



SSGR: Topic Specific Expert Finding in Twitter

Pooja Mallasure¹, Rashmi R. Rachh²

PG Scholar, Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, India¹

Faculty, Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, India²

Abstract: With proliferation of social networking sites, large number of tweets, posts and messages are available which can be analysed to extract some useful information. Identifying topic experts from these posts is a daunting task. Collecting and storing various types twitter data in terms of real world dataset is difficult and analyzing of data is too tedious. Earlier approaches for expert finding partially utilized relations among user and twitter list. In this project, a semi-supervised graph based ranking (SSGR) solution has been implemented to compute the global authority of users in the offline mode. The local relevance between users and the given query is computed in the online mode. By taking the advantage of the local relevance and global authority of users, all the users are ranked and the top users with highest ranking are selected.

Keywords: twitter, expert search, micro-blogging, list, and graph based ranking.

I. INTRODUCTION

The data mining is process of extracting some meaningful information from a data set and transforms it into an explicit structure for further use [1]. Mining is used in multiple regions. The number of commercial data missing system available today and yet these are many challenges in this field [5], [6]. Data mining is generally used in intrusion detection, biological data analysis, scientific applications, social networking [1], [2]. There are numerous classes of social media along with, however no longer constrained to, social networking [3] like, facebook or LinkedIn, microblogging (Twitter), photo sharing (Photobucket, or Picasa, Flickr), information aggregation (StumbleUpon, or Feedburner, Google reader), sharing video (MetaCafe, YouTube), livecasting (Ustream or Twitch.tv), digital worlds (Kaneva), social gaming (global of Warcraft), social search (Google), and immediate messaging (Skype, or Yahoo!, Google speak). In the field of mining useful information from the messages and tweets and posts on social networking sites is presented.

Twitter has gained large popularity since the first day that it became launched. It has also drawn increasing interests from research network [1]. There proceeding work to study the topological and geographical homes of the social network shaped by the twitterers and their followers we're interested by figuring out the influential twitterers. The benefit of fixing this trouble is multifold. First, it doubtlessly brings order to the real-time web in that it lets in the quest consequences to be sorted by means of the authority have an impact on of the contributing twitterers giving a timely update of the mind of influential twitterers [3]. Second, according to twitter is likewise an advertising platform. Focused on those strong users will grow the efficiency of the advertising and marketing campaign. For instance, a handphone manufacturer can have interaction those twitterers influential in subjects about IT gadgets to doubtlessly influence more humans [1], [2]. There are also packages that utilize Twitter to accumulate opinions and records on specific subjects. Identifying influential twitterers for thrilling topics can improve the excellent of the evaluations collected. Mining in social networking is an important data mining task with wide applications. Expert finding problem has gained increasing attention in social media, such as micro-blogging services [4] like Twitter, social media is providing a messages can be no longer than 140-character short messages (i.e., tweets) [3],[10]. Twitter has gained huge amount of popularity and collected a tremendous amount of tweets in recent years. These tweets are cover extremely open and multiple topics, such as everyday action or skill, top news, automation, and infinite of other highly functional areas, etc. thus, users in Twitter have rich expertise on various topics and finding these topic specific experts covers a way to enable others to recall or follow the related and accurate information on a specific topic in micro-blogging services [1], [2]. It builds a weighted graph by considering both the topic similarity between two users and followers graph, and then employ page Rank algorithm to find topic specific influential users. Extracts user's features from follower graph and users posted tweets, and then employ a Gaussian-based mixture model to cluster users for ranking. The Twitter Rank and pal's work only consider user-user single relation [5]. This provides to utilize Twitter List to analyze the attributes of Twitter users. [2] In their subsequent work, they develop a system named Cognos to infer the topical expertise of users by utilizing only user-list relation in Twitter List, which captures the wisdom Twitter crowds. Cognos represents each user by the meta-data of Twitter lists this contain the user, and then employs a similarity measure to compute the similarity score between each user and topical question, which is used to rank



**International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering**

ISO 3297:2007 Certified

Vol. 5, Issue 8, August 2017

customers for search. This method jointly use three types of relations (follower relation, user list relation and list-list relation) for find topic experts. A Semi-Supervised Graph-based Ranking approach (SSGR) to offline calculates the global authority of users. In this SSGR, [3], [6] employ a normalized Laplacian regularization term to jointly explore the three relations, which is subject to the supervised information derived from Twitter crowds. Then local relevance between users and the given query this is for online calculation. By taking the advantage of the global authority and local relevance of users, rank all of users and find top-N users with highest ranking scores. There are different types of relations among two types of information to target Twitter expert finding problem, specifically: i) Local Relevance: the similarity between users produced tweets and the given query, and ii) Global Authority: the global expertise results of users in a twitter on a given topic. The implemented approach successfully integrate different types of user-related information (that is the crowd sourced Lists information, follower graph and users profiles) into a unified ranking framework for accurately inferring the topical expertise of users.

II. RELATED WORK

J.weng et al. [1] have proposed an approach called TwitterRank, which works in two steps. Finding topic it propose starts towards twitter rank. It works in two steps; first one is employs Latent Dirichlet Allocation (LDA) model. It notices the topics of independent based on their tweets. Second one for each topic it builds a graph weighted by taking both the topical similarity in between two users and follower graph, then also enrol page rank algorithm for find topic specific influential users.

N. k.sharma [2] has focused on “inferring who is who in the twitter social network” twitter list: they propose to use twitter list to identify the quality of twitter users by twitter crowd. List contain the users and to compute similarity between each user and given topic query. This is used to search and rank all the users. Using cognos move to choose the users that users contained in more number of lists those Meta data contain the query. It use twitter list to identify the quality of twitter users.

L. Chen et al. [4] have focused on “expert finding for microblog misinformation identification” which expert finding for misinformation identification. The process of increasing in size of social media it provides a convenient communication scheme for people, at the same time cradle of misinformation. Spreading the misinformation over social media is harmful to public interest. So they design a framework, which intelligence and machine intelligence, it helpful for identify misinformation. The basic point is (1) list the expert users according to their microblog satisfied. (2) It matches the experts with specified presence misinformation by sending the truth misinformation to suitable experts. They collect the analysis of expert and to decide the quality of information, and it helpful for prove the misinformation have propose a tag based method to list the experts of microblog users with the social tags. And match guess misinformation it is based on a real world dataset indicate.

A.pal et al. [5] have targeted on “identifying topic authorities in microblog” proposed methods and features as it may be used to produce top authors ranked list by given topic for finding topical authorities in microblogging environment. They also showed that probabilistic clustering it is a way to filter a large chunk of outliers in the feature space. At last they allow that Gaussian based ranking it is helpful to rank users and more effective way for finding top1 ranked authors.

B.Gao et al. [6] have worked in “Semi-Supervised ranking on very large graphs with rich metadata” planned general framework and conjointly economical rule for ranking graph. The semi supervised learning framework for ranking of nodes for a really giant graph and acquire within our framework is named semi supervised page rank.

III. EXISTING SYSTEM

Existing systems find experts in social networks based on the influential users identified from different social networks. These methods do not consider topical dimensions. Existing approaches use either follower relation or user-list relation alone, and therefore they are not suitable for finding topic experts in twitter.

IV. PROPOSED SYSTEM

This project aims at constructing a model for finding topic specific expert in order to resolve the issues identified in the existing systems such as using only one relation or influential users for expert finding. Topic specific expert finds the rank of a set of candidate experts based on the relevance of their expertise on the topic query. Semi-supervised method will be used to calculate the global authority of users in offline mode and also to compute the local relevance between users and the given query in online mode.



V. SYSTEM DESIGN AND IMPLEMENTATION

We present architecture (as shown in fig. 2) of our model to addressing the topic specific expert finding problem. In particular, it consists of two components, that is, an offline graph based ranking algorithm called SSGR. To learn the global authority of each user and an online ranking model to select the top experts on the given query.

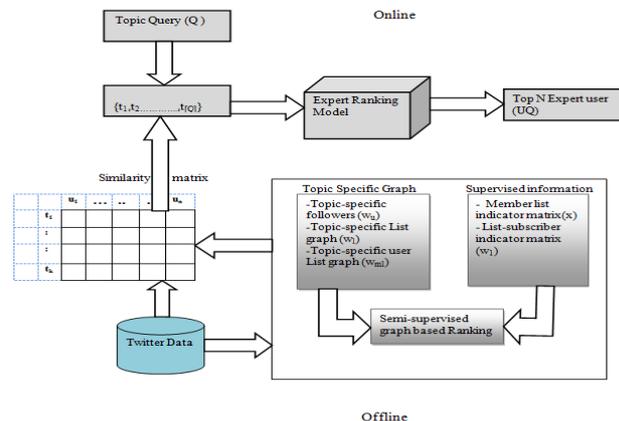


Fig. 1: A Semi Supervised Graph Based Ranking (SSGR)

A. OFFLINE GRAPH-BASED RANKING MODEL

First construct an authority matrix R over the twitter corpus. Specifically, each row R_i belongs to R is offline computed by semi supervised graph-based ranking for each term t in twitter, in which jointly exploit the three different relations of users and twitter lists for inferring the global authority of each candidate on t in twitter.

- Semi-supervised Graph-based Ranking:-

To learn the global authority of candidate users on a single term topic query (Q_t), present a semi-supervised graph-based ranking method, called SSGR [3]. It is capable of exploiting the different relations (i.e. follower relation, user-list relation and list-list relation) among users and lists to mutually reinforce the ranking of users and lists for inferring the global expertise scores of users on Q_t .

- Topic Specific Graph

There are three types of topic-specific graphs related to a given topic (Q_t), which are:

W_u :- a n -by- n symmetric topic-specific follower graph, in which denotes the similarity between user and her follower for a given topic. Each entry in W_u only considers the symmetric relation between two users.

W_l :- m -by- m symmetric topic-specific list graph, which is generated based on the mutual k -nearest neighbor graph.

W_{ml} :- a n -by- m topic specific user-list graph, in which denotes the similarity between user and list containing user for a given topic. Each entry in W_{ml} refers to MEM-OF relation, i.e. user is included in list.

- Supervised Information

In supervised information use two types of relations as the supervised information, namely, MEM-OF relation and SUB-TO relation. Let n -by- m indicator matrix (i.e. X) encodes the MEM-OF relations for supervising the ranking of users and, m -by- n indicator matrix (i.e. Y) encode the SUB-TO relations for supervising the ranking of lists.

B. ONLINE RANKING MODEL

For a given topic query = $\{t_1, \dots, t_{|Q|}\}$, use an online ranking model, based on the corresponding rows in R for terms contained in Q , to select top- N users as expert on Q , by taking into account the global authority and local relevance of candidates.

METHODOLOGY

1) Methods of Data Collection:

Users: The twitter data consists of user profiles, followers, tweets, user-list membership information, and user-list subscribe information. In particular, a user-centric strategy to collect data as a brute-force crawling of all users for all lists would be prohibitively expensive and would not scale. In addition, the dataset contained a mixture of different languages, e.g. Chinese, English, German, Italian and etc. In this filtered the non-english characters, stopwords, punctuations as well as high frequency words in twitter.



**International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering**

ISO 3297:2007 Certified

Vol. 5, Issue 8, August 2017

After processing, the users without any context information are removed.

Queries: In this use 28 sample queries for evaluation, whose topics are from general to specific.

- <https://dev.twitter.com/>.
- <http://wiki.engr.ellinois.edu/dislay/forward/Dataset-UDI-TwitterCrawl-Aug2012>
- http://twitter.com/who_to_follow/siggestions.

2) Methods of Data Analysis

A semi-supervised graph-based ranking method used for computing the global authority of a user on the given topic. The regularization term in the semi-supervised graph-based ranking method used to smooth the ranking scores on the graph. By using the approach of the semi-supervised graph-based ranking method improves the effectiveness of finding the topic-specific experts. Ranking method can effectively exploit the three different types of relations among users and lists (i.e. follower relation, user-list relation, list-list relation). The results demonstrate the effectiveness and superiority of proposed method as compared to the sate-of-art method Cognos and Twitterrank.

VI. EXPERIMENTAL RESULTS

Before finding topic expert in twitter, stores data from the twitter. The admin finds top topic expert by giving keyword admin prepare similarity matrix, search experts, search topic in tweets, and also user registered list. Fig.2 shows top experts and Fig.3 shows that Search topic in tweets.



Fig.2 Top experts

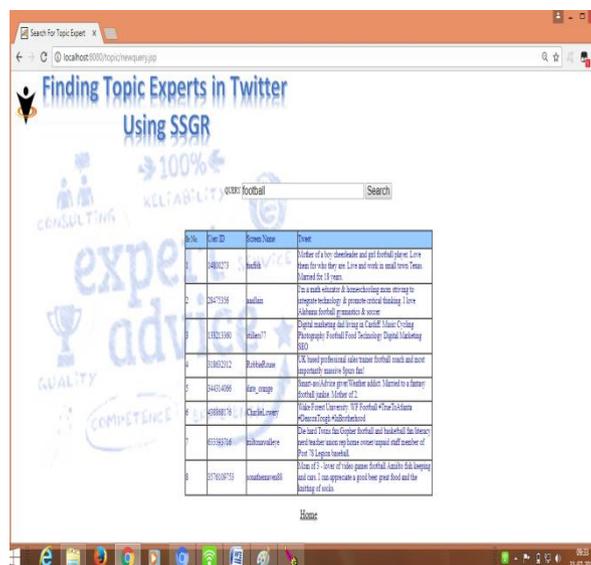


Fig.3 Search topic in tweets



**International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering**

ISO 3297:2007 Certified

Vol. 5, Issue 8, August 2017

VII. CONCLUSION AND FUTURE WORK

In this project, semi-supervised graph based ranking method has been implemented to find the rank of a set of experts based on the relevance of their expertise to the given query. In this, the global authority of the users is calculated offline and in the online mode, local relevance between user and the given query is calculated. This helps the users in finding expertise with accurate results. Follower relations, list-list relation and user-list relations are used in topic specific expert finding method to efficiently identify the topic experts.

The method implemented in this project to find topic expert in twitter can also be extended to other social networking sites. The implementation can be further improved to increase the efficiency in terms of time taken. The views of the topic expert differs from one person to another, this problem needs to be addressed in the upcoming future.

REFERENCES

- [1] Q. He, J.weng, E.-P Lim, and J. Jiang, "Twitterrank: Finding topic-sensitive influential Twitterers," in Proc. ACM Int. Conf. Web Search Data Mining, 2010
- [2] N.K. Sharma "inferring who-is-who in the twitter social network 2012
- [3] Wei Wei, Gao Cong, Chunyan Miao, Feida Zhu, and Guohui Li "Finding the Topic Experts in Twitter using SSGR," IEEE transactions JULY 2016.
- [4] Z.-Y. Liu, L.Chen,"Expert finding for microblog misinformation identification," in Proc. Int. Conf. Comput. Linguistics, 2012
- [5] A. Pal and S. Counts, "Identifying topical authorities in microblog," in Proc. ACM Int. Conf. Web Search Data Mining, 2011
- [6] B.Gao,T.-Y. Liu, W. Wei, 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
- [7] C. S.Campbell, P.-P. Maglio, A. Cozzi, and B. Dom,"Expertise identification using email communications," in Proc. ACM Conf. Inf. Knowl. Manag, 2003
- [8] E. Rosengren, V. Qazvinian, and Q.-Z. Mei, "Rumor has it: Identifying misinformation in microblog," in Proc. Conf. Empirical Methods Natural Language Process., 2011
- [9] J. Lehmann, C. Castillo, M. Lalmas, and E. Zuckerman, "Finding news curators in Twitter," in Proc. Int. Conf. World Wide Web, 2013
- [10] J. Li, J.Zhang, and J.Tang, "Expert finding in a social network. Advances in databases: Concepts, systems and applications," in Proc. 12th Int. Conf. Database Syst. Adv. Appl., 2007
- [11] M. J. Welch, U. Schonfeld, D. He, and J. Cho, "Topical semantics of twitter links," in Proc. ACM Int. Conf. Web Search Data Mining, 2011