Personalized Query Suggestion Diversification

Wanyu Chen
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China
wanyuchen@nudt.edu.cn

Fei Cai
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China
caifei@nudt.edu.cn

Honghui Chen
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China
chh0808@gmail.com

Maarten de Rijke
Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

ABSTRACT

Query suggestions help users refine their queries after they input an initial query. We consider the task of generating query suggestions that are personalized and diversified. We propose a personalized query suggestion diversification model (PQSD), where a user's long-term search behavior is injected into a basic greedy query suggestion diversification model (G-QSD) that considers a user's search context in their current session. Query aspects are identified through clicked documents based on the Open Directory Project (ODP). We quantify the improvement of PQSD over a state-of-the-art baseline using the AOL query log and show that it beats the baseline in terms of metrics used in query suggestion ranking and diversification. The experimental results show that PQSD achieves the best performance when only queries with clicked documents are taken as search context rather than all queries.

ACM Reference format:
Wanyu Chen, Fei Cai, Honghui Chen, and Maarten de Rijke. 2017. Personalized Query Suggestion Diversification. In *Proceedings of SIGIR '17, Shinjuku, Tokyo, Japan, August 07-11, 2017,* 4 pages.
DOI: http://dx.doi.org/10.1145/3077136.3080652

1 INTRODUCTION

Modern search engines offer query suggestions to help user formulate a good query and thus to get their intended search results to address their information needs. Previous work on query suggestion mainly focuses on recommending semantically related queries in response to a user's input query [14]. Such strategies cannot handle queries with uncertain search aspects, especially for different users with diverse search intents. Hence, diversification of query suggestions has been studied [11], where suggested queries try to cover multiple search aspects. In existing models for diversifying query suggestions, a user's personal information has hardly been explored. We combine the advantages of personalization and diversification and propose a Personalized Query Suggestion Diversification (PQSD) model, where personalization ensures that suggested queries are close to a user's specific search intent and diversification helps to generate multiple-aspect queries to increase the likelihood of a suggested query being clicked, that is helpful to diversifying web search results [10].

PQSD consists of two stages. In the first, we develop a greedy query suggestion diversification model (G-QSD) where a user's search context, e.g., queries and clicks, is used to generate a diversified ranked list of queries; to this end, we use co-occurrences as well as semantic similarity. In the second stage, we inject a user's long-term search behavior information into the G-QSD model using Bayes' rule. To determine a query's aspects, we collect clicked documents and extract descriptions of those documents based on the Open Directory Project (ODP).[1] Then, we use topic modeling [2] to obtain a topic distribution of document descriptions and queries.

We compare the performance of PQSD against a state-of-the-art query suggestion baseline on the AOL query log. The results show the effectiveness of PQSD in terms of query suggestion ranking and diversification. In particular, our PQSD model gains an improvement of around 1.35% and 6.39% in terms of MRR and α-nDCG, respectively, over a competitive baseline [11].

Our contributions are: (1) A model for personalized query suggestion diversification (PQSD) that incorporates a user's short-term search context in their current session and their long-term search history to detect search interests; (2) An analysis of the performance of PQSD under various search context selection strategies; PQSD yields better performance when the search context consists of queries with clicked documents rather than all queries.

2 APPROACH

2.1 Greedy query suggestion diversification

Our method for query suggestion diversification assumes that an initial list of query suggestion candidates R_I produced for the user

SIGIR '17, Shinjuku, Tokyo, Japan

978-1-4503-5022-8/17/08…$15.00
DOI: http://dx.doi.org/10.1145/3077136.3080652

[1] http://www.dmoz.org

query q_0 with length $|R_I| = L_I$ is given. We use a relevant term suggestion method [9] to generate this initial query ranking list.

We being by simplifying the problem of query suggestion diversification. The aim of query suggestion diversification is to satisfy the average user who enters the query q_0 by finding at least one acceptable query suggestion among the top N query suggestions returned. This can be achieved by maximizing the following function:

$$P(R_S \mid q_0, S_C) = 1 - \prod_{q_c \in R_S}(1 - P(q_c \mid q_0, S_C)), \quad (1)$$

where S_C denotes the search context in a given session of a user who inputs the initial query q_0 and R_S is a ranked list of queries that contains the top N query suggestion candidates to be returned. Obviously, we have $R_S \subseteq R_I$ with $|R_S| = N$, such that $N \leq L_I$.

Intuitively, the probability $P(q_c \mid q_0, S_C)$ in (1) denotes the likelihood that the suggested query candidate q_c satisfies a user who enters query q_0. With the assumption of query independence, the right-hand side of (1) denotes the probability that at least one query suggestion can satisfy the user. We further interpolate (1) at the aspect level and thus we have

$$P(R_S \mid q_0, S_C) = \sum_a \left(1 - \prod_{q_c \in R_S}(1 - P(q_c \mid q_0, a, S_C))\right), \quad (2)$$

where a ranges over possible aspects.

To maximize the objective in (2), we propose a natural greedy algorithm for generating a diverse ranking of query suggestions. We follow a greedy selection process as follows:

$$q^\star \leftarrow \operatorname*{arg\,max}_{q_c \in R_I \setminus R_S} \sum_a P(q_c \mid q_0, a, S_C) \prod_{q_s \in R_S} (1 - P(q_s \mid a, q_0, S_C)), \quad (3)$$

which guarantees that a suggested query that is the most different from previously selected query suggestions in R_S is selected at each step. Thus, it can minimize the redundancy of the ranked list of query suggestions by iteratively filling the list R_S until $|R_S| = N$.

The expression $P(q_c \mid q_0, a, S_C)$ in (3) is the probability that a query candidate q_c addresses the query aspect a given the input query q_0 and the session context S_C. We estimate this probability based on the following two parts, with a trade-off λ_1 ($0 \leq \lambda_1 \leq 1$) controlling the contribution of each part [16]:

$$P(q_c \mid q_0, a, S_C) \leftarrow \lambda_1 P(q_c \mid q_0) + (1 - \lambda_1) P(q_c \mid a, S_C). \quad (4)$$

Here, $P(q_c \mid q_0)$ denotes the probability that a suggested query q_c is relevant to the input query q_0, which can be estimated by the semantic similarity S_{q_0,q_c} between q_c and q_0, which is weighted by the normalized co-occurrence count C_{q_c,q_0} of q_c and q_0 in search sessions as: $P(q_c \mid q_0) \leftarrow C_{q_c,q_0} \cdot S_{q_0,q_c}$. Intuitively, a higher co-occurrence of two queries q_c and q_0 in search sessions would result in a higher relevance probability of q_c and q_0. Following [9], C_{q_c,q_0} can be estimated by

$$C_{q_c,q_0} = \frac{co_{q_c,q_0}}{f_{q_0} + f_{q_c} - co_{q_c,q_0}}, \quad (5)$$

where f_{q_0} and f_{q_c} denote the number of search sessions containing query q_0 and q_c, respectively; co_{q_c,q_0} indicates the number of search sessions containing both query q_c and q_0. For calculating S_{q_0,q_c}, we take the cosine similarity between two queries, represented by the average of the cosine similarity between query terms w returned by the word2vec model [15]:

$$S_{q_0,q_c} \leftarrow cos(q_0, q_c) = \frac{1}{W} \sum_{w_k \in q_0} \sum_{w_j \in q_c} \cos(w_k, w_j), \quad (6)$$

where $W = |q_0| \cdot |q_c|$ and $|q|$ is the number of terms in query q.

Turning to the right-hand side of (4), we make the query independence assumption [4] and decompose $P(q_c \mid a, S_C)$ to obtain:

$$P(q_c \mid q_0, a, S_C) \leftarrow \lambda_1 P(q_c, q_0) + (1 - \lambda_1) \prod_{q_t \in S_C} P(q_c \mid a, q_t). \quad (7)$$

The probability $P(q_c \mid a, q_t)$ in (7) can be estimated by the distance between query suggestion q_0 and query q_t in the search context given the aspect a. As queries that are submitted within a short temporal interval are bound to share common query aspects [4], we estimate the probability $P(q_c \mid a, q_t)$ as:

$$P(q_c \mid a, q_t) \leftarrow \theta_t \times \left(1 - \frac{|v_{q_c}(a) - v_{q_t}(a)|}{\sqrt{\sum_{i=1}^{M}(v_{q_c}(a_i) - v_{q_t}(a_i))^2}}\right), \quad (8)$$

where $\theta_t = \frac{1}{D(q_t)+1}$ and $D(q_t)$ refers to the interval between previous query q_t and the last query q_T in the search context S_C; M denotes the number of aspects of a query and $v_{q_c}(a_i)$ denotes the relevance of query q_c to its i-th aspect; see Section 3. This explains how the term $P(q_c \mid q_0, a, S_C)$ in (3) can be estimated.

Next, for calculating $P(q_s \mid q_0, a, S_C)$ in (3), which denotes the probability of query suggestions that have been chosen in the list R_R addressing query aspect a given the search context S_C and input query q_0, based on the query independence assumption we can simplify $P(q_s \mid q_0, a, S_C)$ in (3) as:

$$P(q_s \mid q_0, a, S_C) \leftarrow P(q_s \mid a, S_C) = \prod_{q_t \in S_C} P(q_s \mid a, q_t), \quad (9)$$

where $P(q_s \mid a, q_t)$ is computed analogously to $P(q_c \mid a, q_t)$ in (8).

2.2 Personalized query suggestion diversification

We generalize the greedy selection rule to a personalized version by considering a user u's long-term search history so that $q^\star$ becomes

$$\operatorname*{arg\,max}_{q_c \in R_I \setminus R_R} \sum_a P(q_c \mid q_0, a, S_C, u) \prod_{q_s \in R_R} (1 - P(q_s \mid a, q_0, S_C, u)). \quad (10)$$

For calculating $P(q_c \mid q_0, a, S_C, u)$, we use Bayes' rule:

$$P(q_c \mid q_0, a, S_C, u) = \frac{P(q_c) P(a, u, q_0, S_C \mid q_c)}{P(a, u, q_0, S_C)}. \quad (11)$$

We rewrite the term $P(a, u, q_0, S_C \mid q_c)$, which can be regarded as the combination of diversification and personalization, as:

$$P(a, u, q_0, S_C \mid q_c) \leftarrow \lambda_2 P(a, q_0, S_C \mid q_c) + (1 - \lambda_2) P(u, q_0, S_C \mid q_c),$$

where λ_2 is a tradeoff controlling the contributions of diversification and personalization. Based on Bayes' rule, $P(a, q_0, S_C \mid q_c)$ and $P(u, q_0, S_C \mid q_c)$ can be interpolated as

$$P(a, q_0, S_C \mid q_c) = \frac{P(q_c \mid a, q_0, S_C) P(a, q_0, S_C)}{P(q_c)} \quad (12)$$

and

$$P(u, q_0, S_C \mid q_c) = \frac{P(q_c \mid u, q_0, S_C) P(u, q_0, S_C)}{P(q_c)}, \quad (13)$$

respectively. The term $P(q_c \mid a, q_0, S_C)$ in (12) can be calculated following (7). Following the independence assumption used in web search [16], we approximate $P(q_c \mid u, q_0, S_C)$ in (13) as

$$P(q_c \mid u, q_0, S_C) \propto \prod_{q_t \in S_C} P(q_c \mid u) P(q_c \mid q_0) P(q_c \mid q_t), \quad (14)$$

where $P(q_c \mid u)$ denotes the probability of suggesting q_c to u according to their long-term search history and is estimated as:

$$P(q_c \mid u) \leftarrow |Q(u)|^{-1} \sum_{q \in Q(u)} S_{q_c, q}, \tag{15}$$

where $Q(u)$ are all queries submitted by user u; $|Q(u)|$ is the size of $Q(u)$; $S_{q_c,q}$ returns the semantic similarity between two queries like (6). Similarly, $P(q_s \mid a, q_0, S_C, u)$ in (10) can be estimated.

2.3 Generating query distribution over topics

In PQSD, a key problem is how to represent queries over topics. As queries are usually short, it makes sense to use clicked documents to generate their topic distribution rather than using the queries directly [4]. First, we extract a document description based on ODP and then generate the topic distribution of documents using Latent Dirichlet Allocation (LDA) [2]. After that, we obtain a query q's topic distribution as: $v_q = \sum_{d \in D(q)} v_d \times f(q, d)$, where $D(q)$ is the set of documents clicked in response to query q, v_d denotes the topic distribution of document d, which is vectorized using LDA, and $f(q, d)$ indicates the number of clicks on d after submitting q. For queries without clicked documents, we generate the query distribution from similar queries that have been vectorized as semantically related queries (or words) often express similar search topics [3]. We find the most similar vectorized query q_{label} for a query q_{nc} without clicks by $q_{label} \leftarrow \arg\max_{q_l \in Q_L} \cos(q_{nc}, q_l)$, where Q_L is a set of vectorized queries. We take the cosine similarity between two queries like (6).

3 EXPERIMENTS

Model summary. The baselines to be compared are (1) MMR: a query suggestion diversification approach based on Maximal Marginal Relevance (MMR) [5]; (2) DQS: a diversified query suggestion (DQS) model based on the query-URL bipartite graph analysis [11]. We consider four variations of the PQSD model that differ in the information used as search context for personalization: (1) PQSD$_{AL+AS}$ uess all queries in a user's long-term search history and in the current session; (2) PQSD$_{AL+CS}$ uses all queries in a user's long-term search history and only queries with clicks in the current session; (3) PQSD$_{CL+AS}$ uses only queries with clicks in a user's long-term search history and all preceding queries in the current session; (4) PQSD$_{CL+CS}$ uses only queries with clicks in a user's long-term search history and in the current session.

Research questions. (**RQ1**) Is the PQSD model able to beat state-of-the-art query suggestion models in terms of query suggestion ranking and diversification? (**RQ1**) What is the impact on the performance of PQSD of the choice of search context, i.e., choosing all queries or only queries with clicks?

Datasets and parameters. We use the AOL query log [12] in our experiments and preprocess the dataset following [8]. In addition, we split the queries into sessions by 30 minutes of inactivity and sessions with at least two queries are kept. Table 1 details the statistics of the dataset used.

For the parameter setup in our experiments, following [16], we fix $\lambda_1 = 0.5$. Regarding λ_2 in (12), we set $\lambda_2 = 0.5$ to give equal weight to diversification and personalization. In the LDA model, following [1], we set the number of topics to $M = 100$, and the distribution parameters $\alpha = 0.5$ and $\beta = 0.1$. We set the number of query suggestions to $N = 10$, which is commonly used [14].

Table 1: Dataset statistics.

Variables	Training	Test
# Queries	7,256,569	2,628,284
# Unique queries	746,796	373,397
# Sessions	1,428,962	714,481
# Users	220,946	110,473
Average # queries with clicks per session	4.37	4.35
Average # queries with clicks per user	28.87	28.91

For generating the ground truth, i.e., the relevance of a query q to an aspect a, we follow [6], and use a 5-grade scale (perfect = 4, excellent = 3, good = 2, fair = 1, and bad = 0) as: $rel_{q,a} \leftarrow \min(\lfloor v_q(a) \rfloor, 4)$. We use MRR [13] and α-nDCG [7] to measure the ranking and diversification performance of query suggestions.

4 RESULTS AND DISCUSSION

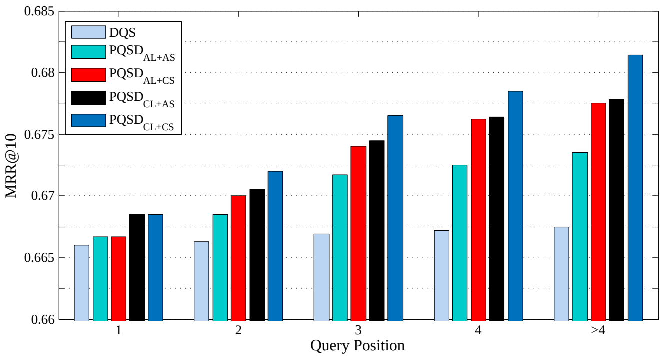
4.1 Performance of query suggestion models

To answer **RQ1**, we examine the query suggestion performance of the baselines as well as our PQSD models, which incorporate a user's search context for personalization. See Table 2 for the results. DQS achieves a better performance than MMR in terms of MRR@10 and α-nDCG@10. Hence, we only use DQS as a baseline from now on. DQS shows a minor MRR improvement against MMR (<1.0%) and a somewhat higher improvement in terms of α-nDCG@10 against MMR (<1.9%). As to the PQSD models, whatever type of search context we consider, PQSD outperforms the baseline, with MRR@10 improvements ranging from 0.8% to 2.0% and α-nDCG@10 improvements ranging from 4.3% to 8.9%. The fact that improvements in α-nDCG are higher than improvements in MRR can be explained by the fact that in some cases, redundant query suggestions ranked lower than the final submitted query are removed from the original query suggestion list; this does not affect the reciprocal rank score but does yield improved diversity scores.

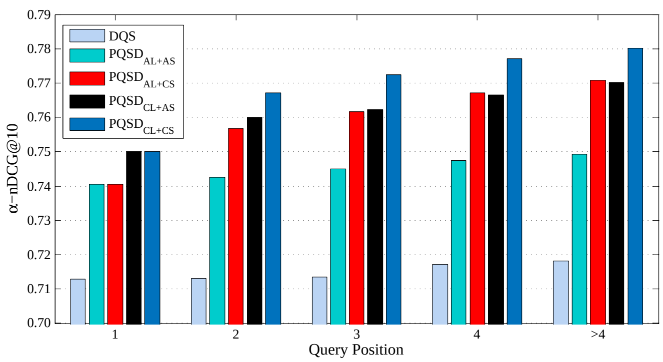
Table 2 shows that PQSD$_{CL+CS}$ achieves the best performance. Significant improvements against the baseline in terms of MRR@10 and α-nDCG@10 are observed for all PQSD models at the α = .01 level except for PQSD$_{AL+AS}$, for which we observe significant improvements at the α = .05 level. Hence, the content of the search context does affect the performance of PQSD model.

Table 2: Performance of query suggestion models. The results produced by the best baseline and the best performer in each column are underlined and boldfaced, respectively. Statistical significance of pairwise differences (PQSD models vs. best baseline) determined by a t-test ($^\blacktriangle$/$^\blacktriangledown$ for α = .01, or $^\vartriangle$/$^\triangledown$ for α = .05).

Models	MRR@10	α-nDCG@10
MMR	.6611	.7021
DQS	<u>.6672</u>	<u>.7152</u>
PQSD$_{AL+AS}$	.6726$^\vartriangle$	.7461$^\vartriangle$
PQSD$_{CL+AS}$	.6763$^\blacktriangle$	.7644$^\blacktriangle$
PQSD$_{AL+CS}$	.6756$^\blacktriangle$	.7686$^\blacktriangle$
PQSD$_{CL+CS}$	**.6807**$^\blacktriangle$	**.7791**$^\blacktriangle$



(a) Performance in terms of MRR@10.



(b) Performance in terms of α-nDCG@10.

Figure 1: Performance of PQSD models and the baseline at different query positions in a session.

4.2 Different personalization strategies

For **RQ2** we fix the search context by using either all previous queries or queries with clicks in the current session as well as the user's long-term search history. In general, PQSD achieves a better performance when it incorporates queries with clicks as search context than when using all previous queries. E.g., as shown in Table 2, PQSD_{CL+AS} beats PQSD_{AL+AS} in terms of both metrics. Similar results can be found when comparing PQSD_{CL+CS} to PQSD_{AL+CS}. Queries with clicks more accurately express a user's search intent, which is helpful for query suggestion personalization.

Results of the PQSD models and the baseline at the query position level (in a session) are shown in Fig. 1. As shown in Fig. 1a, as the search context becomes richer, the performance in terms of MRR@10 of query suggestion models improves too. E.g, at a late query position in a session (> 4), PQSD_{CL+CS} improves MRR@10 over earlier query positions (= 2). In addition, as indicated by the results of the PQSD models at the start of a session (query position = 1), when user's short-term search context in current session is unavailable, PQSD achieves neglible improvements over the baseline, especially for PQSD_{AL+AS} and PQSD_{AL+CS}.

Regarding the evaluation of diversity, similar results can be found in Fig. 1b when reporting the performance of query suggestion models in terms of α-nDCG@10. PQSD achieves relatively larger improvements against the baseline in terms of α-nDCG@10 than MRR@10 at each query position, which is consistent with the findings in Table 2. To sum up, search context consisting of queries with clicks, whether in a user's long-term or short-term search history, can help generate more accurate and diversified query suggestion rankings.

5 CONCLUSIONS AND FUTURE WORK

We have addressed the task of combining personalization and diversification of query suggestions and proposed a personalized query suggestion diversification model (PQSD) that is based on a basic greedy selection algorithm and incorporates a user's previous queries as search context for personalization. A variant of the PQSD model using queries with clicks achieves the best performance in terms of query ranking accuracy and diversification. As future work, we plan to evaluate our models on other datasets so as to verify the effectiveness of our proposal. In addition, we want to investigate the sensitivity of involved parameters, e.g., the cutoff number of query suggestion N and the tradeoff λ_2 controlling the contributions of personalization and diversification.

Acknowledgments. This research was supported by the National Advanced Research Project under no. 6141B08010101, Ahold Delhaize, Amsterdam Data Science, the Bloomberg Research Grant program, the Dutch national program COMMIT, Elsevier, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the Microsoft Research Ph.D. program, the Netherlands Institute for Sound and Vision, the Netherlands Organisation for Scientific Research (NWO) under project nrs 612.001.116, HOR-11-10, CI-14-25, 652.002.001, 612.001.551, 652.001.003, and Yandex. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES

[1] A. Asuncion, M. Welling, P. Smyth, and Y. W. Teh. On smoothing and inference for topic models. In *UAI*, pages 27–34, 2009.
[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(4):993–1022, 2003.
[3] D. Bollegala, Y. Matsuo, and M. Ishizuka. Measuring semantic similarity between words using web search engines. In *WWW*, pages 757–766, 2007.
[4] F. Cai, R. Reinanda, and M. de Rijke. Diversifying query auto-completion. *ACM Trans. Inf. Syst.*, 34(4):1–33, June 2016.
[5] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, pages 335–336, 1998.
[6] O. Chapelle, D. Metzler, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *CIKM*, pages 621–630, 2009.
[7] C. L. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Buttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR*, pages 659–666, 2008.
[8] J. Guo, X. Cheng, G. Xu, and X. Zhu. Intent-aware query similarity. In *CIKM*, pages 259–268, 2011.
[9] C.-K. Huang, L.-F. Chien, and Y.-J. Oyang. Relevant term suggestion in interactive web search based on contextual information in query session logs. *J. Am. Soc. Inf. Sci. Technol.*, 54(7):638–649, May 2003.
[10] S. Liang, F. Cai, Z. Ren, and M. de Rijke. Efficient structured learning for personalized diversification. *IEEE Trans. Knowl. Data Eng.*, 28(11):2958–2973, 2016.
[11] H. Ma, M. R. Lyu, and I. King. Diversifying query suggestion results. In *AAAI*, pages 1399–1404, 2010.
[12] G. Pass, A. Chowdhury, and C. Torgeson. A picture of search. In *InfoScale '06*, pages 1–7, 2006.
[13] C. Shah and W. B. Croft. Evaluating high accuracy retrieval techniques. In *SIGIR*, pages 2–9, 2004.
[14] Y. Song, D. Zhou, and L.-w. He. Post-ranking query suggestion by diversifying search results. In *SIGIR*, pages 815–824, 2011.
[15] M. Tomas, C. Kai, C. Greg, and D. Jeffrey. Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*, 2013.
[16] D. Vallet and P. Castells. Personalized diversification of search results. In *SIGIR*, pages 841–850, 2012.