Literature Review on Answer Processing in Community Question Answering System

Saman Qureshi[(✉)] and Sri. Khetwat Saritha

Department of CSE, MANIT, Bhopal 462003, India

Abstract. Community question answering (CQA) websites like Quora, Yahoo!Answers, Reddit enables users to ask questions as well as to answer questions. These sites are online communities that are popular now a days on the internet due to the increase of Question Answering (QA) websites and covers a wide variety of topics. Answer Processing task is classified as the ranking of answers, selection of answer through voting correlation, predicting the answer, selecting an appropriate answer from the candidate answers by classifying answer in good, bad, and potential category and then performing Yes/No task on selected answers or through best answer prediction or best answer selection. The shortcomings in the current approaches are the lexical gap between text pairs, dependency on external sources, and manual features which leads to a lack of generalization ability and to learn the associate patterns among answers. These shortcomings are resolved by already proposed work but they lack generalization ability and their performance is not satisfying. Feature extraction based methods mostly involve manual featurization which are not generalized form, therefore it can be avoided by deep learned feature. Whereas to focus on rich quality answers attention mechanism can be integrated with the neural network.

Keywords: Answer processing · Community question answering · Online community · Ranking · Voting · Answer selection

1 Introduction

CQA systems are a powerful mechanism that expects to give the most reasonable answers in the shortest possible time to the posted questions. Every day a colossal number of new questions posted and to answer these questions CQA systems can control the explicit knowledge or tacit knowledge so that it can be used effectively. Nevertheless, the user's request can be overloaded without appropriate collaboration support, due to which the CQA system would not able to achieve its main goal as askers would not be able to answer in the shortest possible time. And thus, to support the process of question answering, many approaches have been already proposed, pertaining to questions, answers, and users several data analyses and case studies have been conducted so far.

The several steps in the typical workflow of CQA portals are as follows. The asker firstly posts a new question in the CQA system, and then other users answer the question. In the CQA, the necessary data can be planned all the more accurately as the question

can be described in natural language and it does not have to be limited to some basic semantics. Therefore, the appropriate answer can be effectively received. After getting some answers to the question by posting remarks or voting the answers, the asker can choose the most appropriate answer and alternatively can be voted how good the answer is by other users.

The community question answering system has three stages: Question Processing, Document processing, and the last stage answer processing. Every stage involves a few steps. Parsing and classifying a question and reformulating query comes under question processing stage whereas document processing will find candidate documents and perform answer identification. And the last stage that is answer processing rank the best one answer or selects the best answer from a candidate answers after extraction. The proposed methods are based on patterns, statistical, and features. The workflow of a community question answering system is depicted in Fig. 1.

This paper is structured as follows. Section 2 states related work on answer processing phase in a CQA system. Then Sect. 3 discusses Answer Selection in detail. While Sect. 4 contains conclusion and future work.
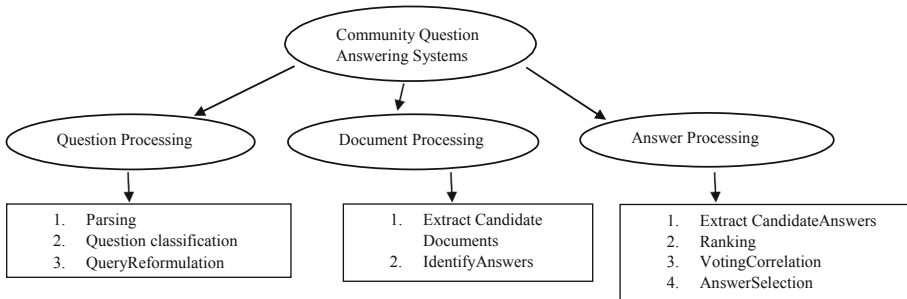
Fig. 1. Workflow of CQA system.

2 Related Work

2.1 Answer Processing

Answer Processing is the final stage of the question answering system where answer extraction is done. It is the most challenging task in CQA systems. When a user posts any question in a community site the answer to the question is given by other users. There can be more than one answer to a question all these answers in an all is called as candidate answers. The main task in answer processing is to select the right and related one answers to a question from these bunch of answers called as candidate answer. The work on answer processing tasks includes semantic similarities between a question and an answer, an answer which is more similar to a question is extracted or through voting correlation or ranking an answer or an answer selection. By these methods, answer processing is done in the CQA system.

2.1.1 Answer Processing Through Voting Correlation

The usability of Community Question Answering (CQA) facilitates the lives of users greatly as day by day its popularity is increasing, where ideas are exchanged and people seek help on the internet. Apart from asking and answering questions, users can provide feedback to these questions/answers through voting or commenting. Like in Stack Overflow forum programmer upload their programming questions and other programmers can give an answer for those questions and then that answer is validated by feedback from others. Such forums are used by millions of programmers when they encounter any kind of programming problems [1]. How to clear the doubts of users by detecting the correct answer? Can a good answer attract for a question? These questions are answered by [2] by voting correlation. [2] correlates the voting score of the answer with its question, and verifies such correlation in two data set that in turn boost the prediction performance. The voting score of a question or answers is characterized as the distinction of the absolute number of upvotes and the total number of downvotes. This voting score acts as an indicator of the intrinsic value of a question or answer.

Other related work is on the measurement of questions and answers by focusing on the quality of question/answer posts [3] in which human annotators label the quality of posts manually. [4] and [5] are proposed frameworks that determine the answer quality. The reiteration of a question is characterized under the estimation of question utility [6]. The methods by the authors: Jeon et al. [5], Suryanto et al.[4], Li et al. [7], Agichtein et al. [8] and Bian et al. [9] are some of the prediction methods for measurements. In the software forums a single question can have more than one answer and to find relevant answers Gottipati et al. [10] focuses on it.

A chunk of co-prediction algorithms is proposed by [2] where the high-impact question is acknowledged by the users in CQA sites through early detection of rich-quality questions/answers. Also to classify a useful answer that can avail positive feedback from users. This paper conjecture two things, one is that an interesting question can get more attention to receive high-score answers from potential answerers and whereas it might be very difficult for a low score question having weak expression in language, or lack of interesting topic to attract high-score answers. Mathematics Stack Exchange and Stack Overflow are the two real CQA sites that are studied for these conjectures. Armed with this verified correlation, the proposed method aims to identify the high-score potentially as soon as it is posted on the CQA sites. The contextual features which is focused are questioners'/answerers' reputation, the number of past questions/answers, length of the body, and title of a question or an answer. These features are extracted at every one hour whenever a question or answer is posted. Over the best contender, this joint forecast strategy accomplishes up to 15.2% net precision improvement and it allows to predict the result of voting for an answer before it appears on site. The effect of question/answer content on its dynamic and correlation is not covered by any proposed methods.

2.1.2 Answer Processing by Answer Ranking

Answer processing task can be considered an answer ranking task. Zhenlei Yan et al. [11] states the problem of the CQA system that many new questions are not able to be solved effectively by a suitable answerer. To resolve this routing task Zhenlei Yan et al.

rank the potential answerers to solve the question by their ability. A novel approach is proposed which simultaneously captures latent semantic relations among question, asker, and answerer by concatenating tensor model and topic model. A new learning procedure is proposed with tensor factorization which optimizes asker-topic-answerer model to execute the optimal answerer ranking task by maximizing multi-class AUC (Area under the ROC Curve). With two real world datasets from Tencent Wenwen (TW) and Yahoo! Answers (YA) this approach outperforms other related approaches.

The two features of new community systems are an ask-reply mechanism and social relations. Due to this researcher's concerns have shifted towards seeking potential answerers from finding existing answers. HAN Wenwen et al. [12] propose a hybrid method to address this problem. The framework considers the user's activity, social status, and authority by partitioning it into three parts question-user network, social graph, and ranking model using an optimized PageRank algorithm.

WikiAnswers, Yahoo! Answers, Baidu Zhida, are some Community web sites where users post a question and the answer to this question is answered manually by other users or it can also be answered automatically from existing community question answer knowledge base. These types of community sites have the CQA knowledge base which consists of question-answer pairs on a large scale. Question retrieval and answer ranking are the two main tasks in this domain. The former task estimates the semantic similarity between question-question pairs to detect similar questions whereas the later one task check the answer responses and rank them on the basis of semantic relatedness between question-answer pairs.

By identifying the major context of the question and some forms of question topic [13] performs the question retrieval task. The author in ref. [14] solve the word mismatch and word ambiguity problems in question by proposing a statistical machine-translation method where other languages are considered to get semantic information between question-question. Whereas for question-answer pairing [15] and [16] authors represent semantic relatedness between question and answer by constructing tree edit models. Considering answer selection task as answering ranking in ref. [17] the author calculate the semantic distance between the question and answer pairs using topic models to rank answers whereas Xiaobing Xue et al. [18] and Zhou et al. [19] uses translation and syntactic based approach. Many cases of semantic similarities are still not captured by these methods and this gap is covered by the authors in ref. [20] and [21] through Convolution Neural Network and Long short term memory deep learning models. The author in ref. [20] works on a question-question pairing task where it uses Ask Ubuntu data which is a part of the StackExchange community and improves accuracy by performing word embedding on different sizes with CNN. An LSTM model is used by the author in ref. [21] for question-answer pairing which sequentially reads words and gives relevance scores to rank answer. A part from these works [23] integrates the two tasks and both are considered as ranking tasks to improve the accuracy of CQA. Two ranking strategies: one is learning-to-rank with ref. to [22] where pairwise training is done and its output is used directly as a ranking score. And second, one train Support Vector Machine and Logistic regression supervised classification model and the probability of confidence score is used as a ranking score. SemEval CQA dataset is used and 45.12% of

MRR value achieved in answering tasks with the help question-question pairing. While propagating from question retrieval to answer ranking, this method reduces errors also.

2.1.3 Answer Selection

In the CQA system answer processing is a critical phase to extract the best answer in a less amount of time. The main problem in a community site is that when a question is posted a bunch of answers is given by users and in these answers, many are not so associated to the question asked and, in certain answers, even shift the topic to the context to a different subject as an example in Fig 1. This issue definition is nowadays considered as this resolves the criticality of answer processing (Fig. 2).
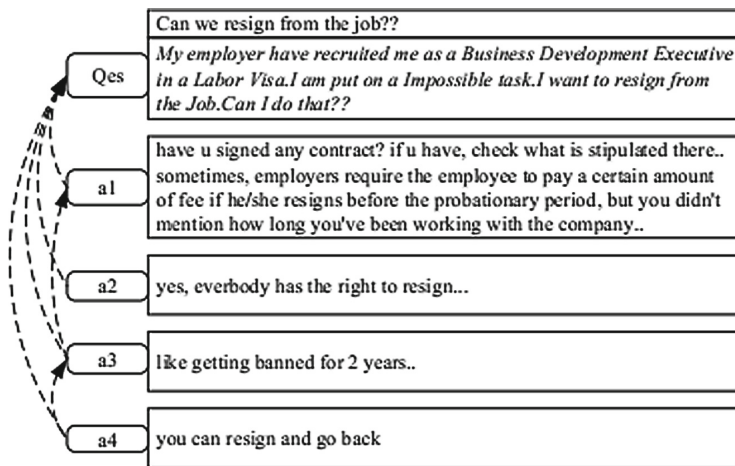

Fig. 2. An example of answers to a question [35].

3 Summary of Answer Selection Based Answer Processing Approaches

Yangsen Zhang et al. [25] removes dependencies of outer assets and manual features as they lack the generalization ability in most cases. These shortcomings can make up by deep learning architecture to catch the semantic data in texts with the utilization of word vector. The two models BLSTM and attention mechanism based on BLSTM is constructed to calculate semantic similarity. InsuranceQA dataset is used to evaluate the proposed approach. The answer with high semantic similarity is selected and accuracy QA-BLSTM achieve is 66.9% whereas QA-Attention Mechanism achieves 68.1%. Baseline models like QA-CNN-S [21], QA- CNN-GESD [21], QA-BLSTM-S [26] and QA-BLSTM-S-A [26] are compared with the two models which prove that BLSTM performs better than CNN as the former one capture a high measure of semantic data from a question and its candidate answers than the later one.

Yin et al. [27], did the comparative study of RNN and CNN. They did the comparison between LSTM and GRU in which they have found out that LSTM is good at modeling the sequence units in long text whereas CNN has an advantage in the short text by extracting invariant features.

Taihua Shao et al. [28] proposed the collaborative learning for answer selection which resolves the drawback of using a single deep neural network that fails to extract the rich sentence features. [28] build a parallel architecture by combining more than one neural network to collaboratively learn there presentations of question and answer. Firstly QA-CL model is built by deploying CNN with BiLSTM which will combine learn word vector matrix of question and answer parallelly. Then, the QA-CL is extended to a hybrid collaborative QA-CLWR model which uses baseline weight removal (WR) to combine the generated sentence embedding with a joint distributed sentence representation. This experiment is conducted on the InsuranceQA dataset. The proposed models are compared with a non-neural network QA-WR [29] model, QA- CNN [30] model, and QA-LSTM/CNN [26] a hybrid model and shows a better performance against them. By achieving the accuracy of 61.22% the experiment performs better only with a medium number of questions as compared to a too small or too large number of questions. Table 1 compare the proposed methods on an InsuranceQA dataset.

Table 1. Result of different methods with InsuaranceQA dataset

Answer processing methods	Accuracy (%)
QA-CLWR [28]	61.22
QA-CL [28]	60.17
QA-LSTM/CNN [26]	59.33
QA-CNN [30]	57.33
QA-WR [29]	36.83
QA-CNN-S [21]	62.8
QA-CNN-S-GESD [21]	65.3
QA-BLSTM-S [26]	63.1
QA-BLSTM-S-A [26]	68.1
QA-BLSTM [25]	67.7
QA-Attention Mechanism [25]	68.5

3.1   Semantic Evaluation (SemEval)-2015 Task3

Semantic Evaluation is a progressing arrangement of assessments to evaluate semantic analysis system, where semantic analysis means analysis of meaning that is the nature of meaning in language is explored. Before SemEval Task 3, the proposed methods are on different independent datasets and to compare these methods results is a complex

task. Therefore, the common framework is provided by Task 3 of SemEval to compare different methods in multiple languages.

The task 3 in SemEval-2015 is related to answer selection in CQA. The feature of the task is a semantic similarity, natural language inference, and textual entailment. This task is initiated to automate the process of identifying the correct answer from the answer thread by classifying the answers as good, bad, and potential and producing all the valid answers by summarizing them as YES/NO.

To identify answer quality, JAIST [31] works on only Task A for English by extracting 16 features which belong to 5 groups (special component features, topic-modeling-based features, word-matching features, translation based features, and non-textual features). The system although achieves high results with 72.52% accuracy and holds rank one but due to heavy dependency on the bag-of-word the potential class is not handled properly.

A hierarchical classification method and a multi-classifier method are proposed by HITSZ-ICRC [32] team for English subtask A, English subtask B, and Arabic task. Two-level hierarchical classification and ensemble learning are proposed to classify answers for all three tasks English subtask A, English subtask B, and Arabic task. Fatwa dataset is used for Arabic task. Three submissions (primary, contrastive1, contrastive 2) were submitted for all three tasks. The Accuracies of English subtask A, English subtask B, and Arabic task is 68.87%, 64%, and 74.53% respectively, and holds the second rank.

QCRI [33], this team also works on the three tasks as HITSZ-ICRC works. In the Arabic task, this team holds the first rank and in the English subtasks the third rank. A supervised Machine learning approach is used considering numerous features i.e. text similarity, the context of a comment, sentiment analysis, word n-grams, and the presence of specific words. For Arabic task logistic regression is used and linear SVM is used for English subtask A. The team has also conducted a Post Experiment without and only a feature to understand the different features performance. The F1 score of Arabic task, English subtask A, and English subtask B is 78.55, 53.74, and 53.60 respectively.

ICRC-HIT [34] proposed a deep learning strategy and present a comment labeling system. To recognize a good comment, a recurrent convolution neural network is used.

The answer selection by Hongjie Fan et al. [35] is done using a multi-dimensional feature combination method. From every question and comment in the dataset, the information is extracted. The total 20 features were extricated dependent on the content description, text similarity, and attribute description. Using the SVM. Gradient Boosting Decision Tree (GBDT) and random forest, a model is built from the extracted features to classify dimensions obtained. Then an experiment is conducted which shows that the three methodologies are more effective than baseline models, and when contrasted with other proposed methods, relatively its ranking is on an all high. The selection of super-parameter of the model is randomly done which are not fine-grained and only 20 features were selected. But despite these limitations, the models ranking is high as compared to others. Different proposed methods for this task are stated in Table 2 and Table 3 for task A and task B respectively with their achieved accuracy.

3.2   Answer Selection by Predicting Best Answer

The objective of Question answering communities is to allow users to share knowledge by means of asking questions or by answering the questions asked by some other user.

Table 2. Result of methods for SemEval Task A

Answer processing methods	Accuracy (%)
JAIST [31]	72.52
HITSZ-ICRC [32]	68.67
QCRI [33]	70.50
ECNU [51]	70.55
ICRC-HIT [34]	67.68
GBDT [35]	68.12
Random Forest [35]	65.89
SVM [35]	43.35
Baseline	50.46

Table 3. Result of methods for SemEval Task B

Answer processing methods	Accuracy (%)
VectorSlu [52]	72.0
GBDT [35]	65.59
ECNU [51]	68.0
QCRI [33]	64.0
HITSZ-ICRC [32]	64.0
Random Forest [35]	65.52
SVM [35]	58.63
Baseline	60

Due to the large flow of information and lots of facilities communities' sites are being widely used nowadays. One of the issues in the answer processing task is to foresee the most fitting answer as not every asker has the capacity or information to choose the most fitting solution for his question.

Dalia Elalfy et al. [37] gives a model based on content feature to select the best answer by prediction method. The learning of the model is based on labeled data and it uses three type of features (1). Answer-answer feature, (2) question-answer features and (3) answer content features. Opposite to this model the [38] model is based on non-content feature where popularity score of the user who is responding to question in the stack overflow portal rather than Yahoo! Answer is measured. Merging these two proposed models with enhancement a hybrid model is build by [36] which consist of 3 different classifiers (Logistic Regression, Random Forest, and Naïve Bayes) to predict the most appropriate answer using some newly added features. The prediction results increase in the hybrid model as compared to the other two models and the accuracy is very promising.

To find autonomously the best answer in CQA services is an essential step. To validate a post voting up and voting down is done by users. The extraction of features is the main challenge while automating the selection of the best answer. Usually, the features are extracted from questions, answer, and metadata. Gkotsis et al. in [40] include comments for each answers as one of the features whereas the variance and average of comments are considered as the main feature by Tiametal.in [41]. [39] considered comments as a feature where text mining technique that is sentiment analysis is applied and answers spell checking is done. The social behavior of users and their activities are considered as informative features. Four big stack exchange websites (Math.SE, English.SE, Ask Ubuntu.SE, and skeptic.SE) from one of the biggest English CQA stack exchanges are considered to verify the work. The model uses 23 features which are selected from three categories Question and answer, comments, and user behavior. The performance of the model is tested on decision tree classifiers (like Adaboost) and some Alternate Decision Trees (ADT) classifier using Weka10. Evaluation of the model is done using F-measure

with a 10-fold cross-validation method. Results show improvement in performance as compared to other models by finding the best blend of different features.

3.3   Answer Selection by Selecting Best Answer

The expansion in utilization of CQA sites within incalculable questions and their relating answers increases the size of contents in this site. Traditionally best answer selection is done manually for the question asked, which is monotonous as to examine such semi-organized and colossal textual contents alongside the associate post score. To automate the selection of answers [42] proposed a model which instead of taking only question-answer related data it takes both answerers and question-answer data into account. This work analyses Stack Overflow Q&A posts, hence the Stack Overflow dataset is used. Based on activity signatures [43–45], domain knowledge [46], and topical similarity [47] the active answerers are identified to the asked questions. Also, topic modeling, topical interest, topical expertise [48], and voting scores are used. Then the relationship between Q&A pairs is found through topic relevance like[47].At last to predict the best answer to the question asked at least five answers of Q&A posts are analyzed to focus on features involved as in [49] and [50] for pattern identification based on topic modeling and classifier. The results are evaluated with Precision-Recall Area Under Curve, Accuracy, Receiver Operating Characteristics Area under Curve, and Accuracy. The accuracy of the two classifiers (Bayes Net and Naïve Bayes) is calculated where Bayes Net outperform Naive Bayes by achieving an overall 69%. The calculation of expertise level and potential experts cannot be done with this model and pre-processing can affect the performance parameter for other CQA sites due to different meta data arrangements.

4   Conclusion and Future Work

Community Question answering websites consist of three phases: question phase, document or passage retrieval phase, and the last one answer processing phase. The answer processing is the challenging one task in Question Answering websites. The selection of the right from candidate answers for a question is the problem stated by CQA systems. The framework or method proposed for this problem is based on pattern matching, static-based, and feature-based. Giving upvote or downvote to an answer is allowed by many community sites and through voting correlation answer extraction is done. And the other ways are ranking the answer, or predicting the answer or answers election to process an answer. Challenges faced by CQA while answer processing is the lexical gap between question and question and a lexical gap between questions and answers and also a deviation from a question. These challenges are covered by proposed frameworks and methods but still their performance lack generalization ability and still, its accuracy can be improved more. Due to the use of external semantic resources and manual features, the generalization of the framework is not achievable and its performance is still can be improved. The probable solutions can be using deep learned feature instead of manual features, the lexical gap can be bridged by deep learning method as it can avoid feature engineering. And also to focus on high quality answers attention mechanism can be integrated with a neural network.

References

1. Osbourn, T.: Getting the most out of the web. Softw. IEEE **28**(1), 96 (2011)
2. Yao, Y., Tong, H., Xie, T., Akoglu, L., Xu, F., Lu, J.: Detecting high-quality posts in community question answering sites. Inf. Sci. **302**, 70–82 (2015). https://doi.org/10.1016/j.ins.2014.12.038, ISSN 0020-0255
3. Harper, F.M., Raban, D., Rafaeli, S., Konstan, J.A.: Predictors of answer quality in online Q&A sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2008), pp. 865–874. Association for Computing Machinery, New York (2008). https://doi.org/10.1145/1357054.1357191
4. Suryanto, M.A., Lim, E.P., Sun, A., Chiang, R.H.: Quality- aware collaborative question answering: methods and evaluation. In: WSDM 2009: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pp. 142–151. Research Collection School of Information Systems (2009)
5. Jeon, J., Croft, W.B., Lee, J.H., Park, S.: A framework to predict the quality of answers with non-textual features. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006), pp. 228–235. Association for Computing Machinery, New York (2006). https://doi.org/10.1145/1148170.1148212
6. Song, Y.I., Lin, C.Y., Cao, Y., Rim, H.-C.: Question utility: a novel static ranking of question search. In: Proceedings of the National Conference on Artificial Intelligence, vol. 2, pp.1231–1236 (2008)
7. Li, B., Jin, T., Lyu, M.R., King, I., Mak, B.: Analyzing and predicting question quality in community question answering services. In: Proceedings of the 21st International Conference on World Wide Web (WWW 2012 Companion), pp. 775–782. Association for Computing Machinery, New York (2012). https://doi.org/10.1145/2187980.2188200.
8. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM 2008), pp. 183–194. Association for Computing Machinery, New York (2008). https://doi.org/10.1145/1341531.1341557
9. Bian, J., Liu, Y., Zhou, D., Agichtein, E., Zha, H.: Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In: Proceedings of the 18th International Conference on World wide web (WWW 2009), pp. 51–60. Association for Computing Machinery, New York (2009). https://doi.org/10.1145/1526709.1526717
10. Gottipati, S., Lo, D., Jiang, J.: Finding relevant answers in software forums. In: Proceedings of the 2011 26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011), pp. 323–332. IEEE Computer Society, USA (2011). https://doi.org/10.1109/ASE.2011.6100069
11. Yan, Z., Zhou, J.: Optimal answerer ranking for new questions in community question answering. Inf. Process. Manage. **51**(1), 163–178 (2015). https://doi.org/10.1016/j.ipm.2014.07.009, ISSN 0306-4573
12. Wenwen, H., Xirong, Q., Siqi, S., Ye, T., Wendong, W.: Ranking potential reply-providers in community question answering system. China Commun. **10**(10), 125–136 (2013). https://doi.org/10.1109/CC.2013.6650325
13. Duan, H., Cao, Y., Lin, C.Y., Yu, Y.: Searching Questions by Identifying Question Topic and Question Focus, pp. 156–164 (2008)
14. Zhou, G., Liu, F., Liu, Y., He, S., Zhao, J.: Statistical machine translation improves question retrieval in community question answering via matrix factorization. In: Proceedings of the Conference on ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, pp. 1.852–861 (2013)

15. Heilman, M. and Smith, N.A.: Tree edit models for recognizing textual entailments, para-phrases, and answers to questions. In: Human Language Technologies: the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT 2010), pp. 1011–1019. Association for Computational Linguistics, USA (2010).

16. Chang, M.W., Goldwasser, D., Roth, D., Srikumar, V.: Discriminative learning over con-strained latent representations. In: Human Language Technologies: the 2010 Annual Confer-ence of the North American Chapter of the Association for Computational Linguistics (HLT 2010), pp. 429–437. Association for Computational Linguistics, USA (2010)

17. Qin, Z., Thint, M., Huang, Z.: Ranking answers by hierarchical topic models. In: Chien, B.C., Hong, T,P., Chen, S,M., Ali, M. (eds.) Next-Generation Applied Intelligence. IEA/AIE 2009. Lecture Notes in Computer Science, vol. 5579. Springer, Heidelberg (2009)

18. Xue, X., Jeon, J., Croft, W.B.: Retrieval models for question and answer archives. In: Proceed-ings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008), pp. 475–482. Association for Computing Machinery, New York (2008). https://doi.org/10.1145/1390334.1390416

19. Zhou, G., Cai, L., Zhao, J., Liu, K.:Phrase-Based Translation Model for Question Retrieval in Community Question Answer Archives, pp. 653–662 (2011)

20. Bogdanova, D., dos Santos, C., Barbosa, L., Zadrozny, B.: Detecting Semantically Equivalent Questions in Online User Forums, pp. 123–131 (2015). https://doi.org/10.18653/v1/K15-1013

21. Wang, D., Nyberg, E.: A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering, pp. 707–712 (2015). https://doi.org/10.3115/v1/P15-2116

22. Trotman, A.: Learning to rank. Inf. Retr. **8**, 359–381 (2005). https://doi.org/10.1007/s10791-005-6991-7

23. Lan, M., Wu, G., Xiao, C., Wu, Y., Wu, J.: Building mutually beneficial relationships between question retrieval and answer ranking to improve performance of community question answer-ing. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 832–839. Vancouver (2016). https://doi.org/10.1109/IJCNN.2016.7727286

24. Zhou, X., Hu, B., Chen, Q., Wang, X.: Recurrent convolutional neural network for answer selection in community question answering. Neurocomputing **274**, 8–18 (2018). https://doi.org/10.1016/j.neucom.2016.07.082. ISSN 0925-2312

25. Zhang, Y., Peng, Y.: Research on Answer Selection Based on LSTM, pp. 357–361 (2018). https://doi.org/10.1109/IALP.2018.8629166

26. Tan, M., Santos, C.D., Xiang, B., Zhou, B.: LSTM-based Deep Learning Models for Non-factoid Answer Selection (2015)

27. Yin, W., Kann, K., Yu, M., Schütze, H.: Comparative study of CNN and RNN for natural language processing (2017). https://arxiv.org/abs/1702.01923

28. Shao, T., Kui, X., Zhang, P., Chen, H.: Collaborative learning for answer selection in question answering. IEEE Access **7**, 7337–7347 (2019). https://doi.org/10.1109/ACCESS.2018.2890102

29. Arora, S., Liang, Y., Ma, T.: A simple but tough-to-beat baseline for sentence embeddings. In: Proceedings of International Conference on Learning Representations, pp. 1–16 (2016)

30. Feng, M., Xiang, B., Glass, M.R., Wang, L., Zhou, B.: Applying deep learning to answer selec-tion: a study and an open task. In: 2015 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 813–820, December 2015

31. Tran, Q.H., Tran, D.V., Vu, T., Le Nguyen, M., Pham, S.B.: JAIST: Combining multiple fea-tures for Answer Selection in Community Question Answering. pp. 215–219 (2015). https://doi.org/10.18653/v1/S15-2038

32. Hou, Y., Tan, C., Wang, X., Zhang, Y., Xu, J., Chen, Q.: HITSZ-ICRC: Exploiting Classi-fication Approach for Answer Selection in Community Question Answering, pp. 196–202 (2015). https://doi.org/10.18653/v1/S15-2035

33. Nicosia, M., Filice, S., Barrón-Cedeno, A., Saleh, I., Mubarak, H., Gao, W., Nakov, P., Martino, G.D.S., Moschitti, A., Darwish, K., Marquz, L.M.: QCRI: Answer Selection for Community Question Answering – Experiments for Arabic and English (2015). https://doi.org/10.18653/v1/S15-2036

34. Zhou, X., Hu, B., Lin, J., Xiang, Y., Wang, X.: ICRC-HIT: A Deep Learning based Comment Sequence Labeling System for Answer Selection Challenge, pp. 210–214 (2015). https://doi.org/10.18653/v1/S15-2037

35. Fan, H., Ma, Z., Li, H., Wang, D., Liu, J.: Enhanced answer selection in CQA using multi-dimensional features combination. Tsinghua Sci. Technol. **24**, 346–359 (2019). https://doi.org/10.26599/TST.2018.9010050

36. Elalfy, D., Gad, W., Ismail, R.: A hybrid model to predict best answers in question answering communities. Egypt. Inform. J. **19**(1), 21–31 (2018). https://doi.org/10.1016/j.eij.2017.06.002. ISSN 1110-8665

37. Tian, Q., Zhang, P., Li, B.: Towards predicting the best answers in community-based question-answering services. In: Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013, pp. 725–728 (2013)

38. Lin, C., Richi, N.: Leveraging the network information for evaluating answer quality in a collaborative question answering portal. Soc. Netw. Anal. Min. **2**(3), 197–215 (2012)

39. Eskandari, F., Shayestehmanesh, H., Hashemi, S.: Predicting best answer using sentiment analysis in community question answering systems. In: 2015 Signal Processing and Intelligent Systems Conference (SPIS), Tehran, pp. 53–57 (2015). https://doi.org/10.1109/SPIS.2015.7422311

40. Gkotsis, G., Stepanyan, K., Pedrinaci, C., Domingue, J., Liakata, M.: It's all in the content: state of the art best answer prediction based on discretisation of shallow linguistic features. In: Proceedings of the 2014 ACM Conference on Web Science, pp. 202–210 (2014)

41. Tian, Q., Zhang, P., Li, B.: Towards predicting the best answers in community-based question-answering services. In: ICWSM (2013)

42. Sahu, T.P., Nagwani, N.K., Verma, S.: Selecting best answer: an empirical analysis on community question answering sites. IEEE Access **4**, 4797–4808 (2016). https://doi.org/10.1109/ACCESS.2016.2600622

43. Anusha, J., Rekha, V.S., Sivakumar, P.B.: A machine learning approach to cluster the users of stack overflow forum. In: Suresh, P.L., Sekhar, D.S.,Ketan, P.B. (eds.) Artificial Intelligence and Evolutionary Algorithms in Engineering Systems: Proceedings of ICAEES 2014, vol. 2, pp.411–418. Springer, India, New Delhi (2015)

44. Sinha, V.S., Mani, S., Gupta, M.: Exploring activeness of users in QA forums. In: Proceedings of 10th Workshop Conference Mining Software Repositories, pp. 77–80 (2013)

45. Bouguessa, M., Dumoulin, B., Wang, S.: Identifying authoritative actors in question-answering forums: the case of Yahoo! answer. In: Proceedings of 14th ACM SIGKDD International Conference Knowledge Discovery Data Mining, pp. 866–874 (2008)

46. Yang, B., Manandhar, S.: Tag-based expert recommendation in community question answering. In: Proceedings IEEE/ACM International Conference Advance Social Network Analysis Mining (ASONAM), pp. 960–963, August 2014

47. Barua, A., Thomas, S.W., Hassan, A.E.: What are developers talking about? An analysis of topics and trends in stack over_ow. Empirical Softw. Eng. **19**(3), 619_654 (2014)

48. Yang, L., et al.: CQArank: jointly model topics and expertise in community question answering. In: Proceedings 22nd ACM International Conference on Information & Knowledge Management, pp. 99–108 (2013)

49. Tian, Q., Zhang, P., Li, B.: Towards predicting the best answers in community-based question answering service. In: Proceedings of ICWSM, pp. 725–728 (2013)

50. Shah, C., Pomerantz, J.: Evaluating and predicting answer quality in community QA. In: Proceedings of 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 411–418 (2010)
51. Zhao, J., Zhu, T.T., Lan, M.: ECNU: One Stone Two Birds: Ensemble of Heterogenous Measures for Semantic Relatedness and Textual Entailment, pp. 271–277 (2014). https://doi.org/10.3115/v1/S14-2044
52. Belinkov, Y., Mohtarami, M., Cyphers, S., Glass, J.: VectorSLU: a continuous word vector approach to answer selection in community question answering systems. In: Proceedings of the 9th International Workshop on Semantic Evaluation, Denver, CO, USA, pp. 282–287 (2015)