International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

SMART RAILWAY SYSTEM USING TRIP CHAINING METHOD

Prem Kumar.T[1], Raaghul Koushik.S[2], Sanjay Srinivas.H[3], Vignesh[4], Senthil.R[5]

[1-4]UG Scholar, Department of Information Technology, Valliammai Engineering College, Chennai-603203,INDIA
[5]Assisstant Professor, Department of Information Technology, Valliammai Engineering College, Chennai-603203, INDIA

Abstract – *This paper proposes a system that makes ease in traveling by metro train through use of smart cards. These smart cards hold the user data that can be used for transactional purposes at the railway terminals. Here the user data means the passengers name, age, time, source and destination station. The trip traveled by the passenger is recovered through their trip chaining behavior. The data read from these cards will be in tremendous amount so use of big data technique is the most appropriate solution for it. After all the details are fetched from these smart cards, these data are collectively stored into databases where the k-means clustering of passenger data is made that is it's like how many passengers of same age have travelled from the source at same time? How many passengers of same age have the destination at same time? How many people of same people of same age are travelling? Then finally naïve bayes classifier determines passengers count at each source and destination stations; the deep insights of passenger data in each station helps the railway department to improve their existing infrastructure and can facilitate personalized services for passengers.*

Key Words: Metro systems, Smart Cards, Trip Chaining, Analysis, Prediction, Big Data

1. INTRODUCTION

Metro Systems is going to be ultimate passenger traveling service in the upcoming days. As the future human population is unpredictable the demands from the passengers side will be rising inch by inch and their expectation and priority will be skyrocketed at that time. So, the smart railway system facilitates efficiency in conducting user surveys and personalized services for travelers through the analysis and prediction of user data from the data card. However, in current scenario, there are two ways in which the passenger can travel by train. First, the user will buy a coin that can be bought at each source destination and drops it at the destination station. Second, the user can buy trip cards that can be used for monthly or weekly basis based on their preference. But in both of these scenarios the obscurity of the passengers data still remains and this is one of the major reason why our railway department have not been developed in the past decades. So the alternative system aims to track every individuals information through usage of smart cards. These smart cards are maturity day by day in all aspects as they are contactless, secure and can be oriented in any direction with the reader. It holds the passengers information such as their name, age, time, source and destination stations. Since every passengers have their trip chaining behavior in traveling from the source and destination stations their original trips are recovered by the analysis in the subsequent steps. The collected data have been clustered on the basis the age, time at source and destination stations. Then, the clustered data has been predicted to determine the passengers count. Therefore, from the final counts of passengers aids the railway department in sorting out the existing systematic problems and can also be helpful in the development of railway infrastructure and services.

1.1 OVERVIEW

For the railway department to improve their existing infrastructure and services is to collect all the user data and to analyze it at a broader range. The valuable parameters such as their age, time at source and destination station should be analyzed and their count has been predicted to be used for surveys, upgrading building and developing new infrastructures.

1.2 APPLICATIONS

The resultant data after analysis and prediction has various applications. Few of them are listed below:

- The collection of user data from smart cards can be used for conducting user surveys in a efficient way.
- The resultant passengers count can be used for the railway department to upgrade their existing infrastructure and services.
- The system on implementation will be a complete automated system and doesn't require much manpower resources

1.3 ALGORITHMS PROPOSED

This section describes the methods used for clustering and classification of passengers data which are k-means clustering for clustering the passengers data and naïve bayes classification to determine the passenger count from the clustered data. A trip chaining method which retrieves the origin and destination trips of each traveller.

1.3.1 K-MEANS CLUSTERING

K-means clustering is a well-known clustering algorithm suited for clustering similar kind of data. It is a kind of unsupervised learning method. This algorithm has a distant relation with k-nearest neighbor classifier which is popular classification technique used in machine learning.

1.3.2 NAIYE BAYES CLASSIFICATION

Naïve Bayes is conditional probabilistic model that is mainly used to determine probabilities of unknown data based on some available data. The advantage of using naïve bayes is because of its scalability and it uses a linear time rather iterative approximation used by other classifiers.

1.3.3 TRIP CHAINING METHOD

Trip chaining is the travel pattern of passengers exhibited by each passenger on a daily routine. Every passenger has their own trip chaining behavior be it daily or weekly travelers. So every passenger before reaching the destination from the source has to travel through series of stations then only they can their desired destination. In the mid of their travel if any passengers detrains in any between stations in their journey then the journey can't be taken as the successful trip.

2. LITERATURE SURVEY

2.1. OVERVIEW

a) Big data challenges in railway engineering

As Big Data becomes part of railroad data analysis, there are many challenges which need to be addressed by the railway industry. This extended abstract highlights some of the challenges from specific examples in railway engineering. This work does not present the challenges of dealing with Big Data in general which is beyond the scope of this paper. The examples provided in this extended abstract cover both the engineering and the management of railroad applications.

b) Large Scale Data Analytics in Transportation and Railway Infrastructure

Currently the increased use of sensors, imaging systems and other emerging techniques in transports and railway infrastructure testing, monitoring and control are enabling massive amount of information and data to be generated at unprecedented scales. Data are generated in such volumes making it very difficult to draw appropriate conclusions. Processing of these large sets which are the magnitude of terabytes to petabytes demand new tools, and new algorithmic, probabilistic and statistical techniques. Despite the advance in large storage, high computational power, mining and drawing inferences from large scale data in critical infrastructure has some challenges: (a) standardization of data format, (b) accurate modeling, (c) clustering and classifying, (d) integrating data from independent sources and finally, (e) uncovering hidden patterns and information, (f) hidden correlation and (g) interpretation.

c) Research on Cloud Computing and Its Application in Big Data Processing of Railway Passenger Flow

Modern railway has a high speed, heavy load and intensive development trend. It not only brings the opportunities of railway transport capacity and volume, but also makes the data of various types of large-scale continuous growth. In the stage of the development of the railway by the production enterprise to the service enterprise, it is necessary to vigorously promote the application of cloud computing, and to plan the cloud computing framework. We study a parallel support vector machine model based on multi-level SVM, and realize the parallel algorithm in cloud computing environment. The algorithm divides the large training data set into a number of small training sets by Map, and then a new SVM is combined with these small training sets. Finally, the data is trained to be a new SVM by Reduce. At the end of the paper, we use the SVM parallel forecasting method to predict the passenger flow of China Railway, and compare the performance of the distributed with that of non-distributed algorithms. Experimental results show that the proposed algorithm has better effect than single machine algorithm in terms of time consumption and classification accuracy. With the increase of nodes, the time consumption is significantly shortened.

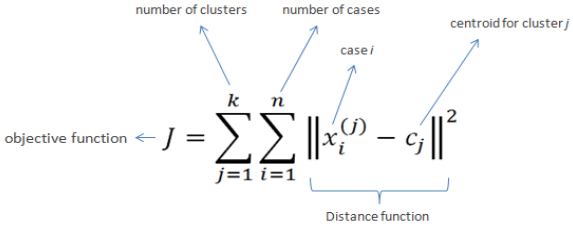
3. PROPOSED METHOD

For the clustering and classification of passengers data we are using a hybrid approach for improving the accuracy and prediction. In this way, first the passengers data such as their age, time at source and destination stations are clustered at the first stage in which we get half of the accurate clustered data. In the second stage, the clustered data are again passed into the classification in which the passengers count in the source and destination station is determined from the clustered data therefore the resultant data can be helpful for the railway department to improve their railway infrastructure and services.

Algorithm proposed:

The two algorithms which we use in our system is mentioned in the following steps:

Step1: K-means clustering approach clusters the passenger data into k clusters that are fetched from each of the smart card holder. The steps involved in k-means clustering are as follows:

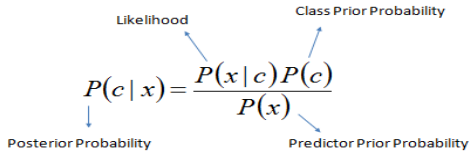


- Clusters passenger data into k clusters where k is pre-defined
- For selecting centroids choose k points in random
- Assign each data point to their closest centroid based on the Euclidean distance function
- After assigning all data points to each centroids, re-calculate the mean of all data points in each cluster

- Steps 2, 3 and 4 are repeated until same points are assigned to each cluster in consecutive iterations.

Step2: Naïve Bayes Classification based on the probabilistic model classifies the clustered data from the k-means clustering determines the probability of an event based on the probability of an event that has been already occurred.

Bayes theorem calculates the posterior probability, $P(c/x)$, from $P(c)$, $P(x)$, and $P(x/c)$. It works on the assumption of class conditional independence that is the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors



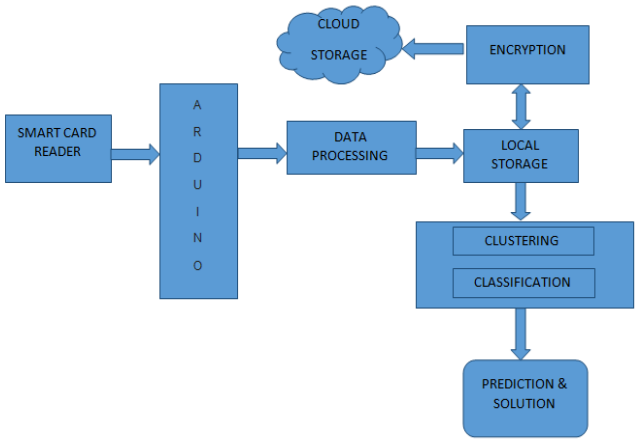
$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- $P(c/x)$ - posterior probability of *class* $P(c)$ given *predictor* p(x)
- $P(c)$ - prior probability of *class*.
- $P(x/c)$ - likelihood which is the probability of *predictor* given *class*.
- $P(x)$ - prior probability of *predictor*.

Advantages of proposed system:

- The proposed hybrid approach of these above methods very well improved the accuracy and prediction.
- Very well suitable for large datasets and scalable systems.

4. ARCHITECTURE DIAGRAM



The above architecture diagram depicts the total functionality of the system. Every passenger taps-in their smart card over the smart card reader at the entry terminal of the railway station. On reaching the destination they again taps-out their card at the exit terminal. The trip-chaining method retrieves the origin and destination trips.

Now the details will be read by the arduino microcontroller and are passed on to the system using RS232 cable for voltage compatibility (+5V to+12V). The details will now fed into the system using COM3 port. After, all the unstructured data will be processed and stored in tabular format in database. The user details are also get encrypted in this stage and simultaneously uploaded to cloud for backup.

The user details imported as an excel file from the database and fed into clustering algorithm where the user details such as their age, time at source and destination stations are clustered and subsequently fed into naïve bayes classifier which determines the passenger count from the clustered data. The resultant data is finally predicted for improving the railway infrastructure and services.

4.1 MODULE DESCRIPTION

4.1.1 SMART CARD AND READER

The RFID card also known as smart card is a chip that works on the mechanism of electromagnetic fields. This chip holds the transactional data of every passenger with it. The transactional data like passengers name, age, time, source and destination stations. The passenger who has this smart card holds over to the smart card readers at the source stations and again taps-out their card at the destination station. In this way, passenger's details are read; these details fetched are transmitted to the microcontroller through a transmitting antenna.

4.1.2. PROGRAMMING MICROCONTROLLER

The information being transmitted by the smart card reader is received by a receiving antenna in the microcontroller. The microcontroller work is to identify each and every passenger and to fetch the user details such as name, age, time, source and destination stations. The details fetched are transferred to the system through a RS232 cable which is +5v to +12V converter used for voltage compatibility.

4.1.3. DATA COLLECTION

The data transferred from the microcontroller through the cable is received in the COM3 port which is a port used for serial communication purpose. After reading the data from the port the user details are processed into a proper tabular form that gets stored in the database and therefore that can be imported as excel file in .csv format and are supplied for analysis and prediction.

4.1.4 ENCRYPTION

The user details fetched are encrypted using the AES encryption mechanism which is the advanced and most popular form of cryptographic technique. With use of this mechanism the user details are properly secured and can avoid security breach instantly.

4.1.5 CLOUD STORAGE

The user details which got stored in the database are uploaded to the cloud storage by the admin. Because, storing in a cloud provides a backup and can avoid data loss and ensure accessibility from anywhere.

4.1.6 ANALYSIS AND PREDICTION

The imported excel file is now fed into the k-means clustering where the every passenger record is now clustered on the basis of passenger's age, time at source and destination stations. subsequently fed into naïve bayes classifier which determines the passenger count from the clustered data.

5. CONCLUSION

This paper proposes a hybrid approach for clustering and classification of passengers data by using smart cards. The passengers data like their age, time at source and destination stations are clustered then they are classified into passengers count, the resultant data is used by the railway department for improving their railway infrastructure and services.

REFERENCES

[1] Abu Abraham Mathews, Amal Babu," Train Ticketing System Using Smart Card", vol.3, Issue.3,2014.

[2]Ayushi Gaur, Nidhi Tawara, Dharmveer Singh Rajpoot," Passengers Segmentation For Metro Run Using Smart Card" vol.2, Issue.8, 2015.

[3] Fan Zhang, Juanjuan Zhao, Chen Tian, Chengzhong Xu, Senior Member, IEEE, Xue Liu, Member, IEEE and Lei Rao," Spatiotemporal Segmentation of Metro Trips Using Smart Card Data", IEEE transactions on vehicular technology, vol.65, no.3, March 2016

[4]Juanjuan Zhao, Fan Zhang, Member, IEEE, Lai Tu, Chengzhong Xu, Fellow, IEEE, Dayong Shen, Chen Tian, Xiang-Yang Li, IEEE and Zhengxi Li," Estimation of Passenger Route Choice Pattern Using Smart Card Data for Complex Metro Systems"19 Apr 2016

[5]Le Minh Kieu," Passenger Segmentation Using Smart Card Data", ieee transactions on intelligent transportation systems, vol.16, no. 3,june 2015

[6]Somnath Sarkar, Dipankar Chatterjee, "Biometric Ticketing System for Railway", Vol.4, Issue.3, 2014.