

Accessing Data Efficiently Through Skyline Technique on User Behavior Analysis

Minal H. Kawalkar¹, Bhagyashree Madan²

¹ Miss. Minal H. Kawalkar W.C.E.M, Nagpur, Maharashtra, India

² Prof. Bhagyashree. Madan W.C.E.M, Nagpur, Maharashtra, India,

ABSTRACT

A smart home is new concepts which have been popular at home and abroad. Due to the rapid development of smart home, a large amount of life data have been accumulated, and these data contain a lot of information about user behavior, they shows user's activity truly and objectively. But these data exist in multiple data sources, such as computers, smart phones, and smart televisions and so on. So we have to gather this data and make the right model to find user behavior. Now we propose technique that searches any record with the help of characteristics of user behavior by using skyline technique. Using it, user gets expected result of file that they always want. We design skyline query for incomplete data set which gives better result

Keywords:- User Behavior, Skyline query

1. INTRODUCTION

Smart home system has always been the focus of global IT people and home appliances manufacture. A fully fledged smart home system is collection of different household appliances. Due to the popularity of smart home, there will emerge sufficient data in our day to day life. Many people started to pay attention to make full use of the data processing capabilities of smart devices, to analyze data of the smart appliances to extract the user's normal patterns of life and the habits, and finally provide users with personalized service and remind. However, using traditional data analysis techniques to deal with the isomerization of data set does have many limitations. Considering the diversification of those data that generated from our daily life, how to use these existing data to analyze user behavior and how to provide personalized services to different people under the specific conditions such as a particular time or place, seems to become a more and more important issue. As an important part of the data mining technology, association rule mining aims at investigating how item sets and iterative search method, through this progressive way it shall identify the relationship among those items then form a rule. This paper is based on the method of Apriori algorithm which is a series of deep research will be conducted including mining towards the massive data. This massive data associated with our life and it find out the potential information that users preferred.

Data mining is the process of extracting interesting (nontrivial, implicit, previously unknown and potentially useful) information or patterns from large information repositories such as: relational database, data warehouses, XML repository, etc. Also data mining is known as one of the core processes of Knowledge Discovery in Database. Data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Although data mining is a relatively new term, the technology is not. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost. Association discovery in databases. Among sets of items in transaction databases, it aims at discovering implicative tendencies that can be valuable information for the decision maker. Database mining is motivated by decision support problems faced by most business organizations and is described as an important area of research. One of the main challenges in database mining is developing fast and efficient algorithms that can handle large volumes of data because most mining algorithms perform computation over the entire database. Association rules identify the set of items that are most often purchased with another set of items. For example, an association rule may state that "95% of customers who bought items A and B also bought C: and D." Association rules may be used for catalog design, 'store layout, product placement, target marketing, etc.

As an important part of the data mining technology, association rule mining aims at investigating how to find out the underlying rules and links through the massive amounts of data. Through this method, it could be helpful with the decision-making. Furthermore, Apriori algorithm is a probability-based Boolean association rules mining algorithm. It uses the knowledge of frequent item sets and iterative search method, through this progressive way it shall identify the relationship among those items then form a rule.

In data mining, association rules are applied very widely and it is the most common method of data mining. There are already many studies analyzing user behavior with association rule algorithm, and successfully applied in a variety of scenarios. There are some instances of detailed algorithm and research units. Our work is based on the existing data mining algorithm and we will further refine the data mining association rules and adapt it to the isomerization data mining in the life data from smart home, finally find out the characteristics of user behavior.

2. LITERATURE REVIEW

Bo Yin ,Zhe Zhang, Xi Wang introduces the association rule mining model for isomerization life data from multiple data sources. Through the pretreatment and the integration of user's life data to form sub data source; Execute association rules algorithm in each sub data source to obtain the data source association rules, and then integrate all association rules, for the given input of user's data, we can calculate the association between knowledge points. The final results show that, this method can get the association rule and handle isomerization multiple data sources of user behavior analyses.

A.Krishna Kumar, D.Amrita , N.Swathi Priya implements a new approach to prune and filter discovered rules. First, we propose to use ontologies in order to improve the integration of user knowledge in the post processing task. Second, we propose the Rule Schema formalism extending the specification language, which help in taking decision. On the one hand, domain ontologies improve the integration of user domain knowledge concerning the database field in the post processing step. On the other hand, we propose a new formalism, called Rule Schemas, extending the specification language. The latter is especially used to express the user expectations and goals concerning the discovered rules.

Bourget C, Kruse et al. This paper actualize technique affiliation guideline prompting strategy additionally called market crate investigation which is utilized to discover regularities in shopping conduct of client of store, online shop. With the assistance of prompting affiliation principle on tries to discover arrangement of item that are habitually united .The significance of tenet is measured by two parameter. One is bolster which is the percentage of exchange that the guideline can be connected. Second is the certainty which is the quantity of cases in which it is relevant.

Bay VO,Bac Le et al. This paper executes another calculation for mining summed up affiliation guideline. The calculation filters database one time just and Tidset to registers backing of generalized item set speedier. This calculation mine thing set hierarchal database with least support.

Rahul Bharuka, P Sreenivasa Kumar et al. This paper preferences quires returns answer which are relevant to the user requirements. The subclass of preference query called skyline query that retrieve result over incomplete data set. It is widely use in personalization, multicriteria decision making or search space pruning .this paper propose efficient SIDS(Sort-based Incomplete Data Skyline) algorithm for computing skyline point over dataset.

Mohamed E. Khalefa Mohamed F. Mokbel Justin J. Levandoski et al .This paper focus on the problem of skyline over incomplete data where multidimensional data are missing some values of their dimension. This paper implements two algorithms Replacement and Bucket algorithm that use variation of traditional skyline algorithms to deal with incomplete data. The propose technique of Iskyline algorithms which is specially designed for incomplete data .it uses two technique virtual point and shadow skyline to exploit properties of incomplete data.

Kian Lee Tan ,Pin –KwangEng et.al This paper focus on the set of interesting answer called skyline from database.In this paper the set of point , comprises the point that are not dominated by other. The better or good point dominate other in all dimension. It uses Bitmap and Index Algorithm.

Yufei Tao, Dimitris Papadias, Greg Fu, Bernhard Seeger et.al In this paper skyline of d-dimension dataset contain the point that are not dominated by other in all dimension .The progressive is the method which can quickly returns initial result without reading all database. This paper use BBS(Branch and Bound Skyline) algorithm.

3. PROPOSED TECHNIQUE

Existing system consist of user behavior model which consist of four parts data acquisition, data pre-processing, data mining, and analysis of result. Data acquisition will collect data from different component unit of smart home and data pre-processing will integrate isomerization data that come from different terminals to improve the quality of data mining. In the part of data mining, we will analyze the data by means of association rules, and get the predictable results ultimately.

1. We can collect user data ,such as bookmarks ,browsing history and message logs of user. This parameter will help us to finding user behavior

2. In next step we use apriori algorithms and k-means algorithm for processing data. By using this algorithm we can achieve existing system.

3. In last step we apply association rule and skyline technique.

3.1 Modul: Registraion

In first module user has to register himself .the basic information about the user such as name, date of birth, gender, monile no, city . he has to set his password for login time. After sucessful registration user informatiom save in databse. The pssword stored in encrypted form. In database the user informtion is stored and it can be differ with user id. As we use different devive such as smart phone, desktop, laptop, tab etc all the data which is stored on different device is collected and should be stored on server.



Fig 1: Home Page



fig 2: Registration Page

To create new registration click on sign in , then registration form open. We have to fill all the basic information of the user .such as name. gender, mobile no. , password city, state etc. if we miss the any field ,registration is not done . message will provide to the user fill the field. Consider if we don't feel field like city then server send the message fill the city name. all the data of the user is stored in server.

3.2 Modul 2: Login

The second module is login module ..as the data is already stored on server we can use this data for next step. The data like history, bookmarks ,message and all types of file are used. First we apply the k- means aloritham to form cluster. It reduces the time required for processing data. Then we apply the apriori aloritham to find out the frequent patern ocuring.

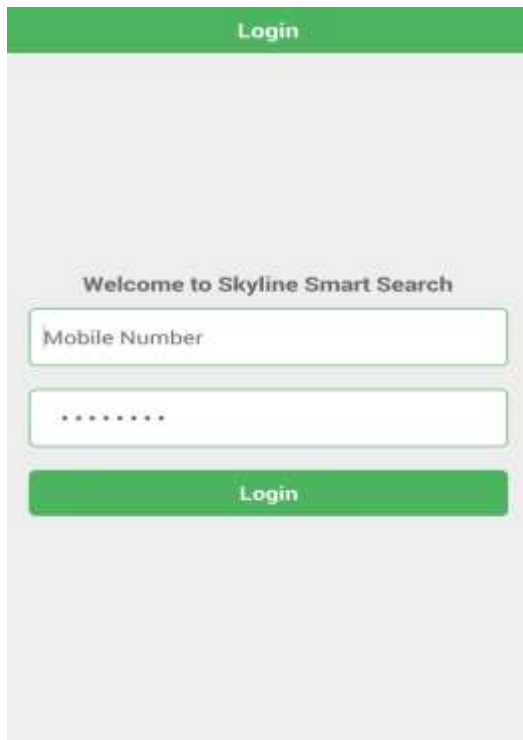


Fig. 3: Login Page

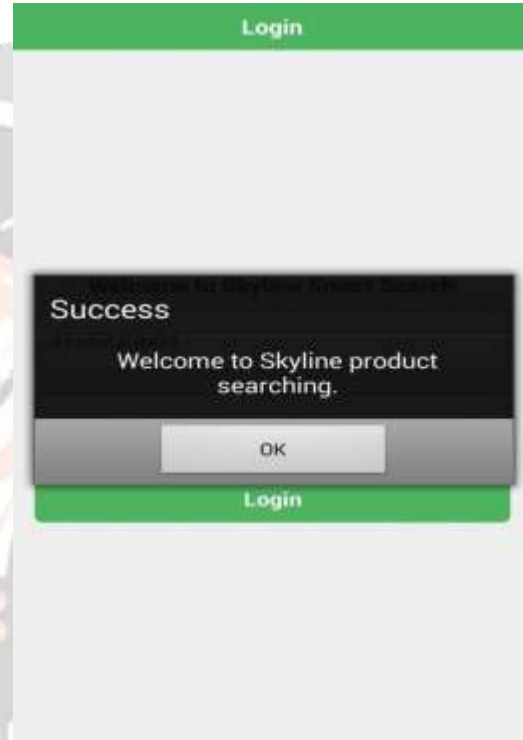


Fig.4: Successful Login

After registration user can easily login with the help of mobile no and password. As the user is already register server check user name and password. If the both parameter are correct sever provide message welcome to skyline searching product. After that server collect data which is present in different device , and apply k-means algorithm to form cluster of the data like makes the cluster of bookmarks, files, history. This cluster returns to the server.

3.3 Modul 3: Feature Add &Search

In third module we can add feature of the object and serch. We can add feature according to the requirment of user. In this module we apply different association rules over the produce data set. Such as

- Candidate rule is generated by merging two rules that share same prefix in the rule consequent.
- Join $(CD \Rightarrow AB, BD \Rightarrow AC)$ would produce the candidate rule $D \Rightarrow ABC$.
- Prune rule $D \Rightarrow ABC$ if subset $AD \Rightarrow BC$ does not have high confidence.
- If (A,B,C,D) is frequent itemset, candidate rule $ABC \Rightarrow D, ACD \Rightarrow B, BCD \Rightarrow A, A \Rightarrow BCD, B \Rightarrow ACD, C \Rightarrow ABC, D \Rightarrow ABC$.

For efficient searching we apply skyline techniuque . The native approach, for computing the skyline over incomplete data, would be to perform exhaustive pair wise dominance checks for all input points and discard the dominated ones. A skyline query returns all the interesting points in a multi-dimensional data set that are not dominated by any other point with respect to all dimensions

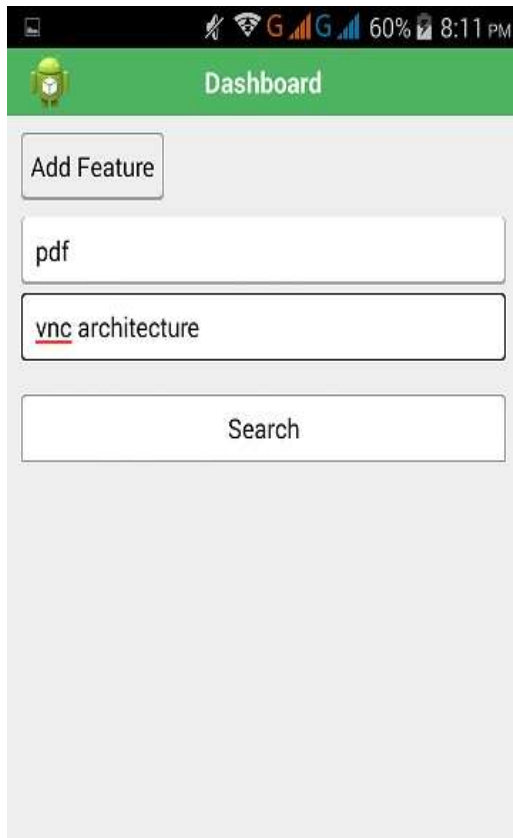


Fig 5: Search Record

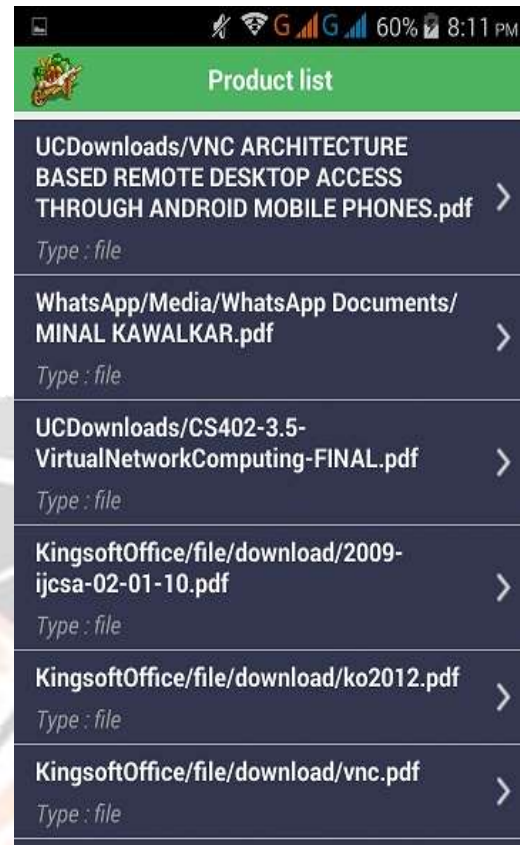


Fig 6: Result Of Searching

After successful login the new page is open for searching record. We can add feature for searching in feature field. We can add more than one field for searching .when the user request server for searching to the server the server apply association rule on cluster. The different association rule are apply on the cluster data. The apriori algorithm is used for finding the frequent pattern . For efficient searching we apply skyline technique . The native approach, for computing the skyline over incomplete data, would be to perform exhaustive pair wise dominance checks for all input points and discard the dominated ones. A skyline query returns all the interesting points in a multi-dimensional data set that are not dominated by any other with respect to other dominated point. Finally we get the respected result for searching.

3.4 GRAPHICAL REPRESENTATION

We make graph for time required for searching in existing system and proposed system by using parameter time(ms) and no of user. This graph show that time required for searching record with k-means and skyline is less than without using k-means and skyline technique .

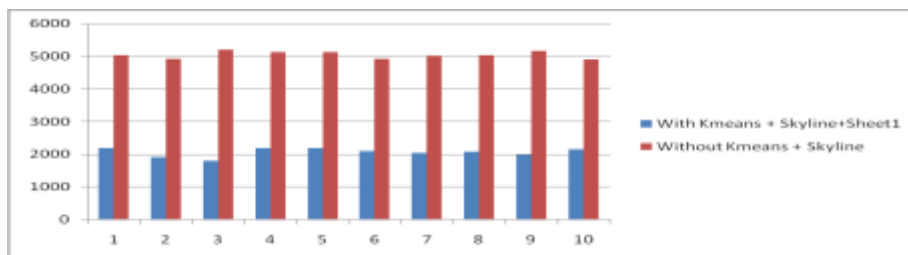


Fig 7: Time Comparison

4. CONCLUSIONS

Generally we cannot retrieve data which is present at different electronic device such as smart phone, smart television, tablet pc etc. by single device. So that our aim is to retrieve that data by single device. This work implement a technique that collect the data which is present in different source and find the record . This will help the user for searching record as the always want .In proposed system we apply skyline technique over present data set for finding efficient result .So our proposed system will give more efficient result than existing system with fixed algorithm . Instead of using those algorithms we are using different one and will provide the better outcome having more advance service than existing system.

5. REFERENCES

- [1]. Bo Yin, Zhe Zhang, Xi Wang, Zhiqiang Wei*" Research and Application of Data Mining Technology Used in the Analysis of Smart Home User Behavior" Sixth International Conference on Measuring Technology and Mechatronics Automation 2014.
- [2]. A.KrishnaKumar, D.Amrita, N.SwathiPriya "Mining Association Rules between Sets of Items in e Databases" International Journal of Science and Modern Engineering (IJISME) ISSN: 2319-6386, Volume-1 2013.
- [3]. IramFatima , Muhammad Fatima, Young-Koo Lee *" A Unified Framework for Activity Recognition-Based Behavior Analysis and Action Prediction in Smart Homes" 2013.
- [4]. Borgelt C, Kruse R. Induction of association rules: Apriori implementation[C]//Compstat. Physica-Verlag HD, 2002: 395-400.
- [5]. Bay vo, Bac Le "Fast Algorithm For Generalized Association Rule" International Journal of Database Theory and Application 2009.
- [6]. Rahul Australasian. Bharuka P Sreenivasa Kumar "Finding Skylines for Incomplete Data Proceedings of the Twenty-Fourth Database Conference (ADC 2013), Adelaide, Australia 2013.
- [7]. Mohamed E. Khalefa Mohamed F. Mokbel Justin J. Levandoski "Skyline Query Processing for Incomplete Data" This work is supported in part by the Grant-in-Aid of Research, Artistry, and Scholarship, University of Minnesota, DTC Digital Technology Initiative Program, University of Minnesota, and DTC Intelligent Storage Consortium (DISC), University of Minnesota
- [8]. Kian Lee Tan ,Pin -KwangEng "Efficient Progressive Skyline Computation" proceeding of the 27th VLDB Conference Roma, Italy 2001.
- [9]. Yufei Tao, Dimitris Papadias, Greg Fu, Bernhard Seeger "Progressive Skyline Computation in Database Systems" To appear in ACM-TODS, special issue on best of SIGMOD/POD2003.
- [10]. Liu, H. & Zeleznikow " Intelligent computation for association rule mining." In F. Berzal, J. Cubero, Z. Ras, T. Sudkamp & R. Yager (Eds.), IEEE International Conference on Data Mining (pp. 49-53). Houston, Texas: IEEE Computer Society