

CQA Post Voting Prediction Using Naïve Bayes Algorithm

Dr.pazhani.U,Atchaya. M,Dhatchinakumari.D S

Date of Submission: 25-06-2020

Date of Acceptance: 13-07-2020

ABSTRACT: Community based question and answering (CQA) sites have accumulated vast amount of questions and corresponding crowd sourced answers over time. In existing system, there will be many users who are posting the question in the sites, the major challenge in CQA tasks is the accurate matching of high-quality answers with respect to given questions. So, in the proposed system, these sites will maintain the freelancer, whenever the user will post the query initially it will be directed to admin. The admin will redirect the question to the freelancer. The freelancer is the people who are expert in their domain. The user's question will end in the freelancers sites; the freelancers can post their answers. The main advantage of the proposed system is the accurate matching of high quality answers according to their question which is given by the freelancer. The main benefits in this proposed system is that the user can outsource their project to the freelancer. The algorithm is proposed, so every time the user post the project description in particular category. The description will be redirected to the particular category freelancer and then freelancer can do the project according to the description and can send the source code when is done.

INDEX TERMS – Freelancer, Voting, Bidding process, CQA

I. INTRODUCTION

Community- based Questions- Answering portals. E.g., Yahoo! Answers can be regarded as variations of online forums. Since their inception, they have rapidly gained popularity. For example, Yahoo! Answers has attracted millions of users and hosts a huge number of questions [2]. The question answering content in online communities provides an alternative for for users to obtain information in the form of answers authored by other users, rather than as lists of results or documents from search engines [1]. The users must passively wait for other users to visit the forums, read their questions and provide answers. It may take hours or days from asking a question in a forum before a user can expect to receive answers. [3] Big Data is a term that describes the large volume of data – both structures

and unstructured – that updates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves. A user who knows well the answer to a particular questions may not answer the questions because the user may not visit the forum frequently or the user may be faced with many open questions. On the other hand, a user who answers a question may just happen to see the question, but is not an expert on the question's subject. [1]

II. LITERATURE STUDY

To avoid multiple answers, we present a novel scheme to rank answer candidates via pair wise comparisons. Alleviate this problems is to make use of the vast amount of data recording user interactions with the search results in the novel systems [1]. Question routing has been proposed which aims at routing new questions to suitable answers, who have high ability to answer the questions. In particular, for a newly posted question, we rank a set of users and with high rankings. A Multi Objective Optimization Learning [2].

III. PROBLEM STATEMENT

In this section, we first define the target problem, and then present its solution to illustrate the relationships between our proposed algorithms and the existing work.

3.1 Problem Definitions

These sites contain rich crowd sourcing knowledge contributed by the site users in the form of questions and answers, and these questions and answers can potentially satisfy the information's need of more users. For example, millions of programmers ask and answer questions on Stack Overflow to seek solutions and even more users now use Stack Overflow to seek solutions for programming problems. In this article, we focus on the voting score prediction of questions/ answers shortly after they are posted in the CQA sites. Such a task is essential for the prosperity and sustainability of the CQA

ecosystem, and it may benefit all types of users, including the information producers and consumer. For example, detecting potentially high score answers can benefit the questioners as well as the people who have similar questions; it would also be helpful to identify high-score questions in the early stage and route them to expert answers.

3.2 Demerits

- Both the contextual features (e.g., the reputation of the user who issues the question).
- The content of the post (e.g., keywords, etc.) might affect its voting score, and the effect of each feature might be beyond the simple linear-relationship.
- The voting of a question might be correlated with that of its associated answers.
- Yet, the questions and answers may reside in different feature spaces

IV. PROBLEM SOLUTION

The prominent sites, for example, Stack Overflow and Yahoo gives the inquiry and answer which is posted by the unknown client by everywhere throughout the world. The main advantages of proposed technique is that it provide the efficiency and correctness. In these locales, the software engineers from everywhere throughout the world will respond in due order regarding the questions which are posted on the sites. In the current framework, if any client needs to know the correct answer they have to mine the appropriate response among answer it was extremely tedious. With the goal that client needs to apply all the response to find the correct solution. So to overcome from this circumstances, we proposing the system in which client can see the correct answer from the sites as it will be shown as a first response for that inquiry through post voting forecast. In the voting procedure, the voting will be given in view of the conditions. On the off chance that answer fulfills the conditions, it will be up voted. Thusly, our proposed method gives the adequacy and proficiency, optimality, rightness and unpredictability.

4.1 Merits

- The exact answer will be uploaded and displayed below the question so user will get the exact information.
- The algorithm will set the condition, if answer satisfies the algorithm then only it will be up voted, and given more preference.

V. PROPOSED ALGORITHM

NAIVE BAYES ALGORITHM

Naïve Bayes is a simple but surprisingly powerful algorithm for predictive modeling.

- The representation used by naïve Bayes that is actually stored when a model is written to a file.
- How you can learn a naïve Bayes model from training data.
- How to best prepare your data for the naïve Bayes algorithm.
- Where to go for more information on naïve Bayes.

In machine learning we are often interested in selecting the best hypothesis (h) given data (d). In a classification problem, our hypothesis (h) may be the class to assign for a new data instance (d). Bayes theorem is stated as:

$$P(h|d) = (P(d|h) * P(h) / P(d))$$

Where

- P(h|d) is the probability of hypothesis h given the data d. This is called the posterior probability.
- P(d|h) is the probability of data d given that the hypothesis h was true.
- P(h) is the probability of the hypothesis h being true (regardless of the data). This is called the prior probability of h.
- P(d) is the probability of the data (regardless of the hypothesis).

5.1 Architecture

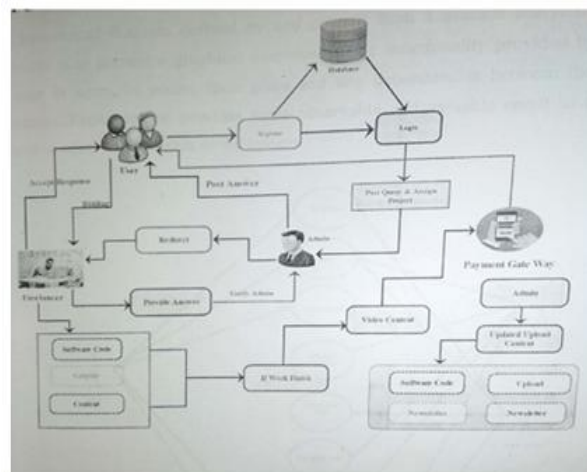


Fig. 1: The whole process of CQA. Shaded boxes are the persons involved and white boxes are the process involved.

VI. CONCLUSION

Communitybased questionanswering Haveattracted substantial users to share knowledge and learn from each other. As the rapid enlargement of CQA platforms, quantities of overlapped questions emerge, which makes users confounded to select a proper reference. There havebeen a growing number of Web information services thatbring Together anetwork of self-declared “experts” to answer other people’s questions. To get fast, relevant answers, the key is getting the right question in front of the right person

REFERENCE

- [1]. T Osbourn, “getting the most out of the web” Software, IEEE, vol. 28, no. 1, pp. 96-96, 2011.
- [2]. Y. Yao, H. Tong, T. Xie, L. Akoglu, F. Xu, and J. Lu, “ Joint voting prediction for Questions and answers in CQA,” In ASONAM, 2014.
- [3]. _____, “Want agood answer? Ask a Good questionfirst!”arXivpreprint arXiv:1311.6876, 2013.
- [4]. C. Saunders, A. Gammerman, andV. Vovk, “Ridge regression learning algorithm in dual variables,” in ICML, 1998, pp. 515-521.
- [5]. S. S. Haykin, Adaptive filter theory, 2005.
- [6]. Y. Engel, S. Mannor, and R. Meir, “The kernel recursive least-squares algorithm,” IEEE Transaction on Signal Processing, vol. 52, no. 8, pp. 2275-2285, 2004.
- [7]. N. Aronszajn, “Theory of reproducing kernels,” Transactions of the American mathematical society, vol.68, no. 3, pp . 337-404, 1950.
- [8]. C. J. Burges, “A tutorial on support vector machines for pattern recognition,” Data mining and knowledge discovery, vol. 2, no. 2, pp. 121-167, 1998.
- [9]. P. Drineas and M. W. Mahoney, “On the Nystrom method for approximating a gram matrix for improved kernel-based learning,” The Journal of Machine Learning Research, vol. 6, pp. 2153-2175, 2005.
- [10]. G. Golub and C. Van Loan, “Matrix computations,” 1996.



**International Journal of Advances in
Engineering and Management**

ISSN: 2395-5252



IJAEM

Volume: 02

Issue: 01

DOI: 10.35629/5252

www.ijaem.net

Email id: ijaem.paper@gmail.com