

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

A Novel Approach to Construct Community QA Systems

Omkar Ghule, Manish Joshi, Srishti Patil, Amey Malunjkar

B.E, Department of Computer Engineering, Smt.Kashibai Navale College of Engineering, Pune, India

ABSTRACT: The data-driven approach can yield high levels of performance and nicely complements traditional question answering techniques driven by information extraction. In order for question answering systems to benefit from this vast store of useful knowledge, they must cope with large volumes of useless data. Question Answering systems (QA) uses similarity in questions and ranking the relevant answer to user. The web gives large data and that require more time as well as no relevancy in answers. To solve this problem proposed system proposed 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. Our model is effective as well as achieves better performance than several existing questions answer selection system.

KEYWORDS: Community-based Question Answer, Answer Selection, Pairwise learning technique

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 In fact, a large number of historical QA pairs, over time, have been archived in the cQA databases.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 [12]. Inspired by this, [11] 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 is 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.

II. RELATED WORK

1. In this paper, we propose a joint segmentation and classification framework for sentence-level sentiment classification. It is widely recognized that phrasal information is crucial for sentiment classification. However, existing sentiment classification algorithms typically split a sentence as a word sequence, which does not effectively handle the inconsistent sentiment polarity between a phrase and the words it contains, such as not bad, bad and a great deal of, great. We address this issue by

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

developing a joint framework for sentence-level sentiment classification. It simultaneously generates useful segmentations and predicts sentence-level polarity based on the segmentation results. Specifically, we develop a candidate generation model to produce segmentation candidates of a sentence; a segmentation ranking model to score the usefulness of a segmentation candidate for sentiment classification; and a classification model for predicting the sentiment polarity of a segmentation [8].

2. 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 retrieve 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[9].
3. In generating a vote, a user's attention is influenced by the answer position and appearance, in addition to real 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 solves this problem 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].
 1. 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].
 2. The author represents how to automatically answer questions posted to Yahoo! Answers community question answering website in real-time. This system combines candidates 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 returned as the final answer. The ranking model is trained on question and answer (Qn A) pairs from Yahoo! Answers archive using pair wise 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].

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 [3].

The author addresses 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 effective semi-supervised Page Rank (SSP) technique to parameterize the derived information

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

within a unified semi-supervised learning framework (SSLF-GR), then simultaneously optimize the parameters and the ranking scores of graph nodes[4].

proposes a general framework based on the learning to rank concepts for QR . 2. Uses SVM-based and Rankings- based methods to rank answers. Disadvantages: Ranking problem[11].

propose a learning to rank (L2R) approach for ranking answers in QA forums . 2. The L2R method was trained to learn the answer rating, based on the feedback users give to answers in QA forums.

Disadvantages: Less integrity, data confidentiality[12]. present an approach to answer ranking and show the usefulness of features that explicitly model answer quality. Furthermore, we introduce the idea of leveraging snippets of web search results for query expansion in answer ranking[13].

III. PROPOSED SYSTEM

System presents a novel scheme to rank answer candidates via pair wise comparisons. The system consists of one offline learning component and one online search component. In the offline learning component, we first automatically establish the positive, negative training samples in terms of preference pairs guided by our data-driven Approach. Present a novel model to jointly incorporate these types of training samples. In the online search component, we first collect a set of answer candidates for the given question via finding its similar questions. We then sort the answer candidates by working on the offline trained model to judge the preference orders. Proposed systems that allow a user to ask a question for computer language and receive an answer quickly and relevantly, with sufficient context to validate the answer. Current search engines can return ranked lists of answer with relevancy. System will check voting and review and according to that it will arrange ascending order. To match the feature we using lavenstine distance algorithm.

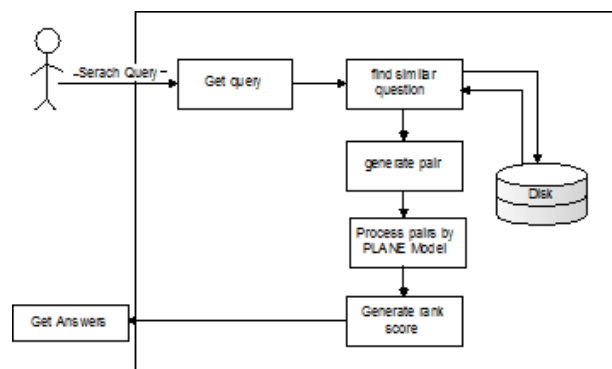


Fig. 1. System Architecture Diagram

A. User

- 1) User_Login();
- 2) User_Registration();
- 3) Enter_query();

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

- 4) Get_answer();
- 5) Give_vote();
- 6) Give_review();

B. Admin

- 1) Admin_login();
- 2) Authorize_user();

IV. PROPOSED METHOD

A. PLANE Model

On user entered query similar questions will find by syn- onym and Lavenstine distance algorithm .Then for each ques- tion, pair will generate and for each pair answer will have vote and review. Model will process a review and give positivity and negativity score of that answer by naive bayes algorithm. Then more vote and more positivity score answers will display to top

Mathematical Model

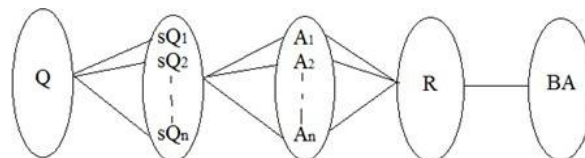


Fig. 2. Venn Diagram

A]Mathematical Equation:

Where,

Q User Entered Question

sQ1, sQ2 to sQn similar questions of user entered Question A1, A2 to An Multiple answers of Similar question

R Ranking of answers

BA Best Answer to user entered Question B] Set Theory

The following terms shows in detail working of project. S is {s, e, X, Y, ϕ }

s = Start of the program. Register to system.

Login to system.

X Input of the program X Q,sQ,R

Where,

Q= User entered Question sQ1,sQ2 to sQn

Similar questions of user entered Question

A1, A2 to An Multiple answers of Similar question Ranking of answers

When user enter question, Find similar questions of entered question.

Establish the samples in terms of preference pairs.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Collect a pool of answer candidates for the given question via finding its similar questions.

Sort the answer candidates by leveraging the offline trained model to judge the preference orders and returns best answer. Y = Output of the program.

$Y = B A$

The output of program is Best answer to User entered Question.

e = End of the program.

ϕ = Failures and Success conditions.

Notation

1. a_j^i = be the jth answer of ith question q

2. a_i^0 = be the best answer

3. q = Entered question.

4. a1 be the votes of answer

5. C is the class of answer.

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

Equation:

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

Gives similar question of entered question.

$$a_i^0 = \text{avg}(a1, C) \text{-----}(2)$$

Best answer by applying naïve bay's e and lda

V. CONCLUSIONS

Present a system for selecting answers in the cqa system. It consists of an online search and online search component. In learning online components, instead of a job that takes a lot of time and effort, it automatically creates positive, neutral, pairs. In the online search component, a particular question is, first of all, gathering a group of answers to find candidates through their similar questions. So system use the offline model to classify candidate responses through pair's comparison

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, 2016.
2. X. Li, Y. Ye, and M. K. Ng, Multivcrank with applications to image retrieval, *TIP*, vol. 25, no. 3, 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, 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, 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.
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.
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.



ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

9. Q. Le and T. Mikolov, Distributed representations of sentences and documents, in Proceedings of ICML'14. Morgan Kaufmann Publishers Inc., 2014,
10. T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, Accurately interpreting clickthrough data as implicit feedback, in Proceedings of SIGIR'05. ACM, 2005,
11. D. H. Dalip, M. A. Goncalves, M. Cristo, and P. Calado, Exploiting user feedback to learn to rank answers in QA forums A case study with stack overflow, in Proceedings of SIGIR,13. ACM, 2013.
12. K. Wang, Z. Ming, and T.-S. Chua, syntactic tree matching approach to finding similar questions in community based QA services, in Proceedings of SIGIR,09. ACM, 2009
13. Felix Hieber, Stefan Riezler Improved Answer Ranking in Social Question-Answering Portals, October 28, 2011, Glasgow, Scotland, UK.