International Journal of Research in Engineering, Science and Management
Volume-2, Issue-2, February-2019
www.ijresm.com | ISSN (Online): 2581-5792

Transport Information Retrieval using Data Mining Technique

S. Ashika[1], V. Anitha[2], M. Jeevitha[3], R. Sathyadevi[4], M. Umapathy[5]
[1,2,3,4]*UG Student, Department of Computer Science and Engineering, TKEC, Salem, India*
[5]*Assistant Professor, Department of Computer Science and Engineering, TKEC, Salem, India*

***Abstract*: In this paper, we propose a design for college transport management system in hierarchical structure which contains (1) Web display layer (2) Business logic layer (3) Data mining layer (4) Database layer. It contains the entire knowledge of bus information, driver information, student information, staff information, routing, stoppage information and location information of the bus using the tracking device. Furthermore, the system is designed effectively to retrieve the information using data mining techniques of clustering and association and also using the Microsoft SQL Server 2005 as data mining tool.**

***Keywords*: College Transport Management System, Hierarchical Structure, Information Retrieval, Data mining techniques, Microsoft SQL Server 2005**

1. Introduction

The growth of the colleges partially depends on their transport system to bring the students from far away locations to the college spot. So, the Continuous scale expansion of the strength of students in the college, have been made the college Transport management system more complicated to handle bulk amount of data about the students, staffs, drivers and buses[1]. In the 21st century, information technology is an important and irreplaceable work currently in the construction of modern universities [2].In general, with the increasing number of students in college ,student transport management systems design is also important[3]. Considering the current situation in college transport management process , transport system is of great importance to satisfy the new requirement of the development of transport management [4].Hence, it is very urgent to develop a practical and effective college transport management system using technological information .The destination of our research is to effectively solve the prominent contradiction and problems between the heavy workload and the scarcity of human resources in current college transport management system[5]. Most of the system only consists of normal database querying technology and it could only utilize time and the information which is already known and relevant data as output. However, considering this current situation, it is very complex to utilize and retrieve information from the college transport management system. Hence, in this paper, we aimed to design and develop college transport management system to efficiently retrieve the information of the system like drivers log, students and staffs information, routing and stoppage details with bus information using data mining techniques.

Clustering and association techniques of data mining has been using to group up the students, staffs and driver into to one group with the characteristics of coming from same location. Association is used to relate the relationship among drivers, buses, staffs and students.

2. Literature survey

In [1], Liangquit MENG discussed about the student management system in hierarchical structure with various logical layers and designed and developed management system with functional modules of users and admin .It uses database layer with normal querying of information from bulk amount of stored data. Majd Ghareeb, designed an efficient bus system with tracking them using tracking devices [5]. According to [6] Changxin Song, When the data is large and continuous in nature then the traditional approaches of mining are not applicable because the real time data is quickly changing and requires fast response as well. Random access to data stream is very expensive so, a single access to streaming data is provided. Also the storage needed is very large. Therefore, clustering and mining techniques for stream data are required [7]. In [8], Ping YU applied data mining techniques on management system to find the relevant and individual services using clustering and association techniques with Microsoft SQL Server 2005 as a data mining tool. Thus it is required to design a transport management system with efficient retrieval of information from the bulk amount of data.

3. Existing system

The system consists of design of transport management system with the hierarchical structure of the following layers (1) Web display layer (2) Business logic layer (3) Data access layer (4)Database layer. In which data access layer consists of Database access interface and database access class. In this existing system, the data access layer retrieve the information of the transport management system whenever needed but it can only query the known relevant data as an output. Select statements are used to retrieve data from SQL tables of the database and Select * is used to retrieve all data fields from the table where Select data field name from table name used to

retrieve particular information. Querying can be based only when the user knows exactly what he is looking for and it can take some efforts to find the exact output as because of SQL querying statements .The most time consuming task is to prepare a normalized data-set from relational database, which is suitable for analysis. In general the database has group of many tables and views that must be joined, aggregated and transformed in order to build the required dataset. It results most long and complex SQL queries written several times independently and in random manner. Therefore, the database develops with many tables and views that are not present as entities in the ER model and similar SQL queries are written multiple times, which complicate the database management, software development and maintenance and this way makes the system insufficient to find the hidden information of the transport management system difficult to find the patterns and relationship among the member of transport management system. The existing system is figured as in Fig. 1

In computer software development, business logic layer (BLL) refer to the part of the program which is able to encode the real-world business rules. Hence, developer can know how to create and display data. In the business logic layer, the software can be correlated with lower-level details of managing a database or displaying the user interface, system infrastructure. On the other hand, data access layer (DAL) refers to a computer program which supports simplified access to data memorized in persistent storage.

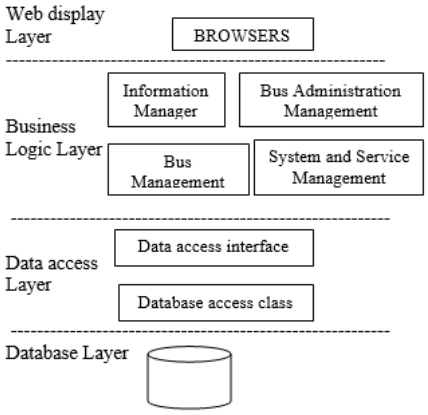

Fig. 1. College transport management database

4. Proposed system

As shown in Fig. 2, our proposed college transport management system is organized by a hierarchical structure, including

- Web display layer,
- Business logic layer,
- Data mining layer and
- Database layer.

The efficient system for college Transport management system has been designed and developed with the technology of data mining as a data access layer between database and Business logic layer from the existing system. Data querying involves retrieving a part of datasets of the existing data as specified by the user but Data mining is mainly about analysis and deals with extracting useful and previously unknown information from raw data. It tries to build models and discover patterns. So it is much more than mere retrieval of data. Often, raw data is stored as databases. Therefore, data querying being a part of the data mining task. But querying can be used only when the user knows exactly what he wants, while data mining is used when the user has an unclear idea and wants to find more interesting features and patterns in the data. In this proposed system, we proposed college transport management system to efficiently retrieve the information of the system like drivers log, students and staffs information, routing and stoppage details with bus information using data mining techniques. The system using Clustering and association techniques of data mining to group up the students, staffs and driver into to one group with the characteristics of coming from same location. Association is used to relate the relationship among drivers, buses, staffs and student. The system also group up the similar characterized students with the similarities of exceeding the due date of pass validity. The students who is not yet payed the bus pass fees can be notified by the admin.

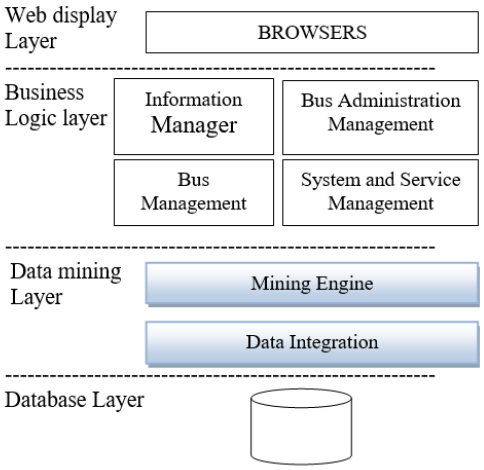

Fig. 2. College Transport Management System

5. Data mining process

The data mining process has been done on the database of the college transport management system. Although data mining regarded as a basic step of database knowledge discovery process is more scientific, but in the industry, the media and the database research field, directly as a data base knowledge discovery is more prevalent. Therefore, data mining has a wider concept: it is a process to extract interesting knowledge from a lot of data stored in databases. Based on this view, a typical data mining system may consist of the following main components, shown in fig 3. It is noted that data mining technology from the outset oriented applications. It is not only simple search, queries and call for specific database, but also micro and macro statistics, analysis, synthesis and reasoning for these data, to

guide for solving practical problems, attempt to find interrelation between events, even use the existing data to predict future events. As a result, increase people on data applications, from the low-level end query operations to decision support for business decision-makers at all levels.

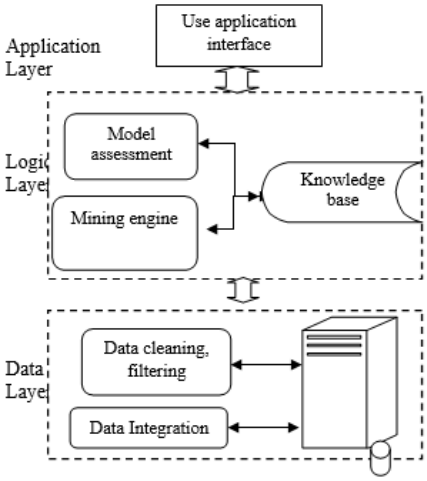

Fig. 3. Data Mining System Structure

6. Data mining in college transport management system

The College Transport management System is based on Microsoft SQL Server 2005.It is selecting the data which related with transport information as a data source, including drivers, buses, students and staffs information.

A. Analysis member's characteristics by clustering.

In this paper, k-means clustering algorithm is one of the Clustering algorithms which is used in analyzing the member's characteristics. K-means algorithm procedure follows a simple and easy way to classify a given dataset through a definite number of clusters (assume k clusters) fixed a priori. The main aim is to define k centroids, one for each cluster. These centroids should be placed in a designing way because of different location makes different result. So, the best way is to place them as much as possible far away from each other. The next step is to take each point corresponds to a given data set and associate it to the nearest centroid. When no point is left, the first step is completed and an early group page is done. At this point we need to re-calculate k new centroids as main centers of the clusters resulting from the previous step. After we have these k new centroids, a new blending has to be done between the same dataset points and the nearest new centroid. A loop has been generated. As a result of this loop we may find that the k centroids change their location step by step until no more changes are done. Finally, this algorithm aims at minimizing an objective function. Selecting circulation data as mining data sources, using the Microsoft Clustering algorithm to create a mining model. Viewing the model, it is characterized by various types of categories as students, staffs, drivers and buses.

B. Finding out relationship among member's by association rules.

Use association rules to find out the member's relationship among each other. The association rules algorithm is finding frequent item sets (common attribute value sets). There are two steps in it. The first step of the algorithm, a calculation intensive phase, is to find frequent item sets. The second step is to generate association rules based on frequent item sets. An item set is a set of items. Frequent item sets are those item sets that are relatively popular in the dataset. The popularity threshold for an item set is defined using support. Support is used to measure the popularity of Internet. Support of an item set {A, B} is made up of the total number of transactions that contain both A and B.Support ({Apattern.rofransactions(A, B) Minimum_Support is a threshold parameter you need to specify before processing an association model. It means that we are interested only in those item sets and rules that represents at least minimum support of the dataset. Probability is a property of an association rule. The probability of a rule A=>B is evaluated using the support of itemset {A, B} divided by the support of {A}.It is defined as follows:

Probability (A => B) = Probability (B|A) = Support (A, B)/ Support (A)

Minimum_Probability is a threshold parameter we need to identify before running the algorithm. It means that the user is interested in only those rules that have a maximum probability rather than a minimum probability. Minimum_Probablity has no impact on itemsets, but it does impact rules. Importance can be used to measure itemsets and rules. The importance of an itemset is defined using the following formula:

Importance ({A,B}) = Probability (A, B)/(Probability (A)* Probability(B)))

If importance = 1, A and B are the independent items. It means that A and B are two independent events. If importance<1, A and B are negatively correlated. If importance>1, A and B are positively correlated.

For rules, the importance is evaluated using the following formula:

Importance (A => B) = log (p(B|A)/p(B|not A) an importance is 0 means that there is no association between A and B.

A positive importance score means that the probability of B goes up when A is true. Select reader information, circulation and borrowing categories as the mining data source, set the value of Mininum_Support and Minimum_Probability as 0.2 and 0.6, and use Microsoft Association Rules to create the data mining pattern.

7. Conclusion

This paper proposes a design for college transport management system with the technology innovation of data mining. The proposed system is organized in a hierarchical way, including: Web display layer, Business logic layer, Data mining layer, Database layer. It contains of entire knowledge of bus information, driver logs, student information, staff

information, routing and stoppage information and location information of the bus using the tracking device. Furthermore, the system uses Microsoft SQL Server 2005 as a data mining tool to effectively retrieve the information of transport management system by the data mining techniques of clustering and association.

References

[1] Liangqiu Meng, " College Student Management System Design Using Computer Aided System", International Journal Of Electrical Engineering Education, 2016.
[2] Cone Trey III, Redus Brady S, "Computer Based Health Management System and Overall Optimism in College Age Students Medicine and Science in Sports and Exercise", 2014.
[3] Horvat Ana, Dobrota Marina, Krsmanovic Maja, Cudanov Mladen, "Student Perception of Moodle Learning Management System: A Satisfaction and Significance Analysis, Interactive Learning Environments", 2015.
[4] Swart Arthur James, "Student Usage of a Learning Management System at an Open Distance Learning Institute: A Case Study in Electrical Engineering", International Journal of Electrical Engineering Education, 2015.
[5] Majd Ghareeb, Athar Ghamlous, Hawraa Hamdan, Ali Bazzi and Samih Abdul-Nabi "Smart Bus: A Tracking System for School Buses", International Journal of Electrical Engineering Education, 2017.
[6] Changxin Song "Research of Association Rule Algorithm based on Data Mining", International Journal of Electrical Engineering Education, 2016.
[7] Garima, Hina Gulati, P. K. Singh "Clustering Techniques in Data Mining: A Comparison", International Journal of Electrical Engineering Education, 2015.
[8] Ping YU, "Data Mining in Library Reader Management", International Journal of Electrical Engineering Education, 2011.