

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

A Survey on Novel Approach to Construct Community QA Systems

Ajinkya Ramchandra Ugale, V.S Wadne

M. E Student, JSPM's Imperial College of Engineering and Research, Wagholi, Pune, India

Professor, JSPM's Imperial College of Engineering and Research, Wagholi, Pune, India

ABSTRACT: On web user post different question and get answer to that question by different view. Proposed system will work on to give answer to question within time and provide relevancy in answer by working on pairwise learning technique. To find the similar questions is very difficult in Question Answering (QA) System. Because each question in the returned candidate pool consist with multiple answers, and hence users get trouble to browse a lot before finding the correct one. To solve this problem proposed a novel approach a novel Pairwise Learning to rANK model i.e PLANE which can quantitatively rank answer candidates from the relevant question pool. Specifically, it uses two components i.e one offline learning component and one online search component. In the online search component, System initially gathers a pool of answer candidates for the given question by means of discovering its comparable or similar questions by proposed algorithm. System at that point sorts the appropriate answer candidates by utilizing the offline trained model to judge the preference orders. Our model is effective as well as achieves better performance than several existing questions answer selection system. In proposed system recommend the question to other similar user to answer the asked question based on past history of users.

KEYWORDS: Community-based Question Answer, Answer Selection, Observation-guided Training Set Construction, Pairwise learning technique.

I. INTRODUCTION

Community question answering system (cQA), one of the fastest-growing user-generated-content (UGC) portals, has risen as an enormous market, so to speak, for the fulfillment of complex information needs. QA enables users to ask/answer questions and search through the archived historical question-answer (QA) pairs. Despite the success of CQA and active user Demand for famine participation exists extensively in CQA sites, which refers to the following two types Phenomena: Firstly, those looking for information usually have to wait long before getting answers to their questions. Secondly, a large percentage of questions do not any response even within a relatively long period. Question starvation is probably caused by several Reasons: First the questions are badly formulated, Ambiguous or less interesting to all. CQA systems are hardly to route the newly posted questions to the appropriate answerers. Second Potential interviewees have experience, But they are not available or overwhelmed by pure volume of incoming questions. This case often occurs in the vertical CQA forums, so only experts can answer these questions. In proposed system recommend the question to other similar user to answer the asked question.

II. RELATED WORK OR LITERATURE SURVEY

1.X. Wei, H. Huang, C. Lin, X. Xin, X. Mao, and S. Wang(2015) have represents Reranking voting-based answers by discarding user behavior biases [5]. In generating a vote, a user's attention is influenced by the answer position and appearance, in addition to right answer quality. Previously, these biases are ignored. As a result, the top answers obtained from this mechanism are not reliable, if the number of votes for the active question is not sufficient. The author solve this issue by analyzing two kinds of biases; position bias and appearance bias. To identify the existence of these biases and propose a joint click model for dealing with both of them.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

2. **Q. H. Tran, V. Duc, Tran, T. T. Vu, M. L. Nguyen, and S. B. Pham (2015) Jaist: Combining multiple features for answer selection in community question answering [6].** The author designed Answer Selection in Community Question Answering. In this task, the systems are required to identify the good or potentially good answers from the answer thread in Community Question Answering collections. This system combines 16 features belong to 5 groups to predict answer quality. This final model achieves the best result in subtask A for English, both in accuracy and F1-score.
3. **Savenkov (2015) has represents Ranking answers and web passages for non-factoid question answering: Emory university at TREC liveQA [7].** The author represents how to automatically answer questions posted to Yahoo! Answers community question answering website in real-time. This system combines candidates that extracted from answers to similar questions previously posted to Yahoo! Answers and web passages from documents retrieved using web search. The candidates are ranked by a trained linear model and the top candidate is given as the _nal answer. The ranking model is trained on question and answer (QnA) pairs from Yahoo! Answers archive using Pairwise ranking criterion. Candidates are represented with a set of features, which includes statistics about candidate text, question term matches and retrieval scores, associations between question and candidate text terms and the score returned by a Long Short-Term Memory (LSTM) neural network model.
4. **W. Wei, Z. Ming, L. Nie, G. Li, J. Li, F. Zhu, T. Shang, and C. Luo (2016) have represents Exploring heterogeneous features for query-focused summarization of categorized community answers[1].** The author proposes a three level scheme, which aims to generate a query-focused summary-style answer in terms of two factors, i.e., novelty and redundancy. Specifically, we first get a set of Qas to the given query, and then develop a smoothed Naive Bayes model to identify the topics of answers, by exploiting their associated category information.
5. **X. Li, Y. Ye, and M. K. Ng (2016) have represents Multivcrank with applications to image retrieval [2].** The author propose and developed a multivisual- concept ranking (MultiVCRank) technique for image retrieval. The main idea is that an image can be displayed by several visual concepts, and a hyper graph is built based on visual concepts as hyper edges, where each edge contains images as vertices to share a specific visual concept. In the proposed hyper graph, the weight between two vertices in a hyper edge is incorporated, and it can be calculated by their affinity in the corresponding visual concept. A ranking technique is proposed to compute the association scores of images and the relevance scores of visual concepts by employing input query vectors to handle image retrieval.
6. **W. Wei, G. Cong, C. Miao, F. Zhu, and G. Li (2016) has represents Learning to find topic experts in twitter via different relations[3].** The author developed a probabilistic method to jointly exploit three types of relations (i.e., follower relation, user-list relation, and list-list relation) for finding experts. Specifically, propose a Semi-Supervised Graph-based Ranking approach (SSGR) to offline calculate the global authority of users. In SSGR, employ a normalized Laplacian regularization method to jointly explore the three relations, which is subject to the supervised information derived from Twitter crowds. Then online compute the local relevant between users and the given query. By leveraging the global authority and local relevance of users, here rank all of users and find top-N users with highest ranking scores.
7. **W. Wei, B. Gao, T. Liu, T. Wang, G. Li, and H. Li (2016) has designed A ranking approach on large-scale graph with multidimensional heterogeneous information[4].** The author 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, propose a technique and effective semi-supervised Page Rank (SSP) technique to parameterize the derived information within a unified semi-supervised learning framework (SSLF-GR), then simultaneously optimize the parameters and the ranking scores of graph nodes.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

III. EXISTING SYSTEM AND DISADVANTAGES

In existing system, the question starvation widely exists in cQA system which refers to the following two types Phenomen:

1. Firstly, those looking for information usually have to wait long before getting answers to their questions.
2. Secondly, a large percentage of questions do not any response even within a relatively long period.

Question starvation is probably caused by several Reasons:

1. The questions are badly formulated, Ambiguous or less interesting to all.
2. cqa systems are hardly to route the newly entered questions to the appropriate answerers
3. Potential interviewees have experience, But they are not available or overwhelmed by pure volume of incoming questions. This case often occurs in the vertical cQA forums, so only experts can answer these questions.

Disadvantages:

1. There is no guarantee to that relevancy of answer
2. Time consuming technique.
3. Difficulty to find similar question.
4. Ranking is not proper.

IV. PROPOSED SYSTEM AND ADVANTAGES

The proposed system, construct a novel Pairwise Learning to rANk model i.e PLANE which can quantitatively rank answer candidates from the relevant question pool. Specifically, it comprises two components i.e one offline learning component and one online search component.

- 1) In the offline learning component, we first consequently set up the positive, neutral, and negative training samples in the forms of preference pairs guided by our data-driven results.
- 2) In the online search component, we initially gather a pool of answer candidates for the given question by means of discovering its comparable or similar questions. We at that point sort the appropriate answer candidates by utilizing the offline trained model to judge the preference orders.

Proposed system get question from user then select similar question for entered query by using similarity of available question then apply Pairwise learning that will processed and within time user will get answer and relevancy of answer will be maintained.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

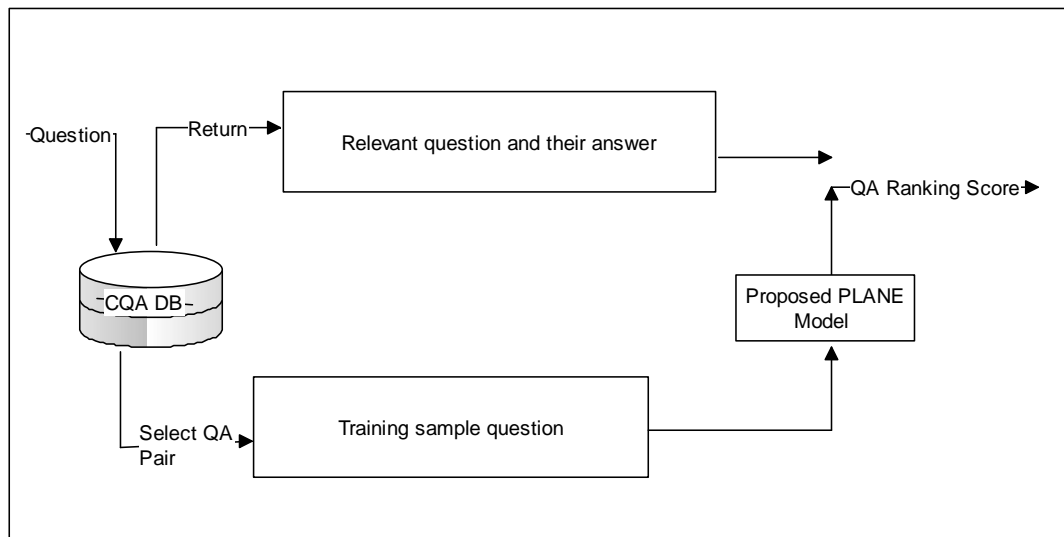


Figure: Proposed System i.e PLANE model

Advantages:

1. Not time consuming process.
2. Achieve better performance.
3. Provide security.
4. Achieve robustness.
5. Make system user friendly.

V. CONCLUSION

We present a novel scheme for answer selection in cqa system. It consists of one online learning and the online search component. In component online learning, instead of timeconsuming and labor-intensive annotation, automatically builds positive, neutral and Negative training samples. In the online search component, A particular question is, first of all, gathering a group of answers find candidates through their similar questions. We then use the offline model to classify candidate answers through Pairwisecomparison. We have conducted extensive experiments to justify The effectiveness of our model in a cqa general data set and a series of vertical cqa data.

REFERENCES

- [1] W. Wei, Z. Ming, L. Nie, G. Li, J. Li, F. Zhu, T. Shang, and C. Luo, "Exploring heterogeneous features for query-focused summarization of categorized community answers," *Inf. Sci.*, vol. 330, pp. 403–423, 2016.
- [2] X. Li, Y. Ye, and M. K. Ng, "Multivcrank with applications to image retrieval," *TIP*, vol. 25, no. 3, pp. 1396–1409, 2016.
- [3] W. Wei, G. Cong, C. Miao, F. Zhu, and G. Li, "Learning to find topic experts in twitter via different relations," *TKDE*, vol. 28, no. 7, pp. 1764–1778, 2016.
- [4] W. Wei, B. Gao, T. Liu, T. Wang, G. Li, and H. Li, "A ranking approach on large-scale graph with multidimensional heterogeneous information," *TOC*, vol. 46, no. 4, pp. 930–944, 2016.
- [5] X. Wei, H. Huang, C. Lin, X. Xin, X. Mao, and S. Wang, "Reranking voting-based answers by discarding user behavior biases," in *Proceedings of IJCAI'15*, 2015, pp. 2380–2386.
- [6] Q. H. Tran, V. Duc, Tran, T. T. Vu, M. L. Nguyen, and S. B. Pham, "Jaist: Combining multiple features for answer selection in community question answering," in *Proceedings of SemEval'15*. ACL, 2015, pp. 215C–219.
- [7] Savenkov, "Ranking answers and web passages for non-factoid question answering: Emory university at TREC liveqa," in *Proceedings of TREC'15*, 2015.
- [8] L. Nie, Y. Zhao, X. Wang, J. Shen, and T. Chua, "Learning to recommend descriptive tags for questions in social forums," *TOIS*, vol. 32, no. 1, p. 5, 2014.
- [9] Z. Ji and B. Wang, "Learning to rank for question routing in community question answering," in *Proceedings of CIKM'13*. ACM, 2013, pp. 2363–2368.
- [10] D. H. Dalip, M. A. Goncalves, M. Cristo, and P. Calado, "Exploiting user feedback to learn to rank answers in qaforums: A case study with stack overflow," in *Proceedings of SIGIR'13*. ACM, 2013, pp. 543–552.