

Smart Borders

Manmeet Kaur
M.Tech Student
Department of computer science
J.S.S Academy of Technical Education,(Noida) India
Manmeet.kaur05@gmail.com

Abstract— Securing the Indian border and transportation is one of the critical issues of our country. To create “*smart borders*” it is required to utilize the information from multiple agencies like police record management system, vehicle registration agencies, driving license agencies, passport authorities, border check post so that the border management agencies have a software tool that will provide accurate and relevant information related to every vehicle crossing the border and this will also save the time of border check post agents as well as people who are waiting to cross the Indian border. In this paper concept of Apriori algorithm has been used to identify the criminal/illegal vehicles that crosses the border frequently an at last pattern analysis has been performed to identify that what pattern is followed by vehicles crossing the border.

Index Terms— Apriori algorithm, Association rule mining, Data mining, Pattern analysis.

1 INTRODUCTION

Data mining is becoming a useful tool for detecting and preventing terrorism. Data mining is the process of posing queries and extracting useful pattern often previously unknown from large amount of data. Recently there has been much interest on exploring the use of data mining for counter-terrorism application. Border and Transportation security is one of the issues of counterterrorism where data mining can be applied.

Safeguarding the Indian borders is critical for security of nation. There could be threats at the borders from illegal immigration to gun and drug trafficking as well as human trafficking to terrorists or criminal entering the country. Millions of vehicles cross the Indian borders each year. These vehicles are monitored at the border check post for the nation’s safety and security. Vehicles are monitored thoroughly for drugs and other contraband goods. If the lines at the border become too long the whole procedure become too time consuming [2].

Border check post agents monitors vehicles entering/leaving the country, recording driver name, driving license number, vehicle registration number, chassis number, goods carried, passport number along with the date and time of entry. In order to identify the vehicles that are involved illegal activities, it is required to utilize the data available in multiple agencies (like police record management system, vehicle registration record, D.L registration records etc). Vehicle crossing the border is having any criminal record can be verified using data available in police record management system. Similarly vehicle is authorized can be identified using vehicle registration data. Person is having valid D.L can be identified using D.L registration details. Similarly persons others details can be verified using information available in other law enforcement agencies. Sharing of data between multiple law enforcement agencies is very essential to help identify the illegal vehicles so that illegal activities across the border can be controlled.

In this research work a software tool has been developed that will be deliver at the border check post, which will reduce time and effort needed to check vehicles at the border and whether to allow them to cross the border or not. We have

proposed to perform association analysis using the concept of Apriori algorithm to identify the vehicles that crosses border frequently. At the end pattern analysis has been done to analyze the pattern of vehicles crossing the border.

2. LITERATURE REVIEW

A. Data Integration

To identify whether a vehicle is involved in any illegal activities and whether to allow it to cross the border or not, it’s very important to utilize the information from multiple sources. Integrating the data from these independently developed sources is a very difficult task [1]. There can following difficulties associated with integration of data from multiple agencies [1][2]: *name difference*: it is possible that same entity have different name in multiple source databases, *missing data*: incomplete data or different data may be available in different databases, *object identification*: lack of global identifiers.

Although it’s very difficult to integrate multiple data sources, but it’s the most important phase of this work. Because by utilizing the information available in multiple databases we can identify, whether a vehicle trying to cross the border is having any criminal record and whether it should be allowed to cross the border or not.

B. Apriori

Association rule mining is one of the most important and well researched techniques of Data mining. It was first introduced in [4]. The main aim of association rule mining is to find the interesting correlations, frequent patterns, and associations among sets of items in the transaction databases. It finds out association rules that satisfy the predefined minimum support and confidence from a

given database. There are many association rule mining algorithms, out of them Apriori is most commonly known algorithm.

Apriori algorithm was first proposed by R.Agrawal and R.Srikant in 1994 for mining the frequent item sets. The name of the algorithm was based on the fact that algorithm uses the prior knowledge of the frequent itemset. Apriori employs an iterative approach known as level-wise search, where k itemsets is use to explore $(k+1)$ -itemsets. First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support [5][7].

Apriori algorithm has been successfully applied as "market basket analysis" in retail stores. It helps in identifying the purchase behavior of customers, which items are frequently purchase together by them on a visit to a retail shop. Based on the analysis shopkeeper can decide which things can be placed together, and which item to put on sale so as to increase the profit margin [6].

Classical Apriori algorithm suffers from the problem that it requires large number of database passes which was equal to size of frequent itemsets. This whole procedure was very time consuming. To solve this problem, more efficient form of Apriori algorithm was proposed by Goswami D.N [7]. This approach requires less time as compared to classical Apriori algorithm.

To find whether vehicle is a frequent vehicle crosser or not concept of Apriori algorithm proposed in [7] is applied on the historical border crossing database. Apriori algorithm is applied to scan the database to search for the frequent 2 -itemset (i.e. vehicle no. and driving license).

3. RESEARCH DATASET

A. DATASET

For creation of smart borders the most important thing is data sharing. To identify whether a vehicle is involved in any illegal activities and whether to allow it to cross the border or not, it's very important to utilize the information from multiple sources. For this research work data from multiple sources has been utilized. Dataset for this research work includes information from police record of the different Indian states, state wise vehicle registration database, state wise driving license database, border crossing data from various Indian international border, border crossing status data, passport department. Since all this data which was required in this research work was highly sensitive and due to the security reasons all this data is not openly available in our country. So to carry out this research work we have created dummy. Police record contains criminal incidents from 2010 to 2012. Police record contains information related to people and vehicles involved in some illegal activity.

Border crossing data includes information related to vehicles crossing the various Indian borders (like indo-pak, indo-china, indo-Nepal etc). Border crossing data includes the name of driver, his driving license no., DOB, vehicle type, vehicle number, chasis number, date, and time for crossings, goods carried, supporting document for the goods carried, citizenship, name of the border crossed and the passport number. Border crossing database contains the record from 2010 to 2012. Whenever vehicle try to cross the border, check post agents at border monitors the vehicles and record all the above details related to the vehicle in border crossing data. Along with the border crossing data, border status database is also maintained that contains the status of vehicle trying to cross the borders. Border status database contains list of illegal vehicles seized from crossing the border and vehicles allowed to cross the border.

B. RESEARCH DESIGN

Fig. 1 shows the design of the software tool proposed in this research work. This software tool will be delivered to the border check post agents at the various Indian borders. Whenever vehicle try to enter/leave the country its details are entered in the software like name of the person, his driving license no., his DOB, vehicle type, vehicle no., chasis no., date and time of entry, goods carried, supporting document for the goods carried, citizenship, name of the border, passport number of the person. All these details (related to every vehicle crossing the border) get stored in the border crossing database. And also these details are verified by utilizing the information from multiple sources to identify that whether to allow vehicle to cross the border or not. Name of the person, DOB, driving license no. or vehicle no. is verified in police records of different Indian states to find whether that person/vehicle crossing the border is having any criminal record or not. Vehicle no and chasis no. are verified by utilizing vehicle registration database and this helps in identifying whether vehicle is legal or is some theft vehicle, then D.L no is verified by using D.L database and then name of the person, DOB, passport no. is searched in passport database to identify that person is carrying a legal passport or not. It is also very important that person must have supporting document for the goods carried. Because it is possible that person might be carrying those things illegally or might be carrying some contraband items. After verifying all the details of the person by utilizing the information from multiple databases, software tool helps the border check post agents to decide whether to allow that person to cross the border or not. Status related to each vehicle trying to cross the border is stored in border status database.

Then Apriori algorithm is applied to identify vehicle trying to cross the border is a frequent border crosser or not. The output of the Apriori algorithm is used to maintain a list of frequent border crossers. At last pattern analysis is done to analyze the behavior of vehicles crossing the border.

having illegal driving license or having illegal passport, vehicle with illegal registration, stolen vehicles, vehicle carrying any contraband item. It is a positive sign that this work will provide more relevant and accurate information to the border check post agents.

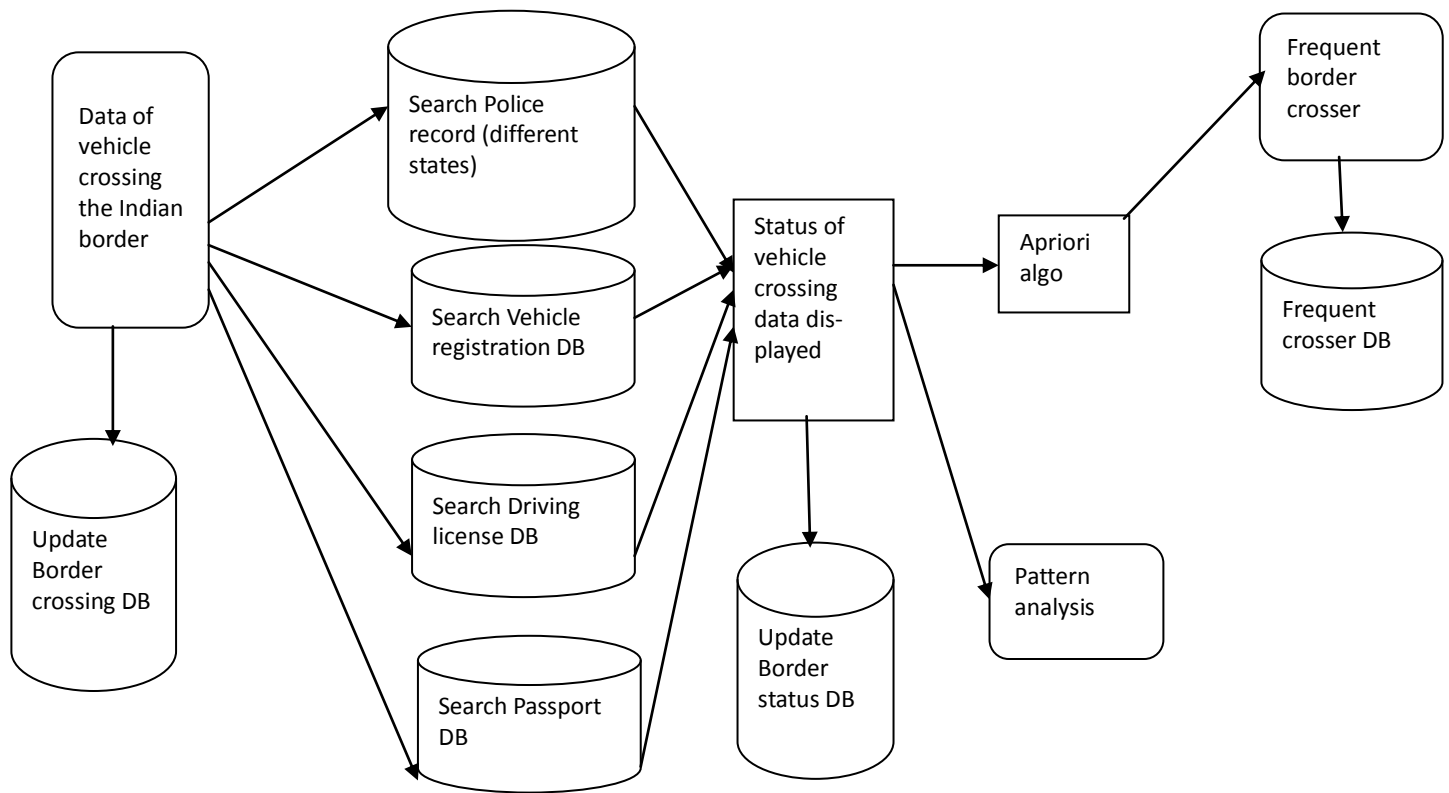


Fig 1: proposed architecture

4. EXPERIMENTAL RESULT

A. Result based on information integration

While conducting the research it was found that integrating the information from multiple sources helps in identifying many vehicles crossing the border have some criminal incident associated with them. On the border it becomes easier to identify people/vehicle crossing with criminal record, people

B. Result based on Apriori algorithm

In order to identify the frequent border crossers Apriori algorithm is applied on each vehicle crossing the border. On each border crossing record Apriori algorithm is applied based on the vehicle no. and driving license number. All available historical borders crossing data is scanned to identify that wheth-

er the vehicle crossing the border is a frequent border crosser or not. If it satisfies minimum support then only it is considered that it is a frequent border crosser and its details are stored in frequent crossers list. This list is then provided to higher authorities in country to keep watch on any suspicious frequent crosser.

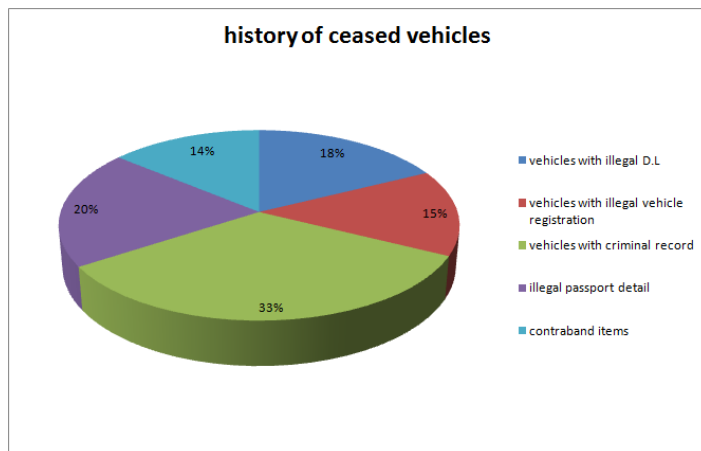


Fig2: history of ceased vehicles

C. Pattern analysis

After applying the Apriori algorithm a pattern analysis has been done to identify pattern followed by the vehicle at the check post. In the fig 2 pattern analysis has been done to show the history of ceased vehicles that were not allowed to cross the border. From the research work following is concluded about the ceased vehicles:

- 18% vehicles trying to cross the border were caught with illegal driving license
- 15% vehicles were ceased for illegal vehicle registration detail
- 33% vehicles trying to cross the border were ceased for having criminal record
- 20% vehicles were ceased for having illegal passport detail
- 14% vehicles were caught for carrying contraband items

In figure 3 pattern analysis has been done to compare the no. of vehicles crossing in different months of year 2010, 2011, 2012. Through this pattern it is analyzed that in all the three consecutive years' number of vehicles crossing the border was maximum in the month of April. Since in all the three years

vehicles crossing was maximum for this month it is possible that some cross border illegal activities might be taking place in this month.

This pattern analysis draws a very useful observation. This observation can be used to warn border check post agents that there might be some illegal activities going around the border in this month. So they must tighten the security and monitor the vehicles thoroughly in this month.

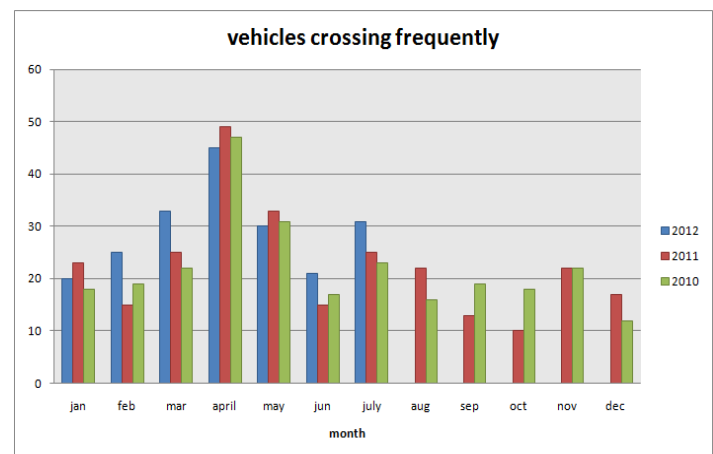


Fig3: yearly comparison of vehicles crossing frequently

5. CONCLUSION AND FUTURE WORK

In this research work a software tool has been implemented in order to solve the problems faced by border check post agents to check each and every vehicle manually. This software helps border check post agent to get more accurate, and relevant information related to every vehicles crossing the border. In this proposed system emphasis is given on utilizing the information from multiple law enforcement databases. These multiple agencies do not share their information frequently; they share information in special case only. Sharing of the information enhances the process of securing the Indian border and transportation. In this work information from different states of Indian police records, different states vehicle registration databases, different states driving license databases, border crossing data, border status database, Passport database has been integrated.

We conclude that the system worked successfully on the data that is used for the analysis. In this software more security features have been added as compared to previous research work. This helps border check post agents to decide whether to allow a vehicle to pass border or not and it also helps in identifying which vehicles are frequent border crosser

and pattern analysis has been done to identify pattern followed by vehicles which . The observation drawn from pattern analysis can be used to warn border check post agents about the illegal activities going around the border, so accordingly border check post agents can take the decision to tighten the security and monitor the vehicles thoroughly.

As we know that border and transportation security is very important for our nation's security. In the future work more security features can be implemented to increase the security at the border. In the system developed information from Indian law enforcement sources has been utilized, so in future information from the different countries can be utilized to enhance the security further at Indian borders. In future performance can be increased by applying some other association rule algorithm.

[8] www.fas.org

[9] <http://www.exinfm.com>

6. REFERENCES

- [1] Hsinchun chen, fei-yue wang, Daniel zeng, 2004 "Intelligence and security informatics for homeland security:information, communication, and transportation", IEEE transactions on intelligent transportation systems vol.5,no.4
- [2] Byron Marshall, Siddharth Kaza, Jennifer Xu, Homa Atabakhsh, Tim Petersen, Chuck Violette, and Hsinchun Chen, 2004 "cross jurisdictional crime activity networks to support border and transportation security"
- [3] Siddharth kaza, Tao wag, Hemanth Gowda, Hsinchun chen, 2005 "Target vehicle identification for border safety using mutual information "in proc. of 8th international conference on intelligent transportation system Vienna, Austrias.
- [4] Agrawal, R., Imielinski, T., and Swami, A. N. 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 207-216.
- [5] Jiawei Han and Micheline Kamber, 2006. "Data Mining: Concepts and Techniques, Second edition, Morgan Kaufmann"
- [6] Pramod Prasad, 2011 "Using association rule mining for extracting product sales pattern in retail store transactions", International journal on computer science and engineering (IJCSE) vol.3, no.5
- [7] Goswami D.N, Chaturvedi Anshu, Raghuvanshi C.S, 2010 "An algorithm for frequent pattern mining based on Apriori", International journal on computer science and engineering (IJCSE) vol.2, no.4