

Predicting User Behavior using Enhanced Data Mining Approach for Prediction in Smarthome

Ms Sonam Sheikh

Department Of Computer science &
Engineering
Nagpur Institute Of Technology, Nagpur
sonamsheikh0@gmail.com

Prof. Abhijeet Bajpayee

Department Of Computer science &
Engineering
Nagpur Institute Of Technology, Nagpur
abhijeet_bajpayee@live.com

Abstract: A fully fledged smart home system contains a wide variety of household appliances. With the popularity of smart home, there will emerge sufficient data in our everyday life. How to use these existing data to analyze users' behavior and how to provide personalized services to different, seems to become a more and more important issue. We proposed a system that will gather all the information a human performed, will apply good clustering that cluster all relevant information of the user.

Keywords: *Data Mining, User Behavior Analysis, K means algorithm, clustering.*

I. Introduction

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.

Data mining is the process of extracting interesting (non-trivial, 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.

II. Literature Survey

Bo Yin and all in 2014, presents "Research and Application of Data Mining Technology Used in the Analysis of smart Home User Behavior" focus on At home and abroad, smart home, after more than ten years development is getting more and more attention and favor. With the rapid development of smart home, a huge amount of life data have been accumulated, and these data include a lot of information about user behavior, they can reflect the user's activity truly and objectively, but these data exist in different data sources, such as computers, smart phones, smart televisions. Technique use :- Apriori Algorithm

Apriori algorithm is a classical algorithm using association rules in data mining[7]. Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of data items with n different items. Given a transaction database D , there are m transactions and n data items in database D , the transaction T in D made by a subset of I .

A. Krishna Kumar and all represents "Mining Association Rules between Sets of Items in Large Databases" in 2013, deals with the In Data Mining. This paper proposes a new interactive 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 proposed by Liu et al. for user expectations

Ashok Savasere and all [3] in 1995 deals with "An Efficient Algorithm for Mining Association Rules in Large Databases" focused on Mining for association rules between items in a large database of sales transactions has been described as an important database mining problem. In

this paper we present an efficient algorithm for mining association rules that is fundamentally different from known algorithms.

Partition Algorithm:- The idea behind Partition algorithm is as follows. Recall that the reason the database needs to be scanned multiple number of times is because the number of possible item sets to be tested for support is exponentially large if it must be done in a single scan of the database.

Iram Fatima , Muhammad Fatima, Young-Koo Lee * and Sungyoung Lee in Sensors 2013, “ A Unified Framework for Activity Recognition-Based Behavior Analysis and Action Prediction in Smart Homes” proposed activity recognition in smart homes is an active research area due to its applicability in many applications, such as assistive living and healthcare. Besides activity recognition, the information collected from smart homes has great potential for other application domains like lifestyle analysis, security and surveillance, and interaction monitoring.

III. Problem Statement

- The problem definition is in the normal home , there are variety of home applying that can represent the behavior of the user.
- The smart home is a home that predict the nature of human by activity that he/she perform on electronic device with the help of information gathering in order to predict the his/her behavior according to that information.
- Here, the Problem is in gathering that information from the number of devices into to the single device and predict the human nature as it is difficult to collect all information and make the data list according to that on which human behavior will be predicted and also deciding the factors which will useful to classifying the human behavior .

IV. Proposed Approach

- We proposed a system that will gather all the information a human performed. In this system we will apply good clustering that cluster all relevant information of the user. After that we will perform efficient association rules that associate a gathered data.
- The smart home is a home that predict the nature of human by activity that he/she perform on electronic device with the help of information gathering in order to predict the his/her behavior according to that information

- Firstly we form cluster of data using k-means algorithm. After association we will apply some fuzzy theory that will predict a human behavior regarding human task that he/ she performed on difference devices according which will predict behavior of user using fuzzy logic. And finally we will analyze the result.

V. Methodology

The figure below shows the flowchart of proposed system.

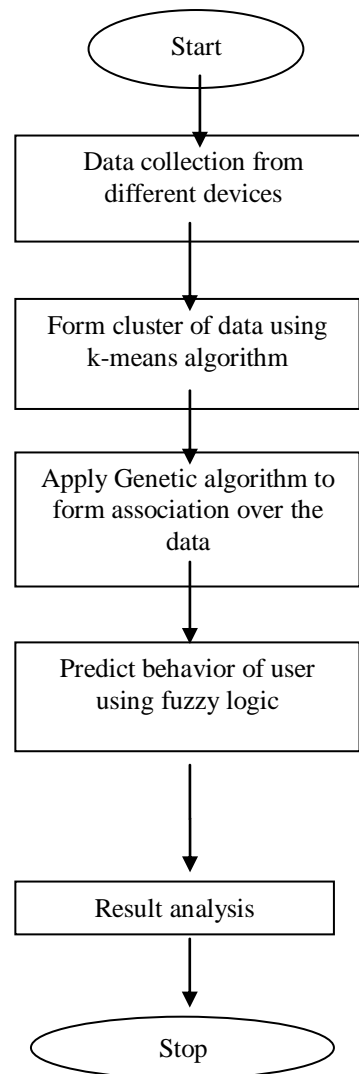


Fig. Basic flowchart of proposed system

For achieving proposed work we apply a k-means algorithm for clustering data. We will use genetic algorithm for forming association rules over the cluster. And we will apply some fuzzy theory to predict the human behavior.

k-means :- -It means clustering is a method of vector quantization, originally from signal processing, that is

popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

Genetic algorithm:- The computer science field of artificial intelligence, a genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. This heuristic (also sometimes called a metaheuristic) is routinely used to generate useful solutions to optimization and search problems.^[1] Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

Fuzzy logic :- Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values), fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false.^[1] Furthermore, when linguistic variables are used, these degrees may be managed by specific functions. Irrationality can be described in terms of what is known as the fuzzy objective.

VI. Tools for development and verification of result

SOFTWARE REQUIREMENT

1. Operating System : Windows- Service Pack2
2. Web Server : IIS Server 7.0
3. Language : ASP.Net, C#
4. Platform : Dot net 2010
5. Database : SQL SERVER 2008
6. Scripting Language : JavaScript
7. Web Tools : HTML, CSS
8. Web Browser : Microsoft Internet Explorer, Firefox

HARDWARE REQUIREMENT :

1. Processor : Intel Pentium IV
2. Clock Speed : 333 MHz
3. RAM : 512 Mb
4. Hard Disk Capacity : 40 GB

VII. Expected Outcome

By applying proposed scheme and algorithm we will get a better cluster that contain related information. And better association over that cluster and a technique that predict the good result of human behavior.

So our proposed system will give more efficient result than existing system which is using some fixed algorithm. Instead of using those algorithm we are using different one and will provide the better outcome having more advanced service than existing system

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” 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation IEEE 2014
- [2] A.KrishnaKumar, D.Amrita, N.Swathi Priya “Mining Association Rules between Sets of Items in e Databases” International Journal of Science and Modern Engineering (IJISME) ISSN: 2319-6386,Volume-1, Issue-5, April 2013.
- [3] Ashok Savasere Edward Omiecinski Shamkant Navathe “An Efficient Algorithm for Mining Association Rules in Large Databases” Technical Report GIT-CC-95- 04, Georgia. Institute of Technology, Atlanta. GA 30332, January 2000.
- [4] Iram Fatima ,Qiankun Zhao “Association Rule Mining: A Survey” Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116 , 2003.
- [5] C.W. Hirt, B.D. Nichols. Volume of fluid (VOF) method for the dynamics of free boundaries. Journal of Computational Physics, 1981, 39:201-225.
- [6] JAlex Berson. Stephen Smith Building Data Mining Applications for CRM .McGraw-Hill Companies, 2000.
- [7] L.A.Zadeh. Fuzzy sets and in formation granularity. Gupta M M, Ragade R K, Yager R R.Advances in Fuzzy Set Theory and Applications. North Holland: Amsterdam, 1979: 3-18.
- [8] Han J.Car Y. Cercone N. Knowledge Discovery m Data base; An attribute oriented approach[A]. In,Proceedings of the 18th VLDR Conference [C], Canada Bruish Columba,Vancouver, 1992.
- [9] Han Jiawei, Kamber Miceline. Fan Ming, Meng Xiaofeng translation. Data mining concepts and technologies,Machinery Industry Press, 2001.
- [10] Agrawal R.Imielinski T.Swami A. Mining association rules between sets of items in large databases, Proceedings of the ACM SIGMOD Conference on Management of data, pp. 207-216, May,1993.
- [11] Wu Jianhua, Song Qinbao, Shen Junyi. A novel association rulemining based missing nominal data imputation method [C]//Proceedings of the Eighth ACIS International Conference onSoftware Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing. 2007:44-249.
- [12] Agrawal R,Imielinski T,Swami A.Mining Association Rules between Sets of Items in Large Databases[C].In:Proc of the ACM SIGMOD International conference on Management of Data, Washington DC,1993: 207-216.

- [13] Agrawal R, Srikant R. Fast algorithms for mining association rules[A]. Proc. 1994 Int'l Conf. Very Large Data Bases (VLDB'94) [C]. Santiago: 1994: 487-499.
- [14] Ball, Robert, and Chris North. Analysis of user behavior on high-resolution tiled displays. Human-Computer Interaction-INTERACT 2005. Springer Berlin Heidelberg, 2005: 350-363.
- [15] Borgelt C, Kruse R. Induction of association rules: Apriori implementation[C]//Compstat. Physica-Verlag HD, 2002: 395-400.