Hashtag Recommendation

Hashtag recommendation suggests hashtags to users while they write microblogs in social media platforms. Although researchers have investigated various methods and factors that affect the performance of hashtag recommendations in Twitter and Sina Weibo, a systematic review of these methods is lacking.

1. Introduction

Social media platforms have become fast-growing and influential media that enable people to communicate with each other easily, share information and search for exciting topics. The number of social media users was 3.6 billion in 2020, and this number is predicted to increase to 4.41 billion users in 2025 (https://www.statista.com/statistics/278414/ (accessed on 10 May 2021)). Twitter (https://about.twitter.com/company (accessed on 10 May 2021)) is a microblogging social media platform that permits users to write and share short messages of 280 characters or less, including hashtags, mentions and URLs. These types of short messages are referred to as "microblogs" and "tweets" [1][2][3][4][5][6][7]. Founded in 2006, Twitter has quickly become an increasingly popular and powerful tool worldwide. According to the Internet Live Stat (http://www.internetlivestats.com/twitter-statistics/ (accessed on 10 May 2021)), 500 million messages on average are posted per day by 330 million active users. In July 2020 (https://www.statista.com/statistics/242606/ (accessed on 10 May 2021)), the United States had the largest audience size, with 62.55 million users, followed by Japan with 49.1 million users, and India ranked third with 17 million users. Sina Weibo (https://www.statista.com/statistics/795303/china-mau-of-sina-weibo/ (accessed on 10 May 2021)), the Chinese equivalent of Twitter, had around 523 million active users in the same year.

With the information overload and increase in technology dependency, social recommendations have become a key research area. Social recommendation systems can be defined as techniques or algorithms that automatically suggest the most relevant and interesting data to social media users. Hashtag recommendation is a branch of the social recommendation systems that proposes contemporary and relevant hashtags to users as they type tweets [5][6][7][8][9][10][11][12][13][14][15][16][17][18][19][20][21][22][23]. Choosing the correct hashtag has several benefits: it enables the user to quickly join a discussion and read tweets written by other users [24]. Using hashtags gives the user a chance for their tweets to be noticed and reach a wider audience [25]. Hashtags also help researchers to analyze users' behaviors or to predict the outbreaks of natural disasters and epidemics [26]; they are also helpful for companies to advertise their products and improve customer services and support through users' complaints and comments [27]. In politics, politicians can communicate with the public and advertise their campaigns [28]. Moreover, people can raise and share their voice nationally and globally [29]. For Twitter and Sina Weibo, recommending hashtags helps to enhance discussion as users are guided to use more accurate and relevant hashtags. Adopting the right hashtags helps Twitter/Sina Weibo to eliminate insignificant and noisy hashtags and reduce information overload. Automatically recommending personalized hashtags to users also helps them save time and effort in searching for relevant hashtags.

Recommending hashtags by analyzing tweets and extracting information from the Twitter/Sina Weibo hashtag universe can be a very challenging task. One of the challenges is that these tweets and hashtags are *user-generated*. Users tend to use informal language when writing their tweets; for example, users use "4U" to mean "for you", "AMA" for "ask me anything" and "BFN" for "bye for now". Spelling and grammatical mistakes are not checked or corrected. Short texts are therefore more difficult to analyze than long texts. The facts that tweets are short texts and are noisy add extra complication to the data. Furthermore, hashtags can be acronyms, shortened or misspelled words or a combination of words, numbers and punctuation marks. Thus, using hashtags as keywords does not necessarily convey the meaning of the discussion. The lack of control over the creation of hashtags has resulted in hundreds of hashtags being associated with a single discussion topic and different discussion topics being associated with a single hashtag.

Hashtag recommendation can be either general, when the suggested hashtags are obtained based on the data of all users, or personalized, when the suggested hashtags incorporate the user's preferences and data. The hashtags from all the tweets in a dataset form a space known as the "hashtag space". The suggested hashtags are said to be novel if they are not in this hashtag space (i.e., not previously used by other users). Otherwise, they are said to be predefined.

2. A New Taxonomy for Hashtag Recommendation of Tweets

The taxonomy classifies hashtag recommendation methods for tweets into three main categories: text-based, hybrid user-based and hybrid miscellaneous methods. Text-based methods find hashtags similar to what a user intends to adopt based on the textual information. This category is further classified into tweet-similarity-based methods, probabilistic methods, classification based methods, graph-based methods and matrix factorization based methods. Since methods of collaborative filtering suffer from the cold-start problem, they are integrated with other methods. Hybrid user-based hashtag recommendation methods recommend hashtags based on the similarity of the users' behavior, interests or relations. This category is further classified into behavioral and social collaborative filtering methods. Hybrid miscellaneous hashtag recommendation take advantage of multi-modalities and multi-factors to recommend the hashtags. Regardless of the specific techniques employed, it has become clear that the best outcome can be achieved using the hybrid methods (user-based or miscellaneous) for their ability to overcome problems occurring with content-based and collaborative filtering methods. It was noticed that understanding various factors that affect the performance of hashtag recommendation and the underlying assumptions have a significant impact on the algorithmic approach that should be considered.

We highlight some open challenges, which can be considered future research directions. These challenges are as follows:

- Despite the advancement of the current methods, further improvements are required to propose more effective methods that are less expensive in terms of time and computation and provide a personalized recommendation that covers a broader range of pre-defined and novel hashtags with higher accuracy. Furthermore, most of the previous research was tested offline. Recommending personalized hashtags in real-time is more difficult where the recommended hashtags need to be accurate and given instantly.

- As an extension to work presented in Alsini et al.'s paper [23], the association of the four networks and their combined effect on the performance of hashtag recommendation can be examined. In addition, rather than considering the mutual tie relationships between users, weighted relationships can be used to construct the networks and detect communities.

- It is challenging to compare newly proposed methods with baseline methods due to the variance in the size of the datasets (i.e., number of tweets, users, and hashtags). It is recommended for future research papers to set a minimum size of the dataset for evaluation.

- Accuracy-based metrics were the primary measures of evaluation for a long time. In recent years, *concepts of evaluation*, which are metrics beyond accuracy, have been studied to evaluate the value of the traditional recommendations. For example, *diversity* is concerned with the variety of items recommended by the system, and *novelty* is concerned with how the recommended items are new to users [30][31]. However, concepts of the evaluation were rarely used to evaluate hashtag recommendation methods. The value of the recommendations also needs to be studied in terms of user satisfaction and expectation.

- With the dynamic nature of social media platforms, studies of hashtag recommendation should focus more on the automatic update of the data on the recommendation.

References

1. Ding, Z.; Zhang, Q.; Huang, X. Automatic Hashtag Recommendation for Microblogs using Topic-Specific Translation Model. In Proceedings of the COLING 2012: Posters, The COLING 2012 Organizing Committee, Mumbai, India, 8–15 December 2012; pp. 265–274.

2. Ding, Z.; Qiu, X.; Zhang, Q.; Huang, X. Learning Topical Translation Model for Microblog Hashtag Suggestion. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI '13, Beijing, China, 3–9 August 2013; pp. 2078–2084.

3. Gong, Y.; Zhang, Q.; Huang, X. Hashtag Recommendation Using Dirichlet Process Mixture Models Incorporating Types of Hashtags. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal, 17–21 September 2015; pp. 401–410.

4. Gong, Y.; Zhang, Q.; Han, X.; Huang, X. Phrase-based hashtag recommendation for microblog posts. Sci. China Inform. Sci. 2016, 60, 012109.

5. Song, S.; Meng, Y.; Zheng, Z. Recommending Hashtags to Forthcoming Tweets in Microblogging. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 1998–2003.

6. Yu, J.; Zhu, T. Combining long-term and short-term user interest for personalized hashtag recommendation. Front. Comput. Sci. 2015, 9, 608–622.

7. Zhang, Q.; Gong, Y.; Sun, X.; Huang, X. Time-aware Personalized Hashtag Recommendation on Social Media. In Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 23–29 August 2014; pp. 203–212.

8. 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, Girona, Spain, 22–26 June 2011; pp. 67–78.

9. Khabiri, E.; Caverlee, J.; Kamath, K.Y. Predicting Semantic Annotations on the Real-Time Web. In Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT '12, Association for Computing Machinery, New York, NY, USA, 25–28 June 2012; pp. 219–228.

10. Chen, C.; Yin, H.; Yao, J.; Cui, B. TeRec: A Temporal Recommender System over Tweet Stream. Proc. VLDB Endow. 2013, 6, 1254–1257.

11. Ma, Z.; Sun, A.; Yuan, Q.; Cong, G. Tagging Your Tweets: A Probabilistic Modeling of Hashtag Annotation in Twitter. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, 1 November 2014; pp. 999–1008.

12. Jeon, M.; Jun, S.; Hwang, E. Hashtag Recommendation Based on User Tweet and Hashtag Classification on Twitter; Springer: Berlin/Heidelberg, Germany, 2014; pp. 325–336.

13. Feng, W.; Wang, J. We can learn your #hashtags: Connecting tweets to explicit topics. In Proceedings of the 2014 IEEE 30th International Conference on Data Engineering, Chicago, IL, USA, 31 March–4 April 2014; pp. 856–867.

14. Al-Dhelaan, M.; Alhawasi, H. Graph Summarization for Hashtag Recommendation. In Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud, Rome, Italy, 24–26 August 2015; pp. 698–702.

15. Zhang, Q.; Wang, J.; Huang, H.; Huang, X.; Gong, Y. Hashtag Recommendation for Multimodal Microblog Using Co-Attention Network. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, Melbourne, Australia, 19–25 August 2017; pp. 3420–3426.

16. Alsini, A.; Datta, A.; Li, J.; Huynh, D. Empirical Analysis of Factors Influencing Twitter Hashtag Recommendation on Detected Communities. In Proceedings of the Advanced Data Mining and Applications—13th International Conference, ADMA 2017, Singapore, 5–6 November 2017; pp. 119–131.

17. Li, Y.; Jiang, J.; Liu, T.; Qiu, M.; Sun, X. Personalized Microtopic Recommendation on Microblogs. ACM Trans. Intell. Syst. Technol. 2017, 8.

18. Kowald, D.; Pujari, S.C.; Lex, E. Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach. In Proceedings of the 26th International Conference on WWW, International World Wide Web Conferences Steering Committee, Geneva, Switzerland, 1 April 2017; pp. 1401–1410.

19. Alsini, A.; Datta, A.; Huynh, D.Q.; Li, J. Community Aware Personalized Hashtag Recommendation in Social Networks. In Data Mining; Islam, R., Koh, Y.S., Zhao, Y., Warwick, G., Stirling, D., Li, C.T., Islam, Z., Eds.; Springer: Singapore, 2019; pp. 216–227.

20. Ma, R.; Qiu, X.; Zhang, Q.; Hu, X.; Jiang, Y.G.; Huang, X. Co-attention Memory Network for Multimodal Microblog's Hashtag Recommendation. IEEE Trans. Know. Data Eng. 2019.

21. Belhadi, A.; Djenouri, Y.; Lin, C.W.; Cano, A. A Data-Driven Approach for Twitter Hashtag Recommendation. IEEE Access 2020, 8, 79182–79191.

22. Javari, A.; He, Z.; Huang, Z.; Jeetu, R.; Chen-Chuan Chang, K. Weakly Supervised Attention for Hashtag Recommendation Using Graph Data. In Proceedings of the Web Conference 2020, WWW '20, Taipei, Taiwan, 20–24 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1038–1048.

23. Alsini, A.; Datta, A.; Huynh, D.Q. On Utilizing Communities Detected From Social Networks in Hashtag Recommendation. IEEE Trans. Comput. Soc. Syst. 2020, 7, 971–982.

24. DeMasi, O.; Mason, D.; Ma, J. Understanding Communities via Hashtag Engagement: A Clustering Based Approach Authors. In Proceedings of the International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016.

25. Laniado, D.; Mika, P. Making Sense of Twitter. In The Semantic Web—ISWC 2010; Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 470–485.

26. Chowdhury, J.R.; Caragea, C.; Caragea, D. On identifying hashtags in disaster twitter data. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020.

27. Xiao, F.; Noro, T.; Tokuda, T. News-Topic Oriented Hashtag Recommendation in Twitter Based on Characteristic Co-occurrence Word Detection. In Web Engineering; Brambilla, M., Tokuda, T., Tolksdorf, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 16–30.

28. Jungherr, A.; Schoen, H.; Jürgens, P. The Mediation of Politics through Twitter: An Analysis of Messages posted during the Campaign for the German Federal Election 2013. J. Comput. Med. Commun. 2015, 21, 50–68.

29. Ince, J.; Rojas, F.; Davis, C.A. The social media response to Black Lives Matter: How Twitter users interact with Black Lives Matter through hashtag use. Ethn. Rac. Stud. 2017, 40, 1814–1830.

30. Silveira, T.; Zhang, M.; Lin, X.; Liu, Y.; Ma, S. How good your recommender system is? A survey on evaluations in recommendation. Int. J. Mach. Learn. Cybernet. 2017, 10.

31. Ziegler, C.N.; McNee, S.M.; Konstan, J.A.; Lausen, G. Improving Recommendation Lists through Topic Diversification. In Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, 10–14 May 2010; Association for Computing Machinery: New York, NY, USA, 2005; pp. 22–32.