Harnessing Semantic Features for Large-Scale Content-Based Hashtag Recommendations on Microblogging Platforms

Fahd Kalloubi, University Sidi Mohamed Ben Abdellah, Fez, Morocco
El Habib Nfaoui, University of Sidi Mohamed Ben Abdellah, Fez, Morocco
Omar El Beqqali, University Sidi Mohamed Ben Abdellah, Fez, Morocco

ABSTRACT

Twitter is one of the most popular microblog service providers, in this microblogging platform users use hashtags to categorize their tweets and to join communities around particular topics. However, the percentage of messages incorporating hashtags is small and the hashtags usage is very heterogeneous as users may spend a lot of time searching the appropriate hashtags for their messages. In this paper, the authors present an approach for hashtag recommendations in microblogging platforms by leveraging semantic features. Moreover, they conduct a detailed study on how the semantic-based model influences the final recommended hashtags using different ranking strategies. Also, users are interested by fresh and specific hashtags due to the rapid growth of microblogs, thus, the authors propose a time popularity ranking strategy. Furthermore, they study the combination of these ranking strategies. The experiment results conducted on a large dataset; show that their approach improves respectively lexical and semantic based recommendation by more than 11% and 7% on recommending 5 hashtags.

KEYWORDS

Content-Based Hashtag Recommendation, Hashtag Recommendation, Information Retrieval, Named Entity Linking, Recommender System, Semantic Similarity

1. 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 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.org¹. 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, 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

DOI: 10.4018/IJSWIS.2017010105

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
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 this paper are:

- The authors present an approach for hashtag suggestions/recommendations 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 their system.
- They study the impact of combining ranking strategies on the hashtags recommendation process.
- To evaluate the effectiveness of their approach they use a real data set harvested from Twitter;
  the experiment results show that their approach improves respectively lexical and semantic based
  recommendation by more than 11% and 7% on recommending 5 hashtags.

The remainder of this paper is organized as follows. In section 2, authors discuss related work
pertaining to hashtags recommendation and retrieval and draw a comparison to their approach;
they describe their dataset in section 3. In section 4, they present their approach for hashtag
recommendations in microblogging platforms, including their semantic similarity, the ranking
functions and their combination. They give the evaluation results of the proposed approach in section
5. In section 6, they conclude this paper and they give future work.

2. RELATED WORK

Twitter has not implemented yet any hashtags recommendation system to assist users to choose the
appropriate hashtags for their posts which is crucial for the popularity of their micro-posts. Many
approaches for tags suggestion in social networking services have been proposed (Sigurbjörnsson &
von Zwol, 2008) (Sood, Owsley, Hammond, & Birnbaum, 2007). In the context of Twitter hashtag
recommendations many approaches have been proposed which can be classified into

two classes.

One class of these approaches focuses on collaborative filtering. Other approaches use techniques
from information retrieval (content-based approaches) to recommend hashtags while there are some
A User-Aware and Semantic Approach for Enterprise Search
www.igi-global.com/article/a-user-aware-and-semantic-approach-for-enterprise-search/210656?camid=4v1a

Semantic Web Services: Towards an Appropriate Solution to Application Integration
www.igi-global.com/chapter/semantic-web-services/35732?camid=4v1a