Chapter 12
Learning to Suggest Hashtags: Leveraging Semantic Features for Time-Sensitive Hashtag Recommendation on the Twitter Network

Fahd Kalloubi
Chouaib Doukkali University, Morocco
El Habib Nfaoui
Sidi Mohamed Ben Abdellah University, Morocco

ABSTRACT
Twitter is one of the primary online social networks where users share messages and contents of interest to those who follow their activities. To effectively categorize and give audience to their tweets, users try to append appropriate hashtags to their short messages. However, the hashtags usage is very small and very heterogeneous and users may spend a lot of time searching the appropriate hashtags. Thus, the need for a system to assist users in this task is very important to increase and homogenize the hashtagging usage. In this chapter, the authors present a hashtag recommendation system on microblogging platforms by leveraging semantic features. Furthermore, they conduct a detailed study on how the semantic-based model influences the final recommended hashtags using different ranking strategies. Moreover, they propose a linear and a machine learning based combination of these ranking strategies. The experiment results show that their approach improves content-based recommendations, achieving a recall of more than 47% on recommending 5 hashtags.

INTRODUCTION
Twitter users broadcast short messages of 140 characters called tweets to users who follow their activities, in this form of communication users often use hashtags (the # symbol concatenated with a short character string) to categorize their posts in order to give meaning or to join communities around particular topics. Thus, users tend to use the appropriate hashtags in their tweets which is crucial for the popularity of their messages, in the measure that hashtags can be seen as a way to give some context to the tweet and DOI: 10.4018/978-1-5225-7186-5.ch012
they make tweets easily exploitable by other users in Twitter sphere. Moreover, they make tweets more accessible by hashtag-based search engines such as hashtags.org. 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 (Kywe Su, Hoang, Lim, & Zhu, 2012). The problem of hashtags suggestions can be defined as follow: given a message entered by a user, retrieve the most reliable hashtags from the top-n similar messages. Indeed, some microbloggers try to create their personalized hashtags in a meaningful way to grant that these hashtags should be widely used for the topic but in many cases the truth may be the contrary, many users add the # symbol prefixed to every word in their tweets; wishing that one of these Hashtags could be widely used by others microbloggers for the topic but that may hinders the quality of the tweet, and the users may be taken as tweet spammers, so it is difficult to users to find the appropriate hashtags for their tweets and they may spend a lot of time to retrieve the appropriate ones, so many efforts have been conducted to assist users in this process, some authors interpreted this problems as a recommender system based on collaborative filtering, other authors have proposed an Information retrieval approaches for hashtags suggestions in Twitter sphere by suggesting hashtags from the most similar tweets using term-based models such as TF-IDF (Zangerle, Gassler, & Specht, 2013) (Zangerle, Gassler, & Specht, 2011) (Mazzia & Juet, 2011). Semantic-based features namely the semantic overlap between the tweet and the query computed using DBpedia are shown to be highly effective for tweet search (Tao, Abel, Hauff, & Houben, 2012).

Previous studies in hashtag suggestions/recommendations use lexical matching between tweets to recommend hashtags to users (Zangerle, Gassler, & Specht, 2013) (Godin, Slavkovikj, De Neve, Schrauwen, & Van de Walle, 2013). Term based models are efficient in term of computation performances and the maturity of term weighting theories make this models very rampant. However, term based approaches often suffer from the problems of polysemy and synonymy and very sensitive to term use variation, especially in the context of micro-posts, due to the limited length of this form of communication, his noisy nature and his highly contextualization. Thus, in their approach the authors leverage semantic features to improve the recommended hashtags by harnessing contextual information which was not studied in previous works.

On the whole, the main contributions of the authors of this chapter are:

- They present an approach for hashtags suggestion based on semantic similarity.
- They use a method for named entity linking for the context of tweets by considering the nature of this form of communication.
- They study the impact of different hashtags ranking strategies on our system.
- They present a linear combination of these ranking strategies.
- They incorporate different hashtag ranking strategies in a learning to rank model, and they study the impact of different combinations on the suggested hashtags, which has not been used in the previous methods.
- They evaluate the effectiveness of our approach with a real data set harvested from twitter.
- Their experiment results show that semantic-based similarity, mainly the overlap score between the semantic meanings of tweets outperforms lexical-based similarity, and using learning to rank model by incorporating different ranking functions with semantic similarity leads to a high performance on the suggested hashtags.