ISSN: 2664-9640

Personalized Search Recommendation for Open Source Projects Based on Bert

Xumeng Fu [a], Hongmin Ren [b]

School of Information Engineering, Shanghai Maritime University, Shanghai 200135, China;

[a] xumengfu03@gmail.com, [b] hmren@shmtu.edu.cn

Abstract

Modern human beings can get hundreds of millions of relevant information for selection just by typing their personal needs into the search engine. However, how to quickly obtain the most valuable and effective information from the massive information has become the most concerned problem. The purpose of this paper is to study the personalized search and recommendation of open source projects based on Bert, briefly describe the existing recommendation technologies, mainly introduce the word vector model of Bert, and introduce the problem of applying the pre training model to the vertical field and the recommendation algorithm of personalized recommendation system. This paper briefly introduces the source of experimental data, the formation of sampling data set, and the operating hardware environment of the experiment. Then the process of the experiment is described in detail, and the personalized search engine based on Bert is compared with the existing Google search engine. The experimental results show that when three keywords are added to the history file, the ranking recommendation results are relatively good. It can be seen that the number of keywords in the history file also has an impact on the sorting results.

Keywords

Bert Word Vector Model, Open Source Project, Personalized Search, Search Recommendation.

1. Introduction

In the era of Internet economy, relying on network information to obtain economic information and make economic decisions is the current development trend of network economy [1-2]. Personalized recommendation system as a product of the times, has a pivotal position [3]. From the social point of view, as an advanced intelligent commerce platform under the new economic form, the recommendation system has created a unique economic operation mode and greatly promoted economic development and social progress [4-5]. From the perspective of enterprises, the recommendation system can obtain users' preference needs from product rating data, help enterprises understand market supply and demand status, and improve product sales [6-7].

Because users are sometimes unwilling or unable to specify a complete CP network, personalized search cannot be carried out effectively [8]. Barman D proposed a personalized search method using incomplete CP network. A preference recommendation scheme is proposed to supplement the user's CP network to improve the accuracy of personalized search [9]. Liz combines the advantages of content-based filtering and collaborative filtering technology, and proposes a new intelligent architecture of personalized search engine. On this basis, a user collaborative recommendation algorithm is proposed. Simulation results show that the architecture and algorithm can effectively recommend relevant information for users, and has good adaptability [10].

This paper designs an open source project personalized search engine based on Bert. It is a search engine based on the Bert word vector model, which can obtain better word vector representation, efficiently and accurately capture the keywords related to the user's interests. The search engine is implemented by using the recommendation algorithm, and the search results are sorted by adding user preferences on the basis of search engine ranking, In addition to the results returned by the search engine, it also generates a recommendation list based on the user's historical behavior information; the website will store the user's use information and feedback information on the recommendation results to continuously improve and optimize the recommendation engine.

2. Research on Personalized Search Recommendation of Open Source Projects Based on Bert

2.1. Bert Word Vector Model

A typical language model calculates the probability of the next word from left to right, as shown in Formula 1:

$$P(s) = p(w_1, w_2,,, w_m) = \prod_{i=1}^{m} p(w_i | w_1, w_2,,, w_m). \quad (1)$$

However, in many cases, when the pre training model is applied to the vertical domain, it does not need a language model, but a context representation of a word, which can represent the polysemy of a word and the syntactic characteristics of a sentence.

Compared with other language models, better word vector representation can be obtained by making full use of the information of word context, so it can be used to differentiate polysemy to some extent.

The input of decoder is the word vector of the word in the standard summary (the target sentence generated) and the hidden state(s_{t-1}) of the previous RNN cell, and the output is the Hidden state(s_t) of the RNN cell, as shown in formula 2:

$$s_t = RNN_{dec}(y_{t-1,}\widehat{s_{t-1}}). \quad (2)$$

The final output probability is calculated by some nonlinear transformations and softmax:

$$p(y_t | y_{<t}, x) = \mathrm{softmax}(w_s \widehat{s_t}). \quad (3)$$

2.2. Personalized Recommendation System

Recommendation system is widely used in many fields, such as media websites, social networking sites, e-commerce, video and music. These websites have joined their own recommendation systems. In this system, the most important part is recommendation algorithm, which determines whether the system can exist for a long time. If the recommendation quality is excellent, it can bring surprise to the users, better retain the old users and attract new users, the bad recommendation will lead to the wrong guidance of the users and cause the users to have the psychology of resistance. At present, the main recommendation technologies include content-based recommendation, association rule recommendation, collaborative filtering recommendation and hybrid recommendation.

3. Experiment of Personalized Search Recommendation for Open Source Projects Based on Bert

The hardware environment of the test is: inter core i5 dual core processor, 2.53GHz CPU, 2GB memory; and the software environment is R × 643.1.1. According to the specific operation steps of icorc collaborative filtering recommendation algorithm, movielens data and jesterdata are operated step by step.

In this experiment, five postgraduates were selected as the experimental subjects, all of them were Computer Science Masters and often used search engines to search web information. The

subject of this experiment is software. The subjects are required to input the information related to the software and interested in it as user profile. In addition, considering that the knowledge file established by the system is small, some keywords may be omitted. Therefore, this experiment designs five query keywords with software as the theme, which are provided to the experimental objects for query, and determines that any words that may become keywords in the web page information obtained by using these five query keywords already exist in the knowledge file.

4. Experimental Analysis of Personalized Search Recommendation for Open Source Projects Based on Bert

4.1. Compare the Ranking Results of Search Engines

The experiment provided five software related query keywords to five experimental subjects. The subjects were asked to input their interests and preferences into the user profile by using Bert's open source project personalized search engine and Google search engine. In this experiment, we recorded the order of search results by personalized search engine, the ranking order of search results by Google search engine, and the average position difference between the best order of experimental object feedback and the ranking order of search engine.

Compared with the experimental results, X-axis represents the number of queries in this experiment, Y-axis represents the ranking results of personalized search engine and user ranking results of Bert's open-source project, and the average position difference between Google search engine ranking results and user ranking results, as shown in Table 1.

Table 1: Comparison between Bert search recommendation and Google search engine

Search Engines	1	2	3	4	5
Recommended learning (automatic search)	2.1	2	1.8	2	1.7
Bert search recommendation	2.6	2.6	2.8	2.4	2.4
Google search	3.1	3.8	3.2	3.3	3.6
Search Engines	1	2	3	4	5

Whether or not the automatic learning mechanism is added, the average position difference of the personalized search engine of the open source project of Bert is smaller than that of the Google search engine, as shown in Figure 2. However, when the automatic learning mechanism is added to the personalized search engine of the open source project of Bert, the ranking result is better than that of the sorting result without the automatic learning mechanism, so it can be seen that the automatic learning mechanism is effective.

Compared with the traditional Google search engine, due to the introduction of user profiles that can reflect the real interests of users, whether or not the automatic learning mechanism is added, the quality of web page recommendation of the personalized search engine of Bert's open-source projects is better than that of Google search engine; when the personalized search engine of Bert's open-source project adds automatic learning mechanism, the quality of web page recommendation of personalized search engine of Bert's open-source project is better than that of Google search engine, The results show that the quality of web page recommendation is better than that without automatic learning mechanism; when there is no information in user profile, the previous ranking recommendation results are not good, but under the action of automatic learning mechanism, the impact will become more and more insignificant with the increase of query times.

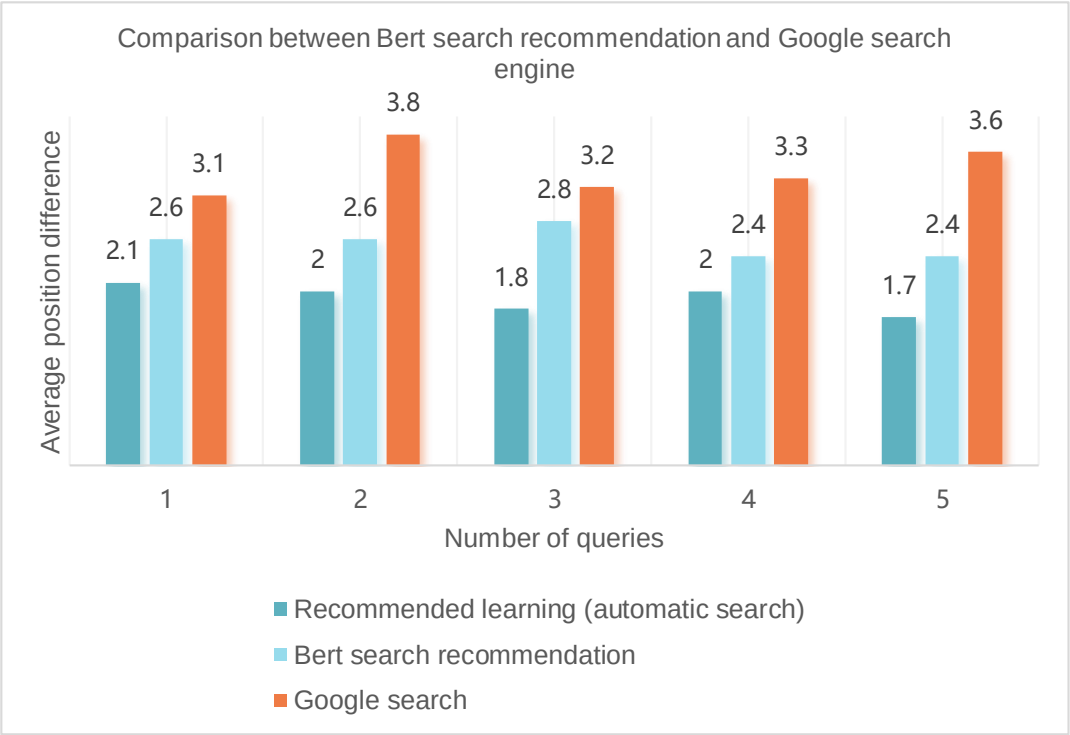

Figure 2: Comparison of Bert search recommendation and Google search engine

4.2. The Influence of The Number of Keywords Added to The History File on The Sorting Results

Record the average position difference of Bert's open source project personalized search engine when four, six or eight keywords are added to the history file after analyzing the keywords that may be related to users' interests and preferences after each user feedback.

Table 1: Comparison of the number of keywords added to the history file

Search Engines	4	6	8
Recommended learning (automatic search)	1.7	1	1.5
Bert search recommendation	2	1.2	1.9
Search Engines	4	6	8

In each query, adding the top six keywords with the highest frequency in the first few pages of interest to the history file will produce better sorting results; when the personalized search engine of Bert's open source project only adds the top four keywords with the highest frequency to the history file, it may be because there are too few keywords added, Some keywords that are really related to users' interests cannot be added to the history file, resulting in poor sorting results; when the personalized search engine of Bert's open-source project adds the eight most frequent keywords to the history file, some keywords that are not related to the user's interest may be added to the history file because of too many keywords, The result of ranking recommendation is poor.

By adding 4, 6 and 8 keywords to the history file, the results show that when adding the most frequent keywords from the first few pages that users are interested in, too few or too many keywords will lead to poor ranking and recommendation results. When six keywords are added to the history file, the ranking recommendation result is relatively good. It can be seen that the number of keywords added to the history file also has an impact on the sorting result.

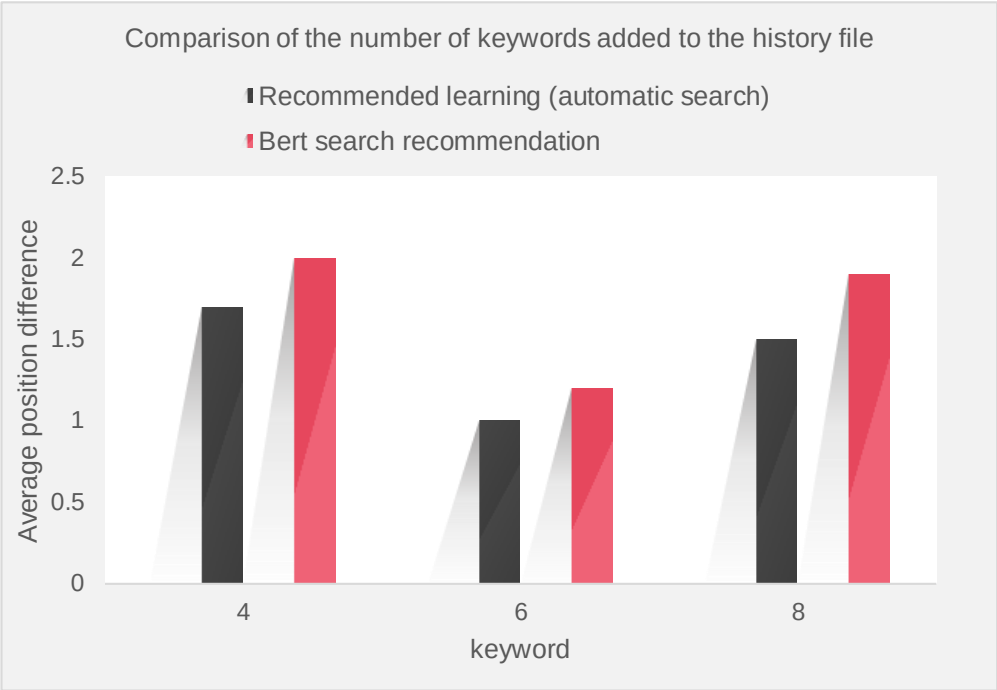

Figure 1: Comparison of the number of keywords added to the history file

5. Conclusion

Personalized recommendation system can not only quickly find users' interests and give suggestions, but also can mine potential user interests. Based on the current development of personalized recommendation system, this paper constructs an open source personalized search engine based on Bert. By using the Bert word vector model, the personalized recommendation level of the recommendation engine can be significantly improved. In addition, compared with the traditional Google search engine, the personalized search engine of open source project based on Bert also adds automatic learning mechanism. Therefore, the personalized search engine for open source projects based on Bert greatly improves the recommendation quality and reader experience. At the end of this paper, the characteristics of the engine are summarized and the future development prospects are prospected.

References

[1] Chen Y, Sun X, Gong D: DPM-IEDA: Dual Probabilistic Model Assisted Interactive Estimation of Distribution Algorithm for Personalized Search, IEEE Access, Vol. 7(2019)No.1, p.41006-41016.
[2] Ke Y, Xiong Y, Hu Y, Liu S: Time-Dependent Demographic Prediction Based on Time-Back-Propagation Method, Chinese Journal of Electronics, Vol.27(2018)No.1, p.35-40.
[3] Wang M, Li Q, Lin Y: A Personalized Metasearch Engine Based on Multi-Agent System, The International Arab Journal of Information Technology,Vol.16(2019)No.6, p.978-987.
[4] Gan, Mingxin, Jiang: FLOWER: Fusing global and local associations towards personalized social recommendation, Future Generations Computer Systems: FGCS, Vol.78(2018)Pt.1, p.462-473.
[5] Jiang M, Fang Y, Xie H: User click prediction for personalized job recommendation, World Wide Web, Vol.22(2019)No.1, p.325-345.
[6] Cui Z, Xu X, Xue F: Personalized Recommendation System Based on Collaborative Filtering for IoT Scenarios, IEEE Transactions on Services Computing, Vol.13(2020)No.4, p.685-695.
[7] Dai W, Wang Y, Ma J: BTR: A Feature-Based Bayesian Task Recommendation Scheme for Crowdsourcing System, IEEE Transactions on Computational Social Systems, Vol.7(2020)No.3, p.780-789.

ISSN: 2664-9640

[8] Le D D, Lauw H W: Stochastically Robust Personalized Ranking for LSH Recommendation Retrieval, Proceedings of the AAAI Conference on Artificial Intelligence, Vol.34(2020)No.4, p.4594-4601.
[9] Barman D, Sarkar R, Tudu A: Personalized query recommendation system : A genetic algorithm approach, Journal of Interdisciplinary Mathematics, Vol.23(2020)No.2, p.523-535.
[10] Li Z, Zou X: A Review on Personalized Academic Paper Recommendation, Computer and Information Science, Vol.12(2019)No.1, p.33-43.