Hashtag Recommendation Based on User Tweet and Hashtag Classification on Twitter

Mina Jeon, Sanghoon Jun, and Eenjun Hwang(✉)

School of Electrical Engineering, Korea University, Seoul, South Korea
{jmina, ysbhjun, ehwang04}@korea.ac.kr

Abstract. With the explosive popularity of various social network services (SNSs), an enormous number of user documents are generated and shared daily by users. Considering the volume of user documents, efficient methods for grouping or searching relevant user documents are required. In the case of Twitter, self-defined metadata called hashtags are attached to tweets for that purpose. However, due to the wide scope of hashtags, users are having difficulty in finding out appropriate hashtags for their tweets. In this paper, we propose a new hashtag recommendation scheme for user tweets based on user tweet analysis and hashtag classification. More specifically, we extract keywords from user tweets using TF-IDF and classify their hashtags into pre-defined classes using Naïve Bayes classifier. Next, we select a user interest class based on keywords of user tweets to reflect user interest. To recommend appropriate hashtags to users, we calculate the ranks of candidate hashtags by considering similar tweets, user interest and popularity of hashtags. To show the performance of our scheme, we developed an Android application named "TWITH" and evaluate its recommendation accuracy. Through various experiments, we show that our scheme is quite effective in the hashtag recommendation.

Keywords: Twitter · Hashtag · User interest · Naïve Bayes classifier · Classification · Ranking · Recommendation · Android

1 Introduction

As a very popular online social networking service, Twitter is currently used by millions of users and organizations to quickly share and discover information. Users can access Twitter through the web or mobile devices and publish a message called tweet of up to 140 characters that can be sent or read by anyone. Each message can have replies from other users, which could lead to a real-time conversation around some hot topic or interesting content.

Furthermore, for grouping and searching of certain topics, users can utilize self-defined hashtags starting with a hash symbol (#) as a prefix to a word or a multi-word phrase without whitespace. A hashtag is a simple and convenient tool for users to categorize their tweets to represent some specific event or topic. In other words, users can attach appropriate hashtags to their tweet to join a relevant conversation about a specific topic on Twitter. The usage of hashtag in tweet is very easy and simple. People use hashtags for diverse purposes: for example, to classify a product (e.g. *#iphone5s*),

Y. Chen et al. (Eds.): WAIM 2014, LNCS 8597, pp. 325–336, 2014.
DOI: 10.1007/978-3-319-11538-2_30

to represent information or event (e.g. *#worldcup2014, #superbowl*), or to express an emotion (e.g. *#happy, #sad*), etc. If users talk about same topic using a specific hashtag, their tweets will appear in the same stream. In this way, hashtags can help to identify messages on a specific topic or event and facilitate conversation among users.

However, due to the wide range of topics discussed on Twitter, it is difficult for users to find out appropriate hashtags for their tweets. Besides, some users include too many hashtags in their tweet. For instance, in the tweet "*#This #is a #tweet #with #lots #of #hashtags,*" its hashtags do not deliver any clear purpose. Such tweets tend to confuse other users and are usually considered as a spam. Accordingly, hashtags in a tweet should be relevant to the topic and reflect user interest. To help users utilize hashtags appropriate to their tweets, it is necessary to understand the topics of tweets and recommended appropriate hashtags relevant to the topic to the users effectively and efficiently. Therefore, in this paper, we propose a hashtag recommendation scheme that helps users to utilize appropriate hashtags for their tweets by analyzing existing tweets and their hashtags. The main components of our proposed scheme are as follows.

- Extracting a set of keywords from each tweet
- Mapping the set of keywords into one of pre-defined classes
- Calculating the ranks of candidate hashtags
- Recommending the most appropriate hashtags to the user

Consequently, when a user completes a tweet with no hashtags, then our system can automatically recommend appropriate hashtags relevant to the tweet.

2 Related Work

So far, many studies have been done for tweet classification and hashtag recommendation. In this section, we first consider several works on tweet classification and then introduce works on hashtag recommendation.

2.1 Tweets Classification

On Twitter, it is important to classify a specific topic of tweet into general categories with high accuracy for better information retrieval or for easier understanding of topics. Therefore, a number of recent papers have introduced the classification of tweets on Twitter. Go et al. [1] introduced an approach for automatically classifying the sentiment of Twitter messages. These messages are classified as either positive or negative using tweets with emotions for distant supervised learning. They showed that machine learning algorithms (Naïve Bayes, Maximum Entropy and SVM) have high accuracy. Wang et al. [2] focused on hashtag-level sentiment classification, instead of presenting the sentiment polarity of each tweet relevant to the topic. They classified hashtags into three categories, a topic which is closely connected to a certain hashtag, sentiment hashtags which express subjective opinions and sentiment-topic hashtags which indicate a certain target word and the sentiment words. To capture the relationships among hashtags, they developed an undirected edge between two nodes if those particular hashtags appeared together in a single tweet. In addition, Lee et al. [3] classified trending topics into general categories such as sports, politics, technology, etc. by using two main supervised learning techniques. They employed a text classification technique called Naive Bayes and proposed a network-based approach to predict the category of a topic knowing the categories of its similar topics.

2.2 Hashtag Recommendation

Sometimes, people attempt to use a hashtag to categorize their tweets as broadcast media for certain topics or events such as elections. However, it might be difficult for them to select hashtags suitable for their tweets. To solve this problem, many recommendation schemes have been proposed for suggesting appropriate hashtags to the users. Zangerle et al. [4] suggested an approach for the recommendation of suitable hashtags to the user during the creation process. The recommender system retrieves a set of similar tweets using TF-IDF. Hashtags are extracted from the retrieved similar tweets and are ranked using their number of occurrences in the whole dataset, their number of occurrences in the retrieved dataset or similarity scores of tweets. Kywe et al. [5] proposed a personalized hashtag recommendation method based on collaborative filtering, which recommends hashtags found in the previous month's data. This method considers both a target user interest and tweet content. Given a user and a tweet, this method selects the top most similar users and top most similar tweets using TF-IDF. Hashtags are then selected from the most similar tweets and users assigned some ranking scores. Furthermore, Gordin et al. [6] proposed a method for unsupervised and content based hashtag recommendation for tweets. This method applied Naïve Bayes, Expectation-Maximization and Latent Dirichlet Allocation (LDA) to model the underlying topic assignment of language classified tweets. Even though a number of related studies have been conducted for the classification and recommendation, only a few works have addressed the problem of individual preferences for the hashtag recommendation.

3 Proposed Approach

In this section, we describe how to recommend appropriate hashtags to the user when he/she creates a tweet. The main steps for the recommendation include keyword extraction, keyword classification, hashtag ranking, and hashtag recommendation. The overall system architecture for hashtag recommendation is shown in Fig. 1.

Four main components for hashtag recommendation are as follows:

(1) Extracting a set of keywords from collected tweets using TF-IDF
(2) Classifying that set of keywords into one of pre-defined classes using Naïve Bayes Classifier
(3) Ranking candidate hashtags
(4) Recommending appropriate hashtags to the user

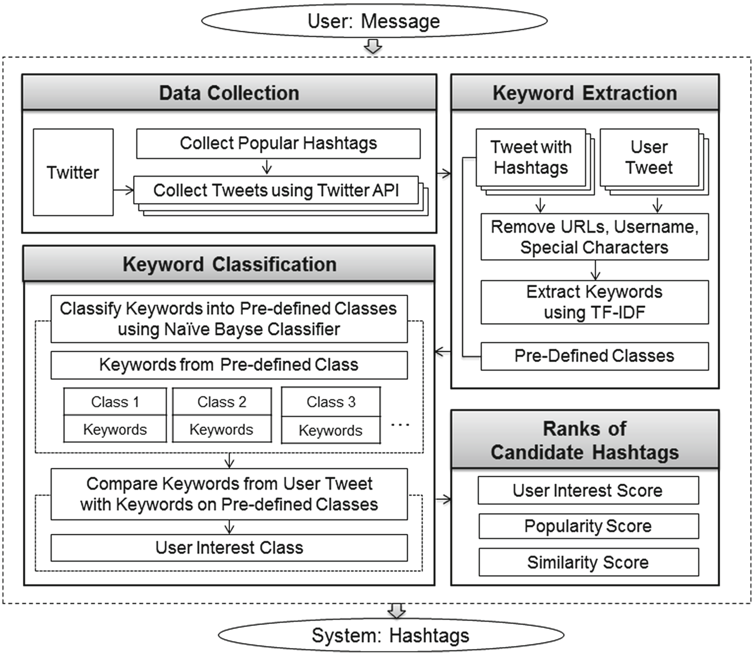

Fig. 1. The overall system architecture

3.1 Data Collection

To build a hashtag recommendation system, we first collected the top 500 most popular hashtags for the year 2013 from Statweestics [7]. By removing non-English and follow-activity hashtags, we have chosen 404 final hashtags with their ranking. Sample hashtag ranks are shown in Table 1. Hashtags are categorized by the preprocessing step into 16 classes defined in [8], which are *art & design, books, business, charity, entertainment, family, fashion, food & drink, funny, health, music, news, politics, science, sports,* and *technology*. Next, we collected 240,000 tweets (OTs) that contain at least one of those hashtags during one week (3-Apr-2014 to 10-Apr-2014) using Twitter Search API. For those tweets, we extract their keywords. For effective extraction of keywords, we remove words starting with the @ character (A username that can be used to send a message), URLs and special characters such as *%!? > $^&<{}.

Table 1. Sample hashtag ranking from Statweestics.com

Ranking	Hashtag
2	#android
18	#oomf
70	#fashion
123	#family
207	#football

3.2 Keyword Extraction

For more effective hashtag recommendation, the first step is to extract keywords from tweets. For that purpose, we use TF-IDF (Term Frequency-Inverse Document Frequency), which is one of the most common weighting methods. This method reflects the importance of each keyword in a tweet and can be defined by the following expression.

$$TFIDF_{t,d} = n_{t,d} \times \log \frac{|D|}{|\{d : t \in d\}|} \tag{1}$$

In the expression, $T = \{t_1, t_2, t_3, \ldots, t_n\}$ and $D = \{d_1, d_2, d_3, \ldots, d_n\}$ indicate the set of keywords *t* (term) and the set of tweets *d* (document), respectively. The result of keyword extraction is shown in Table 2.

3.3 Keywords Classification

Based on the extracted keywords of a tweet, we can decide the class type of the tweet. To do that, we first define the class type of each hashtag in the preprocessing step manually. As a result, each class contains a set of distinct hashtags. Also, by analyzing the keywords of each tweet and its hashtags, we can decide the relationship between keywords and hashtags. As a result, we can get a set of relevant keywords for each class type.

For more accurate hashtag recommendation, we calculate a user interest class by considering all the tweets created by a specific user. By analyzing all the user tweets, we can get a set of keywords and use a Naïve Bayes Classifier [9] to calculate the user interest class from those keywords [10], which is known to be simple and easy to implement. In some types of probability models, Naïve Bayes Classifiers can be trained efficiently in a supervised machine learning. Therefore, we employ this classifier to determine the class of keywords. In this paper, the probability of a tweet t being in class c is computed as follows:

$$C_{map} = \underset{c \in C}{\operatorname{argmax}}\ P(c) \prod_{i=1}^{n_t} P(k_i|c) \tag{2}$$

Here, P(c) is the probability of a document occurring in class c and $P(k_i|c)$ indicates the conditional probability of term k_i occurring in class c. $\{k_1, k_2, \ldots, k_{nt}\}$ are the keywords in t that are part of the vocabulary we use for classification and n_t is the number of such keywords in t. In some cases, keywords that do not exist in the training data might appear in the classification process. This can lead to zero probability and interrupt classification process. To prevent this problem, we use the Laplace smoothing (also called "add one smoothing") which simply adds 1 to the word count. Some of classified hashtags and their keywords are shown in Table 2.

Table 2. Classified hashtags and extracted keywords

Hashtags	Keywords	Class
#coffee, #food, #delicious, …	recipe, dinner, cook, yummi, snack, cake, …	Food & Drink
#nowplaying, #mtvhottest, #music,…	music, musicvideo, femaleartist, radio, …	Music
#oomf, #me, #lol, #cute,…	love, smile, awesome, want, friend,…	Funny

User interest classes play an important role in the personalized recommendation. Therefore, we perform the analysis of user tweets for recommendation based on the assumption that the user would frequently write tweets about the topic that are represented by the user interest class. Hence, by analyzing user tweets and associating their relationship with classes, we can decide the most relevant class, which is the user interest class. Keywords extraction from user tweets are already described in Sect. 3.2.

If the user has not published any tweets before, then we use timeline (following's tweets of the user). Table 3 shows user tweets, their keywords, and calculated user interest class.

Overall, by calculating user interest class of a specific user, our proposed scheme can achieve more effective personalized recommendation on Twitter.

Table 3. Extracted keywords and user interest class

User	User tweets	Keyword	Class
A	Just saying…… #soda #health #nutrition #diet	soda, health, nutrition, diet	Health
	Someone asked, "Do you have any tips for dieting?" My answer: "Step 1, Learn this photo, Step 2: Practice it daily."	tips, dieting, practice, daily	
B	Men's #streetstyle #menswear	men, streetstyle, menswear	Fashion
	Red and black together works! Agree? #fashion	red, black, fashion	

3.4 Ranks of Candidate Hashtags

To recommend hashtags for the user who creates a message more accurately, we further define the ranks of the candidate hashtags for recommendation. Here, candidate hashtags indicate those hashtags that are relevant to the keywords of the tweet. Subsequently, user interest class and the popularity of hashtags are applied. Consequently, top-n recommendations [11] can be provided to the user, where n indicates the number of recommended hashtags presented to the user. For detecting the most suitable top-n hashtags, the candidate hashtags have to be ranked. By detecting the similar keywords which belong to tweets and the most suitable class from the candidate hashtags, we proposed a ranking scheme that is composed of three steps. Their flow chart is shown in Fig. 2.

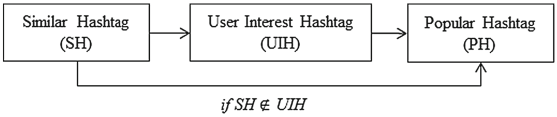

Fig. 2. Flow chart for ranking

(1) Calculating similar hashtags: This is based on the similarity between keywords of hashtags from collected tweets and keywords of user tweets. We detect the similar hashtags which belong to the keyword to candidate hashtags.

$$\text{Similiar Hashtag}(\text{SH}) = \frac{T_{keyword} \cdot H_{keyword}}{\left\|T_{keyword}\right\| \cdot \left\|H_{keyword}\right\|} \tag{3}$$

(2) Finding the class of each similar hashtag: This method is based on classification for user interest class as described in Sect. 3.3. It applies such classification as it analyzes published tweets by the user. If both SH class and UIH class are same, they are included in the candidate hashtags. If not, this step should be ignored and proceed to the next method (PH).

$$\text{User Interest Hashtag (UIH)} = \text{SH}_{class} \cap \underset{class \in c}{\operatorname{argmax}}\, p(class) \prod_{i=1}^{n_t} p(keyword_i | class) \tag{4}$$

(3) Selecting the most popular hashtags: This method reflects the most popular hashtag ranking from Statweestics.com. In other words, it is based on the popularity of candidate hashtags. According to our collected hashtags, we consider those hashtag rankings as addressed in Sect. 3.1.

$$\text{Popular Hashtag (PH)} = \max\left(Hashtag_{popularity}\right) \tag{5}$$

The ranks of the candidate hashtags are important for hashtag recommendation because it can represent user interest and enable to find out more suitable hashtags to the user.

4 Experiments

For each experiment, we use three datasets as shown in Table 4. We use our collected 240,000 tweets (*OTs*) for training dataset as described in Sect. 3.1. Additionally, we collected tweets of 80 users which are 5 users for each class (*UTs*) for similarity between classes with a user interest. We also collected tweets containing 80 new hashtags (*NTs*) that did not appear in the classification process. This dataset is used for evaluating hashtag classification accuracy. Using these datasets, our experiments are described as follows.

Table 4. Description of our datasets

Dataset	Description	Number of data
OTs	Our collected Tweets for training	240,000 tweets
UTs	User Tweets	14,840 tweets from 5 users each class (total 80 users)
NTs	Tweets with New Hashtag	15,360 tweets

4.1 Hashtag Classification Accuracy

To evaluate the hashtag classification, we use *NTs* as a dataset in this experiment and measured how accurately the class of new hashtag is calculated. To do that, we first decide the class of a new hashtag manually and see whether the class from the classification is matched with it. The results are shown in Table 5. After modeling the classifier, it can classify new keywords into proper class by calculating the highest posterior probability. In the classification, a hashtag is classified into a class having the highest posterior probability. The figure shows that the precision is 62.9 % and recall is 71.5 %. Furthermore, we observed that a hashtag can belong to multiple classes. In that case, we decided to consider two classes for each hashtag by calculating two highest posterior probabilities. Precision and recall of such hashtag classification are shown in Fig. 3. In this experiment for the hashtag classification, recall is 77.1 %.

Table 5. The new hashtag classification result

New hashtag	Keyword	Expected Class	Class 1	Class 2
#superbowlxlviii	seahawks, watching, champions, broncos, …	Sports	Sports	Entertainment
#worldcancerday	love,cancer, awareness, strong, …	Health	Health	Family

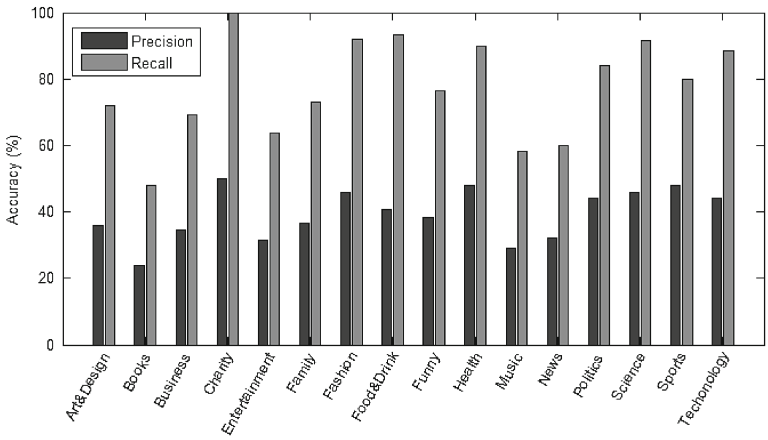

Fig. 3. Precision and recall for hashtag classification

4.2 Similarity Keywords of User Tweet

Users are likely to publish tweets to express their interest. By classifying keywords of user tweets, we can find a user interest class. Accordingly, we used *UTs* and extracted keywords from user tweets and classified them into pre-defined class. To evaluate the performance of the user interest classification, we compared the similarity between keywords of user tweet and a user interest class (i.e. pre-defined class). After that, we measured the similarity between the class based on user interest class and its pre-defined class by Cosine Similarity. The result is shown in Fig. 4. In this figure, *News* class has shown low similarity since *News* class usually covers a variety of topics such as sports, technology, business etc. By performing user interest classification, we achieved an average similarity of 0.67. *Sports* class gives the highest similarity score because tweets related to the sports usually have specific topics.

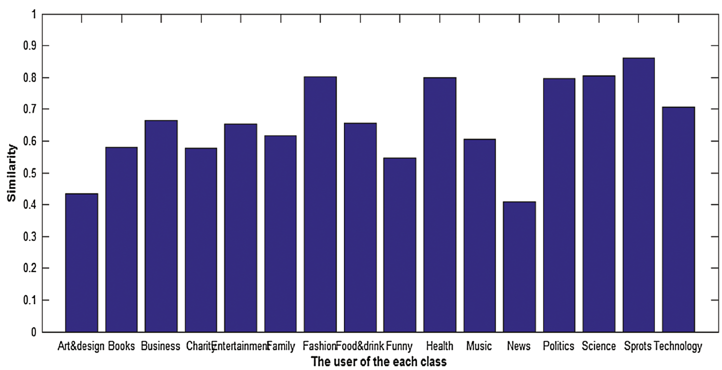

Fig. 4. Similarity between user interest class and its pre-defined class

4.3 Hashtag Ranking and Recommendation

This experiment is based on user tweets (*UTs*). In this experiment, we remove the existing hashtags from user tweets and investigate whether those hashtags are calculated by the recommendation or not. Messages without hashtags are used as the input data for this experiment. We carry out the experiment for ranking methods and top-n recommendation. We consider three ranking steps described in Sect. 3.4. For the evaluation of the recommendation, we considered three ranking methods respectively, Similar Hashtag (*SH*), User Interest Hashtag (*UIH*), Popular Hashtag (*PH*) method and their combination. We calculated precision and recall of the top-n recommendations with n = 1, n = 2, …, n = 5 (the number of recommended hashtags). In this way, we experiment our three ranking methods with top-n hashtags. Precision and recall for top-n recommended hashtags can be seen in Fig. 5. As shown in this figure, the result of precision decreases with an increasing n and recall increases with an increasing n. In addition, our proposed ranking methods are suitable for hashtag recommendations rather than calculating a similarity of hashtag. We determined that hashtags can be recommended up to three, considering an average number of hashtags per message is one or two hashtags in [4]. Among three hashtags, a user can select hashtags via our user interface as described in Sect. 5.

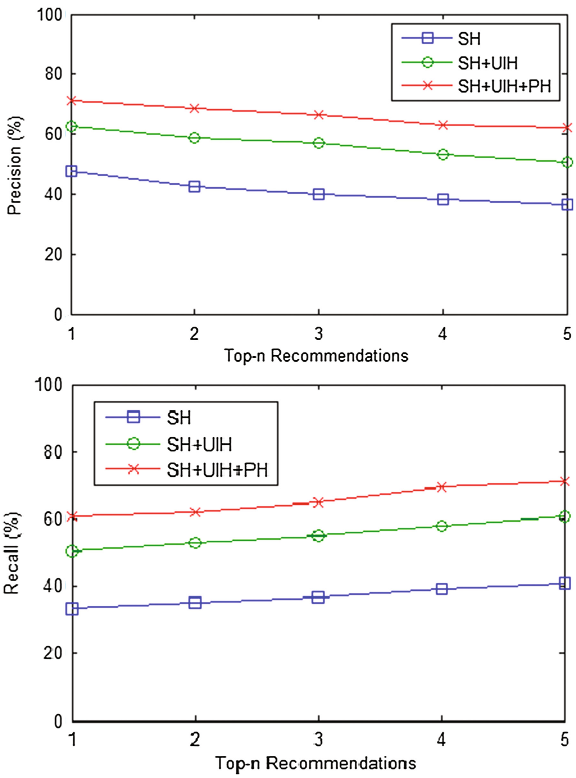

Fig. 5. Precision and Recall for top-n recommendations

5 Application

As shown in Fig. 6, we developed an application named "TWITH" (Twitter Hashtag) for our hashtag recommendation for Android. When a user writes a message without a hashtag, our system recommended and displayed three hashtags as shown on the left side of this figure. After that, a user can select the hashtags on the list and these hashtags are included in the message as shown on the right side of this figure. Then, a user can post a tweet with selected hashtags on Twitter.

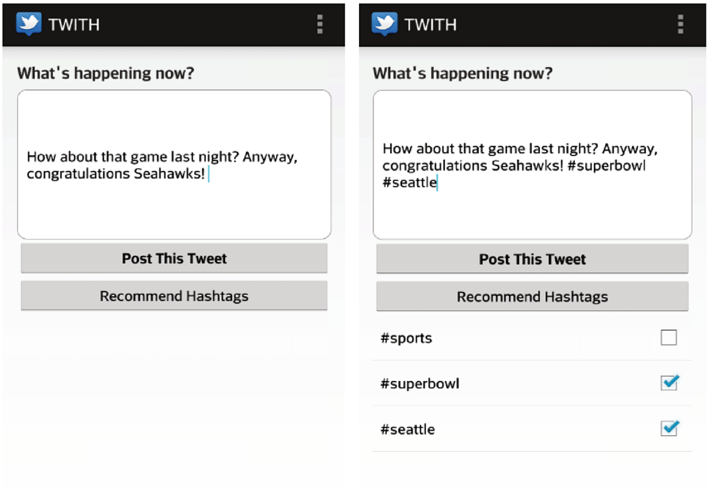

Fig. 6. Screenshot of user application for Android

6 Conclusion

Hashtag is a great tool for organizing information on Twitter. If a tweet is interesting to the users who follow a certain hashtag, then the tweet can be attached with the hashtag. In this paper, we proposed a new hashtag recommendation scheme for user tweets based on user tweet analysis and hashtag classification. To do that, we performed keyword extraction, keyword classification, and hashtag ranking. Especially, the ranks of candidate hashtags are calculated by considering similar tweets, user interest and popularity of hashtags. To show the performance of our scheme, we developed an Android application named "TWITH" and evaluated its recommendation accuracy. In conclusion, by recommending the most suitable hashtags up to three to the user our scheme achieved reasonable performance. This recommendation system can be helpful to the user which wants to know appropriate hashtags of their tweets.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2013R1A1A2012627) and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2014-H0301-14-1001) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224 N Project Report, Stanford, pp. 1–12 (2009)
2. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 1031–1040. ACM, New York (2011)
3. Lee, K., Palsetia, D., Narayanan, R., Patwary, M.M.A., Agrawal, A., Choudhary, A.: Twitter trending topic classification. In: 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW), pp. 251–258 (2011)
4. Zangerle, E., Gassler, W., Specht, G.: Recommending#-tags in twitter. In: Proceedings of the Workshop on Semantic Adaptive Social Web (SASWeb 2011). CEUR Workshop Proceedings, pp. 67–78 (2011)
5. Kywe, S.M., Hoang, T.-A., Lim, E.-P., Zhu, F.: On recommending hashtags in twitter networks. In: Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C. (eds.) SocInfo 2012. LNCS, vol. 7710, pp. 337–350. Springer, Heidelberg (2012)
6. Godin, F., Slavkovikj, V., De Neve, W., Schrauwen, B., Van de Walle, R.: Using topic models for twitter hashtag recommendation. In: Proceedings of the 22nd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp. 593–596 (2013)
7. Statweestics. (http://statweestics.com)
8. Hong, L., Davison, B.D.: Empirical study of topic modeling in twitter. In: Proceedings of the First Workshop on Social Media Analytics, pp. 80–88. ACM, New York (2010)
9. McCallum, A., Nigam, K.: A comparison of event models for Naive Bayes text classification. Presented at the in AAAI-98 Workshop on Learning for Text Categorization (1998)
10. Jeon, M., Jun, S., Hwang, E.: Tweet-based hashtag classification scheme. In: Proceedings of the 2014 International Conference on Platform Technology and Service (PlatCon 2014), pp. 85–86 (2014)
11. Deshpande, M., Karypis, G.: Item-based top-N recommendation algorithms. ACM Trans. Inf. Syst. **22**, 143–177 (2004)