International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 03, Issue – 04, April 2018, PP – 55-58

Expert Non-Expert Classifier

NishantVhatkar[1], Tejas Mudholkar[2], Prathamesh Nadgouda[3], Aishwarya Patil[4], Prof. Sagar Patil

[1]*(Computer science, Bharati vidyapeeth/ ShivajiUniversity, India)*
[3]*(Computer science, Bharati vidyapeeth/ Shivaji University, India)*
[3]*(Computer science, Bharati vidyapeeth/ Shivaji University, India)*
[4]*(Computer science, Bharati vidyapeeth/ Shivaji University, India)*
[5]*(Computer science, Bharati vidyapeeth/ Shivaji University, India)*

Abstract: Now a day's use of Question Answering (QA) sites is increasing rapidlyit provides all the possible answers by faculty experts. Stack overflow, Quora, yahoo, Super User etc. are some QA sites providing great service in QA. Quora is one of the popularsite where users ask some questions and the other users answer them. The quality of answer is checked by using the upvote and downvote. Unfortunately, some users avoid voting there or use wrong vote that will not help user to find **better** answers. To fill the gaps, in this existing system we (1) analyze properties & features of experts and non-experts on some. popular topics; (2) Then that features are tested by using Linguistics rules which differentiate the experts and non-expert on some marking and (3) develop statistical models based on the features to automatically detect experts. Our experimental results show that our module identify experts in general topics and a specific topic.

Keywords: liws, random forest, web crawler

I. INTRODUCTION

As increase in users of QA sites the answer giving people also increased. Quora is one of the question answer site which is different than other sites. In this site user can follow topics in which he has interest and has some knowledge and can follow some user which will help us to find the answers of Queries the ask on site. To find the quality of answer there is a upwoting and downwoting policy which will help them to find which answer is better one.

Generally, the answering people are the experts which are always active which give the proper answer, which update his answer according to questions requirement regularly, which have more followers, which have great knowledge of the topics he followed.

In the previous system that upwoting and downvoting may lead to wrong guidance as some user are not aware of that policy and some don't used to participate in voting. Hence only that voting is not helpful to decide which one is expert and which one is non expert.

As Q&A sites have become popular, people have desire to quickly identify experts in general topics or a specific topic. New users are not familiar with the community, but they want to find experts who could give them relevant answers. Also, expert finding can be used for an expert recommendation service in a social Q&A site.

In our paper we tried to find answers of some questions like do experts and non-expert behave different? Do they change behavior according to time? Can we detect best answer automatically? Can we differentiate experts and non-experts in general and specific topic? For that wedo the following things.

- As Quora don't provided their official API we have to parse that site by creating crawler.
- Analysis them.
- Create dictionaries of LIWS Words
- Write algorithm which will classifyexpert and non-expert for unknown user.

To analyse behaviours of experts and non-experts on Quora, the first step is to collect user information. As mentioned earlier since there is no publicly available official APIs, we developed our own crawler which collected user information on Quora.Our crawling strategy is to first manually given some users answers URL. Crawler contains the answers and their features. As shown in Fig (1) user profile contain their personal information, followers, following, answers given, posts, questions asked, Blogs etc. By running our crawler we collect information of some user. First database contains answers of the users second database contain there features. In this way we just cleanthe data and pick the required data only.

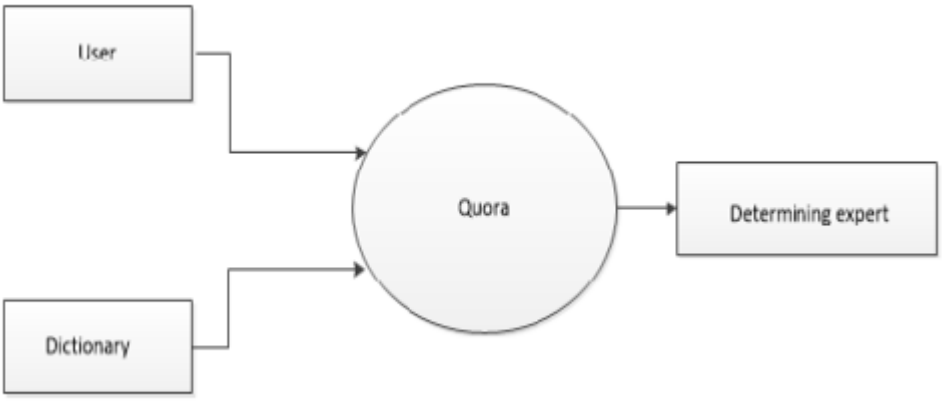

Fig.1 Labeling expert – Non expert

Next step would be label the dataset according to expert and non-expert.In general, an expert is a person who has a great knowledge of or skill in a particular area. To differentiate the expert and non-expert we use the LIWS rules which gives some marking according to features of users. We store that score in database and according to that score user will be labeled as expert or non-expert manually. (refer Fig.1)

Next step will be extracting questions and answers from that user profile and evaluate features of that answers according to LIWS rules (refer Fig.2). This is nothing but counting the number of words given in LIWS rules used by the user in answer. This features will be saved in database and later on converted as a text document which will be input to training set.

Same procedure will be followed to the unknown user and then both text files will be given input to the algorithm where we will classify unknown user is expert or non-expert. (refer Fig.3)

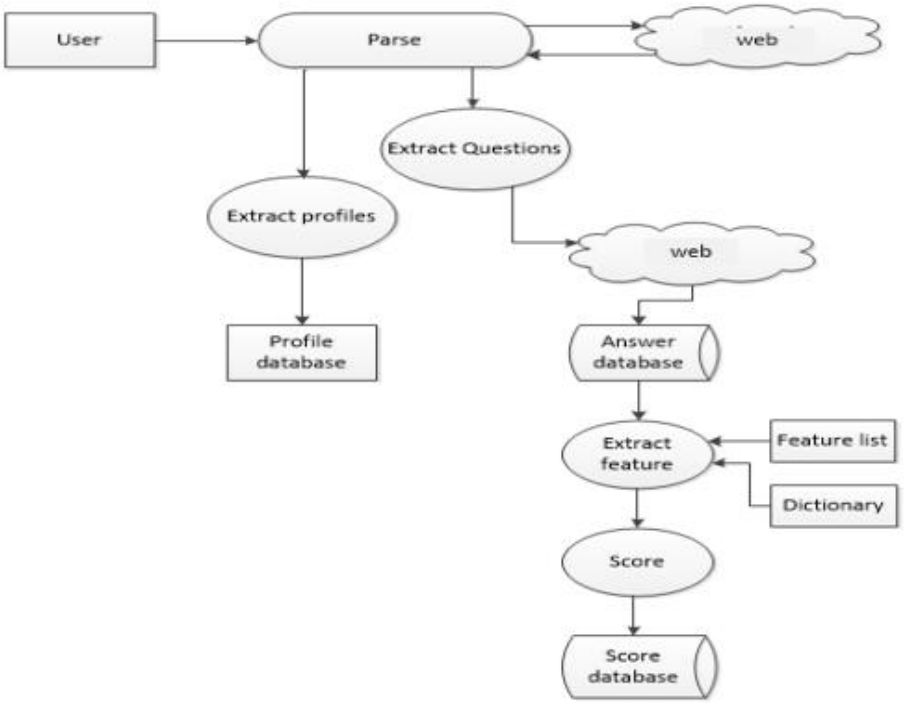

Fig.2 Question and Answer extraction

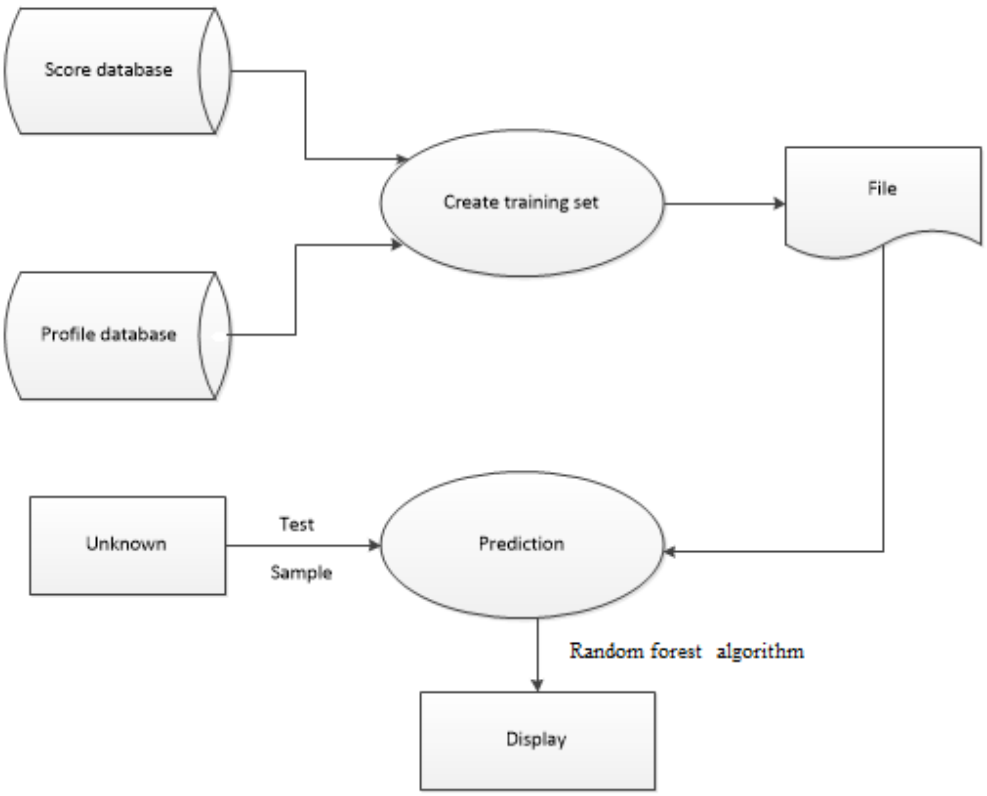

Fig.3 Algorithm Implementation & Classification

There are lots of algorithm available for classification namely navies bias,Random forest, J48, SMO etc.
Out of that random forest algorithm will give more accuracy.

II. LITERATURE REVIEW

Firstly, it is important to understand what are the Q & A sites? How they work? What people do with this site? Q&A sites are categorized in 3 types digital reference services, ask an expert service, community Q&A sites. Sites are also categorized according to responsiveness of sites about that questions.

For this we take reference of the paper [1]. Paper [1] give us idea of some Q&A sites. Their behaviour, their working and how people deal with these sites. This helps us to understanding how to deal with these sites and how they work.

Secondly, we have to check quality of answer on different parameters or features like number of upper case words used, number of pronounce used, number of first singular pronoun and so on.

For this we take reference of paper [2]. Paper [2] gives us idea of such parameters which we can use to check quality of answer. It gives us brief description of linguistic rules and shows some ways to check quality of answer.

Lastly, we have to find expert users according to question asked. For that we have to check all the parameters like answers given, followers, upvotes and so on.

For this we take reference of paper [3] and [4]. In that paper we get different parameters to differentiate expert and non-expert.

Since our goal is to find expert by using linguistic rules that rules need to be studied well. That all 68 LIWC rules can be studied from paper [4].

III. CONCLUSION

We have presented web application which reduce efforts of the user of finding best answer from the given answer on Quora. This will reduces manual efforts. This system is support to eliminate and reduce the hardships faced by existing system. This system is prepared to predict the best answer and help to find experts to new community which don't know about them. This system is showing the accurate prediction that user get expert advice in less time. It is user friendly system. As we are using classifier algorithm this app can be used to any application where we need to classify something eg. In Gmail received mail can categorized into spam inbox updates like categories with little change in code.

REFERENCES

Journal Papers:

[1]. Bouguessa M, Dumoulin B, Wang S (2008) Identifying authoritative actors in question-answering forums: The case of yahoo! answers. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining.

[2]. Pal A, Farzan R, Konstan JA, Kraut RE (2011) Early detection of potential experts in question answering communities. In: Proceedings of the 19th international conference on User modeling, adaption, and personalization.

[3]. Zhou G, Liu K, Zhao J (2012) Joint relevance and answer quality learning for question routing in community QA. In: Proceedings of the 21st ACM international conference on Information and knowledge management

[4]. Sumanthpatil, Kyumin Lee (2015) Detecting expert on Quora: by their activity, quality of answers, linguistic characteristics and temporal behaviours.

[5]. Jeon J, Croft WB, Lee JH, Park S (2006) A framework to predict the quality of answers with non-textual features. In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval.

[6]. Kao WC, Liu DR, Wang SW (2010) Expert finding in question-answering websites: a novel hybrid approach. In: Proceedings of the 2010 ACM Symposium on Applied Computing.

[7]. Schall D, Skopik F (2011) An analysis of the structure and dynamics of large-scale q/a communities. In: Proceedings of the 15th international conference on advances in databases and information systems (ADBIS)

[8]. Zhu Z, Bernhard D, Gurevych I (2009) A multi-dimensional model for assessing the quality of answers in social Q & A sites. In: Proceedings of 14th International Conference on Information Quality

