2581-4575

International Journal For Recent Developments in Science & Technology

Crossref

A Peer Reviewed Research Journal

DESIGN AND DEVELOPMENT QUESTION AND ANSWER SYSTEM IN ONLINE SOCIAL NETWORK

[1]PAYASAM RAMYA, [2]CH.N.D.CHAMUNDESWARI

[1]Department of MCA, SKBR PG College, Amalapuram

[2]Lecturer, Department of MCA, SKBR PG College, Amalapuram

Abstract: Question and Answer (Q&A) systems play a vital role in our daily life for information and knowledge sharing. Users post questions and pick questions to answer in the system. Due to the rapidly growing user population and the number of questions, it is unlikely for a user to stumble upon a question by chance that (s)he can answer. Also, altruism does not encourage all users to provide answers, not to mention high quality answers with a short answer wait time. The primary objective of this paper is to improve the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To this end, we have designed and implemented SocialQ&A, an online social network based Q&A system. SocialQ&A leverages the social network properties of common-interest and mutual-trust friend relationship to identify an asker through friendship who are most likely to answer the question, and enhance the user security. We also improve SocialQ&A with security and efficiency enhancements by protecting user privacy and identifies, and retrieving answers automatically for recurrent questions. We describe the architecture and algorithms, and conducted comprehensive large-scale simulation to evaluate SocialQ&A in comparison with other methods. Our results suggest that social networks can be leveraged to improve the answer quality and asker's waiting time. We also implemented a real prototype of SocialQ&A, and analyze the Q&A behavior of real users and questions from a small-scale real-world SocialQ&A system.

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

I.INTRODUCTION

The Internet is an important source of information, where the amount of data is vast and constantly growing. Users rely on search engines to find specific information in this knowledge base. Search engines such as Google and Bing use keywords provided by the users to perform searches. Recently, industrial research and development activities, such as Microsoft and Facebook's social-featured Bing search endeavor, try to combine search engines and online social networks for higher search performance. As previous research has indicated [1, 2], search engines perform well in indexing web pages and providing users with relevant content to their search but are not suited for non-factual questions such as "Which is the best local auto shop?". To address this particular class of non-factual questions, many Question and Answer (Q&A) systems such as Yahoo! Answers, Baidu Zhidao, StackExchange, Quora and Ask have been developed. Since their inception, Q&A systems have proved to be a valuable resource for sharing expertise and consequently are used by a large number of Internet users. For example,

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


Yahoo! Answers was launched at the end of the year 2005 and attracted more than 10 million users in February of 2007 [3], and hit 200 million users in December of 2009 [4, 5]. Q&A systems also preserve all questions and answers, thus acting as a repository for information retrieval. They are not only important for sharing technical knowledge, but also as a source for receiving advice and satisfying one's curiosity about a wide variety of subjects [6].

With a vast population in a Q&A system, a large number of questions are posed online every day. For example, there are 823,966 questions and answers posed to Yahoo! Answers per day [4]. Then, when a user intends to answer a question, (s)he may be overwhelmed by the plethora of questions. Moreover, simply relying on altruistic users to provide answers cannot encourage all users to provide answers and to answer questions quickly. To locate appropriate answer providers, current Q&A systems allow users to choose tags (i.e., interest categories) for their questions. However, it may not be easy to determine the appropriate tag(s) for a question such as "how is the computer organization class at our university?".

As a result, current Q&A systems may not meet the requirement of providing high quality answer with a short answer wait time, though users wish to receive satisfactory answers quickly. This is confirmed by the study in [5]. It found that for Yahoo! Answers, only 17.6% of questions were answered satisfactorily; for the remaining 82.4%, one fifth of the questions remained unanswered. For Baidu Zhidao, 22.7% of questions were successfully answered, and 42.8% of the unresolved questions were not answered at all. Thus, there is an increasing need for an advanced Q&A system that can decrease the number of unanswered questions, enhance the answer quality and decrease the response time. In addition, the privacy of the Q&A system is very important nowadays. Many users may ask or answer questions related to sensitive topics such health problem, political activism or even sexual orientation [7].

Although the user may want the response as soon as possible, he/she still needs the privacy protection to avoid potential disclosure of personal information [8]. Since Social Q&A is built upon social networks. The asker and answerer are social close to each other. Therefore, protecting the privacy is important and challenge. To meet this need, we propose SocialQ&A, an online social network based Q&A system, that actively forwards questions to those users with the highest likelihood (capability and willingness) of answering them with expertise and interest in the questions' subjects. The design of SocialQ&A is based on two social network properties. First, social friends tend to share similar interests (e.g., lab members majoring in computer systems) [9]. Second, social friends tend to be trustworthy and altruistic due to the property of "friendship fosters cooperation" [10].

Accordingly, SocialQ&A favors routing queries among friends and identifies a question's potential answerers by considering two metrics: the interest of the friend towards the question and the social closeness of the friend to the asker/forwarder. Thus, the answer receivers have high probability of providing high-quality answers in a short time [11, 12]. Different from the existing Q&A systems, due

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


to the importance of users privacy, we future introduce security and efficiency enhancement to protect users privacy while users using social network answering questions. The contributions of this work are as follows:

• The design of SocialQ&A. SocialQ&A is composed of three components: User Interest Analyzer, Question Categorizer, and Question-User Mapper. User Interest Analyzer associates each user with a vector of interest categories. Question Categorizer associates a vector of interest categories to each question. Then, based on user interest and social closeness, QuestionUser Mapper identifies potential answerers for each question.

• The design of security and efficiency enhancement methods. SocialQ&A incorporates three methods to enhance its security and efficiency performance. The bloom filter based personal information exchange method protects users' privacy including friendship and interest information. The onion routing based answer forwarding method protects the identities of the asker and the answerer from being exposed. The answer retrieval for recurrent questions automatically finds the answers for recurrent questions.

• Comparative trace-driven experiments. We conducted comprehensive large-scale simulation to evaluate SocialQ&A in comparison with other methods. Our results suggest that SocialQ&A improves the quality of answers and reduces the wait time for answers.

• The development of a real-world SocialQ&A. We have prototyped the SocialQ&A system with user interfaces, and conducted a real-world small-scale test with real users from India, the United Kingdom, and the United States for a period of approximately one month.

• The analysis of the data from real SocialQ&A. We have analyzed the features of the questions posted, the questioning and answering activities of users, the quality of answers, and the wait time for answers. Analytical results show the benefits of SocialQ&A in enhancing answer quality and wait time.

II. RELATED WORK

The growing importance of Q&A systems demands an effort to better understand these systems and to improve them [13]. The works in [14–19] studied the influence of different factors (e.g., users' profiles, messages prediction, system interactions and community size) in the social networks on Q&A performance. These study results lay the foundation of SocialQ&A to leverage social network properties [20] in the design. Note that the existing social network based on the asker-answerer relationship in current Q&A systems [17] is different from online social network based on the social relationship, which is used in SocialQ&A. The works in [21–24] concentrated on locating experts and authoritative users. Instead, SocialQ&A aims to find normal users that can answer questions including opinion-type questions. Some studies have been conducted to create reputation models in Q&A systems [25, 26] to increase the credibility of answers, and to determine the relationship between the reputation of the users and the quality of their provided answers [27]. SocialQ&A directly utilizes the social network property of mutual-trust friendship to motivate users to provide answers without relying on an

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


additional reputation model. SocialQ&A shares similarity with other peer-assistant systems such as [28] in leveraging the collective power of peers for a certain goal. Some research [29–31] 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. [30] proposed three new supervised term weighting schemes for question categorization, and evaluated each scheme using a trace from Yahoo! Answers. Song et al. [31] proposed a sequential process including topic-wise word identification and weighting, semantic mapping, and similarity calculation. Text mining techniques also have been used to provide better answers [5, 32–36]. These categorization and text mining methods can be used in SocialQ&A to more accurately derive user interests and question interests. Li et al. [5] proposed a language model by combining expertise estimation and availability estimation, and later proposed category-sensitive language models [32] for expert identification, which helps route questions to available and capable experts.

Zhou et al. [33] 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. [34] leveraged question category to enhance question retrieval in communitybased Q&A systems. Guo et al. [35] proposed a topic-based model to identify appropriate answerers by calculating the similarities between questions' topics and users specialists. Nie et al. [36] proposed a scheme which can annotate social questions automatically to unravels the incomplete and biased problems of question tags. Compared to previous Q&A system works, SocialQ&A also leverages both the common-interest and mutual-trust social network properties to improve the QoS erformance.

It incorporates different algorithms to determine user interest, question interest and the question-user mapping. Unlike previous Q&A system works, it does not assume that friendship is always trustable and incorporates algorithms that avoid revealing personal information to others as little as possible. Different from previous Q&A system works, our previously proposed SOS [39] is also a Q&A system based on a social network. However, SOS focuses on realizing a mobile Q&A system in a distributed manner and using knowledge engineering techniques. Also, it assumes that social closeness is already provided by users. Instead, SocialQ&A focuses on how to leverage social network properties in better identifying potential answerers with predefined interest categories and showing its benefits through the analysis on real users' Q&A activities.

III. PROPOSED SYSTEM

3.1 The Rationale of SocialQ&A Design

A real-life social network is formed by regarding each person as a node and linking two nodes with a social relationship. This network is featured by social communities such as the football club and ECE department at a university. In real life, the people we rely on for answers to questions such as "how is the computer organization class at our university?" are usually those in our social communities. Persons in the same social community share common interests and trust each other on

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

answering questions on their common interests, and are willing to answer the questions from community members. An online social network connects friends with real-life relationship and online friendship, which shares similarity to the real-life social network. Friends in an online social network tend to share similar interests and trust each other [9, 40, 10]. Taking advantage of these properties, we design and develop SocialQ&A that incorporates an online social network to improve the quality of answers and decrease answer wait time. It forwards a user's questions to his/her social friends that have common interest and a close social relationship.

3.2 The Design of SocialQ&A

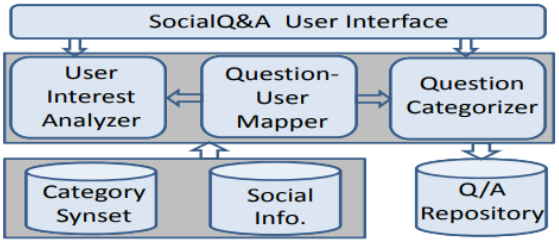

Fig. 1: The architecture of SocialQ&A.

Like all online social networks, the one in SocialQ&A has user profiles that record users' interests, education, hobbies and etc. Like Yahoo! Answers, SocialQ&A also predefines interest categories and subcategories. A total of 4 categories (music, movies, television, and books) and 32 subcategories (e.g., books: novel, drama) derived from Yahoo! Answers were used to implement SocialQ&A. We used these 4 categories as an example and will add more categories in our future work. Figure 1 shows the high-level architecture of SocialQ&A and the interaction between the core components: User Interest Analyzer, Question Categorizer, and Question-User Mapper. User Interest Analyzer analyzes data associated with each user in the social network to derive user interests. Question Categorizer categorizes the user questions into interest categories based on the Category Synsets, which stores the synonyms of all categories' keywords from WordNet [41]. Question-User Mapper connects these two components by identifying potential answerers who are most likely to be willing to and be able to provide satisfactory answers. The data from user questions and answers is stored on Q/A Repository to serve subsequent similar questions. Below, we present each component and user interface.

3.2.1 User Interest Analyzer

User Interest Analyzer utilizes each user's profile information in the social network and user interactions (answers provided and questions asked) to determine the interests of the user in the predefined interest categories. This is because if a user asks or answers questions in an interest category, (s)he is likely to be interested in this particular category.

3.2.2 Question Categorizer

The primary task of Question Categorizer is to categorize a question into predefined interest categories based on the topic(s) of the question. We also allow users to input selfdefined tags associate with questions, which are analyzed in question parsing.

3.2.3 Question-User

Mapper Question-User Mapper identifies the appropriate answerers for a given question. The potential answer providers are chosen from the

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


asker's friends in the online social network. Note that the changes in a user's friends in the online social network do not affect the performance of SocialQ&A as it always uses a user's current friends. To check the appropriateness of a friend (Uk) as an answer provider for a question, two parameters are considered:

i) the interest similarity between the interest vectors of the friend and the question and

ii) the social closeness between the friend and the asker. The former represents the potential capability of a friend to answer the question, and the latter represents the willingness of a friend to answer the question.

IV. SECURITY AND EFFICIENCY ENHANCEMENT

4.1 Secure Personal Information Exchange and Answer Forwarding

The friendship through online social networks may not be always trustable. It is important for users to reveal personal information to each other as little as possible. Besides, the askers and answerers for some questions, such as political sensitive questions, may want to be anonymous to the public. Therefore, a Q&A system should support secure question forwarding process through untrustable friendships. In the following, we propose bloom filter based personal information exchange method and onion routing based answer forwarding method to achieve a certain degree of security.

The bloom filter results are stored in an integer array of t entries. Each hash function encrypts the feed information into an integer m within [0, t], and the mth entry of the integer array is increased by 1. To search whether an information item is stored in a bloom filter, the information item is encrypted by each hash function of the bloom filter. If for each hashed result m, the value at mth entry in the array is larger than 0, this information item has a higher probability of being stored in the bloom filter; otherwise, it is not stored in the bloom filter.

In all possible user IDs or interests, a malicious user Ui can check the existence of each user ID or interest in the bloom filter result of his/her friend Uk in order to derive Uk's friends and interests. Note that a bloom filter is generated with a predefined false positive rate and an expected maximum number of feed inputs (i.e., interests and user IDs). The generated bloom filter has an actual false positive rate no larger than the predefined false positive rate if the number of actual feed inputs is no larger than the expected maximum number. Therefore, for the same number of feed inputs, in order to increase the false positive rate to protect the users' privacy, we can also reduce the expected maximum number of inputs. However, a larger false positive rate generates a higher probability of choosing some friends falsely regarded with many common interests and friends. Therefore, the answer quality is sacrificed to a certain extent but the personal information is better protected. In reality, the false positive rate needs to set according to the requirement of security and answer quality performance to break the tie.

4.2 Answer Retrieval for Recurrent Questions

A large amount of daily questions in a Q&A system usually are recurrent. For example, among 15% of English questions crawled from Yahoo! Answers, 25% questions are recurrent

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

[44]. Therefore, we can save users' efforts and system resources to answer recurrent questions by providing satisfying answers of the former same questions in repository. In order to release the workload of the centralized server to search recurrent questions, each asker stores the former questions and their associated answers, and users depend on nearby users in the social network for searching the recurrent questions. A straightforward way to search a recurrent question of a newly asked question is to broadcast the question to all friends of the asker or question forwarder if TTL>0. However, it generates high network traffic among social friends and friends-of-friends, and high workload for similar question searching in inquired users. Therefore, we introduce our bloom filter based similar question searching method. In this method, each user feeds his/her questions with satisfying answers into a bloom filter, denoted by Bq . Since the recurrent questions may not be exactly the same, the success rate to find the former similar question may not be high if we directly feed the whole new question into the bloom filter.

V. RESULTS

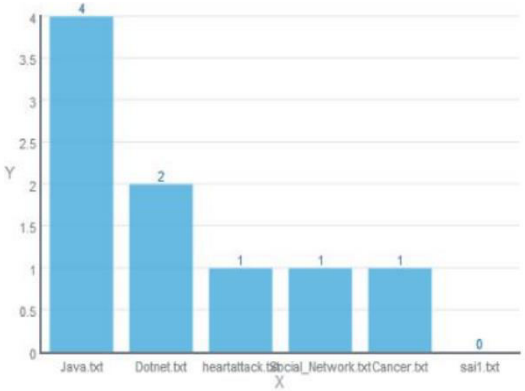

Fig. 2: Document Rank

VI. CONCLUSION

Q&A systems are used by many people for purposes such as information retrieval, academic assistance, and discussion. To increase the quality of answers received and decrease the wait time for answers, we have developed and prototyped an online social network based Q&A system, called SocialQ&A. It utilizes the properties of a social network to forward a question to potential answer providers, ensuring that a given question receives a high-quality answer in a short period of time. It removes the burden from answer providers by directly delivering them the questions they might be interested in, as opposed to requiring answer providers to search through a large collection of questions as in Yahoo! Answers or flooding a question to all of an asker's friends in an online social network. The bloom filter based enhancement methods encrypt the interest and friendship information exchanged between users to protect user privacy, and record all n-grams of answered questions to automatically retrieve answers for recurrent question. The onion routing based answer forwarding protects the identities of askers and answers. Our comprehensive tracedriven experiments and analysis results on the real-world Q&A activities from the SocialQ&A prototype show the promises of SocialQ&A to enhance answer quality and reduce answer wait time in current Q&A systems, and demonstrate the secure and efficiency improvement achieved by the enhancements. Since same questions may be presented very differently and the same question may be answered differently in different situation. In the future, we will cooperate with other techniques (e.g. topic

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal


modeling [48] and word embedding [49]) into SocialQ&A to find the redundant question with a large scale user set. Due to the dynamic of user behavior, SocialQ&A can cooperate a machine learning method to adjust three parameters appropriately, which needs a large user base and much more usage. We will conduct tests on a large user base in the real-world experiment.

VII. REFERENCES

[1] M. R. Morris, J. Teevan, and K. Panovich. A Comparison of Information Seeking Using Search Engines and Social Networks. In 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] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina. Questioning Yahoo! Answers. In Proc. of QAWeb, 2008.

[4] Yahoo!Answers Team. Yahoo! Answers BLOG. http://yahooanswers.tumblr.com/, [Accessed on 10/20/2014].

[5] B. Li and I. King. Routing Questions to Appropriate Answerers in Community Question Answering Services. In Proc. of CIKM, 2010.

[6] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In Proc. of WWW, 2008.

[7] G. Drosatos, P. Efraimidis, A. Arampatzis, G. Stamatelatos, and I. Athanasiadis. Pythia: A privacy-enhanced personalized contextual suggestion system for tourism. In COMPSAC, 2015.

[8] S. Li, Q. Jin, X. Jiang, and J. Park. Frontier and Future Development of Information Technology in Medicine and Education: ITME 2013. Springer Science & Business Media, 2013.

[9] A. Mtibaa, M. May, C. Diot, and M. Ammar. Peoplerank: Social Opportunistic Forwarding. In Proc. of Infocom, 2010.

[10] E. Pennisi. How Did Cooperative Behavior Evolve? Science, 2005. [

11] H. Shen, Z. Li, G. Liu, and J. Li. Sos: A distributed mobile q&a systembased on social networks. TPDS, 2014.

[12] A. Spagnolli and L. Gamberini. Interacting via sms: Practices of social closeness and reciprocation. British Journal of Social Psychology, 2007.

[13] 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.

[14] M. Richardson and R. White. Supporting Synchronous Social Q&A Throughout the Question Lifecycle. In Proc. of WWW, 2011.

[15] 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.