ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

# A Ranking Approach to Extracting Relevant Answer for QA System

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. Each question in the group of candidates returned consists of several answers, here the user has to wait for a long time. To solve this problem, a ranking approach was proposed, a new model of Peer Learning to Learn, that is, rAnk Model, which can significantly classify candidates from the list of relevant questions.Specifically, it uses two components i.e one offline learning element and second one online search element. In the online searching system get a collection of answer candidates for the particular given question by means of discovering its comparable or similar questions by proposed algorithm. System at that point sorts the appropriate answer option by utilizing the offline searching to calculate the 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-**Answer similarities, Community Question Answering, Question-Answer pairs, Pairwise learning technique,PLANE Model.

## I. INTRODUCTION

In the web user, often, the hunger for questions is probably due to several reasons: 1) the questions are poorly written, ambiguous or not at all interesting; 2) cQA systems can hardly address the newly published questions to the appropriate respondents; and 3) potential respondents have the corresponding experience, but are not available or are overwhelmed by the large volume of incoming questions. This case often occurs in vertical cQA forums, whereby only authorized experts can answer these questions. Regarding the first case, the quality model of the application has been well studied, which can assess the quality of the application and serve to remind the respondents to reformulate their questions. Routing applications work by exploring the resources of the current system, in particular human resources. Beyond that, we can reuse the resolved questions from the past to answer new questions. In fact, a large number of historical QA pairs, over time, have been archived in the cQA databases. Therefore, information seekers have a good chance of getting direct answers looking for repositories, instead of waiting long. Inspired by this, Wang et al. They have transformed the quality control task into the task of finding relevant and similar questions. However, candidates returned from the main application are generally associated with multiple answers and research on how to choose the correct answers from the relevant question group are relatively poor. When a question is asked, instead of naively choosing the best answer to the most pertinent question, In this paper, we present a new Pairwise Learning to Run model, dubbed PLANE, which can quantitatively classify candidates from the relevant question group. Figure 1 shows the workflow of the PLANE model, which consists of two components: offline learning and online research.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

**A. Comparision with similar system**

Previous system Developed a hierarchical framework toidentify the predictive factors for obtaining a high quality answer based on textual and non-textual features. Existing system introduced a general classification framework to combine the evidence from different views, including the graph based relationship, content, and usage-based features.

**B. Scope:**

1.The system automatically give relevant answer to user within a time.
2.Will promote other similar user to answer on questions of similar user.
3.Will use in community question answer system.

## II. MOTIVATION

The main motivation is to overcome the problem of to find the similar questions, 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. So i motivate to construct a ranking approach and Pairwise Learning to rank model which can quantitatively rank answer candidates from the relevant question pool.

## III. REVIEW OF LITERATURE

**1.** 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**[5].**
2.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**[6].**

3.The author represents hor w 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 final 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[7].

4.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 gets a set of Qas to the given query, and then develop a smoothed Naive ayes model to identify the topics of answers, by exploiting their associated category information[1].
5.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

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

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[4].

6.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[1].

7.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 an 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[2].

## IV.SYSTEM OVERVIEW

The proposed system, construct a Pairwise Learning model to ranking,model that give relevant answer to each question. Specifically, it comprises two components i.e offline and online search element.

1. In the offline Searching element, we first consequently set up the neutral, positive, and negative training patterns in the forms of desire pairs advised by data-driven collection of results.

2.In the online searching element, system first get group of answers for the given question by means of discovering its comparable or similar questions.

We at that point categories the appropriate answer by theoffline trained model to find the ranking orders. Proposed system get question from user then select similarquestion 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. System recommend other user to answer the newly arrived question that has no available answer in database. That will reduce user waiting time to get answer. For that askers past question are matched with other askers past matched question and then system ask that question to matched askers by email .So it will reduce users waiting time to get result.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

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

*Website: www.ijircce.com*

**Vol. 6, Issue 4, April 2018**

**SYSTEM ARCHITECTURE**

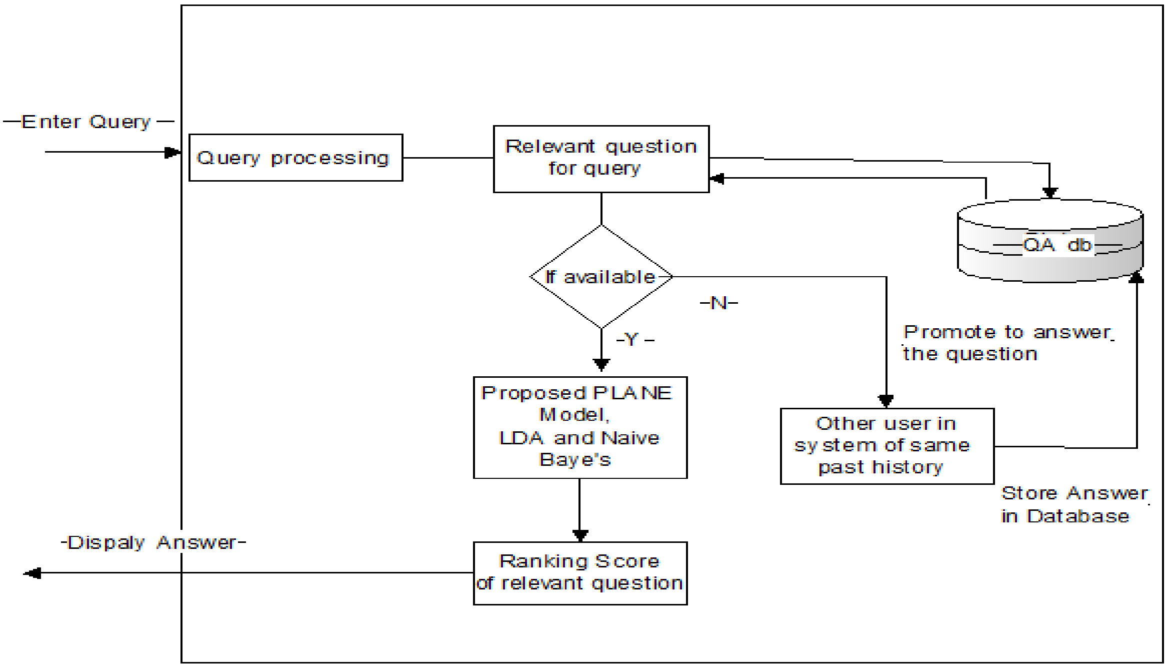


**Fig. 01 System architecture**

## V. MATHEMATICAL MODEL

**Notation**

1.q=Entered question.

2.a1 be the votes of answer

3.C be the class of answer.

4.$a_j^i$ =be the j th answer of i'th question q

5.$a_i^0$ =be the best answer

6.$A_{11}$ =all similar question of q

**Equation:**

$A_{11}$ =avg(feature (all matched question))-------------------------(1)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

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

*Website: www.ijircce.com*

**Vol. 6, Issue 4, April 2018**

Gives similar question of entered questionand Lavenstine Distance Algorithm.

$a_i^0$ =avg(a1,C)---------------------(3)

User gets best answer by applying Naïve Baye's.

Collabrativereccomendation:

$b^0_i$=sim(C1,C2)————————(3)
Eq.3 A similar user wlll find that will give answer for that question has not any answer available in database.It is get by similarity of two users of past asked question preference.

## VI.ALGORITHMS

**Levenshtein distance algorithm.**
The Levenshtein algorithm (also called Edit-Distance) calculates the least number of edit operations that are necessary to modify one string to obtain another string. The most common way of calculating this is by the dynamic programming approach.In proposed system we using this to match user entered question with available question in database.
Input.: Get user entered question.
Working:
Step1.Select user entered question
Step 2:.Select all question from available database
Step3. Pass the distance to match entered question with available question.
System will check question with according to entered question with available question word by word with available answer.
Step4:One by one question will gets by visiting each question to specified distance.
Output:Get matched similar questions.

**Algorithm 2: Naive Bayes**
This algorithm is used classify review is positive or negative that will used find best answer in plane model. This is used to get review is positive or negative according to that we get relevancy inanswer.
Input:Review
Output: Predicated class of review.

Working:
Step 1: Take review
Step 2: Preprocess the review
Step 3: Pass to naive bayes class.
Step 4: Get positive and negative score according tospecify its dictionary.
Step 5:Get max score and declare as positive or negative.
Step 6: Predicated class of all review.

## VII.EXPERIMENT RESULT

**A. Efficiency calculation**
1.Precision:relevant answer-reterived answer=reterivedanswer.
2.Recall:relevant answer-retrived answer=relevant answer.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

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

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

**B. Performance measure used**

1.Proposed system get relevant question using LavenstineDistance Algorithm.Lavenstine Distance Algorithm gives the similarity between two questions.So by using this we can easily find out the relevant questions

2.Vote and review processing is used to get relevant answer in system.

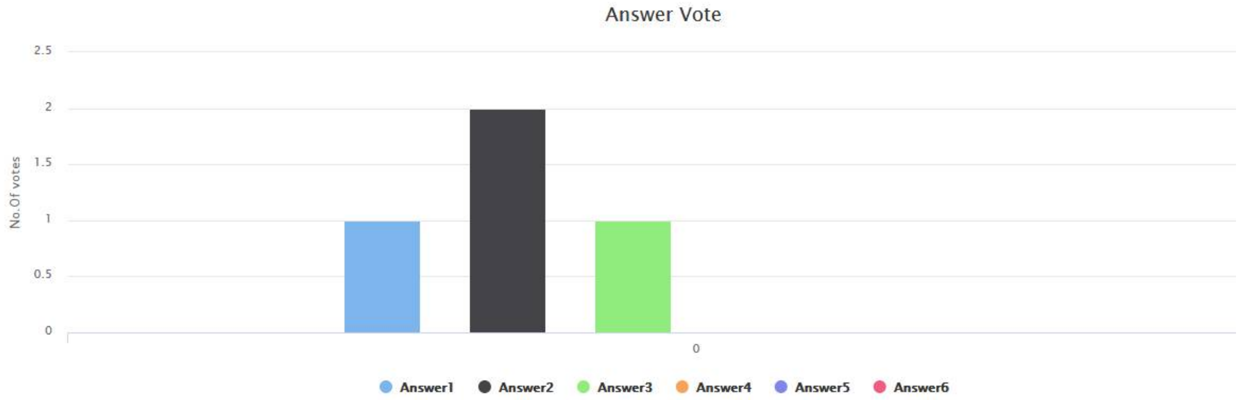


**Fig 02. Shows on X-Axis Que.ID On Y-Axis votes for that answers**

Explanation:Fig 2 the graph shows no. of vote given for each answer by other user.

**Table: 1**

| Ans.ID | No. of votes per question |
|---|---|
| Anwer 1 | 10 |
| Anwer 2 | 20 |
| Anwer 3 | 7 |
| Anwer 4 | 0 |
| Anwer 5 | 0 |

**Table1. shows no. votes for each answer**

## VIII. CONCLUSION

Present a ranking approach for answer selection in cqa system. It consists of one online learning and the online search component. In component online learning, instead of time consuming and labor-intensive annotation, automatically builds positive, neutral and Negative training samples. In the online search component, A particular question is, first of

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

all, gathering a group of answers find candidates through their similar questions. We then use the offline model to classify candidate answers through Pairwise comparison.System recommendother user to answer the newly arrived question that has no available answer in database.That will reduce user waiting time to get answer.

## REFERENCES

[1]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

[2]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.

[3]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.

[4]X. Li, Y. Ye, and M. K. Ng, "Multivcrank with applications to image retrieval," TIP, vol. 25, no. 3, pp. 1396–1409, 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]A Joint Segmentation and Classification Framework for Sentence Level Sentiment Classification Duyu Tang, Bing Qin, Furu Wei, Li Dong, Ting Liu, and Ming Zhou.

[9] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in*Proceedings of ICML'14*. Morgan Kaufmann Publishers Inc., 2014, pp. 1188–1196.

[10] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicitfeedback," in *Proceedings of SIGIR'05*. ACM, 2005, pp. 154–161.