Report on the SIGIR 2017 Workshop on eCommerce (ECOM17)

Jon Degenhardt

eBay inc., USA

Surya Kallumadi

Kansas State University, USA

Yiu-Chang Lin

Rakuten, USA

Maarten de Rijke

University of Amsterdam, The Netherlands

Luo Si

Alibaba Group Inc USA, China

Andrew Trotman

University of Otago, New Zealand

Sindhuja Venkatesh

Amazon, USA

Xu Yinghui

Alibaba Group Inc USA, China

Abstract

The SIGIR 2017 Workshop on eCommerce (ECOM17), was a full day workshop that took place on Friday, August 11, 2017 in Tokyo, Japan. The purpose of the workshop was to serve as a platform for publication and discussion of Information Retrieval and NLP research and their applications in the domain of eCommerce. The workshop program was designed to bring together practitioners and researchers from academia and industry to discuss the challenges and approaches to product search and recommendation in the eCommerce domain. Another goal of the workshop was to examine the building of a benchmark data set to facilitate research into this topic. The workshop drew contributions from both industry as well as academia, in total the workshop received a total of twenty one submissions, and accepted thirteen papers. In addition to presentation of a subset of accepted submissions, the workshop had two keynotes by invited speakers from the industry, a poster session where all the accepted submissions were presented, a breakout session, a panel discussion, and a group discussion.

1   Introduction

eCommerce Information Retrieval has received little attention in the academic literature, yet it is an essential component of some of the largest web sites (such as eBay, Amazon, Airbnb, Alibaba, Taobao, Target, Facebook, and others). SIGIR has for several years seen sponsorship from these kinds of organisations, who clearly value the importance of research into Information Retrieval. The SIGIR 2017 Workshop on eCommerce (ECOM17)[1] brought

---

[1] Workshop website: http://sigir-ecom.weebly.com/

together researchers and practitioners of eCommerce IR to discuss topics unique to it, to set a research agenda going forward, and to examine how to build a data set for research. Our primary motivation as organizers of this workshop was to create a community and act as a forum to discuss interesting research challenges in the eCommerce domain.

Search, ranking, and recommendation have applications ranging from traditional web search to document databases to vertical search systems. This workshop explored approaches for search and recommendations of products. Although the task is the same as web-page search (fulfill a users information need), the way in which this is achieved is very much different. On product sites (such as eBay, Flipkart, Amazon, and Alibaba), the traditional web-page ranking features are either not present or are present in a different form. The entities that need to be discovered (the information that fulfills the need) might be unstructured, associated with structure, semi-structured, or have facets such as: price, ratings, title, description, seller location, and so on.

Domains with such facets raise interesting research challenges such as a) relevance and ranking functions that take into account the tradeoffs across various facets with respect to the input query b) recommendations based on entity similarity c) recommendations based on user location (e.g. shipping cost), and so on. In the case of eCommerce IR these challenges require inherent understanding of product attributes, user behavior, and the query context. Product sites are characterized by the presence of a dynamic inventory with a high rate of change and turnover, and a long tail of query distribution.

Outside of search but still within Information Retrieval, the same feature in different domains can have radically different meaning. For example, in email filtering the presence of Ray-Ban along with a price is a strong indication of spam, but within an auction setting this likely indicates a valid product for sale. Another example is natural language translation; company names, product names, and even product descriptions do not translate well with existing tools. Similar problems exist with knowledge graphs that are not customised to match the product domain.

The workshop also examined the problem of data availability. As the purpose of a product site is to make data on entities available, the same security concerns that plague other search domains may not exist. However sales and seller information is private and proprietary and likely to be unavailable. We hope that the continuing discussion on data will result in both a proposal to release data by an eCommerce company, as well as some tasks that can be examined on that data set.

We invited submissions that included the following topics:

- Machine learning techniques such as online learning and deep learning for eCommerce applications

- Semantic representation for users, products and services & Semantic understanding of queries

- Structured data and faceted search, converting unstructured data to its structured form

- The use of domain specific facets in search and other IR tasks, and how those facets are chosen

- Temporal dynamics for Search and Recommendation

- Models for relevance and ranking for multi-faceted entities

- Deterministic (and other) sorting of results lists (e.g. price low to high including postage)

- Personalized search and recommendations

- Inventory display issues (for example: legal, ethical, and spam)

- Cold start issues

- Personalization and the use of personal facets such as age, gender, location etc.

- Indexing and search in a rapidly changing environment (for example, an auction site)

- Scalability

- Diversity in product search and recommendations

- Strategies for resolving extremely low (or no) recall queries

- Query intent

- The use of external features such as reviews and ratings in ranking

- User interfaces and personalization

- Reviews and sentiment analysis

- The use of social signals in ranking and beyond

- The balance between business requirements and user requirements (revenue vs relevance)

- Trust

- Live experimentation Desktop and mobile issues

- Questions and answering, chatbots for eCommerce

All submissions were reviewed single blind and each submission was reviewed by at least 3 reviewers.

2   Proceedings

We had a broad mix of invited talks, paper presentations, breakout session, group discussion, and a panel discussion with active contribution and participation from both industry and academia[2]. A diverse set of four papers were presented orally, and we organized a poster session where all accepted papers (including the four) were presented. The purpose of the poster session was to allow for more interactions between the workshop attendees and the authors.

2.1   Invited Talks

The first invited keynote was presented by Mohit Kumar, who is a principal data scientist at Flipkart, an indian eCommerce company. Mohit's talk was titled *"E-Commerce for a billion people: challenges and opportunities"*. Mohit heads the query understanding and search ranking efforts at Flipkart. In his keynote, Mohit discussed the challenges for search in the Indian eCommerce context, and highlight 4 key patterns:

- Diversity of the user base with particular shift towards tier-2/3 cities

---

[2]Workshop program: http://sigir-ecom.weebly.com/program.html

- Diversity/long tail of products particularly in lifestyle (like non branded women's wear, electronic accessories)

- Shift of traffic from desktop to mobile

- Less reliable/low internet connectivity

He then discussed the challenges that are encountered due to these patterns throughout the entire life cycle of query processing in eCommerce ranging from AutoSuggest/Query Solicitation to Ranking as well as overall search experience through the interface design.

The second invited keynote was presented by Xuemei Gu, who is the head of Alibaba Group's ads business unit – Alimama. Xuemei's talk was titled *"Personalized shopping experience for 400M users: machine learning and data mining in e-commerce"*. In her keynote, Xumei discussed how Alibaba understands users and items, and optimizes user experiences and platform efficiency. She discussed Alibaba's recent work in the areas of recommendations and search, and presented interesting research topics and challenges to better serve personalized needs of the users in the eCommerce domain.

2.2   Papers and Poster Session

The workshop received twenty one submissions from both industry and academia, and accepted thirteen submissions (62%). The submissions were reviewed by an international program committee of experts in the field formed from representatives of several eCommerce companies and academic institutions. Each submission was reviewed by at least three reviewers. Of these 13 accepted submissions, four papers were chosen for oral presented and all 13 were presented in a poster session.

The first paper presented was by Trotman et al., *"The Architecture of eBay Search"* [8] in which the authors presented the architecture of *Cassini* – eBay's search engine. The paper discusses certain domain specific attributes that are unique to eBay and describes how these attributes dictate the engineering choices made while building their search engine. The authors discuss various aspects of eBay search engine such as ranking and the indexing pipeline. Finally they discuss updates to sellers of items, a difficult and unique problem eBay faces.

The second paper was *"Predicting Sales from the Language of Product Descriptions"* by Pryzant et al. [2] In this paper the authors try to predict the purchasing behavior of consumers by looking at the textual descriptions of the items. The authors then show how they can use this to identify actionable phrases that will help drive their sales. The authors conclude that product descriptions that contain seasonal, polite, authoritative and informative phrases lead to better sales outcomes.

The third paper was *"Document Reordering is Good, Especially for e-Commerce"* by Ramaswamy et al. [6] In this work the authors discuss the benefits of document-id reordering for eCommerce search engines. They present a document-id reordering approach based on item category that is applicable to eCommerce domains with structured and categorical data. This approach to reordering leads to a smaller index and reduced query latency.

The fourth paper chosen for presentation was *"A content-based recommender system for e-commerce offers and coupons"* by Xia et al. [12] in which the authors propose a content based recommender system that provides a personalized coupon recommendation to customers to increase click through rates and conversion rates.

A poster session with all the 13 accepted paper submissions [1–13] was organized during the lunch break. The poster session facilitated one-on-one interaction between the attendees and the authors and was well attended (indeed, drawn to a premature close as the next session started).

2.3 Breakout Session

We formed 4 groups for breakout session. The groups were a mix of industry and academia. Each of the groups were provided discussion guidelines. The groups were asked to discuss the following topics: collaboration between academia and industry, release of data sets by industry to facilitate academic research, and identify interesting problems and challenges in their domain.

Following are the reports from each of the breakout session groups:

Breakout Session – Group 1

Participants

The first group consisted of five people from industry and one from academia. Participants were from Flipkart, Rakuten, eBay, Schibested, Booking.com, and Kyoto University.

Breakout Focus

How can industry and academia improve collaboration in eCommerce search environments?

General Statement

Industry needs to identify conceptual problems not being addressed in a formal manner by academia. For eCommerce search, this means identifying challenges unique to the problem domain. Two examples: Non-relevance sorts like price low-to-high. This requires identifying appropriately matching search results independent of the ranking order. Query segmentation. Open question: Do product queries benefit from a different segmentation and analysis than informational queries?

ROI Projection

All industry members described a key issue limiting adoption of academia derived solutions: It is typically quite difficult to project business ROIs based on evaluation metrics reported in academic venues. Approaches requiring meaningful upfront investment often face significant barriers to adoption as a result.

The group discussed whether techniques involving efficiency (hardware cost) were less an issue, as there is a typically a more direct ROI calculation that can be made. Consensus was that though reliable ROI estimation is easier, businesses generally prioritize business growth metrics (market share, revenue, etc) ahead of hardware cost savings.

Privacy

Privacy was discussed as an area of possible collaboration, the idea being to engage academic research to improve protection of user privacy. This topic was introduced in the context of

online systems. However, it could potentially be used to protect privacy in the context of academic data set sharing. A possible related idea would be protection of company proprietary information, such as market information. Such ideas would simplify data set construction for academic sharing.

Additional Work Areas

There were a number of problem areas mentioned by industry members as having potential for academic collaboration. These included cross-market and cross-language product search, fraud detection, item-catalog matching, geo-location information utilization.

Breakout Session – Group 2

Participants

This group was a mix of people from industry and academia. From industry there were 3 participants from Yahoo Japan, 1 from Alibaba and Shopify each. From academia, the group had 2 students (Kansas State University and University of North Carolina).

Important Problems

- Ensuring item diversity in product search and diversity measurement.
- Identifying similar items and similar products. In the case of Shopify, which has multiple merchants, transfer learning is an important task, making the platform capable of learning from user behavior on one platform to improve the experience on another platform without compromising on privacy and data.
- Use of personalization and recommendation and balancing it with the long term sales generated from the user. Recommendation is a way of cannibalizing future sales. How does one balance the need to provide the most appropriate recommendation with the life cycle management of the user. Would it be okay to not provide recommendations if it results in the user viewing a lot more products?
- Serendipity in recommendations is something that contributes to delight. This is still an open research area.
- Use of reinforcement learning to improve sales.

Collaboration between Academia and Industry and Data Release

The participants from the industry were of a unanimous view that the primary goal of people in the industry is to improve the product, publishing papers is just a side effect. While, academia is oriented towards publishing papers. Industry is bottom up where as academia is top down in terms of approaching a problem. Everyone agreed that both the sides would want to collaborate, but there are multitude of reasons this does not happen:

- Creating a data set is too much effort and can take significant amount of resources away from building the product. Needs a dedicated person to curate and sanitise the data.
- Preserving anonymity of entities within the data and preventing privacy leaks. This again is a lot of work. The negative publicity of a privacy violation is not worth it.

- Companies like Shopify are too small to afford to take the time to generate the data sets.

- Fear that the data provided by the companies can be used to improve the products of their competitors. How does one ensure that this does not happen?

- Industry does not want to give away their trade secrets.

- Industry does contribute some times in the form of tech blogs.

- Alibaba would love to contribute compute resources to academia to foster better collaboration.

- Instead of releasing generic data sets, industry needs to release goal specific data sets, with a well defined problem and evaluation criteria.

Deep Learning Hype / Usefulness

Deep learning definitely is not a hype, it has proven to be useful and effective in a lot of applications. But there is a general overkill in the way it is being used for everything without considering its applicability. There are many traditional approaches that would be useful and appropriate. Interpretability and transparency of deep learning models is definitely an issue.

Breakout Session – Group 3

Participants

This group consisted of people mostly from the industry side including Rakuten, Yahoo Japan, Recruit Technology, Vipshop, Chicken Smoothie, etc. Each of the participants raised a challenging problem they faced. Several topics were discussed, ranging from search engine ranking quality, ranking models, evaluation metrics, to recommendation system and even user implicit/explicit feedback. The discussion is summarized as follows:

Data Set

Compared to most data sets being used in the academia, a common theme between different companies is that the data is usually on a very large scale, i.e., millions of users and products and they are always super dirty. As a result, most algorithms that we learned from or proposed by the academy do not necessarily apply to business, either due to the computation complexity or the accuracy.

Ranking

Do we want to rank our items according to relevancy or diversity? Or do we want to rank the most clicked item first or most purchased item first? An example is, for search query iphone 7, the most clicked item is the real iphone 7 product but the most purchased item is iphone 7 accessory (e.g. case or wire) because the former is much more expensive than the latter. This highly depends on the company's goal in the sense that whether the company wants to maximize the number of purchased items or the user experience.

Deep Learning Models

Nowadays deep learning models dominate the Natural Language Processing, Information Retrieval, and Machine Learning community and according to some group members experience, several companies have started adapting deep learning models. However, there are two issues. One is that it is always the case that we have to further fine-tune the models to apply them to each companys task setting and the improvement may not be as astounding as expected. Furthermore, how to put deep learning models into production is a big challenge as well.

Evaluation Metrics

In Information Retrieval, the mostly used evaluation metrics are MRR, MAP, NDCG, etc. However, do these metrics really contribute to high CTR or CVR? Usually they are not aligned with the business KPI either. How do we find a sweet spot between our algorithms and other business units' goal inside the companies? This topic is also closely related to the previously discussed ranking issue since the ranking algorithms are designed to maximize the evaluation metrics.

Breakout Session – Group 4

Participants

This group consisted of 8 people with representation from industry (Amazon, eBay, Rakuten etc.) and academia (University of Melbourne, University of Otago)

Problems Faced by Industry

Scale is a big problem i.e. in terms of finding the right item for the right person. Queries are not specific and well structured, for example *"4k tv"* and *"red shoes"*. Queries tend to get more abstract, as the customer expectations grow. Industry is faced with tricky choice of resorting to simple IR or conversational systems or guidance mechanisms to solve these issues. Conversational systems are an open research area, and if not implemented properly could cause friction with customer experience, and could lead to a loss of revenue. Ensuring catalog quality is another problem. eCommerce companies optimize for long term user engagement and satisfaction both short term and long term.

Data Set Availability

Big companies would not like to make data available publicly. Even if the data is anonymized as it is a big risk. Internship programs are a good way to get access to this data. Companies such as eBay do have active collaboration with academic researchers. Evaluation over any publicly released data sets is hard, and especially performing A/B tests to validate the results.

Breakout Session Closing Remarks

We have not discussed the economic and societal consequences of systems not working. Ecommerce is a primary source of livelihood for a lot of people and the consequences of, say, eBay shut down for a few days is enormous.

2.4 Panel Discussion

The topic of panel discussion was: *"eCommerce does not need academia"*. The purpose of the panel discussion was to identify ways to improve collaboration between the eCommerce industry and academia and to explore the reasons as to why a close relationship does not already exist. To this end, a mix of people were chosen to represent academia and industry; the panel discussion was moderated by Jimmy Lin(University of Waterloo), who has a history of working in both academia and industry. The other panelists were Ian Soboroff(N.I.S.T.), Chengxiang Zhai(U.I.U.C.) representing academia, and Roberto Konow(eBay inc.), Parikshit Sondhi(Walmart labs) representing the eCommerce industry.

The moderator started the discussion by asking the panelists to state a few success stories of tech transfer from academia to industry, approaches that are utilized in production environments. The panelists gave examples such as learning to rank, partitioning and compact data structures, click models, memex, NDCG, BM25, Dirichlet processes, entity search. The panelists and audience from industry were of the opinion that most academic papers do not work as they are supposed to work or are claimed to work. This is an issue of reproducibility, for example document reordering is known to work very well on .gov2, but not so well most elsewhere.

The discussion then moved to how industry can help academia by identifying research areas in eCommerce. Zhai was of the opinion that industry does not provide valuable feedback such as success / failure of reproducing academic work. A positive feedback would act as a motivation, while a negative feedback would point to cases where approaches fail and open up areas of research to explore.

Another issue discussed was the way academia formulates problems, as they tend to overcomplicate them, while the problem is framed in a much simpler way in an industry setting. But the general consensus of the panelists and the audience was that industry needs to identify the problems that they encounter and by doing this they will bring to light things that academic researchers haven't contemplated. Industry also needs to specify which of the academic papers are reproducible and which ones are not, with SIRIP and IRJ being good venues to publish such papers.

The next question asked by the moderator was: can industry survive if they do not publish, to which Soboroff responded that academia will not survive if they don't. Both Soboroff and Lin were of the opinion that industry needs to publish papers about how they solve problems, and that regular publication venues need to have such tracks. When ever its not possible, industry needs to cite the papers that they utilize without saying exactly what they did.

The next topic was on identifying aspects of academia that are not performed in the right manner.

- Zhai was of the opinion that instead of improving results in papers by a tiny fractions, academia needs to focus on the right problems, industry can help academia do this.

- In addition to introducing things that work, academia also needs to know what does not work and reasons for this. Academia will benefit from a negative results track at conferences.

- Soboroff brought up the issue of failure analysis being not properly performed by academia as they are not trained to do so and that this skill needs to be taught to students.

The next topic of discussion was about training students to become good researchers. Sondhi was of the opinion that details and knowledge about how eCommerce systems are built and how they are designed are not available to students in academia. Engineering oriented conferences usually deal with these issues of system architecture and design. Students need to start with simple ideas as the way academia defines success is convoluted, for academia success is dependent on the complexity of the solution. Students also need to have the requisite know how to build systems. The audience was of the opinion that academic research needs to be more disciplined and that they are publishing papers that are not useful outside academia. Lin was of the opinion that students need to learn about taste in research and interestingness of the problem. Soboroff added that TREC helps with experiment methodology and students who contribute to TREC are taught this and some students come back to create new TREC challenges.

The discussion then moved to the availability of data sets in the eCommerce domain. Zhai stated that academia generally has either a misunderstanding of the problem, or misunderstanding of the problem dataset, or is not sure of the exact data set to use. Academic researchers need to know what exists in the real world, and academia will benefit from industry papers. Industry need to show what real problems are and what the real constraints are. These insights will foster better research collaboration between academia and industry. The panelists in general were of the opinion that for new problems we need to have new data sets, and should not use old data sets tailored to solve new problems.

The varied background of the panelists and the audience generated an engaging discussion. The main takeaways from the panel discussion were that industry and academia have a symbiotic relationship and that both the sides need to have an increase in engagement to identify good ideas as well as better research areas, and we need new data sets to increase academic research in eCommerce.

3   Conclusions and Future Directions

The SIGIR 2017 Workshop on eCommerce (ECOM17) had a rich and diverse set of contributions and discussions ranging from eCommerce systems architecture to product search and recommendations to fraud detection. SIGIR was an excellent venue for the workshop – and we thank them for their assistance in running the workshop. This workshop was a first and a much needed forum for bringing together practitioners from industry and academia working in the eCommerce domain, and acted as a venue for eCommerce research. The need to have an active community to discuss problems and potential opportunities in the eCommerce domain came to the forefront in the group discussion and panel sessions. Availability of high quality data sets coupled with well defined problem statements is another aspect that would move the field forward. Future versions of this workshop might address this in the form of a data challenge. A number of participants from industry were willing to contribute data sets and identify interesting research areas. The workshop attendees were enthusiastic in their agreement that there should be future workshops on this topic.

References

[1] Damiano Spina, Maria Maistro, Yongli Ren, Sargol Sadeghi, Wilson Wong, Timothy Baldwin, Lawrence Cavedon, Alistair Moffat, Mark Sanderson, Falk Scholer, and Justin

Zobel. *Understanding User Behavior in Job and Talent Search: An Initial Investigation.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[2] Reid Pryzant, Young-Joo Chung, and Dan Jurafsky. *Predicting Sales from the Language of Product Descriptions.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[3] Haifa Alharthi, Diana Inkpen, and Stan Szpakowicz. *Unsupervised Topic Modelling in a Book Recommender System for New Users.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[4] Arthur Toth, Louis Tan, Giuseppe Di Fabbrizio, and Ankur Datta. *Predicting Shopping Behavior with Mixture of RNNs.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[5] Zehong Tan, Canran Xu, Mengjie Jiang, Hua Yang, and Xiaoyuan Wu. *Query Rewrite for Null and Low Search Results in eCommerce.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[6] Vishnusaran Ramaswamy, Roberto Konow, Andrew Trotman, Jon Degenhardt, and Nick Whyte. *Document Reordering is Good, Especially for e-Commerce.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[7] Yiu-Chang Lin, Giuseppe Di Fabbrizio, and Ankur Datta. *Query Segmentation via RNN Encoder Decoder Framework.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[8] Andrew Trotman, Jon Degenhardt, and Surya Kallumadi. *The Architecture of eBay Search.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[9] Rachana Nget, Yang Cao, and Masatoshi Yoshikawa. *How to Balance Privacy and Money through Pricing Mechanism in Personal Data Market.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[10] Haijing Liu, Pin Lv, Shiqiang Geng, Xueyu Duan, Xinwei Zhao, and Junxing Hu. *Argument based Online Deceptive Review Spam Detection.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[11] Amber Madvariya, and Sumit Borar. *Discovering Similar Products in Fashion E-commerce.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[12] Yandi Xia, Giuseppe Di Fabbrizio, Shikhar Vaibhav, and Ankur Datta. *A content-based recommender system for e-commerce offers and coupons.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.

[13] Ravindra Babu T., and Vishal Kakkar. *Address Fraud: Monkey Typed Address Classification for e-Commerce Applications.* In *Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)*, 2017.