ENHANCING PREDICTIVE CRIME MAPPING MODEL USING ASSOCIATION RULE MINING FOR GEOGRAPHICAL AND DEMOGRAPHIC STRUCTURE

Nur Izzatul Abidah Roslin[1] Siti Azirah Asmai [2],

[1] *Universiti Teknikal Malaysia Melaka (MALAYSIA)*
[2] *Universiti Teknikal Malaysia Melaka(MALAYSIA)*

abidahroslin@gmail.com, azirah@utem.edu.my,

ABSTRACT

This research project is to enhanced predictive crime mapping model with data mining technique to predict the possible rate of crime occurrence. Few specific objectives are stated in order to achieve the aim of this research project. This project proposed a data mining technique called Association Rule Mining. Basically Association Rule Mining is to investigate the rules according to the predefined parameter. This technique considered useful if it can satisfy both minimum confidence and support. Apriori is a popular algorithm in finding frequent set of items in data and association rule. Dataset of Communities and Crime from UCI Machine Learning Repository is used in order to setup the experiment. 60% of the dataset is used for training to generate association rules by using WEKA. The association rules generated shows the prediction of the rate of crime occurrence. The other 40% of the dataset is used to test generated rules. A simple program of C++ is implemented using Microsoft Visual Studio to test generated rules until accuracy of performance is obtained. At the end of the project, generated rules tested and come out with difference accuracy according to predefined minimum support.

Keywords: Data Mining, Association Rule, Apriori Algorithm, Crime, Prediction.

1.0 INTRODUCTION

Crime has been highly increased in cases nowadays. This is important to the Malaysia Government in making the country safer and improves the quality of life in line with the requirements of the country to achieve the Government Transformation Programed (GTP). Based on this, the Malaysian government has been visible making various efforts in crime prevention in order to reduce crime. However, the implementations of crime prevention make highly extensive use of resources such as control unit, time and finances. Therefore, it is critical to identify a better approach to facilitate crime prevention solutions that correspond to specific future crime location.

The predictive crime mapping is a one of the solutions can be used to analyse the relatively high future crime location that can improve the crime prevention implementation.

2.0 PROBLEM STATEMENT

There are various preventive methods that have been proposed to overcome the increasing of the crime. One of the approaches is the use of predictive mapping. Predictive mapping is one of the specific features to provide identification of the areas in which to focus interventions in order to improve the execution decision. By having predictive mapping in crime prevention, it can support law enforcement to anticipate in the criminal target location to reduce the possibility of its occurrence. Even though the crime can be systematic and entirely random occurrences, geographic location also can be an attractive criterion

of the crime occurrence. Thus, it is critical to identify the spatial patterns and capture the information from the pattern to predict the possible crime occurrence for a better crime prediction solution. In addition, some of the predictive mapping models may only provide the future crime location without appropriate data representation. Hence, these solutions lack of delivering a rich set of demographic features and aggregating information.

3.0 OBJECTIVE

1. To analyze the spatial data parameter that involve in the prediction of crime.
2. To propose a new predictive crime mapping model using association rule to predict future crime location based on crime geographical and demographic patterns data.
3. To obtain the accuracy and usability of enhanced algorithm.

4.0 LITERATURE REVIEW

According Nath (2007), solving crime is a complex task that requires human intelligence and experience for analysing the crime pattern data. In addition the knowledge that is gained from data mining approaches is a very useful tool which can help and support police forces (Keyvanpour et al., 2011). However, data mining in the context of crime and intelligence analysis for national security is still a new field in Malaysia and yet to be utilized in crime mapping solution for geographical and demographic data. Crime mapping has been recognized as a tool for analysing the crime and increasingly implemented in law enforcement and criminal investigation recently (Schaaik et al., 2009).

Association rule mining is a data mining technique used for the identification of interesting structure in data, where structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data (Herawan et al., 2011). The association rules are considered useful if it can satisfy minimum confidence constraint (Qodmanan et al., 2011).

5.0 ASSOCIATION RULE TECHNIQUE AND DATA ANALYSIS

According to Kotsiantis and Kanellopoulos (2006), association rule mining is to investigate the rules according to the predefined minimum support and confidence. The rules must satisfy both the minimum support and confidence. There are two main issues; to find itemsets that exceed the threshold and to generate strong association rules from the itemsets with minimal threshold. From the large itemsets, one of them supposed to be L_k, $L_k = \{I_1, I_2, \dots, I_k\}$. Association rules that generated must be in the following way: $\{I_1, I_2, \dots, I_{k-1} \rightarrow I_k\}$, with checked confidence whether the rule can be interesting or not.

Apriori is an algorithm for mining association rules which iteratively reduces the minimum support until it finds the required number of rules with predefined minimum confidence. The algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemsets. The algorithm has an option to mine class association rules and generates strong association rules from the frequent itemsets.

Data collection includes the collection of Communities and Crime dataset and all the information related with the Communities and Crime which consists 14 attributes and 389 instances for experimentation. The dataset were split into 60% - 40%. 60% is used for training while the other 40% of the dataset is used for testing until the accuracy of performance obtained.

6.0 RESULTS

The experimental result of applying the apriori algorithm is revealed. Figure 6.1 shows the setting used before run the dataset. The car setting was set to true to make the last attribute in the data as the class attribute. The maximum number of rules to be displayed is 150. But the rules obtained were 115 rules and all the rules were the best that can be found from the dataset. The minimum support was set to 0.01 while, the maximum support was set to 0.9. Apriori algorithm will start creating rules with max support ending with either the number of rules specified or the minimum support. The other settings were set to default.

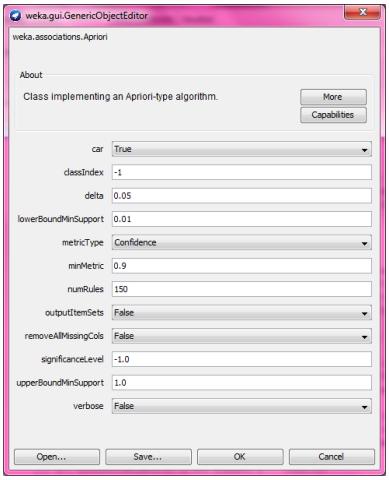

Fig. 6.1: Parameter Setting for Apriori Algorithm

Fig. 6.2: Visualisation after pre-processing.

For Fig. 6.1, we focus on the class attribute which it is the predicted attribute. It rates the crime occurrence at a specific location. The red bar at the class attribute related to the other 13 attributes. We can see that the red bars are the majority of all attributes. This show that rate is the most popular or highest among all the attributes. Other rates also as same as explained.

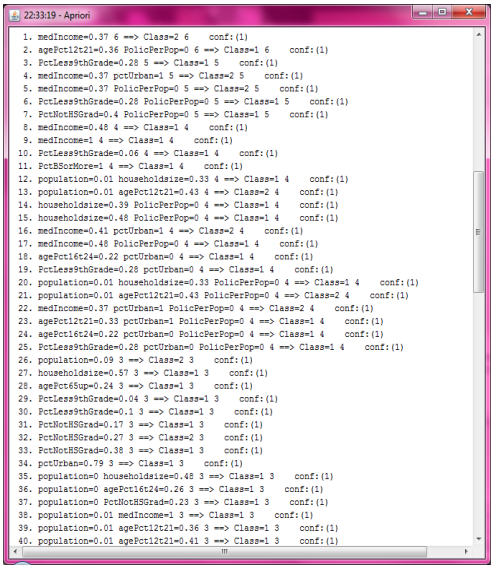

Figure 6.3: Sample of rules generated

Minimum support	Number of generated rules	Accuracy	Number of instances
0.01	115	17.31%	27
0.02	7	0.64%	1

Figure 6.4: Testing result

Figure 6.3 shows the sample rules generated from association rule mining. See appendices to look for the full list of the rules. The Apriori algorithm generated the sets of large itemsets found for each support size that been considered. The rules obtained were all 115 rules and all the rules were the best that can be found from the dataset. As stated, the confident value for all the rules were 1 which means 100% good. Figure 6.4 shows the comparison accuracy performance of two different minimum supports. As we can see, it is clearly stated 0.01 is a better minimum support that satisfied to find all the itemsets compared to 0.02 minimum support where it is less satisfied to find the frequent itemsets.

7.0 DISCUSSION

Apriori algorithm is the popular approaches to discover association rules based on frequent itemsets by setting the same or single minimum support for all itemset. Minimum support value is calculated based on percentage of support. A low minimum support will generate a large number of frequent itemset which will satisfy minimum support criteria. While a high minimum support could be fail to satisfy the minimum support criteria. Therefore, a low minimum support will generate a large number of rules and will obtain a high percentage of accuracy.

8.0 CONCLUSION

In conclusion, this project has been carried out according to the methodology proposed. This project has meets all of the three objectives of this project. However, future enhancement can be made to improve its effectiveness and its accuracy performance in future problem.

ACKNOWLEDGEMENTS

The authors would like to thank the participating students for their co-operations. We would also like to express our gratitude to Universiti Teknikal Malaysia Melaka for supporting these research activities.

REFERENCES

[1] Keyvanpour, M., Javideh, M. and Ebrahimi, M. (2011) Detecting and investigating crime by means of data mining: a general crime matching framework, Procedia Computer Science. World Conference on Information Technology, Elsvier BV.
[2] Schaaik, J. G. and Kemp, J. J. (2009) Real Crimes on Virtual Maps: The Application of Geography and GIS in Criminology. Geospatial Technology and the Role of Location in Science, 217-237.
[3] Herawan, T. and Deris, M. M. (2011) A soft set approach for association rules mining. *Knowledge-Based Systems*, 24**,** 186-195.
[4] Qodmanan, H. R., Nasiri, M. and Minaei-Bidgoli, B. (2011) Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence. *Expert Systems with applications*, 38**,** 288-298.
[5] Kotsiantis, Sotiris, and Dimitris Kanellopoulos. "Association rules mining: A recent overview." GESTS International Transactions on Computer Science and Engineering 32.1 (2006): 71-82.