International Journal of Management, IT & Engineering
Vol. 8 Issue 10, October 2018,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

PREDICTION OF ACCURATE AND EXPERT ANSWER SELECTION IN COMMUNITY QUESTION ANSWER SYSTEM

Mrs.Veera S. Raval*
Mr. Chetan J. Awati**
Ms.Sonam Kharade***

Abstract

Question noting framework is further catalyst for the clients this days, clients ask the inquiry on the web and accordingly they will get the react of that inquiry, however as perusing is prime require for every one an individual, the measure of clients request question and framework will invest with answer yet the calculation time enlarged and in addition holding up time intensified and same kind of inquiries are asked by various clients, framework need to give comparative answers continually to various clients. To avoid this we propose 'PLANE' system which may quantitatively review react applicants from the huge inquiry pool. On the off chance that clients ask a few questions, at that point framework offer answers for that inquiry in positioning structure, at that point framework propose most elevated review reply to the client. We proposing ERS (Expert Recommendation System), which will give reply of the inquiry which is asked through the client and we in addition set in motion sentence level grouping strategy in which a lone inquiry have amiable answers. The framework gives essentially fitting response to the inquiry which is asked by the client.

Keywords:

Community- Based question answering;
Answer selection;
Expert answer;
similarity score;

* **Mrs. Veera S. Raval,Department Of Technology, SU,Kolhapur, India**
** **Mr. Chetan J. Awati,Co- ordinator : CST,Department Of Technology, SU,Kolhapur, India**
*** **Ms.Sonam Kharade, *DYPCOE*, Pune, India**

1. Introduction

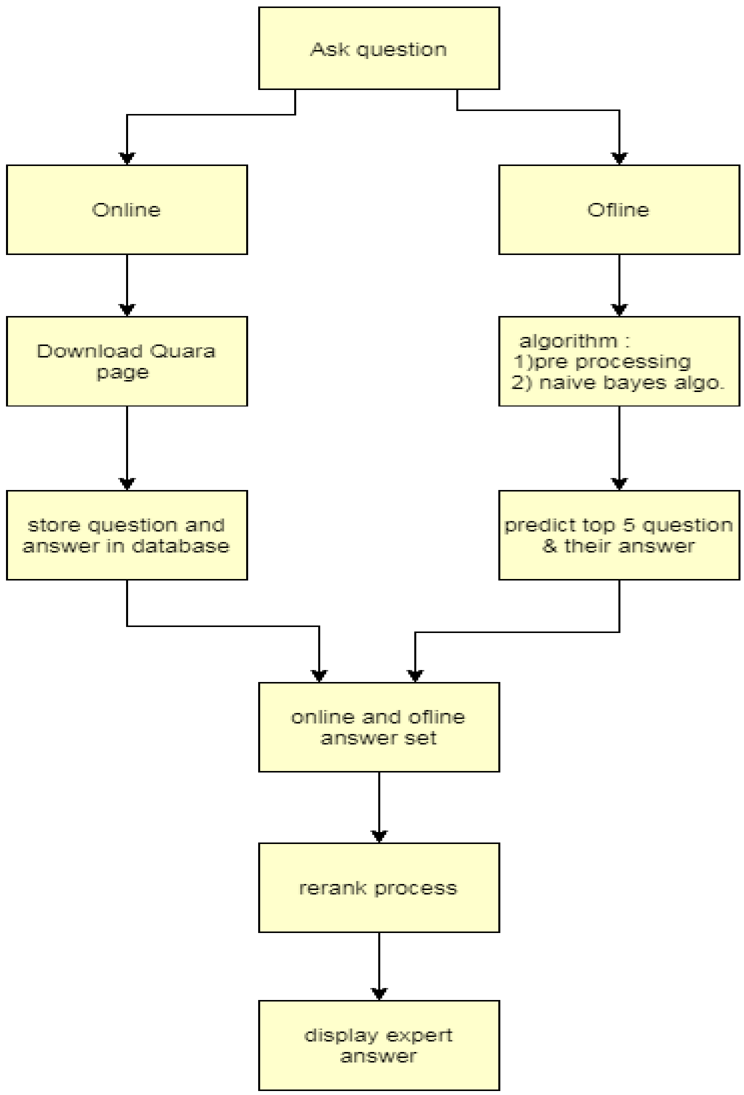
Network Question Answering (CQA) discussions are picking up prevalence on the web. They are rarely directed, rather open, and therefore they have couple of limitations, assuming any, on who can post and who can answer an inquiry. On the positive side, this implies one can uninhibitedly make any inquiry and expect some great, fair answers. On the negative side, it requires push to experience every single conceivable answer and to comprehend them. For instance, it isn't unordinary for an inquiry to have many answers, which makes it extremely tedious to the client to investigate and to winnow. The test we propose may help robotize the way toward discovering clever responses to new inquiries in a network made exchange discussion (e.g., by recovering comparable inquiries in the gathering and distinguishing the posts in the appropriate response strings of those inquiries that answer the inquiry well).

Question replying (QA) is a software engineering discipline inside the fields of data recovery and common dialect preparing (NLP), which is worried about building frameworks that consequently answer questions postured by people in a characteristic dialect. A QA usage, more often than not a PC program, may develop its answers by questioning an organized database of learning or data, as a rule an information base. All the more regularly, QA frameworks can pull answers from an unstructured accumulation of normal dialect reports.

QA is exceptionally subject to a decent pursuit corpus - for without archives containing the appropriate response, there is minimal any QA framework can do. It in this manner bodes well that bigger gathering sizes by and large loan well to better QA execution, except if the inquiry area is symmetrical to the accumulation. The thought of information excess in huge accumulations, for example, the web, implies that chunks of data are probably going to be expressed in a wide range of courses in contrasting settings and documents,leading to two advantages: By having the correct data show up in numerous structures, the weight on the QA framework to perform complex NLP systems to comprehend the content is decreased. Rectify answers can be separated from false positives by depending on the right response to seem a larger number of times in the archives than occurrences of off base ones. Some inquiry noting frameworks depend intensely on mechanized thinking. There are various inquiry noting frameworks outlined in Prolog, a rationale programming dialect related with man-made brainpower.

System Architecture:

In this system, take input as question from user, process on that to removes stopwords and stemming words i.e. less frequency words. Output given to aspect result to find expert answer by more number of similarity score as well as voting score for each answer.The question answer system to provide the best answer from analysis of more number of question or aswer .



System Architecture

There are following modules in our proposed system.

a. Data collection

b. Preprocessing of data

c. Transformation

d. Aspect Category

e. Evaluation

Module1:Data Accumulation

getInputFile() method is used to take the question from more number of user . This method is available in java. This module take input as different question or statement on which system able to process and analyze.

Module2:Preprocessing of data

This module removes stopwords and streeming words from user question sentence. Stopwords are like a, an, am, and, are, as, at, be, been, both, did, do, so, some, was etc. Streeming words like ed, ing, ation , lly etc. All stopwords and streem words are removed from question sentence so deletion of this words are not affected on system as well as processing time is also minimized.

Module3:Transformation

In the transformation process, the score for each sentence is calculated in the presented question .This score helps to detect the similarity scoring.

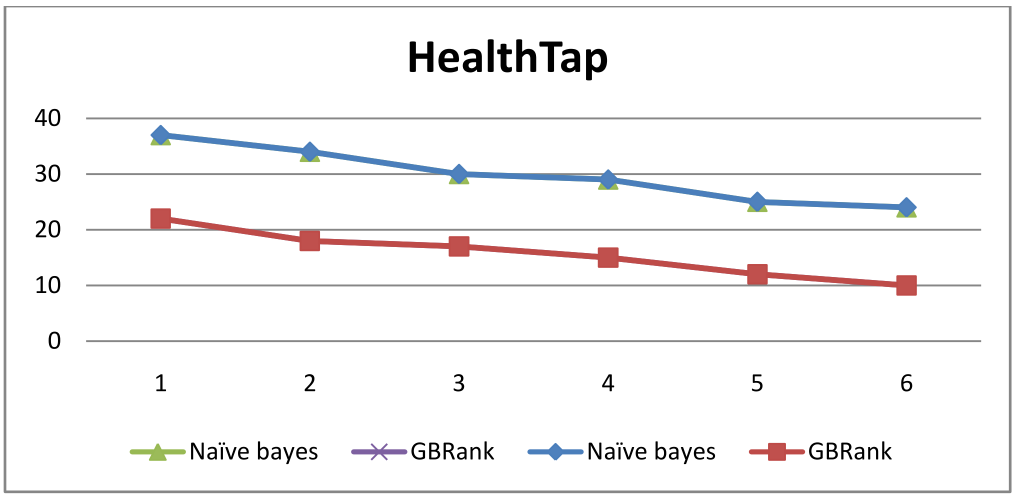
Module4:Aspect Category

Naïve bayes classifier algorithm is used to predict top 5 question and their answers . This algorithm generates config value for each relation exist for particular input categories. Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis and it is robust to noise in input data.

Module5:Evaluation

It gives the result analysis in graph.The representation of similarity scoring is important on every question and answer. This helps to improve accuracy of question as well as answer.There are one graph to represent ther naïve bayes algorithm and GBRank algorithm for each question.

2. Results and Analysis



Comparison of naïve bayes classifier and GBRank for Aspect Category

Que_no	Naïve bayes	GBRank
5	37	22
6	34	18
7	30	17
8	29	15
9	25	12
10	24	10

Aspect category comparison

In this area, we demonstrate the trial after effect of our framework which is contrasted and the current frameworks. In this we done examination over likeness list/score utilizing well being tap

dataset. This examination contrasts closeness record and Naïve bayes, and GBRank techniques. From examination we can
infer that likeness score is conversely corresponding to the no. of inquiries.

4. Conclusion

To provide an unique system for solution option in cQA setups. It consists of an offline learning and also an on the internet search element. In the offline learning element, rather than lengthy and also labor-intensive comment, An instantly build the favorable, neutral, unfavorable training examples the forms of choice sets directed by our data-driven monitorings. Then suggest a durable pairwise learning to rank model to integrate these 3 kinds of training examples. On the internet search part, for a provided question, initially gather a pool of solution prospects through locating its comparable questions. Then use the offline learned model to rank the solution prospects through using pairwise contrast. Actually to carried out comprehensive experiments to validate the efficiency of model on one basic cQA dataset and also one upright cQA datasets.

References

[1]LiqiangNie, Xiaochi Wei, Dongxiang Zhang, Xiang Wang, Zhipeng Gao, and Yi Yang, "Data-Driven Answer Selection in Community QA Systems", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 29, NO. 6, JUNE 2017

[2] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T. S. Chua," Disease inference from health-related questions via sparse deep learning," IEEE Trans. Knowl. Data Eng., vol. 27, no. 8, pp. 2107–2119, Aug. 2015.

[3] 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.

[4] X. Li, Y. Ye, and M. K. Ng, "MultiVCRank with applications to image retrieval," IEEE Trans. Image Process., vol. 25, no. 3,pp. 1396–1409, Mar. 2016

[5] X. Li, M. K. Ng, and Y. Ye, "MultiComm: Finding community structure in multi-dimensional networks," IEEE Trans. Knowl.Data Eng., vol. 26, no. 4, pp. 929–941, Apr. 2014.

[6] 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. 29thAnnu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2006,pp. 228–235.

[7] X.-J. Wang, X. Tu, D. Feng, and L. Zhang, “Ranking community answers by modeling question-answer relationships via analogical reasoning,” in Proc. 32nd Annu. Int. ACM SIGIR Conf. Res.Develop. Inf. Retrieval, 2009, pp. 179–186.

[8] 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.

[9] B. Li and I. King, “Routing questions to appropriate answerers in community question answering services,” in Proc. 19th ACMInt. Conf. Inf. Knowl. Manage., 2010, pp. 1585–1588.

[10] C. Shah and J. Pomerantz, “Evaluating and predicting answer quality in community QA,” in Proc. 33rd Int. ACM SIGIRConf. Res. Develop. Inf. Retrieval, 2010, pp. 411–418.

[11] J. Zhang, M. S. Ackerman, and L. Adamic, “Expertise networks in online communities: Structure and algorithms,” in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 221–230.

[12] A. Agarwal, et al., “Learning to rank for robust question answering,” in Proc. 21st ACM Int. Conf. Inf. Knowl.Manage., 2012,pp. 833–842.

[13] M. K. Ng, X. Li, and Y. Ye, “MultiRank: Co-ranking for objects and relations in multi-relational data,” in Proc. 17th ACM SIGKDD Int.Conf. Knowl. Discovery Data Mining, 2011, pp. 1217–1225.

[14] Y. Cao, J. Xu, T.-Y.Liu, H. Li, Y. Huang, and H.-W.Hon,“ Adapting ranking SVM to document retrieval,” in Proc. 29thAnnu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2006,pp. 186–193.

[15] F. Hieber and S. Riezler, “Improved answer ranking in social question-answering portals, ” in Proc. 3rd Int. Workshop Search Mining User-Generated Contents, 2011, pp. 19–26.

[16] 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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] Q. H. Tran, V. D. Tran, T. T. Vu, M. L. Nguyen, and S. B. Pham, “JAIST: Combining multiple features for answer selection in community question answering,” in Proc. 9th Int. Workshop Semantic Eval., 2015, pp. 215–219.

[22] W. Wei, et al., “Exploring heterogeneous features for query focused summarization of categorized community answers,” Inf.Sci., vol. 330, pp. 403–423, 2016.

[23] A. Shtok, G. Dror, Y. Maarek, and I. Szpektor, “Learning from the past: Answering new questions with past answers,” in Proc. 21stInt. Conf. World Wide Web, 2012, pp. 759–768.

[24] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne,“ Finding high-quality content in social media,” in Proc. Int.Conf. Web Search Data Mining, 2008, pp. 183–194.

[25] S. Tellex, B. Katz, J. Lin, A. Fernandes, and G. Marton,“Quantitative evaluation of passage retrieval algorithms for question answering,” in Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2003, pp. 41–47.

[26] H. Cui, R. Sun, K. Li, M.-Y.Kan, and T.-S. Chua, “Question answering passage retrieval using dependency relations”, in Proc.28th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2005,pp. 400–407.