International Journal of Innovative Research in Computerand Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.194 |

||Volume 8, Issue 4, April 2020||

The Potential of using Twitter as a Data Source for User Profiling in News Recommendation

Fathima Shanaz

Lecturer, Department of Computer Science and Engineering, South Eastern University of Sri Lanka, Sri Lanka

ABSTRACT: Online news reading has become widely popular and the news recommender systems (NRS) are built to provide most relevant news stories readers based on their interest and needs without any manual search effort by the users. It is a non-trivial task that requires access to information about the user, the news item and the general context. One of the main challenge in news recommendation is User profiling, i.e. understanding user's reading preference and behave in accordance of them. News reading preferences of a user may also change over time. Hence, NRS should be able to modify itself as per the needs of the user by employing different recommendation methods. Our work suggests the potential of utilizing the Twitter data for user profiling in news recommendation based on previous studies. Twitter is one of several social media platform contains a large amount of publically accessible user generated content. People use Twitter as a medium for real-time information propagation. Besides, it also acts as an entry point to news for many readers. Hence, observing the user on Twitter tells a lot about person's domain of interest. However, only a little research has been done to profile users' news interest from Twitter data. This paper will survey the potential of using Twitter as a datasource for user profiling in news recommendation and we also aim to investigate different methods to build user profiles using information obtained from Twitter to provide personalized news recommendation.

KEYWORDS: User profiling; News recommendation; Twitter

I. INTRODUCTION

With the development of the Internet, people spend more time online and choose to read news articles on the internet. Due to the presence of a massive amount of news articles, online news readers find it challenging to quickly select relevant news stories based on their interest and preferences. The news recommender systems (NRS) are built to provide users with the right news at the right time based on their interest and needs. Times, Google News, Daily learner, News360 and NTNU SmartMedia are examples of both commercial and research-oriented news recommender systems [1]. For example, Google News is a news aggregator that aims to help users find personalized news stories from thousands of publishers and magazines all over the world. High-quality recommendations make customers satisfied and increase news organization revenue. This requires access to information about the user, the news item and the general context [2]. The more the NRS understands about users and items, the better results can be expected.

NRS are becoming widely popular in recent days, and more weight is given to the research in this area. Similar to NRS, there are also different types of other recommender systems which are used in various domains: such as in movie domain, Netflix recommends movies to users on the basis of what he/she has watched and what he /she has rated. Another example is Amazon product recommendation, where different items are proposed to the user based on their shopping history. However, NRSs are different in several ways from other recommender systems such as for music, movies, products and books. It has to deal with some unique challenges and need a more in-depth analysis of the user, content and their relationships [3]. Researchers have identified the challenges in news recommendation. Some of the most important challenges are listed below.

- Dynamic environment: Every hour hundreds of new articles are published by different sources [3] and older articles quickly become redundant [4]. This makes news domain different from other web objects, requiring much more computation for the recommendation [5].
- Unstructured format: the news domain is rich in text and unstructured in nature. This textual nature of them still make it difficult to analyze the content in the news domain and turn recommendations unreliable [6].
- Recency: people tend to read recent news, instead of old ones [3]. So the importance of news articles decreases in time [2].
- Faster changing user interests compared to other domains: In news domain, it is really hard to predict the user's interest. Some people may read the news not because of their interest in the topic but because of its importance [2].
- User profiling (Knowledge of user preferences): User profiling is an important component of any recommender system. A NRS needs to deal with a large number of users, and each user has their own reading

preferences [7]. In order to make more individual specific recommendations, it is needed to construct a user profile [2]. News reading preferences of a user may change over time [2], [8], [9]. Hence, NRS should be able to modify itself as per the needs of the user by employing different recommendation methods. Therefore, user profiling or user modelling is an important step for solving problems of NRS [10], which helps to recognize user requirements and behave in accordance of them [2], [7], [11].

In the following section, we describe the user profiling for news recommendation in detail.

A. *User profiling for News Recommendation*

People find it difficult to make a choice when they come across a massive amount of information. Therefore, not only news media but also other many organizations use recommender systems to suggest specific products/items to users they might be interested in. These recommendations are usually based on users' past purchase/browsing history and preferences. Therefore, user profiling is normally either knowledge-based (already known/factual) or behavior-based [12], [13]. Information about an individual user is essential because users differ in their preferences, interests, background and goals when using software applications [7]. A typical user profile can include information about user's interests and preferences and it can also contain various user characteristics, such as age, gender, ethnicity, location, etc. for each user of a website [14]. Plumbaum [15] states modeling users is usually an application-dependent approach. In an online newspaper domain, the user profile would contain information such as the types of news (topics) the user likes to read, the types of news (topics) the user does not like to read, the newspapers he usually reads, and the user's reading habits and patterns [7].

There are many approaches to build user profiles for news recommendation. However, accurate user profiling is important for an online recommender system to provide proper personalized recommendations to its users [8]. Cufoglu and Gauch et al. [16], [17] describe two fundamental ways to generate user profiles for personalized recommendations: explicit user profiling and implicit user profiling.

B. *Explicit user profiling*

A user can explicitly give ratings to news articles (example, numbers ranging from 1 to 10) to express how much the user like the article. The most common method to estimate the user's interest by estimating the rating that the user would give to an item. Billsus and Pazzani [18] present a newsagent that learns about the user's interests from user rated news stories (interesting vs. not interesting). Explicit user profiling also makes use of user-generated contents such as user's comments to news articles. Q. Li et al. [19] utilize reader's comments on the news stories to improve the recommendation of related news stories. Shmueli et al. [20] also propose a method to personalized news recommendation based on readers' comments. However, quality comments are needed to make good recommendation. Generally, user's does not provide explicit ratings or comments for the news stories they read. Because it brings an extra burden to users. Hence, explicit user information are generally very sparse and lead to inaccurate user profiles.

C. *Implicit user profiling*

It is also possible to obtain information about users by observing their activities with the underlying application [7]. User activities are recorded in weblog files while they browse a web site and they are maintained by the web servers. It includes web usage information such as a user's click streams and user's browsing history. Web usage mining techniques are used to analyze and discover interesting usage patterns from these weblog files [21]. Thus, most systems depend on web log file as the main source for implicit user information [15].

Here we brief some of previous works which adopt implicit user profiling approach for news recommendation include, Carreira et al. [22] implemented a news browser for PDAs which monitors the users' reading behaviors for each article (such as total reading time, total number of lines, number of lines read by the user) to infer the user's interest on the particular article and update the profile accordingly. J.Liu et al. [23] calculated user's news interest from the user activities and the news trends. The user's genuine news interest was predicted based on readers' past click behaviors and the click distribution of the general public (in a short current time period) was used to measure the current news trend. Further, [24] also proposed a personalized news recommendation algorithm based on consumers' click behavior and it was represented using consumer id, news id and browsing time. Nevertheless, there are a few drawbacks to this implicit user profiling approach,

- Web log files are generally huge in size and contain various kinds of noise.
- Users can access the news sites from different devices or many users can use the same shared device. This further complicates the ability to track the user's browsing history [25], [26].

- Plumbaum [15] explains that in times of highly dynamic websites, most of the user actions are not necessarily send back to the server but handled by the client. Hence, log files only capture a small amount of user interactions. Therefore, it becomes difficult to profile users accurately only with web log data from news web sites.
- This type of user information collection can also rise privacy concerns. Because most of the time users are not aware of the usage of their personal information.
- Usage data can also be collected from user's computers by using adapted browsers or by using techniques such as cookies [27]. This kind of user information collection also cause a lot of privacy issues.

Therefore, additional user information sources need to be explored to construct satisfactory user profiles. It is well recognized that a huge amount of user-generated content in web 2.0 helps to obtain deeper knowledge about users such as users' opinions, perspectives, or tastes towards items or other users. This provides a new way to construct user profiles accurately compared to traditional approach and to mitigate the cold start (of new users) and malicious user rating problems (of news article) considerably [28].

D. *User profiling in Web 2.0*

Web 2.0 tools allows users to interact and collaborate with each other as content creators. Which includes wikis, blogs, folksonomies, social networks such as Twitter or video sharing sites such as YouTube, and the like. This increasing amount of user generated contents in web 2.0 have been used by researchers to extract useful information about users. [29]–[31] use folksonomies to generate user profiles for personalized recommendation. Researchers have discovered that a user's social behaviors can also be used to build user profiles. Farseev et al. [32] use multiple data sources such as Foursquare, Twitter, Instagram and Facebook to learn user profiles. [33], [34] make use of information obtained from Twitter to build user profiles for news recommendation.

Social media data offer huge benefits in terms of its quick response, the scale at which results are delivered and its cost-effectiveness [35]. Twitter is one of several social media platform contains a large amount of publically accessible user generated content. Ahmed [36] reports that Twitter's unique infrastructure and near-total availability of its data have made its popularity among researchers. Previous studies on Twitter suggest the advantages of using Twitter as a datasource.

- Twitter is the most popular microblogging service with a large number of monthly active users. Although Facebook and WhatsApp have more monthly active users than Twitter, They do not make their data available to academics and researchers on a similar scale to Twitter [36]–[38].
- Twitter is public to anyone by default, unlike on most online social networking sites, such as Facebook or MySpace, a Twitter user can follow any other user. However, users can adjust privacy settings to make their accounts protected.
- It is easy to find and follow conversations on Twitter through its search feature and Google search results.
- Twitter has hashtag norms which make it easier gathering, sorting, and expanding searches when collecting data [36].
- Also, it is easy to retrieve major incidents, news stories and event from Twitter as they are tend to be centered around hashtag [36].
- Unlike other Social networking services, Twitter provides open API for free of cost. Which facilitates researchers to acquire and analyze data for various purposes [39].

II. USER INTEREST MODELLING FROM READERS TWEETS FOR NEWS ARTICLE RECOMMENDATION

Various reports suggest the potential of utilizing the Twitter data for news recommendation. According to Wikipedia definition[1], Twitter is an online news and social networking service where users post and interact with messages, called "tweets." These messages were originally restricted to 140 characters, but on November 7, 2017, the limit was doubled to 280 characters for non-Asian languages. Tweets are created at a dramatic rate, as of the first quarter of 2019, the Twitter has 330 million monthly active users[2]. Averagely 500 million tweets are sent every day and this is around 200 billion tweets per year[3]. People use Twitter as a medium for real-time information propagation [40], [41]. This has transformed the way people convey information especially in the areas of news [42]. Unlike on most online social networking sites, such as Facebook or MySpace, a Twitter user can follow any other user. Being a follower on Twitter

[1]https://en.wikipedia.org/wiki/Twitter
[2]https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[3]https://www.internetlivestats.com/twitter-statistics/

means that the user receives all the messages (called tweets) from those the user follows [43]. These tweets cover various topics, including comments on events and trending topics, personal activities, politics and many others [44].

Twitter also acts as an entry point to news for many readers [45]. It has been used by news media for real-time information dissemination [46]. Kwak et al. [43] discovered that over 85% of topics mentioned in tweets are related to news. Researchers have investigated the extensive use of Twitter by journalist as a news reporting tool and as a medium for rapid news propagation [47]–[50]. Lerman and Ghosh [47] express that Twitter users having many followers influence significantly on the dissemination of news, Therefore, news stories through Twitter network spread faster than traditional news sources. Broersma and Graham [48] mention the growing popularity of Twitter among journalists and their extensive use of Twitter as a source for reporting. Because it helps them to find new stories in real time and facilitates a rapid dissemination of information. Vis [49] describe that how Twitter can be used as a reporting tool during a breaking news. Armstrong and Fangfang Gao [50] indicate that news organizations employ Twitter as a tool of information dissemination. Their findings reveal that Crime and public affairs were the most frequently tweeted topics and they suggest that 67% of tweets from news organizations were the same as headlines in the linked news stories.

III. RELATED WORK

The short length and the frequent updating nature of Twitter has attracted increasing interests among researchers and started to discover various insights from large volume of Twitter data. Many studies suggest the potential of utilizing the Twitter data for extracting characteristics that influence user's news interest such as, trending news topics [40], [51], breaking news [40], [42], [49], social media popularity of news articles [45], temporal behaviour of trending topics and user participation [43], knowing topical expertise of Twitter users [52], detection of crime and disaster related events [53], individual users' mobility behaviours [54], identification of influential users of Twitter [55], investigation on the cold-start news popularity [56], inferring user demographic information [57]–[59] earthquake detection system [60], Tweet location detection [61], user interest identification [62], User's personality prediction [63], Clustering users based on Interests [35] and so on.

While related works expose the potential of Twitter in the area of news recommendation and user modelling, only a little research has been done to profile users from Twitter data for news recommendation. Such as [40] describe an approach to news recommendation using real time Twitter data for ranking and recommending articles from a collection of RSS (Really Simple Syndication) feeds. Users provide their Twitter account information and a list of RSS feeds they wish to follow during user registration process. Then the RSS stories are ranked based on the co-occurrence of popular terms within the user's RSS and tweets. Furthermore, [34] utilize tweets, re-tweets, and hashtags, from which important keywords (noun phrases and hashtags) are extracted to build the personal profile to provide personalized news recommendation. They adopted the bag-of-words approach for user profiling in which the bag of words are extracted from the tweets (or re-tweets) written by the user. To rank the upcoming news articles they use cosine similarity score between this news article profile and the user profile. The effectiveness of their approach was validated by implementing a prototype news recommendation service and by performing a user study.

Abel et al. [64] study semantic user profiling based on Twitter posts. They validate that enriching the semantics of tweets have strong impact on the construction of user profile. As automatically inferring the semantic meaning of Twitter posts is a non-trivial task, they link Tweets and related news articles in order to contextualize Twitter activities. This way they enrich the semantics of Twitter messages based on the relationship between a Tweet and the news articles. Two strategies have been proposed to find the correlation between the Tweet and the news article such as URL based strategies and content based strategies. Besides, [65] measured and compared the performance of three different types of user modelling strategies in the context of a personalized news recommendation, namely hashtag-based, entity-based and topic-based. They exploited both user tweets and linked news articles for creating user profiles. Their results reveal that entity-based user profiling enables better recommendation quality due to the number of distinct entities that occur in entity-based profiles. Also analyzed the temporal dynamics of the different types of profiles and observed the change of profiles over time. They show that the consideration of such temporal characteristics is beneficial to news recommendation. Moreover, [51] recommends interesting news stories to the user based on their popularity as well as their relevance to the user favourite news topics. Where hot stories for a day is obtained by analysing Twitter public timeline. The user's favourite news topics (among 7 news categories from news site) is extracted from web forms.

IV. DISCUSSION

We understand that these tweets have become a valuable source of information for user profiling in news recommendation application. However, mining tweets poses unique challenges due to their short, noisy, context-dependent, and dynamic nature [66]. Nevertheless, previous studies suggest that the semantic enrichment of twitter messages have a strong impact on the construction of user profiles. Gao et al. [67] study user modeling on Twitter and investigate the relationship between personal interests and public trends. They show that for news recommendation personal interests are more important than public trends and by combining trend and personal interest recommendation quality can be improved. An important work by Abel et al. [33] compare entity-based and category-based user profiles with the hashtag-based profiles and reveal that an understanding of the semantic meaning of Twitter messages is key for generating high-quality user profiles. Similarly, Ran et al. [68] also state that a critical step in understanding and mining information from Twitter is to disambiguate entities in tweets, i.e., tweet entity linking. Therefore, tweet entity linking is significantly beneficial when using Twitter data for user profiling in news recommendation.

The task to link the named entity mentions detected from tweets with the corresponding real world entities in the knowledge base is called tweet entity linking [69]. However, Tweets pose special challenges to entity linking. Because, when extracting entity mentions from tweets the same entity mention can refer to multiple entities. The term 'entity' in this paper refers to real-world objects; it can be a personal name, a product, a book title, an organization name, a location and more. Semantic meaning of Twitter messages can be inferred by mapping entity mentions in tweets to the corresponding entities in a given knowledge base, e.g., Wikipedia. This task is challenging due to name variations, entity ambiguity and incompleteness of knowledge bases. Several approaches have been proposed to tackle these challenges. Knowledge bases are a prerequisite for any entity linking task. Knowledge bases provide information about the world's entities, their semantic categories and the mutual relationships between entities. Four of such knowledge bases which have been widely exploited in the field of entity linking include Wikipedia, DBpedia [70], YAGO [71] and Freebase [72].

V. CONCLUSION

To review, people use Twitter as a medium for real-time information propagation; accordingly, it acts as an entry point to news for many online news readers. Therefore, users' Twitter activities can be exploited to collect variety of interested entities of targeted users as well as the period of validity (as news reading preferences of a user change over time). Further, trending news stories also influence to user's news reading preference. Because, readers can choose news articles to read not only for their personal interests but also based on their profession, community the person belongs to, etc. hence incorporating these factors into consideration which influence users' news reading preferences we can improve the timeliness of news recommendation. However, only a very little work has been done on entity-based user profiling for news recommendation using Twitter. But an accurate profile of users' interests is critically important for the success of content-based news recommendation. Therefore, more research is needed to address the problems related to entity-based user profiling for news recommendation and using Twitter.

REFERENCES

1. M. Harandi and J. A. Gulla, "Survey of User Profiling in News Recommender Systems", 3rd International Workshop on News Recommendation and Analytics (INRA 2015), 2015.
2. Ö. Özgöbek, J. A. Gulla, and R. C. Erdur, "A Survey on Challenges and Methods in News Recommendation", Proceedings of the 10th International Conference on Web Information Systems and Technologies, 2014, pp. 278–285.
3. J. A. Gulla, L. Martens, Ö. Özgöbek, N. Shabib, and T. De Pessemier, "4th International Workshop on News Recommendation and Analytics (INRA 2016)", 2016.
4. D. Doychev, A. Lawlor, R. Rafter, and B. Smyth, "An Analysis of Recommender Algorithms for Online News", CLEF 2014 Conference, 2014.
5. K. G. Saranya and G. S. Sadhasivam, "A Personalized Online News Recommendation System", International Journal of Computer Applications, vol. 57, no. 18, pp. 6–14, 2012.
6. J. A. Gulla, B. Yu, Ö. Özgöbek, and N. Shabib, "3rd International Workshop on News Recommendation and Analytics (INRA 2015)", Proceedings of the 9th ACM Conference on Recommender Systems, 2015, pp. 345–346.
7. S. Schiaffino and A. Amandi, "Intelligent User Profiling," in Artificial Intelligence An International Perspective, vol. 5640, Springer, Berlin, Heidelberg, 2009, pp. 193–216.
8. Z. Lu, S. J. Pan, Y. Li, J. Jiang, and Q. Yang, "Collaborative Evolution for User Profiling in Recommender Systems", Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016, pp. 3804–3810.
9. F. Abel, Q. Gao, G. Houben, and K. Tao, "Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web", Proceedings of the 3rd International Web Science Conference on - WebSci '11, 2011, pp. 1–8.
10. L. Li, L. Zheng, F. Yang, and T. Li, "Modeling and broadening temporal user interest in personalized news recommendation", Expert Syst. Appl., vol. 41, no. 7, pp. 3168–3177, 2014.
11. S. Kanoje, S. Girase, and D. Mukhopadhyay, "User Profiling Trends, Techniques and Applications", International Journal of Advance Foundation and Research in Computer, vol. 1, no. 1, 2014.

12. S. Kanoje, D. Mukhopadhyay, and S. Girase, “User Profiling for University Recommender System Using Automatic Information Retrieval”,International Conference on Information Security and Privacy, vol. 78, pp. 5–12, 2016.
13. S. Natarajan and M. Moh, “Recommending News Based on Hybrid User Profile, Popularity, Trends, and Location”, 2016 International Conference on Collaboration Technologies and Systems (CTS), 2016, pp. 204–211.
14. G. Adomavicius and A. Tuzhilin, “Towards the Next Generation of Recommender Systems : A Survey of the State-of-the-Art and Possible Extensions”,IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.
15. T. Plumbaum, “User Modeling in the Social Semantic Web”, Technical University Berlin, 2015.
16. S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli, “User Profiles for Personalized Information Access”, The Adaptive Web,vol. 4321. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
17. A. Cufoglu, “User Profiling-A Short Review”,International Journal ofComputer Applications, vol. 108, no. 3, pp. 1–9, 2014.
18. D. Billsus and M. J. Pazzani, “A Hybrid User Model for News Story Classification”, Proceedings of the Seventh International Conference on User Modeling, 1999, pp. 99–108.
19. Q. Li, J. Wang, Y. P. Chen, and Z. Lin, “User comments for news recommendation in forum-based social media”, Information Sciences, vol. 180, no. 24, pp. 4929–4939, 2010.
20. E. Shmueli, A. Kagian, Y. Koren, and R. Lempel, “Care to Comment? Recommendations for Commenting on News Stories”, Proceedings of the 21st international conference on World Wide Web - WWW ’12, 2012, pp. 429–438.
21. R. K. Shukla, S. Silakari, and P. K. Chande, “Existing Trends and Techniques for Web Personalization,” Int. J. Comput. Sci., vol. 9, no. 4, pp. 430–439, 2012.
22. R. Carreira, J. M. Crato, D. Gonçalves, and J. A. Jorge, “Evaluating adaptive user profiles for news classification”, Proceedings of the 9th international conference on Intelligent user interface - IUI '04, p. 206, 2004.
23. J. Liu, P. Dolan, and E. R. Pedersen, “Personalized News Recommendation Based on Click Behavior”,Proceedings of the 15th international conference on Intelligent user interfaces, 2010, pp. 31–40.
24. Y. Wang and W. Shang, “Personalized News Recommendation Based on Consumers’ Click Behavior”,12th International Conference on Fuzzy Systems and Knowledge Discovery, pp. 634–638, 2015.
25. I. Mohallick and Ö. Özgöbek, “Exploring privacy concerns in news recommender systems”, Proceedings of the International Conference on Web Intelligence - WI ’17, 2017, pp. 1054–1061.
26. B. Kille, F. Hopfgartner, T. Brodt, and T. Heintz, “The plista dataset”, Proceedings of the 2013 International News Recommender Systems Workshop and Challenge on - NRS ’13, 2013, pp. 16–23.
27. D. Anandhi and M. S. Irfan Ahmed, “An improved web log mining and online navigational pattern prediction”,Research Journal of Applied Sciences, Engineering and Technology, vol. 8, no. 12, pp. 1472–1479, 2014.
28. X. Zhou, Y. Xu, Y. Li, A. Josang, and C. Cox, “The state-of-the-art in personalized recommender systems for social networking”,Artificial Intelligence Review, vol. 37, no. 2, pp. 119–132, 2012.
29. J. Stoyanovich, S. Amer-Yahia, C. Marlow, and C. Yu, “Leveraging Tagging to Model User Interests in del. icio. us”,AAAI Spring Symposium: Social Information Processing, pp. 104–109, 2008.
30. B. Tahar-Rafik and A.-O. Rachid, “Towards a new approach for generating user profile from folksonomies”, 2014 4th International Symposium ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb), 2014, pp. 1–6.
31. H. Liang, “USER PROFILING BASED ON FOLKSONOMY INFORMATION IN WEB 2.0 FOR PERSONALISED RECOMMENDER SYSTEMS”, Queensland University of Technology, 2010.
32. A. Farseev, L. Nie, M. Akbari, and T. Chua, “Harvesting Multiple Sources for User Profile Learning : a Big Data Study”,Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp. 235–242, 2015.
33. F. Abel, Q. Gao, G. J. Houben, and K. Tao, “Twitter-based user modeling for news recommendations”,IJCAI International Joint Conference on Artificial Intelligence, pp. 2962–2966, 2013.
34. W.-J. Lee, K.-J. Oh, C.-G. Lim, and H.-J. Choi, “User profile extraction from Twitter for personalized news recommendation”, 16th International Conference on Advanced Communication Technology, 2014, pp. 779–783.
35. M. Brambilla and B. Rahdari, “Event-based User Profiling in Social Media Using Data Mining Approaches”, School of Industrial and Information Engineering, 2017.
36. W. Ahmed, “Using Twitter as a data source: An overview of current social media research tools”, Impact of Social Sciences Blog, pp. 1–6, 2015.
37. B. Batrinca and P. C. Treleaven, “Social media analytics: a survey of techniques, tools and platforms”,AI and Society, vol. 30, no. 1, pp. 89–116, 2014.
38. H. Liang, F. Shen, and K. Fu, “Privacy protection and self- disclosure across societies : A study of global Twitter users”,new media & society, vol. 19, no. 9, 2017 2016.
39. M. G. Kim, Y. O. Kang, and J. H. Koh, “Evaluating residential location inference of twitter users at district level: focused on Seoul city”,Spatial Information Research, vol. 24, no. 4, pp. 493–502, 2016.
40. O. Phelan, K. McCarthy, and B. Smyth, “Using twitter to recommend real-time topical news”,Proceedings of the third ACM conference on Recommender systems - RecSys '09, p. 385, 2009.
41. Z. Xu, Y. Zhang, Y. Wu, and Q. Yang, “Modeling user posting behavior on social media”,Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, pp. 545–554, 2012.
42. S. Phuvipadawat and T. Murata, “Breaking news detection and tracking in Twitter”, Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp. 120–123.
43. H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a Social Network or a News Media ?”, Proceedings of the 19th international conference on World wide web, 2010, pp. 591–600.
44. Z. Ma, A. Sun, Q. Yuan, and G. Cong, “Tagging Your Tweets : A Probabilistic Modeling of Hashtag Annotation in Twitter”, Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, pp. 999-1008, 2014.
45. V. Dimitrova, A. Piotrkowicz, J. Otterbacher, and K. Markert, “Headlines Matter: Using Headlines to Predict the Popularity of News Articles on Twitter and Facebook”, Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), 2017, pp. 656–659.
46. K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal, and A. Choudhary, “Twitter trending topic classification”, 11th IEEE International Conference on Data Mining Workshops Twitter, 2011, pp. 251–258.
47. K. Lerman and R. Ghosh, “Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks”,Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010, pp. 90–97.
48. M. Broersma and T. Graham, “Twitter as a news source how dutch and british newspapers used tweets in their news coverage, 2007-2011,” Journalism Practice, vol. 7, no. 4, pp. 446–464, 2013.

49. F. Vis, "Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots",Digital Journalism, vol. 1, no. 1, pp. 27–47, 2013.
50. C. L. Armstrong and Fangfang Gao, "Now Tweet This: How News Organizations Use Twitter",Electronic News, vol. 4, no. 4, pp. 218–235, 2010.
51. N. Jonnalagedda and S. Gauch, "Personalized News Recommendation Using Twitter," 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), vol. 03, pp. 21–25, 2013.
52. C. Wagner, V. Liao, P. Pirolli, L. Nelson, and M. Strohmaier, "It's not in their tweets: Modeling topical expertise of twitter users", 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, 2012, pp. 91–100.
53. R. Li, K. H. Lei, R. Khadiwala, and K. C. C. Chang, "TEDAS: A twitter-based event detection and analysis system", 2012 IEEE 28th International Conference on Data Engineering, 2012, pp. 1273–1276.
54. Q. Yuan, G. Cong, and Z. Ma, "Who, where, when and what: discover spatio-temporal topics for twitter users", 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 2013, pp. 605–613.
55. J. Weng, E. P. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers", 3rd ACM International Conference on Web Search and Data Mining (WSDM 2010), 2010, pp. 261–270.
56. I. Arapakis, B. Barla Cambazoglu, and M. Lalmas, "On the feasibility of predicting news popularity at cold start", International Conference on Social Informatics, 2014, vol. 8851, pp. 290–299.
57. W. Liu and D. Ruths, "What's in a Name? Using First Names as Features for Gender Inference in Twitter", 2013 AAAI Spring Symposium, 2013, pp. 10–16.
58. D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, "Classifying Latent User Attributes in Twitter", 2nd international workshop on Search and mining user-generated contents, 2010, pp. 37–44.
59. X. W. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu, and X. Li, "We Know What You Want to Buy: A Demographic-based System for Product Recommendation on Microblogs", 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 1935–1944.
60. T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: Real-time Event Detection by Social Sensors", Proceedings of the 19th international conference on World wide web - WWW '10, 2010, pp. 851–860.
61. B. R. Heravi and I. Salawdeh, "Tweet Location Detection", Computation + Journalism, 2015.
62. P. Kapanipathi, P. Jain, C. Venkataramani, and A. Sheth, "User Interests Identification on Twitter Using a Hierarchical Knowledge Base", European Semantic Web Conference, 2014, no. 1, pp. 99–113.
63. J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter", 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, 2011, pp. 149–156.
64. F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web", Extended Semantic Web Conference, 2011, pp. 375–389.
65. F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Analyzing User Modeling on Twitter for Personalized News Recommendations", International Conference on User Modeling, Adaptation, and Personalization, 2011, pp. 1–12.
66. S. Perera, P. Mendes, A. Alex, and A. Sheth, "Implicit Entity Linking in Tweets", European Semantic Web Conference, 2016, pp. 118–132.
67. Q. Gao, F. Abel, G. J. Houben, and K. Tao, "Interweaving trend and user modeling for personalized news recommendation", 2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2011, vol. 1, pp. 100–103.
68. C. Ran, W. Shen, and J. Wang, "An Attention Factor Graph Model for Tweet Entity Linking", 2018 World Wide Web Conference on World Wide Web - WWW '18, 2018, vol. 2, pp. 1135–1144.
69. W. Shen, J. Wang, P. Luo, and M. Wang, "Linking Named Entities in Tweets with Knowledge Base via User Interest Modeling", 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 2013, pp. 68–76.
70. S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A Nucleus for a Web of Open Data", International Semantic Web Conference, 2007, vol. 4825, pp. 722–735.
71. F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge Unifying WordNet and Wikipedia", 16th international conference on World Wide Web - WWW '07, 2007, pp. 697–706.
72. K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A Collaboratively Created Graph Database For Structuring Human Knowledge", 2008 ACM SIGMOD international conference on Management of data - SIGMOD '08, 2008, pp. 1247–1250.

BIOGRAPHY

Abdul Lathif Fathima Shanaz is a Lecturerat Department of Computer Science and Engineering, Faculty of Engineering, South Eastern University of Sri Lanka. She is currently working toward the M.Phil degree in Postgraduate Institute of Science, University of Peradeniya, Sri Lanka. She received bachelor's degree in Computer Engineering in 2007 from University of Peradeniya, Sri Lanka. Her research interests include Text Mining, Natural Language Processing, Machine Learning, Recommender Systems etc.