12-2012

On recommending hashtags in Twitter networks

Su Mon KYWE
Singapore Management University, monkywe.su.2011@smu.edu.sg

Tuan-Anh HOANG
Singapore Management University, tahoang.2011@smu.edu.sg

Ee Peng LIM
Singapore Management University, eplim@smu.edu.sg

Feida ZHU
Singapore Management University, fdzhu@smu.edu.sg

DOI: https://doi.org/10.1007/978-3-642-35386-4_25

On Recommending Hashtags in Twitter Networks

Su Mon Kywe, Tuan-Anh Hoang, Ee-Peng Lim and Feida Zhu

Singapore Management University, Singapore
{monkywe.su.2011,tahoang.2011,eplim,fdzhu}@smu.edu.sg

Abstract. Twitter network is currently overwhelmed by massive amount of tweets generated by its users. To effectively organize and search tweets, users have to depend on appropriate hashtags inserted into tweets. We begin our research on hashtags by first analyzing a Twitter dataset generated by more than 150,000 Singapore users over a three-month period. Among several interesting findings about hashtag usage by this user community, we have found a consistent and significant use of new hashtags on a daily basis. This suggests that most hashtags have very short life span. We further propose a novel hashtag recommendation method based on collaborative filtering and the method recommends hashtags found in the previous month's data. Our method considers both user preferences and tweet content in selecting hashtags to be recommended. Our experiments show that our method yields better performance than recommendation based only on tweet content, even by considering the hashtags adopted by a small number (1 to 3)of users who share similar user preferences.

Keywords: Twitter, hashtag, recommendation systems

1 Introduction

Motivation. Recommendation systems have been widely used at e-commerce websites to identify from a huge range of products the most appropriate ones for each user. Various recommendation methods have been proposed to predict the wanted products based on users' preferences as well as preferences of similar users [12]. A rising star of social information networks, Twitter presents to its users the same challenge of finding the most appropriate people to follow, tweets to read, as well as hashtags to use in their tweets [7]. Recommendation systems are therefore pertinent in these scenarios [1, 2].

In Twitter, users write tweets which are short messages containing no more than 140 characters. A hashtag is a word prefixed by a # symbol and one or more hashtag can be inserted into a tweet. Past empirical research shows that hashtags have been used for different purposes. Some people use hashtags to categorize their tweets. Others use hashtags to tag content related to disasters or special events such as elections. Hashtags are also used for brand promotion or micro-meme discussions [6]. Hashtags make tweets easily searchable by other relevant users and this facilitates conversations among the users. Moreover, hashtags make tweets more accessible by hashtag-based search engines such as hashtags.org[1]. In [4], hashtags have been used to help users tag other social media sites. Since hashtags are neither registered nor controlled by any user or group, it will be hard for some users to find appropriate hashtags for their tweets.

Research objectives and contributions. In this paper, we therefore address the personalized hashtag recommendation task in Twitter. The objective is to recommend a list of hashtags appropriate for a given user who has just written a new tweet. We do not consider hashtag recommendation for retweets (i.e., "forwarded" tweets) as they often contain the same hashtags as the corresponding original tweets. It is therefore relatively easy to derive hashtags for retweets.

Hashtag recommendation should be personalized as we would like to consider the user preferences in the choice of hashtags. Twitter users adopt different styles and preferences in writing tweets. For example, users from UK may prefer hashtags in British spellings. Classical music lovers may prefer using composer names as hashtags for musical pieces. Knowing their personal preferences will help to predict the appropriate hashtags.

We begin this research by first analyzing a Twitter dataset consisting of tweets written by more than 150,000 Singapore users over a three-month period from October 2011 to December 2011. This is a reasonably large user community with 44M tweets. We examine the usage of hashtags among these users and their tweets, and highlight several interesting findings about the dataset.

The second part of the paper focuses on our proposed hashtag recommendation method. Our proposed method selects hashtags from both similar tweets (of the target tweet) and similar users (of the user who writes the target tweet). The selected hashtags are ranked and the top ranked hashtags are then recommended to the target user. We evaluate our proposed method and compare it with the recommendation method which only considers the hashtags from the most similar tweets. The results show that our method outperforms the latter approach by about 20 percent.

On the whole, this paper makes a number of contributions to hashtag analysis and recommendation as shown below:

- For the first time, a very large user community and its tweets have been used in a study on hashtag usage and recommendation. We have observed in this dataset that less than 8% of tweets contain hashtags and 40% of users ever use hashtags in our three-month data.
- Our study shows that hashtag usage by a user community is very skewed. Very few hashtags enjoy high popularity in tweets and users, while the vast majority of them are used in one tweet or by one user. This observation is consistent with the earlier studies.
- For any given day, we observe that 40% of the hashtags are not used by the user community in the last 30 days. This suggests that a lot of hashtags used

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

are new to the users. This observation is only possible as we track the tweets from the same user community over time.
- We have developed a personalized hashtag recommendation method considering both user preferences and tweet content. The former has not been used in the previous methods.
- Our experiments show that user preferences from very few similar users can help to improve recommendation accuracy significantly.

Paper outline. Our paper is structured as follows. Section 2 provides a quick summary of the related recommendation research in Twitter. We describe the Singapore's user community and Twitter data collected from its users in Section 3. We also present our analysis results in this section. Section 4 describes our proposed hashtag recommendation method and its experiment results. We conclude the paper in Section 5.

2 Related Works

In this section, we give a brief overview of the traditional recommendation systems followed by previous recommendation research on Twitter network.

2.1 Traditional Recommendation systems

Recommendation systems are information filtering systems which predict the preference of a user towards items (such as songs, books, or movies) or social elements (e.g. people or groups) that she has not considered before [5, 12]. There are two types of recommendation systems – *personalized* and *non-personalized*. Non-personalized recommendation systems rank the items without considering the individual user's preferences. For example, one may recommend the top ten popular songs of the current month. On the other hand, personalized recommendation systems consider the preferences of an individual user. The focus of our paper is on personalized recommendation. There are essentially two major approaches to perform personalized recommendation, namely collaborative filtering and content-based recommendation.

Collaborative Filtering Approach The underlying assumption of the collaborative filtering approach is that if a person X has adopted several common items as adopted by another person Y previously, X is more likely to adopt other Y's items than the items of a random person. In the context of product rating recommendation, collaborative filtering has been used to predict the rating a target user assigns to an item using the ratings on the item assigned by other users who share similar rating preferences as the target user. This type of collaborative filtering is referred as the "user-to-user" collaborative filtering [13].

Another type of collaborative filtering approach is "item-to-item" collaborative filtering. In this approach, we first derive the correlation between two items which is measured by the portion of common users who purchase both items.

We then recommend a new item to a target user using its correlation with other items already adopted by the target user.

Beyond user and item level information, collaborative filtering approach can also be performed in the latent factor space as a user or item can be represented by a set of latent factors through matrix factorization techniques[8]. It has been shown that matrix factorization techniques can yield very accurate recommendation results albeit higher algorithmic complexity.

Content Based Approach Content-based recommendation approach measures similarity between items by comparing their features and characteristics. The recommendation of an item is made to a targeted user if the item is similar to other items adopted by the user before. Unlike item-to-item collaborative filtering, the content based approach makes use of item content only to determine similarity between items.

Other Approaches More recently, community-based recommendation systems have been introduced to recommend items based on the preferences of a user's friends. Such a recommendation approach is only possible when users are connected with one another by friendship links or other forms of social relations[3]. There are many other recommendation systems using demographic information, such as age, profession, country, language, etc., to predict the user's preferences. Such systems use domain specific knowledge about how item features meet the user' needs and preferences or how the items are useful for the user. There are also hybrid systems which combine the above approaches.

2.2 Hashtag Recommendation for Twitter Users

Currently, Twitter has not implemented any hashtag recommendation system which suggests appropriate hashtags for the users' tweets. In the research literature, there are works related to hashtag recommendation and hashtag prediction. We found two hashtag recommendation approaches that are relevant and both of them use *only* tweet content[15, 11]. They will be described below in greater detail. Hashtag prediction refers to predicting the hashtag to be used by a user in the future. In [14], Yang et. al proposed to solve hashtag prediction by training a SVM classifier using a variety of features. Note that this task does not involve any target tweet.

Tweet similarity approach. Zangerle et al. [15] assumed that the primary purpose of the hashtags is to categorize the tweets and facilitate the search. The paper recommends suitable hashtags to the a target user, depending on the content that the user enters without considering user's preference for specific hashtags. Preliminary analysis of hashtags usage in a Twitter data collection obtained by a set of search queries shows that 86% of unique hashtags are used less than five times within 3,209,281 tweets with hashtags. The five most popular hashtags (#jobs, #nowplaying, #zodiacfacts, #news and #fb) appear in 8% of all tweets with hashtags. In other words, a few popular hashtags are used intensively while most of the other hashtags are used very sparsely. The paper also finds out the use of hashtags by spammers (e.g. assigning 17 hashtags to a single spammed tweet).

Zangerle et al. proposed a hashtag recommendation system that retrieves a set of tweets similar to a user given tweet. Similarity score is calculated by TF-IDF scheme. Then, the hashtags are extracted from the retrieved similar tweets and are ranked using one of the proposed score functions: (a) OverallPopularityRank score: number of hashtag occurrences in the whole dataset; (b) RecommendationPopularityRank score: number of hashtag occurrences in the retrieved similar tweet dataset; or (c) SimilarityRank score: similarity score of the most similar tweets containing the hastag. Experiments showed that SimilarityRank score is the best among them and the performance of the recommendation system is the best when only five hashtags are recommended.

Naive Bayes method. In [11], Mazzia et al. recommended hashtags by observing the content produced by the target user. Instead of TF-IDF to find similar tweets, the method proposes to use Bayes model to estimate the probabilities of using different hashtags. In the experiments, the Twitter dataset used is first cleaned by removing micro-memes and spams. Micro-memes are detected by identifying tweets which use the same hashtags but are very dissimilar. Spams are filtered by removing users who have too many tweets using the same hashtag. The Bayes model used in this paper is represented by the following formula.

$$p(C_i|x_1, ..., x_n) = p(C_i)p(x_1|C_i)...p(C_i)p(x_n|C_i)/p(x_1...x_n)$$

where C_i represents the i^{th} hashtag and $x_1, ..., x_n$ represent the words. $p(C_i|x_1, ..., x_n)$ is the probability of using hashtag C_i given the words that the user generates and the hashtags with the highest probabilities are recommended to the user. $p(C_i)$ is the ratio of the number of times hashtag C_i is used to the total number of tweets with hashtags. $p(x_1|C_i)...p(x_n|C_i)$ is calculated from the existing data of tweets.

3 Hashtag Usage Analysis

3.1 Twitter Data for Usage Analysis

In our study, we collect the Twitter data generated by a community of Singapore users. A complete analysis of hashtag usage in the entire Twitter network is not possible as such a dataset is not publicly available. Most researchers in the past chose to analyze Twitter data collected using some forms of data sampling on the stream of Twitter data returned by the APIs provided by the company. For example, Zangerle et. al used a set of query keywords to gather tweets[15]. Inevitably, the analysis results will be biased by the query relevant tweets.

As there is yet a comprehensive analysis of hashtag usage in the tweets generated by user communities, and how the usage patterns may affect hashtag recommendation, we first perform a detailed analysis on the three-month data (October 2011 to December 2011) generated by this Singapore user community.

Our analysis aims to answer the following research questions: (a) How often are hashtags used in tweets? (b) How many hashtags do we expect in a tweet? (c) How familiar are users in using hashtags? (d) Do the hashtags assigned already appear in earlier tweets? Providing answers to the above questions will give us a good understanding of the hashtag usage patterns of a user community and their changes over time.

We collect the Twitter data generated by more than 150,000 Singapore users who are identified by the location field in their user profiles. Every user is at least directly or indirectly connected to a small set of carefully selected seed users so as to prevent spammers to be included. The seed users are popular political bloggers, commentators, election candidates and news media during Singapore Election 2011. Since election is a big socio-political event, we believe that we cover the majority of Singapore Twitter users. We crawl all tweets of these Singapore users on a daily basis. In this manner, we are assured that almost all tweets from this user community have been completely downloaded for our study. Table 1 shows the important statistics found in this dataset. There are more 65,000 users who have written some original tweets during the three-month period. The remaining users (nearly 60% of total user population) do not write original tweets. They could perform retweeting or simply reading tweets from others. Our dataset also contains nearly 450,000 distinct hashtags and 45M original tweets.

# users	65,410
# users using hashtags	46,244
# distinct hashtags	449,206
# original tweets	44,997,784
# original tweets containing hashtags	3,534,869

Table 1. Data Statistics

3.2 Hashtag Usage Analysis

There are substantial fraction of users (about 39%) using hashtags in their original tweets, and very small fraction of original tweets containing hashtags (<8%) as shown in Table 1. This suggests that many users know how to use hashtags but very few actually tweet a lot using hashtags. Figure 1 shows that the fraction of users using hashtags and the fraction of tweets containing hashtag over the three-month period remain very stable for this user community.

We define *tweet popularity* of a hashtag by the number of tweets containing the hashtag. We show the scatterplot of tweet popularity of hashtags in Figure 2(a). Each point in the figure represents the number of hashtags with the same tweet popularity. The distribution is power law-like showing that most hashtags appear in one tweet each and very few tweets enjoy very high tweet popularity. In a similar way, we define *user popularity* of a hashtag by the number

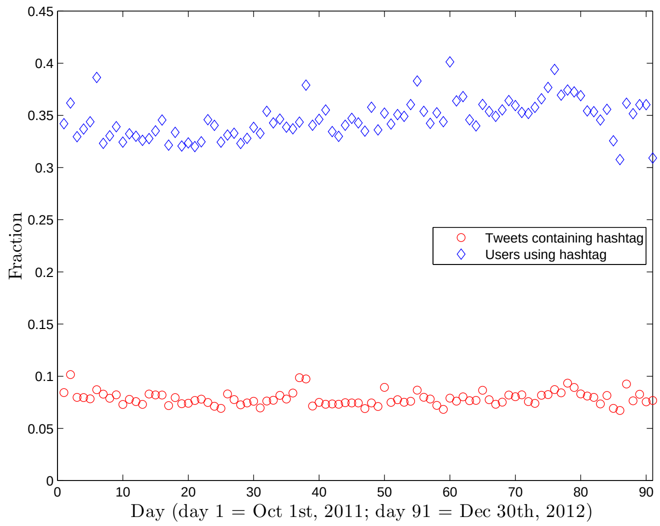


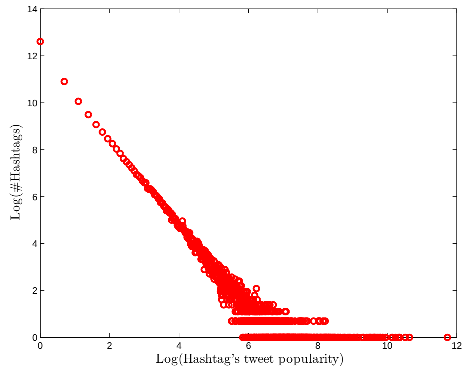
Fig. 1. Hashtag usage

of users using the hashtag. Figure 2(b) shows that the user popularity distribution of hashtags also follows the power law distribution. This suggests that only a few hashtags enjoy high popularity while most hashtags are used by a single user.

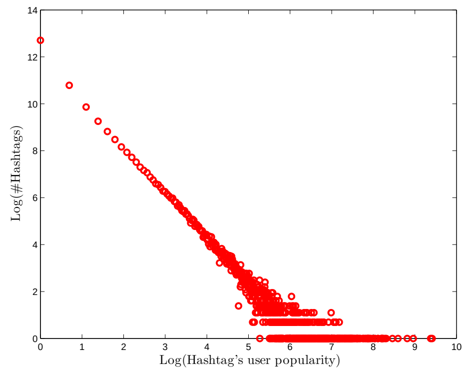
Next, we analyze how frequently users write tweets with hashtags. As shown in Figure 2(c), most users write only one tweet containing hashtag(s) during the observed period. Very few users write many tweets that contain hashtags. Finally, we found out most tweets with hashtag(s) contain only one hashtag as shown in Figure 3. There are very few tweets containing more than one hashtag. This is not a surprise given the short tweet length.

Finally, we want to know if the hashtags are new as users assign them to tweets. Unfortunately, the verification of new hashtags is very costly and may not be viable due to the lack of all historical twitter data. We therefore introduce the definition of "fresh hashtag". A hashtag is said to be fresh to a user community if it has not been used by any user in the community in the last k months. This definition constrains the freshness verification to only k previous months of data generated by a user community. To reduce the verification cost, we have $k = 1$ in our current study.

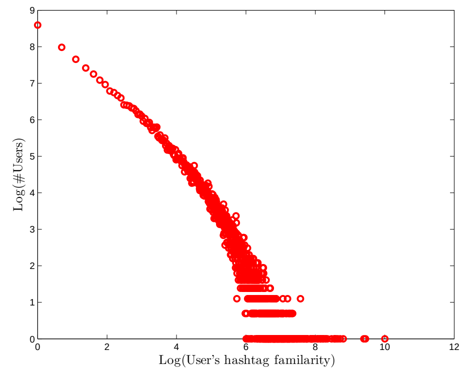
Figure 4 depicts the fraction of fresh hashtags, the fraction of tweets containing fresh hashtags and the fraction of users using fresh hashtags for each day. It is interesting to find 40% fresh hashtags are introduced each day. This suggests that another 40% hashtags are replaced each day. The life expectancy of many hashtags are therefore very short. Less than 30% of tweets contain fresh hashtags and around 40% of users use fresh hashtags each day. These observations lead us to believe that hashtag recommendation is an important task as it helps



(a) Tweet Popularity



(b) User Popularity



(c) User Hashtag Familarity

Fig. 2. Data Distribution

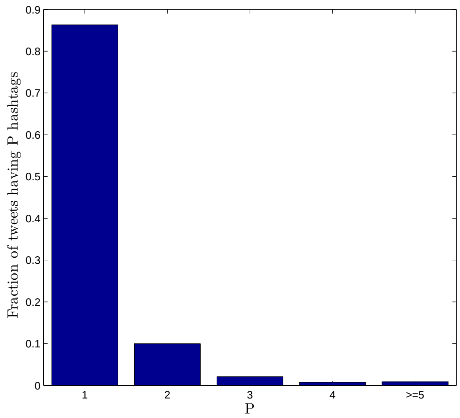


Fig. 3. Number of Hashtags in Each Tweet

users to adopt more hashtags and makes their tweets easily searchable by other relevant users. The recommendation should also involve recent past data so as to recommend fresher hashtags.

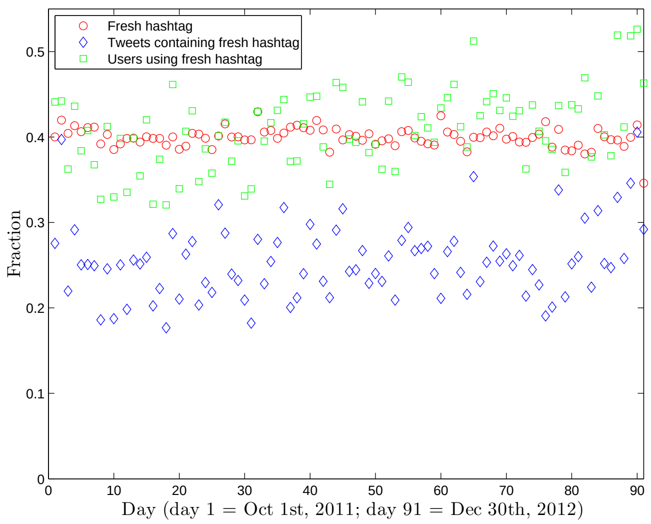


Fig. 4. Fresh hashtag usage

4 Personalized Hashtag Recommendation

Unlike previous methods that recommend hashtags found in similar tweets [15, 11, 10], we propose a new recommendation method that recommends hashtags which are not only appropriate for the tweet but also match the target user's taste. In other words, given a user-tweet pair, we would like to find other similar user-tweet pairs and recommend the hashtags from those user-tweet pairs. We believe that this approach will be able to personalize the recommended hashtags to the user's perferences. In the following, we first describe our proposed method followed by its evaluation.

4.1 Our Proposed Method

Finding similar user-tweet pairs involves three subtasks: (a) selecting hashtags from users with preferences similar to the target user, (b) selecting hashtags from tweets that are similar to the target tweet, and (c) deriving ranking scores for the selected hashtags. In both subtasks (a) and (b), we adopt a TF-IDF scheme to find similar users and tweets as described below.

Selecting hashtags from similar users. We represent a user by her preference weights for each hashtag in our hashtag dictionary H. Formally, a user u_j is represented by a weight vector:

$$u_j = \{w_{1j}, w_{2j}, w_{3j}, \ldots, w_{i|H|}\}$$

where w_{ij} is the preference weight of user u_j towards hashtag h_i and can be defined by the TF-IDF scheme.

$$w_{ij} = TF_{ij}.IDF_i$$

$$TF_{ij} = \frac{Freq_{ij}}{Max_j}, IDF_i = log\left(\frac{N_u}{n_i}\right)$$

where $Freq_{ij}$ = usage frequency of hashtag h_i by user u_j, Max_j = maximum hashtag usage frequency by u_j, N_u = total number of users, and n_i = number of users who use h_i before.

The intuition of TF_{ij} is that if a user uses a hashtag a lot, more preference weight is given to the hashtag. At the same time, this weight is normalized by the maximum hashtag frequency of the user. IDF_i assigns higher weight to a hashtag if the latter is rarely used by other users.

Given a target user u and another user u_i, we can measure the cosine similarity between them as follows.

$$Sim(u, u_i) = \frac{u \cdot u_i}{||u|| \cdot ||u_i||}$$

We then rank the users by similarity score and the most similar X users are selected. Let $TopXUsers(u)$ denote the X users most similar to u, and

$Hashtags(u_i)$ be the set of hashtags previously used by u_i. We combine the hashtags from these top-X users to be our candidate hashtag set $HTofUsers(u)$.

$$HTofUsers(u) = \cup_{u_i \in TopXUsers(u)} Hashtags(u_i)$$

Selecting hashtags from similar tweets. In a similar manner, we represent a tweet t_k can be represented by a weighted vector of words in a word vocabulary W.

$$t_k = \{w_{k1}, w_{k2}, w_{k3}, \ldots, w_{k|W|}\}$$

where

$$w_{kl} = TF_{kl}.IDF_l$$

$$TF_{kl} = \frac{Freq_{kl}}{Max_k}, IDF_l = log\left(\frac{N_t}{n_l}\right)$$

where $Freq_{kl}$ = frequency of word w_l in tweet t_k, Max_k = maximum word frequency in t_k, N_t = total number of tweets, and n_l = number tweets in which w_l appears.

The similarity score between the target tweet t and another tweet t_k is defined by:

$$Sim(t, t_k) = \frac{t \cdot t_k}{||t|| \cdot ||t_k||}$$

We now select the top-Y tweets most similar to the target tweet t, denoted by $TopYTweets(t)$. Let $Hashtags(t_k)$ denote the set of hashtags in tweet t_k. We derive a second set of candidate hashtags $HTofTweets(t)$ from $TopYTweets(t)$ as follows.

$$HTofTweets(t) = \cup_{t_k \in TopYTweets(t)} Hashtags(t_k)$$

Ranking candidate hashtags. The candidate hashtags to be recommended for the target user u and tweet t can be obtained by the union of hashtags from top-Xsimilar users and top-Y similar tweets.

$$SuggestedHashtags(u, t) = HTofUsers(u) \cup HTofTweets(t)$$

After that, hashtags in $SuggestedHashtags(u, t)$ are ranked by frequency. The hashtag frequency is defined by adding the number of times the hashtag is used by top-X users with the number of times it appears in top-Y tweets. Finally, the top ranked hashtags are finally recommended to the user u.

4.2 Experiment

To evaluate our hashtag recommendation method, we conduct experiments using the tweets generated by Singapore users in November and December of 2011. Tweets that do not contain hashtags are removed from the dataset. The remaining dataset in November contains 2,264,801 tweets and 37,617 unique users and is used as training data. To evaluate the recommendation results, we randomly selected 5606 original tweets from the December data with authors in the training set. These tweets form our target tweet set. The hashtags actually used in the target tweets serve as the ground truth. Since the hashtags to be recommended are from November, we expect that they are still relatively fresh.

Since other previous methods recommend hashtags purely based on similar tweets, our experiment varies the number of similar users (i.e., X) used in our method. When $X = 0$, our method will recommend only hashtags from similar tweets. We also want to evaluate the different number of similar tweets Y used in recommendation.

For each target user-tweet pair, we consider the top *five* and top *ten* recommended hashtags and measure the performance of our method using **hit rate** as defined below.

$$Hit\ Rate = \frac{\text{Number of Hits}}{\text{Number of Target User-Tweet Pairs}} \quad (1)$$

A hit occurs when the recommended hashtags for a target tweet t include at least one of the ground truth hashtags. Although multiple hashtags may be used in a target tweet, such cases are rare. Hence, it is reasonable to use the above hit rate measure.

We use Apache Lucene[2] to derive the similarity scores and retrieve the hashtags of the top-X similar users and hashtags of the top-Y similar tweets as Lucene is very efficient in such computation and retrieval.

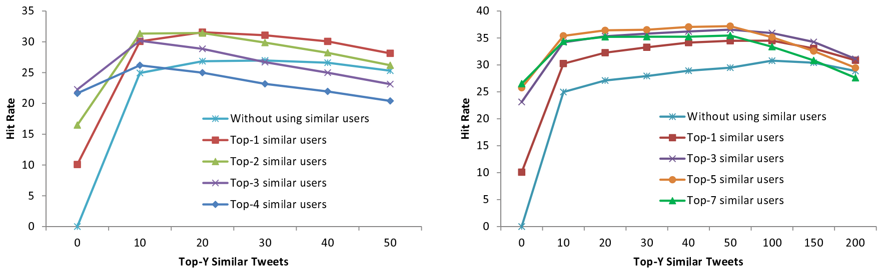
4.3 Results

Figure 5(a) shows the hit rate (in percentage) of top five recommended hashtags. We vary the number of top similar tweets Y used from 0 to 50, and measure the performance of our method with top $X = 0$ to 4 similar users. The figure shows that as we increase the number of similar tweets from 0 to 10, the hit rate improves significantly. The improvement beyond 10 similar tweets is however very small or even negative. We can also observe that considering top 1 to 3 similar users can help to further improve the hit rate when the number of similar tweets are small, i.e., 10 and 20. The improvement percentage of recommendation using top 1 similar user over recommendation without similar user at $Y = 10$ is about 20%. Our method performs best with hit rate = 31.56% when $X = 1$ and $Y = 20$.

Figure 5(b) shows the hit rate (in percentage) of top ten recommended hashtags. We vary the number of top similar tweets Y used from 0 to 200, and measure the performance of our method with top $X = 0, 1, 3, 5$ and 7 similar users. On the whole, the hit rate has improved as we recommend more hashtags. Again, most significant improvement in hit rate occurs between $Y = 0$ and $Y = 10$. Beyond $Y = 10$, the improvement is small. On the other hand, using similar users is almost always better than not using similar users. The improvement margin of recommendation using top 1 similar user over recommendation

[2] `http://lucene.apache.org/core/`

without similar user at $Y = 10$, i.e., 21%, is similar to that observed for top 5 recommended hashtags. This time, our method performs best with hit rate = 37.19% when $X = 5$ and $Y = 50$.



(a) Hit Rate (%) for Five Recommended Hashtags

(b) Hit Rate (%) for Ten Recommended Hashtags

Fig. 5. Hit Rate

5 Conclusions

Hashtag recommendation is a novel problem in Twitter. It is also important as most tweets do not carry hashtags and most hashtags do not have long life span. Our hashtag usage study on a three-month Twitter data generated by over 150,000 users in Singapore confirms the above observations. Our study shows that 40% of the hashtags in any day are fresh, i.e, not used in the last 30 days. We also observe that the usage patterns are stable over the period.

Our paper also proposes a personalized hashtag recommendation method that considers both target user preferences and target tweet content. Given a user and a tweet, our method selects the top most similar users and top most similar tweets. Hashtags are then selected from the most similar tweets and users and assigned some ranking scores. Experiment results show that using user preferences and tweet content will give us better recommendation than just using tweet content alone.

Beyond this early and promising results, there are several other interesting future directions to explore for hashtag recommendation. We can further divide hashtags into different categories, e.g., by freshness or by topic, and study their recommendation accuracies. In [9], popular hashtags have been clustered into four categories by their before-peak, after-peak, and during-peak popularity. For each hashtag category, it will be interesting to propose different recommendation methods that work well. So far, our proposed method is based on simple collaborative filtering. More sophisticated methods such as matrix factorization can also be used in the future.

6 Acknowledgements

This research/project is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

References

1. Armentano, M.G., Godoy, D.L., Amandi, A.A.: Recommending information sources to information seekers in twitter. In: International Workshop on Social Web Mining (2011)
2. Armentano, M.G., Godoy, D.L., Amandi, A.A.: A topology-based approach for followees recommendation in twitter. In: The 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (ITWP'11) (2011)
3. Chua, F.C.T., Lauw, H.W., Lim, E.P.: Predicting item adoption using social correlation. In: SIAM Conference on Data Mining. pp. 367–378 (2011)
4. Correa, D., Sureka, A.: Mining tweets for tag recommendation on social media. In: The 3rd International Workshop on Search and Mining User-generated Contents (2011)
5. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 22(1), 5–53 (2004)
6. Huang, J., Thornton, K.M., Efthimiadis, E.N.: Conversational tagging in twitter. In: The 21st ACM Conference on Hypertext and hypermedia. pp. 173–178 (2010)
7. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: The 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (2007)
8. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer 42(8), 30–37 (2009)
9. Lehmann, J., Gonçalves, B., Ramasco, J.J., Cattuto, C.: Dynamical classes of collective attention in twitter. In: The 21st International Conference on World Wide Web. pp. 251–260 (2012)
10. Li, T., Yu Wu, Y.Z.: Twitter hash tag prediction algorithm (2011), `http://cerc.wvu.edu/download/WORLDCOMP%2711/2011%20CD%20papers/ICM3338.pdf`
11. Mazzia, A., Juett, J.: Suggesting hashtags on twitter `http://www-personal.umich.edu/~amazzia/pubs/545-final.pdf`
12. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: Recommender Systems Handbook, pp. 1–35. Springer (2011)
13. Schafer, J.B., Konstan, J.A., Riedl, J.: E-commerce recommendation applications. Data Mining and Knowledge Discovery 5(1-2), 115–153 (2001)
14. Yang, L., Sun, T., Zhang, M., Mei, Q.: We know what @you #tag: does the dual role affect hashtag adoption? In: The 21st International Conference on World Wide Web. pp. 261–270 (2012)
15. Zangerle, E., Gassler, W.: Recommending #-tags in twitter. In: Proceedings of the CEUR Workshop (2011), `http://ceur-ws.org/Vol-730/paper7.pdf`