

Implementation of Designing Community Aware Ranking Algorithms for Expert Recommendation in Question Answer Forums

Shraddha A. Borkar¹, Dr. Sudhir W. Mohod², A. D. Gotmare³

¹Department of Computer Science & Engineering, BDCE, Sevagram Wardha, Maharashtra, India

²Professor & HOD Department of Computer Science & Engineering, Amravati, Maharashtra, India

³Assistant Professor, Department of Computer Science & Engineering, BDCE, Sevagram, Wardha, Maharashtra, India

Abstract

A huge growth in the number of online question answering (QA) systems has occurred as a result of the vast amount of information that is freely available on the World Wide Web. The community-based question answering (QA) systems are a subset of quality assurance systems that has witnessed tremendous growth in recent years. Our system, described in this paper, is proactive enough to provide replies to inquiries while also providing word definitions, with the goal of eliminating the time lag that occurs as a result of askers needing to wait for answers to a query from a variety of users. The usefulness of the proposed technique is demonstrated by the evaluation results on the expected response from the computing-related datasets that were used in the evaluation.

Keywords: Question-Answer forums, Expert identification; overlapping community detection; Ranking algorithms

1. Introduction

Question and Answer (Q&A) systems, which facilitate the exchange of information and knowledge, have a substantial impact on our quality of life. Users can ask questions and then choose which questions to answer within the system, according to their preferences. Because of the rapidly increasing number of users and, consequently, the variety of questions, it is quite rare that a user will come across a question by coincidence that he will be able to answer. Altruism does not encourage all users to provide responses, much alone high-quality answers with a short wait period for a response.

Every day, a big number of queries are posed online to a Q&A system that caters to a large number of people. In one day, for example, there are 823,966 questions and answers posted to the Yahoo! Answers website. Once this happens, the user may find himself or herself overwhelmed by the sheer number of questions that are being asked. Furthermore, depending just on altruistic individuals to supply answers will not incentivize all users to submit answers and to respond to inquiries as rapidly as possible. Current Q&A systems allow users to tag their queries (i.e., categorise them according to their interests) in order to find the most appropriate answer providers. The proper tag(s) for a question such as "How is the computer organisation class at our university?" may, on the other hand, be difficult to determine.

Consequently, conventional Q&A systems may fall short of the criterion of giving high-quality responses in a short amount of time, despite the fact that users need to obtain appropriate answers as quickly as possible. This is supported by the findings of the study in. It was discovered that only 17.6 percent of queries on Yahoo! Answers were

adequately addressed; for the remaining 82.4 percent, one-fifth of the inquiries remained unanswered, according to the study. A total of 22.7 percent of inquiries were satisfactorily answered by Baidu Zhidao, while 42.8 percent of the unsolved issues received no response at all. In order to address this growing requirement, an innovative Q&A system that can reduce the number of unanswered questions, improve the quality of answers, and shorten response times is being developed.

Furthermore, the anonymity of the Q&A system is becoming increasingly vital nowadays. Many users may ask or answer questions about sensitive topics such as health problems, political activism, or even sexual orientation, which may be considered offensive by others. Despite the fact that the user may require a response as soon as possible, he or she still requires privacy protection in order to avoid the potential revelation of personal information. Due to the fact that Social Q&A is founded on social networks. Both the questioner and the answerer are socially close to one another. As a result, protecting personal information is both necessary and difficult.

Social Q&A is an online social network-based question-and-answer system that actively passes inquiries to those users who are most likely (in terms of capacity and willingness) to answer them in a knowledgeable and interesting manner about their respective topics. The design of Social Q&A is based on the characteristics of social networks.

2. Literature Survey

In order to better understand and create Q&A forums systems, it is necessary to make an effort to better understand and develop these systems. A variety of characteristics (such as task users, system interactions, and social size) were investigated in the context of Q&A operations during the activities. It is important to note that the existing social network, which is based on responding relationships to present Q&A forum systems, is distinct from the online social network, which is based on social media and is employed in the Social Q&A forum.

The majority of the activities are geared toward identifying qualified experts and users. Instead, the Social Q&A forum seeks to identify general users who are capable of answering questions of any nature, including opinion-based inquiries. Other research have been carried out in order to develop respectable models for use in Q&A systems in order to boost the reliability of responses, as well as to evaluate the association amongst users' reputation and also the quality of the answers that they supply.

Social Q and A effectively use social networking sites social media technologies to encourage users to submit feedback, rather than depending on a separate reputation mechanism to accomplish this. Some studies categorise questions into subcategories that have already been completed, making it easier for consumers to find answers that have been already asked and for specialists to find problems to solve in a more efficient manner. Three surveillance strategies for the observation of interrogative terms were devised, and each programme was tested using Yahoo! Answers, as described in the paper.

Text mining techniques have also been employed to improve the quality of the replies. It is possible to utilize categorization and text mining methods in Social Q&A to derive more accurate information about users' and questions' interests. The authors proposed a language model that combined expertise estimation with availability estimation, and they later proposed category-sensitive language models for expert identification, which aides in routing questions to available and capable experts.

The questions were classified by Zhu et al. based on a variety of local and global characteristics of the questions as well as user relationships in order to route a classified question to its potential answerers.

In order to improve the retrieval of questions in community-based Q&A systems, Aso et al. created a new question category. Guo and colleagues proposed a tie-based model to identify appropriate answers by calculating the similarities between questioning ties and users specialists. Social networks were also investigated by researchers as part of their work in the field of search engines. Identifying searching as a group activity, Evans et al. proved that social interactions can aid in the improvement of search results.

Instead of utilising web search engines, Morris and colleagues explored the growing tendency of submitting searches as social network statuses rather than using web search engines. However, sending a flood of questions to all or some of a user's pals may cause the buddies to become overburdened, resulting in dissatisfaction.

Horowitz and Kamvar introduced a social computer simulation that locates the most appropriate person to satisfy a user's information requirement while also providing trust-based social intimacy for that user. The difference between this study and earlier Q&A system efforts is that this study concentrates on software architecture by utilising social network aspects and demonstrates its potential for performance improvement. SOS is also a Q&A system that is accompanied by a social networking site.

SOS, on the other hand, is concerned with the implementation of a mobile Q&A system in an extremely distributed manner, while also employing knowledge engineering methodologies. Furthermore, it is presumptively assumed that users have already provided social closeness. As an alternative, Social Q&A focuses on the benefits of leveraging social network capabilities in better identifying possible answerers with established interest categories and demonstrating the benefits of this approach through an examination of real users' Q&A activities.

The primary objective for this project is to shorten the wait time for answers by mapping the questions to the asker's close friends.

- Because of the close personal interactions and their expertise/interest in the issues, close friends are more likely to respond promptly.
- Friends should refrain from immediately communicating such sensitive information with one another in order to maintain a certain level of privacy.
- Rather than exchanging their buddy lists and interest vectors, they should trade Bloom Filter-based Personal Information (BFPI) derived from those lists and vectors.
- Text mining techniques have also been employed to improve the quality of the replies.
- The onion routing-based answer forwarding method is used by the asker to search for answers utilising the previously mentioned method of searching for replies.
- Unlike traditional systems, these systems are designed in a distributed fashion, meaning that each node chooses its friends to transmit and pass inquiries autonomously.

3. Proposed Work

Q&A systems play an important role in our daily lives for information and knowledge sharing, but in the past they were not as accurate or satisfied to users as they should have been, and the waiting time was longer. To overcome this problem, we are enhancing the structure by optimizing the productivity of Q&A systems by actively transmitting inquiries to users who are ready and willing to answer the questions, and the waiting time is shorter. In order to do this, we have added and the solution Social Q&A, an online Q&A system based on social networks. When used in conjunction with social network qualities such as common interest and mutual trust friend relationships, social Q&A can be used to select an asker through friendship who is most available to respond the question, as well as to increase the user's security. We also make improvements to Social Q&A in terms of security and efficiency by safeguarding user privacy and identifying information, as well as obtaining responses automatically for frequently asked queries. We detail the structure and algorithms, and we ran a complete large-scale simulator to evaluate Social Q&A in comparison to alternative methods. Download the full paper here. Our findings imply that social networks can be used to increase the quality of answers and the amount of time an asker has to wait. In addition, we developed a real-world Social Q&A prototype and conducted an analysis of the Q&A behaviour of real-world users and queries from a small-scale genuine Social Q&A system. The design of Social Question & Answer is composed of three components:

- The design of security and efficiency enhancement methods.
- Comparative trace-driven experiments, the development of a real-world Online QA.
- The analysis of the data from real Online QA

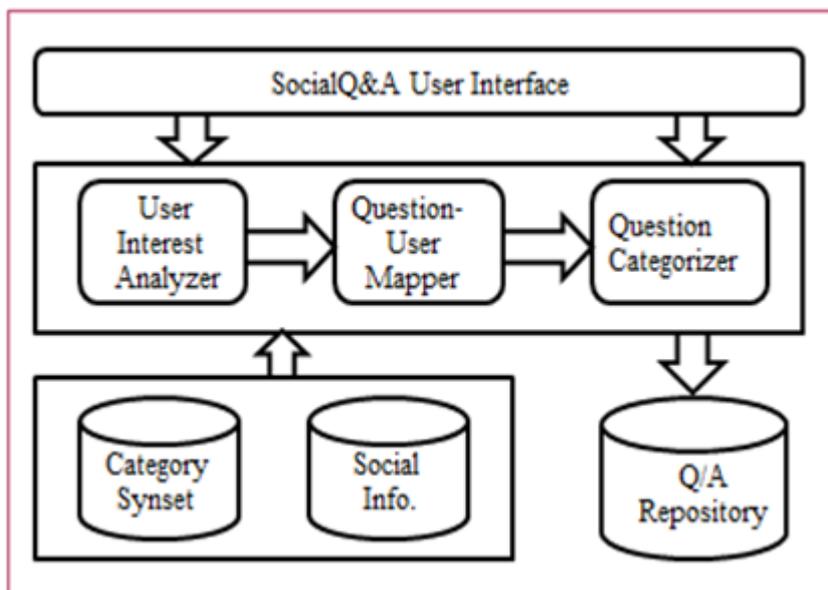


Fig.1 Block Diagram

4. Algorithm Implementation

1. TF_IDF

The TF-IDF value increases proportionally to the number of times a word appears in the document, but is often offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general. Nowadays, TF-IDF is one of the most popular term-weighting schemes.

For example:

Document 1: The game of life is a game of everlasting learning

Document 2: The unexamined life is not worth living

Document 3: Never stop learning

Step 1: Term Frequency (TF):

Term Frequency also known as TF measures the number of times a term (word) occurs in a document. Given below are the terms and their frequency on each of the document.

Step 2: Inverse Document Frequency (IDF):

The main purpose of doing a search is to find out relevant documents matching the query. In the first step all terms are considered equally important. In fact certain terms that occur too frequently have little power in determining the relevance. We need a way to weigh down the effects of too frequently occurring terms. Also the terms that occur less in the document can be more relevant. We need a way to weigh up the effects of less frequently occurring terms.

Let us compute IDF for the term game:

$$Idf(t, D) = \log |D| + 1 - |\{d \in D: t \in d\}|$$

Step 3: TF * IDF

Remember we are trying to find out relevant documents for the query: life learning. For each term in the query multiply its normalized term frequency with its IDF on each document. In Document1 for the term “life” the normalized term frequency is 0.1 and its IDF is 1.405507153. Multiplying them together we get 0.140550715 (0.1 * 1.405507153). Given below is TF * IDF calculations for life and learning in all the documents.

Step 4: Vector Space Model –

Cosine Similarity From each document we derive a vector. If you need some refresher on vector refer here. The set of documents in a collection then is viewed as a set of vectors in a vector space. Each term will have its own axis. Using the formula given below we can find out the similarity between any two documents.

$$\cos(\theta) = \frac{v \cdot w}{\|v\| \|w\|} = \frac{\sum_{i=1}^n v_i w_i}{\sqrt{\sum_{i=1}^n v_i^2} \sqrt{\sum_{i=1}^n w_i^2}}$$

In, the calculation formula is used to find the similarity between the documents. Term Frequency also known as TF measures the number of times a term (word) occurs the document. The TF-IDF esteem builds relatively to the circumstances a word shows up in the record, however is regularly balanced by the recurrence of

the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general.

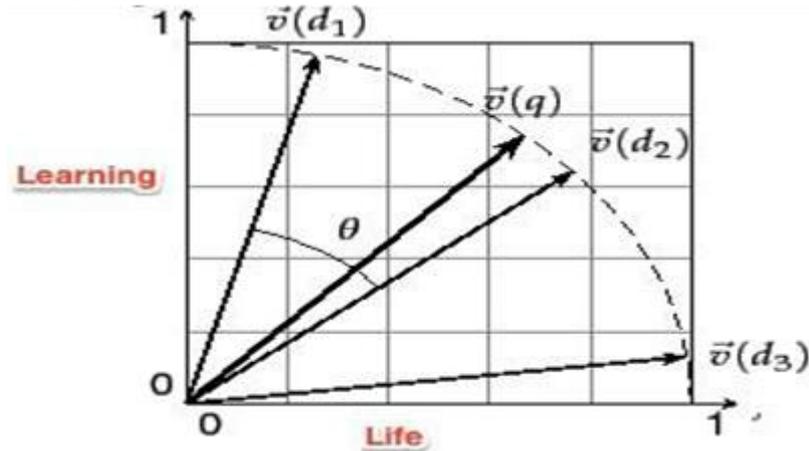


Fig. 2 Vector Space Model of Life and learning in documents.

In [Fig. 2], the vector space model has been implemented for calculations for life and learning in all the documents using the similarity formula. From each document we derive a vector. The set of documents in a collection then is viewed as a set of vectors in a vector space. Each term will have its own axis.

2. KNN algorithm:

Steps of the Algorithm are:

1. Determine parameter k = number of nearest neighbours.
2. Calculate the distance between the query instance and all the training samples.
3. Sort the distance and determine nearest neighbours based on the k^{th} minimum distance.
4. Gather the category y of the nearest neighbours.
5. Use simple majority of the category of nearest neighbours as the prediction value of the query instance

3. Design Goals

Our design goal can be summarized as follows:

1. To reduce the wait time of Answers, the questions are mapped to the asker's close friends, they tend to respond quickly due to the close social relationship and their expertise/interest on the questions.
2. To protect user privacy to a certain extent, friends should avoid exchanging such personal information directly. Instead, they should exchange the Bloom Filter based Personal Information of their friend lists and interest vectors.
3. Text mining techniques also have been used to provide better answers.
4. The onion routing based answer forwarding, the asker searches the answers using the previously introduced method.
5. These systems is implemented in a distributed manner, that is, each node selects its friends to send/forward questions autonomously.

5. Conclusion

The Question Answer Forum has altered the way in which the question and answer method used to function on social media. We have put in place a method that allows a large number of users to receive the answers they require while yet keeping total anonymity of their identities. In addition, when compared to other series of questions systems available, our system has received a high level of customer satisfaction. When all elements are taken into consideration, we have effectively built a comprehensive question and response system that outperforms its competitors.

References

- [1] Haiying Shen; Guoxin Liu; Haoyu Wang; Nikhil Vithlani "Online social network based question and answer system IEEE 2017".
- [2] H. Shen, Z. Li, G. Liu, and J. Li, "SOS: A distributed mobile Q&A system based on social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 4, pp. 1066–1077, Apr. 2014
- [3] Answers.com. Accessed: Feb. 2017. OY. Lin and H. Shen, "SmartQ: A question and answer system for supplying high-quality and trust worthy answers," in *Proc. IEEE Trans. Big Data*, Aug. 2017
- [4] H. Shen, Z. Li, G. Liu, and J. Li, "SOS: A distributed mobile Q&A system based on social networks" *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 4, pp. 1066–1077, Apr. 2014
- [5] G. Liu and H. Shen, "iASK: A distributed Q&A system incorporating social community and global collective intelligence," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 5, pp. 1–14, May 2016.
- [6] I. Ali, R. Y. Chang, J.-C. Chuang, C.-H. Hsu, and C. M. Yetis, "Optimal question answering routing in dynamic online social networks," in *Proc. IEEE VTC*, Sep. 2017, pp. 1–7.
- [7] L. Zhang, X.-Y. Li, J. Lei, J. Sun, and Y. Liu, "Mechanism design for finding experts using locally constructed social referral Web," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 8, pp. 2316–2326, Aug. 2017