

LOAD BALANCING AND SPAM DETECTION ALGORITHM FOR ONLINE SOCIAL NETWORK BASED SYSTEM

S.Arvindsha

UG Scholar
SSE, Saveetha University
sarvinds2@gmail.com

N.Deepa

Assistant Professor
SSE, Saveetha University
deepa23narayanan@gmail.com

Abstract— Question and Answer (Q&A) systems take part in a vital position in our everyday life for information and facts contribution. Users post questions and pick questions to answer in the scheme. Due to the quickly rising user populace and the number of questions, it is improbable for a user to stagger upon a question by chance that (s)he can answer. Furthermore, humanity does not support all users to offer answers, not to talk about high excellence answers with a small answer wait time. The most important purpose of this paper is to look up the performance of Q&A systems by enthusiastically forwarding questions to users who are talented and prepared to answer the questions. To this closing stages, we have intended and done Social Q&A, an online communal network based Q&A system. Social Q&A leverages the social network property of ordinary interest and mutual-trust friend relationship to recognize an asker's friends who are most likely to answer the question. We make clear the structural design, algorithms and user border of Social Q&A, and analyze the Q&A performance of real users and questions from a small-scale real-world communal Q&A system. We as well conducted inclusive large-scale replication to assess Social Q&A in contrast with previous methods. Our results put it to somebody that social network can be leveraged to get better the answer quality and asker's waiting instance.

Keywords — Q&A system, questions, answers, social network, algorithms, Big data.

I. INTRODUCTION

Users have confidence on search engines to find accurate information during this info foundation. Search engines akin to Google and Bing build use of keywords provided by the users to hold out searches. In recent times, industrial investigate and growth performance, like Microsoft and Facebook's social-featured Bing search endeavour, attempt to unite search engines and on-line social networks for advanced search performance.

As previous analysis has indicated, search engines carry out well in compartmentalization web content and as long as users with relevant satisfied to their search on the other hand aren't harmonized for non-factual queries. to handle this unambiguous category of non-factual query, several Question and Answer (Q&A) systems like Yahoo! Answers, Baidu Zhidao, Stack-Exchange, Quora and raise are developed. For illustration, Yahoo! Answers was launched at the top of the year 2005 and taking note over ten million users in February of 2007, and hit two hundred million users in December of 2009. Q&A systems additionally preserve all queries and answers, therefore acting as a repository for info retrieval. They're not solely vital for contribution technological information, however in addition as contribution for receiving recommendation and satisfying one's curiosity a few wide selection of subjects.

With a massive population in a very Q&A system, an outsized figure of queries area unit posed on-line on a daily basis. for instance, there area unit 823,966 queries and answers posed to Yahoo! Answers per day. Then, once a user intends to answer a matter, (s) he could also be overpowered by the over plus of queries. Moreover, merely counting on selfless users to supply answers cannot encourage all users to supply answers and to answer queries quickly.

To find applicable answer suppliers, current Q&A systems allow users to settle on tags (i.e., interest categories) for his or her queries. However, it should not be simple to work out the suitable tag(s) for a matter like “how is that the laptop organization category at our university?” As a result, current Q&A systems might not meet the need of providing prime quality answer with a brief answer wait time, although users want to receive satisfactory answers quickly. This is often established by the study in.

Thus, there's Associate in nursing increasing want for a sophisticated Q&A system that may decrease the quantity of unrequited queries, enhance the solution quality and reduce the latent period. To satisfy this want, we tend to propose Social Q&A, a web social network primarily based Q&A system, that actively forwards inquiries to those users with the best chance(potential and compliance) of responsive them expertly and interest within the questions' subjects. The planning of Social Q&A relies on 2 social network properties. First, social friends tend to share similar interests (e.g., research laboratory members majoring in laptop systems). Second, social friends tend to be trustworthy and selfless owing to the property of “friendship fosters cooperation”. Consequently, Social Q&A favours routing queries among friends and identifies a question's potential answerers by considering 2 metrics: the interest of the friend towards the question and therefore the social closeness of the friend to the asker/forwarder.

The assistance of this effort is as follows:

- 1) The plan of Social Q&A. Social Q&A consists of three parts. User Interest Analyzer connects each user with a vector of notice category. Question Categorizer acquaintances a vector of interest categories to each query. after that, based on user interest and social proximity, Question-User Mapper identify probable answerers for every question.
- 2) Relative trace-driven experiments. We conducted wide-ranging large-scale simulation to weigh up Social Q&A in comparison with other methods. Our results advocate that SocialQ&A improve the excellence of answers and reduce the wait time for answers.
- 3) The growth of a real-world Social Q&A. We have prototyped the SocialQ&A scheme 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 time of just about one month.
- 4) The examination of the data from real Social Q&A. We have analyzed the description of the questions posted, the questioning and answering recital of users, the superiority of answers, and the wait time for answers. systematic results

show the payback of SocialQ&A in enhancing answers value and wait time.

II. EXISTING SYSTEM

In this scheme, planned three innovative supervise conditions weighting schemes for query classification, and evaluate each scheme using a trace from Yahoo! Answers. The 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. These categorization and text mining methods can be used in Social Q&A to more accurately derive user interests and question interests. The Proposed a language model by combining expertise estimation and availability estimation, and later proposed category sensitive language models for expert identification, which helps route questions to available and capable experts. The new proposed a topic based model to identify appropriate answerers by calculating the similarities between questions' topics and user specialists.

DISADVANTAGES

- High level competence
- High level safety measures
- Clear answer to user providing reply.
- Full content is provided to the user.

III. PROPOSED SYSTEM

The proposed Social Q&A, an online communal network based Q&A scheme, that vigorously onwards questions to those users with the uppermost probability (capability and willingness) of answering them with know-how and attention in the questions' subject. The plan of Social Q&A is based on two social network properties. Initial, social associates are likely to contribute to alike interests (e.g., lab members majoring in computer systems) [7]. Next, social acquaintances tend to be truthful and humane due to the belongings of “friendship fosters cooperation” [8]. Accordingly, Social Q&A favours direction-finding queries among friends and identifies a question's potential answerers by considering two metrics: the interest of the friend in the direction of the question and the social nearness of the friend to the asker/forwarder. Thus, the answer receivers have high likelihood of provided that high-quality answers in a short occasion. . LDA algorithm is second-hand for suggestion. Decision tree algorithm used to attain the answer for query.

ADVANTAGES

- Boost the excellence of answers arriving and reduce the wait time for answers.
- It utilize the properties of a social network to presumptuous a question to possible answer provider, ensuring that a given query receives a high- excellence answer in a small phase of time.

IV. PROBLEM STATEMENT

The Main scope of this project is that social networks can be leveraged to improve the answer quality and asker's waiting 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.

V. ARCHITECTURE

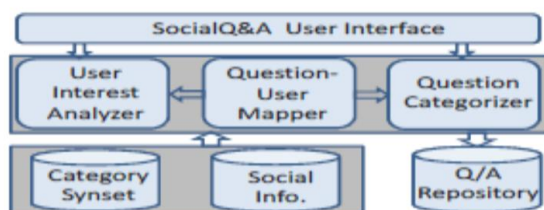


Fig.1. Architecture diagram

Outline shows the high-level structural design of Social Q&A and the interaction amid the central part machinery: User Interest Analyzer, Question Categorizer, and Question-User Mapper. User Interest Analyzer analyze data connected with each user in the social network to derive user interests. Query folder groups the customer questions into attentiveness categories based on the Category Synsets, that provisions the synonyms of all categories keywords from WordNet. Query-User Mapper connect these two mechanism by identifying budding answerers who are most liable to be keen to and be able to supply good enough answers. The information as of user questions and answers is stored on Q/A Repository to hand out successive comparable questions. Also, we present each component and user interface.

VI. COMPONENTS

User Interest analyzer:

User Interest Analyzer utilizes each user's report info contained by the social network and user

interactions (answers provided and queries asked) to determine the interests of the user within the predefined interest categories.

This is often as a result of if a user asks or answers queries in AN interest class, (s) he is probably going to have an interest during this particular class. The order of phrases doesn't essentially represent the different preferences of a user.

Algorithm 1 Pseudocode for the User Interest Analyzer.

Input: A user's profile, questions and answers
Output: The user's interest vector $V_{U_j} = \langle I_i, W_{I_i} \rangle$

- 1: Parse the "interests" field to generate a token stream T_I
- 2: Parse the "activities" field to generate a token stream T_a
- 3: Use the inputs from the user's selection from the Music, Movie, Television and Book fields to generate token streams T_{mu} , T_{mo} , T_t and T_b
- 4: **for** each token stream T_x ($T_x = T_I, T_a, T_{mu}, T_{mo}, T_t, T_b$) **do**
- 5: Check each token in the Synset
- 6: **if** a matching interest category I_i exists **then**
- 7: Update interest weight: $W_{I_i}++$ (e.g., $W_{music}++$)
- 8: **end if**
- 9: **end for**
- 10: Keep updating W_{I_i} based on questions asked and answered and profile update
- 11: Periodically update W_{I_i} using $W_{I_i} = \alpha * W_{I_{old}}$

Fig.2. Algorithm 1

Algorithm 1 shows the pseudo code for the User Interest Analyzer. When a user registers for Social Q&A, (s)he is given the option of entering his/her wellbeing and activities and to mark predefined interest categories to add to his/her interest list. Social Q&A uses WordNet to parse these text fields to token streams (Steps 1-3). For every token, its matching interest category is located in the Synset and corresponding weight is updated (Steps 4-9).

Question Categorizer:

The most important task of Question Categorizer is to classify a question into predefined notice categories based on the theme(s) of the question. We also permit users to input self-defined tags correlate with questions, which are analyzed in question parsing. Question Categorizer generates a vector of question Q_i 's interests, denoted by V_{Q_i} , using a comparable algorithm as Algorithm 1. While processing a question, Social Q&A uses WordNet to inspect the tags and text of the question and generates a token

string. The tokens are compared to Social Q&A's Synset to determine the categories where the question belongs.

Question-User Mapper:

Question-User Mapper identifies the appropriate answerers for a given question. The potential answer providers are chosen from the 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 Social Q&A as it always uses a user's current friends. To check the appropriateness of a friend (U_k) 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 (denoted by Ψ_I, U_k);
- ii) the social closeness between the friend and the asker (denoted by Ψ_C, U_k).

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.

Algorithm 2 Pseudocode for the Question-User Mapper.

Input: Interest vectors of a user, his/her friends and question

Output: A list of potential answer providers

```

1: for each friend  $U_k$  in the friend set of  $U_j$  do
2:   Compute  $\Psi_{I, U_k}$  based on Eq. (1)
3:   Compute  $P_{S, U_k}$ ,  $P_{A, U_k}$  and  $P_{C, U_k}$  based on Eq. (2)
4:   Compute  $\Psi_{C, U_k}$  based on Eq. (3)
5:   Compute  $\Psi_{U_k}$  based on Eq. (4)
6: end for
7: Order the friends in descending order of  $\Psi_{U_k}$ 
8: Notify the top  $N$  friends

```

Fig.3. Algorithm 2

Algorithm 2 shows the pseudo code of the Question-User Mapper. If no one responds during a specific time period, Social Q&A can try the nodes in 2-hop social distance from the asker, and then in 3-hop social distance, until the nodes in Time-To-Live (TTL)-hop social distance have attempted. A question receiver can forward the question if (s)he cannot answer it. The question-user mapper algorithm is called while asking or forwarding questions. When forwarding a question, the asker's information is replaced by the forwarder's information. The Question-User Mapper can be executed in either a centralized manner or a decentralized manner. In the centralized manner, the centralized server

selects the potential answerers for each question and sends the question to them. In the decentralized execution, each node autonomously determines the potential answerers for the question initialized or received by itself to send the question. If there are not enough N selected friends through the Question-User Mapper, the remaining answerers are randomly selected from all users having such interests.

User Interface:

Social Q&A allow users to catalog and adjust user information, put intake away associates, ask/reply/onward questions and make sure question notification. Think about a theoretical user named Mike. When Mike registers, he is required to supply indispensable information regarding himself, such as his individual information, area of study/expertise, his up to date interests, and his attachment in other actions.

VII. SYSTEM CONFIGURATION

HARDWARE CONFIGURATION

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 1 GB.

SOFTWARE CONFIGURATION

- Operating system : Windows XP/7/8.
- Coding Language : JAVA/J2EE
- IDE : Eclipse
- Database : MYSQL

VIII. CONCLUSION

These systems play a significant role in our standard of living for info and information sharing.

Users post queries and choose inquiries to answer inside the system. Attributable to the promptly growing user population and therefore the variety of queries, it's unlikely

for a user to come across a matter by chance that (s) he will answer.

We have a tendency to conjointly improve Social Q&A with security and potency enhancements by protective user privacy and identifies, and retrieving answers mechanically for continual queries.

IX. FUTURE SCOPE

We used these 4 categories as an example and will add more categories in our future work. Evaluation on a large-scale user base remains as our future work. We will conduct tests on a large user base in the real-world experiment. Social Q&A is mainly designed to answer a user's questions in his/her interests. However, a user may sometimes ask questions outside of his/her interests. We will explore a method to handle this case. We will also study how to protect the privacy of questions as requested by the askers.

X. 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] 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://yahoanswers.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.

