Research on Mining Association Behavior of Smart Home Users Based on Apriori Algorithm

Yongjun Zhang[1, a], Tingting Qi[2,b]

[1] Beijing University of Posts and Telecommunications, Beijing100876, China;

[2]Beijing University of Posts and Telecommunications, Beijing100876, China.

[a]yjzhang@bupt.edu.cn, [b]andrea_tt@163.com

Keywords: Apriori algorithm, Smart home, user behavior preference.

Abstract. This paper mainly studies the improvement of Apriori algorithm based on smart home system. Association rule algorithm has the promotion effect to the analysis of the smart home user behavior. Apriori algorithm is one of the classical algorithms of association rules. However, the efficiency of the Apriori algorithm is not high, there is stillroom for improvement in smart home. Based on the characteristics of smart home system and terminal operation platform, this paper optimizes the Apriori algorithm. By introducing the auxiliary matrix, the number of scanning database is reduced, and the speed of the support solving is accelerated. Simulation experiment results show that the Apriori algorithm optimized is more efficient than the optimal algorithm.

1. Introduction

In rapid advocacy of the Internet of things,with the improvement of people's living standards and housing conditions, digital intelligent have walk into people's life. Smart home products have already entered the vision of ordinary families and greatly simplified people's life. Which make it possible to develop a new lifestyle. Smart home products can not just stay in the original remote control mode. Packing multiple behaviors ofusers and concurrent operation will further streamline operations that improve the intelligence level. Custom profile is a manifestation. Technology product is continually improving the experience of users and to simplify user operations, which is the next target of smart products. According to the user's behavior,mining high degree of user behavior associatedthrough data mining analysis and then recommended as a personalized user profile. Which replacing the custom profiles of trouble, adding intelligence andmeet user preferences in operation.

In data mining, association ruleis an importantissue. Association rules can find the existence of things that may exist or contactfrom the data behind. Which could be applied to the association study of user behavior in the smart home. The most classical association rule algorithm is Apriori algorithm. Apriori algorithm is an iterative manner based mining method by searching frequent item set. The mining process of Apriori algorithm is complex because of the iterative. Based on the characteristics of smart home system, this paper proposes an improved method of Apriori algorithm, which makes it more suitable for the implementation and application of the terminal platform.

2. User Behavior and Associated Algorithm

Data are mainly from the user's command to the smart home in the terminal equipment. The data storage format is in the form of linear linkedlistand circular queue. Data mainly includes deviceid number and status. Data mainly refers to output device data. Data recording processfor frequen to perations considered to be user errors, not record sand finally deleteduplicates. Due to the above operation is the basic procedure, not in thisrepeat, it mainly introduces the algorithm of association rules analysis of user behavior.

2.1 Association rules and Apriori Algorithm

Given a set of itemset $I=\{i_1,i_2,\ldots,i_m\}$, a set of affairs set $D=\{T_1,T_2,\ldots,T_n\}$, where $T_i\subseteq I(i=1,2,\ldots,n)$, that means T_i is a subsetof I. The so-called association rule is the implication

of the form $X \Rightarrow Y\left(X,Y \subseteq D, X \cap Y = \varnothing\right)$.Association rules havetwo important concepts,which called support and confidence. Support denoted by $Sup\left(X \Rightarrow Y\right)$, is the number of affairs set contain X and Y percentage of the total number of affairs set. Confidence denoted $Conf\left(X \Rightarrow Y\right)$, is the number of affairs set contain X and Y percentage of the number of affairs set contain X, which indicate the intensity of $X \Rightarrow Y$. Formula as follows:

$$Sup\left(X \Rightarrow Y\right) = P\left(X \cup Y\right) \quad (1)$$

$$Conf\left(X \Rightarrow Y\right) = P\left(Y \mid X\right) \quad (2)$$

Apriori is a kind of association rulealgorithm mining frequent itemsets, the core idea is that iterative mining frequent itemsetsby generating candidate itemsets and pruning. It used to explore the $K+1$item set by K item set, finally exhausted all frequent itemsets in dataset.Apriori algorithm flow is shown in Fig.1.

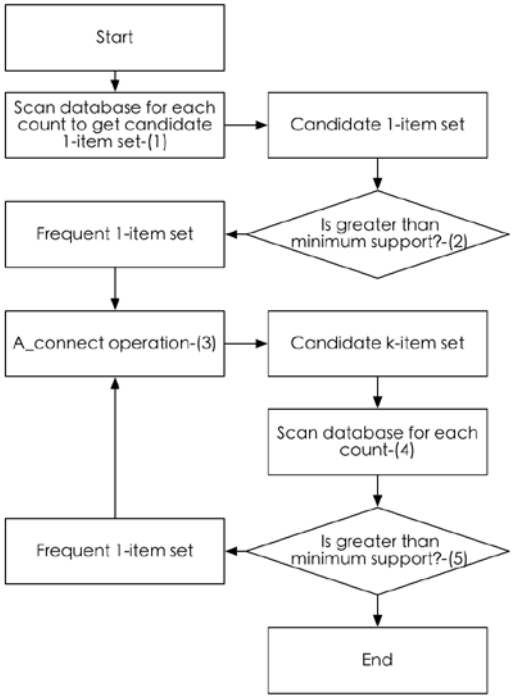

Fig. 1

Step as follows:

1) Count the number of each item set by scanning affairs set database, then get the candidate 1-itemset C_1.

2) Determine whether each item set in C_1 is greater than the minimum support and result in the frequent 1-itemset L_1.

3) Connection step: get frequent 2-itemset C_2 by connecting item sets in L_1(A_connect).

4) Prune step: count the number of each itemset to get the itemset support by scanning affairs set database.

5) Delete itemsets which support is less than the minimum support to get frequent 2-itemset L_2.

6) Repeat step 3,4 and step 5 until frequent k+1-itemset doesn't match the condition. Then get the wanted frequent k-itemset.

Through the above steps can be seen that iterative process of two steps connection and pruning in the mining process of Apriori algorithm. It is required to judge the condition that former k-2 item sets are equal and the k-1 itemset is unequal when connecting itemsets in L_{k-1} to get C_k. It will generate large set of candidate itemsets that reduces the efficiency of the algorithm. It is also required to get the support of each itemset in C_k by scanning affairs set database after get C_k which is inefficient. Therefore, taking into account the characteristics of the smart home system and

the operating background of the mobile platform, the Apriori algorithm could be optimized to improve the efficiency and reduce the computation time.

2.2 Apriori Algorithm improvements

According to the characteristics of Apriori's iteration operation mentioned above, we optimize Apriori algorithm targeting on the two steps of connection and pruning.

1) Scan each item set in the affairs set database and store the content and locate information into a matrix of order n*2, which denoted as $IMP[n][2]$(n depending on specific circumstances). $IMP[n][2]$is used to record the total number of item sets, so the item set of records starting from the second row. The first column of the matrix is used to store the contents of each item set.The second column is used to store the location information for each item set.Location information is calculated, assuming the base is zero, plus 2^k on the basis of base if an item set is concentrated in affairs set k.Location information is helpful for theconnection operation. $IMP[n][2]$ is the candidate 1-itemset. Then pruning to get the frequent 1-itemset according to the minimum support. It only recordthe subscript of frequent item set and doesn't move the content of the item set when generate the frequent 1-itemset. Which will save memory and accelerate the operation speed. This is because the content of item set maybe large. Then subscript of frequent item set is the row number of matrix which is helpful to simplify the connect operation when generating the candidatek-itemset.

The content and locate information of each item set should be statistics by scanning the affairs item database and stored in the matrix $IMP[n][2]$. For example if an item exit in the number one, three and four affairs set. Its locate information should be $2+2^3+2^4=26$.

2) When candidate item set $C_k$is generated by the frequent item set L_{k-1}, some item sets can be determined in advance for non-frequent item sets could be deleted directly. Which could reduce the number of candidate item sets. Specific principle: each item set is in order because the frequent item set L_{k-1} record the subscript of each item set in $IMP[n][2]$ matrix. If the former k-2 item of any two item sets has unequal item, the connection operation can be stopped. This is because the connections afterwards do not meet the linking condition. It also reduce the number of connect operation. For example $L_2=\{1,2;1,3;1,4;2,3;2,4;3,4\}$, in the operation of L_2connection, $L_2[1]$ don't have to compare with $L_2[4]$ because $L_2[1].[1]<L_2[4].[1]$. Because of the order of item set, $L_2[4].[1]$ must less than or equal to the first item of the item sets below. Experiment show that when affairs item number is fifty and the minimum support is 80%, the operation time of connection accounted for 36.2% of the whole algorithm, while is 4.3% of the optimized algorithm.

3) After determine the candidate item set C_k, the support of each item set should be calculate. Which is a time consuming process. In order to reduce the number of database scan, matrix $IMP[n][2]$ could be used. For example, an item set is $\{1,2,4\}$, the items is the subscript of matrix $IMP[n][2]$. The locate information of each item could be picked up according to the corresponding subscript. And these locate information to get it of the new item set. The number of 1 in the binary representation is the support of the new item, which calculation is simple.

On one hand the number of item sets is limited to approximately 64 because of the principle of the optimized algorithm. But because of the characteristics of the smart home system, the limit number will not impact on the system. On the other hand the number of affairs item sets is limited to approximately 60 because of the limitation of mobile computing power. Which will also not impact on the system because of the recent request for the data.

3. Simulation Experiment Design and Analysis

The Apriori algorithm and the optimization algorithm are implemented in the simulation experiment, and the two algorithms are compared by simulation.Starting from the background of

the application of smart home system, the simulation experimental environment choose iOS mobile terminal as platform, with Xcode integrated development environment,programming language choice mixed C and Objective-C language.

Standard test data of Apriori algorithm was chosen for the experiment test data.Because of the characteristics of smart home system, which is characterized by the characteristics of the recent data and the characteristics of the mobile phone platform, part of the test data was chosen.

Experiment one: compare the implementation time of the two algorithms under different support, and consider thenumber of affaires set is 60, the number of item set is 60 which is within the allowable range. Experiment shown in Fig.2.

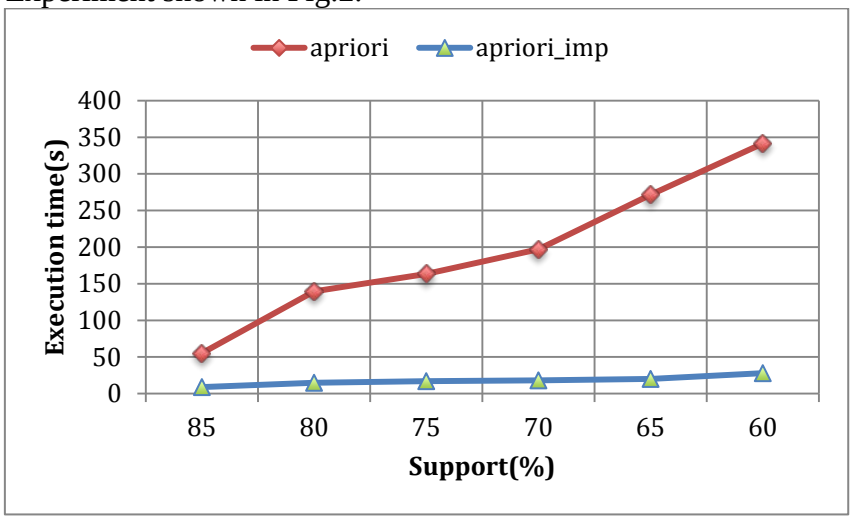

Fig. 2

As can be seen from the simulation experimental results of experiment one, in the premise of the same number of item sets and affaires sets, execution time increased with the decreasing support. However the execution time of the optimized Apriori algorithm is much less than the original Apriori algorithm which growth rate of the execution time is alsoslower.

Experiment two: compare the implementation time of the two algorithms under different numbers of affaires set, and consider the minimum support is 0.6, the number of item set is 60. Experiment shown in Fig.3.

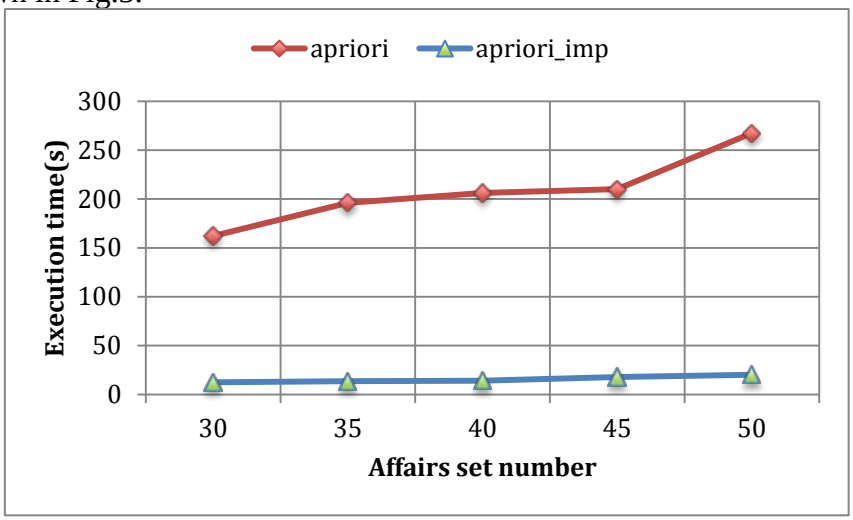

Fig. 3

As can be seen from the simulation experimental results of experiment two, in the premise of the same number of item sets and the same minimum support, execution time increased with the increasing number of affaires sets. However the execution time of the optimized Apriori algorithm is much less than the original Apriori algorithm which growth rate of the execution time is alsoslower.

As can be seen from the above two experiments, the execution speedof Apriori algorithm optimized is faster, which is more suitable for operating under the terminal platform for smart home system.

4. Summary

In summary, this paper takes the smart home system as the research background, takes the iOS terminal as the platform, and optimizes the Apriori algorithm.By introducing the auxiliary matrix, the number of scan database is reduced, the speed of the algorithm is improved, and the generationcandidate itemsets are reduced, as toimprove the efficiency of the algorithm.Experimental results also prove the fact. This has some significance for the analysis of user behavior, intelligent development of the smart home system, and simplify the operation of the user.

References

[1]. Li Xiaocong, TengShaohua. Frequent itemset mining improved Apriori algorithm [J]. Jiangxi Normal University Journal (NATURAL SCIENCE EDITION), 2011,05:498-502.

[2]. Liu Huating, GuoRenxiang, Jiang Hao. Research and improvement of Apriori algorithm for association rule mining [J]. computer application and software, 2009,01:146-149.

[3]. JiaRuiyu, QianGuangchao, Zhang Ran, Li Longshu. An optimization method for the Apriori algorithm [J]. Computer Engineering, 2008,23:196-198.

[4]. Gong Jing, FeiFei Zhou, Lu Jia, Chen Wei Liu. Based on the content preference of mobile Internet customers access behavior analysis [J]. Computer technology and development, 2012,12:149-152.

[5]. Zhang Deying. The design and implementation of a mobile user preference acquisition system based on context and user behavior [D]. Beijing University of Posts and Telecommunications, 2012

[6]. Chen Liang. Research and development of smart home client system IOS platform [D]. Wuhan University of Technology, 2014

[7]. GeQinghua, Fu Wei, Xue Zhen. Research and implementation of smart home software based on IOS system [J]. world electronic components, 2013,09:43-45.