International Journal of Engineering & Technology, 7 (2.33) (2018) 832-834

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Finding of experts using behavioral aspects

Sreelekshmi.U [1] *, Gopu Darsan [2]

[1] *Mtech scholar Sree Buddha College of Engineering Alappuzha, India*
[2] *Asst. Professor Sree Buddha College of Engineering Alappuzha, India*
**Corresponding author E-mail: lekshmishibin@gmail.com*

Abstract

As the social networks emerged like Twitter, the process of exploring experts has become an interesting topic. However, previous methods can never be used to learn about topic experts in Twitter. Some of the new methods make use of the relations existing between Twitter lists and users for exploring experts. A probabilistic method has been developed to explore the relations (i.e. follower, user-list and list-list relations) for finding experts. A Semi-Supervised Graph-based Ranking (SSGR) method is used to find the users global authority. Between users and given query a local relevance is computed. By understanding the global authority and local relevance of users, all of them are ranked and those with high scores for the ranking are retrieved which constitute the expert extraction. On the other hand a behavior extraction is done with respect to understandability, level of detail and writing style which contributes to the feature set. This feature extraction leads to the SVM (Support Vector Model) classification. Finally a behavioral oriented expert ranking is done by uncovering expert extraction and SVM classification which constitute the topic experts in Twitter.

Keywords: *Twitter; Expert Finding; SSGR (Semi-Supervised Graph-Based Ranking) SVM (Support Vector Model) Behavioral Oriented Expert Ranking.*

1. Introduction

Twitter [1] is one of the online news and social networking services in which the users can post and interact with messages called "tweets". Those users who are registered can post tweets, but those who are not registered can only read the tweets. The users can reach Twitter via its website, SMS or mobile device app. The head offices of Twitter Inc. are situated at San Francisco, California , States. More than 25 offices reside across the world. In March 2006 Jack Dorsey, Noah Glass, Biz Stone, and Williams developed Twitter. It was accessible in July, by which the service in no time achieved worldwide popularity. By 2013, it became one of the ten top viewed websites and yet it was defined as "SMS of the Internet". By its signature bird logo, it has become internationally identifiable.

Expert finding [2] is the phenomenon of searching people having expertise, which is relevant, or experience on a certain topic. The problem of finding experts has become an interesting topic with the emergence of social networks like Twitter. The Twitter is a rich source of topic experts and it shows a way to follow appropriate and trustful information on a certain topic. In some of the processes like mining of opinions and Name Entity Recognition (NER) [3], identifying topic experts is a preprocessing step. For instance, opinions drawn from beautician's tweets would favor a manufacturer of cosmetics than users in common.

Some of the prior methods like Page Rank [4] and clustering methods [5] are used to identify impact of users for various topics which is based on follower relations. Recent studies show that to identify experts for certain topics with the help of the data in the form of Meta of Lists for Twitter is more beneficial. A user to accumulate her followings based on a criteria e.g. owning experience for "data mining" forms the list for Twitter. In a list, the meta-data (e.g. title) can be explained in theform of annotations of users in that list. For example, in a list if a user is titled "machine learning" [6] then that user will have experience on machine learning.

Roadmap. The remaining of the work is put forward in the following way. The proposed system is presented in Section2.Finally, Section3 concludes this paper.

2. Proposed system

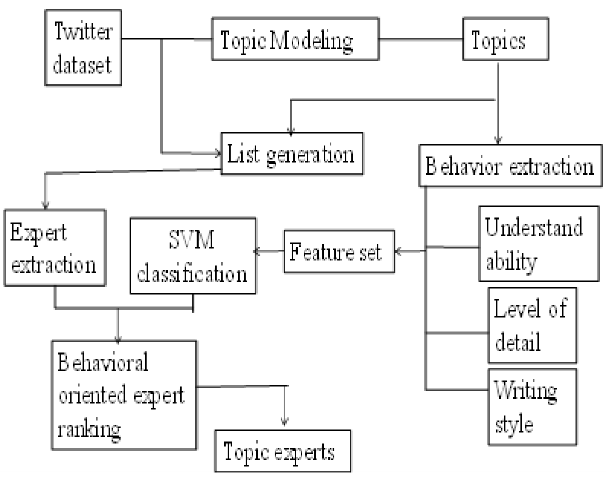

Fig. 1: Architecture of the Proposed System.

In the proposed system [7] a test input in the form of topic is given as a query to the system. Then a topic modeling process is applied on it which retrieves relevant topics from twitter database. Based on the behavior analysis over the retrieved topics there arises different relations lists like follower list, user list and list-list. The follower list gives the various followers for a particular topic. The user list gives details of similar users over a set of similar topics.

The list-list provides similar topics handled by various users. Then a Semi-Supervised Graph based Ranking (SSGR) algorithm is used to perform ranking process. After the ranking procedure we obtain ranking scores for various users for a particular topic. The top N users who have high ranking scores are selected as the authoritative users of a particular topic. Thus expert extraction is performed.

This paper also attempts to extract behavioral aspects of the topics. The Natural Language Processing (NLP) tool is used to perform text processing such as identifying Parts Of Speech (POS) and tokenization .Some of the tags for POS are nouns, adjectives, preposition,conjunctions,verbs,adverbs and pronouns.

The tweets can be checked semantically with respect to the various operational features of the behavioral aspects like understandability, level of details, writing style and cognition indicators. The understandability is the degree at which a tweet is clear and easy to understand. It can be analyzed from familiar words, characters per word, words, long words and words per sentence. The level of details is the degree at which a tweet is rich in information. It can be analyzed from nouns, adjectives, prepositions, articles, conjunctions, verbs, adverbs, pronouns, visual words, aural words, feeling words, spatial words, temporal words and function words. The writing style is the way by which certain kinds of words are used to express opinions in a tweet comfortably. It can be analyzed from emotiveness, positive emotion words, negative emotion words, past tense, present tense, future tense, firm words, upper case characters, exclamation points, question marks and all punctuations. The cognition indicators are the linguistic information that could be absorbed as a result of ignorance in expressing tweet. It can be analyzed from discrepancy, fillers, tentative words, causal words, insight words, motion words and exclusion words. These form the feature set for SVM (Support Vector Model) classification. This process involves clustering, classification and feature selection. Now based on the expert extraction and supervised SVM learning, a behavioral oriented expert mining occurs. This expert ranking procedure gives ranking scores for various experts based on the behavioral or semantic aspects of the tweeters. Finally those users with the highest ranking scores constitute the topic experts in twitter.

3. Evaluation and results

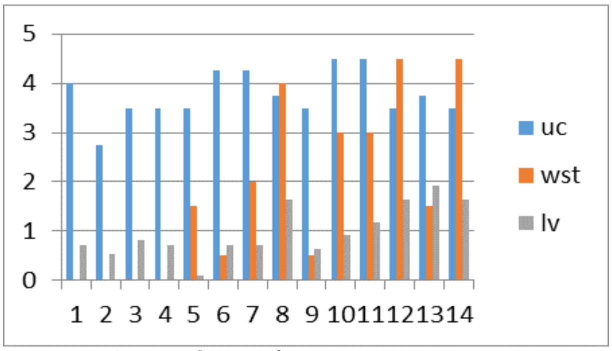

Fig. 2: Evaluation of Using Linguistic Features.

The concept of SVM based author finding is implemented with an experimental setup consisting of an Intel based processor with 4 GB primary memory. The tweets collected from twitter dataset were pre-processed and fed to the experiment. The libSVM3.0 is used for the classification process. The classic concept was first tested with list based algorithms (list-list and user-list interpretations). Then the pre-processed reviews were undergone to the Linguistic process. Stanford NL tools were used to extract the linguistic properties of the tweets. The properties categorized as Level of Details, Understandability etc are measured from each review to form a numerical dataset of linguistic properties. The criteria for class detection are formalized by going through the training set. The SVM classification has formulated a new set of classes, with in two ranges (0-1) indicating the good or bad tweets. The total number of good reviews and the corresponding topic experts were traced out. A comparison of experts detected through SVM based classification to the original method is performed.

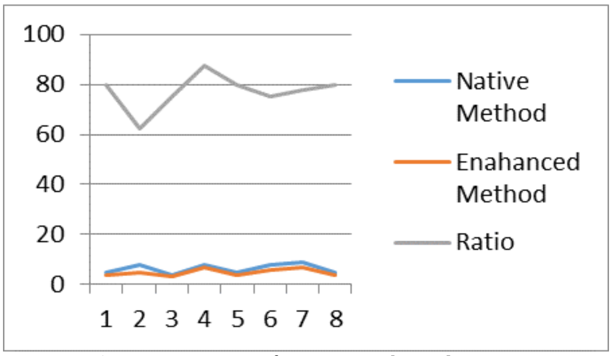

Fig. 3: Comparison of Native & Enhanced System.

The Ground Truth to evaluate the quality of the expert search of two methods was cross checked with different number of datasets. Then the aggregated list of the class -1 (linguistically good) users returned as the output of experiment. The outputs revealed the number of authors was considerably streamlined by minimizing the first generated list. This indicates the linguistic qualities have best impact on tweets which will decide a user as the topic authority. Based on the computed SVM scores, we select the top-N relevant users for any given topic.

4. Conclusion

The methods of recommendation used by the previous papers were based on a Semi-Supervised Graph-based Ranking method and relevance which may be a local one that can exist among users and a particular query. The main aim of the proposed system is to explore an enhanced expert finding using behavioral aspects. Firstly, it focus on the behavior extraction of the topics based on features like understandability, level of detail and writing style .Secondly, it focus on a Support Vector Model (SVM) which is a classification model of users with respect to the feature set generated from the behavioral analysis. Thirdly, it focuses on a Behavioral Oriented Expert Ranking procedure to achieve accurate and reliable recommendations for generation of topic experts in Twitter. The work employed in this paper is to develop a recommendation system to increase the accuracy of previous recommendation by incorporating content classification, to modify the expert ranking process, and to overcome current problems in the process of finding topic experts in Twitter with the help of different relations. The author detection acts as a decision support system. The information given by twitter helps users to depend on it to infer relevant information on a particular topic. From the evaluation and results it was found that the proposed approach outperforms well.

Acknowledgement

I am indebted to Prof. Anil A.R; Head of the Department, Computer Science & Engineering who guided me in the research process. I want to acknowledge the contributions of my guide Prof. Gopu Darsan, Assistant Professor in the department of Computer Science & Engineering. His co-operations and patience as I formed the paper work has to be sincerely appreciated. He has helped me a lot to materialize this seminar. I am very much obliged to our seminar coordinator, Prof. Minu Lalitha Madhavu, Assistant Professor in the department of Computer Science & Engineering who was instrumental in familiarizing me with the technologies.

References

[1] V. Qazvinian, E. Rosengren, D.-R. Radev, and Q.-Z. Mei, "Rumor has it: Identifying misinformation in micro-blogs," in Proc. Conf. Empirical Methods Natural Language Process.2011, pp. 1589–1599.

[2] L. Chen, Z.-Y. Liu, and M.-S. Sun, "Expert finding for micro-blog misinformation identification," in Proc. Int. Conf. Comput. Linguistics, 2012, pp. 703-712.

[3] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential Twitterers," in Proc. ACM Int. Conf. Web Search Data Mining, 2010, pp. 261–270.

[4] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in Proc. ACM Int. Conf. Web Search Data Mining, 2011, pp. 45–54.

[5] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: Crowdsourcing search for topic experts in microblogs," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2012, pp. 575–590.

[6] X. Liu, S. Zhang, F. Wei, and M. Zhou, "Recognizing named entities in tweets," in Proc. 49th Annu. Meet. Assoc. Comput. Linguistics:Human Language Technol., 2011, pp. 359-367.

[7] Wei Wei,Gao Cong, Chunyan Miao, Feida Zhu and Guohui Li, "Learning to find Topic Experts in Twitter via Different Relations", IEEE Transactions on Knowledge and Data Engineering,Vol.28,No.7,JULY 2016.

[8] X. Meng, F. Wei, X. Liu, M. Zhou, S. Li, and H. C. Wang, "Entity- centric topic-oriented opinion summarization in twitter," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 379-387.

[9] P. Desislava and W.-B. Croft, "Proximity-based document representation for named entity retrieval," in *Proc. 16th ACM Conf. Inf. knowl. Manag,* 2007, pp. 731-740.

[10] K. Balog, L. Azzopardi, and M. De Rijke, "Formal models for expert finding in enterprise corpora," in *Proc. 29th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2006, pp. 43–50.

[11] H. Deng, I. King, and M.-R. Lyu, "Formal models for expert finding on DBLP bibliography data," in *Proc. Int. Conf. Data Mining*, 2008, pp. 163–172.

[12] N. Craswell, A. P. de Vries, and I. Soborof, "Overview of the TREC 2005 enterprise track," in *Proc. Text Retrieval Conf.*, 2005, pp. 199–205.

[13] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: Crowdsourcing search for topic experts in microblogs," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2012, pp. 575–590.

[14] G. Demartini, D. E. Difallah, and P. Cudré-Mauroux, "Zencrowd: Leveraging probabilistic reasoning and crowd sourcing techniques for large-scale entity linking," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 469–478.

[15] X. Liu, W. B. Croft, and M. Koll, "Finding experts in community-based question-answering services," in *Proc. ACM Conf. Inf. Knowl. Manag.* 2005, pp. 315–316.

[16] A. Pal and J. A. Konstan, "Expert identification in community question answering: Exploring question selection bias," in *Proc. ACM Conf. Inf. Knowl. Manag*, 2010, pp. 1505–1508.

[17] Z. Zhao, L.-J. Zhang, X.-F. He, and W. Ng, "Expert finding for question answering via graph regularized matrix completion," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 4, pp. 993–1004, Apr. 2015.

[18] R. Yeniterzi and J. Callan, "Analyzing bias in CQA-based expert finding test sets," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2014, pp. 967–970.

[19] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitter rank: Finding topic-sensitive influential Twitterers," in *Proc. ACM Int. Conf. Web Search Data Mining*, 2010, pp. 261–270.

[20] N. Agarwal, H. Liu, L. Tang, and P.-S. Yu, "Identifying the influential bloggers in a community," in *Proc. ACM Int. Conf. Web Search Data Mining*, 2008, pp. 207–218.

[21] W. Wei, B. GAO, T.-Y. Liu, T.-F. Wang, H.-G. Li and H. Li. "A ranking approach on large-scale graph with multidimensional heterogeneous information," *IEEE Trans. Cybern.*, vol. pp, no. 99, pp. 1– 15, Apr. 2015.

[22] B. GAO, T.-Y. Liu, W. Wei, T.-F. Wang, and H. Li, "Semi-supervised ranking on very large graphs with rich metadata," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 96–104.

[23] C. L. Clarke, G. V. Cormack, and E. A. Tudhope, "Relevance ranking for one to three term queries," *Inform. Process. Manage* vol. 36, no. 2, pp. 291–311, 2000.

[24] L. Page, S. Brin, R. Motwani, and T. Winograd, "The page rank citation ranking: Bringing order to the web," *Stanford Digit. Libr. Tech- nol. Project*, Stanford, CA, USA, Tech. Rep. 1999-66, Nov. 1999.

[25] Y. Fang, S. Luo, and O. Etzioni, "Discriminative models of integrating document evidence and document-candidate associations for expert search," in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2010, pp. 683-690.