Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 6, June 2017, pg.27 – 35

A Result Paper on Temporal-Topic Model for Friend Recommendations in Chinese Micro Blogging Systems

Reshma Bhondave[1], Nilesh Biradar[2], Poonam Patil[3], Pallavi Malik[4], Prof. Prashant Dongare[5]

Dept.Of Comp, SIT Lonawala, Pune

Abstract— Because of the temporary frame and developing omnipresence the Microblogging is popping into people's most attention-grabbing selection for seeking the data and expressing opinions. Messages got by a user mainly trust on whom user follows. Therefore, recommending user with comparable interest may enhance the expertise quality for data receiving. Since messages posted by Microblogging users replicate their hobbies or interest and the essential keywords within the messages show their main focus to a large extent, we will realize users' preferences by investigation the user generated contents. Besides, user's hobbies, interest are not static; despite what can be expected, they change as time passes by. In light of such instincts, we projected a temporal-topic to analyze user's doable behaviors' and predict their potential friends in Microblogging. The model takes in users' latent preferences by extracting keywords on aggregated messages over a stretch of time by suggests that of a subject model, and after that the impact of your time is taken into account to deal interest.

Keywords— Microblogging; Latent Dirichlet Allocation (LDA); Singular Value Decomposition (SVD); Crawler

I. Introduction

Micro blogging has become a convenient way for web surfers and average users to communicate with their friends and members of the family, or to express intimate emotions or feelings. Using a microblog conjointly has step by step become a habit for a huge quantity of users, which leads to associate exponential explosion of knowledge within the virtual microblog society on the web, making retrieving and distinctive required microblog or connected data very troublesome. Therefore, additional and more microblog services square measure developing novel engines dedicated to recommending user-specific data. Early researchers mainly centered on the characteristics of Micro blogging and social network analysis. Recently, there has been an increasing interest in the field of knowledge retrieval, such as event detection and tracking, identification of influential folks, sentiment analysis, and personalized recommendations. Traditional recommendation systems will primarily be classified into 3 categories: CF-based, content-based, and hybrid recommendation systems Probabilistic topic models have been proved to be the powerful tools for distinctive latent text patterns within the content. Latent Dirichlet allocation (LDA) achieves the capacity of generalizing the topic distributions so the models are often wont to generate unseen documents further. LDA has also been applied to numerous works on Twitter to demonstrate its utility. Users' interests are not static;

contrarily, their interests may amendment as time goes by. Since the real-time and brevity options of Micro blogging lead to frequent updates of micro blog, users' interests are additional in depth and changeable over time.

II. LITERATURE SURVEY

*A. **Barbosa and Feng (2010)***

The significant effort for sentiment classification on Twitter information is given by Barbosa and Feng. They use polarity predictions from three websites as rackety labels to train a model and use one thousand manually labeled tweets for standardization and another one thousand manually labeled tweets for testing. They however ever do not mention how they collect their check information. They propose the use of syntax features of tweets like retweet, hash tags, link, punctuation and exclamation marks in conjunction with features like previous polarity of words and POS of words. We extend their approach by victimization real valued previous polarity, and by combining prior polarity with POS. Our results show that the options that enhance the performance of our classifiers the most are features that mix previous polarity of words with their components of speech. The tweet syntax features facilitate however solely marginally.

B. Gamon (2004)

Perform sentiment analysis on feedback data from international Support Services survey. One aim of their paper is to analyze the role of linguistic features like POS tags. They perform extensive feature analysis and feature choice and demonstrate that an abstract linguistic analysis option contributes to the classifier accuracy. In this paper we perform in depth feature analysis and show that the employment of solely a hundred abstract linguistic options performs yet as a tough unigram baseline.

C. Brin & Page

Brin& Page has introduced the Page Rank algorithm. Pre-computes a rank vector that provides a priori authority estimates for all of the nodes during a given graph. The node authority is independent of the attributes of every node associate degreed such an authority lives solely emerges from the topological structure of the graph. In particular, the authority of a node m depends on the number of incoming links and on the authority of the nodes that purpose to m with forward links. In this paper, a PageRank based model is planned to discover the most fashionable topics in micro-blogging supported users' interest relationship. The model first detects the favorite topics of every user with vote theory, then creates the links between topics with users' attentiveness relationship to create the 'topic graph' in the entire micro-blogging social network, finally, ranks those topics with Page Rank algorithm to notice the foremost fashionable ones in micro-blogging.

D. Chakrabart

Link-based ranking has contributed meaningfully to the success of source search. Chakrabarti et al. built associate automatic resource compiler by analyzing link structure and associated text. Given a topic that's broad and well represented on the net, it will hunt down and come a listing of net resources that it considers the foremost authoritative for that topic.

E. Giles

Giles proposed the Cite Seer, a system for searching research papers and creating a digital library. This system offers the users features similar to DBLP1 with a different approach to generate the library: it retrieves new documents, automatic tags and links metadata information inherent in an academic documents syntactic structure.

F. Zheng&Xie

Zheng & Xie proposed a HITS-based model to infer the interest level of a location and a user's travel expertise (knowledge), for recommending a user with top attention-grabbing locations and travel categorizations in a given geospatial area. Different from earlier link-based ranking researches, which were chiefly primarily based on specific links among entities, we discover implicit links among topics considering users' interest relationships to produce a subject graph for link-based topic ranking. Topic ranking aims to order a list of the topics by their popularity, semantic connectedness or some alternative evaluating indicators. Wang et al. proposed associate degree automatic on-line news topic ranking algorithmic rule that is in a position to rank topics from on-line news streams through the construct of burstiness. This algorithm is primarily based on inconsistency analysis between media focus and user attention. News stories are organized into topics, which are graded in terms of each media focus and user attention. Wu et al. offered a new perspective by exploring the potential inter-modal relationships derived between near duplicate and matter info for topic chase and re-ranking.

III. PROPOSED SYSTEM

We recommend a temporal-topic model to predict user's potential friends. The model first mainly extracts user's topic distributions from keyword usage patterns of aggregative messages exploitation temporal method. Then, it calculates user likenesses over time established on users topic distributions. Finally, users potential interests on others are foreseen according to user similarities over completely different periods of your time via temporal functions supported topic model, we conduct friend recommendation to user foreseen scores.If a user reports others messages without any comments, then system will add "forwarding microblogs" mechanically. Such a denotation does not have any result on user's interests; so, we take away it from messages, since reposts messages, but keep the content of the reposted messages, since reposts represents users interests on the related content.

Advantages of Proposed System

1. It is effective recommendation system for recommending friends to user.
2. It takes less time because of the effectiveness of LDA algorithm.

IV. System Architecture

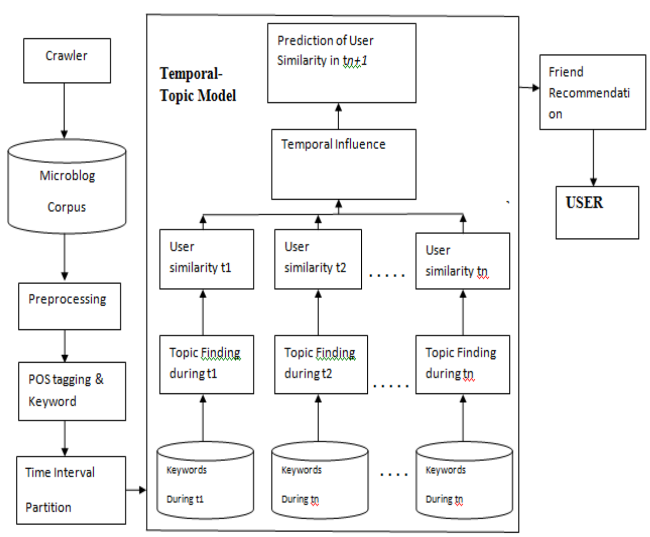


Fig 1. System Architecture of Proposed System

V. Mathematical Model

Let W be the whole system which consists:

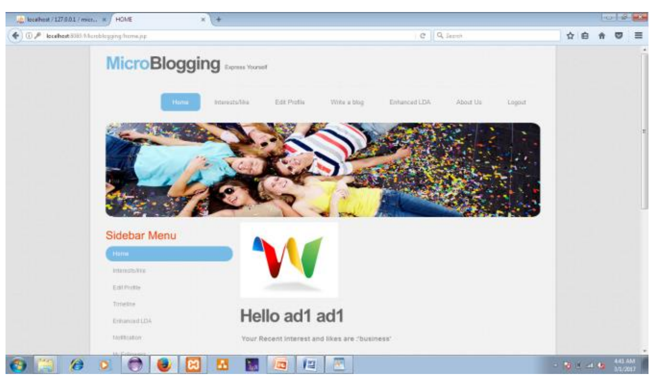
W= {U, W, Nu ,Nw, W, t, T , β, α, θ, γ, δ, I, n, w, S, M}.

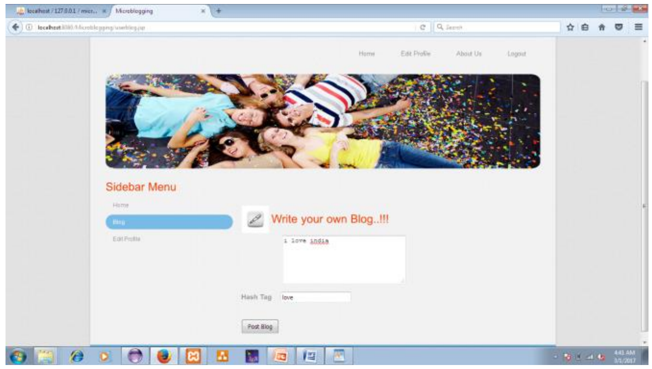
Where,

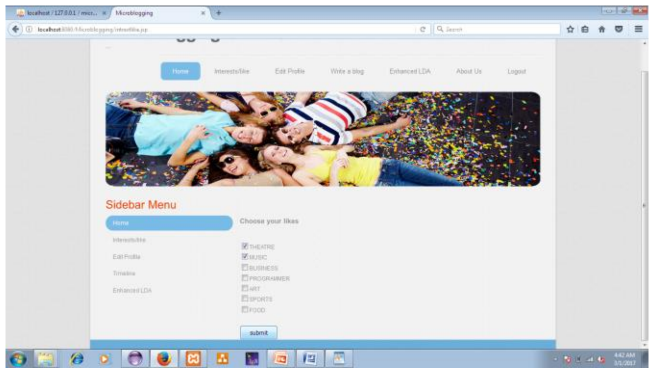
1. U is the set of user.
2. W is the set of keywords.
3. Nu is the set of total number of user.
4. Nw is the set of total number of keywords.
5. t is the time interval.

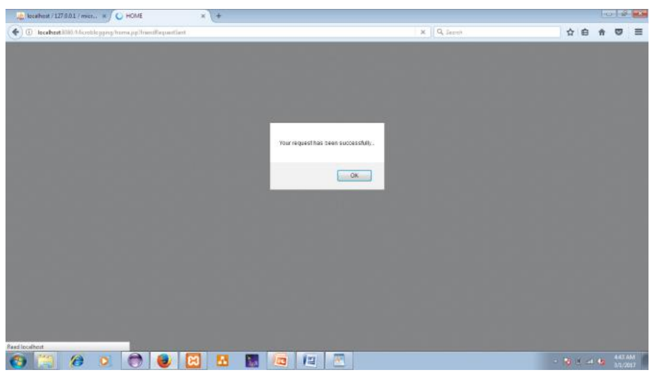
6. $N_w^u\ (t)$is the total number of keywords of user u at time t.
7. T is the set of number of topics.
8. n is the total number of time intervals.
9. β is the Dirichlet prior for user.
10. α is the Dirichlet prior for user
11. $\alpha^t$is the Dirichlet prior for users at time t.
12. $\beta^t$is the Dirichlet prior for hidden topics at time t.
13. γ is the kernel parameter in the exponential decay function.
14. $\delta$is the size of time interval.
15. $w_i^t$is the unique word associated with the i-th token of useru at time t.
16. $z_i^t$is the topic associated with w_i^t.
17. θ is the multinomial distribution of particular topic.
18. θ_u(t) is the multinomial distribution topic specific to the user u at time t.
19. θ_z(t) is the multinomial distribution words specific to the topic z at time t.
20. S be the similarity matrix.
21. St is the users topical similarity matrix at time t.
22. I is the number of iterations in LDA model.
23. M is the keyword matrix.
24. Mt be the users keyword matrix at time t.

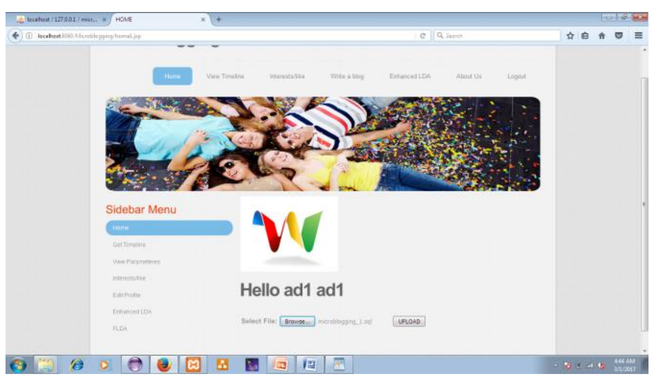
VI. Result Analysis



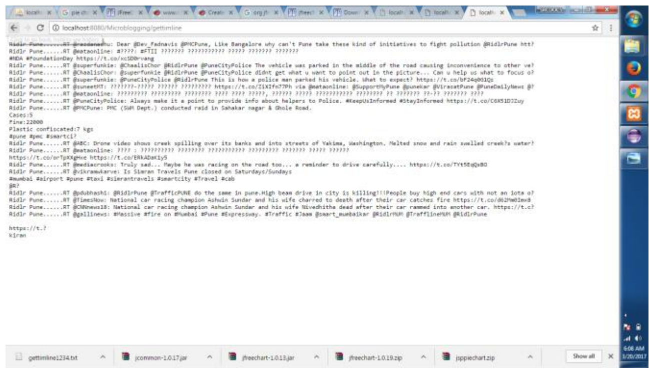

Sidebar Menu
Write your own Blog..!!!

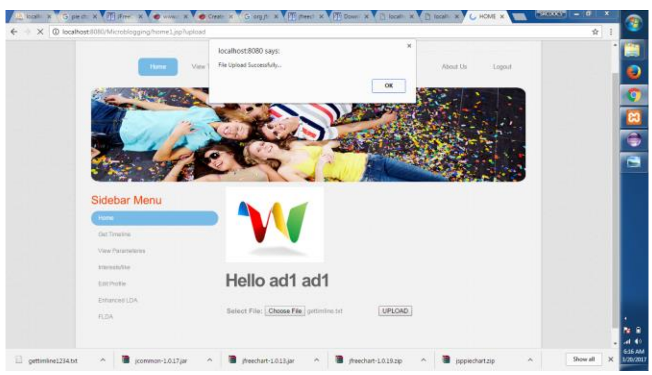

Sidebar Menu

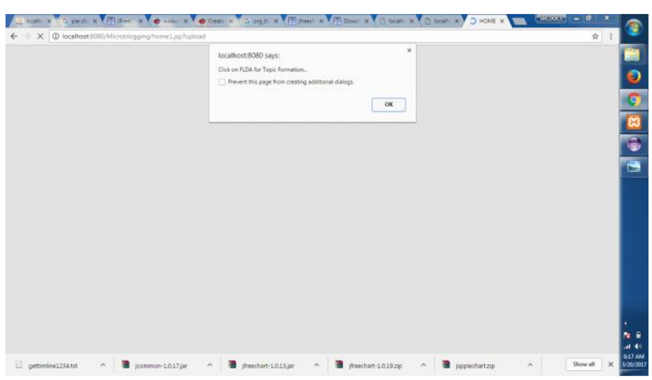



Sidebar Menu
Hello ad1 ad1

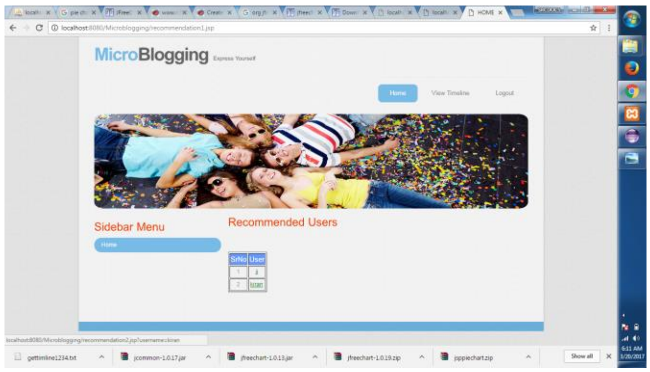

MicroBlogging
Sidebar Menu
Hello ad1 ad1

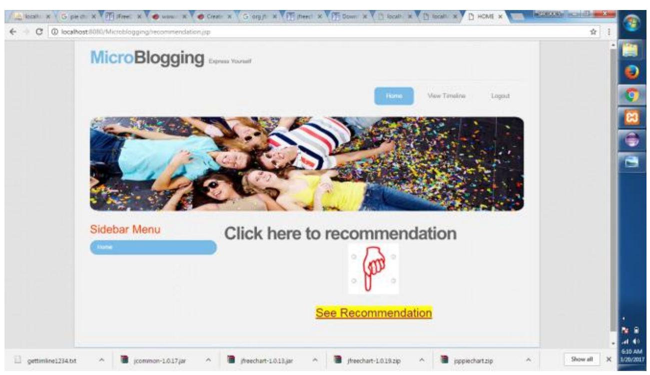


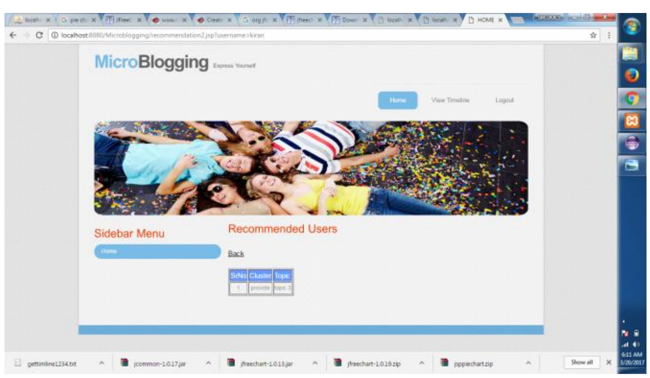

Sidebar Menu
Hello ad1 ad1
UPLOAD




Sidebar Menu
Hello ad1 ad1
UPLOAD


MicroBlogging
Sidebar Menu
Recommended Users


MicroBlogging
Sidebar Menu
Click here to recommendation
See Recommendation


MicroBlogging
Sidebar Menu
Recommended Users
Back

VII. Conclusion

In this project, we propose a temporal-topic model for friend recommendations in Chinese microblogging systems. The model first discovers users' latent preferences throughout completely different time intervals primarily based on keywords extracted from the aggregative microblogs through a subject model. Then, it calculates user similarities in each time interval supported temporal topic distributions. After that, an exponential decay operate is employed to live interest drifts. Finally, users' potential interests on others can be foreseen supported the sequence of users' interests on the timeline. Based on the model, we conducted friend recommendations and the experimental results showed that our model is effective.

For future work, we arrange to conduct our experiments on users who have less friends and followers to point out if our model is helpful for the cold-start downside of customized recommendations. We additionally aim to unearth different factors to enhance the performance of the projected model, such as social relationships among users (i.e., followers, follows), the sentiment of microblogs, users' location information, etc. We additionally arrange to investigate different progressive models with temporal evolvement and compare the performances of various ways on friend recommendations. Other datasets such as Twitter are going to be tested for the utility and effectiveness of the model.

Acknowledgement

I would prefer to give thanks the researchers likewise publishers for creating their resources available. Additionally, I appreciative to commentator for their vital recommendations moreover I am conjointly grateful to reviewer for their valuable suggestions and also thank the college authorities for providing the required infrastructure and support.

References

[1] F.-Y. Wang, "Toward a paradigm shift in social computing: The ACP approach," *IEEE Intell. Syst.*, vol. 22, no. 5, pp. 65–67, Sep./Oct. 2007.

[2] F.-Y. Wang, K. M. Carley, D. Zeng, and W. Mao, "Social computing: From social informatics to social intelligence," *IEEE Intell. Syst.*, vol. 22, no. 2, pp. 79–83, Mar./Apr. 2007.

[3] C. L. P. Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Inf. Sci.*, vol. 275, pp. 314–347, Aug. 2014.

[4] M. Moricz, Y. Dosbayev, and M. Berlyant, "PYMK: Friend recommendation at myspace," in *Proc. ACM SIGMOD Int. Conf. Manage. Data.*,Indianapolis, IN, USA, 2010, pp. 999–1002.

[5] L. Hong and B. D. Davison, "Empirical study of topic modeling in Twitter," in *Proc. 1st Workshop Soc. Media Anal.*, Washington, DC, USA, 2010, pp. 80–88.

[6] D. Ramage, S. Dumais, and D. Liebling, "Characterizing microblogs with topic models," in *Proc. Int. AAAI Conf. Weblogs Soc. Media*, Menlo Park, CA, USA, 2010, pp. 130–137.

[7] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. 19th Int. Conf. World Wide Web*, Raleigh, NC, USA, 2010, pp. 591–600.

[8] D. Zhao and M. B. Rosson, "How and why people Twitter: The role that micro-blogging plays in informal communication at work," in *Proc. ACM Int. Conf. Support. Group Work*, Sanibel Island, FL, USA, 2009, pp. 243–252.

[9] A. Java, X. Song, T. Finin, and B. Tseng, "Why we Twitter: Understanding microblogging usage and communities," in *Proc. 9thWebKDD 1st SNA-KDD Workshop Web Min. Soc. Netw. Anal.*, San Jose, CA, USA, 2007, pp. 56–65.
[10] W. X. Zhao *et al.*, "Comparing Twitter and traditional media using topic models," in *Advances in Information Retrieval*. Berlin, Germany: Springer, 2011, pp. 338–349.

