Expert System for Problem Solution in Quality Assurance System

Adesh Vikram Singh
DY Patil Institute of
Engineering and Technology
Pune, Maharashtra
India

Ashish Kaul
DY Patil Institute of
Engineering and Technology
Pune, Maharashtra
India

Ankit Kumar Thakur
DY Patil Institute of
Engineering and Technology
Pune, Maharashtra
India

Abstract: To find a particular solution for a given problem or to answer any question we need to find a solution from the given set of solution by pair wise comparisons. A historical set of data help us to find different answers for the same question asked multiple times. In our proposed system there are two models one is "offline model" and another one is "online model". In offline learning component we establish training samples with the help of data driven observations. In online search component a pool of candidate answer is collected for a given question by finding similar types of questions. With the help of offline trained model candidate answer are sorted in preference order. A comparative demonstration of the experiments on the real-world community-based question answering is shown.

Keywords: pair-wise**;** offline mode; online model; training samples; comparative.

1. INTRODUCTION

As our project purely based on data mining, the data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database system. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and management aspects, data pre-processing and interface considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Community Question Answering (CQA) is gaining popularity online. They are seldom moderated, rather open, and thus they have few restrictions, if any, on who can post and who can answer a question. On the positive side, this means that one can freely ask any question and expect some good, honest answers. On the negative side, it takes effort to go through all possible answers and to make sense of them.

For, example, it is not unusual for a question to have hundreds of answers, which makes it very time consuming to the user to inspect and to winnow. The challenge we propose may help automate the process of finding good answers to new questions in community created forum.

2. LITERATURE REVIEW

1) "Data-Driven Answer Selection in Community QA Systems", Liqiang Nie, Xiaochi Wei, Dongxiang Zhang, Xiang Wang, Zhipeng Gao, and Yi Yang, we present a novel scheme to rank answer candidates via pairwise comparisons. In particular, it consists of one offline learning component and one online search component. In the offline learning component, we first automatically establish the positive, negative, and neutral training samples in terms of preference pairs guided by our data-driven observations. We then present

a novel model to jointly incorporate these three types of training samples. The closed-form solution of this model is derived. In the online search component, we first collect a pool of answer candidates for the given question via finding its similar questions.

2) "Disease inference from health-related questions via sparse deep learning" L. Nie, M. Wang, L. Zhang, S. Yan,B. Zhang, and T. S. Chua, [2] Proposed a paper aims to build a disease inference scheme that is able to automatically infer the possible Diseases of the given questions in community-based health services. In this paper, we first report a user study on the information needs of health seekers in terms of questions and then select those that ask for possible diseases of their manifested symptoms for further analytic. We next propose a novel deep learning scheme to infer the possible diseases given the questions of health seekers. The proposed scheme comprises of two key components.

3) "Multi-VC Rank with applications to image retrieval", X. Li, Y. Ye, and M. K. Ng, propose and develop a multi-visual concept ranking (Multi-VC-Rank) scheme for image retrieval. The key idea is that an image can be represented by several visual concepts, and a hypergraph is built based on visual concepts as hyperedges, where each edge contains images as vertices to share a specific visual concept.

4) "Beyond text QA: Multimedia answer generation by harvesting Web information", L. Nie, M. Wang, Y. Gao, Z. Zha, and T. Chua, in this paper, we propose a scheme that is able to enrich textual answers in cQA with appropriate media data. Our scheme consists of three components: answer medium selection, query generation for multimedia search, and multimedia data selection and presentation. This approach automatically determines which type of media information should be added for a textual answer. It then automatically collects data from the web to enrich the answer.

5) "A ranking approach on large-scale graph with multidimensional heterogeneous information," IEEE Trans. W. Wei, B. Gao, T. Liu, T. Wang, G. Li, and H. Li, address the large-scale graph-based ranking problem and focus on how to effectively exploit rich heterogeneous information of the graph to improve the ranking performance. Specifically, we propose an innovative and effective semi-supervised Page Rank (SSP) approach to parameterize the derived information within a unified semi-supervised learning framework (SSLF-GR), and then simultaneously optimize the parameters and the ranking scores of graph nodes.

3. EXISTING SYSTEM

To make question respondent system time effective and to cut back user's efforts to search out actual answers for his question by suggesting him antecedent answered same form of queries with its ranking. to beat this downside, we have a tendency to use sentence level bunch, this system provides multiple answers which is able to be the precise match for that question. However, the actual queries have multiple answers. So, it's tough to outline a selected declare single question.

4. PROPOSED SYSTEM

Offline Learning: -
In the offline learning component, instead of long and labour-intensive annotation, we tend to tend to automatically construct the positive, neutral, and negative coaching job samples inside the kinds of preference pairs guided by our data-driven observations.

Online search: -
We initial collect a pool of answer candidates via finding its similar queries.

Database: -
A tremendous vary of historical QA pairs, as time goes on, area unit archived inside the cQA databases. information seekers thus have large prospects to directly get the answers by trying from the repositories, rather than the long waiting.
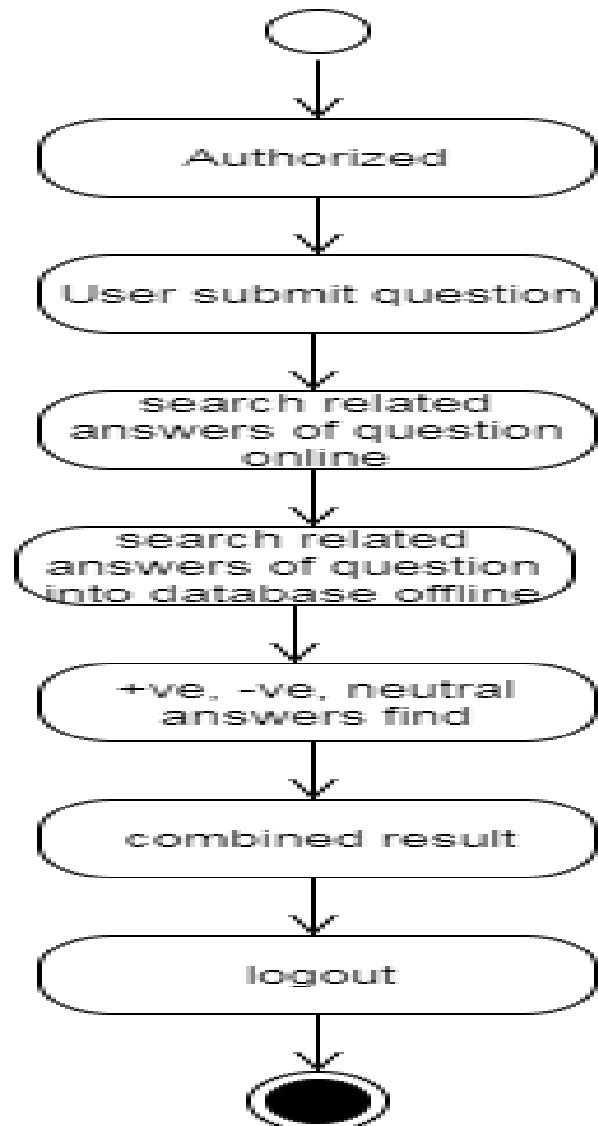
Sentence level clustering:
A question that has multiple forms of answers, however providing best suited answer of that question.

Expert Recommendation system:
Same sort of question that is answered by consultants those consultants are recommending to user for more queries.
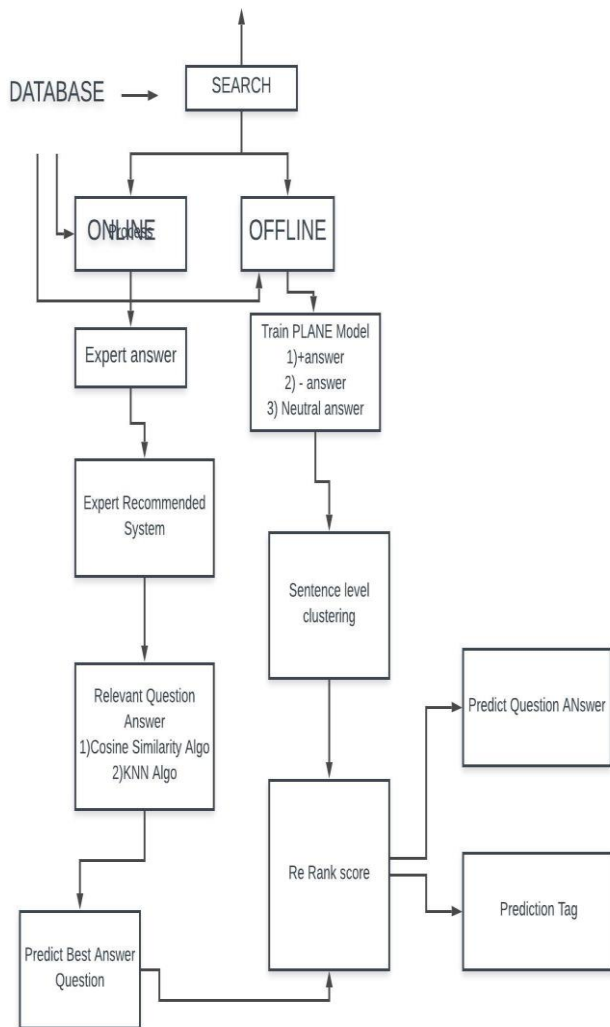
5. FLOW DIAGRAM

6. SYSTEM ARCHITECTURE

Fig. System Architecture

7. FUTURE SCOPE AND CONCLUSION

Attribute-based encoding has been wide employed in cloud computing wherever information suppliers source their encrypted information to the cloud and might share the info with users possessing mere credentials. On the opposite hand de-duplication is a vital technique to avoid wasting the space for storing and network information measure, that eliminates duplicate copies of identical information.

8. REFERENCES

[1] J. Jeon, W. B. Croft, J. H. Lee, and S. Park, "A framework to predict the quality of answers with non-textual features," in Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2006, pp. 228–235.

[2] Z. Ji and B. Wang, "Learning to rank for question routing in community question answering," in Proc. 22nd ACM

Int. Conf. Inf. Knowl. Manage., 2013, pp. 2363–2368.

[3] T. C. Zhou, M. R. Lyu, and I. King, "A classification-based approach to question routing in community question

answering," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 783–790.

[4] L. Yang, et al., "CQArank: Jointly model topics and expertise in community question answering," in Proc. 22nd

ACM Int. Conf. Inf. Knowl. Manage., 2013, pp. 99–108.

[5] B. Li and I. King, "Routing questions to appropriate answerers in community question answering services," in Proc.

19th ACM Int. Conf. Inf. Knowl. Manage., 2010, pp. 1585–1588.

[6] K. Wang, Z. Ming, and T.-S. Chua, "A syntactic tree matching approach to finding similar questions in community-based QA services," in Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2009, pp. 187–194.

[7] Y. Liu, J. Bian, and E. Agichtein, "Predicting information seeker satisfaction in community question answering," in

Proc. 31st Annu. Int. ACMSIGIR Conf. Res. Develop. Inf. Retrieval, 2008, pp. 483–490.

[8] M. J. Blooma, A. Y. K. Chua, and D. H.-L. Goh, "A predictive framework for retrieving the best answer," in Proc.

ACM Symp. Appl. Comput., 2008, pp. 1107–1111.

[9] L. Nie, M. Wang, Y. Gao, Z. Zha, and T. Chua, "Beyond text QA: Multimedia answer generation by harvesting Web information," IEEE Trans. Multimedia, vol. 15, no. 2, pp. 426–441, Feb. 2013.