State of Art Survey of Various User Recommendation Techniques

Divya Sharma
M.Tech Scholar
Department of Computer Engineering, YMCAUST,
Faridabad, India

Dr.Jyoti
Assistant Professor
Department of Computer Engineering, YMCAUST,
Faridabad, India

Abstract: User Recommendation is a technique which helps users in recommendations on the basis of their interests and preferences. The overall objective of user recommendation is to help users deal with the excess information by giving them appropriate recommendations.Various algorithms have been proposed for user recommendation by several researchers. This paper is a state-of-art survey of such techniques.

Keywords: recommendations; sentiments; twittomender; collaborative; signals

I. INTRODUCTION

Microblogging social networking sites have become popular in the past few years for exchanging and sharing different kinds of information among people. Twitter is far more than the total of its 500 million tweets. Due to enormously increasing volume of messages and increasing number of users , information overload has become a serious problem, so to overcome this problem , recommender systems came into picture.

A recommender system is a software tool that predicts the potential choices of user's that he/she may be interested in. Recommender system helps to get valuable information based on user's personal interests.

Various strategies have been proposed in the past in order to find most valuable users by using real Twitter data. Researchers explored different features of Twitter such as the relationships between users, the generated content, sentiments, signals etc. Twitter is a directed social graph that has follower-followee relationship between users. Several studies in the literature use Twitter's social graph for recommending new users in Twitter. Previous studies demonstrated that topological closeness had a positive effect on Twitter users.

Processing the growing volume of Twitter data to extract topics became significantly important for the recommender systems in order to provide tailored suggestions. Further analysis based on content was done to get more personalised recommendations and for finding relevant people . In content analysis part of the study , topic mining and sentiment analysis methods were used. In topic mining part , topics are extracted from tweets in order to find topical similarities between users and in sentiment analysis part, sentiment values of tweets are calculated in order to find similarities of opinion between users.

Given below is the literature review which discusses various user recommender systems required for predicting interesting users for the target user. Research in the area of user recommendation had been going on from the last few years. Next section is an effort to provide the survey of all the major works that has been done in recommendation of users.

II. LITERATURE REVIEW

This paper is an effort to bring the various approaches followed by researchers to make user recommender systems more effective. Various survey papers had also been published in this area. The next subsection highlights the various approaches in user recommendation.

A. *Recommendation based on content filtering*

[1] **Pazzani and Bilsus[1]** used user generated content on the web to extract preferences for recommendation.

Advantages : Requires less up-front classification work, since user models can be created automatically.
Disadvantages : Gives inaccurate recommendations if the content doesnot contain enough information to distinguish items that the user likes from the items that user doesn't like.

[2] **Chen et al.[2]** compared content based and relationship based algorithms in making user recommendations.

Advantages: Content based is better at finding known contacts whereas the relationship based is stronger at discovering new friends.
Disadvantages : Content based approach cannot discover new friends.

[3] **Chen et al.[3], Phelan et al.[4] and Esparza et al.[5]** used user generated content on Twitter to filter information streams.

Advantages : Allows a user based personalization, so the recommender system can determine the best recommendations for each user individually.
Disadvantages : Requires more computing power.

[4] **Guy et al.[6]** propose a user recommendation engine in the context of an enterprise social network scenario. They aggregate information about relationships as reflected in different sources within an enterprise such as co-authorship of papers , patents, projects, organizational chart relationships and others. These sources derive factors that might affect the similarity measure.

Advantages : Content based approaches are effective as opposed to relationship based algorithms, especially if histories of usage data in the social network are available.
Disadvantages : Approach limited to just an enterprise.

B. *Recommendation based on topology*

[1] **Weng et al.[8], Yamaguchi et al.[9]** used the topology of the social structure for recommending followees other than the content of the tweets. TwitterRank[8] tries to

identify influential users in certain topics by taking into account the topical similarity between users as well as the link structure, TURank[9] considered the social graph and the actual tweet flow.

Advantages : Suitable for information seekers.
Disadvantages : No assessment about target user's interest in recommendation.
  [2] **Kwak et al.[14]** analyzed the topological characteristics of Twitter and its power for information sharing. Some divergences were found between traditional human social networks and this follower/followees network like follower distribution exhibit a non-power law(users have more followers than predicted by power-law), the degree of separation is smaller than expected and there is a low reciprocity (most users in Twitter do not follow their followers back).

Advantage : Indegree proved to be a good representation of user's influence in Twitter using only its topology.
Disadvantage : Gives worst results for followee recommendation since people who are popular on Twitter will not necessarily match particular user interests.
  [3] **Armentano et al.[13]** combine collaborative filtering information, popularity and number of mentions to make user recommendations. The three features are combined naively by either taking their average or their product.

Advantages : Recommendations can be found quickly based on simple analysis of network structure.
Weighting feature, considering the number of occurences of user in the list of recommendations, gave better precision scores.
Disadvantages : Affect the final result negatively when one of the algorithms being combined is substantially inferior to others.
Ignores attributes of individuals, therefore cannot sufficiently explain preferences of users.

C. Recommendation based on sentimental analysis

  [1] Gaspare et al.[17] propose a user recommendation technique based on a novel weighting function, named sentiment volume-objectivity function (SVO) function, which takes into account not only user interests, but also his sentiments. The main idea behind this work is that users may share similar interests but have different opinions about them. Therefore, content based recommendation is extended by means of the sentiments and opinions extracted from the user micro-posts in order to improve the accuracy of the suggestions. This leads to define a novel weighting function in order to enrich content-based user profiles.

Advantages : Accuracy of suggestions is high.
To build better and complete user profiles as compared to traditional content based approach.
Disadvantages : No discussion of user's interest on topic being discussed by users.
Sentiment analysis is not suitable for short text(tweets).

D. Recommendation based on signals

  [1] **Arru at al.[16]** propose a user recommendation system on Twitter which relies on a signal based model, which explicitly includes a time dimension in the representation of user interests.Specifically, this model takes advantage of a signal processing technique, namely the wavelet transform for defining an efficient pattern-based similarity function among users. Such a system is based on a novel user model, termed bag-of-signal, which makes use of signal processing techniques to represent not only the number of occurences of the informative entities (concepts), but also the related time use patterns. The bag-of-signal user model involves modeling the user interests through a set of signals and the adoption of suitably defined similarity functions.

Advantage : Harnessing the time dimension guarantees better results in user profiling.
Disadvantage : Short term approach as user interests vary with time.

E. Diffusion based recommendation

  [1] **Sun et al.[7]** proposed a diffusion based microblogging user recommendation framework that identifies some users playing the role of news reporters and recommends them to information seekers during emergency events.

Advantage : Suitable for information seekers.
Disadvantage : It is a non personalised approach for non emergency events.

F. Hybrid approaches for recommendation

  [1] **Hannon et al.[10]** proposed a system for followee recommendation named Twittomender. Twittomender[10, 12] lets users find relevant profiles on Twitter exploiting different strategies : content based, collaborative based as well as a number of hybrid approaches after the user submitted an initial query of interest. Multiple profiling strategies were considered based on how users are represented in a content based approach (by their own tweets, by the tweets of their followees, by the tweets of their followers, by the combination of the three), a collaborative filtering approach (by the IDs of their followees, by the IDs of their followers or a combination of the two) and two hybrid algorithms. Indexing of user profiles and recommendations generated using a search engine. Then receiving a ranked list of relevant Twitter users based on a target user profile or a specific set of query terms. The way they combine their collaborative and content algorithms is by either combining their scores (authors do not specify how), or by using what seems to be a variation of Borda count(again, not fully specified). Hannon et al[11] presents a faceted profile structure that makes different types of interest more explicit.

Advantages : Delivers high quality recommendations based on real user data.
Analysis suggests that, more noisy the Twitter content is, more it provides a useful profiling and recommendation system.
Disadvantages : Disregard user measures in favor of content and collaborative techniques.
The hybrid versions of Twittomender is not as good as collaborative algorithm in terms of precision.
  [2] **Marco et al.[15]** proposed a holistic hybrid algorithm that takes into account content based, collaborative based and user based information. By using a Logistic Regression model, the algorithm learns how to combine different sources of evidence (including the ouput from other algorithms) from the data itself. Therefore, instead of manually determining the importance of each source, or worse weighting all the sources equally, the importance given to each of the sources in our model, comes

from the data. The algorithm is different from previous hybrid methods in three ways :

It is able to combine content (both word based and topic based), collaborative and user features simultaneously.

Instead of relying on naive ways of combining different sources of evidence, the algorithm learns how to combine the different features from the data in an optimal way.

It is straightforward to add or remove sources of evidence from the algorithm.

Advantage : Content based algorithms in this paper are more efficient than traditional baselines.

Disadvantage : It is not suitable when content doesnot contain sufficient information.

[3]**Cui et al.[18]** proposed a top-k followee recommendation over microblogging systems by exploiting diverse information sources. Since, explicit user to user ratings are not available on microblogging systems , thus existing collaborative schemes are not applicable to followee recommendation over microbloging systems.

To solve this problem, this paper proposed a novel followee ranking scheme using the variation of the latent factor model, which influence implicit users feedback including both tweet content and social relation information.

To achieve good top-k recommendation , a rank based criterion is introduced to latent factor model(LFM).

Advantages : LFM is proven to be effective for leveraging implicit information for efficient recommendation with scarcity of information.

Converges fast with sufficiently small number of rounds for model training.

Disadvantage : Not suitable for new users and inactive users having little information.

[4] **Yuan et al.[19]** proposed a microblog followee recommendation algorithm based on user interest degree and attribute characteristics after analyzing the structure of microblog recommendation, and this algorithm mainly works on the basis of user collaborative filtering recommendation algorithm to recommend new followee to user. It works as follows :

Firstly, users are grouped through calculating users interest degree to followee.

Secondly, users preferences value to followee is found with bayesian algorithm when they are with different attribute characteristics.

Finally, target user's nearest neighbor set is calculated using a new optimized similarity degree method to form the recommendation list.

This paper basically optimize the traditional user-based collaborative filtering recommendation algorithm.

Advantages : Enhance the effectiveness and accuracy of nearest neighbor set.

Recommendation quality improved.

Advantages : Enhance the effectiveness and accuracy of nearest neighbor set.

Recommendation quality improved.

Disadvantages : Data sparsity increases with increase in number of followers.

[5] **Liu et al.[20]** combined collaborative information and deep learning techniques to explore the user representations latent behind the topology and content. Topology and content are two important kinds of information available for modeling users in followee recommendation. In this work two neural networks are designed :

First, explore the topology-based latent representation of users and calculate the topology relevance scores between users.

Second, mine users content based representations and produce the content based relevance scores.

The improvement of these two neural networks is that the collaborative information including followers of followees and followees of followers is naturally modelled. Further, an adaptive layer is designed to tune the contribution of topology and content in followee recommendation and dynamically predict the following relation between any two users.

Advantage : Experiments verify that this method can efficiently model users.

Disadvantage : Not suitable for inactive and new users.

[6] **Mangal et al.[21]** proposed an approach that combines both sentiment analysis and classification. Thus, the topic is extracted in which users are interested. The sentiments of emoticon and acronym are also considered.

Advantage : Improves the recommendation list by combining sentiments with classification.

Disadvantage : Difficult to classify the topic when the content has limited information.

A critical survey of the state of the art user recommendation systems in Twitter brings out the comparisons which are highlighted in the next section. A comparison table has been presented for this purpose.

III. COMPARISON TABLE OF VARIOUS RECOMMENDATION APPROACHES

This section consists of a comparison tables which highlights the different aspects used in various user recommendation techniques and discussed above in the literature survey.

Table I : Comparison Table of Factors Used in User Recommendation System

S.N	Approach	User Based features	Content based	Collaborative based	Topology based
1.	Pazzani [1]		✓		
2.	Chen et al.[3], Phelan et al.[4] Esparza et al. [5]		✓		
3.	Guy et al. [6]		✓		
4.	Weng et al.[8] Yamaguchi et al..[9]				✓
5.	Hannon et al[10]		✓	✓	
6.	Kwak et al. [14]				✓
7.	Armentano et al [13]				✓
8.	Marco et al.	✓	✓	✓	

	[15]					
9.	Cui et		✓		✓	

IV. CONCLUSION

User Recommendation techniques are used to minimize the user's web access latency using the keywords on the web document.Various user recommendation techniques using content filtering, topology, sentimental analysis, hybrid based have been discussed in this paper with their pros and cons. A lot of research has been going on in this area from the past few years. Various algorithms have been proposed but still there is a lot of scope for improvement.

V. REFERENCES

[1] Pazzani, M.J. and D. Billsus, 2007. Content-Based Recommendation Systems. In: The Adaptive Web: Methods and Strategies of Web Personalization, Brusilovsky, P., A. Kobsa and W. Nejdl (Eds.). Springer, New York, pp: 325-341.

[2] [Chen et al., 2009] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy. Make new friends, but keep the old: recommending people on social networking sites. In Proceedings of the 27th International Conference on Human Factors in Computing Systems, pages 201–210, 2009.

[3] Chen, J., R. Nairn, L. Nelson, M. Bernstein and E. Chi, 2010. Short and tweet: Experiments on recommending content from information streams. Proceedings of the 28th International Conference on Human Factors in Computing Systems, April 10-15, 2010, Atlanta, Georgia, USA., pp: 1185-1194.

[4] Esparza, S.G., M.P. O'Mahony and B. Smyth, 2010. On the real-time web as a source of recommendation knowledge. Proceedings of the 4th ACM Conference on Recommender Systems, September 26-30, 2010, Barcelona, Spain, pp: 305-308.

[5] Phelan, O., K. McCarthy and B. Smyth, 2009. Using Twitter to recommend real-time topical news. Proceedings of the 3rd ACM Conference on Recommender Systems, October 23-25, 2009, New York, USA., pp: 385-388.

[6] Guy, I., I. Ronen and E. Wilcox, 2009. Do you know?: Recommending people to invite into your social network. Proceedings of the 14th International Conference on Intelligent User Interfaces, February 8-11, 2009, Sanibel Island, Florida, USA., pp: 77-86

[7] [Sun et al., 2009] A. R. Sun, J. Cheng, and D. D. Zeng. A novel recommendation framework for micro-blogging based on information diffusion. In Proceedings of the 19th Workshop on Information Technologies and Systems, 2009.

[8] Weng, J., E. Lim, J. Jiang and Q. He, 2010. TwitterRank: Finding topic-sensitive influential twitterers. Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, February 3-6, 2010, New York, USA., pp: 261-270.

[9] Yamaguchi, Y., T. Takahashi, T. Amagasa and H. Kitagawa, 2010. TURank: Twitter user ranking based on user-tweet graph analysis. Proceedings of the 11th International Conference on Web Information Systems Engineering, December 12-14, 2010, Hong Kong, China, pp: 240-253.

[10] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In Proceedings of the fourth ACM Conference on Recommender Systems, RecSys '10, pages 199–206, New York, NY, USA, 2010. ACM.

[11] J. Hannon, K. McCarthy, M. P. O'Mahony, and B. Smyth. A multi-faceted user model for twitter. In Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization, UMAP'12, pages 303–309, Berlin, Heidelberg, 2012. Springer-Verlag.

[12] J. Hannon, K. McCarthy, and B. Smyth. Finding useful users on twitter: twittomender the followee recommender. In Proceedings of the 33rd European Conference on Advances in Information Retrieval, ECIR'11, pages 784–787, Berlin, Heidelberg, 2011. Springer-Verlag.

[13] Armentano, M.G., D.L. Godoy and A.A. Amandi, 2011. A topology-based approach for followees recommendation in Twitter. Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems, July 16, 2011, Barcelona, Spain

[14] [Kwak et al., 2010] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In Proceedings of the 19th International Conference on World Wide Web (WWW'10), pages 591–600, 2010.

[15] Sara Guimarães, Marco Túlio Ribeiro, Renato Assunção and Wagner Meira Jr.,2013. A Holistic Hybrid Algorithm for User Recommendation on Twitter. Journal of Information and Data Management, Vol. 4, No. 3, October 2013, Pages 341–356.

[16] Giuliano Arru, Davide Feltoni Gurini, Fabio Gasparetti, Alessandro Micarelli and Giuseppe Sansonetti 2013. Signal-Based User Recommendation on Twitter. In Proceedings of the ACM Conference on Recommender Systems, May 13–17, 2013, Rio de Janeiro, Brazil.

[17] Davide Feltoni Gurini, Fabio Gasparetti, Alessandro Micarelli and Giuseppe Sansonetti 2013. A Sentiment-Based Approach to Twitter User Recommendation.

[18] Hanhua Chen∗,Xiaolong Cui,Hai Jin 2014 Top-k followee recommendation over microblogging systems by exploiting diverse information sources.

[19] Rongsheng Wang, Aihui Liu∗, Fuyong Yuan, 2014. The Followee Recommendation Algorithm Based on Microblog User Interest and Characteristic. Journal of Information & Computational Science 11:5 (2014) 1585–1596 March 20, 2014.

[20] Yang Liu1, Xuan Chen2, Sujian Li1,3 and Liang Wang1 ,2015 A User Adaptive Model for Followee Recommendation on Twitter.

[21] Nimita Mangal1, Rajdeep Niyogi1(&), and Alfredo Milani2 ,2016 .Analysis of Users' Interest Based on Tweets. Springer International Publishing Switzerland 2016, pp. 12–23, 2016.

[22] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy. Make new friends, but keep the old: recommending people on social networking sites. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, pages 201–210, New York, NY, USA, 2009. ACM.