

Fusing Retrieval Techniques for Enhanced Personalized Community Question Answering^{*}

Vraj Patel^{1,*}, Palak Vanpariya^{1,†}, Kandarp Gajjar^{1,†} and Himani Trivedi^{1,†}

¹LDRP Institute of Technology & Research, Kadi Sarva Vishwavidyalaya, Gandhinagar, Gujarat, India

Abstract

In the current informational age, personalized information retrieval (PIR) has proved to be useful in addressing the problem of information overload. We introduce a new Personal Information Retrieval architecture utilizing the SE-PQA (Stack Exchange – Personalized Question Answering) dataset and a community question answering task model. Our approach leverages rich user relationship level social features and social interactivity data contained in the SE-PQA, which spans over 1 million questions and 2 million answers across various Stack Exchange communities. The proposed model integrates three critical components: initial BM25 retrieval, MiniLM semantic reranking, and user-specific ranking through the TAG model, combining the strengths of traditional information retrieval, efficient language models, and personalized ranking. Extensive experimentation with both the Base SE-PQA and Personalized SE-PQA datasets demonstrates the efficacy of this methodology, with significant improvements in performance metrics. On the Personalized SE-PQA dataset, which incorporates user-selected answers rather than community-upvoted answers as ground truth, our model achieved a P@1 of 0.339, NDCG@10 of 0.465, and MAP@100 of 0.428. These results suggest that incorporating both traditional and neural approaches, along with user-specific features, can contribute to more effective personalized Community Question Answering (CQA) systems, while demonstrating the potential of SE-PQA data in developing and evaluating Personalized Information Retrieval (PIR) frameworks.

Keywords

Personalized Information Retrieval, Community Question Answering, SE-PQA Dataset, Personalized Ranking

1. Introduction

The exponential surge in online information necessitates increasingly personalized retrieval mechanisms to effectively cater to users' diverse needs. Community Question Answering (CQA) platforms, exemplified by StackExchange, face unique challenges in balancing the specificity of user queries with the collective expertise within their specialized domains. Enhancing personalization in these environments holds the potential to significantly elevate user experience by delivering search results that align more closely with individual preferences, expertise levels, and interaction histories.

The recent advancement in PIR has been restricted, since there has not been sufficient availability of larger, real-world datasets that reflect the richness of user behavior and content relevance. The release of the SE-PQA dataset StackExchange - Personalized Question Answering [1] has provided researchers with a source to develop and test PIR models against the context of CQA. This dataset has over 1 million questions and 2 million answers from different StackExchange communities with rich user-level features and social interaction data, thus making it amenable to building complex personalization techniques.

Traditional approaches to PIR in the context of CQA platforms have primarily relied on basic user profiling techniques or rudimentary forms of collaborative filtering. While these methodologies demonstrate some effectiveness, they are inherently limited in their capacity to capture the complex, multi-faceted nature of user preferences and the dynamic social interactions characteristic of CQA environments. Furthermore, the application of advanced language models coupled with sophisticated

Forum for Information Retrieval Evaluation, December 12-15, 2024, India

^{*}Corresponding author.

[†]These authors contributed equally.

✉ vrajp2108@gmail.com (V. Patel); palakvanpariya13@gmail.com (P. Vanpariya); kandarp.gajjar.460@gmail.com (K. Gajjar); himani_ce@ldrp.ac.in (H. Trivedi)

ORCID 0009-0004-9368-3556 (V. Patel); 0009-0009-0137-7188 (P. Vanpariya); 0009-0004-2284-2951 (K. Gajjar); 0000-0002-5039-6424 (H. Trivedi)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

personalization techniques remains an under-explored avenue of research within the CQA domain, presenting significant opportunities for enhancing the accuracy and relevance of information retrieval in these platforms.

The rest of this paper is structured as follows: Section 2 provides a comprehensive review of related work in PIR and CQA, highlighting current limitations and research gaps. Section 3 delineates our implementation, detailing the integration of diverse personalization components. We present and discuss our results and analysis in Section 4, including comparisons with baseline methods and ablation studies. Section 5 concludes the paper by succinctly summarizing our key findings and proposing potential avenues for future research in this rapidly evolving field. Finally, Section 6 encompasses the references cited throughout this work.

2. Related Work

Research in Community Question Answering (CQA) began with retrieval challenges. Othman et al. [2] addressed the word mismatch problem using word embeddings and k-means clustering to achieve semantic similarity in question retrieval. However, issues like out-of-vocabulary (OOV) words, high computational overhead, and biases in embeddings limited its scalability to larger datasets and multilingual contexts.

Yang et al. [3] explored expert recommendation systems, proposing a framework to classify recommendation methods. While this eased limitations in earlier retrieval systems, it introduced new challenges such as sparse data and inconclusive results, highlighting the need for advanced personalization techniques and accurate profiling mechanisms.

Building on these gaps, Zhang et al. [4] proposed a personalized chatbot model that learns implicit user profiles from dialogue histories, mitigating sparse data issues. However, it introduced noise-related challenges in extended dialogues, emphasizing the trade-off between performance and efficiency.

Further advancements were made in "SE-PQA: Personalized Community Question Answering" [1], which introduced multi-domain user interaction features to enhance personalization on large-scale datasets. Although promising, the study relied on a simple user model and called for more extensive experimentation to unlock its full potential across diverse domains.

Recently, the use of Large Language Models (LLMs) has been explored as a transformative step in CQA. A study titled "Large Language Models and Future of Information Retrieval" [5] addressed persistent challenges like OOVs, scalability, and intent understanding. Despite their capabilities, LLMs bring concerns about bias and ethical considerations, necessitating careful deployment in real-world applications.

Ongoing research in Personalized Information Retrieval (PIR) continues to push the boundaries of the field. The 'Overview of the PIR Track at FIRE 2024' [6] provides a comprehensive evaluation of state-of-the-art personalized retrieval techniques, while highlighting the collaborative efforts that have driven advancements in balancing precision, effectiveness, and user relevance. These developments, discussed in the proceedings of FIRE 2024 [7], underscore the growing importance of refining retrieval systems to meet the demands of dynamic and evolving online environments.

These studies collectively illustrate the evolution of CQA systems—from word-embedding techniques to sophisticated LLM-powered solutions—reflecting continuous progress toward enhancing scalability, personalization, and user-centricity.

3. Proposed Model

In our approach for personalized information retrieval, we make use of three main models: BM25 as the primary model for retrieval, MiniLM for reranking, and the TAG model for personalized ranking. BM25 can be regarded to be the retrieval model at the first step of ranking potential answers to the queries raised by the users based on their relevance. It considers query term frequencies in documents

and the overall importance of the documents in the entire dataset so that the most relevant answers are surfaced first.

Next, we employ MiniLM as a neural re-ranker. This model tunes the initial output of BM25 in terms of the context and semantics of a question and its answer. Deep analysis of language patterns and meanings residing in them with the help of MiniLM ensures that answers provided to users are not only relevant but also contextual.

Lastly, we add the TAG model to refine the rank. This assigns scores to the right answer according to the user’s past behaviors and interests, which are depicted through tags. At the moment of query submission, the TAG model calculates the tags associated with the user’s past questions and matches them against the tags associated with the answers. If that result is coming from a user who has the same interest as us, it gains more points, which in turn enhances the personalization aspect of our retrieval system.

Together, the three models-BM25, MiniLM, and TAG form a more robust framework that improve on relevance as well as personalization of answers provided. That actually addresses the problem of information overload in community question-answering platforms and allows users to get answers tailored uniquely to their preferences and context.

4. Experiments and Results

We performed experiments on the SE-PQA dataset to evaluate our proposed model. It combines three components: BM25 for initial retrieval, MiniLM for semantic re-ranking, and the TAG model for personalized ranking. The model aims to combine the strengths of traditional information retrieval, efficient language models, and user-specific personalization.

We evaluated our model on two variants of the SE-PQA dataset: the Base SE-PQA and the Personalized SE-PQA. Table 1 presents the results for the cQA task on the Base SE-PQA dataset.

Table 1

Results for the cQA task on Base SE-PQA.

#	Model	P@1	NDCG@3	NDCG@10	Recall@100	MAP@100
a	BM25	0.330	0.325	0.359	0.615	0.320
b	BERT	0.163	0.184	0.237	0.615	0.199
c	TAG	0.084	0.110	0.160	0.615	0.131
d	BM25 + TAG	0.355	0.349	0.383	0.615	0.342
e	BM25 + BERT	0.364	0.361	0.398	0.615	0.355
f	BM25 + BERT + TAG	0.382	0.378	0.414	0.615	0.370

The results in Table 1 show that the combined method, BM25 + MiniLM + TAG, outperforms individual models and simpler combinations on all metrics. It achieves significant improvements of 15.8% for P@1 and 15.6% for MAP@100 compared to the BM25 baseline. MiniLM proves effective at increasing semantic understanding for both queries and answers, serving as a considerably lighter alternative to larger language models.

To further validate our approach, we evaluated the models on the Personalized SE-PQA dataset, which incorporates richer user-level features. Table 2 presents these results.

As shown in Table 2, the results on the Personalized SE-PQA dataset are even more pronounced. Our integrated model achieves a P@1 score of 0.339, an NDCG@10 score of 0.465, and a MAP@100 score of 0.428. These findings highlight the importance of personalization in CQA systems and demonstrate the effectiveness of the proposed integrated approach, which leverages the semantic understanding and efficiency of MiniLM.

Table 2

Results for the cQA task on Pers SE-PQA.

#	Model	P@1	NDCG@3	NDCG@10	Recall@100	MAP@100
a	BM25	0.279	0.353	0.394	0.707	0.362
b	BERT	0.139	0.209	0.268	0.707	0.232
c	TAG	0.082	0.123	0.175	0.707	0.153
d	BM25 + TAG	0.306	0.383	0.425	0.707	0.392
e	BM25 + BERT	0.315	0.396	0.439	0.707	0.403
f	BM25 + BERT + TAG	0.339	0.422	0.465	0.707	0.428

5. Conclusion

Our work on personalized information retrieval for Community Question Answering systems using the SE-PQA dataset gives good results. In fact, we notice that the combination of BM25, MiniLM, and the TAG model performs better than all individual approaches, as well as any simpler combination. This could be seen particularly in the Personalized SE-PQA dataset, where our best model BM25 + MiniLM + TAG achieved the highest overall scores: P@1=0.339, NDCG@10=0.465, and MAP@100=0.428. The potential of this proposed combination lies in its ability to exploit the advantages that are rooted within each of the different areas: the benefits of the traditional information retrieval algorithms (BM25), efficient language modeling (MiniLM), and user-specific personalization (TAG). The addition of MiniLM proved particularly useful, as it yielded very strong semantic understanding without the computational burdens that often accompany heavier language models.

While these results are promising, there is still quite a lot of room to improve and explore in this area. Future work may further push the personalization aspects toward even higher levels, perhaps introducing more dynamic user behavior modeling or even advanced few-shot learning techniques. The investigation can also be conducted with even more efficient language models or making the existing approach optimized for real-time applications. As these CQA sites are developed, so must be our perceptions of retrieval and our methods of information retrieval, continually seeking to balance precision, effectiveness, and user relevance in ever more complex and diverse online communities.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT-4 for text polishing (Activity: Editing). Additionally, Quillbot was used for grammar correction (Activity: Checking). After using these tools, the author(s) reviewed and edited the content as needed and take full responsibility for the publication's content.

Acknowledgments

We express our sincere gratitude to the Department of Computer Engineering, LDRP Institute of Technology and Research, affiliated with Kadi Sarva Vishwavidyalaya (KSV), for their continuous support, guidance, and encouragement throughout this work. We also thank the University of Milano-Bicocca for providing the results that made this paper possible.

References

- [1] P. Kasela, M. Braga, G. Pasi, R. Perego, Se-pqa: Personalized community question answering, in: Companion Proceedings of the ACM Web Conference 2024, volume 9 of WWW '24, ACM, 2024, pp. 1095–1098. URL: <http://dx.doi.org/10.1145/3589335.3651445>. doi:10.1145/3589335.3651445.

- [2] N. Othman, R. Faiz, K. Smaïli, Enhancing question retrieval in community question answering using word embeddings, *Procedia Computer Science* 159 (2019) 485–494. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919313857>. doi:<https://doi.org/10.1016/j.procs.2019.09.203>, knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 23rd International Conference KES2019.
- [3] Z. Yang, Q. Liu, B. Sun, X. Zhao, Expert recommendation in community question answering: a review and future direction, *International Journal of Crowd Science* 3 (2019) 348–372. URL: <https://doi.org/10.1108/IJCS-03-2019-0011>. doi:10.1108/IJCS-03-2019-0011.
- [4] H. Qian, Z. Dou, Y. Zhu, Y. Ma, J.-R. Wen, Learning implicit user profile for personalized retrieval-based chatbot, in: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management, CIKM '21*, Association for Computing Machinery, New York, NY, USA, 2021, p. 1467–1477. URL: <https://doi.org/10.1145/3459637.3482269>. doi:10.1145/3459637.3482269.
- [5] C. Zhai, Large language models and future of information retrieval: Opportunities and challenges, in: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '24*, Association for Computing Machinery, New York, NY, USA, 2024, p. 481–490. URL: <https://doi.org/10.1145/3626772.3657848>. doi:10.1145/3626772.3657848.
- [6] P. Kasela, M. Braga, E. Sokli, G. C. Milanese, G. Peikos, S. Modha, A. Raganato, M. Viviani, G. Pasi, Overview of the pir track at fire 2024: Evaluation of personalised information retrieval, in: *FIRE 2024: Proceedings of the 16th Annual Meeting in the Forum for Information Retrieval Evaluation*, Association for Computing Machinery, New York, NY, USA, 2024.
- [7] K. Ghosh, T. Mandl, P. Majumder, D. Ganguly, Forum for information retrieval evaluation (working notes), in: *FIRE 2024*, CEUR-WS.org, Gandhinagar, India, 2024.