Learning for Search Results Diversification in Twitter

Ying Wang, Zhunchen Luo, and Yang Yu

China Defense Science and Technology Information Center,
100142, Beijing, China
{suneony,zhunchenluo}@gmail.com, fengqy911@163.com

Abstract. Diversifying the results retieved is an effective approach to tackling users' information needs in Twitter, which typically described by query phrase are often ambiguous and have more than one interpretation. Due to tweets being often very short and lacking in reliable grammatical sytle, it reduces the effectiveness of traditional IR and NLP techniques. However, Twitter, as a social media, also presents interesting opportunies for this task (for example the author information such as the number of statuses). In this paper, we firstly address diversitication of the search results in Twitter with a learning method and explore a series of diversity features describing the relationship between tweets which include tweet content, sub-topic of tweet and the Twitter specific social information such as hashtags. The experimental results on the Tweets2013 datasets demonstrate the effectiveness of the learning approach. Additionally, the Twitter retrieval task achieves improvement by taking into account the diversity features. Finally, we find the sub-topic and Twitter specific social features can help solve the diversity task, especially the post time, hashtags of tweet and the location of author.

1 Introduction

Twitter provides a platform to allow users to post text messages known as tweets to update their followers with their findings, thinking and comments on some topics [1]. Twitter users often conduct search on the posted tweets to fulfill their information needs. As of 2014, the number of queries submitted to Twitter per day is reported to be more than two billion [2]. However, users' information needs, typically described by keyword based queries, are often ambiguous and have more than one interpretation. For example, as for the query "dreamliner battery", there are three relative tweets[1]:

– $tweet_1$*: Boeing 787 battery fire was difficult to control: An investigation of a battery fire aboard a Boeing 787*
– $tweet_2$*: Boeing 787 Dreamliner battery was miswired, Japan says - CTV News*
– $tweet_3$*: Rockford-Area News Boeing proposes battery fix to FAA for 787 Dreamliner planes*

The $tweet_1$ is an introduction to the dreamlinear battery accident while the $tweet_2$ analyzes the reason of the accident and the $tweet_3$ refers to the solution of the accident. The three tweets refer to different aspects of the query and it is difficult to disamiguate the user intent without further information.

[1] Tweets come from Tweets2013 corpus

The typical web search also suffers this problem. Search results diversification has attracted considerable attention [3]. The key idea is to provide a diversified results list, in the hope that different users will find some results that can cover their information needs. Different methods on web search results diversification have been proposed in literature [4] [5] [6]. While, search results diversification in Twitter can be harder than typical web search largely due to tweet being often very short and lacking in reliable grammatical style and quality [7]. Morever, queries in Twitter are even shorter (1.64 words on the average). These factors reduce the effectiveness of traditional IR and NLP techniques.

Fortunately, Twitter also presents interesting opportunities. The rich environment presents us with a myriad of social information over-and-above just using terms in a post all of which potentially can improve on search results diversification performance in Twitter. For example, people uaually use hashtags to indicate the topics of the tweet and the hashtags can be used for diversifying tweets.

In this paper, we firstly address the diversification of results in Twitter search with a learning method and propose a series of features including tweet content, the sub-topic of tweet and Twitter specific social information such as hashtags. Benefited from the learning method, we can easily incorporate aforementioned features into a ranking model easily. With each of these features describing the property of tweet in one perspective, we can diversify the tweets from multiple aspects. Obviously, the pure content is meaningful for diversity. Additionally, different tweets may associate with different sub-topics and two tweets may refer to the same sub-event if their sub-topics are similar. Hence, the sub-topics may help diversify the tweets. The rich social informations in Twitter can also reflect the differences between tweets. We investigate the effects of these features and produce a diverse ranking system [6].

With a series of experiments, we demonstrate the effectiveness of the learning method in tackling the search results diversification problem in Twitter. Our approach achieve comparable performance with the traditional diversification approaches [8] in the evaluations of most measures. Furthermore, we find the Twitter retrieval task achieves improvement by taking into account the diversity features which consider the relationship between tweets in the evaluations of α-nDCG, Precision-IA and Subtopic Recall. Finally, we find sub-topic and the Twitter specific social features can help solve the search results diversification, especially the post time, hashtags of tweet and the location of author.

The contributions of this paper can be summarized as follows:

1) The proposal of a new learning method for search results diversification in Twitter.
2) The exploration of a series of diversity features.
3) The verification of the effectiveness of the proposed approach based on public datasets.

2 Related Work

We review related works on three main areas: Twitter search, diversifying web search results and diversifying Twitter search results.

2.1 Twitter Search

Relevance to the search query is the major ranking criteria in most of the work on tweet. Jabeur, Tamine, and Boughanem model the relevance of a tweet to a query by a Bayesian network that intergrates a veriety of features[9]. Zhang et al train machine learning models for ranking tweets against a query[10]. Luo et al introduce Twitter Building Blocks and their structural combinations, they use this structural information as features into a learning to rank scenario for Twitter retrieval[11]. Luo et al integrate social and opinionatedness information for tweets opinion retrieval[12].

TREC 2011 introduced the Microblog Track which addressed one single pilot task, entitled real-time search task, where the user wished to see the most recent but relevant information to the query. A total 59 groups participated in the track from across the world, with 184 submitted runs. The experimental results indicate the large gap between the best and medians evaluation score (e.g., MAP value) per-topic for 59 participated groups. It shows that Tweets retrieval is far from being a solved problem.

All these methods are simple ad-hoc retrieval approaches that only consider the relevance between tweet and query, while not taking into account the relationship between tweets.

2.2 Diversifying Web Search Result

There are many existing search results diversification methods in web search which can be mainly divided into two categories: implicit approaches and explicit approaches.

The implicit methods assume that similar documents cover similar aspects and model inter-document dependencies. For example, Maximal Marginal Relevance (MMR) method proposes to iteratively select a candidate document with the highest similarity to the user query and the lowest similarity to the already selected documents, in order to promote novelty [4].

The explicit methods explicitly model aspects of a query and then select documents that cover different aspects. The aspects of a user query can be achieved with a taxonomy [3], top retrieved documents [13], query reformulations [14], or multiple external resources [15].

All these methods are non-learning methods and utilize a heuristic predefined utility function. Zhu et al address search results diversification as a learning problem where a ranking function is learned for diverse ranking[6].

None of the aforementioned methods are aimed at search results diversification in Twitter but the traditional way.

2.3 Diversifying Twitter Search Result

Tao, Hauff, and Houben create a microblog-based corpus (Tweets2013) for search result diversification experiments and a comprehensive analysis of the corpus showed its suitability for this purpose[16]. In this paper, we use this corpus to evaluate the performace of our approach.

Ozsoy, Onal, and Altingovde present an empirical analysis of a variety of search result diversification methods adopted from the text summarization and web search domains for the task of tweet ranking. Their experiments revealed that the implicit diversification methods outperform a popular explicit method, xQuAD, due to the vocabulary gap between the official query sub-topics and tweets[8].

Nevertheless, social information of Twitter was not taken into account by those methods.

3 Problem Definition

The diversification problem in Twitter can be naturally stated as a tradeoff between finding relevant tweets and diversifying the retrieved results:

> Given an initial ranking R for a query q, find a re-ranking S that has the maximun coverage and the minimum redundancy with respect to the different aspects underlying q.

4 Overview of Our Approach

We adopt a learning approach to learn a diverse ranking function for tweets and use the tweet content, sub-topic of tweet and Twitter specific social information as features for diversifying tweets.

4.1 Learning to Rank Framework

Learning to rank is a data driven approach which effectively integrates a bag of features in the model. Figure 1 shows the paradigm of learning for tweet diverse ranking.

Firstly, a set of queries Q with related tweets and the ground-truth of the ranking in the form of a vector of ranking scores or a ranking list were used as training data. A bag of features related to the tweets is extracted to form a feature vector. Then a learning to rank algorithm is used to train a diverse ranking model. For a new query, their related tweets, which extract the same features to form feature vector, can be ranked by the diversie ranking function based on this model. The ranking performance of the model using a particular of feature sets in testing data can reflect the effect of these features for search results diversification in Twitter.

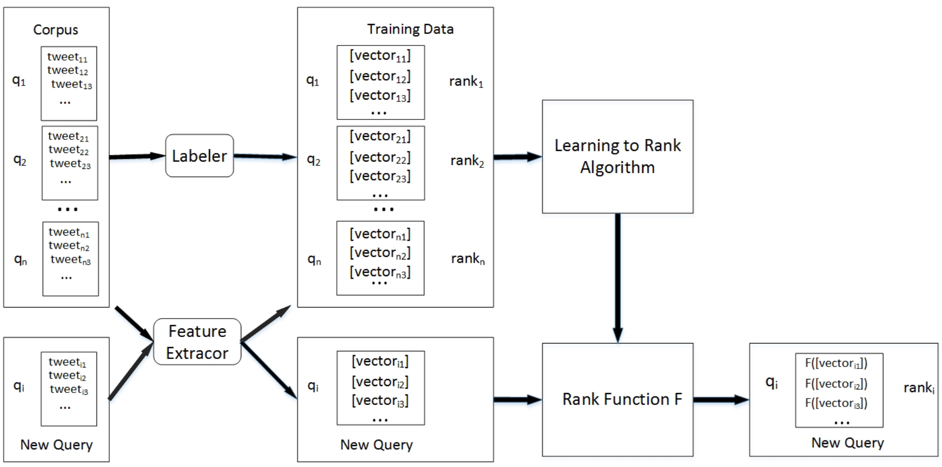

Fig. 1. Framework of learning to diverse rank in Twitter

4.2 Features for Diverse Ranking

We exploit three types of features for tweet diverse ranking:

1) Tweet content feature refers to the one which describes the content diversity between tweets
2) Sub-topic of tweet feature refers to the one that represents the sub-topic difference between tweets
3) Twitter specific social features refer the ones that represent the particular characteristics of Twitter and the author's information.

In the next section, we will describe these three types of features in detail.

5 Feature Description

5.1 Tweet Content Feature

We use text dissimilarity as the content feature between tweets.

Text Dissimilarity Obviously, text dissimilarity is meaningful for diversity. Two tweets may refer to the same sub-event if their texts are similar. We propose to represent it as the cosine dissimilarity based on weightd term vector representations, and define the feature as Eq.1:

$$TT_1 = 1 - \frac{t_i \cdot t_j}{||t_i||||t_j||} \tag{1}$$

where t_i, t_j are the weighted tweet vector based tf*idf. The tf denotes the term frequency and idf denotes inverse document (tweet) frequency.

5.2 Sub-topic Feature

Sub-topics are the emerging topics related to an ongoing real-world topic. For instance, among all tweets[2] refer to the query "hillary clinton resign", we can discover "last, politician, popular, obama, clinton" and "look, woman, india, resign, speech" as different sub-topics which are characterized by a set of words. The sub-topic differs from sub-event in that sub-topics mostly happen across the same broad time interval, and with considerable everlap. However, intuitively, there are also some relationships between sub-topic and sub-event. Different tweets may associate with different sub-topics and two tweets may refer to the same sub-event if their sub-topics (characterized by a set of words) are similar. Therefore, we include the sub-topic of tweet as a feature for the diversity task and use Hierarchical Dirichlet Processes (HDP) to model implicit sub-topics distribution of candidate objects. For a pair of tweets, we define the implicit sub-topic feature as Eq.2:

$$TT_4 = \sqrt{\sum_{k=1}^{m} (p(z_k|t_i) - p(z_k|t_j))^2} \tag{2}$$

where $p(z_k|t_i)$ is the probability distribution of $tweet_i$ on sub-topic z_k. We use hierarchical topic modelling to detect sub-topics in Twitter and conduct the sub-topics distribution using the HAC software[3]. Table 1 shows some sub-topics of tweets related to the query "*hillary clinton resign*" detected using HCA software and the probability distribution of tweet *"Hillary Clinton tops Obama as most popular US politician"* on the detected sub-topics.

[2] Tweets come from Tweets2013 corpus

[3] The software can be downloaded from http://mloss.org/software/view/527/

Table 1. Sub-topics detected using HCA software and probability distribution of tweet on the detected sub-topics

subtopic	**represented words**	tweet
Topic1	last, politician, popular, obama, clinton	0.9807
Topic2	look, woman, india, resign, speech	0.0030
Topic3	run, go, want, hat, snap	0.0026
Topic4	justin, send, bieber, gonna, lift	0.0034
Topic5	clinon, dick, give, email, check	0.0017
Topic6	better, miss, bank, wemen, bill	0.0032
Topic7	first, john, visit, iraq, kabul	0.0002
Topic8	week, sworn, post, know, vote	0.0020
Topic9	support, former, samesex, syria, hurrah	0.0027

5.3 Twitter Specific Social Features

Twitter has many special characteristics of social media. We exploit them and extract Twitter specific social features as follows:

Time Window Size Two tweets may refer to the same sub-event if their post times are close. Time dissimilarity score between two tweets is based on the difference between their normalized timestamps (using Min-Max Normalization), and computed by Eq.3:

$$TT_2 = |t_{norm}(t_i) - t_{norm}(t_j)| \tag{3}$$

where $t_{norm}(t_i)$, $t_{norm}(t_i)$ are the normalized timestamps of two tweets. For example, given the minimal timestamps *"Fri Feb 01 00:09:29 +0000 2013"* and the maximal timestamps *"Sun Mar 31 23:57:58 +0000 2013"*, the normalized timestamp of *"Tue Mar 25 14:45:00 +0000 2008"* is 0.387101.

Hashtags Dissimilarity A hashtag refers to a word in the tweet that begins with the "#" character. It is used to indicate the topic of the tweet. Two tweets may refer to the same sub-event if they contain the same hashtags. For a given pair of tweets, we compute the Jaccard dissimilarity of the set of of hashtags, which is defined as Eq.4:

$$TT_3 = 1 - \frac{|Terms(t_i) \cap Terms(t_j)|}{|Terms(t_i) \cup Terms(t_j)|} \tag{4}$$

where $Terms(t_i)$, $Terms(t_j)$ are the sets of hashtags of two tweets.

Mentions Dissimilarity In a tweet, people usually use "@" preceding a user name to reply other user (Mention). If two tweets related to one event mention the same person, two authors may talk about the same sub-event with the person they mention both. Therefore, we use a binary feature indicating whether two tweets contain the same "@username".

URL Dissimilarity Sharing links in tweets is very popular in Twitter. Most tweets containing a link usually give an introduction to the links. Two tweets may refer to the same sub-event if they contain the same links. Therefore, we use a binary feature indicating whether two tweets contain the same links.

Author Dissimilarity Twitter is a social network. The rich author information can also be used for the task. We use the following features related to the author of tweet: **location**, **language**, **verified**, **statuses**, **followers**, **friends**, **listed**. People living in the same area may care about the same event. Therefore, we consider the location information of authors and use a binary feature indicating whether two authors live in the same area. Intuitively, people use the same language are more likely to care about the same event than those use different languages. We use a binary feature indicating whether two authors use the same language. Additional information related the authority and activeness of users may also reveal the dissimilarity between events people care about. We use a binary feature indicating whether two authors are both verified users. As for the number of statuses, followers, friends and listed, we map them into an interval from 0 to 1 and calculate the dissimilarity between them.

6 Experiment

6.1 Dataset and Experimental Settings

For our evaluations, we use the Tweets2013 corpus[4] that is specifically built for tweet search results diversification problem. The dataset includes tweets collected between Februlary 1, 2013 and March 31, 2013. There are forty seven query topics and each topic has 9 sub-topics on the average.

For learning method, we use the relational learning to rank algorithm proposed by [6]. Different from the traditional learning to rank algorithm where a ranking function is defined on the content of each individual document (tweet), the relational learning to rank algorithm considers both contents of individual document (tweet) and relationships between documents (tweets). Therefore, the diverse ranking function is defined as follows:

$$f_s(x_i, R_i) = w_r^T x_i + w^T h_s(R_i) \tag{5}$$

where x_i denotes the relevance features vector of the candidate tweet x_i, R_i stands for the matrix of relationships between tweet x_i and other selected tweets. For relevance, we use several fetures as summarized in Table 2.

In particular, our experimentation focus on three main questions:

Table 2. Relevance features for relational learning to rank algorithm

Feature	**Description**
$REL_{content}$	Content similarity between tweet and query
$REL_{hashtag}$	Whether a tweet contains a hashtag
$REL_{mention}$	Whether a tweet contains "@username"
REL_{url}	Whether a tweet contains links
$REL_{verified}$	Whether the author is a verified user
$REL_{statuses}$	The number of statuses
$REL_{friends}$	The number of friends
$REL_{followers}$	The number of followers
REL_{listed}	The times the author of a tweet has been listed

1) Is the learning to rank approach effective on search results diversification problem in Twitter?
2) Do the diversity features help in improving the Twitter retrieval?
3) How does the individual feature effect on search results diversification in Twitter?

Five-fold cross validation is used in our experiments. We choose tweets of 27 queries as the training data. The remaining tweets are divided into evaluation data and validation data equally.

[4] The corpus is publicly available at http://wis.ewi.tudelft.nl/airs2013

6.2 Evaluation Metrics

We evaluate diversification methods using the ndeval software [5] employed in TREC Diversity Tasks. We report results utilizing three popular metrics, namely, α-nDCG [17], Precision-IA [3], and Subtopic-Recall [18] at the cut-off values of 10 and 20, as typical in the literature.

6.3 Results

We first investigate whether the learning method is effective on search results diversification problem in Twitter. To evaluate the performance of learning method, we compare it with the state-of-the-art approachs [8], which are introduced as follows:

Max Marginal Relevance (MMR) is a classical implicit diversity method in the diversity research. It employs a linear combination of relevance and diversity [4].

eXplicit Query Aspect Diversification (xQuAD) is an explicit diversification method based on the assumption that aspects of a query can be known apriori [5].

Simple Yet (Sy) [16] present a framework for detecting duplicate and near-duplicate tweets and define a simple yet effective diversification method, so-called Sy [16].

While implementing these approaches in Twitter, [8] use three types of features, namely, the content, hashtags, time features for computing the similarity between tweets and the only content feature for computing the relevance between query and tweet. Therefore we choose to conclude those features in our learning method, marked as *RLTR*. The performaces of *RLTR* and baselines are shown in Table 3.

From the results, we can see that our learning method achieves comparable

Table 3. Performance of Diverse Ranking Methods

Methods	**α-nDCG**		**Prec-IA**		**ST-Recall**	
	@10	**@20**	**@10**	**@20**	**@10**	**@20**
MMR	0.341	0.374	0.066	0.056	0.417	0.539
xQuAD	0.235	0.263	0.050	0.041	0.302	0.419
Sy	0.384	0.383	0.083	0.069	0.419	0.542
RLTR	0.382	0.417	0.058	0.051	0.351	0.483

performance with the traditional diversification approaches in the evaluations of most measures. Specially, in the evaluation of α-nDCG@20, the relative improvement of our learning approach over the best of traditional diversification approaches is up to 8.9%. The results demonstrate the effectiveness of the learning method in tackling the search results diversification problem in Twitter.

Next, we investigate whether the diversity features can help in improving the Twitter retrieval. To demonstrate the effectiveness of diversity features in Twitter retrieval, we made a contrast experiment. The first experiment only considers the relevance features between tweet and query, marked as *REL*, while the second takes into account all the diversity features on the basis of the first, marked as *REL+DIV*. The results of performace between *REL* and *REL+DIV*

[5] http://trec.nist.gov/data/web10.html

are shown in Table 4.

From the results, we can see that the *REL+DIV* outperforms the *REL* in

Table 4. Performance of Ranking Methods. A significant improvement over the baseline with bold respectively ($p < 0.05$)

Methods	α-nDCG		Prec-IA		ST-Recall	
	@10	@20	@10	@20	@10	@20
REL	0.282	0.322	0.045	0.043	0.273	0.445
REL+DIV	**0.491**	**0.510**	**0.074**	**0.069**	**0.441**	**0.578**

the evaluations of all measures. In particualr, the relative improvement of the Precision-IA@10 and Precision-IA@20 is up to 64.4% and 60.4% , the improvement of the α-nDCG@10 and α-nDCG@20 is up to 71.4% and 58.4%. The results demonstrate the effectiveness of the diversity features in Twitter retrieval.

Finally, we investigate the effects of sub-topic feature and each of individual Twitter specific social features. We use the learning method which takes into account the relevance features and leverages diversity feature, tweet content as a baseline, marked as *BASE*. We combine each individual feature with the *BASE*. Table 5 shows the performance of each ranking model.

From the results, we can see that sub-topic feature and the Twitter specific

Table 5. Performance of Each Ranking Model. A significant improvement over the baseline with bold respectively ($p < 0.05$)

Methods	α-nDCG		Prec-IA		ST-Recall	
	@10	@20	@10	@20	@10	@20
BASE	0.417	0.428	0.056	0.049	0.349	0.465
+subtopic	**0.454**	**0.475**	**0.067**	**0.061**	**0.421**	**0.574**
+time	**0.445**	**0.463**	**0.064**	**0.057**	**0.393**	**0.536**
+hashtags	**0.431**	**0.459**	**0.063**	**0.059**	**0.401**	**0.519**
+location	**0.433**	**0.455**	**0.060**	0.051	**0.387**	**0.518**
+url	0.419	0.431	0.058	0.050	0.349	0.462
+verified	0.402	0.421	0.056	0.048	0.341	0.471
+language	0.413	0.420	0.054	0.049	0.339	0.466
+statuses	0.420	0.431	0.058	0.051	0.355	0.470
+friends	0.417	0.426	0.053	0.049	0.346	0.462
+followers	0.399	0.405	0.051	0.047	0.347	0.453
+listed	0.401	0.409	0.050	0.049	0.353	0.457
best	**0.499**	**0.514**	**0.075**	**0.068**	**0.447**	**0.580**

social features can help tackle the search results diversification task in Twitter, especially the post time, hashtags of tweet and the location of author.

Comparing with the baseline, the relative improvement of the model taking into account the sub-topic feature is up to 8.8% and 10.9% in the evaluations of α-nDCG@10 and α-nDCG@20 respectively. As for the sub-topic feature, different tweets may associate with different sub-topics and two tweets may refer to the same sub-event if their sub-topics are similar. For example, as for the query "North Korea nuclear", there are three relative tweets[6] and their distributions on sub-topics are shown as Table 6:

[6] Tweets come from Tweets2013 corpus

– $tweet_4$: *US and China agree on North Korea sanctions after nuclear test - Fox News: San Francisco ChronicleUS and China http://t.co/9qs4CGruYb*
– $tweet_5$: *NEWSFLASH: Diplomats say US, China agree on new sanctions to punish North Korea for nuclear test. More as we get*
– $tweet_6$: *How Powerful Was #NorthKorea;s #Nuclear Test? #SouthKorea believes 7 kilotons; #Japan,10 & #Germany institute*

Table 6. Probability Distributions of Tweets on Sub-Topics

sub-topics	$tweet_4$	$tweet_5$	$tweet_6$
Topic1	0.0151	0.0212	0.0031
Topic2	0.0061	0.0868	0.0127
Topic3	0.8705	0.6907	0.0869
Topic4	0.0215	0.0248	0.0056
Topic5	0.0106	0.0140	0.0007
Topic6	0.0238	0.0141	0.0006
Topic7	0.0129	0.0133	0.0013
Topic8	0.0213	0.0203	0.8885
Topic9	0.0178	0.0648	0.0006

The $tweet_4$ and $tweet_5$ both refer to the "reaction of China to the nuclear test", while the $tweet_6$ refers to the "details of the third nuclear test". We can see from Table 6 that the distributions of the $tweet_4$ and $tweet_5$ are similar, while the $tweet_6$ is not. Therefore, the sub-topic of tweet can indeed help diversifying tweets.

We can also observe that the ranking model yield significant improvement over the baseline when taking into account time feature. The relative improvement is up to 6.7% and 8.2% in the evaluations of α-nDCG@10 and α-nDCG@20 respectively. As for the time feature, two tweets may refer to the same sub-event if their post times are close. For example, as for the query "hillary clinton resign", there are three relative tweets[7]:

– $tweet_7$: *TN China: Secretary of State Hillary Clinton formally resigns: Her resignation is effective upon the swearing[created_at:Fri Feb 01 20:39:04 +0000 2013]*
– $tweet_8$: *Secretary of State Hillary Clinton formally resigns: Her resignation is effective upon the swearing-in of John[created_at:Fri Feb 01 20:43:07 +0000 2013]*
– $tweet_9$: *Hillary Clinton: As Hillary Clinton leaves office after four years, John Kerry prepares to take over[created_at:Fri Feb 01 10:00:14 +0000 2013]*

The $tweet_7$ and $tweet_8$ both refer to the "details of resignation" while the $tweet_9$ to "who follows Clinton as secretary of state". Obviously, the post times of the $tweet_7$ and $tweet_8$ are close, while the the $tweet_9$ is far from them. Therefore, the time feature can improve the diversifying performance.

From Table 5, we can see that the model taking into account hashtags features outperforms the baseline in all settings. Especially, in the evaluations of α-nDCG@10 and α-nDCG@20, the relative improvement is up to 3.4% and 9.6%. As for the hashtags feature, two tweets may refer to the same sub-event if they contain the same hashtags. For example, as for the query "syria civil war", there are three relative tweets[8]:

[7] Tweets come from Tweets2013 corpus
[8] Tweets come from Tweets2013 corpus

– $tweet_{10}$*: RT @SyriaDayofRage: (02-21-13) #Damascus #Syria l Rebels Move Closer to Central Damascus as Clashes continue and Rebel Rockets hitting*
– $tweet_{11}$*: So far today 17 martyrs were reported in #Damascus and its suburbs, 3 in #Idlib, 3 in #Hama,3 in #Aleppo,2 in #Homs*
– $tweet_{12}$*: RT @lysdeschamps: #Syria 13 02 Killed: 231. 20 children, 98 women, 59 rebels. Aleppo 50 Idlib 33 Damascus 30 Daraa 21 Homs 18 Deir Azzo*

The $tweet_{10}$ and $tweet_{11}$ both refer to the "cities where the fighting is" while the $tweet_{12}$ refers to the "casualties of the war". we can see that the $tweet_{10}$ and $tweet_{11}$ both contain hashtag "#Damascus" which is a city, while the $tweet_{12}$ not. Therefore, the hashtag feature can also help tackling the task.

As for the location of author, the relative improvement of the model taking into it over the baseline is up to 3.8% and 6.3% respectively. It is due to that peoples who live in the same area may concern about the same sub-topic of event.

Finally we add all the features which can significantly improve the diversifying performance into a ranking model. They are **sub-topic**, **time**, **hashtag** and **location of author** features. Table 5 shows the **best** result of method which improves significantly over the baseline.

7 Conclusion

To the best of out knowledge, we are the first to propose the learning method for search results diversification in Twitter. We explore a series of diversity features. The experimental results on the tweets2013 datasets demonstrate the effectiveness of our approach. Additionally, the Twitter retrieval task achieves improvement by taking into account the diversity features. Finally, we find the sub-topic and Twitter specific social features can help solve the diversity task, especially the post time, hashtags of tweet and the location of author.

References

1. Java, A., Song, X., Finin, T., Tseng, B.L.: Why we twitter: An analysis of a microblogging community. In: Advances in Web Mining and Web Usage Analysis, 9th International Workshop on Knowledge Discovery on the Web, WebKDD 2007, and 1st International Workshop on Social Networks Analysis, SNA-KDD 2007, San Jose, CA, USA, August 12-15, 2007. Revised Papers. (2007) 118–138
2. Busch, M., Gade, K., Larson, B., Lok, P., Luckenbill, S., Lin, J.: Earlybird: Real-time search at twitter. In: IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012. (2012) 1360–1369
3. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying search results. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, ACM (2009) 5–14
4. Carbonell, J.G., Goldstein, J.: The use of mmr, diversity-based reranking for reordering documents and producing summaries. In: SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 24-28 1998, Melbourne, Australia. (1998) 335–336
5. Santos, R.L.T., Macdonald, C., Ounis, I.: Exploiting query reformulations for web search result diversification. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010. (2010) 881–890

6. Zhu, Y., Lan, Y., Guo, J., Cheng, X., Niu, S.: Learning for search result diversification. In: The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014. (2014) 293–302
7. Teevan, J., Ramage, D., Morris, M.R.: #twittersearch: a comparison of microblog search and web search. In: Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011. (2011) 35–44
8. Ozsoy, M.G., Onal, K.D., Altingovde, I.S.: Result diversification for tweet search. In: Web Information Systems Engineering - WISE 2014 - 15th International Conference, Thessaloniki, Greece, October 12-14, 2014, Proceedings, Part II. (2014) 78–89
9. Jabeur, L.B., Tamine, L., Boughanem, M.: Uprising microblogs: a bayesian network retrieval model for tweet search. In: Proceedings of the ACM Symposium on Applied Computing, SAC 2012, Riva, Trento, Italy, March 26-30, 2012. (2012) 943–948
10. Zhang, X., He, B., Luo, T., Li, B.: Query-biased learning to rank for real-time twitter search. In: 21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012. (2012) 1915–1919
11. Luo, Z., Osborne, M., Petrovic, S., Wang, T.: Improving twitter retrieval by exploiting structural information. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada. (2012)
12. Luo, Z., Osborne, M., Wang, T.: Opinion retrieval in twitter. In: Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012. (2012)
13. Carterette, B., Chandar, P.: Probabilistic models of ranking novel documents for faceted topic retrieval. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November 2-6, 2009. (2009) 1287–1296
14. Radlinski, F., Dumais, S.T.: Improving personalized web search using result diversification. In: SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, August 6-11, 2006. (2006) 691–692
15. He, J., Hollink, V., de Vries, A.P.: Combining implicit and explicit topic representations for result diversification. In: The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012. (2012) 851–860
16. Tao, K., Hauff, C., Houben, G.: Building a microblog corpus for search result diversification. In: Information Retrieval Technology - 9th Asia Information Retrieval Societies Conference, AIRS 2013, Singapore, December 9-11, 2013. Proceedings. (2013) 251–262
17. Clarke, C.L.A., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon, I.: Novelty and diversity in information retrieval evaluation. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, Singapore, July 20-24, 2008. (2008) 659–666
18. Zhai, C., Lafferty, J.D.: A risk minimization framework for information retrieval. Inf. Process. Manage. **42**(1) (2006) 31–55