ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

Mobile Question and Answer System Based on Social Network

G.Manoranjitham[1], S.J.Veeraselvi[2]

PG Scholar, Computer Science and Engineering, Kalaignar Karunanidhi Institute of Technology, Coimbatore, India [1]

PG Scholar, Computer Science and Engineering, Kalaignar Karunanidhi Institute of Technology, Coimbatore, India [2]

Abstract: Question and Answer (Q&A) system based on social network gains more attention recently. The social-based Q&A systems answer non-factual questions, which cannot be easily resolved by web search engines. These systems either rely on a centralized server for identifying friends based on social information or broadcast a user's questions to all of its friends. Mobile Q&A systems, where mobile nodes access the Q&A systems through internet, are very promising considering the rapid increase of mobile users and the convenience of practical use. However, such systems cannot directly use the previous centralized methods or broadcasting methods, which generate high cost of mobile internet access, node overload, and high server bandwidth cost with the tremendous number of mobile users. We propose a distributed social-based mobile Q&A system with low overhead and system cost as well as quick response to question askers. It enables mobile users to forward questions to potential answerers in their friend lists in a decentralized manner for a number of hops before restoring to the server. It leverages lightweight knowledge engineering techniques to accurately identify friends who are able to and willing to answer questions, thus reducing the search and computation costs of mobile nodes. The trace-driven simulation results show that Q&A system can achieve a high query precision and recall rate, a short response latency and low overhead.

Keywords: question and answer systems, online social networks, non-factual questions

I. Introduction

Traditional search engines such as Google and Bing are the primary way for information retrieval on the internet. To improve the performance of search engines, social search engines have been proposed to determine the results searched by key words that are more relevant to the searchers. These social search engines group people with similar interests and refer to the historical selected results of a person's group members to decide the relevant results for the person.

Although the search engines perform well in answering factual queries for information already in a database, they are not suitable for non-factual queries that are more subjective, relative and multi-dimensional (e.g., can anyone recommend a professor in advising research on social-based question and answer systems?), especially when the information is not in the database. One method to solve this problem is to forward the non-factual queries to humans, which are the most "intelligent machines" that are capable of parsing, interpreting and answering the queries, provided they are familiar with the queries. Accordingly, a number of expertise location systems have been proposed to search experts in social networks or internet aided by a centralized search engine. Also, web Q&A sites such as Yahoo!Answers and Ask.com provides high quality answers and have been increasingly popular.

The social-based Q&A systems can be classified into two categories: broadcasting-based, which broadcast the questions of a user to all of the user's friends, and centralized server, which constructs and maintains the social network of each user, it searches the potential answerers for a given question from the asker's friends, friends of friends and so on.

In respect to the client side, the rapid prevalence of smart phones has boosted mobile internet access, which makes the mobile Q&A system a very promising application. The mobile Q&A systems enable users to ask and answer questions anytime and anywhere at their fingertips. However, the previous broadcasting and centralized methods are not suitable to the mobile environment, where each mobile node has limited resources. Broadcasting questions to a large number of friends cannot guarantee the quality of the answers.

To solve the previous social-based Q&A system, in this paper, we propose a distributed Social-based mObile Q&A System (SOS) with low node overhead and system cost as well as quick response to question askers. It achieves lightweight distributed answerer search, while still enabling

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

a node to accurately identify its friends that can answer a question. The analytical results of the data from the real application show the highly satisfying Q&A service and high performance of SOS.
SOS leverages the lightweight knowledge engineering techniques to transform users' social information and closeness, as well as questions to Ids, respectively, so that a node can locally and accurately identify its friends capable of answering a given question by mapping the question's ID with the social IDs. The node then forwards the question to the identified friends in a decentralized manner. After receiving a question, the users answer the questions if they can or forward the question to their friends. The forwarded along friend social links for a number of hops, and then to the server. The cornerstone of SOS is that a person usually issues a question that is closely related to his/her social life. As people sharing similar interests are likely to be clustered in the social network, the social network can be regarded as social interest clusters intersecting with each other. By locally choosing the most potential answerers in a node's friend list, the queries can be finally forwarded to the social clusters that have answers for the questions.

In a nutshell, SOS is featured by three advantages:
(1) Decentralized. Avoids query congestion and high server bandwidth and maintenance cost problem.
(2) Low cost. Reducing the node overhead, traffic and mobile internet access.
(3) Quick response. An asker identifies potential answerers from his/her friends based on their past answer quality and answering activeness to his/her questions.

The contributions of this work are summarized as follows:
(1) Design a distributed Q&A mobile system based on social networks, which can be extended to low-end mobile devices.
(2) We propose a method that leverages lightweight knowledge engineering techniques for accurate answerer identification.
(3) We use answer quality to represent both the willingness of a node to answer another node's questions and the quality of its answers. We propose a method that considers both interest similarity and answer quality based on past experience.

The Google earns a little higher user satisfaction degree than SOS on factual questions, SOS gains much higher satisfaction degree for non-factual questions than Google.
Note that SOS still has a centralized server to support Q&A activities for questions that are difficult to find answerers in the user social network. SOS also can collect previous questions and answers in the centralized server to improve the Q&A system performance.

II. SYSTEM DESIGN

A. Question Routing

SOS incorporates an online social network, where nodes connect each other by their social links. A registration server is responsible for user registration. Each user has an interest ID, which represents his/her interest. Users who have been willing to answer questions and provided high quality answers to node *i*'s questions previously are more likely to be willing to answer node i's questions and provide high quality answers. Thus, SOS has a metric best answerer ($BA_{(qi,j)}$) that measures likelihood of node *j* to be able and willing to answer node i's question q_i with a high quality answer. It is determined by the interest similarity ($S_{(qi,j)}$) between the questionq_i's interest and node j's interest as well as the answer quality ($Q_{(i,j)}$) of node j to node i's previous questions.

B. Question/User Interest Representation

When a user first uses the SOS system, s(he) is required to complete his/her social profile such as interests, professional background and so on. Based on the social information, the registration server recommends friends to the user, and the user then adds friends into his/her friend list. Each user locally stores his/her own profile and interest ID, and friend list and their interest IDs and answer quality values. Each user calculates his/her own interest ID on his/her social information and sends it to their friends. To calculate interest ID, a node first drives the first-order logic representation (FOL) from its social information, then conducts first-order logic inference to infer its interests, from which it decides the interest ID.
To parse a question, the node first processes the question using natural language processing (NLP), and then represents the question in the FOL format and uses the FOL inference to infer the question's interests. Finally, it transforms the question to a question ID in the form of a numerical string. After a node *i* parses its initiated question *qi* to a question ID, it calculates the interest similarity $S_{(qi,j)}$ for each of its friends $j \in Fi$, where F_i denotes the set of node i's friends. It then calculates the best answerer value ($BA_{(qi,j)}$) for each friend *j* by combining $S_{(qi,j)}$ and answer quality from friend *j* ($Q_{(i,j)}$). Finally, node *i* choose top *K* friends that have the highest $BA_{(qi,j)}$ values to send the question. By comparing the similarity between a question's ID and its friend's interest ID, a node can identify its friends that are able to answer questions.

C. First-order Logic Inference

The FOL inference component consists of three parts: (1) fuzzy database, (2) rules and axioms, (3) inference engine. The goal of the inference is to identify node interests

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

represented by a numerical string that can accurately represent the capability of a node to answer questions. The fuzzy database is used to store words that have relationships, including subset, alias(x), related, with the information in profiles. For example, related (cinema) =movie, subset (computer science, algorithm), alias (USA) =US.
The rule and axioms provide basic formulas for the inference. The inference engine checks the rules and finds related but not obvious information. It sets each interest as an inference goal and builds lattice inference structure, as shown in Figure 1, to connect all the FOL symbols with the goals. Each node in the lattice is an FOL syntax symbol and the arrows represent the connective symbols that connect the symbols.

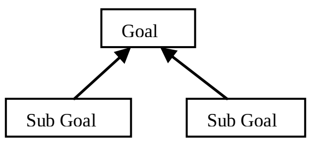

Fig. 1 Lattice in an inference engine.

The question ID and interest ID is generated in the form of a numerical string using the interest arrays.

III. SIMILARITY VALUE CALCULATION

After user's social information and questions are transformed into numerical strings, the similarity between a user and a question can be calculated based on two parts: interest similarity between the user and question, and answer quality between the question sender and receiver.

A. Interest similarity calculation

To evaluate the interest similarity of a question of user i (q_i) and a user j, we use a method proposed in [25]. We use ID_{qi} and ID_j to denote the interest strings of question q_i and user j, respectively. We use $n_{(qi,j)}$ to denote the number of interests owned by ID_{qi} but not by ID_j; use $l_{(qi,j)}$ to denote the number of categories of interest elements owned both by ID_{qi} and ID_j , and $m_{(qi,j)}$ the number of categories of interest elements owned by ID_j but not by ID_{qi}. Then the interest similarity of question q_i and user j is defined as:

$$S_{(qi,j)} = \frac{l_{(qi,j)}+1}{2}\left(\frac{1}{l_{(qi,j)}+n_{(qi,j)}+2} + \frac{1}{l_{(qi,j)}+m_{(qi,j)}+2}\right) \qquad (1)$$

The value of $S_{(qi,j)}$ ranges in the classical spectrum [0, 1], and it represents the level of likelihood that two strings under comparison are actually similar. If two strings have complete overlapping (n=m=0), $S_{(qi,j)}$ approaches 1 as the number of common features grows. The underlying idea of Equation (1) is that two strings with longer complete overlapping should have higher similarity than the two strings with less complete overlapping. In the case of no overlapping (*l=0*), the function approaches to 0 as long as the number of non-shared entries grows. It indicates that two strings with a larger number of entries and share no common entries are more likely to have smaller similarity than the two strings with a smaller number of entries and share no common entries.

B. Answer quality calculation

Social closeness value calculation mechanisms are based on the whole social network topology, which are energy consuming. It is even worse when the social network dynamically changes. Therefore, the topology base social closeness calculation methods are not suitable for energy-stringent mobile devices in SOS. Performance of the SOS largely depends on the activeness and the knowledge base of the users, user i considers the number of received answers from user j and their associated quality ratings when calculating the answer quality of user j. we call it as feedback mechanism. For each received answer, an asker can rate the quality of the answer within rating scale R=[1,5]. The answer quality value is updated based on the number of answers received from friend j during each period T and the associated quality rating ($r\in$ [1, 5]). For the k^{th} question sent from node I to node j, if node I receives an answer from node j during T, x_k=1; otherwise, x_k=0. The parameter x_k is used to represent the willingness of node j to answer questions from node i. Then, the answer quality $Q_{(i,j)}$ is calculated by:

$$Q_{(i,j)} = \alpha. Q_{(i,j)} + (1-\alpha). \textstyle\sum_k (xk.rk\,/R)\ (xk = 0,1) \ (2)$$

Where α ∈ [0, 1] is a damping factor, r_k is node i's quality rating for the k^{th} answer received from node j. A larger $Q_{(i,j)}$ implies that user j is willing and able to provide high-quality answers to user i.
Considering the high dynamism of the social networks, in which the willingness of users to answer questions and the quality of answers from a user to another user may change over time, we add damping factor α into the answer quality calculation.

C. Best answer metric calculation

Based on above sections, for its generated or received question q_i that it cannot answer, node i calculates the best answer metric of each of its friends. That is,

$$BA_{(qi,j)} = \beta S_{(qi,j)} + (1-\beta) Q_{(i,j)} \qquad (3)$$

Where $\beta \in [0,1]$ is a parameter used to adjust the weight of the similarity and answer quality. Node i then selects the top K friends that have the highest $BA_{(qi,j)}$ values and forwards the question to them. Social trust between two nodes decrease exponentially with distance. This relationship has been confirmed by other studies [28, 29]. A reduction in social distance between two persons significantly increases the trust between them.

Algorithm 1 shows the pseudo code of the process for the best answerer metric calculation and best answerer selection conducted by node *i*. If node *i* does not receive answers for its created question during the time corresponding to TTL, it resorts to the centralized server for the answers, where all users conduct Q&A activities in online Q&A sites.

Algorithm 1 Pseudo code of the best answerer identification executed by node *i*.

1: **Input:** *IDi*, *IDj* , *Q*(*i*,*j*) (*j* ∈ *Fi*)
2: **Output:** top-*K* best answerers
3: //*Periodically update* *Q*(*i*,*j*) (*j* ∈ *Fi*)
4: **for** each friend *j* in friend list *Fi* **do**
5: Update *Q*(*i*,*j*) based on Equation (2)
6: **end for**
7: **if** create a question or receive a question it cannot answer **then**
8: **if** TTL>0 **then**
9: **for** each friend *j* in friend list *Fi* **do**
10: Calculate *S*(*qi*,*j*) using *IDqi* and *IDj* based on Equation (1)
11: Calculate *BA*(*i*,*j*) using *Q*(*i*,*j*) and *S*(*qi*,*j*) based on Equation (3)
12: Add *BA*(*i*,*j*) to a list *List*
13: **end for**
14: QuickSort partition around the *Kth* largest element in *List*
15: Find the top-*K* friends having the highest *BA*(*i*,*j*)
16: TTL-=1
17: Send the question to the identified *K* friends
18: **end if**
19: **end if**
20: **if** does not receive answers for its created question during the time corresponding to TTL **then**
21: Resort to the centralized server for the answers
22: **end if**

Line4-Line6 are used to periodically update answer quality of each of its friends. Line8-Line13 calculates each friend's best answerer metric and generates a list including all metric values. Line14-Line17 identifies the top-*K* friends with the highest best answerer metric values and send question to them. Answer quality $Q_{(i,j)}$ is pre-processed and only interest similarity $S_{(qi,j)}$ need to be calculated at run time. The $S_{(qi,j)}$ calculation has a time complexity of $O(|F_i|)$. As the number of keywords in a question is generally very small, the calculation of $S_{(qi,j)}$ should take a short time and costs little computation resources of the mobile devices. This top-*K* friend selection algorithm has a time complexity of $O(|F_i|)$.

IV. CONCLUSION

In this paper, we present the design and implementation of a distributed Social-based mObile Q&A System(SOS). SOS is a novel in that it achieves lightweight distributed answerer search, while still enables a node to accurately identify its friends that can answer a question. SOS uses the FOL representation and inference engine to derive the interests of questions, and interests of users based on user social information. A node considers both its friend's parsed interests and answer quality in determining the friend's similarity value, which measures both the capability and willingness of the friend to answer/forward a question. Compared to the centralized social network based Q&A systems that suffer from traffic congestions and high server bandwidth cost, SOS is a fully distributed system in which each node makes local decision on question forwarding. Compared to broadcasting, SOS generates much less overhead with its limited question forwarding hops. Since each user belongs to several social clusters, by locally selecting most potential answerers, the question is very likely to be forwarded to answerers that can provide answers. The results show that SOS can accurately identify answerers that are able to answer questions. Also, SOS earns high user satisfaction ratings on answering both factual and non-factual questions. In the future, we will study the combination of SOS and cloud-based Q&A system. We will also release the application in the App Store and study the Q&A behaviors of users in a larger-scale social network.

REFERENCES

[1] Google. http://www.google.com
[2] B. M. Evans and E. H. Chi. An elaborated model of social search. Information Processing & Management, 2009.
[3] L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter. Phoaks: A system for sharing recommendations. Comm. of the ACM, 1997.
[4] L. G. Terveen, P. G. Selfridge, and M. D. Long. Living design memory: Framework, implementation, lessons learned. Human-Computer Interaction, 1995.
[5] E. Amitay, D. Carmel, N. Har'El, S. Ofek-Koifman, A. Soffer, S. Yogev, and N. Golbandi. Social search and discovery using a unified approach. In *Proc. of HT*, 2009.
[6] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'el, I. Ronen, E. Uziel, S. Yogev, and S. Chernov. Personalized social search based on the user's social network. In Proc. of CIKM, 2009.
[7] S. Kolay and A. Dasdan. The value of socially tagged URLs for a search engine. In *Proc. of WWW*, 2009.
[8] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *Proc. of WWW*, 2007.
[9] H. H. Chen, L. Gou, X. Zhang, and C. L. Giles. Collabseer: A search engine for collaboration discovery. In *Proc. of JCDL*, 2011.
[10] C. Y. Lin, N. Cao, S. X. Liu, S. Papadimitriou, J. Sun, and

X. Yan. Smallblue: Social network analysis for expertise search and collective intelligence. In *Proc. of ICDE*, 2009.

H. Kautz, B. Selman, and M. Shah. Referral web: combining social networks and collaborative filtering. *Communications* of the ACM, 1997.

D. W. McDonald and M. S. Ackerman. Expertise recommender: a flexible recommendation system and architecture. In Proc. of CSCW, 2000.

F. Harper, D. Raban, S. Rafaeli, and J. Konstan. Predictors of answer quality in online Q&A sites. In *Proc. of SIGCHI*, 2008.

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.

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 AAAI*, 2011.

R. W. White, M. Richardson, and Y. Liu. Effects of community size and contact rate in synchronous social Q&A. 2011.

M. Richardson and R. W. White. Supporting synchronous social q&a throughout the question lifecycle. In *Proc. ofWWW*, 2011.

D. Horowitz and S. D. Kamvar. The anatomy of a largescale social search engine. In *Proc. of WWW*, 2010. Mobile internet stats roundup.http://econsultancy.com/us/blog.

A. Mtibaa, M. May, C. Diot, and M. Ammar. Peoplerank: Social opportunistic forwarding. In *Proc. of infocom*, 2010.

J. Raacke and J. Bonds-Raacke. MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites. *CyberPsychology & Behavior*, 2008.

M. R. Morris, J. Teevan, , and K. Panovich. A comparison of information seeking using search engines and social networks. In *Proc. of ICSWM*, 2010.

Z. Li, H. Shen, G. Liu, and J. Li. SOS: A Distributed Context-Aware Question Answering System Based on Social Networks. In *Proc. of ICDCS*, 2012.

Ni Lao, Tom Mitchell, and William W. Cohen. Random walk inference and learning in a large scale knowledge base. In *Proc. of EMNLP*, 2011.

M. Kirsten and S. Wrobel. Extending K-Means Clustering to First-Order Representations. In *Proc. of ICILP*, 2000.

Z. Li, H. Shen, and K. Sapra. Leveraging Social Networks to Combat Collusion in Reputation Systems for Peer-to-Peer Networks. In *Proc. of IPDPS*, 2011.

Z. Li and H. Shen. SOAP: A social network aided personalized and effective spam filter to clean your E-mail box. In Proc. of INFOCOM, 2011.

C. Binzel and D. Fehr. How Social Distance Affects Trust and Cooperation: Experimental Evidence in a Slum. In Proc. of ERF, 2009.

[11] E-commerce. http://en.wikipedia.org/wiki/E-commerce.

[12] M. F. Bulut, Y. S. Yilmaz, and M. Demirbas. Crowdsourcing location-based queries. In *Proc. of PERCOMW*, 2011.

B. M. Evans, S. Kairam, and P. Pirolli. Exploring the cognitive consequences of social search. In *Proc. of CHIEA*, 2009.

[13] M. R. Morris B. Hecht, J. Teevan and D. Liebling. Searchbuddies: Bringing search engines into the conversation. In Proc. of CHI, 2012.

C. Lampe, J. Vitak, R. Gray, and N. B. Ellison. Perceptions of facebook's value as an information source. In *Proc. Of CHI*, 2012.