

Query Dependent Time-Sensitive Ranking Model for Microblog Search^{*}

Shuxin Wang^{1,3}, Kai Lu², Xiao Lu^{1,3}, and Bin Wang^{3,4}

¹ University of Chinese Academy of Sciences, P.R. China

² School of Engineering, University of California, Santa Cruz, USA

³ Institute of Computing Technology, Chinese Academy of Sciences, P.R. China

⁴ Institute of Information Engineering, Chinese Academy of Sciences, P.R. China
{wangshuxin,luxiao,wangbin}@ict.ac.cn, kailu@soe.ucsc.edu

Abstract. Previous works show that one main difference between web search and microblog search is that most microblog queries are time-sensitive. Therefore, many existing works based on one straightforward temporal assumption have tried to incorporate the temporal factors into ranking model to improve the retrieval effectiveness. However, our study show that temporal role in ranking is complicated and hard to be summarized into one straightforward assumption. In addition, temporal influence is different among queries. To address these problems, we propose a query-dependent time-sensitive microblog ranking model, which use learning to rank to combine both temporal and entity evidences into the ranking process as the basic ranking model. In order to leverage the query difference, the k most similar training queries are used to train the ranking model. Experimental results on the public TrecMicroblog2011 data set show that comparing with the existing time-sensitive models, our models can significantly improve the performance of microblog search.

Keywords: Time-sensitive, Query dependent ranking, Microblog search.

1 Introduction

Microblog is a user-generated content system, which allows its users to publish and share short messages. Microblog search has become a hot research topic in recent years. Related works [1] show that microblog search queries are time sensitive. User's motivation of microblog search is time sensitive, whether event searches including keep up with what was happening and understanding trends or real-time information searches. Hence, incorporating temporal information in relevance judgement has been studied by many researchers. Based on different assumptions of temporal role played in ranking, different methods are proposed to incorporate temporal information to the ranking process [6, 7]. However, the

^{*} This work is supported by the National Science Foundation of China under Grant No. 61070111 and the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDA06030200.

temporal influence in ranking of microblog search is very complicated, and hard to be summarized into one simple straightforward assumption.

In this paper, we first analyze the temporal influence over ranking in microblog search, and come up with two important hypotheses. Then, in order to capture the complex pattern of the temporal influence, we propose to employ learning to rank framework to combine both temporal and entity evidences into ranking process. Experimental result shows that time-sensitive learning to rank model improves the retrieval effectiveness. Moreover, in order to leverage the query difference information, the k nearest neighbors with the most similar temporal distribution patterns are found to build a query dependent model. Experimental results show that the query dependent model discriminates different queries in the temporal influence, and achieve the best performance.

2 Related Works

Previous works have tried to incorporate temporal information into the ranking process. An important temporal assumption is how fresh the results are, also known as “the newer the document, the more important”, [6] and [8] tried to retrieve not only relevant but also fresh documents. [9] and [7] considered the other important temporal assumption, i.e., the event’s peak time point. [7] proposed a method to give more weight to these documents around peak points. [9] tried to combine query expansion on temporal variation with recency in order to improve the retrieval performance. However, these studies have strong assumption on the temporal influence over ranking, which in fact, the influence could be complex, and it’s inaccurate to be summarized into one simple assumption.

[10] proposed to use learning to rank for time-sensitive web search, however, there’re different specialties in microblog search, and most of their features are not applicable in the microblog search any more.

The problem of the large differences among different queries can be addressed by query dependent ranking approaches. There’re different loss functions to be employed to leverage the query difference to improve the retrieval effectiveness, such as query-classification-based approach [11], query clustering-based approach [12] and nearest neighbor-based approach [13]. As for web search, queries can be navigational, informational or transactional [15]. However, the queries in time-sensitive microblog search may have different patterns.

3 Our Methods

In this section, we first analyze the temporal distributions of query relevant documents and pseudo relevant documents, and come up with two hypotheses, which are the cornerstones of this paper, and then introduce our methods.

3.1 Query Temporal Distribution Analysis

We study the queries released by TREC Microblog Track 2011-2012. The corpus consists of microblogs between January 23rd and February 7th 2011, in total 17

days. Given a query, we want to find out the temporal distributions of relevant documents and pseudo relevant documents. Regard ground-truth labeled documents as relevant collection, and the top 500 documents retrieved by the query likelihood model as pseudo relevant collection¹. For a collection, the temporal distribution is computed as follows: Firstly, we divide the relevant documents by different days, then we count the ratio of documents fallen into each day.

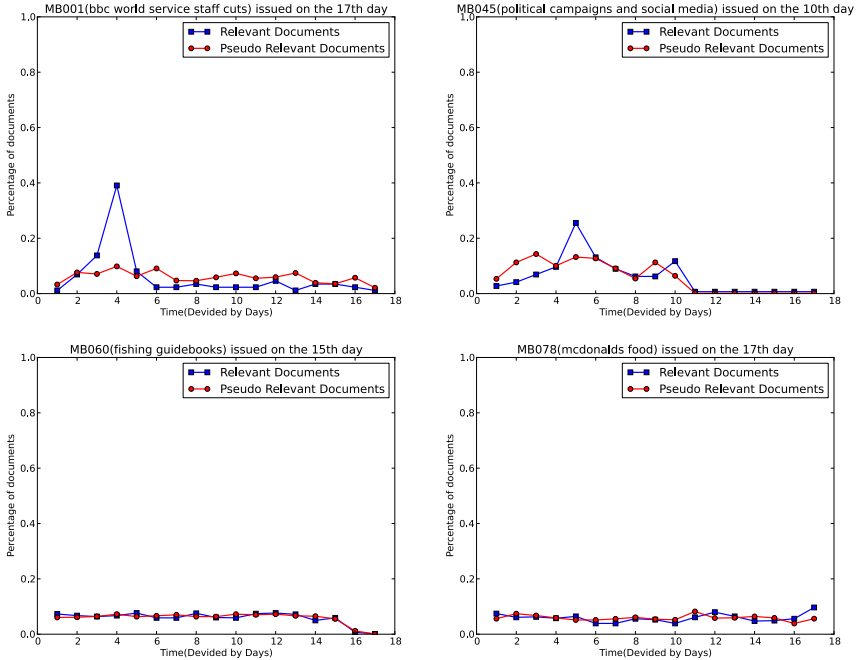


Fig. 1. Temporal Distribution of Queries

To demonstrate the temporal property of microblog, we take four queries used in the TREC 2011 Microblog Track as examples. Temporal distributions are plotted in the Fig. 1 using the method mentioned above.

To interpret this figure, we need to pay attention to the trends of pseudo relevant documents comparing with the relevant documents. We can see in this figure every subfigure’s trend of pseudo relevant documents coordinate with relevant documents curve respectively. This phenomenon implies that the temporal feature between pseudo relevant is quite similar. Its obvious that queries may vary in temporal distribution.

¹ For the real time ad-hoc queries, the documents returned newer than the time when query is issued are removed.

When the curve of relevant documents is higher than that of the pseudo relevant documents, this means the ranking model should give more weight to the documents at that point. From the figure, we can see every subfigure implies different weighting schema. For query MB001, the document around peak point should be weight more, While for query MB078, the newer document should be weighed more. The situation of MB060 and MB045 are complex situations.

Consider if a query's temporal distribution is similar to the query MB001, it implies the pattern of temporal influence over ranking is the same with MB001. To sum up:

Hypothesis 1. *The temporal features between relevant documents set and pseudo relevant documents set are quite similar. A query can be projected in the temporal feature space of its relevant or pseudo relevant documents set.*

Hypothesis 2. *Temporal influence over ranking is complicated and it's inaccurate to summarize into one straightforward assumption. The temporal influence is different among queries, and similar temporal distribution of query imply similar pattern of temporal influence.*

3.2 Query Temporal Distribution Dependent Ranking Method

Query Dependent Learning to Rank Model. It will be beneficial to combine temporal information into the learning to rank framework. We utilize a simple linear score function which can be adopted to the point-wise, pair-wise or list-wise learning to rank process. In this paper, we use listnet[14] as a basic learning to rank model, the typical list-wise learning to rank model whose loss function is defined by the probability distribution over permutations.

In[13], Geng et al. performed query-dependent ranking using K Nearest Neighbor method. In our study, according to Hypothesis 1, query can be projected into as a temporal distribution. Since the most similar temporal distribution implies the similar impact of temporal information in ranking, k nearest neighbors used to training ranking models would leverage the useful information of similar training queries, and avoid the negative effects from dissimilar ones. The similarity is computed by the distance of two queries' temporal distributions. The distance of two distributions can be computed by Kullback-Leibler Divergence(KLD). Given two arbitrary queries q_i and q_j , The KLD of two queries can be defined as:

$$KLD(q_i, q_j) = \sum_k^n \log \frac{P_{q_i}^k}{P_{q_j}^k}. \quad (1)$$

where $P_{q_i}^k$, $P_{q_j}^k$ indicate temporal distribution probability of query q_i , query q_j on k th time unit separately.

For queries in training set, the temporal distribution can be calculated by relevant documents with ground-truth labels. For queries in test set, according to Hypothesis 1, the pseudo relevant documents' distribution is quite similar to relevant documents' for a given query, therefore, pseudo relevant documents' temporal distribution are used to represent this query.

Features. As we know, the quality of features is vital to the learning to rank framework. We divide the features into two categories: the first one is entity related, the other one is temporal related.

Entity based features are aimed at measuring the textual similarity between a query and a document. Since our study focus on time-sensitive problem, we only choose representative entity features. The proposed features include qlSim, tfSim, hashTagSim, senLen and hashTagLen. The qlSim measures the query likelihood probability, which is calculated by $\log(p(Q|M_D))$. The tfSim measures the *tf-idf* similarity of query-document pair in vector space model. hashTagSim measures the similarity between query and hashtags, which is calculated by word Jaccard distance after stemmed. senLen is the length of tweet after removing the hashtags, hasTagLen is the length of hashtags. Among these features, hashTagSim, SenLen and hashTagLen can be computed offline, while the qlSim and tfSim were computed by the retrieval system (E.g., lemur).

Temporal features try to capture temporal influence over ranking. In our work, we extract two types of temporal features. The global features including *h*, *tc*, *ta* and *tp* indicate the feature extracted from whole collection. While the local features including *ti*, *tde*, *tai* and *tpi* denote the current document's characteristics.

h indicates whether the temporal distribution is concentrated or uniform. The probability distribution is calculated by the ratio of documents fallen into a time unit. $p(x)^i$ stands for the distribution probability of the *i*th time unit.

$$h(x) = - \sum_{i=0}^n p(x)^i \log p(x)^i. \quad (2)$$

ta indicates the average time of the whole collection. td_i stands for the time of document *i*.

$$ta = \frac{\sum_i^m td_i}{m}. \quad (3)$$

tp is the peak point of whole collection[7], is computed by finding the highest ratio in the distributions of the whole collection in different days. $c(i)$ stands for the count of documents on the *i*th day.

$$tp = \operatorname{argmax}_i |c(i)|. \quad (4)$$

tc indicates the newest time of the whole collection.

$$tc = \operatorname{argmax}_i td_i. \quad (5)$$

ti measures the time interval between the query and the current document. *tq* stands for query time.

$$ti = |tq - td_i|. \quad (6)$$

tde is exponential decay to indicate the time decay from the newest document's time stamp to the current document's time stamp. λ is parameter to be tuned.

$$tde = \lambda^{\lambda(tc - td_i)}. \quad (7)$$

The interval between the average time of pseudo relevant documents and the time of current document is defined as tai .

$$tai = |td_i - ta|. \quad (8)$$

The interval between the peak point and the current document's time stamp is defined as thi .

$$thi = |td_i - th|. \quad (9)$$

The combined feature could be meaningful in linear formed score function in learning to rank, this enlarges the dimension of feature space, and make the score function non-linear. We combine ti, tde, tmi, thi with h to a series of new features as $ti * h, tde * h, tai * h, thi * h$.

4 Experiments

To empirically evaluate our proposed hypotheses and approaches, we conduct a series of experiments on public data set TrecMicroblog2011 in this section.

4.1 Experimental Setup

The experimental corpus is distributed by TREC 2011 Microblog Track². After preprocessing(remove the mentions, url information, delete non-english tweets and repost tweets), the documents set size is 9,679,710. The query set contains 110 queries. The evaluation metrics include precision at 30 (P@30) and mean average precision(MAP), the same as in the TREC Microblog Track [2].

In our two approaches, the first one combines a series of our proposed features into the learning to rank framework, which is named as TLTR model. The other one is the query-dependent ranking model(QDLTR). We use the basic query likelihood model (QLM) as the baseline model, and also evaluate the methods proposed in [6](TLM) and [7] (HTLM-AdaptiveMultiML).

In our experiments, the QLM is smoothed by Jelinek-Mercery, and all the parameters are tuned using 5-fold cross validation one by one based on P@30 metric. The parameters are shown in Tab.1.

Table 1. Parameters in experiments of best performance

Model	Parameter	Description	Value
QL	λ	JK smooth parameter	0.4
TLM	λ	exponential decay parameter	0.3
HTLM-AdaptiveMultiML	α	threshold of peak point	0.1
	λ	exponential decay parameters	0.5
TLTR&QDTLTR	λ	exponential decay parameter	0.8

² <http://sites.google.com/site/microblogtrack/>

4.2 Experimental Results

The experimental results are shown in Tab.2. Statistical significant tests are performed using paired T-test at 0.05 level of significance. From the table, we can see that:

Table 2. Experimental Results(The significance of QDLTD compare with QLM is marked as *, and with TLTR is marked as ★)

<i>model</i>	<i>MAP</i>	<i>P@30</i>
QLM	0.2283	0.3019
TLM	0.2344	0.2972
HTLM-AdaptiveMultiML	0.2358	0.3062
TLTR	0.2415	0.3327
QDLTR(K=20) ³	0.2463**	0.3428**

As a baseline and benchmark of this evaluation, QLM model only measures textual similarity, and performs the worst on both metrics. Then TLM uses recency information as document’s prior which ranks the newer documents higher. The results reflect that TLM has a little improvement on the MAP comparing QLM. After leveraging the temporal specialty of microblog, where lots of queries are event-related and the documents around peak point are more favorable. HTLM-AdaptiveMultiML brings positively effect on the P@30 and MAP, while the improvement is still limited. This reflects that the queries in microblog search is time-sensitive and very complex, although incorporating the temporal feature can bring a little improvement, it is limited if we only rely a simple assumption.

Through incorporating a series of features including the temporal features and entity features to the learning to rank framework, we can see that TLTR effectively improves the performance on both of evaluation metrics compared with TLM and HTLM-AdaptiveMultiML. It shows without any assumption of temporal influence, learning to rank is powerful to combine temporal evidence into ranking process. Furthermore, we can see QDLTR shows the best performance on both of evaluation metrics. Comparing with the baseline models, it outperforms QLM by 7.88% on the MAP score. Moreover, it shows even more significant improvement on the P@30 score which increases by 13.55%. Through considering the differences among different queries, the QDLTR model significantly outperforms the other methods.

Through the experiments, we can see that by combining temporal information into ranking can effectively improve retrieval effectiveness, and by further considering the query dependent will have much better performance. Therefore, our two hypotheses are very meaningful in the time-sensitive microblog retrieval tasks.

³ K is chosen by 5-fold cross validation.

5 Conclusion and Future Work

In this paper, we propose a query dependent learning to rank model to address the time-sensitive problem in microblog search. First, we extract the temporal and entity features from microblogs, and use learning to rank framework to combine these diverse temporal evidences which are able to capture the temporal influence in ranking. Second, to leverage the query difference over time distribution, we employ k-nearest neighbor algorithm to find temporal patterns similar to given query, and use this model to rank the documents for a given query. Experimental results show that our models can significantly improve the retrieval effectiveness of microblog search.

As for future work, there're lots of words in microblog indicates time, such as "yesterday", "last week" and "Jan 5th". These temporal information can also be extracted as temporal evidence used for ranking.

References

1. Teevan, J., Ramage, D., Morris, M.R.: #TwitterSearch: a comparison of microblog search and web search, pp. 35–44 (2011)
2. Ounis, I., Macdonald, C., Lin, J.: Overview of the trec-2011 microblog track. In: Proceedings of the 20th Text REtrieval Conference (TREC 2011) (2011)
3. Metzler, D., Cai, C.: USC/ISI at TREC 2011: Microblog Track. In: TREC (2011)
4. Miyanishi, T., Okamura, N., Liu, X., Seki, K., Uehara, K.: TREC 2011 Microblog Track Experiments at Kobe University. In: TREC (2011)
5. Zhang, X., He, B., Luo, T., Li, B.: Query-biased learning to rank for real-time twitter search. Presented at the CIKM 2012: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, New York, New York, USA (October 2012)
6. Li, X., Croft, W.B.: Time-based language models, pp. 469–475 (2003)
7. Wei, B., Zhang, S., Li, R., Wang, B.: A Time-Aware Language Model for Microblog Retrieval, trec.nist.gov
8. Efron, M., Golovchinsky, G.: Estimation methods for ranking recent information. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 495–504 (2011)
9. Miyanishi, T., Seki, K., Uehara, K.: Combining recency and topic-dependent temporal variation for microblog search. In: Serdyukov, P., Braslavski, P., Kuznetsov, S.O., Kamps, J., Rüger, S., Agichtein, E., Segalovich, I., Yilmaz, E. (eds.) ECIR 2013. LNCS, vol. 7814, pp. 331–343. Springer, Heidelberg (2013)
10. Kanhabua, N., Ng, K.: Learning to rank search results for time-sensitive queries, pp. 2463–2466 (2012)
11. Kang, I.-H., Kim, G.: Query type classification for web document retrieval. ACM, New York (2003)
12. Banerjee, S., Dubey, A., Machchhar, J.: Efficient and accurate local learning for ranking. In: SIGIR Workshop: Learning to Rank for Information Retrieval (2009)
13. Geng, X., Liu, T.-Y., Qin, T., Arnold, A., Li, H., Shum, H.-Y.: Query dependent ranking using K-nearest neighbor. In: SIGIR 2008, pp. 115–122 (2008)
14. Cao, Z., Qin, T., Liu, T.Y., Tsai, M.F., Li, H.: Learning to rank: from pairwise approach to listwise approach. In: Proceedings of the 24th International Conference on Machine Learning (2007)
15. Broder, A.: A taxonomy of web search. ACM SIGIR Forum 36, 3–10 (2002)