IJCAT - International Journal of Computing and Technology
Volume 1, Issue 1, February 2014
www.IJCAT.org

Effective Use of Pattern Discovery for Detection of Fraudulent Patterns in Railway Reservation

[1] Ashish Gaigowad, [2] Prachi Deote, [3] Prathamesh Badge, [4] Rahul Giradkar

[1, 2, 3, 4] Dept. of Computer Technology, Yeshwantraochavan college of Engineering, Nagpur, India

Abstract - Corruption is the biggest issue that is increasingly affecting India and making its economy weak. The corrupt activities adversely effects the people and makes their living cumbersome. Fraudulent activities are increasing every year. The development of new technologies has also provided further ways in which criminals may commit fraud. Data mining concepts and techniques can be used to control these frauds. Data mining is the process of extracting useful knowledge from a large set of data. This hidden knowledge may help decision makers to detect illegal activities. In this paper, we have proposed a mechanism which uses these data mining techniques to detect unusual patterns in railway reservation dataset. This will help Railways to take precautionary measures when something bad happens. In our study we have used association rule mining techniques to detect fraudulent activities like fake bookings and unusual patterns like ticket booking by same person of different trains departing at same time on same date.

Keywords **- Data mining, fraudulent transactions, Association rule mining, pattern discovery**

1. Introduction

The primary intent of our research is to evaluate the use of data mining techniques in railway reservation systems and to explore the ways by which data mining can be used in this system. The main aim of this research is to identify different techniques to detect fraudulent activities in railway reservation dataset and to report the frauds to the user so that corrective measures can be taken. Indian Railways is one of the world's largest railway networks comprising 115,000 km of track over a route of 65,000 km and 7,500 stations. In 2011-12 Indian Railways had revenues of US$18 billion which consists of US$11 billion from freight and US$4.6 billion from passenger's tickets. [6] The major problem which is faced by the Indian Railways passengers is the unavailability of tickets. In recent times, we have come across cases where the agents have illegitimately hacked into the reservation system and done fake reservations. These frauds are usually done with the help of fake ID's, fake multiple reservations etc. This results in unavailability of tickets for normal users. One more example of the frauds is the "tatkal ticket" scam which was exposed recently. These types of cases cannot be easily tracked by a person. Here we can make effective use of rule mining techniques.

Data mining refers to extracting or "mining" knowledge from large amounts of data. The term is actually a misnomer. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long. It has been described as "finding hidden information in a database. Alternatively, it has been called exploratory data analysis, data driven discovery, and deductive learning" [1]. The data mining process has several steps which mainly deals with data cleaning, transformation and pattern extraction. These steps are explained in the section 2.

"In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness." [2] When these rules are observed to be violated, it may be an indicator of unusual or fraudulent pattern. Here in our research we have identified some rules which finds the unusual patterns in the ticket reservation data. The users are immediately informed about these wrongly done reservations.

The rest of the paper introduces the complete description of data mining process followed by the main methodology for a fraudulent pattern discovery.

2. Methodology

2.1 Problem definition

Lately we have seen lots of fraudulent activities in the process of railway ticket reservations. These activities results into unavailability of railway tickets and causes inconvenience to the customer. In current scenario, the railway employee who is working with the reservation system is sole responsible to detect such fraud. The main aim of our research is to develop an automated system which is more accurate than conventional systems. This proposed system makes use of pattern discovery for detection of fraudulent patterns in Railway reservation dataset. Here we have made effective use of association rule mining techniques to detect the fraudulent and unusual patterns. The detected frauds in reservation may be reported to the user, so that corrective measures can

be taken. The overall problem can be defined as designing the association rules, detecting the fraudulent patterns and reporting these frauds to take corrective action.

2.2 Defining Data mining

Data Mining is also known as Knowledge Discovery in Databases (KDD). "Knowledge Discovery in Databases (KDD) is an automatic, exploratory analysis and modeling of large data repositories. KDD is the organized processof identifying valid, novel, useful, and understandable patterns from large and complex data sets. Data Mining (DM) is the core of the KDD process, involving the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns."[3]

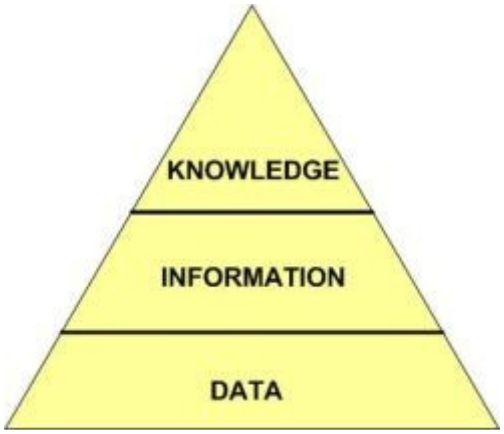

Fig. 1. Data, information and Knowledge pyramid

Thus, the use of Data Mining is intended to support the discovery of patterns in databases in order to transform information in knowledge, to assist the decision making process or to explain and justify it. According to Witten and Frank "Data Mining can be defined as an automatic or semiautomatic patterns discovery in great amounts of data, where these patterns can be perceived as useful". [3]The stages of data mining can be divided in three iterative stages: pre-processing, pattern extraction and post-processing.

2.3 Pre-processing

The data mining is usually carried out on large and complex data that are stored in one central database, dispersed through several databases, or distributed in many formats. The main tasks performed in pre-processing are:

1) Database Collection: Data collection is the most important step in application of any data mining algorithm. The quality of data governs the efficiency of the algorithm to be applied. Here for our research, we created the railway reservation database with the data collected from Indian Railways web site. The railway reservation transactional database mainly consist of following master files:

alltrain
pnrtbl
Train_info
Train_name
Agent_reservation
Agent_table

The overall structure of above database files is shown in following figure:

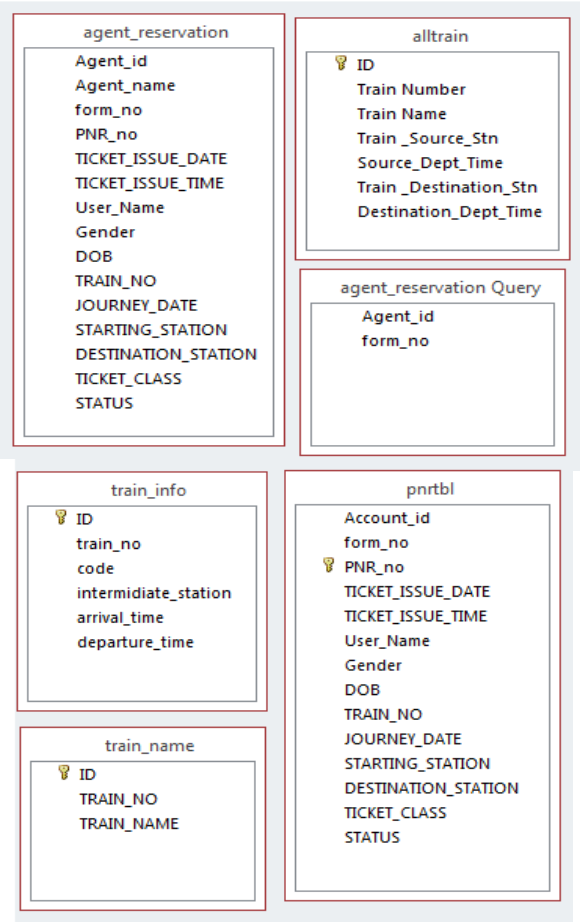

Fig. 2 structure of tables in the ticket reservation database

2) Data integration: For efficient application of data mining algorithms the dispersed data must be integrated and extracted into a single source. The data collected by us was in different formats like Text files and HTML pages. It was then integrated into a single access database.

3) Cleaning: In this process, the data is cleaned by filling in the missing values and correcting the wrong values of the attributes. For example in the railway ticket reservation database, we corrected the type of some attributes to make it easier to compare them during data mining. One more example of data cleaning is the filling

of missing values. For example, the missing values of intermediate station were filled.

4) Data reduction: Data mining on a large amount of data takes a very long time. Most of the time it is not needed to use all the available data items. Sometimes the use of the complete amount of data, or the use of all existing attributes can affect the processing and even the results in a bad way. Data reduction techniques can be used to reduce the size of the data. Here we have used the dimension reduction technique using which redundant dimensions are eliminated.

5) Transformations: During transformation data can be codified, normalized, enriched with external information etc. aiming at a better adjustment to the problem goal, data mining algorithms and techniques. In this stage, the quality of data is improved. In the normalization step, we have normalized the data up to second level of normalization.

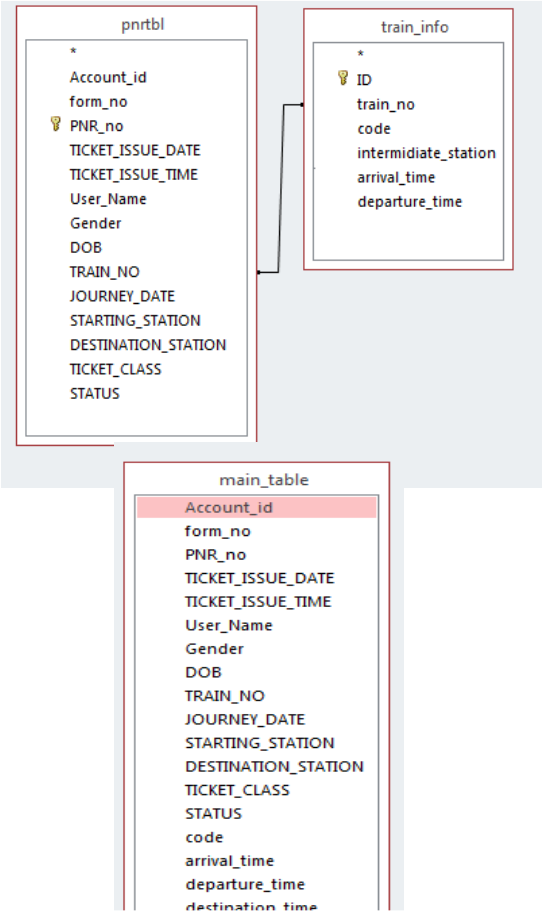

Fig. 3 Structure of main table

2.4 Pattern discovery

In this stage the data is mined to extract the rules that identifies the unusual and fraudulent patterns. Normally, the characteristic of the problem, its domain and expected results define the model to be used for the pattern discovery. "The models can be predictive or descriptive. A predictive Model calculates some value that represents a level in the future activity, a descriptive model discovers rules that are used to group items in categories". [4] Therefore, the problem classification, the expected results and the characteristics of the data must guide the choice of the mining techniques and the most appropriate algorithms. We are using the association rule mining technique. The association rule technique is related to descriptive models, it looks for the identification of patterns that occur simultaneously in collections of data.[5] After analyzing the railways reservation database, we have identified fraudulent activities in railways reservation system and from that we have designed various rules for detection of frauds. Some of the rules found are:

Rule 1) Ticket booking of different trains departing at same time.

Rule 2) Ticket booking of same person in different class in same train,

Rule 3) Tickets booked before opening of ticket counter.

Rule 4) Booking of tickets for two stations on same path.

Rule 5) Booking of two different trains tickets in busy time.

Rule 6) Booking of tickets by agents with serial reservation form.

Rule 7) More than 10 reservation on one form in one day

Rule 8) Ticket issued by agent between illegal timings.

The association rule mining is implemented by using Java. The mining process works in the background so that there is no need of any user interaction. So this makes the system completely automated.

Fig 6.1 GUI of the system

2.5 Post-processing

After the process of pattern discovery, in the last stage, the results are analyzed. Many times, different patterns are found in data and a deep analysis of the results is needed in order to filter the less important or very abnormal results. Usually, a problem domain expert conducts this analysis. In the post processing stage, the knowledge which is extracted in the pattern extraction stage is used. Here in our project we have designed a system which alerts the user if his/her ticket is booked by any illegal means. This system alerts the user by sending a message. The user is requested to take corrective actions within 24 hours. Otherwise his/her ticket will be cancelled.

3. Conclusion

Data mining concepts and techniques can help in solving many problems. This paper suggests that data mining should be studied and applied in railway management. Using these techniques we can effectively identify the fraudulent activities. These techniques used in real time can even proliferate the user experience by improving the interactivity. The main goal is to evaluate the usability and the efficiency of data mining algorithm in the context of the pattern discovery.

4. Scope and Limitations

This research only studies the application of a set of techniques that are used for fraud detection. Here we have not compared the performance and efficiency of various algorithms. As per our knowledge no extensive research has been done in this application area. So this can be seen as a cornerstone work done in Railway reservations. Data is a major factor in the process of achieving the objectives of data mining algorithms. So the scope of the research completely depends on the Data availability. Lack of dataset can limit the scope of this research.

Acknowledgment

We would like to thank Ms. Amreen Khan, Professor at YeshwantraoChavan College of engineering, Nagpur University, India for valuable comments on the first draft of this paper. However, the authors should be held responsible for any remaining errors.

References

[1] Margaret H. Dunham, “Data mining: Introductory and advanced topics”, Dorling Kindersley (India) Pvt.Ltd. Pearson, 2006.

[2] Pei Jian, Han Jiawei and Lakshmanan, “Mining frequent item sets with convertible constraints”, in Proceedings of the 17th International Conference on Data Engineering, Heidelberg, Germany, 2001.

[3] OdedMaimon and LiorRokach, “introduction to knowledge discovery in databases”, Department of Industrial Engineering, Tel-Aviv University.

[4] Jiawei Han, MichelineKamber, “Data Mining: Concepts and Techniques”, Second Edition.

[5] Rasika Ingle, ManaliKshirsagar, “An Approach for Effective Use of Pattern Discovery for Detection of Fraudulent Patterns in Railway Reservation Dataset”, IJCER march -2013.

[6] www.en.m.wikipedia.org/wiki/Indian_Railways