User Modeling on Twitter with Exploiting Explicit Relationships for Personalized Recommendations

Abdullah Alshammari[1], Stelios Kapetanakis[1], Roger Evans[1], Nikolaos Polatidis[1]
Gharbi Alshammari[1]

[1]School of Computing, Engineering and Mathematics, *University of Brighton*, Brighton, United Kingdom
(A.Alshammari1, S.Kapetanakis, R.P.Evans, N.Polatidis, G.Alshammari)@brighton.ac.uk

Abstract. The use of social networks sites has led to a challenging overload of information that helped new social networking sites such as Twitter to become popular. It is believed that Twitter provides a rich environment for shared information that can help with recommender systems research. In this paper, we study Twitter user modeling by utilizing explicit relationships among users. This work aims to build personal profiles through a alternative methods using information gained from Twitter to provide more accurate recommendations. Our method exploits the explicit relationships of a Twitter user to extract information that is important in building the user's personal profile. The usefulness of this proposed method is validated by implementing a tweet recommendation service and by performing offline evaluation. We compare our proposed user profiles against other profiles such as a baseline using cosine similarity measures to check the effectiveness of the proposed method. The performance is measured on an adequate number of users.

Keywords: recommender systems, user modeling, user profiling, explicit relationships, Twitter, influence score.

1 Introduction

The real-time web is growing as a new technology that allows users to communicate and share information in multi-dimensional contexts such as Twitter. Twitter is a very well-known social media micro-blogging platform used by hundreds of millions of users. It is a social network that allows users to post and exchange short messages, called tweets, of up to 280 characters [20]. It has become a very important source of shared information and breaking news, and it works quickly and effectively. Also, it can be considered an example of a distinctive type of social networking platforms that presents relationships based on the following strategy, which makes it different from other classic social networking platforms that based on reciprocal network structure such as Facebook. The relationships among users of Twitter can be social or informational or both because users follow other users not only to maintain social links but also to receive interesting information generated by others. Twitter features lead to the possibility of using it as a primary source for modeling a specific user who participates in a network characterized by relationships and interactions [1, 20].

According to [1], some studies have shown that Twitter is seen as a valuable resource for many powerful approaches such as recommender systems. Recommender systems have proven to be powerful parts of many web and mobile applications. The main goal of such systems is to provide real-time, context-aware, personalized information to help increase sales and user satisfaction. Many studies have used Twitter to model users, build user profiles, and deliver accurate recommendations.

In this paper, we investigate modeling Twitter users by exploiting relationships to improve recommendations in recommender systems based on short-text-based profiles within short-term (recent tweets such as tweets within the last 2 weeks). User profiles in our approach are built from tweets of the user's following list (i.e., friends). Our model redefines the rule of influence and use it to identify influential friends and use their tweets to build the user profile by examining all incoming tweets within short-term time. As a result, our approach can deliver more accurate recommendations to users in comparison with profiles from the literature.

2 Related work

The authors of [12] proposed a method to improve the accuracy of recommended news articles based on Twitter tweets. The user profile was built by extracting the nouns in the user's tweets and re-tweets. Tweet refers to the action of writing a post whereas re-tweet refers to the action of sharing a tweet of another user. Their results showed that Twitter-activity-based recommendations were more accurate than the random recommendations. The TRUPI system was proposed in [9], combining the history of the user's tweets and the social features. It also catches the dynamic level of user's interests in several topics to measure changes in the user's interests over time.

In [1], the authors analyzed temporal dynamics in Twitter profiles for personalized recommendations in the social web. They build two different types of profiles, based on hashtags and entities (e.g., places and celebrities), taking into consideration time-sensitivity, enrichments (using external resources such as Wikipedia), and the activity of the user. The results showed that the entity-based profile, which was built with short-term time (i.e., last two weeks) and enrichment, performed better than other profile types in the news recommender system based on Twitter activities. Furthermore, the problem is that for many users there is not enough data in their recent activities to create a reliable user profile. Authors in [15] showed that using a decay function in long-term profiles that gives higher weight to recent topics of interest than older topics of interest showed better performance in delivering recommendations than the long-term profiles without the decay function. Moreover, authors in [2] proved that a short-term profile (last week, for example) is better than the complete profile.

One of the solutions is to enrich user profile with other data. The authors in [2] modeled user profiles in Twitter with different dimensions and compared the quality of each one and enrichments was one of the dimensions. Results showed that using external resources, such as news articles, is better than relying solely on Twitter.

Enriching user profile with data have been done in different ways: exploiting URLs in tweets or using textual external resources (such as articles or Wikipedia). In exploiting URLs in tweets, a CatStream system was proposed in [10] that uses traditional classification techniques to profile Twitter users based on URLs in their tweets. However, the system only focuses on URLs in tweets and it is not suitable for users who do not have enough tweets that include URLs. In [3], the authors categorized a set of tweets as interesting or uninteresting by using crowdsourcing. The method showed that the existence of a URL link is a successful feature in selecting interesting tweets with high accuracy. However, the shortcoming of this factor is that it might incorrectly categorize an uninteresting tweet that links to useless content [11].

In using external resources, authors in [1, 10] used external resources such as Wikipedia and news articles. The enriched user profiles with external resources performed better than profiles that built with only Twitter activities. These methods can be useful to supply the user profile with more information and thus improve the accuracy of the recommender system. However, some data gained from external resources might have no relevance to the user's interests and it might affect the performance of the recommender system. Also, many users do not provide enough URL links in their tweets.

A field that remains to be investigated is exploiting the network of relationships between users in Twitter to characterize a specific user and improve the performance of recommender systems based on short-text activities. It is obvious that a single user who generates short-text data (i.e., tweets and retweets) can be characterized based on his history and behavior by collecting history data (i.e., timeline) that the user himself has generated. However, to acquire sufficient information for profiling, this method may need to go a long way back into the past and the collected information might not form a very coherent set and up to date. Furthermore, the problem is that for many users there is not enough data and URLs in their recent activities to create a reliable profile. To address this lack of data, we propose using explicit links between users (e.g., following links) to expand the set of relevant recent activities. The advantage of this method is that there is more recent data to build profiles from, which will allow us to improve the performance of short-text-based recommender systems.

Exploiting following links can be achieved by finding influential users in the friends list. Most of the researches focuses on how influential a user is based on his popularity, as indicated by the number of followers and followees (friends) and his interactions with others [4, 5, 8, 21]. Authors in [18] collected and classified different Twitter influence measures in the literature. However, from the perspective of a normal user who follows other identified influential users, it is believed that the influence rule can vary from one user to another in the list of followees (friends). For example, some accounts followed by a user, such as close and active friends, that are not identified as influential by the influence rule, but they might be more influential than the identified influential users such as celebrities. In [18] authors stated that naming a person in a social network as an influential is a conceptual problem and there is no agreement what the influential user should be. Therefore, there is a need to create a method that can generate an influence score from the user's perspective relating to the user's behavior and interactions.

After identifying the influence score, there is a need to classify incoming tweets into categories such as relevant and not relevant. Researchers proposed techniques to predict the likelihood of a tweet being retweeted based content-based features [14], collaborative tweet ranking (CTR) [7] and a coordinate ascent (CA) algorithm [19].

3 Proposed Method

In general, the recommender system includes two stages: (1) user profiling and (2) item ranking. In this work, user profiles are built by extracting information from tweets in the user's timeline and tweets by his following list (friends). Then, recommendation items are ranked based on the user profile. Figure 1 shows the general process of our proposed work. The user's information is collected using the Twitter API. This collected information is processed to identify important keywords posted by the user's friends. The following subsections explain in detail the steps of the two stages.

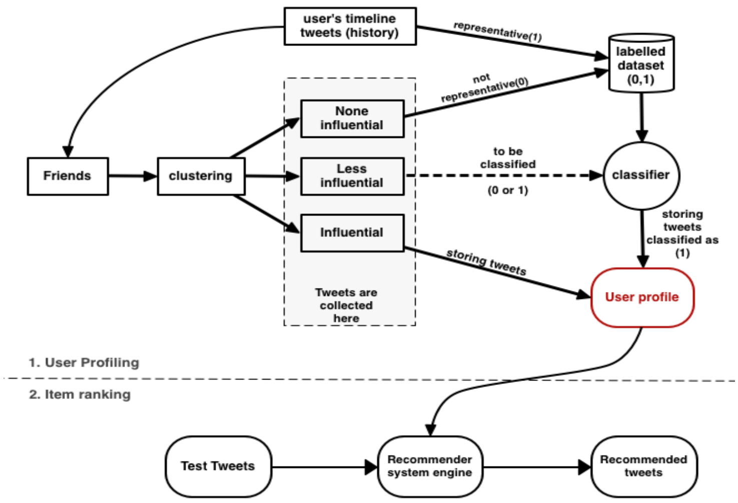


Fig. 1. The general steps of the proposed method.

3.1 User Profiling Stage

In this stage, we develop user profiles that contains important information about the user. These profiles are built from tweets by other users with explicit relationships to the user in Twitter (e.g., user's friends list). All profiles are built as key-words profiles (bag of words); pre-processing must be carried out before the recommendation process based on the steps suggested by Micarelli et al. [13]. The aim is to filter tweets and then extract important content. The tweets generated by the user himself and re-tweeted tweets explicitly represent the user's interests, whereas the received and collected tweets from explicit links need to be examined and classified.

Before building the profiles, and for each user, data about the user is collected from the user's Twitter timeline, including friends list, timeline tweets (posted and re-tweeted tweets), and favorited tweets. Then, we calculate an influence score between the user and his/her friends to help us rank his/her friends based on their importance and then choose and collect appropriate content. The influence score, which considers actions such as following, re-tweeting, replying, and favoriting, might be a good attribute to find friends who are influential to the user. The influential users, who are followed by the examined user, are found by using the influence score from the user's perspective. The following simple Eq. (1) calculates the influence score, where the original user, denoted user1 (u1) follows user2 (u2):

$$\text{Influence Score (u1, u2)} = \left(\frac{\sum \text{RTs (u2)}}{\sum \text{RTs p(u1)}} + \frac{\sum \text{RTs (u2)}}{\sum \text{Ts p(u2)}} + \frac{\sum \text{MT(u2)}}{\sum \text{MT p(u1)}} + \frac{\sum \text{FV(u2)}}{\sum \text{FV p(u1)}}\right) \times \frac{1}{4} \quad (1)$$

∑RTs (u2) represents the total number of tweets posted by user2 and re-tweeted by user1. ∑RTs p(u1) represents the total number of re-tweets in the user1 profile. ∑Ts p(u2) represents the number of tweets in the user2 profile. ∑MT(u2) represents the number of replies (mentions) that user1 posted to user2. ∑MT p(u1) represents the total number of mentions in user1's profile. ∑FV(u2) represents the total number of tweets from user2 that user1 has favorited. ∑FV p(u1) represents the total number of favorited tweets in user1's profile. Finally, 1/4 is used to normalize the score between 0 and 1.

The friends list is divided into three groups based on influence score: influential users, less influential users, and non-influential users, by K-means clustering algorithm.

All tweets from the influential group are added to the user profile whereas tweets from the non-influential group are not added. The tweets from less influential users are classified into representative (re-tweetable) or not representative (not re-tweetable). Different classifiers are used in this process: Naïve Bayes, Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbor and Neural Networks. All classifiers trained using the same labeled dataset from the user's timeline (the history of the user's tweets and re-tweets) and the tweets of the non-influential users. The tweets from the user timeline are labeled as representative (re-tweetable) and the tweets from non-influential users are labeled as not representative (not re-tweetable). The dataset is divided into training and testing sets. The latter is used to calculate the accuracy of each classifier. The classifier with the highest accuracy is chosen automatically to classify less influential user tweets. This step ensures that tweets are classified with the most accurate classifier. After classifying less influential users' tweets, the tweets classified as representative (re-tweetable) are stored in the user's profile with influencers' tweets.

3.2 Items Ranking Stage

At this stage, the recommendation items are a set of tweets that the user might show some interest in by re-tweeting. Vector space model representation is used, which considers user profile and recommendation items as vectors and then calculates the angle between them. The closer the item is to the user profile, the more relevant it is. Cosine similarity is used here to calculate the angles.

The user profiles built as described in the previous subsection are used as the base for ranking the set of tweets in the recommendation items. Every tweet in the recommendation items is evaluated based on its similarity to the user profile. This process is applied following the steps of [13] and additionally, tweets are excluded if the remaining text contains fewer than three words. Finally, the similarity score between the tweet profile and the user profile is calculated by the cosine similarity equation. All the recommendation items are ranked, and top-k tweets are recommended to the user.

4 Experiment and Results

To validate the advantages of our proposed method, a tweets recommender system was implemented, and an offline evaluation was performed with an adequate group of users. Using the Twitter API provided in the development section of Twitter's website, the timelines of 29 randomly chosen users were collected and examined. For the recommendation items, test tweets (recommendation items) in which the user had shown some interest by re-tweeting were collected from the user's timelines. The next subsection will explain how these test tweets were collected.

4.1 Experiment Setup

After collecting the timeline of the examined user and before calculating the influence score and clustering users into the three mentioned groups, the dataset is divided into three time periods as it is seen in Figure 2.

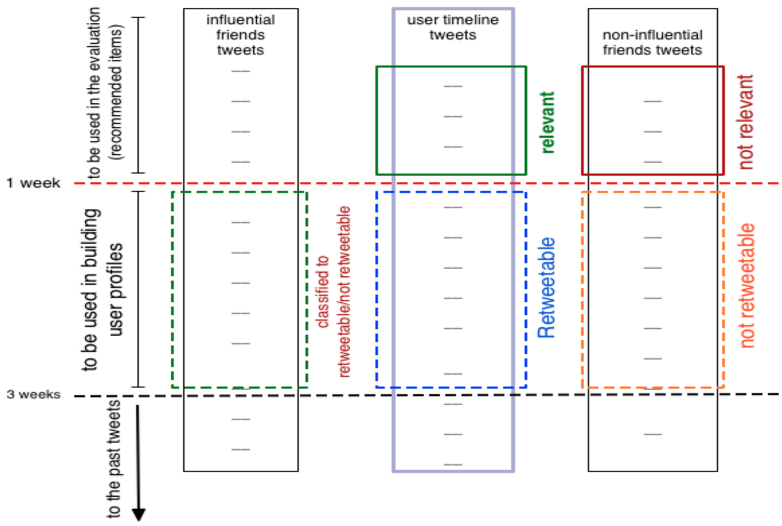


Fig. 2. Dividing the user timeline into three evaluation time periods.

The tweets in the first period (the previous week) are used as test items (test tweets); it is already known that the user showed interest in them by re-tweeting them. This is like going back on a time machine into the past to predict a future that is already known, which can help in the evaluation process. The tweets from the second period (between 1 week ago and 3 weeks ago) are used to build user profiles from different sources. The third period (more than 3 weeks ago) is used along with the second period to calculate the influence score from the timeline of the examined user. Also, the user' timeline in this period is used in the machine learning classification.

Profiles: For each user of the examined users, and after calculating the influence score between the examined user and his friends and classifying them as influential, less influential, and non-influential, tweets are collected from influential and less influential friends to build different profiles that are examined and compared in the experiment. These profiles are:

- Baseline: includes all tweets from the user's timeline. It contains both the tweets that the user posted and the tweets the user re-tweeted.
- BLCinf: includes all tweets from the user's timeline, short-term tweets by influential friends (second time period) and short-term tweets by less influential friends that are classified as representative (second time period).
- STBLCinf: includes tweets only from the second time period (short-term), and includes tweets from the user's timeline, tweets by influential friends, and tweets by less influential friends that are classified as representative.
- STBLinf: includes tweets only from the second time period (short-term) and includes tweets from the user's timeline and all tweets by influential and less influential friends without consideration of classification.
- BLinf: includes all tweets from the user's timeline and all short-term tweets by influential and less influential friends. Classification of the tweets of less influential friends is not considered in this profile.
- Followers count: this profile is built in the same way of STBLCinf. In contrast, the clustering is applied based on followers count instead of influence score. Some literature researches used it in their experiments such as [5, 6, 16, 18, 21].

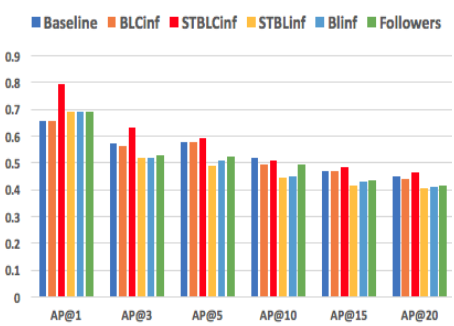
Test tweets: Test tweets are used to evaluate the accuracy of the recommender system. They are a collection of tweets from the first period (1 week ago) as explained previously, and they are used as recommendation items in the recommender system. This collection contains both relevant and non-relevant tweets. From the timeline of the user, it is already known which items the user showed some interest in by re-tweeting, and these are considered relevant items. Tweets from non-influential users in the same period are used as non-relevant items. Thus, the set of the recommended items contains a mixture of relevant and non-relevant items, which can help the recommender system to measure the accuracy when it runs on different user profiles. It also allows a comparison between the built profiles and the baseline profiles.

Evaluation metrics: In this study, offline evaluation is used to measure the accuracy of the recommender system with different user profiles [19]. As in our methodology, various user profiles are used in the recommender system and then compared. The metrics are used in this research to measure the accuracy of the performance of the system are average precision @ k (AP) and mean average precision (MAP). Average precision is used to measure how good the system is in retrieving top-k relevant items and MAP is used to measure how good the system in retrieving all relevant items.

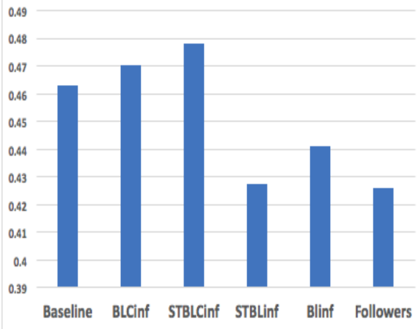
4.2 Results

In the metric of the Average Precision @top-k recommendations, the tested values of k are: 1, 3, 5, 10, 15 and 20. (See Figure 3a).

In the top-k = 1, 3 and 5, results showed that the profile STBLCinf outperformed all other strategies of building user profiles. On the other hand, the baseline and BLCinf profiles has the lowest average precision on top-1.



(a) Average Precision



(b) Mean Average Precision (MAP)

Fig. 3. The average precision @1, 3, 5, 10, 15 and 20, and the Mean Average Precision (MAP) of profiles.

In the top-k = 10, the results showed that the baseline outperformed all profiles. Whereas the STBLCinf came in the second place. On the other hand, STBLinf achieved the lowest Average Precision.

When the top-k is set to 15 and 20, profile STBLCinf outperformed all other profiles. However, again profile STBLinf achieved the lowest average precision, which might mean that non-relevant tweets affected the accuracy of the recommender system. STBLCinf and STBLinf were built similarly but the difference is that the latter included all tweets from influential and less influential users. Therefore, this small difference can affect the performance and make it the worst profile.

Table 1. Average number of tweets in each profile.

	Baseline	BLCinf	STBLCinf	STBLinf	BLinf	Followers
Average	1991	4195	2491	4335	6039	1098

In the Mean Average Precision (MAP), Figure 3b shows the results that the profile STBLCinf achieved the highest performance against all profiles. Also, profile BLCinf achieved better performance than the baseline and that might mean enhancing the base-line profiles with some related data can improve the performance of the recommender system. Also, profile BLinf was built similarly to BLCinf achieved less MAP than the latter and the baseline profile. This may clarify that enriching profile with none related data (unclassified tweets in this case) can reduce the quality of recommendations even worse than the baseline. Also building profiles based only on timeline (baseline) cannot deliver more relevant recommendations. On the other hand, profiles STBLinf and followers achieved the lowest MAP. Table 1 shows the average number of tweets in each profile. Followers count profile has the lowest tweets whereas BLinf has the highest tweets. In contrast, enriching the profile with good chosen tweets can improve its quality such as STBLCinf.

To validate how strong our proposed influence score that the profile STBLCinf built based on, we included the followers count profile and we built it in the same way of the former profile. Results shows that our influence score achieved better performance than the classic method of considering users as influential based on their followers.

5 Conclusion

To conclude, in this work we presented a new way to build users' profiles by exploiting a Twitter-explicit network structure to improve the performance of short-text-based recommender systems. This new user profile exploits the following links between users and their friends to collect a set of short-term tweets and then uses these tweets to build profiles. Furthermore, the influence rule in Twitter is redefined, which helped us to cluster the following list into three groups: influential, less influential, and non-influential. As a result, influential users' tweets are stored directly to the user profile whereas non-influential users' tweets are excluded. Classifiers classify less influential users' tweets to store the representative tweets into the users' profiles.

The advantage of this method has been validated by an offline evaluation experiment over a prototype tweets-recommender system. The discriminative power of our method has been presented by testing and comparing our method against baseline profiles and followers count profiles. Therefore, the proposed method outperformed other profiles. In the future, regarding explicit relationships, we may explore additional types of relationships between a user and friends of his/her friends. Also, we may consider the similarity between the user and his followers to expand the set of tweets that represent the user's interests.

References

1. Abel, F., Gao, Q., Houben, G. J., & Tao, K.: Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In: Proceedings of the 3rd International Web Science Conference. p. 2 ACM (2011).
2. Abel, F., Gao, Q., Houben, G. J., & Tao, K.: Twitter-based user modeling for news recommendations. IJCAI. 13, 2962-2966 (2013).
3. Alonso, O., Carson, C., Gerster, D., Ji, X., & Nabar, S. U.: Detecting uninteresting content in text streams. In SIGIR Crowdsourcing for Search Evaluation Workshop (2010).
4. Anger, I., & Kittl, C.: Measuring influence on twitter. In Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies. p. 3 ACM (2011).
5. Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J.: Everyone's an influencer: quantifying influence on twitter. In Proceedings of the fourth ACM international conference on Web search and data mining. pp. 65-74 ACM (2011).

6. Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K.: Measuring user influence in twitter: the million follower fallacy. Icwsm 10(10-17), 30 (2010).
7. Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., & Yu, Y.: Collaborative personalized tweet recommendation. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 661-670 ACM (2012).
8. Chen, C., Gao, D., Li, W., & Hou, Y.: Inferring topic-dependent influence roles of twitter users. In Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1203-1206 ACM (2014).
9. Elmongui, H. G., Mansour, R., Morsy, H., Khater, S., El-Sharkasy, A., & Ibrahim, R.: TRUPI: twitter recommendation based on users' personal interests. In International Conference on Intelligent Text Processing and Computational Linguistics. pp. 272-284 Springer (2015).
10. Garcia Esparza, S., O'Mahony, M. P., & Smyth, B.: Catstream: categorising tweets for user profiling and stream filtering. In Proceedings of the 2013 International Conference on Intelligent User Interfaces. pp. 25-36 ACM (2013).
11. Karidi, D. P., Stavrakas, Y., & Vassiliou, Y.: A personalized tweet recommendation approach based on concept graphs. In Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences. pp. 253-260) IEEE (2016).
12. Lee, W. J., Oh, K. J., Lim, C. G., & Choi, H. J.: User profile extraction from twitter for personalized news recommendation. Proc. 16th Advanced Communication Technology. 779-783 (2014).
13. Micarelli, A., & Sciarrone, F.: Anatomy and empirical evaluation of an adaptive web-based information filtering system. User Modeling and User-Adapted Interaction. 14(2-3), 159-200 (2004).
14. Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C.: Bad news travel fast: a content-based analysis of interestingness on twitter. In Proceedings of the 3rd International Web Science Conference. p. 8 ACM (2011).
15. Piao, G., & Breslin, J. G. (2016, September). Exploring dynamics and semantics of user interests for user modeling on twitter for link recommendations. In Proceedings of the 12th International Conference on Semantic Systems. pp. 81-88 ACM (2016).
16. Razis, G., & Anagnostopoulos, I.: InfluenceTracker: rating the impact of a twitter account. In IFIP International Conference on Artificial Intelligence Applications and Innovations. pp. 184-195 Springer (2014).
17. Ricci, F., Rokach, L., & Shapira, B.: Recommender systems: introduction and challenges. In recommender systems handbook. Springer (2015).
18. Riquelme, F., & González-Cantergiani, P.: Measuring user influence on twitter: a survey. Information Processing & Management. 52(5), 949-975 (2016).
19. Uysal, I., & Croft, W. B.: User oriented tweet ranking: a filtering approach to microblogs. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management. pp. 2261-2264 ACM (2011).
20. Vosoughi, S.: Automatic detection and verification of rumors on twitter (Doctoral dissertation). Massachusetts Institute of Technology (2015).
21. Weng, J., Lim, E. P., Jiang, J., & He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In Proceedings of the Third ACM International Conference on Web Search and Data Mining. pp. 261-270 ACM (2010).