

System to Rank Results in CQA Sites

Vishwajeet Rapte¹, Amit Khebade², Sawani Sapkal³, Shaheen Yasmin⁴, Prof. Amol Lachake⁵
^{1,2,3,4,5}*Dept. of Computer Engineering, Dr. D Y Patil School of Engineering and Technology, Lohegaon, Pune*

Abstract - Question Answering (CQA) is a specially designed and an appropriate form to retrieve information. An available set of credentials, a Question Answering system bids to fetch the correct answers to questions which are posed. For the purpose to discover a similar question from available data means historical data records has been put to question answering, with great hypothetical and sensible achievement and hence users have to allusion with credentials or even identical passage as the majority of the data recovery systems before finding the correct one answer. To deal with such issues, there is a need of a system that permits users to ask a question in daily language as well as get an answer quickly and concisely, with enough contexts to constitute the answer. Since users move violently to find the way the prosperity of online data currently obtainable, the required for automatic question answer system becomes most vital. A proposed system defines a new system to grade answer candidates via pair wise comparison. In proposed system, it includes two type of elements, Offline Learning as well as Online Search. The system species out the answer candidates through influencing the offline trained form to examine the priority orders. The data-driven draw can produce higher level presentation and complement conventional question answering methods which are motivated through data withdrawal.

Index Terms - Community question answering, Rake_nltk, K-nearest neighbor, Word embeddings.

I. INTRODUCTION

Community question answering (CQA) system provides a platform for people to share knowledge and is used by enormous number of users every day to find an answer on numerous varieties of questions. There is a wide variation of questions asked like truth-seeking, knowledge of general world questions, which have definite answers, and also judgement-seeking questions, questions about personal belief which do not have any explicit answers but they can be judged as either acceptable or unacceptable. In the current scenario, people generally have to explore a lot before

getting a correct answer to their queries and also a vast segment of questions do not get any feedback in return for an extensive duration of time.

Due to this, there is an absolute need of a system that permits any user to search for any question on any CQA website and get their answers, quickly and accurately. Most of the users are impatient to find the solution for their queries online. Thus, their requirement for automatic question answering system becomes vital. The scope of our proposed system is to find similar questions, their multiple answers from past obtained data and then prioritize all those answers, with an effective conceptual base and attain empirical success. Moreover, each and every question is often associated with copious amount of answers, due to which users have to explore a good deal, prior to obtaining the unerring one. Therefore, to lessen these problems, we bring forward our trail-blazing project which ranks each of those answers of candidates via pairwise comparison.

II. LITERATURE REVIEW

¹*Hanyin Fang, Fei Wu, Zhou Zhao, Xinyu Duan and Yueting Zhuang* used a framework Heterogenous Social Network Learning (HSNL) which encodes the contents of question-answer as well as the signs of communal interconnection in the society to improve the CQA tasks by using a random walker.

²*Geerthik S, K. Rajiv Gandhi, and Venkatraman S* proposed RespondRank, which ranks the user answers by recognizing accurate answers better compared to contemporary techniques. Also, the answers given by different users is accepted properly. Experiments which they have performed on a popular CQA called Quora, conveys that the improvement in ranking exhibited by their system is remarkable compared to the current ranking methods.

³Shivani Singh, Nishtha Das, Rachel Michael and Dr. Poonam Tanwar worked on a smart learning system where the input will be a text file and the knowledge will be received from the given text. Thus, their system will try to answer questions asked by the users using this knowledge. Their system (QAS) will help users to find particular answers to particular questions in a restricted domain. The segregated group open-domain comprises of a variety of questions about practically everything i.e. World Wide Web whereas the questions under a particular domain like education, weather, politics, music etc. is dealt by the closed domain.

⁵Liqiang Nie, Xiaochi Wei, Dongxiang Zhang, Xiang Wang, Zhipeng Gao and Yi Yang proposed a system which consists of one offline learning module and one online search module. The system automatically establishes the positive, negative and neutral training samples regarding the preference pairs in the offline learning module which is guided by their data-driven experimentations. A list of answer candidates was collected for a given question in the online search module. The model then sorts the answer candidates by utilizing the offline training module to decide the orders having high preference.

⁶Sunny S. Shah, Tejas S. Bavaskar, Sourabh S. Ukhale, Akash S. Kalyankar and Rahul A. Patil proposed a system where the repetition of questions having similar interpretations will be removed and on the basis of the number of comments and the likes received, the ranking of answers will be done. Characteristics like bookmarking the answer liked by user and sending those question and answer to the user via email have also been added in the proposed system. Also, their proposed system helps the user to get suggestions based on the user's interest and activity.

⁹Nouha Othman, Rim Faiz and Kamel Smaili uses word embeddings in their system to remove the complication of word mismatch between questions which occurs when the same question is asked by the user using different words. But the answers from the past similar questions will be the same for the new questions asked in the present.

¹⁰Vaibhav Mishra and Nitesh Khilwani developed Question Answering System for Education (QUASE) which contains a restricted set of records which in turn

helps to find answers in a closed domain for any given question. Several NLP techniques are used in the system like Question Taxonomy for question categorization, Parts of Speech, Lemmatization etc. for document processing to make search better. Their observations have shown a remarkable improvement in allocating the questions into correct answer types with approximately 91% accuracy, thus providing a better execution as a Question-Answering system in closed domain search.

III. PROPOSED WORK

The proposed algorithm is based on Rake_nltk which helps in generating tags for questions and allows users to rank answers. Also, we can keep all the answers by different users. Our system is divided into six components: home page, contact page, my-question page, ask question page, sign-up page and login page.

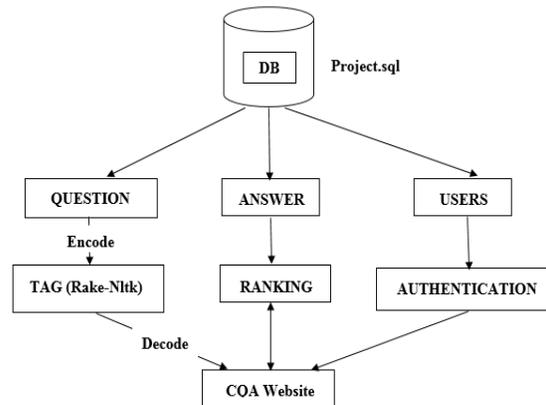


Figure 1: System Architecture

For python tag function to work, we imported various system functions which are system, base 64, json, rake_nltk. We will not be able to directly process the questions, so we have used json functionality to import the question in rows so that the system can extract the keywords from the text of the question, there after we can extract the keywords from the text. To resend the tags which are processed by rake_nltk we convert list to json format so that it is readable by php.

The processing of extraction of tags for question is a necessity. To use this functionality in our system, we use library from python. While we would have used nltk, it had many complex operations and every different command to be used will give its speed of execution less. So, we have used rake_nltk.

Rake_nltk is the external library that we have used by installing via command prompt in windows for tag generation. The functionality works by using extraction commands which are specified.

The logic of ranking answers lies between the logic of likes and dislikes count. The likes will prioritize the answers by +1 and dislike will -1 (decrease) the priority of answers and pushes it to the bottom.

Management of connection in php and python functionality: usage of json encode as the database readability is in json format which decodes it via json library available in python. It is used for decoding every question that is dynamically updated.

KNN Algorithm: For pairwise answer comparison, we chose KNN algorithm. for each question and answer, different clusters are formed.

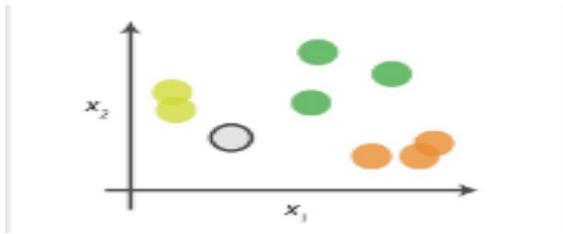


Figure 2: Look at The Data

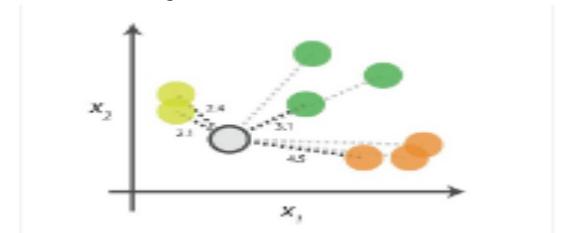


Figure 3: Calculate Distances

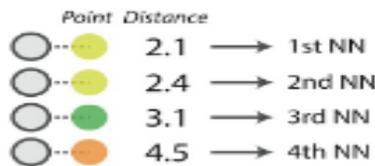


Figure 4: Find Neighbors

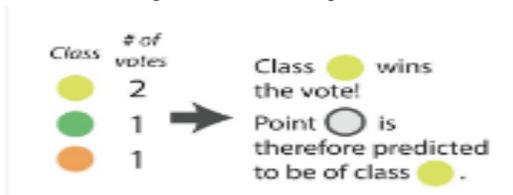


Figure 5: Vote on Labels

System Requirements

Software Requirements:

- Language used: Python 3.7.6
- IDE: Spyder
- Database: MySQL
- Operating System: Windows 7 & above

Hardware Requirements:

- Processor: Intel Pentium Processor Core II or higher
- Hard Disk: 2 GB minimum
- RAM: 2 GB minimum
- Processor Speeds: 3.2 GHz or faster processor

Advantages:

- Achieves better performance.
- It is robust to the noises caused by enlarging the number of returned similar questions.
- It improves the scalability and performance of searching.

IV. RESULT AND ANALYSIS

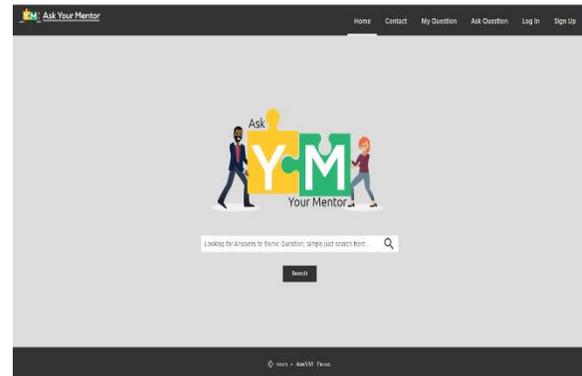


Figure 6: Home Page

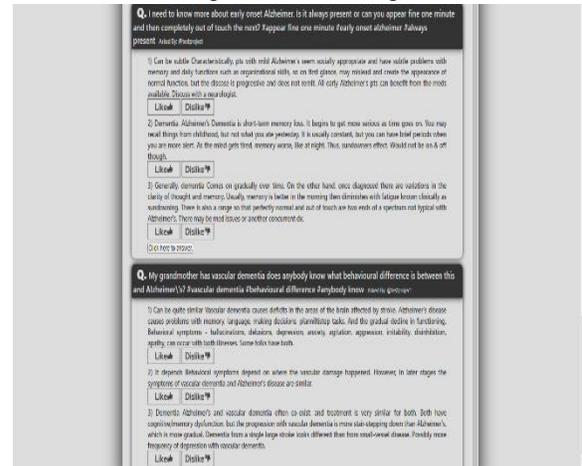


Figure 7: Search Results

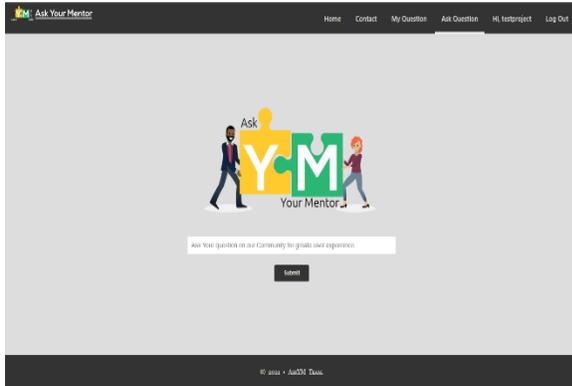


Figure 8: Ask Question

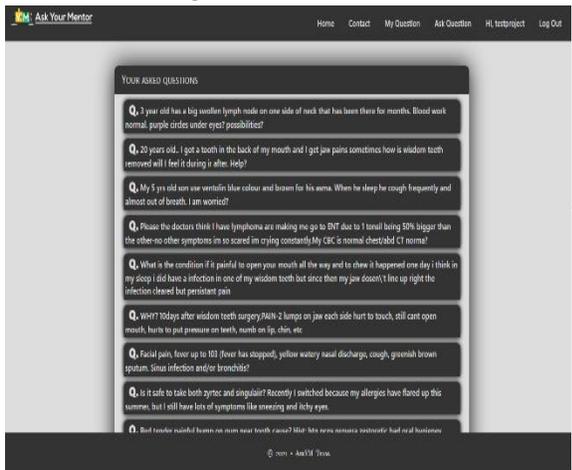


Figure 9: My Question

REFERENCES

- [1] Hanying Fang, Fei Wu, Zhou Zhao, Xinyu Duan, Yueting Zhuang, “Community Based Question Answering via Heterogeneous Social Network Learning”, AAAI’16, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 2016.
- [2] Geerthik S, K. Rajiv Gandhi, Venkatraman S, “Respond Rank: Improving Ranking of Answers in Community Question Answering”, International Journal of Electrical and Computer Engineering (IJECE), Vol. 6, No. 4, August 2016.
- [3] Shivani Singh, Nishtha Das, Rachel Michael, Dr. Poonam Tanwar, “The Question Answering System Using NLP and AI”, International Journal of Scientific and Engineering Research, ISSN 2229-5518, Vol. 7, Issue 12, December 2016.
- [4] Lorena Kodra, Elinda Kajo Mece, “Question Answering Systems: A Review on Present Developments, Challenges and Trends”, International Journal of Advanced Computer Science and Applications, Vol. 8, No. 9, 2017.
- [5] Liqiang Nie, Xiaochi Wei, Dongxiang Zhang, Xiang Wang, Zhipeng Gao, and Yi Yang, “Data Driven Answers Selection in Community QA Systems”, IEEE Transactions on Knowledge and Data Engineering, Vol. 29, No. 6, June 2017.
- [6] Sunny S. Shah, Tejas S. Bavaskar, Sourabh S. Ukhale, Akash S. Kalyankar, Rahul A. Patil, “Answer Ranking in Community Question Answer (QA) System and Questions Recommendation”, Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), August 2018.
- [7] Bolanle Ojokoh and Emmanuel Adebisi, “A Review of Question Answering Systems”, Journal of Web Engineering (JWE), January 2019.
- [8] Zhengfa Yang, Qian Liu, Baowen Sun, Xin Zhao, “Expert Recommendation in Community Question Answering: A Review and Future Direction”, International Journal of Crowd Science, Vol. 3, Issue 3, September 2019.
- [9] Nouha Othman, Rim Faiz, Kamel Smaili, “Enhancing Question Retrieval in Community Question Answering Using Word Embeddings”, Proceedings of the 23rd International Conference on Knowledge-Based and Intelligent Information

V. CONCLUSION

A. Future Work

- Introducing the voting for answers,
- Showing the rank count for answers.

B. Conclusion

A proposed system generates a novel scheme in Community Question Answering settings for answer selection. System compass two components that is offline learning and other one is online search module. If the keyword given for the question asked has no match in the database, it automatically moves to the online search component. The usage of KNN algorithm speeds the ranking of the answers based on the likes and dislikes. In place of protracted and manual annotation, system axiomatically build up the sanguine, apathetic and unbiased training sample in the way of partiality pairs in an offline learning component which is suggested by our data driven explanation.

& Engineering Systems (KES2019), September 2019.

- [10] Vaibhav Mishra, Nitesh Khilwani, “QUASE: An Ontology-Based Domain Specific Natural Language Question Answering System”, International Journal of Recent Technology and Engineering (IJRTE), ISSN 2277-3878, Vol. 8, Issue 4, November 2019.