mathbf{y}}} n_{w,\mathbf{y}} \\ \frac{1}{N_{\mathbf{x}}} \sum\limits_{w \in \mathbf{x}} \phi(w|z_2, t) \cdot \frac{1}{N_{\mathbf{y}}} n_{w,\mathbf{y}} \\ \vdots \\ \frac{1}{N_{\mathbf{x}}} \sum\limits_{w \in \mathbf{x}} \phi(w|z_K, t) \cdot \frac{1}{N_{\mathbf{y}}} n_{w,\mathbf{y}} \end{bmatrix}, \tag{6.10}$$

where $n_{w,\mathbf{y}}$ indicates the number of times word w exist in $\mathbf{y}$ for the past $t-1$ periods; $N_{\mathbf{x}}$ refers to the number of words in documents $\mathbf{x}$ whereas $N_{\mathbf{y}}$ is the number of words in $\mathbf{y}$.

Given multiple structural SVMs $\mathcal{F}_{t,sc}$ that have been updated at time $t-1$, the target of our prediction is to select $\mathbf{y}_{t,sc}$ for instance $\mathbf{x}_t$ from the root chunk $r_{\mathcal{S}} \in \mathcal{S}$ to $\mathcal{S}$'s bottom level. Our selection procedure is shown in Algorithm 8. After prediction and learning at time t, our classifiers are given document set X_{t+1} at time $t+1$. Given a document $\mathbf{x}_{t+1} \in X_{t+1}$, we traverse the whole chunk structure $\mathcal{S}$ from root chunk $r_{\mathcal{S}}$ to leaves, and output the predicted classes that $\mathbf{x}_{t+1}$ belongs to. Parameters in discriminants $\mathcal{F}_{t+1,sc}$ are updated afterwards.

Algorithm 8: Greedy Selection via Chunk Structure $\mathcal{S}$

Input: $\mathcal{S}$, $\mathbf{x}_t$ $\mathbf{w}_{t-1} = \{\mathbf{w}_{t-1,sc}\}_{sc \in \mathcal{S}}$
$\mathbf{y} = \emptyset$;
for $sc = 1, 2, ..., SC$ **do**
  if $sc \in c_{\mathbf{y}_{t,p(sc)}}$ **then**
    $\mathbf{y}_{sc} = \arg\max_{y \in \mathcal{Y}_{sc}, y \neq \mathbf{y}_{sc}} (w^T \Psi(\mathbf{x}_t, \mathbf{y}_{sc} \cup y))$;
  end
  if *sc is leaves chunk in $\mathcal{S}$* **then**
    $\mathbf{y} = \mathbf{y} \cup \mathbf{y}_{sc}$;
  end
end
return ***y***

6.3 Experimental Setup

In §6.3.1, we divide our main research question **RQ4** into five research questions to guide our experiments; we describe our dataset in §6.3.2 and set up our experiments in §6.3.3; §6.3.4 gives details about our evaluation metrics; the baselines are described in §6.3.5.

6.3.1 Research questions

We divide our main research question **RQ4** into five research questions, **RQ4.1** to **RQ4.5**, to guide the remainder of the chapter.

RQ4.1 As a preliminary question, how does our chunk-based method perform in stationary HMC? (See §6.4.1)

RQ4.2 Is our document expansion strategy helpful for classifying documents in a HMC setting? (See §6.4.2)

RQ4.3 Does *topic drift* occur in our streaming short text collection? Does online topic extraction help to avoid *topic drift* on HMC-SST? (See §6.4.3)

RQ4.4 How does our proposed method perform on HMC-SST? Does it outperform baselines in terms of our evaluation metrics? (See §6.4.4)

RQ4.5 What is the effect of we change the size of chunks? Can we find an optimized value of the size of chunks in HMC-SST? (See §6.4.5)

6.3.2 Dataset

General statistics We use a dataset of tweets related to a major public transportation system in a European country. The tweets were posted between January 18, 2010 and June 5, 2012, covering a period of nearly 30 months. The dataset includes $145,692$ tweets posted by 77,161 Twitter users. Using a state-of-the-art language identification tool [38], we found that over 95% tweets in our dataset is written in Dutch, whereas most other tweets are written in English. The dataset has human annotations for each tweet. A diverse set of social media experts produced the annotations after receiving proper training. In total, 81 annotators participated in the process.

Table 6.1: The 13 subsets that make up our dataset, all annotations are in Dutch. The second column shows the English translation, the third column gives the number of tweets per subset, the fourth indicates whether a subset was included in our experiments.

Tag (in Dutch)	Translation	Number	Included
Berichtgeving	Communications	$208,503$	Yes
Aanbeveling	Recommendation	$150,768$	Yes
Bron online	Online source	$2,505$	No
Bron offline	Offline source	$179,073$	Yes
Reiziger	Type of traveler	$123,281$	Yes
Performance	Performance	$28,545$	Yes
Product	Product	$82,284$	Yes
Innovation	Innovation	$114,647$	Yes
Workplace	Workplace	$16,910$	Yes
Governance	Governance	$11,340$	Yes
Bedrijfsgerelateerd	Company related	$15,715$	Yes
Citizenship	Citizenship	628	No
Leadership	Leadership	$10,410$	Yes

The annotation tree for the dataset has 493 nodes. The annotations describe such aspects as reputation dimensions and product attributes and service. All annotators use Dutch during the annotating process. Unlike many other Twitter datasets with human annotations, e.g., Amigó et al. [14], in our dataset those labels are not independent from each other. Instead, each tweet is labeled by multiple hierarchical classes. From the root class, we divide the dataset into 13 individual subsets following the root node's child classes, which are shown in Table 6.1. In our experiment, not all subsets are included in our experiments: we ignore the subset with the fewest tweets: `Citizenship`. As all instances in `Online Source` are annotated by the same labels, we also omit it.

Author and temporal statistics Figure 6.5 shows the number of authors for different numbers of posted tweets in our dataset. Most users post fewer than 200 tweets. In our dataset, $73,245$ users posts fewer than 10 tweets within the whole time period, and the maximum number of tweets posted by one user is $9,293$: this is a news aggregator that accumulates and retweets information about public transportation systems.

One of the most interesting parts of the corpus is the possibility to analyze and test longitudinal temporal statistics. We can display the trends of tweets with various ways of binning. We can look at general developments over long periods of time and bin documents per day and per week. Figure 6.6 shows the total number of tweets posted at each hour over 24 hours. Clearly, people commute in the train: the rush hours between 6am and 8am and between 4pm and 5pm correspond to a larger output of tweets. Figure 6.6 also gives us statistics on the number of tweets posted per day; many more tweets are posted within the period from November 2011 to March 2012, and a peak of the number of tweets happening around February 18, 2012, a day with a lot of delays (according to the uttered tweets).

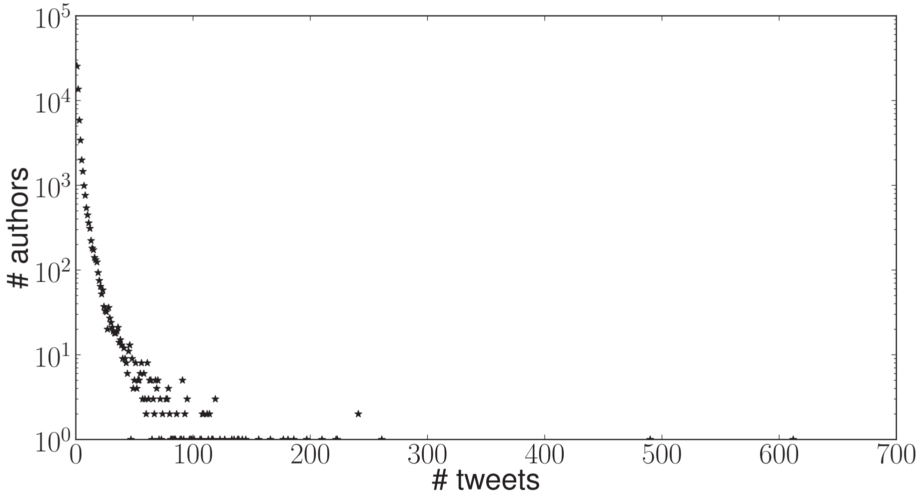
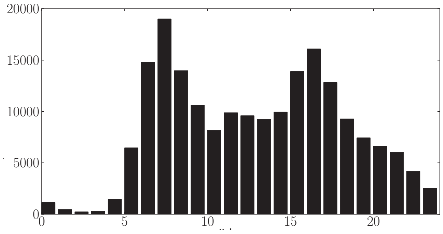
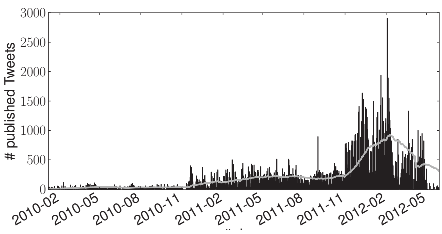


Figure 6.5: Number of tweets per user in our dataset, where the y-axis denotes the number of tweets and the x-axis denotes the corresponding number of tweets the author posted in our dataset. One user with more than 9000 tweets is omitted to improve readability.



(a) Tweets per hour



(b) Tweets per day

Figure 6.6: Number of tweets in our dataset. (Left): number of published tweets published per hour. (Right): number of published tweets published per day.

6.3.3 Experimental setup

Following [190], we set the hyper parameters $\alpha = 50/\left(K^g + K^l\right)$ and $\beta^l = \beta^g = 0.5$ in our experiments. We set $\lambda = 0.2$ and the number of samples to 5000 in our experiment for both document expansion and topic modeling. The number of topics in our topic modeling process is set to 50, for both Z_0^u and Z_0^{com}. For our chunk-based structural SVM classification, we set parameter $C = 0.0001$. For simplicity, we assume that each chunk in our experiments has at most 4 child nodes.

Statistical significance of observed differences between two comparisons is tested using a two-tailed paired t-test. In our experiments, statistical significance is denoted using $^\blacktriangle$ ($^\vartriangle$) for strong (weak) significant differences for $\alpha = 0.01$ ($\alpha = 0.05$). For the stationary HMC evaluation, all experiments are executed using 10-fold cross validation combining training, validation and test sets.

Table 6.2: Baselines and methods used for comparison.

Acronym	Gloss	Reference
C-SSVM	Chunk-based structural learning method	This chapter
NDC-SSVM	C-SSVM without document expansion	This chapter
GTC-SSVM	C-SSVM only with global topics	This chapter
LTC-SSVM	C-SSVM only with local topics	This chapter
Stationary		
CSSA	Kernel density estimation based HMC method	[28]
CLUS-HMC	Decision tree-based HMC method	[237]
H-SVM	Hierarchical SVM for multi-label classification	[50]
Streaming		
H-SVM	Hierarchical SVM for multi-label classification	[50]
CSHC	Structural multi-class learning method	[44]
NBC	Naive Bayesian method	[120]

6.3.4 Evaluation metrics

We adapt *precision* and *recall* to hierarchical multi-label learning following [28]. Given a class $i \in C$, let TP_i, FP_i and FN_i be the number of true positives, false positives and false negatives, respectively. Precision and recall for the whole output tree-structure are:

$$P = \frac{\sum_{i \in C} TP_i}{\sum_{i \in C} TP_i + \sum_{i \in C} FP_i}; \quad R = \frac{\sum_{i \in C} TP_i}{\sum_{i \in C} TP_i + \sum_{i \in C} FN_i} \tag{6.11}$$

We evaluate the performance using *macro* F_1*-measure* (combining precision and recall) and *average accuracy*. The *macro* F_1*-measure* measures the classification effectiveness for each individual class and averages them, whereas *average accuracy* measures the proportion correctly identified. For simplicity's sake, we abbreviate *average accuracy* as accuracy and acc. in §6.4.

6.3.5 Baselines and comparisons

We list the methods and baselines that we consider in Table 6.2. We write C-SSVM for the overall process as described in §6.2, which includes both document expansion and topic tracking. To be able to answer **RQ4.1**, we consider NDC-SSVM, which is C-SSVM without document expansion. Similarly, in the context of **RQ4.2** we consider GTC-SSVM and LTC-SSVM for variations of C-SSVM that only have global topics and local topics, respectively.

There are no previous methods that have been evaluated on the hierarchical multi-label classification of streaming short text. Because of this, we consider two types of baseline: stationary and streaming. For stationary hierarchical multi-label classification, we use CSSA, CLUS-HMC and H-SVM as baselines. We implement CSSA [28] by using kernel dependency estimation to reduce the possibly large number of labels to a manageable number of single-label learning problems. CLUS-HMC [237] is a method

Table 6.3: **RQ4.1:** macro F_1 values for stationary comparisons.

	C-SSVM	CSSA	CLUS-HMC	H-SVM
Communications	**0.5073**	0.5066	0.4812	0.4822
Recommendation	0.4543	**0.4612**	0.4421	0.4452
Offline source	**0.4245**	0.4176	0.4164	0.4161
Type of traveler	0.4623	**0.4677**	0.4652	0.4615
Performance	**0.5221**	0.5109	0.5054	0.5097
Product	**0.4762**	0.4722	0.4686	0.4609
Innovation	**0.4991**	0.4921	0.4822	0.4812
Workplace	0.4645	**0.4725**	0.4687	0.4623
Governance	0.4932	**0.5025**	0.4987	0.4923
Company related	0.4922	**0.4972**	0.4901	0.4852
Leadership	**0.4672**	0.4654	0.4624	0.4602

based on decision trees. H-SVM [50] extends normal SVMs to a hierarchical structure, where the SVM is trained in each node if, and only if, its parent node has been labeled positive. As CSSA and CLUS-HMC need to predefine the number of classes that each document belongs to, we employ MetaLabeler [227] to integrate with those two baselines.

For the streaming short text classification task, besides H-SVM, we implement NBC and CSHC, a naive bayesian classifier framework, which has proved effective in streaming classification [120], and a structural multi-class learning method. Since NBC and CSHC are designed for single-label classification, we introduce a widely-used "one vs. all" strategy on multi-label situation [227]. We evaluate their performance after document expansion (§6.2.2)

6.4 Results and Discussion

In §6.4.1, we compare C-SSVM to other baselines for stationary hierarchical multi-label classification; in §6.4.2 we examine the performance of document expansion. §6.4.3 details the effect of topic modeling on overcoming topic drift; §6.4.4 provides overall performance comparisons; §6.4.5 evaluates the influence of the number of items per chunk.

6.4.1 Performance on stationary HMC

We start by addressing **RQ4.1** and test if our C-SSVM is effective for the stationary HMC task, even though this is not the main purpose for which it was designed. Table 6.3 compares the macro F_1 of C-SSVM to the three HMC baselines. C-SSVM and CSSA tend to outperform the other baselines: for 6 out of 11 tags C-SSVM provides the best performance, while for the remaining 5 CSSA performs best. The performance differences between C-SSVM and CSSA are not statistically significant. This shows that, when compared against state of the art baselines in terms of the macro F_1 metric, C-SSVM is competitive.

Table 6.4: **RQ4.2:** An example of document expansion.

Short text
I'm tempted to get that LG Chocolate Touch. Or at least get a touchscreen phone
Extension
The original LG Chocolate KV5900 was released in Korea long before the UK or U.S. version.
The LG VX8500 or "Chocolate" is a slider cellphone-MP3 player hybrid that is sold as a feature phone.
The sensory information touch, pain, temperature etc., is then conveyed to the central nervous system by afferent neurones ...

Table 6.5: **RQ4.2:** Effect of document expansion in HMC.

	C-SSVM		NDC-SSVM	
Subset	macro-F_1	Acc.	macro-F_1	Acc.
Communication	**0.5073**$^{\blacktriangle}$	**0.5164**$^{\blacktriangle}$	0.4887	0.4972
Recommendation	**0.4543**	**0.4663**	0.4542	0.4655
Offline source	**0.4245**$^{\blacktriangle}$	**0.4523**$^{\blacktriangle}$	0.4112	0.4421
Type of traveler	0.4623	0.4731	**0.4647**	**0.4791**
Performance	**0.5221**$^{\blacktriangle}$	**0.5321**$^{\blacktriangle}$	0.5013	0.5111
Product	**0.4762**$^{\vartriangle}$	**0.4823**$^{\vartriangle}$	0.4612	0.4721
Innovation	**0.4991**$^{\blacktriangle}$	**0.5121**$^{\blacktriangle}$	0.4522	0.4612
Workplace	**0.4645**$^{\vartriangle}$	**0.4724**$^{\vartriangle}$	0.4601	0.4695
Governance	**0.4932**$^{\blacktriangle}$	**0.5072**$^{\blacktriangle}$	0.4787	0.4944
Company related	**0.4922**$^{\blacktriangle}$	**0.5072**$^{\blacktriangle}$	0.4772	0.4921
Leadership	**0.4672**$^{\vartriangle}$	**0.4754**	0.4601	0.4707

6.4.2 Document expansion

Next, we turn to **RQ4.2** and evaluate the effectiveness of document expansion for HMC-SST. As described in §6.2, we extend a short text into a longer document by extracting sentences from linked Wikipedia articles. Table 6.4 shows an example of the document expansion where the new sentences are relevant to the original text.

Table 6.5 contrasts the evaluation results for C-SSVM with that of NDC-SSVM, which excludes documents expansion, in terms of macro-F_1 and average accuracy. We find that C-SSVM outperforms NDC-SSVM for most subsets of stationary HMC comparisons. In terms of macro F_1, C-SSVM offers an increase over NDC-SSVM of up to 9.4%, whereas average accuracy increases by up to 9.9% significantly. We conclude that document expansion is effective for the stationary HMC task, especially for short text classification.

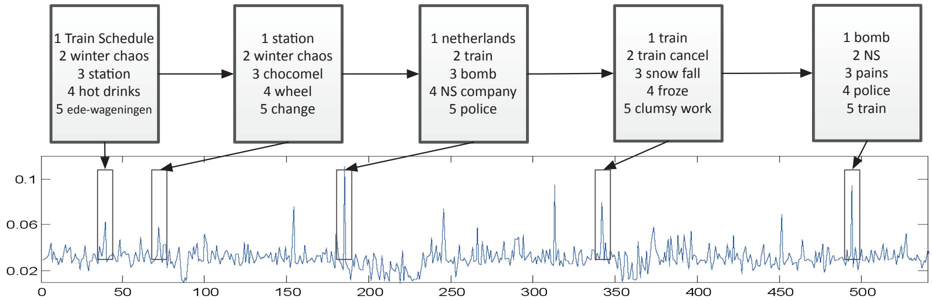


Figure 6.7: **RQ4.3:** An example local topic propagation in the subset "Communication." The text blocks at the top indicate the top 5 representative terms for the topic being propagated at a specific time period; the bottom side shows the topic distribution over the whole timeline.

6.4.3 Time-aware topic extraction

Our third research question **RQ4.3** aims at determining whether topic drift occurs and whether topic extraction helps to avoid this. Figure 6.7 shows the propagation process of an example local topic for the subset "Communication." The upper part of Figure 6.7 shows the 5 most representative terms for the topic during 5 time periods. The bottom half of the figure plots fluctuating topical distributions over time, which indicates topic drift between two adjacent periods.

Figure 6.8 shows the macro F_1 score over time for C-SSVM, C-SSVM with only local topics (LTC-SSVM), and C-SSVM with only globale topics (GTC-SSVM). This helps us understand whether C-SSVM is able to deal with topic drift during classification. We see that the performance in terms of macro F_1 increases over time, rapidly in the early stages, more slowly in the later periods covered by our data set, while not actually plateauing. We also see that the performance curves of LTC-SSVM and GTC-SSVM behave similarly, albeit at a lower performance level. Between LTC-SSVM and GTC-SSVM, LTC-SSVM outperforms GTC-SSVM slightly: local topic distributions are more sensitive, and hence adaptive, when drift occurs.

6.4.4 Overall comparison

To help us answer **RQ4.4**, Table 6.6 lists the macro F_1 and average accuracy for all methods listed in Table 6.2 for all subsets over all time periods. We see that our proposed methods C-SSVM, NDC-SSVM, GTC-SSVM and LTC-SSVM significantly outperform the baselines on most of subsets.

As predicted, NBC performs worse. Using local topics (LTC-SSVM) performs second best (after using both local and global topics), which indicates the importance of dynamic local topics tracking in our streaming classification. C-SSVM achieves a 3.2% (4.5%) increase over GTC-SSVM in terms of macro F_1 (accuracy), whereas the macro F_1 (accuracy) increases 1.9% (2.2%) over LTC-SSVM. Compared to CSHC, C-SSVM offers a statistically significant improvement of up to 7.6% and 8.1% in terms of macro

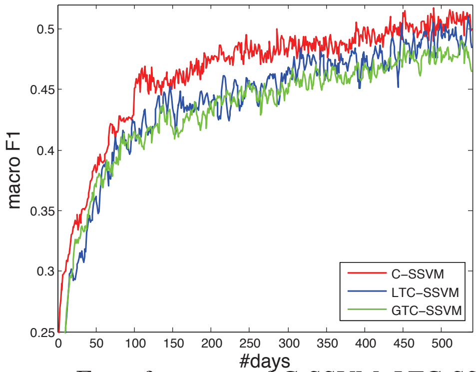


Figure 6.8: **RQ4.3:** macro F_1 performance of C-SSVM, LTC-SSVM and GTC-SSVM over the entire data set.

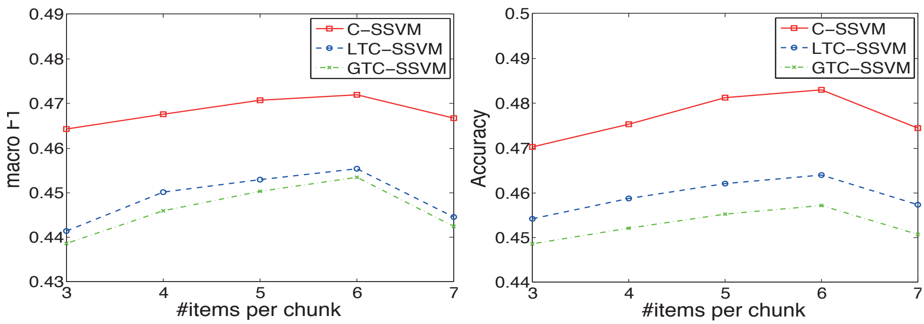
F_1 and accuracy, respectively.

6.4.5 Chunks

We now move on to **RQ4.5**, and analyse the influence of the number of items per chunk. Figure 6.9 plots the performance curves for C-SSVM, LTC-SSVM and GTC-SSVM with varying numbers of items per chunk. While not statistically significant, for both metrics and all three methods, the performance peaks when the number of items equals 6, i.e., higher than our default value of 4.

6.5 Conclusion and Future Work

We have considered the task of hierarchical multi-label classification of social text streams. We have identified three main challenges: the shortness of text, topic drift, and hierarchical labels as classification targets. The first of these was tackled using an entity-based



(a) macro F_1 (b) Accuracy

Figure 6.9: **RQ4.5:** Performance with different numbers of items of each chunk, in terms of macro F_1 (a) and Accuracy (b).

Table 6.6: **RQ4.4:** Performance of all methods on all subsets for all time periods; macro F_1 is abbreviated to m-F_1, average accuracy is written as Acc. We use ▲ and △ to denote significant improvements over **CSHC**. Best performance per subset is indicated in boldface.

	C-SSVM		NDC-SSVM		GTC-SSVM		LTC-SSVM		CSHC		H-SVM		NBC	
Subset	m-F_1	Acc.	m-F_1	Acc.	m-F_1	Acc.	m-F_1	Acc.	m-F_1	Acc.	m-F_1	Acc.	m-F_1	Acc.
Communication	**47.21**▲	**48.16**▲	44.24	45.42	46.44▲	47.68▲	46.25▲	47.82▲	44.12	45.31	45.22	46.62	44.02	45.18
Recommendation	**41.28**▲	**42.52**▲	40.44▲	41.52▲	39.88△	40.24△	40.52▲	41.47▲	38.53	39.42	38.22	39.71	34.31	35.26
Offline source	**40.69**▲	**41.61**▲	39.52▲	40.42▲	39.62▲	41.15▲	40.33▲	41.72▲	36.98	37.43	37.41	38.42	33.21	34.51
Type of traveler	43.73▲	44.61▲	**44.02**▲	**44.96**▲	43.12▲	44.25▲	43.45▲	44.49▲	38.83	40.01	41.07	41.92	38.62	39.38
Performance	**49.52**△	**50.81**△	47.62	48.45	48.86	49.63	48.93	50.02	48.74	49.26	48.84	49.52	46.42	47.32
Product	**44.88**▲	**45.24**▲	43.16▲	44.09▲	44.26▲	45.02▲	44.01▲	45.22▲	41.92	42.85	41.55	42.34	39.21	40.42
Innovation	**46.89**△	**47.68**△	45.58	46.64	45.97	46.81	46.52△	47.51△	45.44	46.56	44.52	45.63	43.41	44.21
Workplace	**43.81**▲	**44.42**▲	43.11▲	44.32▲	42.21▲	43.15▲	42.63▲	43.41▲	36.94	37.22	36.24	37.01	36.59	37.41
Governance	**47.71**▲	**48.44**▲	47.19▲	48.46▲	46.42△	47.35△	47.22△	48.19△	45.61	46.21	46.25	47.36	43.48	44.51
Company related	**47.20**▲	**48.52**▲	46.52▲	47.38▲	46.12▲	47.51▲	46.54▲	47.43▲	43.31	44.99	43.06	44.12	40.91	41.75
Leadership	**44.15**△	**45.88**▲	43.67	44.59	41.75	42.82	42.34	43.21	42.51	43.44	42.15	43.51	40.35	41.27

document expansion strategy. To alleviate the phenomenon of topic drift we have presented a dynamic extension to topic models. This extension tracks topics with topic drift over time, based on both local and global topic distributions. We combine this with an innovative chunk-based structural learning framework to tackle the hierarchical multi-label classification problem. In our experiments, we have provided answers to the main research question raised at the beginning of this chapter:

RQ4: Can we find a method to classify short text streams in a hierarchical multi-label classification setting? How should we tackle the *topic drift* and *shortness* in hierarchical multi-label classification of social text streams?

To answer this research question, we use a dataset of tweets related to a major public transportation system. Because there are no previous methods that have been evaluated on the hierarchical multi-label classification of streaming short text, we consider two types of baseline: stationary and streaming. We have found that local topic extraction in our strategy helps to avoid the topic drift. We have verified the effectiveness of our proposed method in hierarchical multi-label classification of social text streams, showing significant improvements over various baselines tested with a manually annotated dataset of tweets.

As to future work, parallel processing may enhance the efficiency of our method on hierarchical multi-label classification of social text streams. Meanwhile, both the transfer of our approach to a larger social documents dataset and new baselines for document expansion and topic modeling should give new insights. Adaptive learning or semi-supervised learning can be used to optimize the chunk size in our task. Finally, we have evaluated our approaches on fixed time intervals. This might not accurately reflect exact topic drift on social streams. A novel incremental classification method focussing on dynamic time bins opens another direction of future research. In the next chapter, we change our research angle to the explainable recommendation task by tracking viewpoints in social text.

7 Social Collaborative Viewpoint Regression

In the previous four research chapters, we discussed summarization and classification methods that can be used to monitor the content of social media. Given social media text, using content analysis to enhance the performance of recommender systems is another challenging research direction. In this chapter, we address the explainable recommendation task by extracting viewpoints, which are described in our previous research on viewpoint modeling (in Chapter 5). Recommender systems are playing an increasingly important role in e-commerce portals. With the development of social networks, many e-commerce sites have become popular social platforms that help users discuss and select items. Traditionally, a major strategy to predicting ratings in recommender systems is based on collaborative filtering (CF), which infers a user's preference using their previous interaction history. Since CF-based methods only use numerical ratings as input, they suffer from a "cold-start" problem and unexplainable prediction results [89, 137], topics that have received considerable attention in recent years.

Explainable recommendations have been proposed to address the "cold-start" problem and the poor interpretability of recommended results by not only predicting better rating results, but also generating item aspects that attract a user's attention [271]. Most current solutions for explainable recommendations are based on content-based analysis methods [43, 137, 242]. Recent work on explainable recommender systems applies topic models to predict ratings and topical explanations [58, 137], where latent topics are detected from user reviews. Each latent topic in a topic model is represented as a set of words, whereas each item is represented as a set of latent topics. These approaches face two important challenges: (1) Most existing methods neglect to explicitly analyze opinions for recommendation, thereby missing important opportunities to explain users' preferences. (2) Trusted social relations are known to improve the quality of CF recommendation [100, 254], however, current methods for explainable recommendations rarely use this information. Hence in this chapter we ask the following research question:

RQ5: Can we devise an approach to enhance the rating prediction in explainable recommendation? Can user reviews and trusted social relations help explainable recommendation? What are factors that could affect the explainable recommendations?

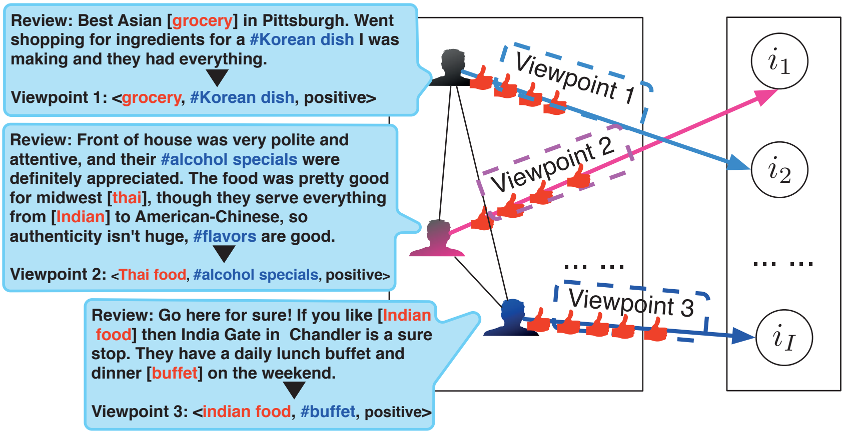


Figure 7.1: An example of trusted social relations, user reviews and ratings in a recommender system. Black arrows connect users with trusted social relations. "ThumpUp" logos reflect the ratings of items. Entities and topics have been highlighted into red and blue color, respectively. Three viewpoints are represented in three different colors.

To answer this research question, our focus is on developing methods to generate viewpoints by jointly analyzing user reviews and trusted social relations. We have already provided the definition of *viewpoint* in Chapter 5. Compared to "topics" in previous explainable recommendation strategies [32, 242, 249], viewpoints contain more useful information that can be used to understand and predict user ratings in recommendation tasks. We assume that each item and user in a recommender system can be represented as a finite mixture of viewpoints. And each user's viewpoints can be influenced by their trusted social friends. In Figure 7.1 we show an example with multiple viewpoints, user reviews, trusted social relations, and ratings in a recommender system.

Three technical issues need to be addressed before viewpoints can successfully be used for explainable recommendations that make use of social relations: (1) the shortness and sparseness of reviews make viewpoint extraction difficult; (2) because of the "bag of words" assumption, traditional topic models do not necessarily work very well in opinion analysis; (3) inferring explicit viewpoint statistics given trusted social relations among users and user reviews is not a solved problem.

In this chapter, we address these technical issues. We propose a latent variable model, called *social collaborative viewpoint regression model* (sCVR), to predict user ratings by discovering viewpoints. Unlike previous collaborative topic regression methods [242], sCVR predicts ratings by detecting viewpoints from user reviews and social relations. sCVR discovers entities, topics and sentiment priors from user reviews. sCVR employs Markov chains to capture the sentiment dependency between two adjacent words; given trusted social relations, in sCVR we assign a viewpoint-bias to each user by considering the social influence of their trusted social relations. Therefore, given a user and an item, sCVR detects viewpoints and predicts ratings by jointly generating entities, topics and sentiment labels in user reviews. Gibbs EM sampling is applied to approximate the posterior probability distributions. We use three real-world benchmark datasets in our experiments: *Yelp 2013*, *Yelp 2014*, and *Epinions*. Extensive experiments on these datasets show that sCVR outperforms state-of-the-art baselines in terms of MAE, RMSE, and NDCG metrics.

To sum up, our contributions in this chapter are as follows:

- To improve rating prediction for explainable recommendations, we focus on generating viewpoints from user reviews and trusted social relations.
- We propose a latent variable model, the social collaborative viewpoint regression model, to predict user ratings by jointly modeling entities, topics, sentiment labels and social relations.
- We prove the effectiveness of our proposed model on three benchmark datasets through extensive experiments, in which our proposed method outperforms state-of-the-art baselines.

We formulate our research problem in §7.1 and describe our approach in §7.2. Then, §7.3 details our experimental setup, §7.4 presents the experimental results, and §7.5 concludes the paper.

7.1 Preliminaries

Before introducing our social collaborative viewpoint regression model for explainable recommendations, we introduce our notation and key concepts. Table 7.1 lists the notation we use.

Similar to the Ratings Meet Reviews model (RMR) [137], we assume that there are U users $\mathcal{U} = \{u_1, u_2, \ldots, u_U\}$; I items $\mathcal{I} = \{i_1, i_2, \ldots, i_I\}$; a set of observed indices $\mathcal{Q} = \{(u, i)\}$, where each pair $(u, i) \in \mathcal{U} \times \mathcal{I}$ indicates an observed rating $r_{u,i}$ with a user review $d_{u,i}$ from user u to item i. For user reviews $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{Q}|}\}$, we assume that each observed rating $r_{u,i}$ is associated with a user review $d_{u,i}$. Given an item i's reviews $\mathcal{D}_i$, each review $d \in \mathcal{D}_i$ is represented as a set of words, i.e., $d = \{w_1, w_2, \ldots, w_{|d|}\}$. If two users u_i and u_j trust each other, as evidenced in a user communities, we define them to be a *trusted social relation* or simply *social relation* with trust value $\mathcal{T}_{u_i,u_j}$. We have already defined the notion of *topic* in Section 2.5, the notion of *sentiment* in Section 4.1 and the notions of *viewpoint* and *entity* in Section 5.1, respectively. In this chapter, we assume that K topics exist in the user reviews on which we focus, we set $z \in \{1, 2, \ldots, K\}$. We use the same assumption in Section 5.1 that the sentiment label l_j for a word w_j depends on the topic z_j. Specifically, we set $l_j = -1$ when the word w_j is "negative," while $l_j = 1$ when w_j is "positive."

Because user reviews are short, we assume that only one viewpoint v_d, represented as a combination of an entity e, a topic z and a sentiment label l, exists in each user review $d \in \mathcal{D}$. We assume that each item $i \in \mathcal{I}$ can be represented as a mixture over viewpoints, thus we set π_i to be a probability distribution of viewpoints in item i, μ to be a probability distribution of topics over viewpoints and λ to be a probability distribution of conceptual features over viewpoints. For words in user reviews, we set ϕ to be a probability distribution over viewpoints, topics and sentiment labels, which is derived from a Dirichlet distribution over hyper-parameter β.

It is common that rating scores are discrete [26, 249]. Unlike much previous work that predicts a decimal rating score given a user and an item, we apply a probabilistic

Table 7.1: Notation used in this chapter.

Symbol	Description
$\mathcal{I}$	candidate items
$\mathcal{U}$	candidate users
$\mathcal{D}$	user reviews
$\mathcal{N}$	vocabulary in review corpus $\mathcal{D}$
$\mathcal{T}$	trust values among users
$\mathcal{R}$	user ratings
$\mathcal{V}$	viewpoints set
$\mathcal{E}$	entities set
$\mathcal{Z}$	topics set in $\mathcal{Z}$
$\mathcal{Q}$	observed indices
u	a user, $u \in \mathcal{U}$
i	an item, $i \in \mathcal{I}$
d	a review, $d \in \mathcal{D}$
v_d	a viewpoint in review d, $v_d \in \mathcal{V}$
e_d	an entity in review d, $e_d \in \mathcal{E}$
w_j	the j-th word present in a review, $w_j \in \mathcal{N}$
z_j	a topic present in word w_j, $z_j \in \mathcal{Z}$
l_j	a sentiment label present in word w_j
f_u	a viewpoint selected by user u
$r_{u,i}$	the rating value from user u to item i
π	distribution of viewpoints
θ^u_v	distribution of viewpoint v for user u
λ	distribution of entities over viewpoints
μ	distribution of topics over viewpoints
$\phi_{v,z,l}$	distribution of words over v, z and l

rating distribution within the exponential family to provide more information to reflect users' rating habits, inspired by [26]. For each user $u \in \mathcal{U}$, we assume that u's ratings in a recommender system can be predicted by their viewpoint distribution over rating values, i.e., $\theta^u = \{\theta^u_{v_1}, \theta^u_{v_2}, \ldots, \theta^u_{v_V}\}$. Given a viewpoint $v \in \mathcal{V}$, $\theta^u_v \in \theta^u$ refers to a probabilistic distribution over each rating value $r \in [1, R]$, thus θ^u can be represented as an R-by-V matrix, shown as follows:

$$\theta^u = \begin{pmatrix} \theta^u_{1,v_1} & \cdots & \theta^u_{1,v_V} \\ \vdots & \ddots & \vdots \\ \theta^u_{R,v_1} & \cdots & \theta^u_{R,v_V} \end{pmatrix} \tag{7.1}$$

where each item $\theta^u_{r,v}$ denotes the probability of rating value r given user u and viewpoint v.

We assume that the viewpoint distribution θ^u_v is derived by a finite mixture over a personalized base distribution $\theta^0_{u,v}$ and viewpoint distributions of u's trusted relations. Given a user u and an item i, we set a multinomial distribution $f_{u,i}$, which derives from the viewpoints distribution π_i for item i, to reflect the viewpoint chosen by u for their rating to item i. If a user u writes a user review $d_{u,i}$ for item i, there is a corresponding rating $r_{u,i} \in [1, R]$ derived from a multinomial distribution over $\theta^u_{f_{u,i}}$.

Given observed indices $\mathcal{Q}$, observed data $\mathcal{R}$, $\mathcal{D}$ and $\mathcal{E}$, our target is to infer the user's viewpoint distribution θ and the item's viewpoint distribution π, which are applied to predict unknown ratings. Represented by tuples of a conceptual feature, a topic and a sentiment label, viewpoints are used to explain our results.

7.2 Method

In this section, we propose our *social collaborative viewpoint regression model*, abbreviated as **sCVR**. We start by detailing the model. We then describe our inference approach and explain our method to predict ratings using posterior distributions from sCVR.

7.2.1 Feature detection and sentiment analysis

We use descriptive keywords in an e-commerce platforms as entities for items. Here we assume that E_i many features exist in an item i's reviews. To discover the entity in a user review $d \in \mathcal{D}_i$, we employ word2vec [161] to calculate the similarity between a given entity $e \in \mathcal{E}_i$ and a user review d. Since the quality of the word vectors increases significantly with the amount of training data, we train a word2vec model using the latest Wikipedia data. Thereafter, we employ our trained model to predict the cosine similarity between a given entity e and each word w in a user review d. Given the cosine similarity $sim(e, w)$ between e and word w, $w \in d$, we calculate the similarity between e and review d following Eq. 7.2:

$$sim(e, d) = \frac{1}{N_d} \sum_{w \in d} sim(e, w) \tag{7.2}$$

where N_d indicates the number of words in d. Given candidate entities $\mathcal{E}_i$, the entity that is most similar to d will be considered as d's relevant entity. By ranking documents according to the similarity between candidate entities and user reviews, we find the relevant entity for each user review.

We employ a state-of-the-art sentiment analysis method [219] to classify user reviews into positive and negative categories. The probability of a sentiment label is set as a prior value in our social collaborative viewpoint regression, which is detailed in §7.2.2.

7.2.2 Social collaborative viewpoint regression

Given observed indices $\mathcal{Q}$, users $\mathcal{U} = \{u_1, u_2, \ldots, u_U\}$, items $\mathcal{I} = \{i_1, i_2, \ldots, i_I\}$, ratings $\mathcal{R} = \{r_1, r_2, \ldots, r_Q\}$ and user reviews $\mathcal{D} = \{d_1, d_2, \ldots, d_Q\}$, our target is to infer distributions of viewpoints to predict unknown user ratings $\mathcal{Q}' = \{(u', i')\}$ from users to items, where $(u', i') \notin \mathcal{Q}$. We propose a latent factor model, *social collaborative viewpoint regression* (sCVR), to tackle this problem. Unlike previous work, sCVR jointly models viewpoints, topics, entities and sentiment labels in $\mathcal{D}$; in addition, sCVR explicitly models influences from a user's social relations on their own viewpoint distribution.

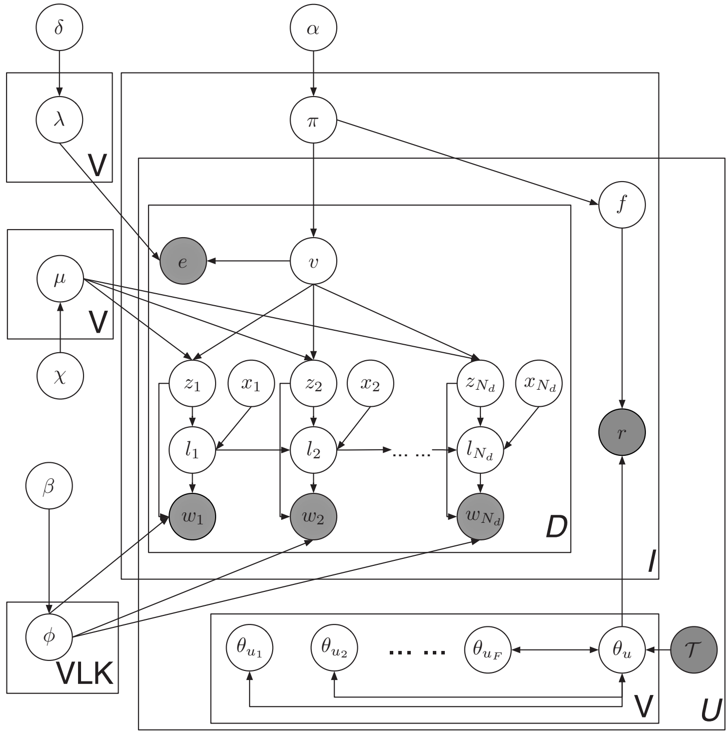


Figure 7.2: Graphical representation of social collaborative viewpoint modeling, sCVR.

Figure 7.2 shows a graphical representation of sCVR, in which we see a number of ingredients. Shaded circles indicate observed variables, whereas unshaded ones are latent variables. Unshaded rectangles are stochastic processes. Capital characters refer to the number of variables, and we use VLK to represent the product of three values V, L and K. Similar to other latent factor models [32], directed arrows show dependency relations between two random variables: for instance, the variables v depend on π; the variables π depend on α; observed variables w depend on the variables z, l, v and ϕ, whereas variables e and z depend on v.

After preprocessing, for each user review $d \in \mathcal{D}$ we assume that there is an entity $e_d \in \mathcal{E}$, and for each word w in d there is a corresponding sentiment label l_w. We assume that there are, in total, V viewpoints and K topics in user reviews. Given an item $i \in \mathcal{I}$, we assume there is a probabilistic distribution π over viewpoints. Given a user review $d \in \mathcal{D}$, for each word $w_j \in d$, there is a topic z_j and a sentiment label l_j. We assume that a viewpoint v in d is derived via a multinomial distribution over a random variable π that indicates a probability distribution over viewpoints in each item; given viewpoint v, an entity e, a topic z and a sentiment label l are derived from probabilistic distributions over v. The probability distribution π is derived from a dirichlet mixture over a hyper parameter α.

Each user $u \in \mathcal{U}$ in sCVR is supposed to have F_u trusted social relations; each trusted

- For each viewpoint $v \in \mathcal{V}$:
  - Draw $\mu_v \sim Dir(\chi)$; $\lambda_v \sim Dir(\delta)$;
  - For each topic z:
    - Draw $\rho_{v,z} \sim Beta(\eta)$;
    - For each sentiment l:
      - Draw $\phi_{z,l,v} \sim Dir(\beta)$;
- For each user $u \in \mathcal{U}$:
  - Draw $\theta_v^u \sim Dir(\theta_{u,v}^0 + \frac{1}{F_u} \sum_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \theta_v^{u'})$;
- For each item $i \in \mathcal{I}$:
  - Draw $\pi_v \sim Dir(\alpha)$;
  - For each user review $d \in \mathcal{D}_{u,i}$ from user u:
    - Draw a viewpoint $v \sim Multi(\pi)$;
    - Draw an entity $e_d \sim Multi(\lambda_v)$;
    - Draw $\sigma \sim Dir(\tau)$;
    - For each word w_j in document d:
      - Draw a topic $z_j \sim Multi(\mu_v)$;
      - Draw $x_j \sim Multi(\sigma)$;
      - If $x_j = 1$, draw $l_j \sim l_{j-1}$
      - If $x_j = -1$, draw $l_j \sim (-1) \cdot l_{j-1}$;
      - If $x_j = 0$, draw $l_j \sim Bern(\rho_{v,z_j})$;
      - Draw word $w_j \sim Multi(\phi_{v,z_j,l_j})$:
  - For each ratings assigned by user u to i:
    - Draw viewpoint $f_{u,i} \sim Multi(\pi)$;
    - Draw rating $r_{u,i} \sim Multi(\theta_{f_{u,i}}^u)$;

Figure 7.3: Generative process in sCVR.

relation u' shares a trust value $\mathcal{T}_{u,u'}$ with user u. For each user $u \in \mathcal{U}$, a probabilistic distribution over viewpoint v, θ_v^u is derived over viewpoint distributions of u's social relations and a base distribution of u, i.e., $\{\theta_v^{u_1}, \theta_v^{u_2}, \ldots, \theta_v^{u_{F_u}}\}$ and $\theta_{u,v}^0$. In sCVR we assume that u's rating $r_{u,i}$ for an item $i \in \mathcal{I}$ is derived from a multinomial distribution over θ_f^u, where f is a sampling viewpoint index derived from u's reviews, i.e., $f \in [1, V]$. In sCVR we consider the sentiment dependency between two adjacent words, as same as the *viewpoint tweets topic model* (See §5.2). The generative process of sCVR is shown in Figure 7.3.

7.2.3 Inference

Similar to previous work [137], because of the unknown relation among random variables, exact posterior inference for sCVR model is intractable. Sampling-based methods for traditional topic models rarely include methods for optimizing hyper parameters. In the sCVR model, since θ_v^u, $\theta_{u,v}^0$, ϕ, π, μ and λ indicate the results for computations, we need to find an optimized process for parameters θ_v^u, $\theta_{u,v}^0$, ϕ, π, μ and λ during our posterior inference. Therefore, unlike much previous work on topic models, to infer weighted priors we apply a Gibbs EM sampler [239] to conditionally approximate the posterior distribution of random variables in sCVR.

We divide our algorithm into two parts: an **E**-step and **M**-step. Given item i and user u, for each user review d the target of our sampling in the **E**-step is to approximate the posterior distribution $p(\mathcal{V}, \mathcal{Z}, \mathcal{L} \mid \mathcal{W}, \mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{F})$. Conceptually, in this step we divide our sampling procedure into three parts. Firstly, given a user u and an item i, during the **E**-step, we sample the conditional probability of viewpoint $f_{u,i}$ given current state of viewpoints, i.e., $P(f_{(u,i)} \mid \mathbf{f}_{-(u,i)}, \mathcal{W}, \mathcal{V}, \mathcal{R})$. Secondly, given the values of inferred topics and sentiment labels, we sample the conditional probability of viewpoint v in each $d \in \mathcal{D}$, i.e., $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}, \mathcal{R})$. Lastly, given the current state of viewpoints, for word w_j we sample the conditional probability of topic z_j with sentiment label l_j transition label x_j, i.e., $P(z_j = k, l_j = l, x_j = x \mid \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{F}, v)$. During the **M**-step, given conditional probabilities derived during the **E**-step, we maximize each user u's viewpoint distribution θ_u, each viewpoint distribution π and the joint probability of viewpoints, entities, topics, and sentiments over words, i.e., ϕ.

We now detail our sampling procedures. Given user u and item i, we first sample $f_{u,i}$ over $\mathbf{f}_{-(u,i)}$ without pair (u, i). So for user u's viewpoint over item i, we obtain $P(f_{(u,i)} \mid \mathbf{f}_{-(u,i)}, \mathcal{W}, \mathcal{V}, \mathcal{R})$ as:

$$P(f_{(u,i)} = y \mid \mathbf{f}_{-(u,i)}, \mathcal{W}, \mathcal{V}, \mathcal{R}) \propto \frac{n_{u,-i}^{r_{(u,i)},y} + \theta^u_{r_{(u,i)},y}}{n_u^y + R_u \cdot \theta^u_{r_{(u,i)},y}} \cdot \frac{n^{i,y}_{f,-(u,i)} + n_v^{i,y} + \alpha}{n^i_{f,-(u,i)} + n^i_v + V\alpha}, \quad (7.3)$$

where R_u indicates how many times user u rates items, and $n^{i,y}_{f,-(u,i)}$ indicates the number of times that variable f has been assigned to y given item i, excluding user u; furthermore, $n_v^{i,y}$ indicates the number of times that viewpoint v in item i has been assigned to y. And $n_{u,-i}^{r_{(u,i)},y}$ indicates the number of times that user u gives rating $r_{(u,i)}$ under $f = y$ for all items, excluding i. We calculate $\theta^u_{r_{(u,i)},y}$ according to Eq. 7.4:

$$\theta^u_{r_{(u,i)},y} = \theta^0_{u,y,r_{(u,i)}} + \frac{1}{F_u} \sum_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \cdot \theta^{u'}_{r_{(u,i)},y}, \quad (7.4)$$

where $\mathcal{T}_{u,u'}$ indicates the trust value between user u and u', $\mathcal{F}_u$ indicates the trusted social relations of user u. For review d written by user u for item i, we infer the conditional probability of viewpoint $v_d = v$ given all other random variables, i.e., $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}, \mathcal{R})$. So we have:

$$\begin{aligned} P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}, \mathcal{R}) \propto \frac{n^{i,v}_{-d} + n^{i,v}_f + \alpha}{n^i_{-d} + n^i_f + V\alpha} \cdot \\ \prod_{e \in \mathcal{E}} \frac{n^{-d}_{v,e} + \delta}{n_v^{-d} + E\delta} \cdot \prod_{z \in \mathcal{Z}} \frac{n^{-d}_{v,z} + \chi}{n_v^{-d} + K\chi} \cdot \prod_{l \in \mathcal{L}} \prod_{w \in \mathcal{N}_d} \frac{n^{w,-d}_{z,l,v} + \beta}{n^{-d}_{z,l,v} + N\beta}, \end{aligned} \quad (7.5)$$

where $n^{i,v}_{-d}$ indicates the number of times that viewpoint v has been assigned to user reviews, excluding d; $n^{-d}_{v,e}$ indicates the number of times that entity e has been assigned to viewpoint v in reviews, excluding d; $n^{-d}_{v,z}$ indicates the number of times that topic z

has been assigned to viewpoint v excluding d; furthermore, $n^{w,-d}_{z,l,v}$ indicates how many words are assigned to topic z, viewpoint v and sentiment l, except for d. Given detected viewpoint $v_d = v$, for each word $w_j \in \mathcal{N}_d$ we sample the conditional probability of topic z_j with sentiment label l_j for word w_j, i.e., $P(z_j = k, l_j = l, x_j = x \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F})$. Given the viewpoint v sampled at the document level, when $x_j \neq 0$ and $x_{j+1} \neq 0$ we can directly sample word w_j's topic z_j and sentiment label l_j using the probability in Eq 7.6:

$$\begin{aligned} &P(z_j = k, l_j = l, x_j = x \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F}) \propto \frac{n^{-j}_{v,k} + \chi}{n^{-j}_{v} + K\chi} \\ &\cdot \frac{n^{w_j,-j}_{k,l,v} + \beta}{n^{-j}_{k,l,v} + N\beta} \cdot \frac{n^{w_j}_{-j,x} + \tau_x}{n^{w_j}_{-j} + \sum\limits_{x \in \mathcal{X}} \tau_x} \cdot \frac{n^{w_{j+1}}_{-(j+1),x_{j+1}} + I(x_{j+1} = x_j) + \tau_{x_{j+1}}}{n^{w_{j+1}}_{-(j+1)} + 1 + \sum\limits_{x \in \mathcal{X}} \tau_x}, \end{aligned} \tag{7.6}$$

where $n^{-j}_{v,k}$ indicates the number of times that topic k has been assigned to viewpoint v, excluding the jth word in d; n^{-j}_{v} indicates how many topics have been assigned to v, not including w_j; $n^{w_j,-j}_{,k,l,v}$ indicates the number of times that word w_j has been assigned to topic z and sentiment l synchronously, excluding current one; $n^{w_j}_{-j,x}$ indicates the number of times that w_j assigned to x, excluding current word; and $I(x_{i+1} = x_i)$ get value 1 if $x_{i+1} = x_i$, otherwise it gets 0. When $x_j = 0$, w_j's sentiment label l_j is derived from a Bernoulli distribution ρ_{v,z_j}; then the conditional probability $P(z_j = k, l_j = l, x_j = 0 \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F})$ becomes:

$$\begin{aligned} &P(z_j = k, l_j = l, x_j = 0 \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F}) \propto \\ &\qquad \frac{n^{-j}_{v,k} + \chi}{n^{-j}_{v} + K\chi} \cdot \frac{n^{w_j,-j}_{,k,l,v} + \beta}{n^{-j}_{k,l,v} + N\beta} \cdot \frac{n^{w_j}_{-j,x} + \tau_x}{n^{w_j}_{-j} + \sum\limits_{x \in \mathcal{X}} \tau_x} \cdot \frac{n^{-j}_{z,l,v} + \eta_l}{n^{-j}_{z,v} + \sum\limits_{l \in \mathcal{L}} \eta_l}, \end{aligned} \tag{7.7}$$

where $n^{-j}_{z,l,v}$ indicates how many words are assigned to viewpoint v, topic z and sentiment label l, excluding current w_j; whereas $n^{-j}_{v,z}$ indicates how many words are assigned to viewpoint v and topic z, excluding current w_j.

In the **M**-step, given conditional probabilities derived in the **E**-step, we estimate the parameters of user u's viewpoint distribution θ_u for each rating r, the viewpoint distribution π_i for each item i, the probability of topics, viewpoints and sentiment over words ϕ, viewpoint distributions over entities λ and viewpoint distributions over topics μ as follows:

$$\begin{aligned} \theta^u_{r,v} &= \frac{n^{r,v}_u + \theta^0_{u,v,r} + \frac{1}{F_u} \sum\limits_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \theta^{u'}_{r,v}}{n_{u,v} + R_u \cdot \left(\theta^0_{u,v,r} + \frac{1}{F_u} \sum\limits_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \theta^{u'}_{r,v} \right)} \\ \pi_{i,v} &= \frac{n_{i,v} + \alpha}{n_i + V\alpha}; \qquad \phi^w_{v,z,l} = \frac{n^w_{v,z,l} + \beta}{n_{v,z,l} + N\beta} \\ \mu_{v,e} &= \frac{n_{v,z} + \chi}{n_v + K\chi}; \qquad \lambda_{v,e} = \frac{n_{v,e} + \delta}{n_v + E\delta}. \end{aligned} \tag{7.8}$$

Algorithm 9: Gibbs EM sampling for sCVR's inference

```
Input: α, β, η, τ, U, I, R, W
Output: θ, φ, μ, λ and π
ite = 0;
if ite¡T then
    E-Step:
    for u = 1 to U do
        for i = 1 to I do
            Draw f_{u,i} = y from Eq. 7.3
            Update n_f^{i,y}, n_v^{i,y} and n_u^{r(u,i),y}
            Draw v_d = v from Eq. 7.5
            Update n^{i,v}, n_{v,e}, n_{v,z} and n^w_{z,l,v} for w ∈ d
            for j = 1 to N_d do
                Draw ⟨z_j, l_j, x_j⟩ from Eq. 7.6
                if x_j ≠ 0 then
                    Update n_{v,z_j}, n^{w_j}_{z_j,l_j,v} and n^{w_j}_{x_j}
                end
                if x_j = 0 then
                    Update n_{v,z_j}, n^{w_j}_{k_j,l_j,v}, n^{w_j}_{x_j} and n_{z_j,l_j,v}
                end
            end
        end
    end
    M-Step:
    Re-estimate θ_u, π, φ, μ and λ from Eq. 7.8;
    Maximize θ̂^0_{u,v} from Eq. 7.9;
    ite = ite + 1 and go to E-Step;
end
```

Given posterior viewpoint distributions, we optimize the value of random variables θ_u^0 for each user u. Using two bounds defined in [162], we derive the following update rule for obtaining each user u's optimized viewpoint distribution in Eq. 7.8 via fixed-point iterations:

$$\hat{\theta}^0_{u,v} \leftarrow \theta^0_{u,v} \cdot \frac{\sum\limits_{v \in V} \Psi(n^u_{r,v} + \theta^u_{r,v}) - \Psi(\theta^u_{r,v})}{\sum\limits_{v \in V} \Psi(n^u_v + R_u \cdot \theta^u_{r,v}) - \Psi(R_u \cdot \theta^u_{r,v})}, \tag{7.9}$$

where $\Psi(x)$ is a digamma function defined by $\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$, and $\theta^u_{r,v}$ is defined in Eq. 7.4. Algorithm 9 summarizes the Gibbs EM sampling inference procedure based on the equations that we have just derived.

7.2.4 Prediction

After Gibbs EM sampling, for each user $u \in \mathcal{U}$, we have a matrix θ_u to describe the conditional probability of ratings given u's viewpoints, i.e., $P(r \mid v, u) = \theta^u_{r,v}$ over ratings. For each item $i \in \mathcal{I}$, we have a viewpoint distribution π_i, i.e., $P(v \mid i) = \pi_{v,i}$. Therefore, given user $u \in \mathcal{U}$ and item $i \in \mathcal{I}$, in order to predict an unknown rating between u and i, we calculate the probability of the rating $r_{u,i} = r$ by Eq. 7.10.

$$P(r_{u,i} = r \mid u, i) = \sum_{v \in \mathcal{V}} \theta^u_{r,v} \cdot \pi_{i,v}. \tag{7.10}$$

By ranking $P(r_{u,i} = r \mid u, i)$ for each candidate rating r, we choose the rating r with the highest probability as the predicted rating for u and i.

7.3 Experimental Setup

7.3.1 Research questions

We divide our main question **RQ5** into the following research questions **RQ5.1**–**RQ5.4** that guide the remainder of the chapter.

- **RQ5.1**: What is the performance of sCVR in rating prediction and top-k item recommendation tasks? Does it outperform state-of-the-art baselines? (See §7.4.1.)
- **RQ5.2**: What is the effect of the number of viewpoints? What is the effect of the number of topics? (See §7.4.2)
- **RQ5.3**: What is the effect of trusted social relations in collaborative filtering? Do they help to enhance the recommendation performance? (See §7.4.3)
- **RQ5.4**: Can sCVR generate explainable recommendation results? (See §7.4.4)

7.3.2 Datasets

We use three benchmark datasets in our experiments: the *Yelp dataset challenge 2013*, *Yelp dataset challenge 2014*[1] and *Epinions.com* dataset.[2] Each dataset has previously been used in research on recommendation [43, 137, 225]. In total, there are over 400,000 users, 80,000 items, 4,000,000 trusted social relations and 2,000,000 user reviews in our datasets. We show the statistics about our datasets in Table 7.2.

Table 7.2: Overview of the three datasets used in the paper.

	Yelp 2013	**Yelp 2014**	**Epinions**
items	15,584	61,184	26,850
reviews	335,021	1,569,264	77,267
users	70,816	366,715	3,474
relations	622,873	2,949,285	37,587

[1] http://www.yelp.com/dataset_challenge
[2] http://epinions.com

Yelp[3] provides a business reviewing platform. Users are able to create a profile that they can use to rate and comment on services provided by local businesses. This service also provides users with the ability to incorporate a social aspect to their profiles by adding people as friends. Our first two datasets ("Yelp challenge 2013" and "Yelp challenge 2014" in Table 7.2) consist of data from the Yelp dataset challenge 2013 and 2014, respectively. The Yelp dataset challenge 2013 contains $15,584$ items, $70,816$ users and $335,021$ user reviews. Between the users, there are $622,873$ social relations. For the Yelp dataset challenge 2014, we find $366,715$ users, $61,184$ items, $1,569,264$ reviews and $2,949,285$ edges in the dataset. The two datasets are quite sparse, which may negatively most collaborative filtering methods based on ratings.

Epinions.com is a consumer opinion website on which people can share their reviews of products. Members of Epinions can review items, e.g., food, books, and electronics, and assign numeric ratings from 1 to 5. Epinions members can identify their own Web of Trust, a group of "reviewers whose reviews and ratings they have consistently found to be valuable." Released by [43], this dataset includes $3,474$ users with $77,267$ reviews for $26,850$ items; there are $37,587$ social edges in this dataset.

7.3.3 Evaluation metrics

We employ three offline evaluation metrics in our experiments: Mean Absolute Error (*MAE*), Root Mean Square Error (*RMSE*) and Normalized Discounted Cumulative Gain (*NDCG*).

Root Mean Squared Error (*RMSE*) and Mean Absolute Error (*MAE*) are two widely used evaluation metrics for rating prediction in recommender systems. Given a predicted rating $\hat{r}_{u,i}$ and a ground-truth rating $r_{u,i}$ from user u to item i, the RMSE is calculated as in Eq. 7.11:

$$RMSE = \sqrt{\frac{1}{R}\sum_{u,i}\left(r_{u,i} - \hat{r}_{u,i}\right)^2}, \tag{7.11}$$

where R indicates the number of ratings between users and items. Similarly, MAE is calculated as follows:

$$MAE = \sqrt{\frac{1}{R}\sum_{u,i}|r_{u,i} - \hat{r}_{u,i}|}. \tag{7.12}$$

These two criteria measure the error between the true ratings and the predicted ratings.

To assess whether sCVR can improve the ranking of item rankings, we use the Normalized Discounted Cumulative Gain (*NDCG*) as our third evaluation metric. NDCG is evaluated over a number of the top items in the ranked item list. Let U be the set of users and r_u^p be the rating score assigned by user u to the item at the pth position of the ranked list. The NDCG value at the n-th position with respect to user u is defined in Eq. 7.13:

$$NDCG_u@n = Z_u \sum_{p=1}^{n} \frac{2^{r_u^p} - 1}{\log(1+p)}, \tag{7.13}$$

[3] `http://www.yelp.com`

Table 7.3: Baselines and methods used for comparison.

Acronym	**Gloss**	**Reference**
CVR	Collaborative viewpoint regression	§7.2
sCVR	Social collaborative viewpoint regression	§7.2
Collaborative filtering methods		
CliMF	Maximize reciprocal rank method for item ranking	[213]
LRMF	List-wise learning to rank method for item ranking	[212]
NMF	Non-negative matrix factorization	[121]
PMF	Probabilistic matrix factorization	[163]
SoMF	Trust propagation matrix factorization	[100]
TrMF	Trust social matrix factorization	[254]
Explainable recommendation methods		
CTR	Collaborative topic regression model	[242]
EFM	Explicit factor model for item recommendation	[271]
HFT	Hidden factors as topics model	[154]
RMR	Ratings meet reviews model	[137]
SCTR	Social-aware collaborative topic regression	[43]

where Z_u is a normalization factor calculated so that the NDCG value of the optimal ranking is 1. NDCG@n takes the mean of the $NDCG_u@n$ of all users, which is computed as follows:

$$NDCG@n = \frac{1}{U} \sum_{u \in \mathcal{U}} NDCG_u@n. \tag{7.14}$$

We apply NDCG@5 and NDCG@10 in our experiments.

Statistical significance of observed differences between the performance of two runs is tested using a two-tailed paired t-test and is denoted using $^\blacktriangle$ (or $^\blacktriangledown$) for strong significance for $\alpha = 0.01$; or $^\vartriangle$ (or $^\triangledown$) for weak significance for $\alpha = 0.05$.

7.3.4 Baselines and comparisons

We list the methods and baselines that we consider in Table 7.3. In this chapter, we propose the social collaborative viewpoint regression model (sCVR); we write sCVR for the overall process as described in Section 7.2, which includes both the viewpoint modeling and social relation modeling. We write CVR for the model that only considers viewpoint modeling in §7.2. Our baselines include recent work on both collaborative filtering and explainable recommendation methods. To evaluate the performance of viewpoint modeling methods in explainable recommendation, we use previous work on explainable recommendation: the hidden factors topic model (HFT) [154], the collaborative topic regression (CTR) [242], and the ratings meet reviews model (RMR) [137] as our baselines. Using a sentiment lexicon analysis tool [271], we use EFM [271] as a baseline in our experiments for explainable recommendation. To evaluate the effect of social communities in explainable recommendation, we use social-aware collaborative topic regression

(SCTR) [43] as another baseline. We also compare sCVR with recent collaborative filtering methods: we use probabilistic matrix factorization (PMF) [163], non-negative matrix factorization (NMF) [121], list-rank matrix factorization (LRMF) [212] and collaborative less-is-more filtering (CliMF) [213] as baselines for collaborative filtering. To compare sCVR with collaborative filtering using trusted social relations, we use trust matrix factorization (TrMF) [254] and social matrix factorization (SoMF) [100] as another two baselines in our experiments.

7.4 Results and Discussion

In §7.4.1, we compare sCVR to other baselines for rating prediction and item recommendation; in §7.4.2 we examine the performance of sCVR for varying numbers of viewpoints and topics; §7.4.3 examines the effect of social relations in sCVR; we also discuss the *explainability* of rating predictions in §7.4.4.

7.4.1 Overall performance

To start, for research question **RQ5.1**, to evaluate the effectiveness of sCVR in personalized recommendation, we examine the performance of sCVR in rating prediction and item recommendation tasks. For the rating prediction task, Table 7.4 lists the performance of all methods in terms of MAE and RMSE. Because our baselines predict decimal rating values based on a Gaussian noise distribution, following Beutel et al. [26], we calculate the predictive probability, i.e., $P(r \mid \hat{r})$, for each predicted rating $\hat{r}$, and we use the discrete rating with highest predictive probability in our experiments. For all three datasets, sCVR outperforms other baselines, and significantly outperforms SCTR on the Yelp 2013 and 2014 datasets. PMF performs worst. The list-wise learning to rank methods (LRMF and CliMF) do not perform well in rating prediction, whereas methods considering social relations outperform other methods. To understand the benefits of viewpoint modeling (and in particular, the addition of entities and sentiment), we compare sCVR with SCTR, which ignores entities and sentiment during topic modeling. On the Yelp 2013 dataset, sCVR achieves a 16.7% and 8.2% decrease over SCTR in terms of MAE and RMSE, respectively, whereas on the Yelp 2014 dataset, it achieves decreases of 11.1% and 5.2%, respectively.

Next, we evaluate the performance of sCVR on the item recommendation task, even though this is not the main purpose for which it was designed. Table 7.5 lists the performance of all methods in terms of NDCG@5 and NDCG@10. Interestingly, we find that sCVR tends to outperform the other baselines: for both the Yelp 2013 and Epinions datasets sCVR provides the best performance, while for the Yelp 2014 dataset sCVR performs almost as good as CliMF, which is a state-of-the-art ranking method for the item recommendation task. For the Yelp 2013 dataset, sCVR achieves a 15.7% increase over NMF in terms of NDCG@5, and a 16.0% increase in terms of NDCG@10. For the Epinions dataset, sCVR achieves a 15.1% increase over NMF in terms of NDCG@5, and a 8.1% increase in terms of NDCG@10. Furthermore, it significantly outperforms NMF on both the Yelp 2013 and Epinions datasets. This shows that, when compared against state-of-the-art baselines in terms of the NDCG metric, sCVR is very competitive.

Table 7.4: **RQ5.1 and RQ5.3: MAE** and **RMSE** values for rating prediction. Significant differences are with respect to SCTR (row with shaded background).

	Yelp 2013		Yelp 2014		Epinions	
	MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
Collaborative filtering						
CliMF	1.109	1.524	1.591	1.912	0.493	0.582
LRMF	1.653	1.944	1.897	2.042	0.517	0.626
NMF	1.130	1.591	1.284	1.763	0.595	0.691
PMF	1.427	1.853	1.424	1.902	0.526	0.688
SoMF	0.912	1.375	0.924	1.402	0.554	0.673
TrMF	1.109	1.524	1.134	1.564	0.542	0.667
Explainable recommendations						
CTR	0.915	1.169	0.971	1.294	0.525	0.612
EFM	0.912	1.182	1.124	1.452	0.532	0.644
HFT	0.844	1.072	1.094	1.336	0.517	0.604
LDA	1.232	1.622	1.294	1.677	0.526	0.612
RMR	0.812	1.013	0.937	1.283	0.514	0.602
SCTR	0.894	1.065	0.907	1.262	0.472	0.584
sCVR	**0.744**$^{\blacktriangle}$	**0.977**$^{\blacktriangle}$	**0.806**$^{\blacktriangle}$	**1.196**$^{\blacktriangle}$	**0.482**	**0.579**

7.4.2 Number of viewpoints and topics

Next we turn to **RQ5.2**. Under the default value of the number of topics $Z = 20$ in sCVR, in Figure 7.4(a) we examine the RMSE performance of sCVR with varying numbers of viewpoints. We find that the performance of sCVR in terms of RMSE hits a minimum when the number of *viewpoints* equals 70 for the Yelp 2013 dataset; with fewer than 70, performance decreases but when the number exceeds 70, due to the redundancy of viewpoints in rating prediction, performance increases. Similar phenomena can be found for the Yelp 2014 dataset and the Epinions dataset. For Yelp 2014, sCVR achieves its best RMSE performance when the number of viewpoints equals 80, whereas for the Epinions dataset, it achieves its best RMSE performance when we set V to 40.

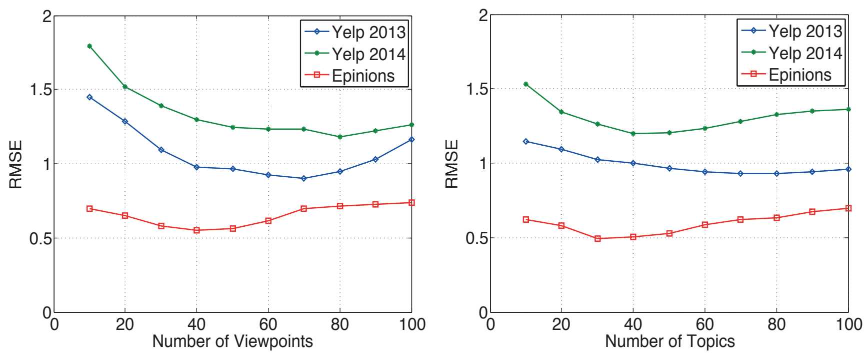
Under the default value of the number of viewpoints $V = 30$, we evaluate the RMSE performance of sCVR with varying numbers of topics in Figure 7.4(b). We find that for the Yelp 2013 dataset, sCVR achieves its best RMSE performance when $Z = 80$, whereas for the Yelp 2014 dataset this value is 40. For the Epinions dataset, sCVR performs best when $Z = 30$.

7.4.3 Effect of social relations

Turning to **RQ5.3**, to determine the contribution of social relations in the rating prediction task, we turn to Table 7.6, where columns 2–3 and 4–5 show the performance of CVR and sCVR, respectively, in terms of MAE and RMSE. Recall that CVR only detects viewpoints without considering social relations. We find that sCVR, which does

Table 7.5: **RQ5.1: NDCG@5** and **NDCG@10** values for item recommendation. Significant differences are with respect to NMF (row with shaded background). N@5 abbreviates NDCG@5, N@10 abbreviates NDCG@10.

	Yelp 2013		Yelp 2014		Epinions	
	N@5	**N@10**	**N@5**	**N@10**	**N@5**	**N@10**
Collaborative filtering						
CliMF	0.741	0.803	**0.482**	**0.562**	0.897	0.921
LRMF	0.712	0.725	0.425	0.491	0.844	0.902
NMF	0.642	0.693	0.472	0.529	0.784	0.853
Explainable recommendations						
EFM	0.722	0.783	0.479	0.532	0.890	0.914
sCVR	**0.743**▲	**0.804**▲	**0.482**	0.544	**0.902**▲	**0.922**▲



(a) RMSE performance with different numbers of viewpoints

(b) RMSE performance on different numbers of topics.

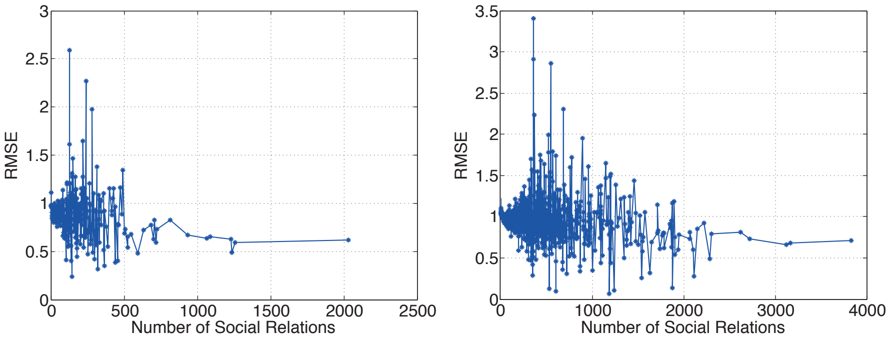
Figure 7.4: **RQ5.2:** RMSE performance with different numbers of viewpoints and topics.

consider social relations, outperforms CVR significantly on all three datasets. From Table 7.4, we also see that methods considering social relations perform quite well in terms of MAE and RMSE. For the Yelp 2013 dataset, sCVR achieves a 6.7% decrease over CVR in terms of RMSE. For the Yelp 2014 dataset, sCVR achieves a 7.4% decrease over CVR in terms of RMSE. In terms of RMSE, on the Epinions dataset, sCVR achieves a significant decrease over CVR of 18.7%. Thus, we conclude that social communities can successfully be applied to enhance the performance of rating prediction.

To evaluate the effect of the number of social relations, Figure 7.5 shows the average RMSE performance for users with different numbers of social relations in the Yelp 2013 and Yelp 2014 datasets. In Figure 7.5 we observe that for both Yelp 2013 and Yelp 2014 datasets, RMSE performance shows a "wave-like" decrease as the number of social relations increases. Thus, we conclude that users with more social relations, in most cases, will get better prediction results using sCVR.

Table 7.6: **RQ5.3:** Effect of social communities in rating prediction in our three datasets.

	CVR		sCVR	
Dataset	MAE	RMSE	MAE	RMSE
Yelp 2013	0.862	1.049	**0.744**$^{\triangle}$	**0.977**$^{\triangle}$
Yelp 2014	0.953	1.291	**0.806**$^{\blacktriangle}$	**1.196**$^{\blacktriangle}$
Epinions	0.641	0.712	**0.482**$^{\blacktriangle}$	**0.579**$^{\blacktriangle}$



(a) Yelp 2013 (b) Yelp 2014

Figure 7.5: **RQ5.3:** RMSE performance with different numbers of social relations on the Yelp datasets.

7.4.4 Explainability

Finally, we address **RQ5.4**. Apart from being more accurate at rating prediction, another advantage of sCVR over collaborative filtering methods is that it provides explainable recommendation results. To illustrate the explainability of outcomes of sCVR, Table 7.7 shows 4 examples of our detected viewpoints. In the example viewpoints in Table 7.7, we see entities with relevant topics and corresponding sentiment labels. For each viewpoint, we find that relevant topics in the second column help to interpret the entity in the first column, and sentiment labels inform users on opinions in the viewpoint. In sum, as we have shown in our experimental results, viewpoints-as-explanations are useful to enhance the accuracy in rating prediction, especially for the "cold-start" problem, e.g., if a user expresses a positive review on "Chinese" cuisine, sCVR would recommend a business that is salient for the same viewpoint. And because of the explainability of sCVR, we also get a better understanding of items and users' preferences by analyzing the viewpoints.

7.5 Conclusion and Future Work

We have considered the task of explainable recommendations. To improve the rating prediction for explainable recommendations, we have identified two main problems: opinions in users' short comments, and complex trusted social relations. We have tackled

Table 7.7: **RQ5.4:** Example viewpoints produced by sCVR in Yelp 2013. Column 1 lists the entities corresponding to the viewpoints; Column 2 list the topics in viewpoints, Columns 3, 4 and 5 list the probabilities of positive and negative labels for each topic, respectively.

Entity	Topic	Positive	Negative
Italian	#topic 2: italian, pizza, well, pasta, menu, wine, favorite, eggplant, dinner, special	0.518	0.482
Fast food	#topic 12: burger, pizza, cheap, bad, drink, sausage, egg, lunch, garden, price	0.224	0.776
Steakhouses	#topic 7: potato, appetizer, good, place, pork, rib, bread, rib-eye, filet, beef	0.797	0.203
Indian	#topic 10: vegetarian, masala, curry, pretty, buffet, busy, delicious, rice, lamb, expect	0.619	0.381
Chinese	#topic 14: dim-sum, chicken, duck, enjoy, spicy, soup, dumpling, worth, flavor, tea	0.652	0.348

these problems by proposing a novel latent variable model, called the social collaborative viewpoint regression model, which detects viewpoints and uses social relations. Our model is divided into two parts: viewpoint detection and rating prediction. Based on the probabilistic distribution of viewpoints, we predict users' ratings of items. Our experiments have provided answers to the main research question raised at the beginning of this chapter:

RQ5: Can we devise an approach to enhance the rating prediction in explainable recommendation? Can user reviews and trusted social relations help explainable recommendation? What are factors that could affect the explainable recommendations?

To answer this question, we work with three benchmark datasets in our experiments. In our experiments, we have demonstrated the effectiveness of our proposed method and have found significant improvements over state-of-the-art baselines when tested with three benchmark datasets. Viewpoint modeling is helpful for rating prediction and item recommendation. We have also shown that the use of social relations can enhance the accuracy of rating predictions. Because of the explainability of our model, viewpoints also yield explanations of items and of users' preferences.

Limitations of our work include the fact that it ignores topic drift over time. Furthermore, as it is based on topic models, the conditional independence among topics may in principle lead to redundant viewpoints and topics. As to future work, we plan to explore whether ranking-based strategies that integrate our sCVR model can enhance the performance of item recommendation. Also, the transfer of our approach to streaming corpora should give new insights. Finally, we would like to conduct user studies to verify the interpretability of the explanations that sCVR generates and to examine their usefulness in different recommendation scenarios. This chapter is the last research chapter of this thesis. The next chapter will summarize the research presented in this thesis, to answer the research questions raised in Chapter 1, and to provide directions for future research based on findings in this thesis.

8 Conclusions

In this thesis, we have devoted five research chapters to address research problems concerning monitoring social media. We have pursued three angles: summarization, classification and recommendation. Specifically, (1) in Chapter 3 we have considered the task of personalized time-aware tweets summarization, based on user history and influences from "social circles;" (2) in Chapter 4, we have considered the task of contrastive theme summarization of multiple opinionated documents; (3) in Chapter 5, we have considered the task of time-aware multi-viewpoint summarization of social text streams; (4) in Chapter 6, we have considered the task of hierarchical multi-label classification of social text streams; (5) in Chapter 7, we have considered the task of explainable recommendations by addressing two main problems: opinions in users' short comments, and complicated trusted social relations.

In this chapter, we list our main findings, with an outlook on our future research directions. In Section 8.1, we provide a detailed summary of the contributions of our research, and answer the research questions we listed in Chapter 1. We discuss directions for future work in Section 8.2.

8.1 Main Findings

We have addressed research problems about social media monitoring from three angles: summarization, classification and recommendation. We began the research part in the thesis by focusing on the personalized time-aware tweets summarization in Chapter 3. In particular, our research question in this first study was:

RQ1: How can we adapt tweets summarization to a specific user based on a user's history and collaborative social influences? Is it possible to explicitly model the temporal nature of microblogging environment in personalized tweets summarization?

To answer this question, we have considered the task of personalized time-aware tweets summarization, based on user history and influences from "social circles." To handle the dynamic nature of topics and user interests along with the relative sparseness of individual messages, we have proposed a time-aware user behavior model. Based on probabilistic distributions from our proposed topic model, the tweets propagation model (TPM), we have introduced an iterative optimization algorithm to select tweets subject to three key criteria: novelty, coverage and diversity. In our experiments we have verified the effectiveness of our proposed method, showing significant improvements over various state-of-the-art baselines.

To illustrate the performance of our model at different time periods, we select 10 contiguous weeks as the time period. We observe that our proposed methods outperform all other strategies in terms of ROUGE metrics for all test period. We observe a "cold-start" phenomenon, which results from the sparseness of the context in the first time period. In that condition, our proposed methods are nearly equivalent to the state-of-the-art baselines since there are neither social circles nor burst topics during the first time period. After the initial time period, the performance of the the tweets propagation model (TPM) based methods keeps increasing over time until it achieves a stable performance. We find that the tweets propagation model (TPM) based strategies are sensitive to time-aware topic drifting. We also find that the performance of TPM changes with the number of social circles, and the value increases and achieves a maximal value between 3 and 5 social circles. We also find that the collaborative topic modeling used in our proposed methods become more effective when there is a bigger data sparseness issue to overcome.

After investigating personalized time-aware tweets summarization by modeling dynamic topics from social media, we then turned to monitor contrastive topics from documents. At the beginning of Chapter 4, we have identified two main challenges: unknown number of topics and unknown relationships among topics. Therefore, our research question here was:

RQ2: How can we optimize the number of topics in contrastive theme summarization of multiple opinionated documents? How can we model the relations among topics in contrastive topic modeling? Can we find an approach to compress the themes into a diverse and salient subsets of themes?

To answer questions about the optimization of the number of topics and the relations among topics, we have combined the nested Chinese restaurant process with contrastive theme modeling, which outputs a set of threaded topic paths as themes. To enhance the diversity of contrastive theme modeling, we have presented the structured determinantal point process to extract a subset of diverse and salient themes. Based on probabilistic distributions of themes, we generate contrastive summaries subject to three key criteria: contrast, diversity and relevance.

In our experiments, we have demonstrated the effectiveness of our proposed method, finding significant improvements over state-of-the-art baselines tested with three manually annotated datasets. Contrastive theme modeling is helpful for extracting contrastive themes and optimizing the number of topics. We have also shown that structured determinantal point processes are effective for diverse theme extraction. Although we focused mostly on news articles or news-relate articles, our methods are more broadly applicable to other settings with opinionated and conflicted content, such as comment sites or product reviews. Limitations of our work include its ignorance of word dependencies and, being based on hierarchical LDA, the documents that our methods work with should be sufficiently large.

Following our research into contrastive theme summarization using non-parametric processes, in Chapter 5 we have considered the task of time-aware multi-viewpoint summarization of social text streams. We identify four main challenges: ambiguous entities, viewpoint drift, multi-linguality, and the shortness of social text streams, resulting in the following questions:

RQ3: Can we find an approach to help detect time-aware viewpoint drift? Can we find an approach to help detect viewpoints from multilingual social text streams? How can we generate summaries to reflect viewpoints of multi-lingual social text streams?

We propose a dynamic viewpoint modeling strategy to infer multiple viewpoints in the given multilingual social text steams, in which we jointly model topics, entities and sentiment labels. After cross-language viewpoint alignment, we apply a random walk ranking strategy to extract documents to tackle the time-aware multi-viewpoint summarization problem. We demonstrated the effectiveness of our proposed method by showing a significant improvement over various baselines tested with a manually annotated dataset. Our viewpoint tweet topic model is helpful for detecting the viewpoint drift phenomenon and summarizing viewpoints over time.

Although we focused mostly on microblogs, our methods are broadly applicable to other settings with opinionated content, such as comment sites or product reviews. Limitations of our work include its ignorance of viewpoint dependencies and, being based on LDA, its predefined number of viewpoints. Neglected by our method, contrastive viewpoints in multilingual text streams still need to get attention.

After investigating summarization of social media documents, we then turned our research angle to the hierarchical multi-label text classification (HMC) of social text streams. Compared to HMC on stationary documents, HMC on documents in social text streams faces specific challenges: topic drift and the shortness of documents in social text streams. In Chapter 6, we address the HMC problem for documents in social text streams. We identified three main challenges: the shortness of text, topic drift, and hierarchical labels as classification targets, thus we asked:

RQ4: Can we find a method to classify short text streams in a hierarchical multi-label classification setting? How to tackle the *topic drift* and *shortness* in hierarchical multi-label classification of social text streams?

To answer this question, we propose a new strategy to address the task of hierarchical multi-label classification of social text streams. We propose an innovative chunk-based structural learning framework to tackle the hierarchical multi-label classification problem. We verified the effectiveness of our proposed method in hierarchical multi-label classification of social text streams, showing significant improvements over various baselines tested with a manually annotated dataset of tweets.

We tackled the shortness of text by using an entity-based document expansion strategy. We find that the method with document expansion outperforms baselines for most subsets of stationary HMC comparisons. Thus we conclude that document expansion is effective for the stationary HMC task, especially for short text classification. To alleviate the phenomenon of topic drift we presented a dynamic extension to topic models. This extension tracks topics with topic drift over time, based on both local and global topic distributions. We have shown that the performance of our proposed method, in terms of macro F_1, increases over time, rapidly in the early stages, more slowly in the later periods covered by our data set, while not actually plateauing.

Finally, in Chapter 7 we zoomed in on studying the problem of explainable recommendation. Explainable recommendations have been proposed to address the "cold-start" problem and the poor interpretability of recommended results. Recent approaches on explainable recommendation face two challenges: (1) Most existing methods neglect to explicitly analyze opinions for recommendation, thereby missing important opportunities to understand users' viewpoints. (2) Trusted social relations are known to improve the quality of CF recommendation, however, but current methods for explainable recommendations rarely use this information.Therefore, we asked the following question:

RQ5: Can we find an approach to enhance the rating prediction in explainable recommendation? Can user reviews and trusted social relations help explainable recommendation? What are factors that could affect the explainable recommendations?

To answer this question, we have tackled challenges in explainable recommendation by proposing a novel latent variable model, called social collaborative viewpoint regression model, which detects viewpoints and uses social relations. Our model is divided into two parts: viewpoint detection and rating prediction. Based on the probabilistic distribution of viewpoints, we predict users' ratings of items. In our experiments, we have demonstrated the effectiveness of our proposed method and have found significant improvements over state-of-the-art baselines when tested with three benchmark datasets. Viewpoint modeling is helpful for rating prediction and item recommendation. We have also shown that the use of social relations can enhance the accuracy of rating predictions. Because of the explainability of our model, viewpoints also yield explanations of items and of users' preferences.

8.2 Future Research Directions

As described in the previous five chapters, the research presented in this thesis has addressed five research problems in monitoring social media from three different angles: summarization, classification and recommendation. A broad variety of future research has also been motivated. In this section we lay out future research directions on monitoring social media. In particular, we list future research directions in three themes: summarization in social media, hierarchical classification in social media, and explainable recommendation in social media.

8.2.1 Summarization in social media

As we have discussed in Chapters 3, 4, and 5, various approaches have been proposed for social media summarization tasks [167, 170, 209, 224, 247, 251]. However, there are still lots of problems that have not been addressed yet, which can be important as future research directions.

The most serious challenge in social media summarization is how to understand the text. In Chapters 3, 4, and 5, we have proposed novel topic models to monitor dynamic latent topics from social media documents. Because of the expandability of topic models, a potential future direction is to take more information and features into account for summarization task, e.g., URLs appearing in social media documents which could enhance the entity linking setup. It will also be interesting to consider other features for modeling, such as geographic or profile information. "Bag of words" assumption hinders the ability of topic models to tackle context-aware information from social media documents. In recent years, approaches based on deep neural networks and word embeddings, such as long short-term memory (LSTM) [83] and word2vec [161], have been proved effective in short text processing [106, 235]. By considering context-aware information from social media documents, using those neural network based methods is an attractive research direction to enhance the effectiveness of summarization in social media. Tracking the *topic drift* is another challenge in social media summarization, in Chapters 3 and 5, our proposed models are evaluated based on fixed time intervals, which might not accurately reflect bursty topics on social media. Therefore, a novel model that includes dynamic time bins instead of the fixed time granularities, will be another direction for future research. Dynamic stochastic processes, such as the Poisson point process [110] and the Recurrent Chinese restaurant process [5], can be considered here. Meanwhile, supervised and semi-supervised learning can be used to improve the accuracy in social media summarization. The large scale data in social media calls for efficient summarization approaches, which become another important future research direction. Parallel processing methods may enhance the efficiency of topic models on large-scale opinionated documents.

As described in Chapters 3, 4, and 5, our approaches for social media summarization still focus on the extractive summarization task. Generating abstractive summaries for social media documents should give new insights. Most of recent approaches on abstractive summarization are proposed based on sentence compression [25, 68], sentence simplification [248] and neural language models [201]. However, those methods have only been shown to be effective on long documents. For short text streams in social media, the *shortness*, *sparseness* and *topic drift* make it difficult to directly apply existing abstractive summarization methods to social media documents. Hence, exploring an effective approach for abstractive summarization of social text streams is becoming an interesting novel task. Because of the multilinguality of social media documents, another challenge for social media summarization is to tackle the cross-language processing problem in social media summarization. Because shortness and sparseness hinder statistical machine translation in social text streams, in Chapter 5, we applied an entity-linking based method to connect related tweets in different languages. Theoretically, we admit that an ideal solution to tackle this problem should still be based on a real-time statistical machine translation model. Multimedia summarization is another research direction of social media summarization. With the development of social media, more and more multimedia documents have been posted on social media. Multimedia documents in social media may include photos, texts, and videos. Understanding and summarizing those multimedia documents has not yet been addressed.

Evaluation of summarization tasks in social media is also a challenge. Traditional evaluation methods for document summarization is based on ROUGE metrics, which relies on the ground truth of the summarization task. However, large-scale candidate documents from social text streams make it difficult and extremely expensive to get the ground truth. User-study annotations can be applied to evaluate the quality of summaries to enhance the accuracy of interest detection, e.g., via an online evaluation. an extrinsic online user evaluation would give a better indication of the performance of the system.

8.2.2 Hierarchical classification in social media

As we have discussed in Chapter 6, our data collection in the experiments is not so large, thus transfer of our approach to a larger social documents dataset should give new insights. Meanwhile, given a huge data collection in which some part of the documents are labeled, our proposed method in Chapter 6 cannot be applied to address the hierarchical classification problem. Therefore, adaptive learning or semi-supervised learning can be used in future work. Most existing hierarchical multi-label classifiers have an efficiency problem, thus parallel processing may enhance the efficiency of methods on hierarchical multi-label classification of social text streams.

Feature selecting is another challenge for hierarchical classification task in social media. The *shortness* and *sparseness* of social media documents make topic models cannot work as well as in long documents. Weakly supervised representation learning from deep neural networks [24, 83] can be applied to extract features from those short text. *Topic drift* is a serious challenge for feature extraction in social text streams. The Recurrent neural network (RNN) [83, 226] has been proved effective to exhibit dynamic temporal behavior, hence it should be helpful to tackle the drift challenge in hierarchical classification. Based on the representation learning strategy, hierarchical multi-label classification of multimedia social text also can be considered as another future direction.

In Chapter 6 we applied document expansion to extend a short text to a long text using a contextualization strategy. In recent years, document expansion have received increasing attention. Generally, approaches for document expansion can be divided into knowledge-based methods [130] and search-based methods [62]. Transfer of hierarchical classification approaches to new baselines for document expansion might enhance the performance of classification.

Finally, in Chapter 6 we only considered a hierarchical topic classification task of social text streams. In realistic applications, e.g., e-commerce portals, new items usually should be labeled as a new class that has not included in predefined classes. Thus, semi-supervised hierarchical topic modeling [166] can be applied as future work to generate new topics of social text streams.

8.2.3 Explainable recommendations in social media

In Chapter 7 we have proposed a novel latent variable model, called social collaborative viewpoint regression model, which detects viewpoints and uses social relations. However, our method ignores topic drift over time. Furthermore, as it is based on topic models, the conditional independence among topics may in principle lead to redundant viewpoints and topics. As to future work, we plan to explore whether ranking-based strategies that integrating the model in Chapter 7 can enhance the performance of item recommendation. Also, the transfer of our approach to streaming corpora should give new insights.

The interpretability of approaches on explainable recommendation is difficult to evaluate, and should be considered as an important research direction of future work. It would be quite interesting to conduct user studies to verify the interpretability of the explanations that explainable recommendation approaches generate and to examine their usefulness in different recommendation scenarios.

Because social media now includes lots of multimedia documents, applying explainable recommendation strategies to multimedia recommendation can be another research direction. In recent years, an increasing number of computer vision (CV) technologies have been proposed to understand and analyze the content of photos and videos [47, 55, 151]. Given those vision features with semantic features and trusted social relations from social media, how to generate a recommender system that can provide explainable recommendation results is still a topic of ongoing research.

Finally, mobile recommendation is also an important direction for future work. In recent years, as mobile devices with positioning functions become pervasive, massive mobile data motivates an increase number of research on mobile recommendation [142, 263, 275, 276]. Unlike traditional recommendation tasks, a key challenge for mobile recommendation is that the data on each individual user might be quite limited, whereas the recommender system might needs extensive annotated location information to make accurate recommendations [275]. In mobile recommendation, most recent work still focuses on traditional matrix factorization strategies [142, 275] that are difficult to provide explainable recommendations. Therefore, we would like to explore new solutions to mobile recommendation tasks to produce explainable mobile recommendation results.

Bibliography

[1] E. Adar and L. A. Adamic. Tracking information epidemics in blogspace. In *WI*, pages 207–214, 2005. (Cited on page 16.)

[2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17 (6):734–749, 2005. (Cited on page 22.)

[3] C. C. Aggarwal and C. Zhai. *Mining text data*. Springer Science & Business Media, 2012. (Cited on pages 1, 15, and 20.)

[4] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *WSDM*, pages 183–194, 2008. (Cited on page 1.)

[5] A. Ahmed and E. P. Xing. Dynamic non-parametric mixture models and the recurrent chinese restaurant process: With applications to evolutionary clustering. In *SDM*, pages 219–230, 2008. (Cited on page 129.)

[6] A. Ahmed and E. P. Xing. Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream. In *UAI*, pages 20–29, 2012. (Cited on page 24.)

[7] A. Ahmed, L. Hong, and A. Smola. Nested chinese restaurant franchise process: Applications to user tracking and document modeling. In *ICML*, pages 1426–1434, 2013. (Cited on page 50.)

[8] T. Aichner and F. Jacob. Measuring the degree of corporate social media use. *International Journal of Market Research*, 57(2):257–275, 2015. (Cited on page 14.)

[9] M. Albakour, C. Macdonald, and I. Ounis. On sparsity and drift for effective real-time filtering in microblogs. In *CIKM*, pages 419–428, 2013. (Cited on pages 24 and 88.)

[10] J. Allan. Introduction to topic detection and tracking. In *Topic Detection and Tracking*, pages 1–16. Springer, 2002. (Cited on pages 2 and 24.)

[11] J. Allan, C. Wade, and A. Bolivar. Retrieval and novelty detection at the sentence level. In *SIGIR*, pages 314–321, 2003. (Cited on page 19.)

[12] L. AlSumait, D. Barbará, and C. Domeniconi. On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking. In *ICDM*, pages 3–12, 2008. (Cited on page 24.)

[13] G. Amati, G. Amodeo, M. Bianchi, G. Marcone, F. U. Bordoni, C. Gaibisso, G. Gambosi, A. Celi, C. Di Nicola, and M. Flammini. FUB, IASI-CNR, UNIVAQ at TREC 2011 microblog track. In *TREC*, 2011. (Cited on page 16.)

[14] E. Amigó, A. Corujo, J. Gonzalo, E. Meij, and M. de Rijke. Overview of RepLab 2012: Evaluating online reputation management systems. In *CLEF*, 2012. (Cited on pages 17 and 98.)

[15] E. Amigó, J. Carrillo de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, T. Martin, E. Meij, M. de Rijke, and D. Spina. Overview of RepLab 2013: Evaluating online reputation monitoring systems. In *CLEF*, pages 333–352, 2013.

[16] E. Amigó, J. Carrillo-de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, E. Meij, M. de Rijke, and D. Spina. Overview of RepLab 2014: Author profiling and reputation dimensions for online reputation management. In *CLEF*, pages 307–322, 2014. (Cited on page 17.)

[17] L. Aroyo and C. Welty. The three sides of crowdtruth. *Journal of Human Computation*, 1:31–34, 2014. (Cited on pages 77 and 78.)

[18] J. Aslam, F. Diaz, M. Ekstrand-Abueg, R. McCreadie, V. Pavlu, and T. Sakai. TREC 2014 temporal summarization track overview. In *TREC*, 2015. (Cited on page 16.)

[19] J. A. Aslam, M. Ekstrand-Abueg, V. Pavlu, F. Diaz, and T. Sakai. TREC 2013 temporal summarization. In *TREC*, 2013. (Cited on page 16.)

[20] R. Baeza-Yates and B. Ribeiro-Neto. *Modern information retrieval*. ACM, 1999. (Cited on page 14.)

[21] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya. Hierarchical multi-label prediction of gene function. *Bioinformatics*, 22(7):830–836, 2006. (Cited on page 21.)

[22] R. M. Bell and Y. Koren. Improved neighborhood-based collaborative filtering. In *KDD Cup and Workshop at the KDD*, pages 7–14, 2007. (Cited on page 22.)

[23] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *SIGCOMM*, pages 49–62, 2009. (Cited on page 17.)

[24] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. (Cited on page 130.)

[25] T. Berg-Kirkpatrick, D. Gillick, and D. Klein. Jointly learning to extract and compress. In *ACL-HLT*, pages 481–490, 2011. (Cited on page 129.)

[26] A. Beutel, K. Murray, C. Faloutsos, and A. J. Smola. Cobafi: Collaborative bayesian filtering. In *WWW*, pages 97–108, 2014. (Cited on pages 109, 110, and 120.)

[27] P. Bhargava, T. Phan, J. Zhou, and J. Lee. Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor factorization on sparse user-generated data. In *WWW*, pages 130–140, 2015. (Cited on page 23.)

[28] W. Bi and J. T. Kwok. Multi-label classification on tree-and dag-structured hierarchies. In *ICML*, pages 17–24, 2011. (Cited on pages 5, 21, 87, 89, 94, and 100.)

[29] M. Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Beyond Personalization 2005: A Workshop on the Next Stage of Recommender Systems Research*, pages 13–18, 2005. (Cited on page 23.)

[30] C. Bishop. *Pattern recognition and machine learning*. Springer, 2007. (Cited on page 20.)

[31] D. M. Blei and J. D. Lafferty. Dynamic topic models. In *ICML*, pages 113–120, 2006. (Cited on pages 24, 71, 72, 88, 90, and 92.)

[32] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003. (Cited on pages 5, 21, 23, 24, 25, 26, 31, 43, 52, 59, 60, 72, 74, 88, 108, and 112.)

[33] D. M. Blei, T. L. Griffiths, and M. I. Jordan. The nested chinese restaurant process and bayesian non-parametric inference of topic hierarchies. *Journal of the ACM*, 57(2):7:1–7:30, 2010. (Cited on pages 24, 50, 53, 59, and 60.)

[34] H. Blockeel, L. Schietgat, J. Struyf, S. Džeroski, and A. Clare. Decision trees for hierarchical multilabel classification: A case study in functional genomics. In *ECML&PKDD*, pages 18–29, 2006. (Cited on pages 5, 21, and 87.)

[35] D. Bollegala, Y. Matsuo, and M. Ishizuka. Measuring semantic similarity between words using web search engines. In *WWW*, pages 757–766, 2007. (Cited on pages 20 and 21.)

[36] A. Borodin. Determinantal point processes. In *The Oxford Handbook of Random Matrix Theory*. Oxford University Press, 2009. (Cited on page 27.)

[37] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW*, pages 491–495, 1998. (Cited on page 15.)

[38] S. Carter, W. Weerkamp, and M. Tsagkias. Microblog language identification: Overcoming the limitations of short, unedited and idiomatic text. *Language Resources and Evaluation*, 47(1):195–215, 2013. (Cited on pages 39 and 97.)

[39] A. Celikyilmaz and D. Hakkani-Tur. A hybrid hierarchical model for multi-document summarization. In *ACL*, pages 815–824, 2010. (Cited on pages 18 and 49.)

[40] N. Cesa-Bianchi, C. Gentile, and L. Zaniboni. Incremental algorithms for hierarchical classification. *Journal of Machine Learning Research*, 7:31–54, 2006. (Cited on pages 5, 21, and 87.)

[41] D. Chakrabarti and K. Punera. Event summarization using tweets. In *ICWSM*, pages 66–73, 2011. (Cited on pages 2, 3, 19, 24, 29, and 42.)

[42] W. Chan, W. Yang, J. Tang, J. Du, X. Zhou, and W. Wang. Community question topic categorization via hierarchical kernelized classification. In *CIKM*, pages 959–968, 2013. (Cited on pages 5 and 87.)

[43] C. Chen, X. Zheng, Y. Wang, F. Hong, and Z. Lin. Context-aware collaborative topic regression with social matrix factorization for recommender systems. In *AAAI*, pages 9–15, 2014. (Cited on pages 2, 23, 107, 117, 118, 119, and 120.)

[44] J. Chen and D. Warren. Cost-sensitive learning for large-scale hierarchical classification of commercial products. In *CIKM*, pages 1351–1360, 2013. (Cited on page 100.)

[45] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu. Collaborative personalized tweet recommendation. In *SIGIR*, pages 661–670, 2012. (Cited on pages 1, 17, 23, 29, and 30.)

[46] M. Chen, X. Jin, and D. Shen. Short text classification improved by learning multi-granularity topics. In *IJCAI*, pages 1776–1781, 2011. (Cited on pages 20 and 21.)

[47] T. Chen, F. X. Yu, J. Chen, Y. Cui, Y.-Y. Chen, and S.-F. Chang. Object-based visual sentiment concept analysis and application. In *MM*, pages 367–376, 2014. (Cited on page 131.)

[48] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, pages 199–208, 2009. (Cited on page 16.)

[49] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *WWW*, pages 925–936, 2014. (Cited on page 16.)

[50] A. Clare. *Machine Learning and Data Mining for Yeast Functional Genomics*. PhD thesis, University of Wales, 2003. (Cited on pages 100 and 101.)

[51] F. Crestani, M. Lalmas, C. J. van Rijsbergen, and I. Campbell. “Is this document relevant?... probably”: A survey of probabilistic models in information retrieval. *ACM Computing Surveys*, 30(4):528–552,

1998. (Cited on page 15.)
[52] A. Dasgupta, R. Kumar, and S. Ravi. Summarization through submodularity and dispersion. In *ACL*, pages 1014–1022, 2013. (Cited on page 18.)
[53] G. De Francisci Morales, A. Gionis, and C. Lucchese. From chatter to headlines: Harnessing the real-time web for personalized news recommendation. In *WSDM*, pages 153–162, 2012. (Cited on pages 1 and 29.)
[54] J.-Y. Delort and E. Alfonseca. DualSum: A topic-model based approach for update summarization. In *EACL*, pages 214–223, 2012. (Cited on pages 19 and 75.)
[55] J. Deng, J. Krause, and L. Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In *CVPR*, pages 580–587, 2013. (Cited on page 131.)
[56] Q. Diao and J. Jiang. Recurrent chinese restaurant process with a duration-based discount for event identification from Twitter. In *SDM*, pages 388–397, 2014. (Cited on pages 2, 24, 72, and 88.)
[57] Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim. Finding bursty topics from microblogs. In *ACL*, pages 536–544, 2012. (Cited on pages 1, 2, 24, 30, and 88.)
[58] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In *KDD*, pages 193–202, 2014. (Cited on pages 1, 23, and 107.)
[59] S. Dori-Hacohen and J. Allan. Detecting controversy on the web. In *CIKM*, pages 1845–1848, 2013. (Cited on pages 3 and 49.)
[60] N. Du, L. Song, M. Gomez-Rodriguez, and H. Zha. Scalable influence estimation in continuous-time diffusion networks. In *NIPS*, pages 3147–3155, 2013. (Cited on page 16.)
[61] Y. Duan, F. Wei, C. Zhumin, Z. Ming, and Y. Shum. Twitter topic summarization by ranking tweets using social influence and content quality. In *COLING*, pages 763–780, 2012. (Cited on page 19.)
[62] M. Efron, P. Organisciak, and K. Fenlon. Improving retrieval of short texts through document expansion. In *SIGIR*, pages 911–920, 2012. (Cited on pages 17 and 130.)
[63] G. Erkan and D. R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22:457–479, 2004. (Cited on pages 1, 18, 40, 59, 60, 75, 80, and 91.)
[64] G. Erkan and D. R. Radev. Lexpagerank: Prestige in multi-document text summarization. In *EMNLP*, pages 365–371, 2004. (Cited on page 74.)
[65] Y. Fang, L. Si, N. Somasundaram, et al. Mining contrastive opinions on political texts using cross-perspective topic model. In *WSDM*, pages 63–72, 2012. (Cited on pages 2, 4, and 67.)
[66] D. Fensel, B. Leiter, and I. Stavrakantonakis. Social media monitoring. *Semantic Technology Institute, Innsbruck*, 2012. (Cited on page 16.)
[67] K. Filippova. Multi-sentence compression: Finding shortest paths in word graphs. In *COLING*, pages 322–330, 2010. (Cited on page 20.)
[68] K. Filippova, E. Alfonseca, C. A. Colmenares, L. Kaiser, and O. Vinyals. Sentence compression by deletion with LSTMs. In *EMNLP*, pages 360–368, 2015. (Cited on page 129.)
[69] T. Finley and T. Joachims. Training structural SVMs when exact inference is intractable. In *ICML*, pages 304–311, 2008. (Cited on page 95.)
[70] S. Fisher and B. Roark. Query-focused supervised sentence ranking for update summaries. *TAC*, 2008. (Cited on pages 19 and 75.)
[71] J. Friedman, T. Hastie, and R. Tibshirani. *The Elements of Statistical Learning*, volume 1. Springer series in statistics Springer, Berlin, 2001. (Cited on page 20.)
[72] N. Fuhr. Optimum polynomial retrieval functions based on the probability ranking principle. *ACM Transactions on Information Systems*, 7(3):183–204, 1989. (Cited on page 15.)
[73] G. P. C. Fung, J. X. Yu, and H. Lu. Classifying text streams in the presence of concept drifting. In *PAKDD*, pages 373–383, 2004. (Cited on page 89.)
[74] K. Ganesan, C. Zhai, and J. Han. Opinosis: A graph-based approach to abstractive summarization of highly redundant opinions. In *COLING*, pages 340–348, 2010. (Cited on pages 2, 3, 4, 18, 20, 49, and 67.)
[75] K. Ganesan, C. Zhai, and E. Viegas. Micropinion generation: An unsupervised approach to generating ultra-concise summaries of opinions. In *WWW*, pages 869–878, 2012. (Cited on pages 2, 3, 20, 49, 59, 60, and 69.)
[76] S. Gao, J. Ma, and Z. Chen. Modeling and predicting retweeting dynamics on microblogging platforms. In *WSDM*, pages 107–116, 2015. (Cited on page 16.)
[77] W. Gao, P. Li, and K. Darwish. Joint topic modeling for event summarization across news and social media streams. In *CIKM*, pages 1173–1182, 2012. (Cited on pages 4 and 67.)

[78] J. Gillenwater, A. Kulesza, and B. Taskar. Discovering diverse and salient threads in document collections. In *EMNLP-CoNLL*, pages 710–720, 2012. (Cited on pages 24 and 27.)
[79] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992. (Cited on page 22.)
[80] M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *KDD*, pages 1019–1028, 2010. (Cited on page 16.)
[81] M. Gomez-Rodriguez, L. Song, N. Du, H. Zha, and B. Schölkopf. Influence estimation and maximization in continuous-time diffusion networks. *ACM Transactions on Information Systems*, 34(2):9, 2016. (Cited on page 16.)
[82] D. Graus, Z. Ren, M. de Rijke, D. van Dijk, H. Henseler, and N. van der Knaap. Semantic search in e-discovery: An interdisciplinary approach. In *ICAIL 2013 Workshop on Standards for Using Predictive Coding, Machine Learning, and Other Advanced Search and Review Methods in E-Discovery (DESI V Workshop)*, 2013. (Cited on pages 10 and 17.)
[83] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):855–868, 2009. (Cited on pages 129 and 130.)
[84] T. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101:5228–5235, 2004. (Cited on pages 26, 40, 43, and 78.)
[85] Q. Guo, F. Diaz, and E. Yom-Tov. Updating users about time critical events. In *ECIR*, pages 483–494, 2013. (Cited on page 16.)
[86] Y. Guo and S. Gu. Multi-label classification using conditional dependency networks. In *IJCAI*, pages 1300–1305, 2011. (Cited on pages 1 and 21.)
[87] X. Han and L. Sun. An entity-topic model for entity linking. In *EMNLP*, pages 105–115, 2012. (Cited on page 24.)
[88] D. Harman. Overview of the first text retrieval conference. In *TREC*, 1992. (Cited on page 15.)
[89] X. He, T. Chen, M.-Y. Kan, and X. Chen. Trirank: Review-aware explainable recommendation by modeling aspects. In *CIKM*, pages 1661–1670, 2015. (Cited on pages 5, 22, 23, and 107.)
[90] T. Hofmann. Probabilistic latent semantic indexing. In *SIGIR*, pages 50–57, 1999. (Cited on pages 15 and 24.)
[91] L. Hong, R. Bekkerman, J. Adler, and B. D. Davison. Learning to rank social update streams. In *SIGIR*, pages 651–660, 2012. (Cited on page 17.)
[92] M. Hu and B. Liu. Mining opinion features in customer reviews. In *AAAI*, pages 755–760, 2004. (Cited on pages 2, 3, 20, and 49.)
[93] M. Hu and B. Liu. Opinion extraction and summarization on the web. In *AAAI*, pages 1621–1624, 2006. (Cited on pages 18 and 20.)
[94] S. Huang, S. Wang, T.-Y. Liu, J. Ma, Z. Chen, and J. Veijalainen. Listwise collaborative filtering. In *SIGIR*, pages 343–352, 2015. (Cited on page 23.)
[95] X. Huang, X. Wan, and J. Xiao. Comparative news summarization using concept-based optimization. *Knowledge and Information Systems*, 38(3):691–716, 2013. (Cited on page 49.)
[96] M. Imran, C. Castillo, F. Diaz, and S. Vieweg. Processing social media messages in mass emergency: A survey. *ACM Computing Surveys*, 47(4):67, 2015. (Cited on page 1.)
[97] O. Inel, K. Khamkham, T. Cristea, A. Dumitrache, A. Rutjes, J. van der Ploeg, L. Romaszko, L. Aroyo, and R.-J. Sips. Crowdtruth: Machine-human computation framework for harnessing disagreement in gathering annotated data. In *ISWC*, pages 486–504, 2014. (Cited on page 77.)
[98] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda. Topic tracking model for analyzing consumer purchase behavior. In *IJCAI*, pages 1427–1432, 2009. (Cited on pages 17, 24, 34, 72, and 73.)
[99] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3): 264–323, 1999. (Cited on page 15.)
[100] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys*, pages 135–142, 2010. (Cited on pages 23, 107, 119, and 120.)
[101] O. Jin, N. N. Liu, K. Zhao, Y. Yu, and Q. Yang. Transferring topical knowledge from auxiliary long texts for short text clustering. In *CIKM*, pages 775–784, 2011. (Cited on page 40.)
[102] T. Joachims. *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998. (Cited on pages 15 and 87.)
[103] T. Joachims. Optimizing search engines using clickthrough data. In *KDD*, pages 133–142, 2002. (Cited on page 15.)
[104] T. Joyce and R. Needham. The thesaurus approach to information retrieval. *American Documentation*, 9(3):192–197, 1958. (Cited on page 15.)

[105] A. M. Kaplan and M. Haenlein. Users of the world, unite! The challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010. (Cited on page 1.)

[106] T. Kenter and M. de Rijke. Short text similarity with word embeddings. In *CIKM*, pages 1411–1420, 2015. (Cited on page 129.)

[107] H. D. Kim and C. Zhai. Generating comparative summaries of contradictory opinions in text. In *CIKM*, pages 385–394, 2009. (Cited on pages 3, 4, 49, and 67.)

[108] H. D. Kim, K. Ganesan, P. Sondhi, and C. Zhai. Comprehensive review of opinion summarization. Technical report, University of Illinois at Urbana-Champaign, 2011. (Cited on pages 2, 3, and 49.)

[109] H. D. Kim, M. G. Castellanos, M. Hsu, C. Zhai, U. Dayal, and R. Ghosh. Ranking explanatory sentences for opinion summarization. In *SIGIR*, pages 1069–1072, 2013. (Cited on page 20.)

[110] J. F. C. Kingman. *Poisson processes*, volume 3. Clarendon Press, 1992. (Cited on page 129.)

[111] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999. (Cited on page 15.)

[112] D. Koller and M. Sahami. Hierarchically classifying documents using very few words. In *ICML*, pages 170–178, 1997. (Cited on pages 20 and 21.)

[113] X. Kong, B. Cao, and P. S. Yu. Multi-label classification by mining label and instance correlations from heterogeneous information networks. In *KDD*, pages 614–622, 2013. (Cited on page 21.)

[114] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 6(8):30–37, 2009. (Cited on page 22.)

[115] E. Kouloumpis, T. Wilson, and J. D. Moore. Twitter sentiment analysis: The good the bad and the omg! *ICWSM*, pages 538–541, 2011. (Cited on page 17.)

[116] A. Kulesza and B. Taskar. Structured determinantal point processes. In *NIPS*, pages 1171–1179, 2010. (Cited on pages 24, 27, 28, 50, and 55.)

[117] A. Kulesza and B. Taskar. Determinantal point processes for machine learning. *Foundation & Trends in Machine Learning*, 5(2–3):123–286, 2012. (Cited on pages 24, 27, and 55.)

[118] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *WWW*, pages 591–600, 2010. (Cited on page 39.)

[119] J. H. Lau, N. Collier, and T. Baldwin. On-line trend analysis with topic models: # Twitter trends detection topic model online. In *COLING*, pages 1519–1534, 2012. (Cited on page 2.)

[120] G. Lebanon and Y. Zhao. Local likelihood modeling of temporal text streams. In *ICML*, pages 552–559, 2008. (Cited on pages 90, 100, and 101.)

[121] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2001. (Cited on pages 22, 119, and 120.)

[122] K. Lerman and R. McDonald. Contrastive summarization: An experiment with consumer reviews. In *NAACL*, pages 113–116, 2009. (Cited on pages 2 and 20.)

[123] J. Leskovec, M. McGlohon, C. Faloutsos, N. S. Glance, and M. Hurst. Patterns of cascading behavior in large blog graphs. In *SDM*, pages 551–556, 2007. (Cited on page 16.)

[124] F. Li et al. Structure-aware review mining and summarization. In *COLING*, pages 653–661, 2010. (Cited on pages 20, 24, 59, 60, 69, and 80.)

[125] L. Li, K. Zhou, G.-R. Xue, H. Zha, and Y. Yu. Enhancing diversity, coverage and balance for summarization through structure learning. In *WWW*, pages 71–80, 2009. (Cited on pages 18, 28, 37, 49, and 88.)

[126] L. Li, K. Zhou, G.-R. Xue, H. Zha, and Y. Yu. Video summarization via transferrable structured learning. In *WWW*, pages 287–296, 2011. (Cited on page 28.)

[127] P. Li, Y. Wang, W. Gao, and J. Jiang. Generating aspect-oriented multi-document summarization with event-aspect model. In *EMNLP*, pages 1137–1146, 2011. (Cited on pages 4 and 67.)

[128] S. Liang. *Fusion and Diversification in Information Retrieval*. PhD thesis, University of Amsterdam, 2014. (Cited on page 15.)

[129] S. Liang, Z. Ren, and M. de Rijke. Fusion helps diversification. In *SIGIR*, pages 303–312, 2014. (Cited on pages 10 and 66.)

[130] S. Liang, Z. Ren, and M. de Rijke. The impact of semantic document expansion on cluster-based fusion for microblog search. In *ECIR*, pages 493–499, 2014. (Cited on pages 10, 17, and 130.)

[131] S. Liang, Z. Ren, and M. de Rijke. Personalized search result diversification via structured learning. In *KDD*, pages 751–760, 2014. (Cited on page 10.)

[132] S. Liang, Z. Ren, W. Weerkamp, E. Meij, and M. de Rijke. Time-aware rank aggregation for microblog search. In *CIKM*, pages 989–998, 2014. (Cited on page 10.)

[133] C.-Y. Lin. Rouge: A package for automatic evaluation of summaries. In *ACL*, pages 74–81, 2004. (Cited on pages 41 and 60.)

[134] C.-Y. Lin and E. Hovy. From single to multi-document summarization: A prototype system and its evaluation. In *ACL*, pages 457–464, 2002. (Cited on page 18.)

[135] J. Lin, M. Efron, Y. Wang, and G. Sherman. Overview of the TREC-2014 microblog track. In *TREC*, 2014. (Cited on page 16.)

[136] W.-H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann. Which side are you on? Identifying perspectives at the document and sentence levels. In *CoNLL*, pages 109–116, 2006. (Cited on page 58.)

[137] G. Ling, M. R. Lyu, and I. King. Ratings meet reviews, a combined approach to recommend. In *RecSys*, pages 105–112, 2014. (Cited on pages 2, 5, 22, 23, 107, 109, 113, 117, and 119.)

[138] B. Liu, M. Hu, and J. Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *WWW*, pages 342–351, 2005. (Cited on pages 4 and 67.)

[139] J. S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966, 1994. (Cited on pages 26 and 93.)

[140] K.-L. Liu, W.-J. Li, and M. Guo. Emoticon smoothed language models for Twitter sentiment analysis. In *AAAI*, pages 1678–1684, 2012. (Cited on page 17.)

[141] S. Liu, S. Wang, F. Zhu, J. Zhang, and R. Krishnan. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In *SIGMOD*, pages 51–62, 2014. (Cited on pages 5 and 87.)

[142] X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *CIKM*, pages 733–738, 2013. (Cited on page 131.)

[143] G. Long, L. Chen, X. Zhu, and C. Zhang. TCSST: Transfer classification of short & sparse text using external data. In *CIKM*, pages 764–772, 2012. (Cited on pages 5 and 87.)

[144] P. Lops, M. De Gemmis, and G. Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*, pages 73–105. Springer, 2011. (Cited on page 22.)

[145] Y. Lu, C. Zhai, and N. Sundaresan. Rated aspect summarization of short comments. In *WWW*, pages 131–140, 2009. (Cited on page 20.)

[146] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159–165, 1958. (Cited on page 18.)

[147] Z. Luo, M. Osborne, S. Petrovic, and T. Wang. Improving Twitter retrieval by exploiting structural information. In *AAAI*, pages 648–654, 2012. (Cited on page 16.)

[148] H. Ma, I. King, and M. Lyu. Learning to recommend with social trust ensemble. In *SIGIR*, pages 203–210, 2009. (Cited on pages 17 and 23.)

[149] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *WSDM*, pages 287–296, 2011. (Cited on pages 17 and 23.)

[150] C. D. Manning, P. Raghavan, H. Schütze, et al. *Introduction to information retrieval*. Cambridge university press Cambridge, 2008. (Cited on pages 14 and 15.)

[151] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *ICCV*, pages 1784–1791, 2011. (Cited on page 131.)

[152] M. E. Maron and J. L. Kuhns. On relevance, probabilistic indexing and information retrieval. *Journal of the ACM*, 7(3):216–244, 1960. (Cited on page 15.)

[153] A. H. Maslow and K. J. Lewis. *Maslow's hierarchy of needs*. Salenger Incorporated, 1987. (Cited on page 14.)

[154] J. McAuley and J. Leskovec. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *RecSys*, pages 165–172, 2013. (Cited on pages 23 and 119.)

[155] R. McCreadie, C. Macdonald, I. Ounis, M. Osborne, and S. Petrovic. Scalable distributed event detection for Twitter. In *International Conference on Big Data*, pages 543–549, 2013. (Cited on page 24.)

[156] R. McCreadie, C. Macdonald, and I. Ounis. Incremental update summarization: Adaptive sentence selection based on prevalence and novelty. In *CIKM*, pages 301–310, 2014. (Cited on pages 16, 18, 19, 75, and 80.)

[157] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *WWW*, pages 171–180, 2007. (Cited on pages 2, 4, 20, and 67.)

[158] E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. In *WSDM*, pages 563–572, 2012. (Cited on pages 4, 39, 67, 69, 71, and 91.)

[159] X. Meng, F. Wei, X. Liu, M. Zhou, S. Li, and H. Wang. Entity-centric topic-oriented opinion summarization in Twitter. In *KDD*, pages 379–387, 2012. (Cited on pages 18 and 20.)

[160] M. Michelson and S. A. Macskassy. Discovering users' topics of interest on Twitter: A first look. In *The Fourth Workshop on Analytics for Noisy Unstructured Text Data*, pages 73–80, 2010. (Cited on page 1.)

[161] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013. (Cited on pages 111 and 129.)

[162] T. Minka. Estimating a dirichlet distribution. *Technical Report, M.I.T*, 2000. (Cited on page 116.)

[163] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *NIPS*, pages 1257–1264, 2007. (Cited on pages 22, 119, and 120.)

[164] C. N. Mooers. The next twenty years in information retrieval. *Journal of the American Society for Information Science*, 11(3):229, 1960. (Cited on page 15.)

[165] A. Nenkova and K. McKeown. Automatic summarization. *Foundation & Trends in Information Retrieval*, 5(2-3):103–233, 2012. (Cited on pages 1, 18, and 19.)

[166] V.-A. Nguyen, J. L. Boyd-Graber, and P. Resnik. Lexical and hierarchical topic regression. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *NIPS*, pages 1106–1114, 2013. (Cited on page 130.)

[167] J. Nichols, J. Mahmud, and C. Drews. Summarizing sporting events using Twitter. In *IUI*, pages 189–198, 2012. (Cited on pages 19 and 128.)

[168] K. Nishida, R. Banno, K. Fujimura, and T. Hoshide. Tweet classification by data compression. In *2011 International Workshop on Detecting and Exploiting Cultural Diversity on the Social Web*, pages 29–34, 2011. (Cited on page 21.)

[169] K. Nishida, T. Hoshide, and K. Fujimura. Improving tweet stream classification by detecting changes in word probability. In *SIGIR*, pages 971–980, 2012. (Cited on pages 2, 5, 17, 20, 87, 88, 89, and 90.)

[170] B. O'Connor, M. Krieger, and D. Ahn. Tweetmotif: Exploratory search and topic summarization for Twitter. *ICWSM*, pages 2–3, 2010. (Cited on pages 3, 17, 19, 29, and 128.)

[171] D. Odijk, E. Meij, and M. de Rijke. Feeding the second screen: Semantic linking based on subtitles. In *OAIR*, pages 9–16, 2013. (Cited on pages 88 and 91.)

[172] A. Oghina, M. Breuss, M. Tsagkias, and M. de Rijke. Predicting IMDB movie ratings using social media. In *ECIR*, pages 333–352, 2012. (Cited on page 16.)

[173] I. Ounis, C. Macdonald, J. Lin, and I. Soboroff. Overview of the TREC-2011 microblog track. In *TREC*, 2011. (Cited on page 16.)

[174] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *EMNLP*, pages 79–86, 2002. (Cited on page 2.)

[175] M. Paul and R. Girju. A two-dimensional topic-aspect model for discovering multi-faceted topics. *AAAI*, pages 545–550, 2010. (Cited on pages 4, 49, 58, 59, 60, 67, and 80.)

[176] M. J. Paul, C. Zhai, and R. Girju. Summarizing contrastive viewpoints in opinionated text. In *EMNLP*, pages 66–76, 2010. (Cited on pages 3, 4, 18, 20, 24, 49, 53, 58, 67, 69, and 79.)

[177] M.-H. Peetz. *Time-Aware Online Reputation Analysis*. PhD thesis, University of Amsterdam, 2015. (Cited on pages 14 and 15.)

[178] M.-H. Peetz, M. de Rijke, and R. Kaptein. Estimating reputation polarity on microblog posts. *Inf. Processing & Management*, 52:193–216, 2015. (Cited on pages 4, 17, and 67.)

[179] M. Pennacchiotti, F. Silvestri, H. Vahabi, and R. Venturini. Making your interests follow you on Twitter. In *CIKM*, pages 165–174, 2012. (Cited on pages 3, 23, 29, and 30.)

[180] J. Petterson and T. S. Caetano. Submodular multi-label learning. In *NIPS*, pages 1512–1520, 2011. (Cited on page 21.)

[181] X.-H. Phan, L.-M. Nguyen, and S. Horiguchi. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *WWW*, pages 91–100, 2008. (Cited on pages 21 and 88.)

[182] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *SIGIR*, pages 275–281, 1998. (Cited on page 15.)

[183] M. F. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980. (Cited on page 39.)

[184] D. Radev, T. Allison, S. Blair-Goldensohn, J. Blitzer, A. Celebi, S. Dimitrov, E. Drabek, A. Hakim, W. Lam, D. Liu, et al. Mead-a platform for multidocument multilingual text summarization. In *LREC*, 2004. (Cited on page 18.)

[185] D. R. Radev, H. Jing, M. Styś, and D. Tam. Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6):919–938, 2004. (Cited on page 42.)

[186] D. Ramage, S. T. Dumais, and D. J. Liebling. Characterizing microblogs with topic models. In *ICWSM*, pages 130–137, 2010. (Cited on pages 2, 24, and 30.)

[187] L. Ren, D. B. Dunson, and L. Carin. The dynamic hierarchical dirichlet process. In *ICML*, pages 824–831, 2008. (Cited on page 50.)

[188] Z. Ren and M. de Rijke. Summarizing contrastive themes via hierarchical non-parametric processes. In *SIGIR*, pages 93–102, 2015. (Cited on pages 9 and 79.)

[189] Z. Ren, J. Ma, S. Wang, and Y. Liu. Summarizing web forum threads based on a latent topic propagation process. In *CIKM*, 2011. (Cited on pages 2, 10, and 19.)

[190] Z. Ren, S. Liang, E. Meij, and M. de Rijke. Personalized time-aware tweets summarization. In *SIGIR*, pages 513–522, 2013. (Cited on pages 9, 14, 17, 19, 24, 49, and 99.)
[191] Z. Ren, D. van Dijk, D. Graus, N. van der Knaap, H. Henseler, and M. de Rijke. Semantic linking and contextualization for social forensic text analysis. In *European Intelligence and Security Informatics Conference*, pages 96–99, 2013. (Cited on pages 11 and 17.)
[192] Z. Ren, M.-H. Peetz, S. Liang, W. van Dolen, and M. de Rijke. Hierarchical multi-label classification of social text streams. In *SIGIR*, pages 213–222, 2014. (Cited on pages 1, 2, and 10.)
[193] Z. Ren, O. Inel, L. Aroyo, and M. de Rijke. Time-aware multi-viewpoint summarization of multilingual social text streams. In *CIKM*, 2016. (Cited on page 10.)
[194] Z. Ren, S. Liang, P. Li, S. Wang, and M. de Rijke. Social collaborative viewpoint regression for explainable recommendations. In *Under submission*, 2016. (Cited on page 10.)
[195] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997. (Cited on page 22.)
[196] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *1994 ACM conference on Computer supported cooperative work*, pages 175–186, 1994. (Cited on page 22.)
[197] S. E. Robertson and K. S. Jones. Relevance weighting of search terms. *Journal of the American Society for Information science*, 27(3):129–146, 1976. (Cited on page 15.)
[198] M. G. Rodriguez and B. Schölkopf. Influence maximization in continuous time diffusion networks. In *ICML*, pages 313–320, 2012. (Cited on page 16.)
[199] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *UAI*, pages 487–494, 2004. (Cited on pages 24, 30, 31, and 42.)
[200] J. Rousu, C. Saunders, S. Szedmak, and J. Shawe-Taylor. Kernel-based learning of hierarchical multilabel classification models. *Journal of Machine Learning Research*, 7:1601–1626, 2006. (Cited on page 21.)
[201] A. Rush, S. Chopra, and J. Weston. A neural attention model for sentence summarization. In *EMNLP*, pages 379–389, 2015. (Cited on page 129.)
[202] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *WWW*, pages 851–860, 2010. (Cited on page 1.)
[203] T. Salles, L. Rocha, G. L. Pappa, F. Mourão, W. Meira Jr, and M. Gonçalves. Temporally-aware algorithms for document classification. In *SIGIR*, pages 307–314, 2010. (Cited on page 1.)
[204] G. Salton and M. E. Lesk. Computer evaluation of indexing and text processing. *Journal of the ACM (JACM)*, 15(1):8–36, 1968. (Cited on page 15.)
[205] S. Sarawagi and R. Gupta. Accurate max-margin training for structured output spaces. In *ICML*, pages 888–895, 2008. (Cited on page 28.)
[206] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, pages 285–295, 2001. (Cited on page 22.)
[207] S. Shalev-Shwartz and Y. Singer. Efficient learning of label ranking by soft projections onto polyhedra. *Journal of Machine Learning Research*, 7:1567–1599, 2006. (Cited on page 21.)
[208] B. Sharifi, M.-A. Hutton, and J. Kalita. Automatic summarization of Twitter topics. In *National Workshop on Design and Analysis of Algorithms*, pages 121–128, 2010. (Cited on pages 1 and 3.)
[209] B. Sharifi, M.-A. Hutton, and J. Kalita. Summarizing microblogs automatically. In *NAACL*, pages 685–688, 2010. (Cited on pages 17, 19, and 128.)
[210] C. Shen and T. Li. Learning to rank for query-focused multi-document summarization. In *ICDM*, pages 626–634, 2011. (Cited on page 18.)
[211] D. Shen, J. Sun, H. Li, Q. Yang, and Z. Chen. Document summarization using conditional random fields. In *IJCAI*, pages 2862–2867, 2007. (Cited on pages 18 and 49.)
[212] Y. Shi, M. Larson, and A. Hanjalic. List-wise learning to rank with matrix factorization for collaborative filtering. In *RecSys*, pages 269–272, 2010. (Cited on pages 119 and 120.)
[213] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, and A. Hanjalic. CLiMF: Learning to maximize reciprocal rank with collaborative less-is-more filtering. In *RecSys*, pages 139–146, 2012. (Cited on pages 22, 119, and 120.)
[214] Y. Shi, M. Larson, and A. Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys*, 47(1):3, 2014. (Cited on page 22.)
[215] L. Shou, Z. Wang, K. Chen, and G. Chen. Sumblr: Continuous summarization of evolving tweet streams. In *SIGIR*, pages 533–542, 2013. (Cited on pages 17, 19, and 24.)
[216] N. Slonim and N. Tishby. The power of word clusters for text classification. In *ECIR*, 2001. (Cited on pages 15 and 87.)

[217] N. Slonim, N. Friedman, and N. Tishby. Unsupervised document classification using sequential information maximization. In *SIGIR*, pages 129–136, 2002. (Cited on page 15.)

[218] I. Soboroff, I. Ounis, C. Macdonald, and J. Lin. Overview of the TREC-2012 microblog track. In *TREC*, 2012. (Cited on page 16.)

[219] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, pages 1631–1642, 2013. (Cited on pages 53 and 111.)

[220] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas. Short text classification in Twitter to improve information filtering. In *SIGIR*, pages 841–842, 2010. (Cited on page 21.)

[221] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009:1–19, 2009. (Cited on page 22.)

[222] A. Sun. Short text classification using very few words. In *SIGIR*, pages 1145–1146, 2012. (Cited on page 21.)

[223] N. A. Syed, H. Liu, and K. K. Sung. Handling concept drifts in incremental learning with support vector machines. In *KDD*, pages 317–321, 1999. (Cited on page 88.)

[224] H. Takamura, H. Yokono, and M. Okumura. Summarizing a document stream. *ECIR*, pages 177–188, 2011. (Cited on pages 17, 19, and 128.)

[225] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin. Learning sentiment-specific word embedding for Twitter sentiment classification. In *ACL*, pages 1555–1565, 2014. (Cited on page 117.)

[226] D. Tang, B. Qin, and T. Liu. Document modeling with gated recurrent neural network for sentiment classification. In *EMNLP*, pages 1422–1432, 2015. (Cited on page 130.)

[227] L. Tang, S. Rajan, and V. K. Narayanan. Large-scale multi-label classification via metalabeler. In *WWW*, pages 211–220, 2009. (Cited on page 101.)

[228] N. Tintarev and J. Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender Systems Handbook*, pages 479–510. Springer, 2011. (Cited on page 23.)

[229] M. Tomasoni and M. Huang. Metadata-aware measures for answer summarization in community question answering. In *ACL*, pages 760–769, 2010. (Cited on pages 2, 17, and 18.)

[230] K. Toutanova, C. Brockett, M. Gamon, J. Jagarlamudi, H. Suzuki, and L. Vanderwende. The pythy summarization system: Microsoft research at DUC 2007. In *DUC*, 2007. (Cited on page 18.)

[231] M. Tsagkias. *Mining Social Media: Tracking Content and Predicting Behavior*. PhD thesis, University of Amsterdam, 2012. (Cited on page 1.)

[232] M. Tsagkias, M. de Rijke, and W. Weerkamp. News comments: Exploring, modeling, and online prediction. In *ECIR 2010*, pages 191–203, 2010. (Cited on page 16.)

[233] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 6:1453–1484, 2005. (Cited on pages 28 and 88.)

[234] D. van Dijk, D. Graus, Z. Ren, H. Henseler, and M. de Rijke. Who is involved? Semantic search for e-discovery. In *The 15th International Conference on Artificial Intelligence & Law*, 2015. (Cited on page 10.)

[235] C. van Gysel, M. de Rijke, and M. Worring. Unsupervised, efficient and semantic expertise retrieval. In *WWW*, pages 1069–1079, 2016. (Cited on page 129.)

[236] O. van Laere, I. Bordino, Y. Mejova, and M. Lalmas. DEESSE: Entity-driven exploratory and serendipitous search system. In *CIKM*, pages 2072–2074, 2014. (Cited on pages 4 and 67.)

[237] C. Vens, J. Struyf, L. Schietgat, S. Džeroski, and H. Blockeel. Decision trees for hierarchical multi-label classification. *Machine Learning*, 73(2):185–214, 2008. (Cited on pages 21 and 100.)

[238] J. Vig, S. Sen, and J. Riedl. Tagsplanations: Explaining recommendations using tags. In *IUI*, pages 47–56, 2009. (Cited on page 23.)

[239] H. M. Wallach. Topic modeling: Beyond bag-of-words. In *ICML*, pages 977–984, 2006. (Cited on pages 30, 34, and 114.)

[240] X. Wan. Update summarization based on co-ranking with constraints. In *COLING*, pages 1291–1300, 2012. (Cited on pages 19, 75, and 80.)

[241] X. Wan and J. Yang. Multi-document summarization using cluster-based link analysis. In *SIGIR*, pages 299–306, 2008. (Cited on pages 15, 18, 49, 59, 60, and 92.)

[242] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *KDD*, pages 448–456, 2011. (Cited on pages 5, 23, 72, 107, 108, and 119.)

[243] D. Wang, S. Zhu, T. Li, and Y. Gong. Comparative document summarization via discriminative sentence selection. *ACM Transactions on Knowledge Discovery from Data*, 6(3):12:1–12:18, 2012. (Cited on pages 4, 67, and 69.)

[244] W. Weerkamp and M. de Rijke. Activity prediction: A Twitter-based exploration. In *SIGIR 2012 Workshop on Time-aware Information Access*, 2012. (Cited on page 17.)
[245] F. Wei, W. Li, Q. Lu, and Y. He. Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In *SIGIR*, pages 283–290, 2008. (Cited on pages 1 and 18.)
[246] X. Wei, J. Sun, and X. Wang. Dynamic mixture models for multiple time series. In *IJCAI*, pages 2909–2914, 2007. (Cited on pages 24 and 34.)
[247] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: Finding topic-sensitive influential twitterers. In *WSDM*, pages 261–270, 2010. (Cited on pages 19, 29, and 128.)
[248] S. Wubben, A. van den Bosch, and E. Krahmer. Sentence simplification by monolingual machine translation. In *ACL*, pages 1015–1024, 2012. (Cited on page 129.)
[249] Y. Xu, W. Lam, and T. Lin. Collaborative filtering incorporating review text and co-clusters of hidden user communities and item groups. In *CIKM*, pages 251–260, 2014. (Cited on pages 108 and 109.)
[250] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user posting behavior on social media. In *SIGIR*, pages 545–554, 2012. (Cited on pages 17, 23, 30, 34, 42, and 43.)
[251] D. Yajuan, C. Zhimin, W. Furu, Z. Ming, and H.-Y. Shum. Twitter topic summarization by ranking tweets using social influence and content quality. In *COLING*, pages 763–779, 2012. (Cited on pages 3, 17, 19, and 128.)
[252] R. Yan, X. Wan, J. Otterbacher, L. Kong, X. Li, and Y. Zhang. Evolutionary timeline summarization: A balanced optimization framework via iterative substitution. In *SIGIR*, pages 745–754, 2011. (Cited on pages 18, 19, and 38.)
[253] X. Yan, J. Guo, Y. Lan, and X. Cheng. A biterm topic model for short texts. In *WWW*, pages 1445–1456, 2013. (Cited on page 17.)
[254] B. Yang, Y. Lei, D. Liu, and J. Liu. Social collaborative filtering by trust. In *IJCAI*, pages 2747–2753, 2013. (Cited on pages 23, 107, 119, and 120.)
[255] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *ICDM*, pages 599–608, 2010. (Cited on page 16.)
[256] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: Joint friendship and interest propagation in social networks. In *WWW*, pages 537–546, 2011. (Cited on pages 16, 17, and 23.)
[257] S.-H. Yang, A. Kolcz, A. Schlaikjer, and P. Gupta. Large-scale high-precision topic modeling on Twitter. In *KDD*, pages 1907–1916. ACM, 2014. (Cited on page 24.)
[258] Y. Yang. A study of thresholding strategies for text categorization. In *SIGIR*, pages 137–145, 2001. (Cited on pages 15 and 87.)
[259] Z. Yang, K. Cai, J. Tang, L. Zhang, Z. Su, and J. Li. Social context summarization. In *SIGIR*, pages 255–264, 2011. (Cited on pages 17 and 19.)
[260] M. Ye, X. Liu, and W.-C. Lee. Exploring social influence for recommendation: A generative model approach. In *SIGIR*, pages 671–680, 2012. (Cited on pages 1, 17, 23, and 29.)
[261] W.-T. Yih and C. Meek. Improving similarity measures for short segments of text. In *AAAI*, pages 1489–1494, 2007. (Cited on pages 20 and 21.)
[262] H. Yin, B. Cui, L. Chen, Z. Hu, and Z. Huang. A temporal context-aware model for user behavior modeling in social media systems. In *SIGMOD*, pages 1543–1554, 2014. (Cited on page 17.)
[263] H. Yin, X. Zhou, Y. Shao, H. Wang, and S. Sadiq. Joint modeling of user check-in behaviors for point-of-interest recommendation. In *CIKM*, pages 1631–1640, 2015. (Cited on page 131.)
[264] Y. Yue and T. Joachims. Predicting diverse subsets using structural SVMs. In *ICML*, pages 1224–1231, 2008. (Cited on pages 28, 88, and 95.)
[265] R. Zafarani, M. A. Abbasi, and H. Liu. *Social media mining: An introduction*. Cambridge University Press, 2014. (Cited on page 1.)
[266] S. Zelikovitz and H. Hirsh. Transductive LSI for short text classification problems. In *FLAIRS*, pages 556–561, 2004. (Cited on page 21.)
[267] K. Zhai and J. Boyd-Graber. Online latent dirichlet allocation with infinite vocabulary. In *ICML*, pages 561–569, 2013. (Cited on page 24.)
[268] S. Zhang, X. Jin, D. Shen, B. Cao, X. Ding, and X. Zhang. Short text classification by detecting information path. In *CIKM*, pages 727–732, 2013. (Cited on pages 5, 21, and 87.)
[269] X. Zhang, S. Lu, B. He, J. Xu, and T. Luo. Ucas at trec 2012 microblog track. In *TREC*, 2012. (Cited on page 16.)
[270] Y. Zhang, M. Zhang, Y. Liu, S. Ma, and S. Feng. Localized matrix factorization for recommendation based on matrix block diagonal forms. In *WWW*, pages 1511–1520, 2013. (Cited on page 22.)

[271] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *SIGIR*, pages 83–92, 2014. (Cited on pages 5, 22, 23, 107, and 119.)

[272] W. X. Zhao, J. Jiang, J. He, Y. Song, P. Achananuparp, E.-P. Lim, and X. Li. Topical keyphrase extraction from Twitter. In *ACL*, pages 379–388, 2011. (Cited on pages 19, 24, 30, and 42.)

[273] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li. Comparing Twitter and traditional media using topic models. In *ECIR*, pages 338–349, 2011. (Cited on pages 2 and 24.)

[274] Y. Zhao, S. Liang, Z. Ren, J. Ma, E. Yilmaz, and M. de Rijke. Explainable user clustering in short text streams. In *SIGIR*, 2016. (Cited on pages 11 and 17.)

[275] V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI*, pages 236–241, 2010. (Cited on page 131.)

[276] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Towards mobile intelligence: Learning from gps history data for collaborative recommendation. *Artificial Intelligence*, 184:17–37, 2012. (Cited on page 131.)

[277] B. Zhu, J. Gao, X. Han, C. Shi, S. Liu, Y. Liu, and X. Cheng. ICTNET at microblog track TREC 2012. In *TREC*, 2012. (Cited on page 16.)

[278] S. Zhu, K. Yu, Y. Chi, and Y. Gong. Combining content and link for classification using matrix factorization. In *SIGIR*, pages 487–494, 2007. (Cited on page 15.)

[279] Y. Zhu, Y. Lan, J. Guo, P. Du, and X. Cheng. A novel relational learning-to-rank approach for topic-focused multi-document summarization. In *ICDM*, pages 927–936, 2013. (Cited on page 18.)

Summary

A key characteristic of social media research is the ambition to monitor the content of social media, i.e., text from social media platforms, social relations among users, and changes in social media data over time. In this thesis, we present research on understanding social media along three dimensions: summarization, classification and recommendation.

Our first line of work concerns summarization of social media documents. Firstly, we address the task of time-aware tweets summarization, based on a user's history and collaborative influences from "social circles." We propose a time-aware user behavior model to infer dynamic probabilistic distributions over interests and topics. Based on probabilistic distributions from our proposed model, we explicitly consider novelty, coverage, and diversity to arrive at an iterative optimization algorithm for selecting tweets. Secondly, we continue our research on summarization by addressing the task of contrastive theme summarization. We combine the nested Chinese restaurant process with contrastive theme modeling, which outputs a set of threaded topic paths as themes. We present the structured determinantal point process to extract a subset of diverse and salient themes. Based on probabilistic distributions of themes, we generate contrastive summaries subject to three key criteria: contrast, diversity and relevance. Lastly, we address the viewpoint summarization of multilingual streaming corpora. We propose a dynamic latent factor model to explicitly characterize a set of viewpoints through which entities, topics and sentiment labels during a time interval are derived jointly; we connect viewpoints in different languages by using an entity-based semantic similarity measure; and we employ an update viewpoint summarization strategy to generate a time-aware summary to reflect viewpoints.

Our second line of work is hierarchical multi-label classification of social text streams. Concept drift, complicated relations among classes, and the limited length of documents in social text streams make this a challenging problem. We extend each short document in social text streams to a more comprehensive representation via state-of-the-art entity linking and sentence ranking strategies. From documents extended in this manner, we infer dynamic probabilistic distributions over topics. For the final phase we propose a chunk-based structural optimization strategy to classify each document into multiple classes.

Our third line of work is explainable recommendation task via viewpoint modeling, which not only predicts a numerical rating for an item, but also generates explanations for users' preferences. We propose a latent variable model for predicting item ratings that uses user opinions and social relations to generate explanations. To this end we use viewpoints from both user reviews and trusted social relations. Our method includes two core ingredients: inferring viewpoints and predicting user ratings. We apply a Gibbs EM sampler to infer posterior distributions of our method.

In our experiments we have verified the effectiveness of our proposed methods for monitoring social media, showing improvements over various state-of-the-art baselines. This thesis provides insights and findings that can be used to facilitate the understanding of social media content, for a range of tasks in social media retrieval.

Samenvatting

Een kerneigenschap van het onderzoek naar sociale media is de ambitie om de inhoud, zoals de tekst, relaties tussen gebruikers en veranderingen door de tijd te monitoren. In dit proefschrift presenteren we langs drie dimensies onderzoek naar het begrijpen van sociale media: samenvatten, classificeren en aanbevelen.

De eerste lijn van onderzoek is het samenvatten van documenten van sociale media. Ten eerste kijken we naar de taak van het tijdsbewust samenvatten van tweets, gebaseerd op de geschiedenis van een gebruiker en collaboratieve invloeden van "sociale kringen." We presenteren een tijdsbewust model van gebruikersgedrag om de dynamische kansverdeling over interesses en onderwerpen af te leiden. Op basis van deze kansverdelingen, beschouwen we "versheid," dekking en diversiteit om tot een iteratief optimalisatie-algoritme te komen voor het selecteren van tweets. Als tweede zetten we de lijn van onderzoek naar samenvatten door met het samenvatten van tegenstrijdige standpunten. We combineren het "Nested Chinese Restaurant Process" met het modelleren van contrastieve standpunten, om tot een set van *threaded topic paths* te komen. We presenteren het *structured determinantal point process* voor het extraheren van diverse en in het oog springende thema's. Gebaseerd op de distributie van thema's genereren we contrastieve samenvattingen op basis van drie kerncriteria: contrast, diversiteit en relevantie. Als laatste kijken we naar het samenvatten van standpunten in meertalige, stromende corpora. We stellen een *dynamic latent factor* model voor om een verzameling van standpunten expliciet te karakteriseren waarbij entiteiten, onderwerpen en sentiment labels gedurende een tijdsinterval gezamenlijk worden afgeleid. We verbinden standpunten in verschillende talen door middel van semantische gelijkenis en leren hoe we een tijdsbewuste samenvatting van standpunten kunnen maken.

Onze tweede onderzoekslijn behandelt multi-label hiërarchisch classificeren van social media streams. Dit is een uitdagend probleem, vanwege concepten die geleidelijk van betekenis veranderen, ingewikkelde relaties tussen verschillende klassen en de geringe lengte van sociale media teksten. Om dit aan te pakken, breiden we de sociale media teksten uit tot meer omvattende representaties met behulp van state-of-the-art entity-linking technologie en het gebruik van strategieën voor het rangschikken van zinnen. Van de teksten die we op deze manier uitbreiden, leiden we de dynamische kansverdelingen af over themas. Als laatste stellen we een chunk-based structural optimization strategy voor om elke tekst te classificeren in meerdere klassen.

Onze derde onderzoekslijn richt zich op het genereren van verklaarde aanbevelingen met behulp van het modelleren van standpunten. Hiervoor moet naast het voorspellen van een waardering voor een item ook een verklaring worden gegeven voor de voorspelde waardering. Hiertoe stellen we een model voor dat gebruik maakt van latente variabelen om de waardering van items te voorspellen, en bovendien de meningen en sociale relaties van gebruikers gebruikt om een verklaring te geven. We gebruiken hiervoor de standpunten uit zowel gebruikersrecensies als sociale relaties. Onze methode bevat twee kern-ingrediënten: het afleiden van standpunten en het voorspellen van waarderingen.

In onze experimenten hebben we de effectiviteit bepaald van onze methoden voor het monitoren van social media. We laten verbeteringen zien over verschillende methoden uit de literatuur. De bevindingen en inzichten in dit proefschrift faciliteren het begrijpen van social media inhoud voor een scala aan taken in social media retrieval.

SIKS Dissertation Series

1998

1 Johan van den Akker (CWI) *DEGAS: An Active, Temporal Database of Autonomous Objects*
2 Floris Wiesman (UM) *Information Retrieval by Graphically Browsing Meta-Information*
3 Ans Steuten (TUD) *A Contribution to the Linguistic Analysis of Business Conversations*
4 Dennis Breuker (UM) *Memory versus Search in Games*
5 E. W. Oskamp (RUL) *Computerondersteuning bij Straftoemeting*

1999

1 Mark Sloof (VUA) *Physiology of Quality Change Modelling: Automated modelling of*
2 Rob Potharst (EUR) *Classification using decision trees and neural nets*
3 Don Beal (UM) *The Nature of Minimax Search*
4 Jacques Penders (UM) *The practical Art of Moving Physical Objects*
5 Aldo de Moor (KUB) *Empowering Communities: A Method for the Legitimate User-Driven*
6 Niek J. E. Wijngaards (VUA) *Re-design of compositional systems*
7 David Spelt (UT) *Verification support for object database design*
8 Jacques H. J. Lenting (UM) *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism*

2000

1 Frank Niessink (VUA) *Perspectives on Improving Software Maintenance*
2 Koen Holtman (TUe) *Prototyping of CMS Storage Management*
3 Carolien M. T. Metselaar (UvA) *Sociaal-organisatorische gevolgen van kennistechnologie*
4 Geert de Haan (VUA) *ETAG, A Formal Model of Competence Knowledge for User Interface*
5 Ruud van der Pol (UM) *Knowledge-based Query Formulation in Information Retrieval*
6 Rogier van Eijk (UU) *Programming Languages for Agent Communication*
7 Niels Peek (UU) *Decision-theoretic Planning of Clinical Patient Management*
8 Veerle Coupé (EUR) *Sensitivity Analyis of Decision-Theoretic Networks*
9 Florian Waas (CWI) *Principles of Probabilistic Query Optimization*
10 Niels Nes (CWI) *Image Database Management System Design Considerations, Algorithms and Architecture*
11 Jonas Karlsson (CWI) *Scalable Distributed Data Structures for Database Management*

2001

1 Silja Renooij (UU) *Qualitative Approaches to Quantifying Probabilistic Networks*
2 Koen Hindriks (UU) *Agent Programming Languages: Programming with Mental Models*
3 Maarten van Someren (UvA) *Learning as problem solving*
4 Evgueni Smirnov (UM) *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*
5 Jacco van Ossenbruggen (VUA) *Processing Structured Hypermedia: A Matter of Style*
6 Martijn van Welie (VUA) *Task-based User Interface Design*
7 Bastiaan Schonhage (VUA) *Diva: Architectural Perspectives on Information Visualization*
8 Pascal van Eck (VUA) *A Compositional Semantic Structure for Multi-Agent Systems Dynamics*
9 Pieter Jan 't Hoen (RUL) *Towards Distributed Development of Large Object-Oriented Models*
10 Maarten Sierhuis (UvA) *Modeling and Simulating Work Practice*
11 Tom M. van Engers (VUA) *Knowledge Management*

2002

1 Nico Lassing (VUA) *Architecture-Level Modifiability Analysis*
2 Roelof van Zwol (UT) *Modelling and searching web-based document collections*
3 Henk Ernst Blok (UT) *Database Optimization Aspects for Information Retrieval*
4 Juan Roberto Castelo Valdueza (UU) *The Discrete Acyclic Digraph Markov Model in Data Mining*
5 Radu Serban (VUA) *The Private Cyberspace Modeling Electronic*
6 Laurens Mommers (UL) *Applied legal epistemology: Building a knowledge-based ontology of*
7 Peter Boncz (CWI) *Monet: A Next-Generation DBMS Kernel For Query-Intensive*
8 Jaap Gordijn (VUA) *Value Based Requirements Engineering: Exploring Innovative*
9 Willem-Jan van den Heuvel (KUB) *Integrating Modern Business Applications with Objectified Legacy*
10 Brian Sheppard (UM) *Towards Perfect Play of Scrabble*
11 Wouter C. A. Wijngaards (VUA) *Agent Based Modelling of Dynamics: Biological and Organisational Applications*
12 Albrecht Schmidt (UvA) *Processing XML in Database Systems*
13 Hongjing Wu (TUe) *A Reference Architecture for Adaptive Hypermedia Applications*

14 Wieke de Vries (UU) *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*
15 Rik Eshuis (UT) *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*
16 Pieter van Langen (VUA) *The Anatomy of Design: Foundations, Models and Applications*
17 Stefan Manegold (UvA) *Understanding, Modeling, and Improving Main-Memory Database Performance*

2003

1 Heiner Stuckenschmidt (VUA) *Ontology-Based Information Sharing in Weakly Structured Environments*
2 Jan Broersen (VUA) *Modal Action Logics for Reasoning About Reactive Systems*
3 Martijn Schuemie (TUD) *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*
4 Milan Petkovic (UT) *Content-Based Video Retrieval Supported by Database Technology*
5 Jos Lehmann (UvA) *Causation in Artificial Intelligence and Law: A modelling approach*
6 Boris van Schooten (UT) *Development and specification of virtual environments*
7 Machiel Jansen (UvA) *Formal Explorations of Knowledge Intensive Tasks*
8 Yongping Ran (UM) *Repair Based Scheduling*
9 Rens Kortmann (UM) *The resolution of visually guided behaviour*
10 Andreas Lincke (UvT) *Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture*
11 Simon Keizer (UT) *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*
12 Roeland Ordelman (UT) *Dutch speech recognition in multimedia information retrieval*
13 Jeroen Donkers (UM) *Nosce Hostem: Searching with Opponent Models*
14 Stijn Hoppenbrouwers (KUN) *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*
15 Mathijs de Weerdt (TUD) *Plan Merging in Multi-Agent Systems*
16 Menzo Windhouwer (CWI) *Feature Grammar Systems: Incremental Maintenance of Indexes to Digital Media Warehouses*
17 David Jansen (UT) *Extensions of Statecharts with Probability, Time, and Stochastic Timing*
18 Levente Kocsis (UM) *Learning Search Decisions*

2004

1 Virginia Dignum (UU) *A Model for Organizational Interaction: Based on Agents, Founded in Logic*
2 Lai Xu (UvT) *Monitoring Multi-party Contracts for E-business*
3 Perry Groot (VUA) *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*
4 Chris van Aart (UvA) *Organizational Principles for Multi-Agent Architectures*
5 Viara Popova (EUR) *Knowledge discovery and monotonicity*
6 Bart-Jan Hommes (TUD) *The Evaluation of Business Process Modeling Techniques*
7 Elise Boltjes (UM) *Voorbeeldig onderwijs: voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes*
8 Joop Verbeek (UM) *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise*
9 Martin Caminada (VUA) *For the Sake of the Argument: explorations into argument-based reasoning*
10 Suzanne Kabel (UvA) *Knowledge-rich indexing of learning-objects*
11 Michel Klein (VUA) *Change Management for Distributed Ontologies*
12 The Duy Bui (UT) *Creating emotions and facial expressions for embodied agents*
13 Wojciech Jamroga (UT) *Using Multiple Models of Reality: On Agents who Know how to Play*
14 Paul Harrenstein (UU) *Logic in Conflict. Logical Explorations in Strategic Equilibrium*
15 Arno Knobbe (UU) *Multi-Relational Data Mining*
16 Federico Divina (VUA) *Hybrid Genetic Relational Search for Inductive Learning*
17 Mark Winands (UM) *Informed Search in Complex Games*
18 Vania Bessa Machado (UvA) *Supporting the Construction of Qualitative Knowledge Models*
19 Thijs Westerveld (UT) *Using generative probabilistic models for multimedia retrieval*
20 Madelon Evers (Nyenrode) *Learning from Design: facilitating multidisciplinary design teams*

2005

1 Floor Verdenius (UvA) *Methodological Aspects of Designing Induction-Based Applications*
2 Erik van der Werf (UM) *AI techniques for the game of Go*
3 Franc Grootjen (RUN) *A Pragmatic Approach to the Conceptualisation of Language*
4 Nirvana Meratnia (UT) *Towards Database Support for Moving Object data*
5 Gabriel Infante-Lopez (UvA) *Two-Level Probabilistic Grammars for Natural Language Parsing*
6 Pieter Spronck (UM) *Adaptive Game AI*
7 Flavius Frasincar (TUe) *Hypermedia Presentation Generation for Semantic Web Information Systems*

8 Richard Vdovjak (TUe) *A Model-driven Approach for Building Distributed Ontology-based Web Applications*
9 Jeen Broekstra (VUA) *Storage, Querying and Inferencing for Semantic Web Languages*
10 Anders Bouwer (UvA) *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*
11 Elth Ogston (VUA) *Agent Based Matchmaking and Clustering: A Decentralized Approach to Search*
12 Csaba Boer (EUR) *Distributed Simulation in Industry*
13 Fred Hamburg (UL) *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*
14 Borys Omelayenko (VUA) *Web-Service configuration on the Semantic Web: Exploring how semantics meets pragmatics*
15 Tibor Bosse (VUA) *Analysis of the Dynamics of Cognitive Processes*
16 Joris Graaumans (UU) *Usability of XML Query Languages*
17 Boris Shishkov (TUD) *Software Specification Based on Re-usable Business Components*
18 Danielle Sent (UU) *Test-selection strategies for probabilistic networks*
19 Michel van Dartel (UM) *Situated Representation*
20 Cristina Coteanu (UL) *Cyber Consumer Law, State of the Art and Perspectives*
21 Wijnand Derks (UT) *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*

2006

1 Samuil Angelov (TUe) *Foundations of B2B Electronic Contracting*
2 Cristina Chisalita (VUA) *Contextual issues in the design and use of information technology in organizations*
3 Noor Christoph (UvA) *The role of metacognitive skills in learning to solve problems*
4 Marta Sabou (VUA) *Building Web Service Ontologies*
5 Cees Pierik (UU) *Validation Techniques for Object-Oriented Proof Outlines*
6 Ziv Baida (VUA) *Software-aided Service Bundling: Intelligent Methods & Tools for Graphical Service Modeling*
7 Marko Smiljanic (UT) *XML schema matching: balancing efficiency and effectiveness by means of clustering*
8 Eelco Herder (UT) *Forward, Back and Home Again: Analyzing User Behavior on the Web*
9 Mohamed Wahdan (UM) *Automatic Formulation of the Auditor's Opinion*
10 Ronny Siebes (VUA) *Semantic Routing in Peer-to-Peer Systems*
11 Joeri van Ruth (UT) *Flattening Queries over Nested Data Types*
12 Bert Bongers (VUA) *Interactivation: Towards an e-cology of people, our technological environment, and the arts*
13 Henk-Jan Lebbink (UU) *Dialogue and Decision Games for Information Exchanging Agents*
14 Johan Hoorn (VUA) *Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change*
15 Rainer Malik (UU) *CONAN: Text Mining in the Biomedical Domain*
16 Carsten Riggelsen (UU) *Approximation Methods for Efficient Learning of Bayesian Networks*
17 Stacey Nagata (UU) *User Assistance for Multitasking with Interruptions on a Mobile Device*
18 Valentin Zhizhkun (UvA) *Graph transformation for Natural Language Processing*
19 Birna van Riemsdijk (UU) *Cognitive Agent Programming: A Semantic Approach*
20 Marina Velikova (UvT) *Monotone models for prediction in data mining*
21 Bas van Gils (RUN) *Aptness on the Web*
22 Paul de Vrieze (RUN) *Fundaments of Adaptive Personalisation*
23 Ion Juvina (UU) *Development of Cognitive Model for Navigating on the Web*
24 Laura Hollink (VUA) *Semantic Annotation for Retrieval of Visual Resources*
25 Madalina Drugan (UU) *Conditional log-likelihood MDL and Evolutionary MCMC*
26 Vojkan Mihajlovic (UT) *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*
27 Stefano Bocconi (CWI) *Vox Populi: generating video documentaries from semantically annotated media repositories*
28 Borkur Sigurbjornsson (UvA) *Focused Information Access using XML Element Retrieval*

2007

1 Kees Leune (UvT) *Access Control and Service-Oriented Architectures*
2 Wouter Teepe (RUG) *Reconciling Information Exchange and Confidentiality: A Formal Approach*
3 Peter Mika (VUA) *Social Networks and the Semantic Web*
4 Jurriaan van Diggelen (UU) *Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach*
5 Bart Schermer (UL) *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance*
6 Gilad Mishne (UvA) *Applied Text Analytics for Blogs*

7 Natasa Jovanovic' (UT) *To Whom It May Concern: Addressee Identification in Face-to-Face Meetings*
8 Mark Hoogendoorn (VUA) *Modeling of Change in Multi-Agent Organizations*
9 David Mobach (VUA) *Agent-Based Mediated Service Negotiation*
10 Huib Aldewereld (UU) *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*
11 Natalia Stash (TUe) *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*
12 Marcel van Gerven (RUN) *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*
13 Rutger Rienks (UT) *Meetings in Smart Environments: Implications of Progressing Technology*
14 Niek Bergboer (UM) *Context-Based Image Analysis*
15 Joyca Lacroix (UM) *NIM: a Situated Computational Memory Model*
16 Davide Grossi (UU) *Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems*
17 Theodore Charitos (UU) *Reasoning with Dynamic Networks in Practice*
18 Bart Orriens (UvT) *On the development an management of adaptive business collaborations*
19 David Levy (UM) *Intimate relationships with artificial partners*
20 Slinger Jansen (UU) *Customer Configuration Updating in a Software Supply Network*
21 Karianne Vermaas (UU) *Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005*
22 Zlatko Zlatev (UT) *Goal-oriented design of value and process models from patterns*
23 Peter Barna (TUe) *Specification of Application Logic in Web Information Systems*
24 Georgina Ramírez Camps (CWI) *Structural Features in XML Retrieval*
25 Joost Schalken (VUA) *Empirical Investigations in Software Process Improvement*

2008

1 Katalin Boer-Sorbán (EUR) *Agent-Based Simulation of Financial Markets: A modular, continuous-time approach*
2 Alexei Sharpanskykh (VUA) *On Computer-Aided Methods for Modeling and Analysis of Organizations*
3 Vera Hollink (UvA) *Optimizing hierarchical menus: a usage-based approach*
4 Ander de Keijzer (UT) *Management of Uncertain Data: towards unattended integration*
5 Bela Mutschler (UT) *Modeling and simulating causal dependencies on process-aware information systems from a cost perspective*
6 Arjen Hommersom (RUN) *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*
7 Peter van Rosmalen (OU) *Supporting the tutor in the design and support of adaptive e-learning*
8 Janneke Bolt (UU) *Bayesian Networks: Aspects of Approximate Inference*
9 Christof van Nimwegen (UU) *The paradox of the guided user: assistance can be counter-effective*
10 Wauter Bosma (UT) *Discourse oriented summarization*
11 Vera Kartseva (VUA) *Designing Controls for Network Organizations: A Value-Based Approach*
12 Jozsef Farkas (RUN) *A Semiotically Oriented Cognitive Model of Knowledge Representation*
13 Caterina Carraciolo (UvA) *Topic Driven Access to Scientific Handbooks*
14 Arthur van Bunningen (UT) *Context-Aware Querying: Better Answers with Less Effort*
15 Martijn van Otterlo (UT) *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains*
16 Henriette van Vugt (VUA) *Embodied agents from a user's perspective*
17 Martin Op 't Land (TUD) *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*
18 Guido de Croon (UM) *Adaptive Active Vision*
19 Henning Rode (UT) *From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search*
20 Rex Arendsen (UvA) *Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven*
21 Krisztian Balog (UvA) *People Search in the Enterprise*
22 Henk Koning (UU) *Communication of IT-Architecture*
23 Stefan Visscher (UU) *Bayesian network models for the management of ventilator-associated pneumonia*
24 Zharko Aleksovski (VUA) *Using background knowledge in ontology matching*
25 Geert Jonker (UU) *Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency*
26 Marijn Huijbregts (UT) *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*
27 Hubert Vogten (OU) *Design and Implementation Strategies for IMS Learning Design*
28 Ildiko Flesch (RUN) *On the Use of Independence Relations in Bayesian Networks*

29 Dennis Reidsma (UT) *Annotations and Subjective Machines: Of Annotators, Embodied Agents, Users, and Other Humans*
30 Wouter van Atteveldt (VUA) *Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content*
31 Loes Braun (UM) *Pro-Active Medical Information Retrieval*
32 Trung H. Bui (UT) *Toward Affective Dialogue Management using Partially Observable Markov Decision Processes*
33 Frank Terpstra (UvA) *Scientific Workflow Design: theoretical and practical issues*
34 Jeroen de Knijf (UU) *Studies in Frequent Tree Mining*
35 Ben Torben Nielsen (UvT) *Dendritic morphologies: function shapes structure*

2009

1 Rasa Jurgelenaite (RUN) *Symmetric Causal Independence Models*
2 Willem Robert van Hage (VUA) *Evaluating Ontology-Alignment Techniques*
3 Hans Stol (UvT) *A Framework for Evidence-based Policy Making Using IT*
4 Josephine Nabukenya (RUN) *Improving the Quality of Organisational Policy Making using Collaboration Engineering*
5 Sietse Overbeek (RUN) *Bridging Supply and Demand for Knowledge Intensive Tasks: Based on Knowledge, Cognition, and Quality*
6 Muhammad Subianto (UU) *Understanding Classification*
7 Ronald Poppe (UT) *Discriminative Vision-Based Recovery and Recognition of Human Motion*
8 Volker Nannen (VUA) *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*
9 Benjamin Kanagwa (RUN) *Design, Discovery and Construction of Service-oriented Systems*
10 Jan Wielemaker (UvA) *Logic programming for knowledge-intensive interactive applications*
11 Alexander Boer (UvA) *Legal Theory, Sources of Law & the Semantic Web*
12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin) *Operating Guidelines for Services*
13 Steven de Jong (UM) *Fairness in Multi-Agent Systems*
14 Maksym Korotkiy (VUA) *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)*
15 Rinke Hoekstra (UvA) *Ontology Representation: Design Patterns and Ontologies that Make Sense*
16 Fritz Reul (UvT) *New Architectures in Computer Chess*
17 Laurens van der Maaten (UvT) *Feature Extraction from Visual Data*
18 Fabian Groffen (CWI) *Armada, An Evolving Database System*
19 Valentin Robu (CWI) *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*
20 Bob van der Vecht (UU) *Adjustable Autonomy: Controling Influences on Decision Making*
21 Stijn Vanderlooy (UM) *Ranking and Reliable Classification*
22 Pavel Serdyukov (UT) *Search For Expertise: Going beyond direct evidence*
23 Peter Hofgesang (VUA) *Modelling Web Usage in a Changing Environment*
24 Annerieke Heuvelink (VUA) *Cognitive Models for Training Simulations*
25 Alex van Ballegooij (CWI) *RAM: Array Database Management through Relational Mapping*
26 Fernando Koch (UU) *An Agent-Based Model for the Development of Intelligent Mobile Services*
27 Christian Glahn (OU) *Contextual Support of social Engagement and Reflection on the Web*
28 Sander Evers (UT) *Sensor Data Management with Probabilistic Models*
29 Stanislav Pokraev (UT) *Model-Driven Semantic Integration of Service-Oriented Applications*
30 Marcin Zukowski (CWI) *Balancing vectorized query execution with bandwidth-optimized storage*
31 Sofiya Katrenko (UvA) *A Closer Look at Learning Relations from Text*
32 Rik Farenhorst (VUA) *Architectural Knowledge Management: Supporting Architects and Auditors*
33 Khiet Truong (UT) *How Does Real Affect Affect Affect Recognition In Speech?*
34 Inge van de Weerd (UU) *Advancing in Software Product Management: An Incremental Method Engineering Approach*
35 Wouter Koelewijn (UL) *Privacy en Politiegegevens: Over geautomatiseerde normatieve informatie-uitwisseling*
36 Marco Kalz (OUN) *Placement Support for Learners in Learning Networks*
37 Hendrik Drachsler (OUN) *Navigation Support for Learners in Informal Learning Networks*
38 Riina Vuorikari (OU) *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context*
39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin) *Service Substitution: A Behavioral Approach Based on Petri Nets*
40 Stephan Raaijmakers (UvT) *Multinomial Language Learning: Investigations into the Geometry of Language*
41 Igor Berezhnyy (UvT) *Digital Analysis of Paintings*
42 Toine Bogers (UvT) *Recommender Systems for Social Bookmarking*

43 Virginia Nunes Leal Franqueira (UT) *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*
44 Roberto Santana Tapia (UT) *Assessing Business-IT Alignment in Networked Organizations*
45 Jilles Vreeken (UU) *Making Pattern Mining Useful*
46 Loredana Afanasiev (UvA) *Querying XML: Benchmarks and Recursion*

2010

1 Matthijs van Leeuwen (UU) *Patterns that Matter*
2 Ingo Wassink (UT) *Work flows in Life Science*
3 Joost Geurts (CWI) *A Document Engineering Model and Processing Framework for Multimedia documents*
4 Olga Kulyk (UT) *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*
5 Claudia Hauff (UT) *Predicting the Effectiveness of Queries and Retrieval Systems*
6 Sander Bakkes (UvT) *Rapid Adaptation of Video Game AI*
7 Wim Fikkert (UT) *Gesture interaction at a Distance*
8 Krzysztof Siewicz (UL) *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments*
9 Hugo Kielman (UL) *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging*
10 Rebecca Ong (UL) *Mobile Communication and Protection of Children*
11 Adriaan Ter Mors (TUD) *The world according to MARP: Multi-Agent Route Planning*
12 Susan van den Braak (UU) *Sensemaking software for crime analysis*
13 Gianluigi Folino (RUN) *High Performance Data Mining using Bio-inspired techniques*
14 Sander van Splunter (VUA) *Automated Web Service Reconfiguration*
15 Lianne Bodenstaff (UT) *Managing Dependency Relations in Inter-Organizational Models*
16 Sicco Verwer (TUD) *Efficient Identification of Timed Automata, theory and practice*
17 Spyros Kotoulas (VUA) *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications*
18 Charlotte Gerritsen (VUA) *Caught in the Act: Investigating Crime by Agent-Based Simulation*
19 Henriette Cramer (UvA) *People's Responses to Autonomous and Adaptive Systems*
20 Ivo Swartjes (UT) *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative*
21 Harold van Heerde (UT) *Privacy-aware data management by means of data degradation*
22 Michiel Hildebrand (CWI) *End-user Support for Access to Heterogeneous Linked Data*
23 Bas Steunebrink (UU) *The Logical Structure of Emotions*
24 Zulfiqar Ali Memon (VUA) *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective*
25 Ying Zhang (CWI) *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines*
26 Marten Voulon (UL) *Automatisch contracteren*
27 Arne Koopman (UU) *Characteristic Relational Patterns*
28 Stratos Idreos (CWI) *Database Cracking: Towards Auto-tuning Database Kernels*
29 Marieke van Erp (UvT) *Accessing Natural History: Discoveries in data cleaning, structuring, and retrieval*
30 Victor de Boer (UvA) *Ontology Enrichment from Heterogeneous Sources on the Web*
31 Marcel Hiel (UvT) *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems*
32 Robin Aly (UT) *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval*
33 Teduh Dirgahayu (UT) *Interaction Design in Service Compositions*
34 Dolf Trieschnigg (UT) *Proof of Concept: Concept-based Biomedical Information Retrieval*
35 Jose Janssen (OU) *Paving the Way for Lifelong Learning: Facilitating competence development through a learning path specification*
36 Niels Lohmann (TUe) *Correctness of services and their composition*
37 Dirk Fahland (TUe) *From Scenarios to components*
38 Ghazanfar Farooq Siddiqui (VUA) *Integrative modeling of emotions in virtual agents*
39 Mark van Assem (VUA) *Converting and Integrating Vocabularies for the Semantic Web*
40 Guillaume Chaslot (UM) *Monte-Carlo Tree Search*
41 Sybren de Kinderen (VUA) *Needs-driven service bundling in a multi-supplier setting: the computational e3-service approach*
42 Peter van Kranenburg (UU) *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*
43 Pieter Bellekens (TUe) *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain*
44 Vasilios Andrikopoulos (UvT) *A theory and model for the evolution of software services*
45 Vincent Pijpers (VUA) *e3alignment: Exploring Inter-Organizational Business-ICT Alignment*
46 Chen Li (UT) *Mining Process Model Variants: Challenges, Techniques, Examples*

47 Jahn-Takeshi Saito (UM) *Solving difficult game positions*
48 Bouke Huurnink (UvA) *Search in Audiovisual Broadcast Archives*
49 Alia Khairia Amin (CWI) *Understanding and supporting information seeking tasks in multiple sources*
50 Peter-Paul van Maanen (VUA) *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention*
51 Edgar Meij (UvA) *Combining Concepts and Language Models for Information Access*

2011

1 Botond Cseke (RUN) *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*
2 Nick Tinnemeier (UU) *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language*
3 Jan Martijn van der Werf (TUe) *Compositional Design and Verification of Component-Based Information Systems*
4 Hado van Hasselt (UU) *Insights in Reinforcement Learning: Formal analysis and empirical evaluation of temporal-difference*
5 Base van der Raadt (VUA) *Enterprise Architecture Coming of Age: Increasing the Performance of an Emerging Discipline*
6 Yiwen Wang (TUe) *Semantically-Enhanced Recommendations in Cultural Heritage*
7 Yujia Cao (UT) *Multimodal Information Presentation for High Load Human Computer Interaction*
8 Nieske Vergunst (UU) *BDI-based Generation of Robust Task-Oriented Dialogues*
9 Tim de Jong (OU) *Contextualised Mobile Media for Learning*
10 Bart Bogaert (UvT) *Cloud Content Contention*
11 Dhaval Vyas (UT) *Designing for Awareness: An Experience-focused HCI Perspective*
12 Carmen Bratosin (TUe) *Grid Architecture for Distributed Process Mining*
13 Xiaoyu Mao (UvT) *Airport under Control. Multiagent Scheduling for Airport Ground Handling*
14 Milan Lovric (EUR) *Behavioral Finance and Agent-Based Artificial Markets*
15 Marijn Koolen (UvA) *The Meaning of Structure: the Value of Link Evidence for Information Retrieval*
16 Maarten Schadd (UM) *Selective Search in Games of Different Complexity*
17 Jiyin He (UvA) *Exploring Topic Structure: Coherence, Diversity and Relatedness*
18 Mark Ponsen (UM) *Strategic Decision-Making in complex games*
19 Ellen Rusman (OU) *The Mind ' s Eye on Personal Profiles*
20 Qing Gu (VUA) *Guiding service-oriented software engineering: A view-based approach*
21 Linda Terlouw (TUD) *Modularization and Specification of Service-Oriented Systems*
22 Junte Zhang (UvA) *System Evaluation of Archival Description and Access*
23 Wouter Weerkamp (UvA) *Finding People and their Utterances in Social Media*
24 Herwin van Welbergen (UT) *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior*
25 Syed Waqar ul Qounain Jaffry (VUA) *Analysis and Validation of Models for Trust Dynamics*
26 Matthijs Aart Pontier (VUA) *Virtual Agents for Human Communication: Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots*
27 Aniel Bhulai (VUA) *Dynamic website optimization through autonomous management of design patterns*
28 Rianne Kaptein (UvA) *Effective Focused Retrieval by Exploiting Query Context and Document Structure*
29 Faisal Kamiran (TUe) *Discrimination-aware Classification*
30 Egon van den Broek (UT) *Affective Signal Processing (ASP): Unraveling the mystery of emotions*
31 Ludo Waltman (EUR) *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*
32 Nees-Jan van Eck (EUR) *Methodological Advances in Bibliometric Mapping of Science*
33 Tom van der Weide (UU) *Arguing to Motivate Decisions*
34 Paolo Turrini (UU) *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations*
35 Maaike Harbers (UU) *Explaining Agent Behavior in Virtual Training*
36 Erik van der Spek (UU) *Experiments in serious game design: a cognitive approach*
37 Adriana Burlutiu (RUN) *Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference*
38 Nyree Lemmens (UM) *Bee-inspired Distributed Optimization*
39 Joost Westra (UU) *Organizing Adaptation using Agents in Serious Games*
40 Viktor Clerc (VUA) *Architectural Knowledge Management in Global Software Development*
41 Luan Ibraimi (UT) *Cryptographically Enforced Distributed Data Access Control*
42 Michal Sindlar (UU) *Explaining Behavior through Mental State Attribution*
43 Henk van der Schuur (UU) *Process Improvement through Software Operation Knowledge*
44 Boris Reuderink (UT) *Robust Brain-Computer Interfaces*

45 Herman Stehouwer (UvT) *Statistical Language Models for Alternative Sequence Selection*
46 Beibei Hu (TUD) *Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work*
47 Azizi Bin Ab Aziz (VUA) *Exploring Computational Models for Intelligent Support of Persons with Depression*
48 Mark Ter Maat (UT) *Response Selection and Turn-taking for a Sensitive Artificial Listening Agent*
49 Andreea Niculescu (UT) *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality*

2012

1 Terry Kakeeto (UvT) *Relationship Marketing for SMEs in Uganda*
2 Muhammad Umair (VUA) *Adaptivity, emotion, and Rationality in Human and Ambient Agent Models*
3 Adam Vanya (VUA) *Supporting Architecture Evolution by Mining Software Repositories*
4 Jurriaan Souer (UU) *Development of Content Management System-based Web Applications*
5 Marijn Plomp (UU) *Maturing Interorganisational Information Systems*
6 Wolfgang Reinhardt (OU) *Awareness Support for Knowledge Workers in Research Networks*
7 Rianne van Lambalgen (VUA) *When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions*
8 Gerben de Vries (UvA) *Kernel Methods for Vessel Trajectories*
9 Ricardo Neisse (UT) *Trust and Privacy Management Support for Context-Aware Service Platforms*
10 David Smits (TUe) *Towards a Generic Distributed Adaptive Hypermedia Environment*
11 J. C. B. Rantham Prabhakara (TUe) *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics*
12 Kees van der Sluijs (TUe) *Model Driven Design and Data Integration in Semantic Web Information Systems*
13 Suleman Shahid (UvT) *Fun and Face: Exploring non-verbal expressions of emotion during playful interactions*
14 Evgeny Knutov (TUe) *Generic Adaptation Framework for Unifying Adaptive Web-based Systems*
15 Natalie van der Wal (VUA) *Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes*
16 Fiemke Both (VUA) *Helping people by understanding them: Ambient Agents supporting task execution and depression treatment*
17 Amal Elgammal (UvT) *Towards a Comprehensive Framework for Business Process Compliance*
18 Eltjo Poort (VUA) *Improving Solution Architecting Practices*
19 Helen Schonenberg (TUe) *What's Next? Operational Support for Business Process Execution*
20 Ali Bahramisharif (RUN) *Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing*
21 Roberto Cornacchia (TUD) *Querying Sparse Matrices for Information Retrieval*
22 Thijs Vis (UvT) *Intelligence, politie en veiligheidsdienst: verenigbare grootheden?*
23 Christian Muehl (UT) *Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction*
24 Laurens van der Werff (UT) *Evaluation of Noisy Transcripts for Spoken Document Retrieval*
25 Silja Eckartz (UT) *Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application*
26 Emile de Maat (UvA) *Making Sense of Legal Text*
27 Hayrettin Gurkok (UT) *Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games*
28 Nancy Pascall (UvT) *Engendering Technology Empowering Women*
29 Almer Tigelaar (UT) *Peer-to-Peer Information Retrieval*
30 Alina Pommeranz (TUD) *Designing Human-Centered Systems for Reflective Decision Making*
31 Emily Bagarukayo (RUN) *A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure*
32 Wietske Visser (TUD) *Qualitative multi-criteria preference representation and reasoning*
33 Rory Sie (OUN) *Coalitions in Cooperation Networks (COCOON)*
34 Pavol Jancura (RUN) *Evolutionary analysis in PPI networks and applications*
35 Evert Haasdijk (VUA) *Never Too Old To Learn: On-line Evolution of Controllers in Swarm- and Modular Robotics*
36 Denis Ssebugwawo (RUN) *Analysis and Evaluation of Collaborative Modeling Processes*
37 Agnes Nakakawa (RUN) *A Collaboration Process for Enterprise Architecture Creation*
38 Selmar Smit (VUA) *Parameter Tuning and Scientific Testing in Evolutionary Algorithms*
39 Hassan Fatemi (UT) *Risk-aware design of value and coordination networks*
40 Agus Gunawan (UvT) *Information Access for SMEs in Indonesia*

41 Sebastian Kelle (OU) *Game Design Patterns for Learning*
42 Dominique Verpoorten (OU) *Reflection Amplifiers in self-regulated Learning*
43 Anna Tordai (VUA) *On Combining Alignment Techniques*
44 Benedikt Kratz (UvT) *A Model and Language for Business-aware Transactions*
45 Simon Carter (UvA) *Exploration and Exploitation of Multilingual Data for Statistical Machine Translation*
46 Manos Tsagkias (UvA) *Mining Social Media: Tracking Content and Predicting Behavior*
47 Jorn Bakker (TUe) *Handling Abrupt Changes in Evolving Time-series Data*
48 Michael Kaisers (UM) *Learning against Learning: Evolutionary dynamics of reinforcement learning algorithms in strategic interactions*
49 Steven van Kervel (TUD) *Ontologogy driven Enterprise Information Systems Engineering*
50 Jeroen de Jong (TUD) *Heuristics in Dynamic Sceduling: a practical framework with a case study in elevator dispatching*

2013

1 Viorel Milea (EUR) *News Analytics for Financial Decision Support*
2 Erietta Liarou (CWI) *MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing*
3 Szymon Klarman (VUA) *Reasoning with Contexts in Description Logics*
4 Chetan Yadati (TUD) *Coordinating autonomous planning and scheduling*
5 Dulce Pumareja (UT) *Groupware Requirements Evolutions Patterns*
6 Romulo Goncalves (CWI) *The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience*
7 Giel van Lankveld (UvT) *Quantifying Individual Player Differences*
8 Robbert-Jan Merk (VUA) *Making enemies: cognitive modeling for opponent agents in fighter pilot simulators*
9 Fabio Gori (RUN) *Metagenomic Data Analysis: Computational Methods and Applications*
10 Jeewanie Jayasinghe Arachchige (UvT) *A Unified Modeling Framework for Service Design*
11 Evangelos Pournaras (TUD) *Multi-level Reconfigurable Self-organization in Overlay Services*
12 Marian Razavian (VUA) *Knowledge-driven Migration to Services*
13 Mohammad Safiri (UT) *Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly*
14 Jafar Tanha (UvA) *Ensemble Approaches to Semi-Supervised Learning Learning*
15 Daniel Hennes (UM) *Multiagent Learning: Dynamic Games and Applications*
16 Eric Kok (UU) *Exploring the practical benefits of argumentation in multi-agent deliberation*
17 Koen Kok (VUA) *The PowerMatcher: Smart Coordination for the Smart Electricity Grid*
18 Jeroen Janssens (UvT) *Outlier Selection and One-Class Classification*
19 Renze Steenhuizen (TUD) *Coordinated Multi-Agent Planning and Scheduling*
20 Katja Hofmann (UvA) *Fast and Reliable Online Learning to Rank for Information Retrieval*
21 Sander Wubben (UvT) *Text-to-text generation by monolingual machine translation*
22 Tom Claassen (RUN) *Causal Discovery and Logic*
23 Patricio de Alencar Silva (UvT) *Value Activity Monitoring*
24 Haitham Bou Ammar (UM) *Automated Transfer in Reinforcement Learning*
25 Agnieszka Anna Latoszek-Berendsen (UM) *Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System*
26 Alireza Zarghami (UT) *Architectural Support for Dynamic Homecare Service Provisioning*
27 Mohammad Huq (UT) *Inference-based Framework Managing Data Provenance*
28 Frans van der Sluis (UT) *When Complexity becomes Interesting: An Inquiry into the Information eXperience*
29 Iwan de Kok (UT) *Listening Heads*
30 Joyce Nakatumba (TUe) *Resource-Aware Business Process Management: Analysis and Support*
31 Dinh Khoa Nguyen (UvT) *Blueprint Model and Language for Engineering Cloud Applications*
32 Kamakshi Rajagopal (OUN) *Networking For Learning: The role of Networking in a Lifelong Learner's Professional Development*
33 Qi Gao (TUD) *User Modeling and Personalization in the Microblogging Sphere*
34 Kien Tjin-Kam-Jet (UT) *Distributed Deep Web Search*
35 Abdallah El Ali (UvA) *Minimal Mobile Human Computer Interaction*
36 Than Lam Hoang (TUe) *Pattern Mining in Data Streams*
37 Dirk Börner (OUN) *Ambient Learning Displays*
38 Eelco den Heijer (VUA) *Autonomous Evolutionary Art*
39 Joop de Jong (TUD) *A Method for Enterprise Ontology based Design of Enterprise Information Systems*
40 Pim Nijssen (UM) *Monte-Carlo Tree Search for Multi-Player Games*
41 Jochem Liem (UvA) *Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning*

42 Léon Planken (TUD) *Algorithms for Simple Temporal Reasoning*
43 Marc Bron (UvA) *Exploration and Contextualization through Interaction and Concepts*

2014

1 Nicola Barile (UU) *Studies in Learning Monotone Models from Data*
2 Fiona Tuliyano (RUN) *Combining System Dynamics with a Domain Modeling Method*
3 Sergio Raul Duarte Torres (UT) *Information Retrieval for Children: Search Behavior and Solutions*
4 Hanna Jochmann-Mannak (UT) *Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation*
5 Jurriaan van Reijsen (UU) *Knowledge Perspectives on Advancing Dynamic Capability*
6 Damian Tamburri (VUA) *Supporting Networked Software Development*
7 Arya Adriansyah (TUe) *Aligning Observed and Modeled Behavior*
8 Samur Araujo (TUD) *Data Integration over Distributed and Heterogeneous Data Endpoints*
9 Philip Jackson (UvT) *Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language*
10 Ivan Salvador Razo Zapata (VUA) *Service Value Networks*
11 Janneke van der Zwaan (TUD) *An Empathic Virtual Buddy for Social Support*
12 Willem van Willigen (VUA) *Look Ma, No Hands: Aspects of Autonomous Vehicle Control*
13 Arlette van Wissen (VUA) *Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains*
14 Yangyang Shi (TUD) *Language Models With Meta-information*
15 Natalya Mogles (VUA) *Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare*
16 Krystyna Milian (VUA) *Supporting trial recruitment and design by automatically interpreting eligibility criteria*
17 Kathrin Dentler (VUA) *Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability*
18 Mattijs Ghijsen (UvA) *Methods and Models for the Design and Study of Dynamic Agent Organizations*
19 Vinicius Ramos (TUe) *Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support*
20 Mena Habib (UT) *Named Entity Extraction and Disambiguation for Informal Text: The Missing Link*
21 Kassidy Clark (TUD) *Negotiation and Monitoring in Open Environments*
22 Marieke Peeters (UU) *Personalized Educational Games: Developing agent-supported scenario-based training*
23 Eleftherios Sidirourgos (UvA/CWI) *Space Efficient Indexes for the Big Data Era*
24 Davide Ceolin (VUA) *Trusting Semi-structured Web Data*
25 Martijn Lappenschaar (RUN) *New network models for the analysis of disease interaction*
26 Tim Baarslag (TUD) *What to Bid and When to Stop*
27 Rui Jorge Almeida (EUR) *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*
28 Anna Chmielowiec (VUA) *Decentralized k-Clique Matching*
29 Jaap Kabbedijk (UU) *Variability in Multi-Tenant Enterprise Software*
30 Peter de Cock (UvT) *Anticipating Criminal Behaviour*
31 Leo van Moergestel (UU) *Agent Technology in Agile Multiparallel Manufacturing and Product Support*
32 Naser Ayat (UvA) *On Entity Resolution in Probabilistic Data*
33 Tesfa Tegegne (RUN) *Service Discovery in eHealth*
34 Christina Manteli (VUA) *The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems*
35 Joost van Ooijen (UU) *Cognitive Agents in Virtual Worlds: A Middleware Design Approach*
36 Joos Buijs (TUe) *Flexible Evolutionary Algorithms for Mining Structured Process Models*
37 Maral Dadvar (UT) *Experts and Machines United Against Cyberbullying*
38 Danny Plass-Oude Bos (UT) *Making brain-computer interfaces better: improving usability through post-processing*
39 Jasmina Maric (UvT) *Web Communities, Immigration, and Social Capital*
40 Walter Omona (RUN) *A Framework for Knowledge Management Using ICT in Higher Education*
41 Frederic Hogenboom (EUR) *Automated Detection of Financial Events in News Text*
42 Carsten Eijckhof (CWI/TUD) *Contextual Multidimensional Relevance Models*
43 Kevin Vlaanderen (UU) *Supporting Process Improvement using Method Increments*
44 Paulien Meesters (UvT) *Intelligent Blauw: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden*
45 Birgit Schmitz (OUN) *Mobile Games for Learning: A Pattern-Based Approach*
46 Ke Tao (TUD) *Social Web Data Analytics: Relevance, Redundancy, Diversity*

47 Shangsong Liang (UvA) *Fusion and Diversification in Information Retrieval*

2015

1 Niels Netten (UvA) *Machine Learning for Relevance of Information in Crisis Response*
2 Faiza Bukhsh (UvT) *Smart auditing: Innovative Compliance Checking in Customs Controls*
3 Twan van Laarhoven (RUN) *Machine learning for network data*
4 Howard Spoelstra (OUN) *Collaborations in Open Learning Environments*
5 Christoph Bösch (UT) *Cryptographically Enforced Search Pattern Hiding*
6 Farideh Heidari (TUD) *Business Process Quality Computation: Computing Non-Functional Requirements to Improve Business Processes*
7 Maria-Hendrike Peetz (UvA) *Time-Aware Online Reputation Analysis*
8 Jie Jiang (TUD) *Organizational Compliance: An agent-based model for designing and evaluating organizational interactions*
9 Randy Klaassen (UT) *HCI Perspectives on Behavior Change Support Systems*
10 Henry Hermans (OUN) *OpenU: design of an integrated system to support lifelong learning*
11 Yongming Luo (TUe) *Designing algorithms for big graph datasets: A study of computing bisimulation and joins*
12 Julie M. Birkholz (VUA) *Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks*
13 Giuseppe Procaccianti (VUA) *Energy-Efficient Software*
14 Bart van Straalen (UT) *A cognitive approach to modeling bad news conversations*
15 Klaas Andries de Graaf (VUA) *Ontology-based Software Architecture Documentation*
16 Changyun Wei (UT) *Cognitive Coordination for Cooperative Multi-Robot Teamwork*
17 André van Cleeff (UT) *Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs*
18 Holger Pirk (CWI) *Waste Not, Want Not!: Managing Relational Data in Asymmetric Memories*
19 Bernardo Tabuenca (OUN) *Ubiquitous Technology for Lifelong Learners*
20 Loïs Vanhée (UU) *Using Culture and Values to Support Flexible Coordination*
21 Sibren Fetter (OUN) *Using Peer-Support to Expand and Stabilize Online Learning*
22 Zhemin Zhu (UT) *Co-occurrence Rate Networks*
23 Luit Gazendam (VUA) *Cataloguer Support in Cultural Heritage*
24 Richard Berendsen (UvA) *Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation*
25 Steven Woudenberg (UU) *Bayesian Tools for Early Disease Detection*
26 Alexander Hogenboom (EUR) *Sentiment Analysis of Text Guided by Semantics and Structure*
27 Sándor Héman (CWI) *Updating compressed colomn stores*
28 Janet Bagorogoza (TiU) *KNOWLEDGE MANAGEMENT AND HIGH PERFORMANCE: The Uganda Financial Institutions Model for HPO*
29 Hendrik Baier (UM) *Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains*
30 Kiavash Bahreini (OU) *Real-time Multimodal Emotion Recognition in E-Learning*
31 Yakup Koç (TUD) *On the robustness of Power Grids*
32 Jerome Gard (UL) *Corporate Venture Management in SMEs*
33 Frederik Schadd (TUD) *Ontology Mapping with Auxiliary Resources*
34 Victor de Graaf (UT) *Gesocial Recommender Systems*
35 Jungxao Xu (TUD) *Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction*

2016

1 Syed Saiden Abbas (RUN) *Recognition of Shapes by Humans and Machines*
2 Michiel Christiaan Meulendijk (UU) *Optimizing medication reviews through decision support: prescribing a better pill to swallow*
3 Maya Sappelli (RUN) *Knowledge Work in Context: User Centered Knowledge Worker Support*
4 Laurens Rietveld (VU) *Publishing and Consuming Linked Data*
5 Evgeny Sherkhonov (UvA) *Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers*
6 Michel Wilson (TUD) *Robust scheduling in an uncertain environment*
7 Jeroen de Man (VU) *Measuring and modeling negative emotions for virtual training*
8 Matje van de Camp (TiU) *A Link to the Past: Constructing Historical Social Networks from Unstructured Data*
9 Archana Nottamkandath (VU) *Trusting Crowdsourced Information on Cultural Artefacts*
10 George Karafotias (VUA) *Parameter Control for Evolutionary Algorithms*
11 Anne Schuth (UvA) *Search Engines that Learn from Their Users*
12 Max Knobbout (UU) *Logics for Modelling and Verifying Normative Multi-Agent Systems*
13 Nana Baah Gyan (VU) *The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach*
14 Ravi Khadka (UU) *Revisiting Legacy Software System Modernization*

15 Steffen Michels (RUN) *Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments*
16 Guangliang Li (UvA) *Socially Intelligent Autonomous Agents that Learn from Human Reward*
17 Berend Weel (VU) *Towards Embodied Evolution of Robot Organisms*
18 Albert Meroño Peñuela (VU) *Refining Statistical Data on the Web*
19 Julia Efremova (Tu/e) *Mining Social Structures from Genealogical Data*
20 Daan Odijk (UvA) *Context & Semantics in News & Web Search*
21 Alejandro Moreno Clleri (UT) *From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground*
22 Grace Lewis (VU) *Software Architecture Strategies for Cyber-Foraging Systems*
23 Fei Cai (UvA) *Query Auto Completion in Information Retrieval*
24 Brend Wanders (UT) *Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach*
25 Julia Kiseleva (TU/e) *Using Contextual Information to Understand Searching and Browsing Behavior*
26 Dilhan Thilakarathne (VU) *In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains*
27 Wen Li (TUD) *Understanding Geo-spatial Information on Social Media*
28 Mingxin Zhang (TUD) *Large-scale Agent-based Social Simulation - A study on epidemic prediction and control*
29 Nicolas Höning (TUD) *Peak reduction in decentralised electricity systems -Markets and prices for flexible planning*
30 Ruud Mattheij (UvT) *The Eyes Have It*
31 Mohammad Khelghati (UT) *Deep web content monitoring*
32 Eelco Vriezekolk (UT) *Assessing Telecommunication Service Availability Risks for Crisis Organisations*
33 Peter Bloem (UVA) *Single Sample Statistics, exercises in learning from just one example*
34 Dennis Schunselaar (TUE) *Configurable Process Trees: Elicitation, Analysis, and Enactment*
35 Zhaochun Ren (UvA) *Monitoring Social Media: Summarization, Classification and Recommendation*