International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 10 | Oct -2017 www.irjet.net p-ISSN: 2395-0072

Semantic expert mining in Social Networks-A Survey

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

[1] *M.tech Scholar, Sree Buddha College of Engineering, Alappuzha, India.*
[2] *Asst. Professor, Sree Buddha College of Engineering, Alappuzha, India*

Abstract - *With the emergence of social networks such as micro-blogging services like Twitter, the expert finding has become an interesting topic. However, previous methods cannot be directly used to learn about topic experts in Twitter. Some of the new methods employ the relations among users and Twitter lists for expert finding. A probabilistic method has been developed to explore the relations (i.e. follower relation, user-list relation and list-list relation) for finding experts. To calculate the global authority of users, a Semi-Supervised Graph-based Ranking (SSGR) approach is used. Then a local relevance between users and given query is computed. By understanding the global authority and local relevance of users, all users are ranked and top-N users with highest ranking scores are retrieved which constitute the topic experts in Twitter.*

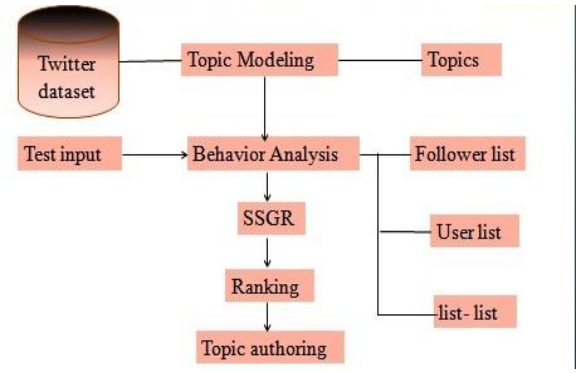
***Key Words*: Twitter, Expert finding, SSGR (Semi-Supervised Graph-based Ranking), Global authority, Local relevance**

1. INTRODUCTION

Twitter is an online news and social networking service where users post and interact with messages, "tweets," restricted to 140 characters. Registered users can post tweets, but those who are unregistered can only read them. Users access Twitter through its website interface, SMS or a mobile device app. Twitter Inc. is based in San Francisco, California , United States , and has more than 25 offices around the world. Twitter was created in March 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams and launched in July, whereby the service rapidly gained worldwide popularity. In 2013, it was one of the ten most-visited websites and has been described as "the SMS of the Internet. Twitter has become internationally identifiable by its signature bird logo.

Expert finding is the process of finding people with relevant expertise or experience on a given topic query. The expert finding problem has become an interesting topic with the emergence of social networks such as micro-blogging services like Twitter. The Twitter is a rich source of topic experts and they paved a way to follow relevant and trustworthy information on a specific topic. For many applications like opinion mining and Name Entity Recognition (NER), identifying topic experts is a preprocessing step. For instance, opinions drawn from beautician's tweets would favor a cosmetic manufacturer than those from common users.

Some of the traditional methods like Page Rank-based method and clustering-based method used follower relations to identify general influence of users on different topics. A recent study proposes to identify topic-specific experts with the help of the Meta –data of Twitter Lists. A Twitter list is created by a user to group her followings based on a criteria e.g. having expertise on "data mining". The meta-data (e.g. title) of a list can be seen as the annotations of users in that list. For instance, a user in a list named "machine learning" will have expertise on machine learning.



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 .Hence we explore the topic experts in twitter.

Road map. The remaining of the paper is organized as follows. In Section 2, we review the literature survey. Finally, Section3 concludes this paper.

2. Literature Survey

Expert finding methods have an assumption that individuals published documents are relevant with respect to their expertise on the knowledge on that particular topic. So they

focus on modeling the associations between documents and candidate experts. This task has a great influence on the information retrieval community.

2.1 Discriminative models of integrating document evidence and document candidate associations for expert search

Generative models are used as a statistical language model to find the relationship between the experts and their expertise in the supporting documents. On the other hand, discriminative models show little attention in expert search research. This paper [1] proposes relevance based discriminative learning in which it automatically integrates the document evidence and document candidate associations into a single model. Discriminative models, also called conditional models, are a class of models used in machine learning for modeling the dependence of unobserved (target) variables y on observed variables x . Within a probabilistic framework, this is done by modeling the conditional probability distribution P(y|x) which can be used for predicting y from x.

The success of the work depends on the collecting expertise evidence from the supporting documents. Let d be the supporting document which acts as a bridge to connect expert e and query q. For a document d it has to be verified that e and q are relevant. This depends on two factors: document evidence and document candidate associations. This method has two steps: 1) Check whether the document d is relevant to the query q; 2) Check whether the expert e is relevant to the document d. The final relevance decision for (e,q) is made by averaging over all the documents.

2.2 Expert finding for micro-blog misinformation identification

L.Chen et.al proposed Expert finding for micro-blog misinformation identification [2] which involves integration of collective and machine intelligence. According to the micro-blog contents, the users are indexed automatically. Then a matching process occurs among users and suspected misinformation. To know the credibility of misinformation, it is sent to respective experts to judge their assessment. A tag based method is used to index experts of micro-blog users with social tags. There are two classes of misinformation: Domain Knowledge Constrained (DKC) and Time Space Constrained (TSC).The former talks about domain specific topics while the later is related to some events that occur in some places and time.

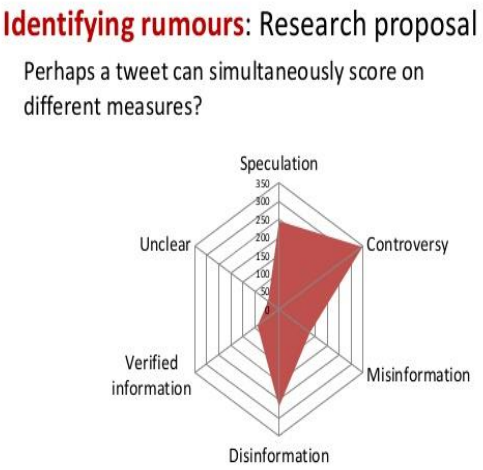


Fig-1: Identifying Rumours

Let E denote all micro-blog candidate experts. Then priori probability of expert, e given a misinformation ,m [Pr(e|m)]is directly proportional to priori probability of m given e [Pr(m|e)] and priori probability of e[Pr(e)].Here Pr(m) indicating misinformation remains same for all users and hence removed from ranking. Therefore only Pr (e) is taken into account.

2.3 Finding experts in community based question answering services

In the paper, finding experts in community based question answering services, [3] there have been a growing number of web information services. One such service called Wondir is a free publicly available live question and answer engine that connects people with questions to people with answers. This type of service are called community based question answering (QA) services. Here the expert finding problem is treated as an IR (Information Retrieval) problem. Given a question, an expert is a person who has answered similar questions in the past in the system. The language models used in this work are: query likelihood, relevance and cluster based language model. Depending on the text used, expert profiles can be built from all previously answered questions by a user, all previously answered questions-question texts only, one of the previously answered questions both questions and answer text and one of the previously answered questions-questions text only.

2.4 Identifying topical authorities in micro-blogs

A.Pal and S.Counts proposed Identifying topical authorities in micro-blogs [4] which proposed a set of features for authors that includes nodal and topical metrics.A probabilistic clustering and within clustering procedures gives a final list of top authors for a particular topic.The algorithm used was found more flexible in the real world scenarios.

The tweets are classified into three: Original Tweet (OT), Conversational Tweet (CT) & Repeated Tweet (RT).The original tweets are those tweets produced by authors. The CT is directed at another user. The RT is produced by someone else but user copies or forwards it. Around a user some metrics like number of original tweets; number of links shared etc are computed. A self similarity score is determined which gives the measure of how many words a user borrows from the previous tweets. Some of the textual features extracted for a user include topical signal (TS) and signal strength (SS).The TS gives a measure of the involvement of user in a topic. The SS gives the strength of TS i.e. true authority of a user on a topic. A Gaussian mixture model is used to cluster users. This clustering aims at reducing the target cluster which consists of authoritative users.

2.5 Cognos: Crowd sourcing search for topic experts in micro-blogs

The paper, Cognos: Crowd sourcing search for topic experts in micro-blogs [5] make use of twitter lists which are created by individual users that includes experts and their topics interested by them. These metadata provides information regarding experts and their domain of expertise. The list information is mined to build a system called cognos to find topic experts in twitter.

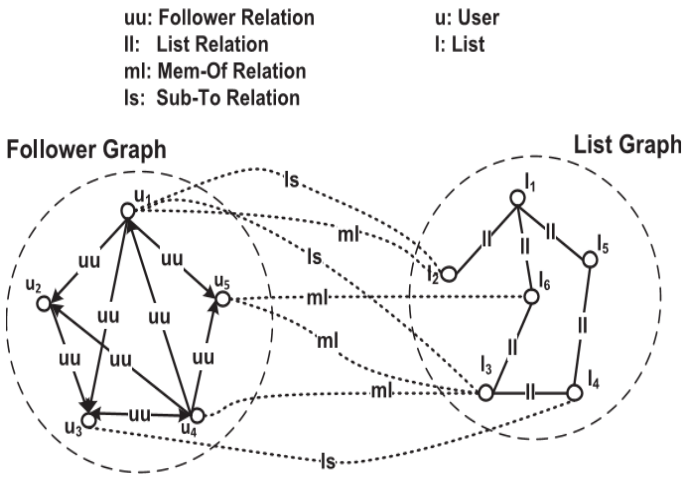


Fig-2: Follower graph and List graph

The twitter list can be seen in the form of a list graph and it can be connected to a follower graph via member of relation and subject to relation. The twitter list can be observed in the form of a table which gives information like list name, description and members. The list name gives the relevant topic; description gives details of topics and members gives the name of experts in the relevant topic.

Since cognos act as a list feature it is indeed a who-to-follow system in twitter .Here ranking procedure is based on list feature .It was found that the performance of cognos was better compared to the conventional methods. The crowd sourced search helps to build future content search. One disadvantage with list based methodology is list spamming where malicious users create fake lists.

2.6 Formal models for expert finding in enterprise corpa

This paper [6] provides two strategies for expert searching for a given document collection: 1) Directly model an expert's knowledge based on the documents that are associated with it.2)Locate documents on topic and then finds the associated expert. By addressing the problem of identifying expertise within an organization has led to the development of a class of search engines known as expert finders. The Hyperlink-induced topic search (HITS) algorithm is used to identify the authorities.

The probability of a query given a candidate can be viewed as the following process: 1) Let a candidate, ca be given 2) Select a document d associated with ca 3) From this document, generate the query with probability p(q|d).By taking the sum of over all documents the probability of the candidate expert ca for a given query

3. CONCLUSIONS

The work employed in this paper is to explore topic experts in Twitter via different relations. The author detection act as a decision support system. The information given by twitter helps users to depend on it to infer relevant information on a particular topic. The proposed approach integrated different user related information and performed a ranking procedure to retrieve topic experts. A semi supervised graph based ranking algorithm was used to find the authoritative users of a particular topic. The ranking scores were based on the supervised information from the Twitter crowds. From the ranking scores the top N users are selected as the topic experts for a particular topic. From the survey 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 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 project coordinator, Minu Lalitha Madhavu, Assistant Professor in the department of Computer Science & Engineering who was instrumental in familiarizing me with the technologies.

REFERENCES

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

[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] 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.

[4] A. Pal and S. Counts, "Identifying topical authorities in micro-blogs," 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 micro-blogs," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2012, pp. 575–590.

[6] 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.

[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.