Research on user behavior recommendation algorithm for business process

Kunying Li*, Yu Ding, Xiaolian Li, Ying Shi, Wenting Zhang

Computer Application Technology Department, PetroChina Research Institute of Petroleum Exploration & Development, Beijing 100083, China
likunying@petrochina.com.cn

Abstract: In this paper, ProcessLink is a user behavior judgment and recommendation algorithm based on the information system business process. The purpose is to provide users with reliable unknown process selection. The algorithm idea is as follows: firstly, collect and analyze historical data, establish the relevance feedback set of “Time-Role -Process”, establish the interest relationship decision table of each role according to the process flow link behavior between users. Starting from each process node, the Apriori algorithm completes a round of initial filtering on these relationships, and obtains the similarity association degree of each role node in the process link; finally, the recommendation sorting is carried out through the recommendation algorithm model matching and scenario similarity calculation. This research realizes the accurate prediction of the user's next link and gives the corresponding opinions. The experimental results show that the algorithm is effective and accurate in the business process.

1 . Introduction
In recent years, driven by the Internet and big data, many enterprises have gradually become one of the important machine learning methods to collect and analyze their accumulated historical data. How to extract effective information from massive data and feed back to users is a problem worthy of exploration in this field. Personalized information analysis for specific problems is a solution. By learning various inputs and outputs, such as user's operation behavior and existing business scenarios, key information features can be extracted to provide valuable content for users. At present, more and more personalized information analysis methods have been widely used in various industries, [1] such as analyzing user interests to recommend contacts, communication circles, shopping goods to users, and even software or applications to analyse user emotions. [2] This paper proposes three data weight functions based on access time, resource type and mood message, analyzes user behavior and recommends corresponding mood recommendation to users. [3] Some researches are based on the analysis of external factors, the matching analysis of the advertising quality and user demand of e-commerce websites, the establishment of scoring matrix, the use of keywords to analyse the real demand behaviour of users, and the realization of personalized content recommendation.

It can be seen that the most widely used recommendation algorithm at present is the collaborative filtering and filtering algorithm based on items. This algorithm matches the correlation similarity of the features of the target products according to the users, and analyzes the degree of agreement of each feature to recommend to the target users. However, this kind of algorithm is not very applicable to the business behavior within the enterprise system. In this scenario, the algorithm itself needs to have good learning ability and scenario computing ability. Therefore, this paper proposes a method to analyze user behavior in the business process scenario, and uses the historical data analysis and comparison results
and role environment factors to modify, and finally gives accurate and reasonable suggestions for the unknown next link.

2. The mining and analysis of behavioral data of multi scene factors

In the actual production process of an enterprise, various uncertain scenarios bring huge pressure to scenario calculation, resulting in the final inaccuracy of data results. In order to simplify the complexity of scenario calculation, the first thing to do is to analyze and filter the historical data in the past, and establish the logical relationship between scenario elements. In the decision-making of behavioral data mining, three situational attributes of "time", "role" and "node" are proposed to preprocess historical data with sequence like constraint processing algorithm. [4] [5] Then the frequent pattern mining algorithm Apriori is used for data mining. Finally, the relationship between the three scenario attributes is extracted and transformed into the behavioural correlation set.

2.1 Situation attribute constraint processing

The first step of attribute constraint and rule extraction is to analyze the centroid value among the three according to each attribute. The calculation formula is as follows:

\[ w(x_i, x_j, x_k) = \begin{cases} 1 & \text{if } (x_i, x_j, x_k) \in Pos_c(D) \text{ and } (x_i, x_j, x_k) \notin Pos_c(D) \\ 0, \text{Other} \end{cases} \]  

Among them, I, J, K represent three kinds of situation attribute condition C and decision attribute D, and the calculation result is matrix \( w(x_i, x_j, x_k) \).

Secondly, formula (2) is used to calculate the difference matrix \( M_s \). Where the element in \( M_s \) is calculated as follows:

\[ m(i,j,k) = \begin{cases} \{[a]| \forall a \in C, f(x_i, a) \neq \phi, f(x_j, a) \neq \phi, f(x_k, a) \neq \phi, w(x_i, x_j, x_k) = 1 \} & \text{if } \phi, \text{other} \end{cases} \]  

Then, the difference function \( f(A_x) \) is calculated by formula (3). According to the calculation rules between disjunction and conjunction, it can be directly simplified in the calculation of three elements. Then \( x_{i,j,k}, y_{i,j,k}, z_{i,j,k} \) are any three relationship values in the matrix. If there is an inclusion relationship between them, you can leave it blank to indicate that they are already represented. Therefore, the simplified results are as follows:

\[ f(A_x) = \prod_{m(i,j,k)=a}(x_i, x_j, x_k) \sum m(i,j,k) \]  

Finally, according to the above calculation results, the rules of the three situation attributes are extracted, combined and reduced against the rules.

2.2 Mining user behavior correlation set

On the basis of the previous calculation idea, input all the historical data, and then use Apriori algorithm to mine the data with length of 2 for decision attribute d frequently. In the process of scanning and calculating the historical data of data, in the operation of event set, supporting length analysis and obtaining frequent set, the relationship between the number of cycles and decision attributes is very strong, which will directly affect the final result of the algorithm. Therefore, in this study, each frequent set \( L \) is obtained and merged in each event set, and then the data with less frequent is deleted to keep a higher frequency fluctuation. Then this event can also be expressed as frequent occurrence state, which is in line with the real intention trend of the data.

The relationship between the collection, support and time complexity of behavior correlation set \( S_{(i,j,k)} \) and event set is shown as follows:

\[ S_{(i,j,k)} = O \left( \frac{t_{m}t_{n}\left(t_{n}-1\right)}{z} \times b \times n \right) \]  

When the algorithm generates two item frequent sets, there is a correction process for the time complexity and the frequency of the event. The number of \( L_n \) directly affects the effect of the algorithm. If the number of \( L_n \) is especially large, it means that the efficiency of the use becomes extremely low.
There is a difference between the correlation and the final situation attribute value, so it will not be adopted in the whole correlation set.

3. Implementation of user behavior recommendation calculation for ongoing process links

The feature of Apriori algorithm is not only to calculate the similarity of all historical data in the database, but also to consider the user behavior in the attribute relationship and the impact of the whole process. [7] According to the effective behavior of user operation steps, the user behavior graph in the whole process is simulated and constructed. All possible user operation nodes diffuse along the role node to the future choice of user concern. Assuming that the user's interest graph in the selected operation is represented as AG(AV,AE), then the selected node is the behavioral set AV(AV∈𝑉), with the user's attention. If the next operation of user x focuses on the user y, a connection is established from the concerned user node V U y. The association graph of this kind of user behavior is a directed graph, and the similarity information between each concerned possible target user node and each node in the process link spreads along the whole process route. Figure 1 is an example of user behavior association. It shows that user x selects a and B from the process, and then C is not being followed. Therefore, C is listed as the abandon node, and no one pays attention to it in this process phase. Therefore, we need to deduce all possible wires of interest after B, and introduce the set of S(I, J, K)) behavior correlation degree to exclude the impossible node relationship chain. In this way, we can get the diffusion information of users' interest along the b-node development, and continuously update ourselves and each node option in the map according to the received correlation information.

Figure 1. User behaviour association diagram

The node calculation method obtained from the information diffusion process of the operation interest association graph of the concerned user node is as follows: assuming that the user set of the user node vn in the whole process is TM vn, the absolute value | TM vn | can be expressed as the number of the concerned user nodes, then the correlation between the X node and the Y node can be expressed as sim(vx, vy). Therefore, the validity of the associated nodes obtained by process node vn is calculated as follows:

\[
\text{ProcessLink}(v_x, v_y) = \frac{1}{|\text{TM} (v_n)|} \times \sum_{v_n \in \text{TM} (v_n)} \text{sim}(v_x, v_y)
\]

(5)

It should be noted that if there is no target user to pay attention to at the last node in the process, then it is necessary to ensure that the correlation degree of this node is minimized during each round of frequent binomial calculation to ensure that other nodes will not be affected by this node.

On the basis of the above, the correlation degree between nodes is used to calculate, so as to get the possibility of connection between each node, so as to calculate all the effective diffusion options under the current node. But even so, the accuracy of target judgment cannot be paid attention to by all users, especially for some users whose personal preferences and working environment are constantly changing, which leads to the decline of recommendation accuracy, which is inevitable. Therefore, [8] this study also designed a random forest model to classify and evaluate the invalid recommendations, which includes tracking the user tuple of the concerned behavior and the real selection record of the user, and then input the data into the context constraint attribute in reverse, as a way to correct the deviation, so that the result orientation of the data will be improved with the reality of the scene.
4. Experimental results and analysis
Through the data calculation of the collected samples, the average error rate of the results fluctuates up and down to 9%. Take part of the data samples in the second month after operation for drawing, and get Figure 2.

![Figure 2. Error rate line chart](image)

Due to business reasons, there will be a relevant business process adjustment at the beginning and middle of the month, so the two peaks of error fluctuation on the graph reach 19% and 16% respectively, and the most accurate range generally appears at the end of the month. At this time, the model has recalculated the deviation data and corrected the wrong results. The average error rate of the experiment is 9.7%, while the accuracy of variance calculation after excluding the highest and lowest values is 0.0034, It shows that the experimental results have little fluctuation with the real selection, which can be used as an effective recommendation option, and the success rate is very high in a stable environment.

5. Conclusion
On the basis of a large number of experiments, this paper proposes a recommended algorithm model ProcessLink, which analyses user behavior and applies it to business process. The algorithm fully considers the main potential factors in the whole process, and effectively analyzes user's role, time and node. For users in the enterprise, the system provides practical and reliable recommendation options for business processes. For each process link, the system considers factors such as work processing time, delay caused by emergencies or rapid audit. So that users can more reasonably consider the trend of the next process, shorten the sunk cost caused by too long business processing, and finally realize the accurate prediction of users' future links and give corresponding opinions. The experimental results also show the reliability and accuracy of the algorithm model. In the next step, we will further improve the algorithm, change the influencing factors, increase the environmental weight and time cycle, and conduct research and exploration in more complex business processes.

Reference
[1] LIU Meng-Juan, WANG Wei, LI Yang-Xi, LUO Xu-Cheng, QIN Zhi-Guang. AttentionRank: A Graph-Based Recommendation Combining Attention Relationship and Multi-Behaviors. Chinese Journal of Computers, 2017, 40(3): 634-648.
[2] Wang Ning, Hu Qingchun. Research of recommendation algorithm based on analysis of user behavior in learning system. Computer Era, 2015(11): 4-7.
[3] Jin shishi, Lu Hanlin. Research on collaborative filtering algorithm based on user behavior. Computer Knowledge and Technology (Academic Exchange), 2019, 15(5): 17-18.
[4] Huang N E, Shen Zheng, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-station time series analysis[C]//Proceedings of the Royal Society of London, 1998: 903-995.

[5] LI Long;LIU Peng;ZHANG Kejia;HUANG Shan;LI Qian. Research and Application of Improved Apriori Algorithm. Computer & Digital Engineering, 2019, 47(6): 1293-1297.

[6] Cover T, Thomas J. Elements of information theory[M]. New York: John Wiley & Sons, Inc., 1991.

[7] WANG Xiao-li;XI Ke-min;LIU Zhan-bo;YAN Shi. Analysis of Association Rules Based on Apriori Algorithms. computer engineering & Software, 2019, 40(2): 23-26.

[8] GUAN Xiao-qiang;PANG Ji-fang;LIANG Ji-ye. Randomization of Classes Based Random Forest Algorithm. Computer Science, 2019, 46(2): 196-201.