

A SURVEY OF HASHTAG RECOMMENDATION SYSTEMS FOR STREAMING NEWS

¹Ms. Sneha S. Kadam, ²Mr. G.V.Otari

¹Student, ²Assistant professor

^{1,2}Department of Computer Science and Engineering,

^{1,2}K.I.T's College of Engineering (autonomous) Kolhapur, Maharashtra, India

Abstract: Nowadays social media and news play a major role in our life. We can't think either domain in isolation. While the Twitter-like social media platform plays an important part in the distribution and utilization of news, the news editors depend on a social media platform to pursue their readers' attention and for crowdsourcing the news stories. Twitter is a social information exchange network where a huge number of users exchange messages using a very easy mechanism. Twitter hashtag naturally names and contextualize the story hence it works as a crucial connection between the Twitter crowd and news media. It also helps in grouping the discussion and defining topic trends. Twitter users either use them to categorize the tweets or as mass broadcast media for disasters or for important events like elections. Currently, hashtags are even used for promotion of different brands or for micro-meme discussions. Hashtags are not recorded anywhere even they are not managed by any user, authority or society. Hence some users may find it difficult to choose an appropriate hashtag for their tweet. Hence recommendation system is needed which will suggest proper hashtags to the user. This paper, try to enlist different approaches proposed in the evolution of hashtag recommender systems so far.

Keywords: Twitter, recommender system, social tags, news, real-time hashtag recommendation system

1. INTRODUCTION

Currently, Twitter is a widely used social media network where users are connected with each other to share information using a very easy messaging system. To maintain social connectivity and to gain information generated by others, Twitter users follow other users. Hence in Twitter, the relation between users can be social or informational. In most of the other online social networks, the user-to-user relationship is purely social. By this, we can say that Twitter is a large information exchange platform where users produce, distribute and access information. As every day new users are getting added to Twitter, the amount of tweets generating each day is also high. Having such a huge volume of information present and a huge number of people to interact with can cause information overload problems to Twitter users. Hence different recommender systems are proposed to help users make the proper selection. In Twitter, there can be different recommendation tasks e.g. Followee recommendation, Tweet recommendation, URL recommendation, hashtag recommendation, mention recommendation, retweet recommendation etc. [1]

1.1 Why we need hashtag Recommendation for News?

TWITTER has taken a principal role in the utilization, creation and spreading of news. A current study shows 90% of Twitter users utilize Twitter for reading news and many of them do it for daily. [2]. Majority of the news articles circulating over Twitter have their special identities, known as hashtags. Hashtags contextualize the stories and provide them its identity. Hashtags come into sight immediately around the latest news or newly emerging news stories. News readers can link to a specific story and community using hashtags and by this, they can obtain focused updates in real-time. Many news organizations use hashtags to attract Twitter readers to promote their authentic contents and to engage newsreaders around the story. Many times Journalists initiate the latest hashtags, but it is the Twitter mob that most frequently creates and directs the utilization of many competing hashtags. A news article can have more than one hashtag. Similarly, it can have diverse hashtags at different phases of its

evolution. News articles and hashtags share a very dynamic relationship, as Twitter users create and adopt new hashtags at a rapid speed.

News story development affects the Twitter discussion, which again affects the news hence understanding and capturing the dynamic concurrent development of news and hashtags is crucial for those applications which aim to use hashtagging [3].

2. RELATED WORK:

For hashtag recommendation, the existing approach uses either class/topic modeling or learning-to-rank algorithm.

2.1 Hashtag Recommendation Systems for Tweets:

Initial work for recommending hashtags for tweets is based on Multi-class Classification modeling (MCC) on static datasets.

The approach suggested by the R.Dovgopol et al. in [4], and A. Mazzia and J. Juett in [5] builds SVM classifiers or Naive Bayes for hashtags, Where a hashtag is viewed as a 'class' and the tweets which are tagged with that hashtag are considered as 'labeled data' for that class. When we assume the headline of any news as a tweet, different approaches proposed for hashtag recommendation for tweets can be used for the hashtag recommendation for news as well. A. Mazzia et al. in [5] propose to recommend hashtags with the help of Naive Bayes based on a binomial distribution. The tweet is represented as a binary feature and the most relevant hashtag for that tweet is derived by a maximum likelihood estimate over these binary features.

The paper proposed by Z. Ding et al. in [6] implements an LDA approach to a group of tweets to recommend hashtags. By using the combination of an LDA model and a translation model, the authors acknowledge the vocabulary gap between hashtags and tweets. Many times LDA- type approaches recommend too general hashtags e.g. #life, #news or hashtags which are not frequently utilized by a Twitter user. Due to this, LDA-type approaches encounter many difficulties regarding the accuracy and scalability of recommendation. These models require constant retraining so that they can readjust to newly arising hashtags hence these models are not effective.

For hashtag recommendation, some methods use MCC combined with deep neural networks. Y. Gong et al. suggest an attention-based Convolutional Neural Network model for recommending hashtags for tweets [7]. Here a new attention-based CNN architecture is proposed to use the 'trigger word' mechanism. Performing a variety of Natural language processing tasks with considerable results is achieved only because of the evolution of word embedding based deep neural networks. In this work, a CNN approach is introduced which in turn uses an attention-based mechanism. The process of hashtag recommendation is viewed as a problem of classification. Attention layer is used to calculate the weight of any word by considering its surrounding context. By considering the attention layer, trigger words are selected and the total microblog is converted into constant length vectors. In the end, a layer is constructed which is fully connected with softmax outputs. This approach uses static dataset which enhances the latest hashtag recommendation system results, but for this, precision is only around 50%.

Some methods use pairwise Learning-to-Rank for recommending hashtags for tweets [8]. Proposed work is suitable for only those tweets which contain minimum one URL and one hashtag in it. Such Tweets make a tiny subgroup of the entire tweet pool having news. Caring out training on such tiny and noisy set of tweets can lead to many difficulties for hashtag recommendation which will cause a low percentage of precision. Each day the data collection process is initiated by a collection of 135 hashtags provided by hashtag.org. Because of this, many hashtags used as initial hashtag or seed does not relate to the news article but just occur to be currently popular or in trend at the collection time. This paper doesn't concentrate on news or efficiency of recommendation which are more important in hashtag recommendation.

2.2 Hashtag Recommendation Systems for News:

Not much prior work is available which especially concentrates on hashtag recommendation for news.

The method proposed by S. Sedhai et al. in [9] depends on using user query manually to get relevant news stories. Some Twitter accounts are manually selected and tweets are collected from those accounts. This approach collects news articles and the

tweets related to them over a certain period of time. News related tweets are concatenated depending on the hashtags and a vector representing each of the news topics is created. When target words are provided, news topics are selected which are related to those target words. Using those news topics, a vector is generated which will represent each news topic. This approach then calculates the similarity score between these two types of vectors. A hashtag is recommended to that news topic which will contain the max similarity score. From such collected tweets a hashtag profile is created. In this approach, the involvement of the user with the hashtag is not taken into consideration as all the experiments are carried upon a static collection of data.

B. Shi et al. suggest a high-precision pointwise Learning-to-Rank framework for recommending hashtags for the news [10]. In this approach, they increased the efficiency as well as coverage of Learning-to-Rank method while maintaining high-precision. In this approach, different methods of retrieving related tweets for corresponding news articles are explored.

The approach suggested by O. Phelan et al. in [11] collects tweets and news articles separately, then utilizes co-occurrence words for the article and tweets datasets (i.e., articles, tweets) to relate the news with tweets. At first, this system extracts the latest RSS feed and tweets and then indexes each tweet and article separately to create two Lucene indexes. The final index terms are created from already created RSS indexes and Twitter indexes. These newly created indexes are used as the starting point to create Twitter and RSS term vector. After that, the suggested approach identifies the common set of terms in both term vectors; they create the starting point for the recommendation process. The collection of articles is retrieved along with their TF-IDF score from RSS index using each term as a query. Each generating term is related to a collection of articles which contain that term. The overall score for each article is calculated by summing up the TF-IDF score of all the terms. Articles with maximum terms in it and having high TF-IDF score are preferred. The recommendation is produced by choosing the top articles with the highest scores. The method used here for tweet collection is inappropriate because it might end in a very small or zero intersection with the news dataset.

M. Tsagkias et al. suggest various techniques for creating queries from news articles [12]. For developing different query models, different techniques are used. These techniques are based on three approaches. 1) Internal document formation 2) Clearly connected social media words 3) different methods for selecting terms. This will create ranked list for every approach and then data fusion methods are used to merge these lists. These various techniques which are used to create a query from news articles are useful but this approach is restricted to tweets containing URLs. News article may or may not contain URLs hence proposed approach is not useful for all type of news articles.

T. Gruetze et al. in [13] concentrates on time-related features of hashtag recommendation and suggests two different content-based models executed in a distributed fashion. Since the stress is on content modeling, instead of the relevancy of hashtag, these methods require continuous retraining to keep pace with the fast switch in the news focus.

B. Shi et al. proposed Hashtagger+, a learning-to-rank model for hashtag recommendation for news in real time [14]. The author compared the learning-to-rank approach with the traditional multi-class approaches. In this paper, the author suggests 4 different ways of query generation. The news article for which hashtags needs to recommend is first converted into query. Query summarizes the whole article into a few words. Using that query past articles or tweets are searched from the tweet bag and using these articles a set of tweets related to the article are extracted from the database (past database). After selecting a collection of related hashtags and preparing a feature vector for every article-hashtag combination, previously trained model m produces recommended hashtags using feature vector. This approach improves efficiency as well as coverage of the best and the most modern hashtag recommendation process by introducing new procedures of data collection and feature computation. But this work doesn't consider the semantic aspects of the tweets as it only compares the tweets or articles text-to-text.

3. CONCLUSION:

As the current era is a social media era, news and social media highly depend on each other. Twitter-like Social media platforms play a major role in the distribution of news. Hence the majority of users find it convenient to use social media to

follow their news interest. Social media have the capacity to produce new news and distribute the news. Hashtags perform contextualization of news, they provide certain identity and fame to the particular news. Hence they are widely used to categorize or refer to the news. News can have more than one hashtag in it; also they can have diverse hashtags at various phases of the news story. A Hashtag is a method for the newsreader to link to the particular news story. Hence hashtag recommendation systems are important in connecting the media social and the news media.

Several hashtag recommender systems have been introduced to help the Twitter user to use appropriate hashtags for news and tweets. Our paper summarizes different hashtag recommendation systems for tweet and news. Our survey shows the evolution of hashtag recommendation systems from class/topic modeling approach to Learning-to-rank approach. Though the hashtag recommendation systems for news are evolving with every new approach, still the development in semantic area lags behind with some extent. Many existing systems do the text-to-text comparison between the existing past tweets and the new tweets for which hashtag recommendation is required. Hence more room is available to work with the semantic aspects of the hashtag recommendation systems for news. We aim to study this semantic approach in our future work.

REFERENCES:

- [1] S. M. Kywe, E.-P. Lim, and F. Zhu, "A survey of recommender systems in twitter," in SocInfo, 2012, pp. 420–433.
- [2] T. Rosenstiel, *Twitter and the News: How People use the Social Network to Learn About the World*. Arlington, VA, USA: American Press Institute, 2015
- [3] B. Shi, G. Poghosyan, G. Ifrim, and N. Hurley, "Hashtagger+: Efficient high-coverage social tagging of streaming news," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 43–58, 2018.
- [4] R. Dovgopol and M. Nohelty, "Twitter hashtag recommendation," *CoRR*, vol. abs/1502.00094, 2015, <http://arxiv.org/abs/1502.00094>
- [5] A. Mazzia and J. Juett, "Suggesting hashtags on twitter," EECS 545 Project, 2011, <http://www-personal.umich.edu/~amazzia/pubs/545-final.pdf>
- [6] Z. Ding, X. Qiu, Q. Zhang, and X. Huang, "Learning topical translation model for microblog hashtag suggestion," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 2078–2084.
- [7] Y. Gong and Q. Zhang, "Hashtag recommendation using attention-based convolutional neural network," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2782–2788.
- [8] S. Sedhai and A. Sun, "Hashtag recommendation for hyperlinked tweets," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2014, pp. 831–834.
- [9] F. Xiao, T. Noro, and T. Tokuda, "News-topic oriented hashtag recommendation in twitter based on characteristic co-occurrence word detection," in *12th Int. Conf. Web Eng.*, 2012.
- [10] B. Shi, G. Ifrim, and N. Hurley, "Learning-to-rank for real-time high-precision hashtag recommendation for streaming news," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 1191–1202.
- [11] O. Phelan, K. McCarthy, and B. Smyth, "Using twitter to recommend real-time topical news," in *Proc. 3rd ACM Conf. Recommender Syst.*, 2009, pp. 385–388.
- [12] M. Tsagkias, M. de Rijke, and W. Weerkamp, "Linking online news and social media," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 565–574.
- [13] T. Gruetze, G. Yao, and R. Krestel, "Learning temporal tagging behaviour," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 1333–1338.
- [14] B. Shi, G. Poghosyan, G. Ifrim, and N. Hurley, "Hashtagger+: Efficient high-coverage social tagging of streaming news," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 43–58, 2018.