ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Friend Recommendation using FLDA Topic Assignment Model in Microblogging System

Prachi Jain[1], Shanthi K. Guru[2]

Master of Engineering, Dept. of C.E., DYPCOE, Savitribai Phule Pune University, Pune, India

Assistant Professor, Dept. of C.E., DYPCOE, Savitribai Phule Pune University, Pune, India

ABSTRACT: As the requirement of emerging real time information, microblogging a form of social media is becoming people's favorite choice for searching information or sharing opinion. In microblogging, user is allowed to publish their post or share opinion but microblogging post are limited to 140 characters. These post depicts the behavior or pattern of user, which can help to find user's area of interest. By finding users social behaviors and dynamics, it is easy for user to find friends with similar interests, which may improve the user's experience, social interactions, and can achieve more business value for corporations. By identifying the topics of words, it is easy to find such common interests. FLDA is applied on microblogging data for topic formation. These topics are represented using Fenwick tree. The purpose is to discover the interests of each user rather than the topics of single messages.

KEYWORDS: Microblogging; friend recommendation; temporal model; interest drifts; fenwick tree

I. INTRODUCTION

Microblogging has become a suitable way for Internet users to interconnect with their friends and family members, or to express their emotions or feelings. Using a microblog also has gradually become a habit for a huge amount of users, which leads to an exponential explosion of data in the virtual microblog society on the Internet which makes retrieval of related data extremely difficult. Therefore, more and more microblog services are developing novel engines dedicated to recommending user-specific information. In the early research the main focus was on the characteristics of microblogging and analysis of social network. But recently, interest has been increased in the area of information retrieval, such as event detection and tracking, identification of influential people, sentiment analysis, and personalized recommendations [1].

Recommender systems have become extremely common in recent years, and are applied in a variety of applications. The most popular ones are probably movies, music, news, books, research articles, search queries, social tags, and products in general. However, there are also recommender systems for experts, collaborators, jokes, restaurants, financial services, life insurance, persons (online dating), and Twitter followers. These system face the social interaction overload problem, by suggesting users or items that a target user might be interested in.

Benefits

- Find things that are interesting
- Narrow down the set of choices
- Increase sales
- Obtain more knowledge about customers
- Deliver Relevant Content
- Reduce Workload and Overhead

As a requirement of emerging real-time information, microblogging is becoming people's favorite choice for seeking information and expressing opinions [1]. Messages received by a user mainly depend on whom the user follows. Thus, to recommend users with similar interests may improve users experience for information they desire to acquire [1]. By mining user's social behaviors and dynamics, it may help them to find friends with similar interests, which may improve the user's experience, social interconnection, and gain high business value for corporations.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

II. RELATED WORK

The popularity of microblogging systems has attracted many researchers attention. Early research focused on the characteristics of microblogging [2], [7], [27] and analysis of social networks [9], [28]. In the area of information retrieval, there is recently interest is increased include tracking and event detection [11], [14], finding of influential people [6], analysis of sentiment [15]- [18], and personalized recommendations [19], [20]. In this paper we focuses on users potential interests prediction, so introduced the related work on topic finding approaches, time-aware recommendations, user-interest analysis and friend recommendations.

A. Topic Finding

Xu et al. [24] modelled user posting behaviour in Twitter by considering three factors: breaking news, posts from social friends and user's intrinsic interest via an extension of author topic (AT) model, where users intrinsic interest was represented by a distribution over latent topics. Hong and Davison [30] conducted standard LDA and extended AT model to predict popular Twitter messages and classified Twitter users and messages into categories. Zhang et al. proposed an evolutionary topic pattern mining approach to discover changing of topic structures on a community question answering platform. The approach first extracted question topics via LDA in each time window, then discovered topic transitions based on cosine similarity, and finally analysed life cycles of the extracted topics. Twitter-Rank [6], approach use the in-degree method to measure the influence of twitterers or the number of followers and then suggestion of friend is done on the basis of users influence.

B. Time-Aware Recommendations

Rafeh and Bahrehmand [23] proposed an adaptive collaborative filtering algorithm which uses time to reflect fluctuations in user's behavior over time. Liu et al. [24] developed a social temporal collaborative ranking model to recommend movies. Recently, Eirinaki et al. [25] proposed a trust-aware system for user recommendations which analyzed the semantics and dynamics of the implicit and explicit connections between users via a discounting factor. In [26], an adaptive exponential forgetting function is used.To cope with current topic recommendations in microblogging system, in [19], a probability matrix factorization is proposed.

C. Recommendation System

There are mainly three approaches in recommendation systems: content-based filtering, collaborative filtering, and hybrid [10]. In [13] friend recommendation and news feed curation algorithm is designed. Content-based filtering methods match the preferences or profiles derived from attribute information of users, such as informational friend recommendation to predict the rating for each post written by user [21]. Collaborative filtering makes automatic predictions (filtering) about the interests of a user by collecting taste information (ratings of items) from many users. In[4] identify usefulness of suggesting activity partners together with items in recommender system. In[22] friend recommendation on the basis of two dimensions: content interest and context(location, time). Friend of friend and number of common followed users, common followers and common joined group of users used for recommendation of friend[12]. Incremental Katz Approximation algorithm uses topology and content based approach to effectively recommend relevant users [8]. In[5] friend recommendation for social network based on the topology of network graph. Friend recommendation on the basis of two phase private friend recommendation protocol [29].

III. FRIEND RECOMMENDATION IN SOCIAL DOMAIN

Existing social networking services recommend friends to users based on their social graphs or to pick friend candidates like friend of friend i.e. mutual friends which may not be the most appropriate to reflect a user's preferences on friend selection in real life. Users potential interests on others are predicted according to user similarities over different periods of time via temporal functions based on topic model [26]. User recommendation in social domain based on the pair of users that are likely to be interested in each other's content or people to follow. User recommender system that operate on social media can be classified on the source of data used to build the recommendation. 1. System based on the analysis of social connection which explore the set of people connected to the target user in order to produce recommendations. These system recommend either closest user in the connection like friends of friend and followees of followees or recommend the user who have highest

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

probability to be crossed in random walk of social connection. 2. Recommendation are produced by identifying users with similar profiles [26].

People's movements on social network or microblogging imply rich information about their life interest and preferences. For example: if a person usually goes to stadium and gym, it denotes that the person might like sports. So if two people have same movements then they might share similar interest and preferences. The more things they share, the more correlated these two users would be. Therefore, based on topic distribution of posts written by user a friend recommender is conducted which provides an individual with some similar user in terms of topics and interests [26]. User recommendation in social domain based on the pair of users that are likely to be interested in each other's content or people to follow. The friend recommendation in social domain can be based on many factors such as:

A. Mining Friends From School and Work

In this type of recommendation, friend recommendation is done on the basis of user's education and work place. If two users are having same professional or educational background, chances of their becoming friend increased and hence friend is recommended.

B. Stalking Facebook Users

In Facebook, if user1 visits another user's profile, at that time system takes the snapshot and consider that user1 is interested in user2 and suggest friend accordingly.

C. Using Facebook Friend Suggestions in Marketing

In this type of recommendation, friend is recommended if users are liking same type of product. Then it consider that people may have same type of interest and likes.

D. Connecting Friends of Friends

In this type of recommendation, friend suggestion is done on the basis of mutual friend or common friend of two users.

E. Suggestion by user only

This kind of friend suggestion is mostly done in Facebook where user can suggest friend to another user.

F. Users area of interest

User's area of interest can be find out by analyzing user's behavior or pattern of publishing post and friend is recommended on the basis of user's interest.

IV. PROPOSED SYSTEM

A collaboration of Fenwick tree and LDA is done for topic formation and to predict user's potential friends. Existing system does not give the good result of topic formation as compared to FLDA model. FLDA uses fenwick tree for topic sampling and achieves better results for friend recommendation.

A. *Problem Definition:*

Keywords are not sufficient for discovering user's interests. As the existence of synonymy, it needs to find the hidden topics from the user's posts. Since the goal is to find topics that each microblogging user is interested in rather than topics that each microblog is about. Interest received by a user mainly depend on whom the user follows. Thus, recommending users with similar interests may improve the experience quality for information receiving. So user's preferences can be discover by analyzing the user-generated contents.

B. *Enhanced LDA Model (FLDA):*

To learn meaningful topic models with huge document collections which contain millions of documents is challenging because of to deal with a large number of topics. In order to handle huge number of topics we proposed

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

FLDA (Fenwick + LDA) which uses an appropriately modified Fenwick tree[3]. This data structure allows us to sample from a multinomial distribution over T items in O (log T) time. A Fenwick tree is a wonderful data structure that supports two operations: increment a given value by a given amount, and find the sum of a segment of values, both in O(log n) time. Topic models offer a way to combine vocabulary from a document corpus to form latent topics. In particular, Latent Dirichlet Allocation (LDA) is one of the most standard topic modelling approach. In LDA, learning is unsupervised because the input data is partial, the corpus make available only the words within documents whereas in FLDA, learning is supervised where training is provided to data for topic formation. We propose a fast sampling algorithm FLDA, where T is the number of topics. F+LDA only costs O(log T) times by utilizing the F+tree data structure. The proposed F+tree sampling method is very efficient in handling large number of topics compared to the simple LDA.

C. *System Architecture:*

All the intermediate steps of the systems are clearly mentioned in this architecture. Here first input file or string is taken from the user. An enhanced model of LDA is proposed which is a collaboration of Fenwick and LDA to predict user's potential friends as given in figure2. The preprocessing step is done initially where stop words are removed from microblog data and trimming of words are done. Then topic formation is done using LDA approach. The topics are represented in form of tree called fenwick tree and hence called FLDA. Then, topic and user similarity matching is performed. Finally, the interest of user are predicted according to user similarities over different periods of time then friend recommendation is performed.

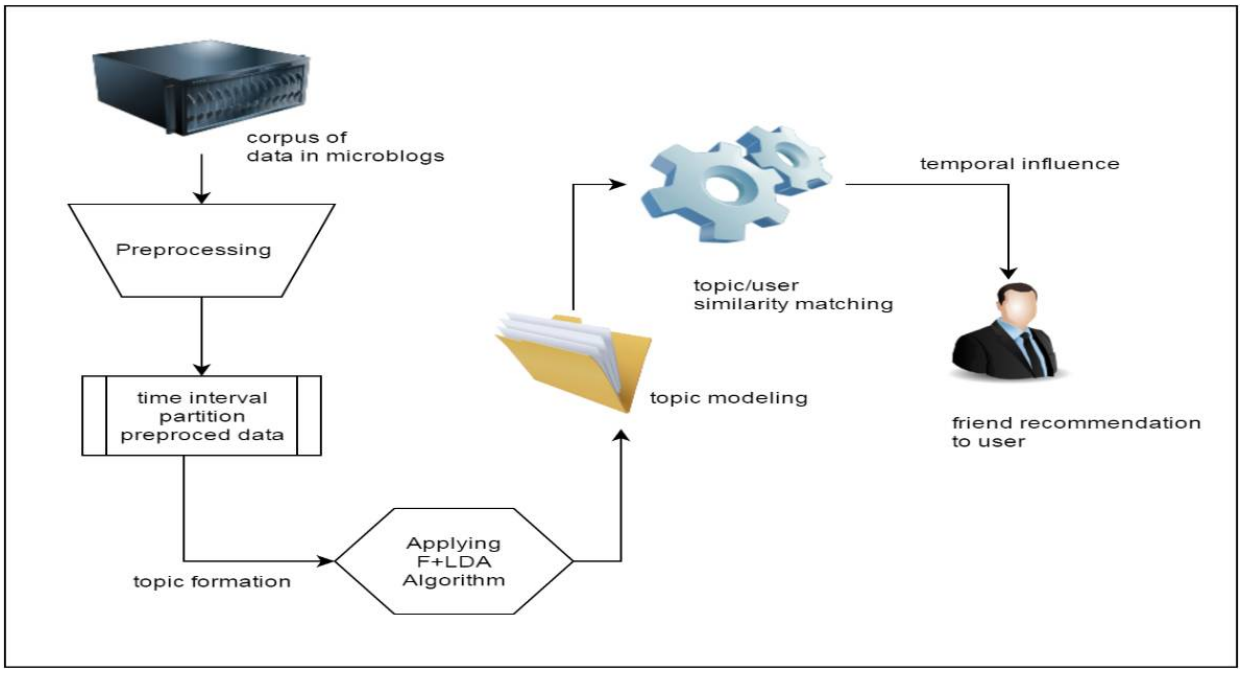

Fig 1. Enhanced LDA Friend Recommendation Framework

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

V. PSEUDO CODE

1. Randomly initialize z and increment counter
2. ***for*** each iteration do
3. ***for*** i = 0 → N-1 do

word← w[i]
topic←z[i]
$n_{d,topic} - =1$; $n_{word,topic} - =1$; $n_{topic} - =1$

for k = 0 →K – 1 do
$p(z = k|.) = (n_{d,k} + \alpha_k) \frac{n_{k,w} + \beta_w}{n_k + \beta \times W}$
end

4. Topic ← sample from p(z|.)
5. z[i] ← topic.
6. $n_{k,topic} + = 1$; $n_{word,topic} + = 1$; $n_{topic} + = 1$
7. ***end***
8. **end**
9. **return** z, $n_{d,k}, n_{k,w}, n_k$
10. **end**

VI. IMPLEMENTATION DETAILS

In this section, the experimental evaluation of the proposed system is initiated. Microblog gradually becomes a habit for a massive amount of users, which leads to an exponential explosion of. Therefore, more and more microblog services are developing novel engines dedicated to recommending user-specific information.

A. Mathematical Model

Let S be the system: S = {I, F, O}
where, S = System, I = Set of Inputs, F = Set of Functions, O = Set of Outputs

1. Input(I) = {M, V, K, α, β, N}
   where,
   - M = Total number of Documents
   - V = Vocabulary Size
   - K = Total number of Topics
   - α = Dirichlet Parameter for users at time t
   - β = Dirichlet Parameter for topic at time t
   - N = Total number of words for f-tree

2. Function(F) = {F1, F2, F3, F4, F5}
   Where,
   - F1= Preprocessing
     Stopword Removal
   - F2= Topic Formation
   - F3= F-tree Generation
     F.Sample (T)
     F.Update (T, α, β)
   - F4= Topic/User Similarity Matching

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

- F5= Friend Recommendation

$$p(z = k|.) = (n_{d,k} + \alpha_k) \frac{n_{k,w} + \beta_w}{n_k + \beta \times W}$$

Where,

- Pn is the probability of potential interests among users
- n_{dksum} = sum for each row in n_{dk}
- n_{kwsum} = sum for each row in n_{kw}
- n = count

3. Output(O) = {O1, O2, O3, O4, O5}

where,

- O1 = Mt be the users keyword matrix at time t.
- O2 = T is the set of number of topics.
- O3 = sim (i, j) is the Topic similarity between user i and user j.
- O4 = P_k is a User's Potential Interest.
- O5 = Friend recommendation.

B. Dataset:

Twitter offers a platform with a number of different ways to interact with it. Web Intents, Tweet Button and Follow Button is the simplest way to bring basic Twitter functionality to the site. It provides features like the ability to tweet, retweet, or follow using basic HTML and JavaScript. While Tweets are still limited to 140 characters, the character counting model has evolved over time. Tweets are UTF-8 string content and any UTF-8 character counts as a single character. Strings that are understood as links, such as http://twitter.com, twitter.com, twitter.com/twitterapi will be automatically converted to t.co.t.co links represent a variant character count currently 22 for HTTP-based links and 23 for HTTPS-based links [31].

The Streaming API allows to stream tweets in real time as they happen. The Search API provides relevant results to ad-hoc user queries from a limited corpus of recent tweets. The REST API allows access to the nouns and verbs of Twitter such as reading timelines, tweeting, and following. To use the REST and Streaming API, one should register an application and get to know the ways of OAuth and explore Twitter Libraries. The acronym API stands for Application Programming Interface. An API is a defined way for a program to accomplish a task, usually by retrieving or modifying data. In Twitters case, API method is about every feature can see on website. Programmers use the Twitter API to make applications, websites, widgets, and other projects that interact with Twitter. Programs talk to the Twitter API over HTTP, the same protocol browser uses to visit and interact with web pages.

VII. RESULTS

Output of LDA system is compared with Enhanced LDA on different parameters such as interest/likes, topic formation and friend recommendation in Table 1. Result of system is calculated in terms of searching result and represented in percentages in given Table 1. Existing approach are less efficient as compare to proposed approach in terms of friend recommendation. Existing approach works average for topic formation and friend recommendation. Topic formation and its sampling achieves good result using proposed system whereas searching results using interest/likes are near about same for both approaches. Enhanced LDA is better than LDA in Friend Recommendation in terms of time and complexity.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Parameters	LDA (Existing) (%)	Enhanced LDA (Proposed) (%)
Interest/Likes	60	70
Topic Formation	60	80
Friend Recommendation	60	85

Table 1 : Efficiency Of Proposed System Against Various Existing Systems On Different Parameters

A Graph is plotted between LDA and Enhanced LDA approach where y-axis shows the searching results and x-axis contains different parameters such as interest/likes, topic formation and friend recommendation in Fig 2.

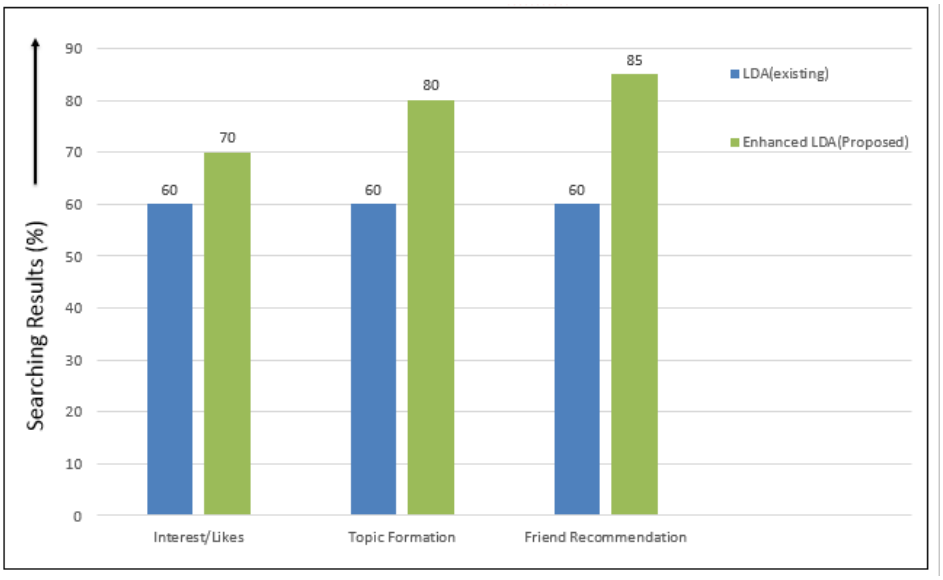

Fig 2. Comparison of Existing System with Proposed System

VIII. CONCLUSION AND FUTURE WORK

AnEnhanced LDA model is proposed which is a combination of Fenwick and LDA for topic finding to discover user preferences in microblogging systems. By identifying the topics of words, it makes easier to find such common interests. Thus, to recommend friend to users who are having same interest may increase users experience for information which they want to acquire. By finding users social behaviors and dynamics, it may help to find friend with similar interests, which improve the experience of user, social interactions, and can achieve more business value for organization. Probabilistic topic models have been proved powerful tools for identifying latent topics in the content. In order to handle huge number of topics FLDA works well which uses an appropriately modified Fenwick tree. Latent Dirichlet allocation model on twitter achieves the capacity of creating the topic distributions so that it can be used to generate unseen documents as well. Thus, FLDA is applied on twitter microblogging data for topic formation and friend recommendation.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

REFERENCES

1. Zheng, Nan, Shuangyong Song, and Hongyun Bao. "A Temporal-Topic Model for Friend Recommendations in Chinese Microblogging Systems." Systems, Man, and Cybernetics: Systems, IEEE Transactions on 45.9 (2015): 1245-1253.
2. Ramage, Daniel, Susan T. Dumais, and Daniel J. Liebling. "Characterizing Microblogs with Topic Models."ICWSM 10 (2010): 1-1.
3. Yu, Hsiang-Fu, Cho-Jui Hsieh, Hyokun Yun, S. V. N. Vishwanathan, and Inderjit S. Dhillon. "A scalable asynchronous distributed algorithm for topic modeling." In Proceedings of the 24th International Conference on World Wide Web, pp. 1340-1350. International World Wide Web Conferences Steering Committee, 2015.
4. Tu, Wenting, David W. Cheung, Nikos Mamoulis, Min Yang, and Ziyu Lu. "Activity-Partner Recommendation." In Advances in Knowledge Discovery and Data Mining, pp. 591-604. Springer International Publishing, 2015.
5. Silva, Nitai B., Ing-Ren Tsang, George DC Cavalcanti, and Ing-Jyh Tsang. "A graph-based friend recommendation system using genetic algorithm." InEvolutionary Computation (CEC), 2010 IEEE Congress on, pp. 1-7. IEEE, 2010.
6. Weng, Jianshu, Ee-Peng Lim, Jing Jiang, and Qi He. "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.
7. Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. "What is Twitter, a social network or a news media?." In Proceedings of the 19th international conference on World wide web, pp. 591-600. ACM, 2010.
8. Kim, Sundong. "Friend recommendation with a target user in social networking services."Data Engineering Workshops (ICDEW), 2015 31stIEEE International Conference on. IEEE, 2015.
9. Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. "Why we twitter: understanding microblogging usage and communities." In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56-65. ACM, 2007.
10. Gou, Liang, Fang You, Jun Guo, Luqi Wu, and Xiaolong Luke Zhang. "Sfviz: interest-based friends exploration and recommendation in social networks." In Proceedings of the 2011 Visual Information Communication-International Symposium, p. 15. ACM, 2011.
11. Song, Shuangyong, Qiudan Li, and Xiaolong Zheng. "Detecting popular topics in micro-blogging based on a user interest-based model."Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE, 2012.
12. Du, Zhao, Lantao Hu, Xiaolong Fu, and Yongqi Liu. "Scalable and explainable friend recommendation in campus social network system." InFrontier and Future Development of Information Technology in Medicine and Education, pp. 457-466. Springer Netherlands, 2014.
13. Wong, Felix Ming Fai, Zhenming Liu, and Mung Chiang. "On the efficiency of social recommender networks."Computer Communications (INFOCOM), 2015 IEEE Conference on. IEEE, 2015.
14. Song, Shuangyong, Qiudan Li, and Hongyun Bao. "Detecting dynamic association among Twitter topics."Proceedings of the 21st international conference companion on World Wide Web. ACM, 2012
15. Lehmann, Janette, Bruno Gonçalves, José J. Ramasco, and Ciro Cattuto. "Dynamical classes of collective attention in twitter." In Proceedings of the 21st international conference on World Wide Web, pp. 251-260. ACM, 2012.
16. Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. "Sentiment analysis of twitter data." In Proceedings of the workshop on languages in social media, pp. 30-38. Association for Computational Linguistics, 2011.
17. Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining."LREC. Vol. 10. 2010.
18. Yang, Dong-Hui, and Guang Yu. "A method of feature selection and sentiment similarity for Chinese micro-blogs."Journal of Information Science (2013): 0165551513480308.
19. Bao, Hongyun, Qiudan Li, Stephen Shaoyi Liao, Shuangyong Song, and Heng Gao. "A new temporal and social PMF-based method to predict users' interests in micro-blogging." Decision Support Systems 55, no. 3 (2013): 698-709.
20. Chen, Kailong, Tianqi Chen, Guoqing Zheng, Ou Jin, Enpeng Yao, and Yong Yu. "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.
21. Wan, Shengxian, Yanyan Lan, Jiafeng Guo, Chaosheng Fan, and Xueqi Cheng. "Informational friend recommendation in social media." InProceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, pp. 1045-1048. ACM, 2013.
22. Xie, Xing. "Potential friend recommendation in online social network." Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on and Int'l Conference on Cyber, Physical and Social Computing (CPSCom). IEEE, 2010.
23. Tang, Lei, and Huan Liu. "Community detection and mining in social Media."Synthesis Lectures on Data Miningand Knowledge Discovery 2.1(2010):1-137
24. Liu, Nathan N., Luheng He, and Min Zhao. "Social temporal collaborative ranking for context aware movie recommendation."ACM Transactions on Intelligent Systems and Technology (TIST) 4.1 (2013): 15.
25. Eirinaki, Magdalini, Malamati D. Louta, and Iraklis Varlamis. "A trust-aware system for personalized user recommendations in social networks."Systems, Man, and Cybernetics: Systems, IEEE Transactions on 44.4 (2014): 409-421.
26. Manca, Matteo, Ludovico Boratto, and Salvatore Carta. "Friend recommendation in a social bookmarking system: Design and architecture guidelines."Intelligent Systems in Science and Information 2014. Springer International Publishing, 2015. 227-242.
27. Zhao, Dejin, and Mary Beth Rosson. "How and why people Twitter: the role that micro-blogging plays in informal communication at work."Proceedings of the ACM 2009 international conference on Supporting group work. ACM, 2009.
28. Zhao, Wayne Xin, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. "Comparing twitter and traditional media using topic models." In Advances in Information Retrieval, pp. 338-349. Springer Berlin Heidelberg, 2011.
29. Samanthula, Bharath K., and Wei Jiang. "Structural and message based private friend recommendation."Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on. IEEE, 2012.
30. Hong, Liangjie, and Brian D. Davison. "Empirical study of topic modeling in twitter."Proceedings of the First Workshop on Social Media Analytics. ACM, 2010.

BIOGRAPHY

Prachi Jain is currently pursuing Masters in Computer Engineering at D. Y. Patil College of Engineering, Akurdi under Savitribai Phule Pune University. She has received Bachelor of Engineering in Computer Science from Dr. D. Y. Patil Institute of Engineering & Technology, SPPU, Pune in 2014. Her areas of interest are Information Retrieval and Social Networking.