

Friend Recommendation in Microblogging System

1. Prof.N.B.Kadu., Asst.prof, Information Technology, PREC Loni, Maharashtra, India
2. Kakad Snehal G., Information Technology, PREC Loni, Maharashtra, India
3. Tambe Harshda S., Information Technology, PREC Loni, Maharashtra, India

ABSTRACT

In the current time, people are becoming more communicative through extension of facilities and multiplatform applications, which establish social and collaborative backgrounds. As the popularity of Web 2.0 technologies are increasing, microblogging sites like Twitter is a new form of online communication in which users talk about their daily lives, publish opinions and share information by posts, has become one of the most popular social networking site today. 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. Probabilistic topic models have been proved to be the powerful tools for recognizing latent text patterns in the content. The purpose is to discover the interests of each user rather than the topics of single messages and on the basis of user's interest, friend is recommended.

Keywords–Micro blogging, friend recommendation, temporal model, interest drifts.

1. INTRODUCTION

Micro blogging 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 properties 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 sentiment analysis, influential people, and personalized recommendations

Because of the brief frame and developing ubiquity the microblogging is turning into people's more interesting choice for seeking the information and expressing opinions. Messages got by a user mainly rely on whom user follows. Therefore, blessing user with comparable interest may enhance the experience quality for information receiving. Since messages posted by microblogging users reflect their hobbies and interest the essential keywords in the messages show their main focus to a huge extent, the users' preferences can be find by investigating the user generated contents. Besides, user's hobbies, interest are dynamic, despite what might be expected, they change as time passes by. In light of such instincts, we proposed a topic model with respect to time to analyze user's possible behaviours' and predict their potential friends in microblogging. The model takes in user's latent preferences by extracting keywords on aggregated messages over a stretch of time by means of a topic typical, and after that the effect of time is considered to deal interest.

1.1 PROBLEM STATEMENT

The Problem is to determine the user's social behaviors and dynamics, it may help them to find friends with similar interests, which may improve the user's experience, social interactions, and gain more business value for corporations.

1.2 LITERATURE SURVEY

The persistent popularity of microblogging systems has attracted many researchers' attention. Early researchers mainly focused on the characteristics of microblogging and social network analysis. Recently, there has been an increasing interest in the field of information retrieval, such as event detection and tracking identification of influential people, sentiment analysis, and personalized recommendations. In this section, related work is introduced on topic finding approaches, user's interest analysis, recommendation system and time-aware recommendations.

2. RELATED WORK

The popularity of microblogging systems has attracted many researchers attention. Early research focused on the characteristics of microblogging and analysis of social networks. In the area of information retrieval, there is recently interest is increased include tracking and event detection finding of influential people, analysis Of sentiment, and personalized recommendations.

2.1 Keyword Extraction

Keyword extraction is most widely used and essential method for user modelling. In previous research simply a bag-of-words is build based on user's microblogs. Discover the interests of Sina Weibo users using keyword extraction. To find structured data from noisy and unstructured data is a challenging task. One of the tasks of information extraction is named entity recognition (NER). To recognize types of entities on text documents, rule-based approaches use heuristic rules to identify named entities from documents in a particular domain. Machine learning is a way to automatically learn to identify complex patterns and make intelligent conclusions based on data. Supervised learning uses only the labelled data to create a model. Semisupervised learning combine both the labelled data as well as useful information from the unlabelled data in learning. Unsupervised learning is designed to learn without or with very few labelled data.

2.2 Topic Finding

To discover the user interest, only keyword extraction is not sufficient. Xu et al. 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 user's intrinsic interest was represented by a distribution over latent topics. In-degree method is used to identify the influential twitterers. Twitter-Rank , 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.

2.3 Times-Aware Recommendations

The temporal information is important when discovering user's interests. Rafehand Bahrehmand proposed an adaptive collaborative filtering algorithm which uses time to reflect fluctuations in user's behaviour over time. Liu et al. Developed a social temporal collaborative ranking model to recommend movies. To cope with current topic recommendations in microblogging system, in a probability matrix factorization is proposed based on the evolution of user's interest. All the above studies have demonstrated the importance of temporal information when discovering user's interests, but little has been done on recommendations in microblogging based on temporal information. So in this work Latent Dirichlet allocation (LDA) achieves the capacity of generalizing the topic distributions so that the model can be used to discover user's preferences.

2.4 System Architecture

A temporal-topic model is proposed to predict user's potential friends. The model first extracts user's topic distributions from keyword usage patterns of aggregated messages using temporal approach. Then, it calculates user similarities over time based on users topic distributions. Finally, user's potential interests on others are predicted according to user similarities over deferent periods of time via temporal functions based on topic model, and then friend recommendation is conducted based on user predicted scores. If a user reports others messages without any comments, then System will add forwarding microblogs automatically. Such a denotation does not have any effect

on user's interests; therefore, we remove it from messages, since reposts Messages, but keep the content of the reposted messages, since reposts represents User's interests on the related content.

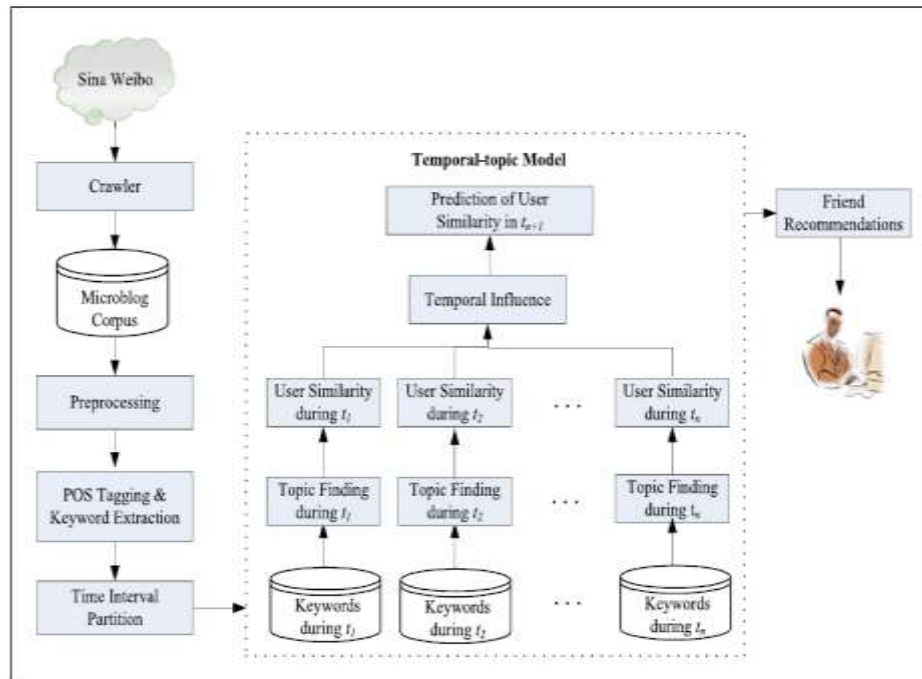


Fig.1 Overview of the temporal topic Friend Recommendation framework

3. METHODS

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 users preferences on friend selection in real life. So to overcome this we propose a enhanced LDA-model to predict users potential friends.

3.1 LDA Model

The general idea of Latent Dirichlet Allocation (LDA) is based on the hypothesis that a person writing a document has definite topics in mind. This model discovers user's topic distributions according to their keyword usage patterns. LDA is a reproductive and an unsupervised model that proposes a stochastic procedure by which the documents which contain words are produced using latent semantic topics in huge collections of text documents. So Latent topics are discovered by recognizing group of words in the corpus that repeatedly occur together within documents. In LDA, learning is unsupervised because the input data is partial, the corpus make available only the words within documents.

3.2 TM-LDA Model

The LDA model takes into account clusters of co-occurrence of words and do not consider temporal information related to these clusters. So in TM-LDA that is temporal LDA consider temporal information associated to cluster of words. Activities discovered in the time range depicts where they are dominant. Unfortunately it is very difficult to choose granularity of time slice which is suitable for all activities. The time slice chosen may be too large for some activities or too small for others. In order to avoid this problem a continuous distribution can be associated over time for each activity. The parameterized distribution chosen for time is Beta distribution, which defines a probability distribution over a normalized time range from 0 to 1.

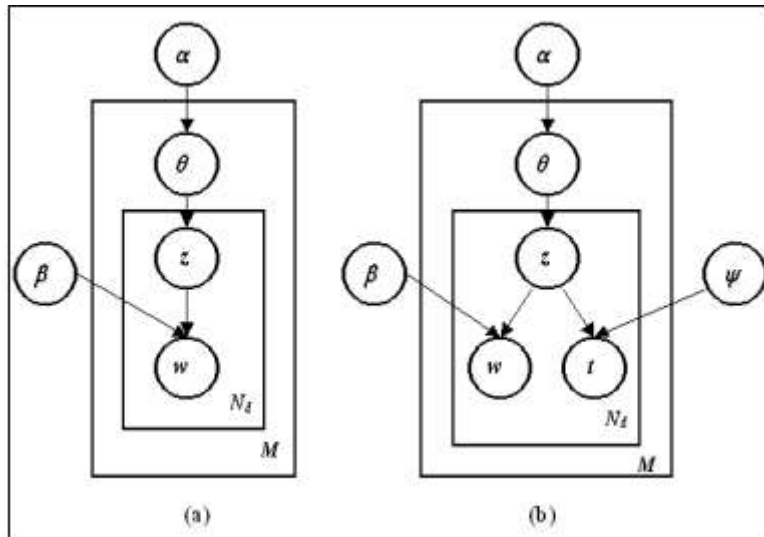


Fig.2 Graphical Representation of Topic Models
(a) LDA (b) Time LDA

3.3. RESULT ANALYSIS

Output of LDA system is compared with Enhanced LDA. Different sample of tweets are taken and on the basis of those sample, user's interest are extracted. User's interest shows their area of interest which helps to find friend with similar interest. Enhanced LDA is better than LDA in Friend Recommendation in terms of accessing data.

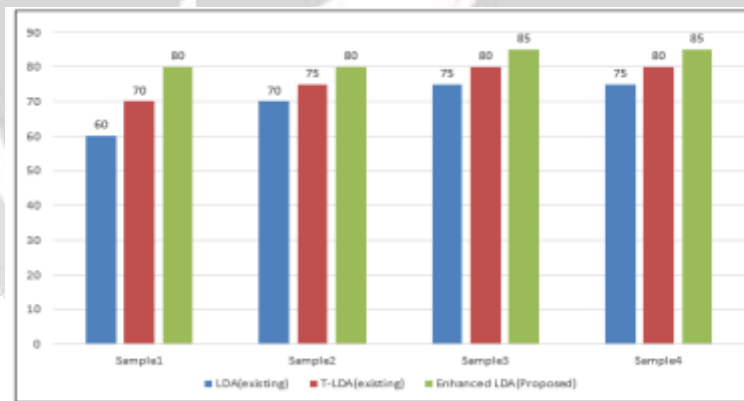


Fig 3: Efficiency of proposed system against various existing systems.

Output of LDA system is compared with Enhanced LDA on different parameters such as interest/likes, topic formation and friend recommendation. 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 topic formation.

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.

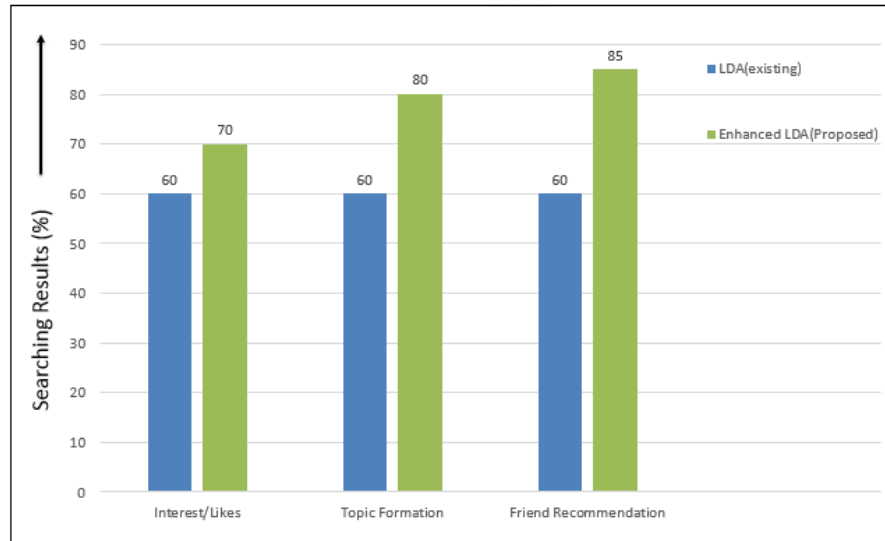


Fig 4: Comparison of Existing System with Proposed System

3.4. Implementation and Experiments

3.4.1. Preprocessing

Data of real world is incomplete, noisy, inconsistent, lacking certain attributes of interest or contain some aggregate data so data preprocessing is required. If user writes some post then the stop words are removed from post. In addition, URLs and other non texts are also removed from microblogs. If user repost same message then, only single data is stored in database and duplicate data is removed.

3.4.2. Time Interal Partition

Users' interests change as time goes by, which reveals in twitter that user's microblogs may focus on different topics at different periods of time. Therefore, users dynamically changing interests can be expressed as a sequence of keyword collections in a single topic. As time changes, the keywords under a topic also change. Microblogs at different time intervals,

3.4.2. Topic Sampling

Topic Sampling is the representation of topic and assignment of keywords to the particular topic. Topics which are formed in topic formation are converted into tree called Fenwick Tree. Tree representation of topic makes insertion, deletion and accessing of keyword easy. Fenwick tree gets updated each time when new keywords added into corpus.

3.4.3. Topic/User Similarity Matching

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. User's potential interests on others are predicted according to user similarities where topics of one user are compared with topics of other users.

3.4.4. Friend Recommendation

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, friend recommendation is based on topic similarity; if topics of user1 are matched with user2 then recommendation takes place.

4. CONCLUSION

An Enhanced LDA model is proposed which is a combination of Fenwick and LDA for topic finding to discover user preferences in microblogging systems. 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. In order to handle huge number of topics FLDA works well which uses an appropriately modified Fenwick tree. Thus, FLDA is applied on twitter microblogging for topic finding and friend recommendation.

5. REFERENCES

- [1] Zheng, Nanning, Seunghyun Song, and Huihui Bao. "A Temporal-Topic Model for Friend Recommendations in Chinese Microblogging Systems."
- [2] Ramage, Daniel, Susan T. Dumais, and Daniel J. Liebling. "Characterizing Microblogs with Topic Models." *ICWSM 10* (2010): 1-1.
- [3] Wang, Fei-Yue, et al. "Social computing: From social informatics to social intelligence." *Intelligent Systems, IEEE* 22.2 (2007): 79-83.
- [4] Bikel, Daniel M., Richard Schwartz, and Ralph M. Weischedel. "An algorithm that learns what's in a name." *Machine learning* 34.1-3 (1999): 211-231.
- [5] Reiss, Frederick, et al. "An algebraic approach to rule-based information extraction." *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on. IEEE, 2008.*
- [6] Weng, Jianshu, et al. "Twitterrank: finding topic-sensitive influential twitterers." *Proceedings of the third ACM international conference on Web search and data mining. ACM, 2010.*
- [7] Kwak, Haewoon, et al. "What is Twitter, a social network or a news media?." *Proceedings of the 19th International conference on World Wide Web. ACM, 2010.*
- [8] Kim, Sundong. "Friend recommendation with a target user in social networking services." *Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference on. IEEE, 2015.*
- [9] Java, Akshay, et al. "Why we twitter: understanding microblogging usage and communities." *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, 2007.*
- [10] Gou, Liang, et al. "Sviz: interest-based friend's exploration and recommendation in social networks." *Proceedings of the 2011 Visual Information Communication-International Symposium. 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, et al. "Scalable and explainable friend recommendation in campus social network system." *Frontier and Future Development of Information Technology in Medicine and Education. Springer Netherlands, 2014. 457-466.*
- [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.*