Hashtag Recommendation for Hyperlinked Tweets

Surendra Sedhai
School of Computer Engineering
Nanyang Technological University, Singapore
surendra001@e.ntu.edu.sg

Aixin Sun
School of Computer Engineering
Nanyang Technological University, Singapore
axsun@ntu.edu.sg

ABSTRACT

Presence of hyperlink in a tweet is a strong indication of tweet being more informative. In this paper, we study the problem of hashtag recommendation for hyperlinked tweets (*i.e.,* tweets containing links to Web pages). By recommending hashtags to hyperlinked tweets, we argue that the functions of hashtags such as providing the right context to interpret the tweets, tweet categorization, and tweet promotion, can be extended to the linked documents. The proposed solution for hashtag recommendation consists of two phases. In the first phase, we select candidate hashtags through five schemes by considering the similar tweets, the similar documents, the named entities contained in the document, and the domain of the link. In the second phase, we formulate the hashtag recommendation problem as a learning to rank problem and adopt RankSVM to aggregate and rank the candidate hashtags. Our experiments on a collection of 24 million tweets show that the proposed solution achieves promising results.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information Filtering*

Keywords

Hashtag recommendation; Microblog, Tweets; Learning to rank

1. INTRODUCTION

Twitter is a popular platform for users to share personal activities to their friends, as well as to share interesting online resources (*e.g.,* news articles, Web pages) to the large audience. It has been reported that many of the trending topics in Twitter are headline news or persistent news, making Twitter a form of news media [3, 4]. Many of such news pieces are shared through tweets with hyperlinks to Web pages. In this study, we call a tweet containing one or more hyperlinks to external documents a ***hyperlinked tweet***.

Hashtags (#-prefixed keywords) are often used to provide topical or contextual information about tweets. For this reason, hashtags are frequently used as queries to search for tweets about a topic or a specific event (*e.g.,* #NBA and #sigir2014). In other words, hashtags not only provide the right context to interpret the tweets and to categorize tweets, but also serve as a medium to promote tweets to reach more readers. Appropriate hashtag usages therefore benefit many applications (*e.g.,* tweet search, tweet classification, and event detection). Despite the importance of hashtags, large portion of tweets, including many hyperlinked tweets, do not contain hashtags.

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.

ACM 978-1-4503-2257-7/14/07 ...$15.00.
http://dx.doi.org/10.1145/2600428.2609452.

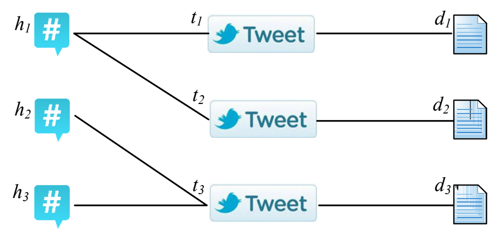


Figure 1: Hashtag, hyperlinked tweet, and linked document

In this paper, we focus on the problem of *hashtag recommendation for hyperlinked tweets*, which is to recommend appropriate hashtags to tweets containing hyperlinks. Figure 1 illustrates a conceptual model consisting of *hashtags*, *hyperlinked tweets*, and the *linked documents*. An edge exists between a hashtag and a hyperlinked tweet if the tweet is annotated with the hashtag (*e.g.,* h_1 and t_1). Similarly, an edge between a tweet and a document exists if the document is linked from the tweet (*e.g.,* t_1 and d_1). We argue that through hyperlinked tweets, the functions of hashtags can be extended to the content of the linked documents. That is, the hashtags convey additional contextual information to the linked documents, facilitate the topical categorization of these documents, and serve as a medium to propagate the documents to more readers in Twitter. It is reported that presence of link in a tweet is strong indication of tweet being more informative [2]. Compared with many non-hyperlinked tweets (*e.g.,* those about daily activities), we believe that a user is more willing to share a hyperlinked tweet with larger audience. However, links in Twitter are mostly encoded in short URLs which neither provide any hint about the information source nor the content of the links. Hence, readers do not have any clue about a link unless the tweet provides a good description. Such description often relies on the right context provided by hashtags. For example, hashtags like #NBA and #sigir2014 provide the right context to understand both the tweet and the link. However, users may not be aware of the relevant hashtags for their tweets due to lack of exposure to the relevant hashtags [10]. Hence, hashtag recommendation system can play an important role to improve user experience in microblogging.

Existing work on hashtag recommendation do not consider the embedded links in tweets, which is an important source of information. In this paper, we recommend hashtags to hyperlinked tweets by first selecting candidate hashtags through five schemes utilizing both the content of tweet and the linked Web pages (*e.g.,* by similar tweets and similar Web pages, by named entities in the Web pages). We then aggregate and rank the candidate hashtags by formulating this problem as a learning to rank problem. Experiments on a collection of 24 million tweets demonstrate the effectiveness of the proposed approach. We show that the proposed solution is effective in combining the strength of the tweet content and the embedded link. Because of the additional information, hashtag recommendation for hyperlinked tweets is expected to achieve better accuracy than that for non-hyperlinked tweets, making hashtag recommendation a more practical application for users.

2. RELATED WORK

In this section, we review hashtag recommendation to tweets and the applications of learning to rank on tweets.

Tag Recommendation. Tags provide description and contextual information to online resources. Similarity based tag recommendation system has been widely adopted to recommend tags for similar resources. A recent study shows that cosine similarity with TFIDF weighting scheme is the most appropriate measure to retrieve similar microblog posts [12]. Using cosine similarity, [5] recommends hashtags by combining hashtags obtained from similar tweets as well as hashtags from similar users. In our proposed solution, we select candidate hashtags by finding similar tweets and similar Web pages by using cosine similarity. Based on the assumption that both tweet content and hashtags are about the same theme but written in different languages, statistical machine translation techniques have been used for hashtag recommendation [1, 7]. In our solution, we also adopt the language translation model. However, we use named entities instead of all words in a Web page based on the observation that some hashtags are named entities in the linked Web pages.

Though tag co-occurrence has been used to recommend tags for many online resources (*e.g.,* photos in Flickr [11]), co-occurrence of hashtags in Twitter is intrinsically different from most others due to the length constraint of tweet. Hence, tag recommendation method suitable for photos in Flickr or other online resources may not be directly applicable for hashtag recommendation for tweets.

Learning to Rank: Learning to rank is a task of automatically constructing a ranking model using training data, such that the model can sort new objects according to their degrees of relevance, preference, or importance [6]. Recently, learning to rank has been applied to many tweet search tasks, *i.e.,* ranking tweets by their relevance to search queries [2, 8, 13]. In our solution, we adopt the pairwise learning to rank framework realized by RankSVM to rank the candidate hashtags obtained from different selection schemes.

3. HASHTAG RECOMMENDATION

Given a hyperlinked tweet t containing link ℓ to a Web page, our task is to recommend a list of hashtags that are most relevant to the tweet.[1] Our proposed hashtag recommendation method has two phases: *candidate hashtag selection* and *hashtag ranking*. We now present the two phases in detail.

3.1 Candidate Hashtag Selection

Candidate hashtag selection is a process of selecting a subset of hashtags from all existing hashtags that have been used to annotated any of the observed tweets with or without hyperlinks. This process effectively reduces the search space of possible hashtags to be recommended to a tweet, facilitating more efficient recommendation. Candidate hashtags are selected as follows.

Hashtags from Similar Tweets and Similar Web Pages. Based on the assumption that topically similar tweets are more likely to be annotated by similar hashtags, we search for candidate hashtags by using the similar tweets to the given tweet t. Cosine similarity and TFIDF weighting scheme are adopted in this search. The top 20 most voted hashtags are selected by using the top 50 most similar tweets following the setting in [5]. Based on the discussion that the functions of hashtags can be extended to the linked documents (see Section 1), we also assume that similar Web pages are annotated by similar hashtags and search for candidate hashtags by using the similar Web pages in a similar manner. Another 20 candidate hashtags are obtained.

Hashtags based on the Domain of the Link. In our dataset, we observe that hashtags often reflect the topics of the linked Web site. Let ℓ_d be the domain of hyperlink ℓ. We select the top 20 hashtags by $P(h|\ell_d)$, which is the probability that h is used to annotate a tweet containing links from domain ℓ_d. For example, hashtags #news, #cnn, and #breaking are candidate hashtags for links from domain *edition.cnn.com*.

Hashtags based on the Named Entities in the linked Web Page. Named entities appearing in a Web page are important topic indicators of the page. We also observe that some of the named entities are directly adopted as hashtags in the hyperlinked tweets. We use two models to select candidate hashtags for a tweet based on the named entities in its linked Web page: Random Walk with Restart (RWR) model [9] and Language Translation (LT) model [7].

Random Walk with Restart (RWR) model. We construct an entity-hashtag graph where there are two sets of nodes N_e and N_h (see Figure 2(a)). $N_e = \{e_i\}$ is the set of nodes corresponding to the named entities in the Web page linked from tweet t. $N_h = \{h_j\}$ is the set of nodes corresponds to a subset of hashtags in the data collection. A hashtag is included in this graph if it can be reached from any of the named entity node $e_i \in N_e$ through the edges. There are three kinds of directed edges: $e_i \rightarrow h_j$, $h_j \rightarrow e_i$, and $h_j \rightarrow h_k$, weighed by $P(h_j|e_i)$, $P(e_i|h_j)$, and $P(h_k|h_j)$, respectively.[2] The first two conditional probabilities are computed by the number of times a hashtag h_j is used to annotate a tweet linking to a document containing an named entity e_i, with respect to the frequencies of e_i and h_j respectively. The last conditional probability is the asymmetric hashtag co-occurrence. Only the edges with positive weights are included in the graph.

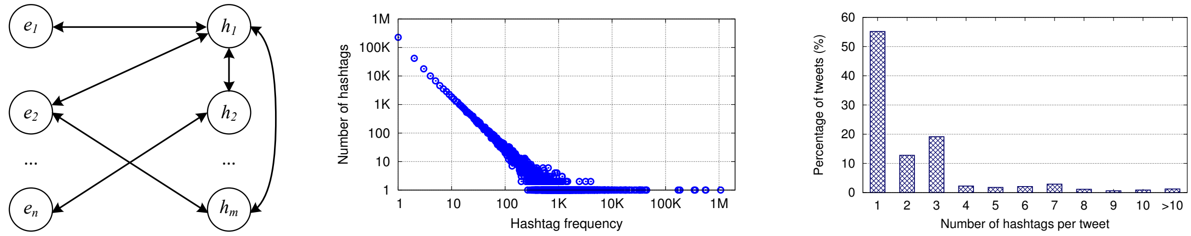
To apply RWR model, let $\vec{r}$ be the restart vector of dimensionality $|N_e| + |N_h|$, where $|\cdot|$ denotes the cardinality of a set. The entries corresponding to named entities in $\vec{r}$ are set to 1 and the entries corresponding to hashtags are set to 0. Let $\mathbf{P}$ be the transition matrix derived from the weights computed above. Let α be the restart probability which was set to 0.75 in our experiments. The steady state value of $\vec{v}$ is computed using the following formula. The top 20 hashtags with highest corresponding values in $\vec{v}$ are selected as candidate hashtags. Note that, $\vec{v}$ is randomly initialized.

$$\vec{v} = (1 - \alpha)\vec{v}\mathbf{P} + \alpha\vec{r}$$

Language Translation (LT) model. In this model we consider the named entities and hashtags as descriptions of the same content in two different languages. Again, let N_e be the set of named enti-

[1] In our data collection of 24 million tweets, more than 96% of hyperlinked tweets each contain only one link. For easy presentation, we assume each tweet contains one link, although the proposed method can be easily extended to handle multiple links.

[2] In our experiments, we have also evaluated the graph with edges between two named entities weighted by $P(e_k|e_i)$. This version of the graph leads to poorer recommendation accuracy hence is not reported due to page limit.



(a) Entity-hashtag Graph (b) Hashtag distribution (c) Number of hashtag per tweet

Figure 2: Entity-hashtag graph, hashtag frequency distribution, and number of hashtags per tweet

ties appearing in the document linked from tweet t, the score of a hashtag h_j is computed by:

$$Score(h_j) = \sum_{e_i \in N_e} P(h_j|e_i)$$

where, $P(h_j|e_i)$ is the same conditional probability as in the RWR model. Similarly, from this model, we obtain 20 candidate hashtags with highest scores.

3.2 Recommendation by Learning to Rank

Observe that each of the candidate hashtag selection methods presented above can be used as a stand-alone hashtag recommendation method, because the candidate hashtags are ranked by their corresponding selection criteria. In particular, candidate hashtag selection by similar tweets is the hashtag recommendation method proposed in [5] (not considering the hashtags by similar users). However, to benefit from all candidate selection methods, we propose to aggregate all selected candidate hashtags and rank them. More specifically, we formulate the hashtag recommendation task as a learning to rank problem.

Pairwise Learning to Rank. We represent each candidate hashtag as a feature vector and use pairwise learning to rank method to find the top ranked hashtags from the candidate set. Given a tweet t, let h_i^+ be a positive hashtag that is selected as a candidate hashtag and is used to annotate tweet t; let h^- be a negative hashtag that is selected as a candidate hashtag but is not used to annotate t. Then the pair $\langle h^+, h^- \rangle$ is a positive instance and $\langle h^-, h^+ \rangle$ is a negative instance in learning the model. Each pair is represented by the vector difference between the feature vectors of the two hashtags, to be detailed shortly. In the test phase, each candidate hashtag is paired with the rest of the candidate hashtags and then classified by using the learned model. Let H_c be the set of candidate hashtags. The recommendation score of candidate hashtag h_i is: $f(h_i) = \sum_{h_j \in H_c, h_i \neq h_j} I(h_i, h_j)$, where $I(h_i, h_j) = 1$ if $\langle h_i, h_j \rangle$ is classified as positive and 0 otherwise.

Features. We use two sets of binary features (*i.e.,* the value of each feature is either 1 or 0).[3] The first set contains 5 features where each feature corresponds to a candidate hashtag selection scheme. Each of the 5 dimensions are set to 1 if the candidate hashtag is selected by similar tweet, similar Web page, domain of the link, named entities with RWR, and named entities by LT, respectively. The second set of 4 features utilizes Wikipedia, category list and domain related information, detailed below.

1. *Wikipedia*. This feature is set to 1 if the hashtag appears as an anchor text in a Wikipedia entry. The list of anchor texts are parsed from the English Wikipedia dump on January 30, 2010.
2. *Is Category*. This feature is set to 1 if the hashtag is one of the top-level category in Yahoo! hierarchy, or it matches a category in any of the four Web sites: bbc, cnn, nytimes and reddit. After duplicate removal, there are 75 categories including sports, technology, business, movie, jobs etc.
3. *Is Popular Domain*. This feature is set to 1 if the hashtag matches one of the top 5000 most popular domains in our dataset.
4. *Matches Domain*. This feature is set to 1 if the hashtag matches the domain of the link contained in the tweet.

The second set of 4 features is proposed based on the observation that hashtags are often used to topically categorize tweet content. We also observe that many Web sites themselves are topically indicative (*e.g.,* Yelp for business review, CNN for news).

4. EXPERIMENTS

Dataset. We collected two months (May 1 to Jun 30, 2013) of sampled tweets from Twitter using twitter streaming API guided by hashtags.org. More specifically, we used the hashtags in "trending up" and "trending down" categories in hashtags.org as query keywords to search for tweets on daily basis. A tweet is returned if it contains the query keyword as its hashtag or as a word in its content. On average 135 hashtags were used as query keywords in each day. In total, we have collected 24 million tweets published by 11.9 million users. The collected tweets contain 6.9 million links with 3.4 million distinct URLs after resolving the short URLs. We managed to download 2.05 million URLs ignoring the deadlinks and links requires login (*e.g.,* Facebook links). Among them, 1.37 million are in English.[4]

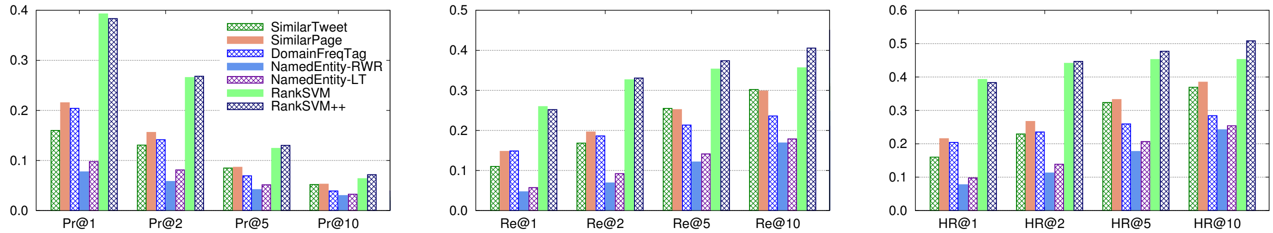
Dataset Analysis. In our dataset, about 30% of tweets contain hashtags probably due to the fact that the tweets were collected by using hashtags as queries. In total, there are about 1 million distinct hashtags which appear for about 20 million times. Figure 2(b) plots the hashtag frequency distribution where a power-law like distribution is observed as expected. Among the tweets containing hashtags, slightly more than half of them each contain one hashtag. Relatively, it is rare to have more than 3 hashtags in a tweet, shown in Figure 2(c). After processing the linked Web pages (*e.g.,* parsing out the textual content and extracting the named entities[5]), we observe that 12.65% of tweets contain named entities from their linked Web pages as hashtags. This accounts only the single-word hashtags without segmentation. Besides named entities, broad categories such as sports, news, business are also widely used hashtags. Moreover, domain names such as NBA, TED, CNN are found to be used as hashtags frequently. Among hyperlinked tweets, more than 96% of tweets each contain only one link.

Methods in Comparison. We evaluated 7 methods in our experiments, summarized in Table 1. Recall that each of the candidate hashtag selection methods can be used to recommend hashtags. We use these 5 methods as baseline in our evaluation. The first method

[3] We have also evaluated the non-binary weighting scheme of the same set of features (*e.g.,* the feature values are normalized to [0, 1]) but observed poorer performance.

[4] Java library: `http://code.google.com/p/language-detection/`

[5] Stanford NER: `http://nlp.stanford.edu/software/CRF-NER.shtml`



(a) Precision@k (b) Recall@k (c) HitRate@k

Figure 3: [Best viewed in color] Hashtag recommendation accuracy of the 7 methods, evaluation with $k = \{1, 2, 5, 10\}$

Table 1: The 7 methods in evaluation

Method name	Description
SimilarTweet	Most frequently used hashtags in similar tweets
SimilarPage	Most frequently used hashtags in similar Web pages
DomainFreqTag	Most frequently used hashtags for the given domain
NamedEntity-RWR	Hashtags selected by named entities using RWR
NamedEntity-LT	Hashtags selected by named entities using LT
RankSVM	RankedSVM using the first set of 5 features
RankSVM++	RankedSVM using the all (two sets of) features

SimilarTweet is similar to the method proposed in [5]. However, because of the different approach adopted in collecting data, the user information used in [5] cannot be adopted here. The last two methods, *i.e.,* RankSVM and RankSVM++ both realize the proposed learning to rank framework with the only difference in the set of features used. RankSVM uses 5 binary features where each feature corresponds to a candidate hashtag selection scheme. RankSVM++ uses all 9 binary features listed in Section 3.2. Both methods were implemented using the liblinear library.[6]

To learn the RankSVM models, we randomly selected 15,000 hyperlinked tweets from the first 40 days of the data to derive the training instances. Another 7,000 hyperlinked tweets randomly selected from the remaining 20 days of the data are used as test tweets. In candidate hashtag selection for both training and testing, we limit the search for similar tweets (resp. Web pages) posted one day before the currently processing tweet to simulate inaccessibility of future data in reality.

Evaluation Metric. We use three metric to evaluate hashtag recommendation accuracy: *Precision@k*, *Recall@k*, and *HitRate@k* (*Pr@k*, *Re@k*, and *HR@k* for short); k={1, 2, 5, 10} is the number of top-ranked recommended hashtags. Let H_k be the set of top-k recommended hashtags and H_g be the set of ground-truth hashtags of a tweet t (likely $1 \leq |H_g| \leq 3$). ***Pr@k*** for tweet t is $|H_k \cap H_g|/k$; ***Re@k*** for tweet t is $|H_k \cap H_g|/|H_g|$; and ***HR@k*** for t is 1 if $|H_k \cap H_g| \geq 1$ and 0 otherwise. The values reported for each method are the averaged values over the 7,000 test tweets.

Experimental Results. Figure 3 reports *Pr@k*, *Re@k*, and *HR@k* of the five baseline methods and the two RankSVM methods for $k = \{1, 2, 5, 10\}$. We make three observations from the results.

1. The two RankSVM methods outperform all the five baseline methods on all three evaluation metrics for all k values. The results clearly show the effectiveness of aggregating candidate hashtags selected by exploiting multiple sources. When $k = 1$, RankSVM is slightly better than RankSVM++ on all three metrics. However, RankSVM++ outperforms RankSVM when k is greater than 2, suggesting the effectiveness of the 4 additional features in the learning to rank model.
2. Among the five baseline methods, SimilarPage performs slightly better than SimilarTweet, followed by DomainFreqTag, particularly when k is greater than 1 on three evaluation matrices. This suggests that similar Web pages are more reliable in recommending hashtags, and the domain of the hyperlink is indeed an important piece of information for hashtag recommendation.
3. There is a large performance degradation for *Pr@k* when $k = 5$ or 10 because most tweets have fewer than 3 hashtags (see Figure 2(c)). However, when $k = 10$, RankSVM++ achieves *Re@10* of 0.4 and *HR@10* of 0.5.

[6] `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`

5. CONCLUSION

Hyperlinked tweets are relatively more informative than non-hyperlinked tweets. We show that the functions of hashtags can be extended to the linked documents from hyperlinked tweets. To facilitate information sharing and organization, we propose to recommend hashtags to hyperlinked tweets. In our proposed solution, we select candidate hashtags by exploiting the content of the tweets, the linked documents, and the domain of the hyperlink. Using the learning to rank models, we show that the aggregation and ranking of candidate hashtags achieves promising hashtag recommendation accuracy, evaluated by precision, recall, and hit rate.

6. REFERENCES

[1] Z. Ding, X. Qiu, Q. Zhang, and X. Huang. Learning topical translation model for microblog hashtag suggestion. In *IJCAI*, pages 2078–2084, 2013.

[2] Y. Duan, L. Jiang, T. Qin, M. Zhou, and H.-Y. Shum. An empirical study on learning to rank of tweets. In *COLING*, pages 295–303, 2010.

[3] M. Hu, S. Liu, F. Wei, Y. Wu, J. Stasko, and K.-L. Ma. Breaking news on twitter. In *CHI*, pages 2751–2754, 2012.

[4] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW*, pages 591–600, 2010.

[5] S. M. Kywe, T.-A. Hoang, E.-P. Lim, and F. Zhu. On recommending hashtags in twitter networks. In *SocInfo*, pages 337–350, 2012.

[6] T.-Y. Liu. Learning to rank for information retrieval. *Found. Trends Inf. Retr.*, 3(3):225–331, Mar. 2009.

[7] Z. Liu, X. Chen, and M. Sun. A simple word trigger method for social tag suggestion. In *EMNLP*, pages 1577–1588, 2011.

[8] R. McCreadie and C. Macdonald. Relevance in microblogs: Enhancing tweet retrieval using hyperlinked documents. In *OAIR*, pages 189–196, 2013.

[9] J.-Y. Pan, H.-J. Yang, C. Faloutsos, and P. Duygulu. GCap: Graph-based automatic image captioning. In *CVPR Workshop*, pages 146–154, 2004.

[10] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *WWW*, pages 695–704, 2011.

[11] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW*, pages 327–336, 2008.

[12] E. Zangerle, W. Gassler, and G. Specht. On the impact of text similarity functions on hashtag recommendations in microblogging environments. *Social Netw. Analys. Mining*, 3(4):889–898, 2013.

[13] X. Zhang, B. He, T. Luo, and B. Li. Query-biased learning to rank for real-time twitter search. In *CIKM*, pages 1915–1919, 2012.