Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 6, Issue 06, June -2019

"Data Driven Answer Selection in Community QA System"

Rachit Mishra, Vibhuti Bhushan, Vaibhav Kumar, Prof. Dhanashree Kulkarni

Dr. D. Y. Patil College of Engineering, Ambi pune-18.
2018-2019

Abstract — *Nowadays, question respondent system is a lot of convenient for the users, users raise question on-line and so they're going to get the solution of that question, however as browsing is primary want for every a private, the amount of users raise question and system can give answer however the computation time accumulated likewise as waiting time accumulated and same sort of queries area unit asked by totally different users, system got to provide same answers repeatedly to totally different users.*

To avoid this we tend to propose PLANE technique which can quantitatively rank answer candidates from the relevant question pool. If users raise any question, then system give answers in ranking type, then system suggest highest rank answer to the user. we tend to proposing professional recommendation system, associate degree professional can give answer of the question that is asked by the user and that we conjointly implement sentence level bunch technique within which one question have multiple answers, system give best suited answer to the question that is asked by the user.

Keywords- *Community-based question answering, answer selection, observation-guided training set construction, expert recommendation system, sentence level clustering.*

I. INTRODUCTION

As our project strictly supported processing, data the information} mining is that the computing technique of discovering patterns. Associate in Nursing mental object subfield of field of study. the goal of info} mining technique is to extract infofrom a data set and rework it into an obvious structure for any use. apart from the raw analysis step, it involves information and information management aspects, info preprocessing, model and management aspects, info pre-processing and interface problems, power metrics, quality problems, post-processing of discovered structures, image, and on-line modification.{data mining knowledge methoding } is that the analysis step of the "knowledge discovery in databases" method, or KDD.Community Question responsive (cQA) is gaining quality on-line. they are seldom tempered, rather open, and then they have fewrestrictions, if any, on global organization agency can post and global organization agency can answer a matter. On the positive facet,this suggests that one can freely raise any question and expect some smart, honest answers. On the negative facet, it takes effort to travel through all potential answers and to make sense of them.For, example, it's commonplace for a touch on own several answers, that produces it very time intense to the user to look at and to winnow. The challenge we've got a bent to propose may facilitate modification the tactic of finding smart answers to new queries in a {very} very community created discussion forum (e.g ., by retrieving similar queries inside the forum and characteristic the posts inside the solution threads of those queries that answer the question well).The accomplishment of cQA and active user participation, question starvation wide exists in cQA forums that ask the next sorts As the vary of data seekers on cQA is exaggerated, the waiting time is extended to induce answers of their question, attributable to waiting time users get unsuccessful.

Expert have to be compelled to answer of all question albeit the question is same, i.e. if a matter is of same type but still skilled have to be compelled to answer of all question.Given a matter, rather than naively selecting the foremost effective answer from the foremost relevant question, throughout this paper we've got a bent to tend to gift a very distinctive Pairwise Learning to rank model, nicknamed PLANE.We gift a very distinctive approach to constructing the positive, neutral, and negative employment samples in terms of preference pairs. This greatly saves the long and labour-intensive labeling methodology.

We tend to propose a pairwise learning to rank model for answer choice in cQA systems. It seamlessly integrates hinge loss, regularization, associate degreed Associate in Nursing additive terms at intervals a unified framework altogether} totally different from the standard pairwise learning to rank models, and ours incorporates the neutral employment sample.

LITERATURE REVIEW

1) "Data-Driven Answer choice in Community QA Systems", Liqiang Nie, Xiaochi Wei dynasty, Dongxiang Zhang, Xiang Wang, Zhipeng United States Government Accounting Office, and Lolo principle, we tend to gift a completely unique theme to rank answer candidates via pairwise comparisons. especially, it consists of 1 offline learning part and one on-line search part. within the offline learning part, we tend to 1st mechanically establish the positive, negative, and neutral coaching samples in terms of preference pairs target-hunting by our data-driven observations. we tend to then gift a completely unique model to conjointly incorporate these 3 styles of coaching samples. The closed-form resolution of this model comes. within the on-line search part, we tend to 1st collect a pool of answer candidates for the given question via finding its similar queries.

2) " Health-related queries via distributed deep learning" L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T. S. Chua, [2] planned a paper aims to create a sickness abstract thought theme that's able to mechanically infer the doable Diseases of the given queries in community-based health services. during this paper, we tend to 1st report a user study on the data desires of health seekers in terms of queries and so choose those who enkindle doable diseases of their manifested symptoms for more analytic. we tend to next propose a completely unique deep learning theme to infer the doable diseases given the queries of health seekers. The planned theme contains of 2 key parts.

3) "MultiVC Rank image retrieval", X. Li, Y. Ye, and M. K. Ng, [22] propose and develop a multi-visual-concept ranking (MultiVCRank) theme for image retrieval. The key plan is that a picture are often drawn by many visual ideas, and a hypergraph is constructed supported visual ideas as hyperedges, wherever every edge contains pictures as vertices to share a particular visual conception.

4) "Multimedia system answer generation ", L. Nie, M. Wang, Y. Gao, Z. Zha, and T. Chua, [13] during this paper, we tend to propose a theme that's able to enrich matter answers in cQA with applicable media knowledge. Our theme consists of 3 components: answer medium choice, question generation for multimedia system search, and multimedia system knowledge choice and presentation. This approach mechanically determines which kind of media info ought to be intercalary for a matter answer. It then mechanically collects knowledge from the net to counterpoint the solution.

5) "A ranking approach on large-scale graph ," IEEE Trans. W. Wei, B. Gao, T. Liu, T. Wang, G. Li, and H. Li, [33] address the large-scale graph-based ranking drawback and concentrate on a way to effectively exploit made heterogeneous info of the graph to enhance the ranking performance. Specifically, we tend to propose associate degree innovative and effective semi-supervised Page Rank (SSP) approach to parameterize the derived info inside a unified semi-supervised learning framework (SSLF-GR), and so at the same time optimize the parameters and also the ranking several graph nodes.

6) "MultiComm: Finding Community Structure", IEEE Trans. Xutao Li, Michael K. Ng, and Yunming Ye,[31] the most aim of this paper is to develop a community discovery theme in an exceedingly multi- dimensional network for data processing applications. In on-line social media, networked knowledge consists of multiple dimensions/entities like users, tags, photos, comments, and stories. we tend to have an interest find a bunch of users World Health Organization move considerably on these media entities. in an exceedingly co-citation network, we tend to have an interest find a bunch of authors World Health Organization relate to alternative authors considerably on publication info in titles, abstracts, and keywords as multiple dimensions/entities within the network.

EXISTING SYSTEM

To make question respondent system time effective and to cut back user's efforts to search out actual answers for his question by suggesting him antecedently answered same form of queries with its ranking. to beat this downside we have a tendency to use sentence level bunch, this system provides multiple answers which is able to be the precise match for that question. however the actual queries have multiple answers. So, it's tough to outline a selected declare single question.

PROPOSED SYSTEM

We gift a very distinctive theme for answer selection in cQA settings. **Offline Learning:-**
In the offline learning component, instead of long and labor-intensive annotation, we tend to tend to automatically construct the positive, neutral, and negative coaching job samples inside the kinds of preference pairs guided by our data-driven observations.

Online search:-
We initial collect a pool of answer candidates via finding its similar queries. **Database:-**
A tremendous vary of historical QA pairs, as time goes on, area unit archived inside the cQA databases. information seekers thus have large prospects to directly get the answers by trying from the repositories, rather than the long waiting.

Sentence level clustering:
A question that has multiple forms of answers, however providing best suited answer of that question. Expert

Recommendation system:
Same sort of question that is answered by consultants those consultants are recommending to user for more queries.

V. SYSTEM ARCHITECTURE

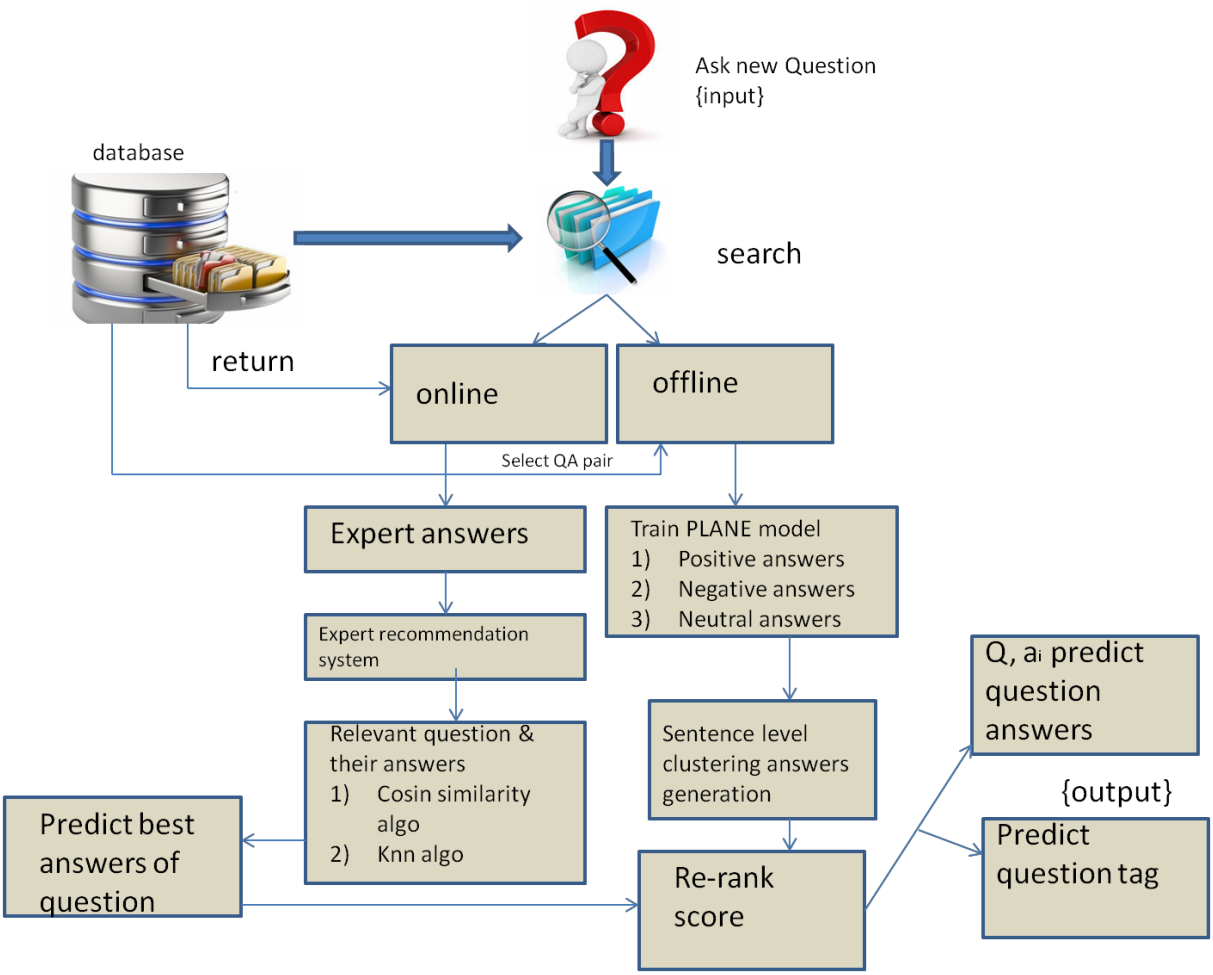

Fig.: System Architecture

VI. CONCLUSION AND FUTURE WORK

Attribute-based encoding (ABE) has been wide employed in cloud computing wherever information suppliers source their encrypted information to the cloud and might share the info with users possessing mere credentials. On the opposite hand, deduplication is a vital technique to avoid wasting the space for storing and network information measure, that eliminates duplicate copies of identical information.

VII. REFERENCES

[1] Liqiang Nie, 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] M. Ali, M. Li, W. Ding, and H. Jiang, Modern Advances in Intelligent Systems and Tools, vol. 431. Berlin, Germany: Springer, 2012.

[3] 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.

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

[5] 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.

[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. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2006,pp. 228–235.

[7] 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.

[8] T. C. Zhou, M. R. Lyu, and I. King, "A classification-based approach to question routing in communityquestion answering," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 783–790.

[9] 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.
[10] B. Li and I. King, "Routing questions to appropriate answerers in community question answering services," in Proc. 19th ACM Int. Conf. Inf. Knowl. Manage., 2010, pp. 1585–1588.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] 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.
[16] W. Wei, et al., "Exploring heterogeneous features for query focused summarization of categorized community answers," Inf. Sci., vol. 330, pp. 403–423, 2016.
[17] 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.
[18] 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.
[19] R. Sun, H. Cui, K. Li, M.-Y. Kan, and T.-S. Chua, "Dependency relation matching for answer selection," in Proc. 28th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2005, pp. 651–652.
[20] M. Surdeanu, M. Ciaramita, and H. Zaragoza, "Learning to rank answers on large online QA collections", in Proc. 46th Annu. Meeting Assoc. Comput. Linguistics: Human Language Technol., 2008, pp. 719–727.
[21] A. Agarwal, et al., "Learning to rank for robust question answering," in Proc. 21st ACM Int. Conf. Inf. Knowl. Manage., 2012, pp. 833–842.
[22] D. Savenkov, "Ranking answers and Web passages for non-factoid question answering: Emory university at TREC LiveQA," in Proc. 24th Text REtrieval Conf., 2015.
[23] 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.
[24] 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.
[25] D. H. Dalip, M. A. Gonc¸alves, M. Cristo, and P. Calado, "Exploiting user feedback to learn to rank answers in QA forums: A case study with stack overflow," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2013, pp. 543–552.
[26] C. Shah and J. Pomerantz, "Evaluating and predicting answer quality in community QA," in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2010, pp. 411–418.

