An Exploration of Tie-Breaking for Microblog Retrieval

Yue Wang, Hao Wu, and Hui Fang

Department of Electrical & Computer Engineering,
University of Delaware, USA
{wangyue,haow,hfang}@udel.edu

Abstract. Microblog retrieval enables users to access relevant information from the huge number of tweets posted on social media. Since tweets are different from traditional documents, existing IR models might not be the optimal choice for this problem. Tie-breaking has been recently proposed as a new way of combining multiple retrieval signals. In this paper, we focus on studying the potential of this approach in microblog retrieval and propose new methods to further improve the performance. Experiment results show that these tie-breaking based methods can achieve comparable performance with the top runs in the TREC Microblog track.

1 Introduction

Microblog has become an important information source in daily life [1]. However, the huge number of tweets generated everyday makes it difficult for users to find useful information. Although traditional IR models could be applied to this domain-specific problem, they might not be the best choice because tweets are different from regular documents in many aspects. Existing studies on microblog retrieval have tried to overcome this limitation and incorporate new retrieval signals such as temporal information [2–5] and quality indicators [6, 7] by either extending existing models or using learning to rank methods.

Tie-breaking approach has been recently proposed as a new way of combining retrieval signals [8]. The basic idea is to prioritize retrieval signals, and then apply them one at a time to rank documents by breaking the ties created by previously applied signals. Its advantage lies in the simplicity and flexibility of combining multiple signals. Previous study showed the promising results on ad hoc retrieval by combining three basic retrieval signals, i.e., document length (DL), term frequency (TF) and inverse document frequency (IDF).

In this paper, we propose to extend the existing framework of tie-breaking to further improve its effectiveness on microblog retrieval. Specifically, we consider two commonly used retrieval strategies, i.e., query expansion and document expansion, and study how to incorporate them into the tie-breaking framework. Our experiments confirm that the tie-breaking approach is more effective than traditional retrieval models and learning to rank methods when combining multiple basic retrieval signals. Moreover, the proposed query expansion and document expansion methods can further improve the retrieval performance while the document expansion method is more effective. Finally, the performances of the proposed methods are comparable with that of top runs in TREC Microblog track.

2 Tie-breaking for Microblog Retrieval

2.1 Basic idea of tie-breaking

The effectiveness of a retrieval function is closely related to how it combines retrieval signals such as TF, IDF and DL. Traditional retrieval models including Okapi BM25 and pivoted normalization often combine these signals in a rather complicated way. However, it remains unclear whether such complicated combinations are necessary.

The basic idea of tie-breaking is to explore a simple yet effective way of combining multiple signals [8]. Specifically, it will first prioritize all the signals and then iteratively apply one signal at a time to rank documents by breaking the ties (i.e., the documents with the same scores) created by the previously applied signals. Figure 1 illustrates this process. Assuming that signal 1 is stronger than signal 2 and both of them are stronger than signal 3, we would then first apply signal 1 and generate a list of ranked documents. Note that the documents in the same row indicate that they receive the same scores. For example, after applying signal 1, d_5 and d_2 have the same relevance score, which is higher than the rest of documents. Similarly, d_3, d_7 and d_1 are also tied. We can then apply the next strongest signal (i.e., signal 2) to break these ties created by the signal 1, e.g., assign a higher score d_5 than d_2. This process can be repeated until all the signals are applied.

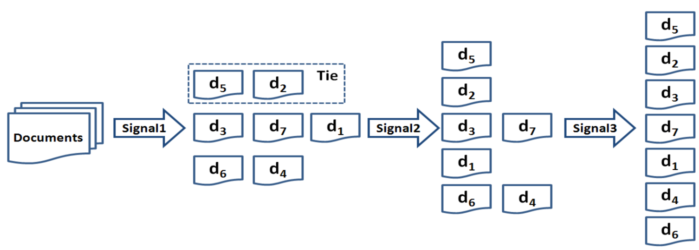

Fig. 1. Example scenario of tie-breaking

2.2 Applying tie-breaking to microblog retrieval

Experiment Setup: TREC Microblog track provides a common platform for the researchers to study the microblog retrieval problem. We conducted experiments on both the TREC Microblog 2011 and 2012 collections, i.e., **MB11** and **MB12**. We leverage the provided API [1] and crawl 10K results returned by API for each query. Due to the nature that the tweets retrieval is sensitive to the time, we build an index for each query in order to ensure that no tweets after the query time is involved. We applied the Krovetz stemming to the tweets, and no stop word is removed. The results are evaluated in terms of both mean average precision (MAP) and precision at 30 (P@30).

Methodology: The first step is to identify the retrieval signals. In additional to the three used in the previous study [8], we also consider the number of followers

[1] https://github.com/lintool/twitter-tools/wiki/TREC-2013-API-Specifications

Table 1. Comparing the strength of the best implementation for signals (MB11)

Retrieval signal	Implementation	$\mathbf{MAP_b}$	$\mathbf{MAP_w}$	$\mathbf{MAP_a}$
TF	$\frac{c}{c+1}$	0.3906	0.2306	0.2842
IDF	$log(1 + \frac{N}{df})$	0.4969	0.2886	0.3631
DL	$\frac{1}{dl}$	0.0319	0.0220	0.0234
NOF	$log(NOF)$	0.0747	0.0562	0.0642

(NOF) because it shows the endorsement from the online community. The more follower a user has, the more influential the user could be, which further means that the content the user posted could be useful and reliable [9]. We do not consider other signals such as the number of retweets because our data sets do not contain detailed information about them.

The second step is to select the best implementation for each signal and then prioritize the signals based on their best implementation. To estimate the strength of each implementation, we would rank documents based on the implementation. Since multiple documents may receive the same score, the retrieval results would be a set of possible rankings where the ties might be broken randomly. To accurately measure the strength of an implementation, we need to look at the best, worst and expected (or average) performance of the set, e.g., MAP_b, MAP_w and MAP_a. Following the previous study [8], the selection and prioritization are done based on the same set of criteria: (1) we prefer the implementation with greater potential (i.e., larger MAP_b) and (2) we prefer the implementation with the better expected performance (i.e., larger MAP_a).

We use **MB11** as the data set to select the implementation and prioritize the signals. We tried different implementation for each signal as in the previous study [8], but only reported the performance of the best implementation for each signal in Table 1. It is clear that IDF is the strongest signal because it has the largest MAP_b and MAP_a. Among the other three signals, TF is the strongest and DL is the weakest. Thus, we should apply the signals in the following order: $IDF \bigoplus TF \bigoplus NOF \bigoplus DL$. This method is denoted as **Tie-breaking**.

Experiment Results: We compare the **Tie-breaking** method with three methods: (1) **Okapi**: Okapi BM25 method; (2) **IDF**: Use only IDF for term weighting; and (3) **L2R**: AdaRank with the same signals shown in Table 1. The results are shown in Table 2. All the methods are trained on **MB11**, and tested on **MB12**. Note that the results with † and * indicate improvement over **Okapi**, **IDF** is statistically significant at 0.05 level based on Wilcoxon signed-rank test.

Table 2. Performance comparison of the different retrieval methods.

	Training (MB11)		**Testing (MB12)**	
Methods	**MAP**	**P@30**	**MAP**	**P@30**
Okapi	0.3238	0.3762	0.1922	0.3034
IDF-only	0.3548	0.4034	0.2092	0.3299
L2R	0.0748	0.0849	0.0411	0.0641
Tie-breaking	$\mathbf{0.3743}^\dagger$	$\mathbf{0.4204}^\dagger$	$\mathbf{0.2299}^{\dagger,\star}$	$\mathbf{0.3684}^{\dagger,\star}$

It is clear that **Tie-breaking** is more effective than the traditional retrieval functions including **Okapi** and **IDF**. Moreover, we find that it is more robust in combining simple retrieval signals than **L2R** and does not require the complicated features to work as existing learning to rank methods. The worse performance of L2R method may come from the use of simple features. In fact, when using more sophisticated features as the ones mentioned in the previous study [9], the performance of learning to rank would be comparable to our method. This further confirms that tie-breaking is a simple yet effective method of combining signals.

3 Extension of Tie-breaking Methods

So far, we have shown that tie-breaking is effective for microblog retrieval. Since tweets are often short, it is necessary to bridge the vocabulary gap between queries and documents. In this section, we explain how to extend existing tie-breaking framework to address this challenge.

3.1 Query expansion for tie-breaking

Query expansion is a common way to improve the performance. One most commonly used method is pseudo relevance feedback, which extract terms from initial retrieval results to enrich the query aiming to bring more relevant query terms. To extend the tie-breaking framework for pseudo relevance feedback, we propose the following strategy. We first conduct the tie-breaking method with all the signals to get a the results. From the initial results, we choose the top N terms with the highest term frequency from the top k ties as the expansion terms. Assume there are m documents in the top k ties. The weight of the expansion term q_{exp} and that of the original query term q_{orig} are computed as:

$$weight_{q_{exp}} = \alpha \cdot \frac{\sum_{i=1}^{m} tf_i(q_{exp})}{\sum_{i=1}^{m} dl_i}, \qquad weight_{q_{orig}} = (1-\alpha) \cdot \frac{qtf(q_{orig})}{query\_length}$$

where the $tf_i(q_{exp})$ denotes the term frequency of q_{exp} in document d_i, the dl_i denotes the length of document d_i, and the $qtf(q_{orig})$ denotes the query term frequency of the original query term q_{orig}. We then combine the original query terms and expanded query terms with the parameter α as a new query to retrieve again using tie breaking method. The α control the weight of the expansion terms. This method is denoted as **TB-PF-TB**.

3.2 Document expansion for tie-breaking

Document expansion is another common way to bridge the vocabulary gap of the document and the query. As the length of the microblogs are limited, users commonly insert a URL to the tweets, which links to the webpage that describes their ideas or expressions. This feature makes it reliable to follow the links as a way of expansion for the original tweets. We propose three methods to utilize the expanded documents in the tie-breaking framework.

Merge with original tweets: One straightforward solution is to merge the expanded document with the original tweets. With the merged document, we could apply the **Tie-breaking** method to retrieve the tweets with the same order. We refer this method as **Merged**. It is clear this method is easy to implement. However, the effectiveness of this method relies on the quality of the

Table 3. Performance of tie breaking with different expansion methods.

	Training(MB11)		Testing(MB12)	
Methods	**MAP**	**P@30**	**MAP**	**P@30**
Tie-Breaking	0.3743	0.4204	0.2299	0.3684
TB-PF-TB	0.3811	0.4401	0.2396	0.3844
Merged	0.2957	0.4109	0.1913	0.3557
Combined	**0.4182**‡	**0.4517**‡	**0.2550**‡	0.4069‡
All	0.4028‡	0.4361	0.2532‡	**0.4201**‡

expanded document. The noisy terms in the expanded document would result in the meaning of the tweet changes after we concatenate the documents.

Tie-breaking on two indexes: One possible way to overcome the limitation of the **Merged** method is to build a separate index for the expanded documents, and then perform tie-breaking on each index respectively. At last we could combine the results from the two indexes into one by using a normalized score or by the ranking. The advantage of this method is that we can control the effect of the expanded document to the original tweets. To be specific, we first applied the **Tie-breaking** on the original tweets index. We then utilize the signals in the order $IDF \bigoplus TF \bigoplus DL$ in the expansion index to perform another round of retrieval. Each document will get a score from both original tweets index and expansion index. Formally, assume we have k signals and document D will get a score $S_i(D)$ with i^{th} signal s_i. Then the score for D in either original or expanded index can be computed as:

$$S(D) = \sum_{i}^{k} S_i(D) \cdot w^{k-i}$$

where the w is a weight to control the difference of each level. w is set to 100 in our experiments because the relevant score of a document would not be greater than 100. At last, we applied a normalized combine method proposed in [10] to combine the scores from these two indexes. We refer it as **Combined**.

Tie-breaking with all signals: Another possible way of combining information from two indexes it to extract retrieval signals from the expanded documents, and then put them together with the signals in the original index to apply **Tie-breaking**. We utilized the TF, IDF, and DL signals from the expanded index as the supplement signals to the original signals. We applied these signals in the order $IDF_{orig} \bigoplus TF_{exp} \bigoplus TF_{orig} \bigoplus IDF_{exp} \bigoplus NOF \bigoplus DL_{orig} \bigoplus DL_{exp}$ based on the preliminary results. This method is denoted as **All**.

3.3 Experiment Results

We use the same experiment set up as described before and evaluate the effectiveness of the proposed expansion methods. Table 3 summarizes the performances. Similar to the previous experiment, parameters or the order of the signals are trained on MB11 and tested on MB12. The results with ‡ indicate improvement over Tie-breaking is statistically significant at 0.05 level based on Wilcoxon signed-rank test.

The results suggest that the query expansion of tie breaking can improve the performance, however, the improvement is not significant. This is probably caused by the limitation of tie-breaking method, which assumes that every query term is important and may not perform well for long queries such as the expanded ones.

Among three document expansion methods, it is less effective to merge the two indexes. And both **Combined** and **All** can improve the performance significantly. In fact, the performance improvement is similar to the improvement when using expansion methods in traditional retrieval models (e.g. the MAP of traditional query expansion is 0.3692 on MB11 and 0.2196 on MB12, and the MAP of document expansion method is 0.4019 on MB11 and 0.2447 on MB12). Moreover, the performance of the best document expansion method is comparable to the best automatic system in TREC 2011 Microblog track. Unfortunately, the overview paper for TREC 2012 Microblog track is not available, but it seems that our method could be ranked among top 3 groups based on the TREC papers about the Microblog track.

4 Conclusions and Future Work

Tie-breaking is a recently proposed new method for combining retrieval signals. In this paper, we study its potential in microblog retrieval and find that the basic tie-breaking method is more effective than traditional retrieval functions. Moreover, we proposed new methods to incorporate query expansion and document expansion into the tie-breaking framework. We find that the proposed document expansion method is more effective than the proposed query expansion method, which may reveal the limitation of tie-breaking methods for the longer queries. In the future, we plan to study how to incorporate temporal-related signals into the tie-breaking framework. Moreover, we will study how to extend the framework for longer queries.

References

1. Golovchinsky, G., Efron, M.: Making sense of twitter search. In: Proc. of CHI'10 Workshop on Microblogging: What and How Can We Learn From It? (2010)
2. Choi, J., Croft, W.B., Kim, J.Y.: Quality models for microblog retrieval. In: Proc. of CIKM'12. (2012)
3. Kumar, N., Carterette, B.: Time based feedback and query expansion for twitter search. In: Proc. of ECIR'13. (2013)
4. Miyanishi, T., Seki, K., Uehara, K.: Combining recency and topic-dependent temporal variation for microblog search. In: Proc. of ECIR'13. (2013)
5. Whiting, S., Klampanos, I.A., Jose, J.M.: Temporal pseudo-relevance feedback in microblog retrieval. In: Proc. of ECIR'12. (2012)
6. Massoudi, K., Tsagkias, M., de Rijke, M., Weerkamp, W.: Incorporating query expansion and quality indicators in searching microblog posts. In: Proc. of ECIR'11. (2011)
7. Choi, J., Croft, W.B.: Temporal models for microblogs. In: Proc. of CIKM'12. (2012)
8. Wu, H., Fang, H.: Tie breaker: A novel way of combining retrieval signals. In: Proc. of ICTIR'13. (2013)
9. Han, Z., Li, X., Yang, M., Qi, H., Li, S., Zhao, T.: Hit at trec 2012 microblog track. In: TREC. (2012)
10. Egozi, O., Markovitch, S., Gabrilovich, E.: Concept-based information retrieval using explicit semantic analysis. ACM TOIS (2011) 8:1–8:34