



AN AUTOMATIC ANSWER RETRIEVING SYSTEM FOR RECURRENT QUESTIONS IN SOCIAL Q&A USING UNSUPERVISED TECHNIQUES

Nisha Bai M¹, Afifa Salsabil Fathima A², Hafsa Fathima³, G. Hemavathi⁴, S.Neha Kouser⁵
^{1,2,3,4,5}Department of Computer Science and Engineering, Dr TTIT, KGF, Kolar Karnataka, India

Abstract

Question and Answering (Q&A) systems have become major part in today's world for gaining and sharing knowledge and information. In general Q&A systems users post questions to get answers and pick the questions to answer in the system. As the user population is growing enormously, these systems are getting flooded with more number of questions. And some disinterested users will post irrelevant answers to posted questions. Hence we are proposing to develop a system which will improve the performance of Q&A systems by actively forwarding questions to experts who are capable and willing to answer the questions. Also, automatically retrieve's answers for recurrent questions. Thus, this system reduces the chances of user getting fake answer and provides satisfactory answers for the posted questions.

Index Terms: Question and answer systems, Social networks, Information search.

I. INTRODUCTION

The internet is an important source through which an individual can easily get information, where the amount of data is vast and its a valuable tool for communication and to gain information. Users rely on search engines such as Google, yahoo, Bing that allows internet users to search for content through world wide web(WWW). As research indicated , search engines performs well indexing with web pages and provides the related content information to the users who rely on search engines. To address the particular information, many Question and answering system such as Yahoo!, Baidu, Quora, Stack Exchange systems have been developed. Question and Answering system is fairly a

information retrieval system in which a query is stated to the system and it relocates the correct or closest answer to the specific question. Q&A system is used by a large number of internet users, most of the search engines rely on Q&A system. This Q&A system stores the questions that are recently asked, thus it acts as a repository for information retrieval.

As the user population increases in social Q&A system, a large no of questions are posted every day. When a user wants to answer a question, he may be overwhelmed by the excess of questions. However depending on disinterested users to provide answers cannot encourage the users to answers quickly To illustrate approximate answer providers, current Q&A systems allows users to choose tags for their questions. Current Q&A system does not overcome the requirement of providing high quality answers in a given period of time in which the users wanted the correct answers rapidly. This lead to propose an advanced Q&A system that reduces the number of unanswered questions, and also boost the answer quality and reduces the waiting time.

It's a common human tendency to get similar kind of doubts so when the users want to know the answer for their questions they tend to take help of these Q&A systems, therefore instead of allowing the users to post same kind of questions we can simply suggest the answers of already answered questions when similar kind of questions are typed. Which reduces the waiting time of user to get an answer.

Also, here only an admin authorized expert can answer any of the question posted by the registered user in the system thus reducing the chances of users getting fake answers and low quality answers which directly means that the users gets answers which are of high quality

which are answered by the particular expert of the field.

II. RELATED WORKS

The importance of the growth of the social Question and answering systems demands wanting to understand these systems and to improve the quality of them [8]. The works in [9-14] studied the influence of difference factors (e.g., message prediction) in the social networks on Q&A performance. The existing social network based on the asker- answerer relationship in current Q&A systems [11] is different from online social network based on the social relationship, which is used in social question answering systems. The works in [21-24] concentrated on locating experts and authoritative users. Certain studies have been conducted to create reputation models in Q&A systems [18] to increase the quality of answers, and to determine the relationship between the reputation of the users and the quality of their provided answers. Some researches such from [19-21] categorizes questions into predefined categories, making it easier for users to locate previously asked questions and for experts to find questions they can answer. Quan et al.

[20] proposed three new supervised term weighting schemes for question categorization, and evaluated each of the scheme using a trace from Yahoo! Answers. Song et al. [21] proposed a sequential process including topic-wise word identification and weighting, semantic mapping, and similarity calculation. Text mining techniques that also have been used to give better answers [8, 22-26]. These categorization and text mining methods can be used in social question answering systems to more accurately derive user interests and question interests. Li et al. [8] proposed a language model by integrating expertise estimation and availability estimation, and later proposed category-sensitive language models [22] for expert identification, which helps route questions to available and capable experts. Zhou et al. [23] classified the questions using a variety of local and global features of questions and users' relationship in order to route a classified question to its potential answerers. Cao et al. [24] leveraged question category to enhance question retrieval in community based Q&A systems. Guo et al. [25] proposed a topic-based model to identify appropriate answerers by calculating the

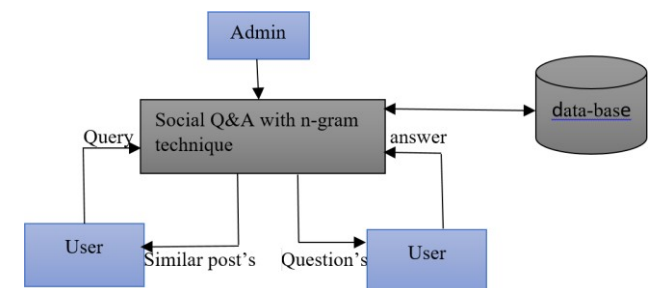
similarities between questions' topics and user's specialists.

Nie et al. [26] proposed a scheme which can add notes to social questions automatically to unravels the incomplete and biased problems of question tags. Compared to previous Q&A system works, social question and answering also determines the experts who are willing and capable of answering the questions posted by the users and retrieve answers for the recurrent questions in the social network properties to improve the quality of service performance (QoS). It incorporates different algorithms to determine the weightage of the questions that were previously asked and provide answers for the recurrent questions. Unlike previous social question and answering system works, it does not allow any users to answer the question to avoid the chances of existence of fake answers in the system hence increasing the quality of the answers provided. That is the system only allows admin authorized users to answer the question posted by the registered users. Hence the social question and answering system only concentrates on the improvement of quality of the answer and retrieval of answers for recurrent questions using weightage schemes.

III. SYSTEM ARCHITECTURE

An Automatic Answer retrieving System for Recurrent Questions in Social Q&A:

Social Q&A's are basically used to gain knowledge and to share information. Where users' ask questions and expect an answer for their queries. Here we are allowing only the experts to post answer hence, increasing the quality of answers. A real life social network is formed by treating user and expert as nodes and linking two nodes with a social relationship to



carry out interactions.

Proposed System Design

Fig1: The System Architecture

Figure 1 shows the high-level architecture of our system and the interaction between the core components: Admin, User and Expert

Here, Admin will first login into the system and allow user registration's and can add experts, later users can login into system using the credentials, and can post queries in the system to obtain an answer. Whereas Experts will login into the system using the credentials and can post answers for the questions asked by the users. Whenever users are typing the questions which are already been answered in the past by experts answer for such questions will be automatically retrieved and will be displayed to users. Hence, reducing the waiting time of users to get answer. Database is used to store all the data from the interactions which are taking place.

Admin

Admin should first login into the system to carry out his functions using his credentials, once logged in he can be able to allow user registrations and can be able to add experts who are capable of answering the queries, and he can also change his password anytime.

Expert

Experts are the individuals who have experience, capability and expertise to provide answers, experts are added by admin. Experts should login into the system to carry out their functions, once logged in they can post answer for the questions posted by experts, they can view the previously posted answers and they can also change their password.

User

Firstly, users should register themselves in the system once their registration is allowed by the admin of the system they can be able to login into the system and they can post their queries to obtain answers they can edit their posts they can view their previously posted questions and can change his password anytime.

IV. LEARNING APPROACH

n-gram Technique for Answer retrieval for recurrent questions:

Most of the question in Social question and answering system are recurrent, for example a question asked by some user is answered the same question can be asked by another user then answer for that question can be filtered from the recurrent questions in Q&A system. Therefore we can save the users effort and system

resources to answer recurrent questions by storing the recurrent questions in the repository. To store and retrieve the recurrent answers we introduce bloom filter method. Bloom filter is a technique for applications where the amount of source data may require an implicitly large amount of memory. Since the recurrent questions are may not be exactly same, the success rate to find the former similar questions may not be high if we directly

feed the whole question into the bloom filter.

To solve this problem we introduce n-gram technique. The n-gram technique is contiguous n words of a question. n-gram model is mainly used for automated text prediction by using the entered text to predict the next word. Tokenization is applied in n-gram to cut the text into pieces from the body text into sentences and words, n-gram will process the files in pieces of 1,000 lines. The n-gram tables are processed and split into (n-1) grams and the final word corresponding to predictor and predicted word. It is uses a simple back off approach in combination with weighting to build a list of probable next words.

From the input text, the last three terms are obtained and searching in the 4-gram table. If one or more matches are found, then the algorithm outputs the best predictions for the next word. If no match is found in the 4-gram table, then the search continues in the 3-gram table using the last two words from the input. And so on. If no match is found, the prediction is then the most common one-gram (single terms) Each user periodically broadcasts his/her bloom filter results B_q for his/her questions with satisfying answers. In the broadcasting, each bloom filter is propagated through the social links. Whenever a user U_k asks a question, before U_k launches question Q_i 's forwarding process, U_k first looks over all bloom filter results received. For each bloom filter result, all n-grams of the newly asked question Q_i are checked in this bloom filter, and the owner of each bloom filter result is scored by the number of successfully found n-grams of Q_i .

The asker then selects the top N users with the highest scores, and sends them the recurrent question searching request for Q_i . When a user, say U_i , receives the request, it then finds the question with the largest number of common n-grams with Q_i , and forwards the question

associated with the satisfying answers back to the asker.

The Tf-idf Technique

TF and IDF (TF) and TF and IDF is an information retrieval technique that weighs a terms frequency and it's a inverse document frequency (IDF). TF and IDF is used by search engines to better understand the content which is undervalued. TF and IDF is used to weigh a keyword in any content and assign the importance to that keyword based on the number of times it appears in the document.

For a term t in a document d , the weight $W_{t,d}$ of term t in document d is given by:

$$W_{t,d} = T_{Ft,d} \log(N/D_{Ft})$$

Where:

- $T_{Ft,d}$ is the number of occurrences of t in document d .
- D_{Ft} is the number of documents containing the term t .
- N is the total number of documents in the corpus.

The onion routing based answer forwarding can also be applied in this answer retrieval process for recurrent questions.

V. CONCLUSION

The Question and Answering systems are used by many people for purpose such as information retrieval, academic assistance, and discussion. To increase the quality of answers received and decrease the wait time for the users to get an answer for a recurrent question. In order to increase the quality of the answers, the prototype developed allows only experts to post the answers. Thus, a high quality answer is obtained. It also forwards the answers to recurrent questions without expert's involvement. The prototype is developed is for a small community of user and involvement of particular experts for the respective community further the system can be enhanced for large community of users and experts where quality answers can be obtained for large number of questions by large number of experts involved.

REFERENCES

[1] M. R. Morris, J. Teevan, and K. Panovich. A Comparison of Information Seeking Using Search Engines and Social Networks. In Proc. of ICWSM, 2010.

- [2] M. R. Morris, J. Teevan, and K. Panovich. What do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. In Proc. of CHI, 2010.
- [3] B. Li and I. King. Routing Questions to Appropriate Answerers in Community Question Answering Services. In Proc. of CIKM, 2010.
- [4] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In Proc. of WWW, 2008.
- [5] G. Drosatos, P. Efraimidis, A. Arampatzis, G. Stamatelatos, and I. Athanasiadis. Pythia: A privacy-enhanced personalized contextual suggestion system for tourism. In COMPSAC, 2015.
- [6] A. Mtibaa, M. May, C. Diot, and M. Ammar. Peoplerank: Social Opportunistic Forwarding. In Proc. of Infocom, 2010.
- [7] H. Shen, Z. Li, G. Liu, and J. Li. Sos: A distributed mobile q&a system based on social networks. TPDS, 2014.
- [8] M. L. Radford, C. Shah, L. Mon, and R. Gazan. Stepping Stones to Synergy: Social Q&A and Virtual Reference. Proceedings of the American Society for Information Science and Technology, 2011.
- [9] M. Richardson and R. White. Supporting Synchronous Social Q&A Throughout the Question Lifecycle. In Proc. of WWW, 2011.
- [10] R. W. White, M. Richardson, and Y. Liu. Effects of Community Size and Contact Rate in Synchronous Social Q&A. In Proc. of SIGCHI, 2011.
- [11] J. Teevan, M. R. Morris, and K. Panovich. Factors Affecting Response Quantity, Quality, and Speed for Questions Asked via Social Network Status Messages. In Proc. of ICWSM, 2011.
- [17] Z. Li and H. Shen. Collective Intelligence in the Online Social Network of Yahoo!Answers and Its Implications. In Proc. of CIKM, 2012.
- [12] J. Bian, Y. Yang, and T. Chua. Predicting trending messages and diffusion participants in microblogging network. In Proc. of SIGIR, 2014.
- [13] X. Geng, H. Zhang, Z. Song, Y. Yang, H. Luan, and T. Chua. One of a kind: User profiling by social curation. In Proc. of Multimedia, 2014.
- [14] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise Networks in Online Communities: Structure and Algorithms. In Proc. of WWW, 2017.

- [15]J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha. Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement. In Proc. of WWW, 2009.
- [16]P. Jurczyk and E. Agichtein. Discovering Authorities in Question Answer Communities by Using Link Analysis. In Proc. of CIKM, 2007.
- [17]M. Bouguessa, B. Dumoulin, and S. Wang. Identifying Authoritative Actors in Question-Answering Forums: the Case of Yahoo!Answers. In Proc. of KDD, 2008.
- [18]W. Chen, Q. Zeng, W. Liu, and T. Hao. A User Reputation Model for a User-Interactive Question Answering System: Research Articles. Concurrency and Computation: Practice and Experience, 2007.
- [19]A. Shtok, G. Dror, Y. Maarek, and I. Szpektor. Learning From the Past: Answering New Questions With Past Answers. In Proc. Of WWW, 2012.
- [20]X. Quan, W. Liu, and B. Qiu. Term Weighting Schemes for Question Categorization. TPAMI, 2011.
- [21]W. Song, W. Liu, N. Gu, X. Quan, and T. Hao. Automatic Categorization of Questions for User-Interactive Question Answering. Information Processing and Management, 2011.
- [22]B. Li, I. King, and M. R. Lyu. Question Routing in Community Question Answering: Putting Category in its Place. In Proc. of CIKM, 2011.
- [23]T. C. Zhou, M. R. Lyu, and I. King. A Classification-Based Approach to Question Routing in Community Question Answering. In Proc. of WWW (Companion Volume), 2012.
- [24]X. Cao, G. Cong, B. Cui, C. S. Jensen, and C. Zhang. The Use of Categorization Information in Language Models for Question Retrieval. In Proc. of CIKM, 2009.
- [25]J. Guo, S. Xu, S. Bao, and Y. Yu. Tapping on the Potential of Q&A Community by Recommending Answer Providers. In Proc. of CIKM, 2008.
- [26]L. Nie, Y. Zhao, X. Wang, J. Shen, and T. Chua. Learning to recommend descriptive tags for questions in social forums. TOIS, 2014.
- [27]D. Horowitz and S. D. Kamvar. The Anatomy of a Large-Scale Social Search Engine. In Proc. of WWW, 2010.
- [28]H. Song, S. Dharmapurikar, J. Turner, and J. Lockwood. Fast hash table lookup using extended bloom filter: an aid to network processing. In Proc. of SIGCOMM, 2005.
- [29]A. Shtok, G. Dror, Y. Maarek, and I. Szpektor. Learning from the past: Answering new questions with past answers. In Proc. Of WWW, 2012.
- [30]H. Shen, G. Liu, and N. Vithlani. Socialq&a: An online social network based question and answer system. In Proc. of ICCCN, 2015.