Improved Method of Classification Algorithms for Crime Prediction

Abba Babakura, Md Nasir Sulaiman and Mahmud A. Yusuf
Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia (UPM),
43400 Serdang, Malaysia
abba_babakura@yahoo.com, nasir@upm.edu.my, mahmudkiru@yahoo.com

***Abstract*—The growing availability of information technologies has enabled law enforcement agencies to collect detailed data about various crimes. Classification is the procedure of finding a model (or function) that depicts and distinguishes data classes or notions, with the end goal of having the ability to utilize the model to predict the crime labels. In this research classification is applied to crime dataset to predict the "crime category" for diverse states of the United States of America (USA). The crime data set utilized within this research is real in nature, it was gathered from socio-economic data from 1990 US census. Law enforcement data from 1990 US LEMAS survey, and from the 1995 FBI UCR. This paper compares two different classification algorithms namely - Naïve Bayesian and Back Propagation (BP) for predicting "Crime Category" for distinctive states in USA. The result from the analysis demonstrated that Naïve Bayesian calculation out performed BP calculation and attained the accuracy of 90.2207% for group 1 and 94.0822% for group 2. This clearly indicates that Naïve Bayesian calculation is supportive for prediction in diverse states in USA.**

Keywords—Pre-processing; Algorithms; Feature selection; Crime prediction; Crime Category

I. INTRODUCTION

Criminologists and statisticians have been using their skills and knowledge trying to analyze these data, with varying degrees of success. However, the volume of crime and the greater awareness of modern criminals have made the process of analyzing the crime data difficult, because human reasoning fails when he is presented with millions of records. Therefore, there is a requirement for a technique to assist in analyzing the crime data. Data mining techniques can be applied to facilitate this task [17].

Data mining has therefore pulled in an incredible arrangement of consideration in the information industry and in society as a whole, because of the wide accessibility of huge amount of data and the imminent need for transforming such data into useful information and knowledge. Subsequently, the objective of data mining is extracting or mining of knowledge from a large amount of data [8]. Fig.1 below shows the data mining process. As illustrated in Fig.1, it clearly describes the stages to be followed in order to achieve the representation of required knowledge. It is a bottom to top procedure; the first stage is the database stage. Here, the database acts as a warehouse where raw data resides. In the second stage called the data selection stage, the target data are selected which leads to the selection of the required attributes or features. When the target data are obtained, then it give ways to the cleaning and processing stage. In this stage the data is processed in order to remove noisy, missing values or incomplete data. It is termed as the processed data when it passes this stage. The data analysis stage utilizes the data mining tasks, algorithms in a more suitable way to obtain actual patterns. This stage is the most important point as it leads to the resulting pattern required i.e. the prediction of future crime. Finally, the last stage interpretation and evaluation describes or interprets the actual pattern into knowledge which is required by the outside world.

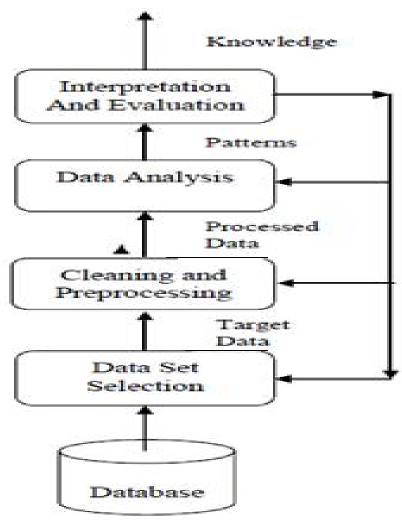

Fig.1 Data mining process

Data mining has numerous techniques yet a standout amongst the most generally utilized and important technique is Classification. Classification is a supervised class prediction technique which permits the prediction of class labels which are nominal rather than numerical [2].classification as a technique has produced success in numerous domains which incorporate homeland security, health care, weather forecasting, medical, financial, business intelligence [3,4]. The main focus of this research is the application of different kind of classification algorithms on the real data and assesses the accuracy of the result obtained in predicting the crime label.

978-1-4799-6444-4/14/$31.00 ©2014 IEEE

Crime is a steadily evolving predicament. Some of the most significant aspects to take into consideration are age, economics and so on. The development of crime in a society is reliant on the characteristics related with the community or society. These characteristics might incorporate - distinctive races in a society, different income groups, different age groups, family structure (single, divorced, married, number of kids), level of education, the locality where individuals live (cheap or expensive housing, size of houses, number of rooms), number of police officers distributed to an locality, number of employed and unemployed individuals and so forth.

In this research, a real crime data set is acquired to perform data mining [5]. The attributes of this dataset are the characteristics related with a community or a society, some of which are highlighted previously. The two distinctive classification algorithms that are utilized to perform classification on the dataset are namely– Naïve Bayesian (naïve bayes classifier) and Back Propagation (BP). By investigation, results of both the algorithm will be compared and examined, and the most efficient algorithm in predicting the target- class type (class label) will be recognized. There are numerous tools accessible for data mining, such as WEKA, Mat Lab, and Rapid Miner etc. For this research, WEKA is utilized, an open source tool written in JAVA [6]. The framework of this paper is as follows: Section 2 describes the related work. Section 3 deploys Methodological framework. Section 4 discusses the experimental results of the classification algorithms for predicting the 'class label' attribute, in diverse states of USA. Finally, Section 5 contains conclusion and future works.

II. Related Work

Data mining is a powerful tool, which enables law enforcement agencies to discover patterns and predict the future crimes. There have been numerous researches on applying data mining techniques to crime data. The necessity to ensure safety in the diverse states of USA can be seen from diverse outlook. Many literatures on crime prediction research highlighted the importance of safety in the environment. One significant work on crime prediction was given emphasize by Aziz et al., pointing the use of decision tree-based classification model for crime prediction [12]. Their proposed model assists law enforcement agencies in discovering crime patterns and predicting future trends.

Another significant work on crime prediction was given emphasize by Rizwan et al., on the experimental study of classification algorithms for crime prediction [13]. In this work, they pointed on the importance of predicting crime on a single group. It also highlighted on the comparison on two algorithms namely: decision tree and Naïve Bayes for classification.

Chen et al. [14], categorized different crime types and proposed some techniques to mine crime data such as entity extraction and association rule mining [14]. They also developed a framework that identifies the relationships between crime types and data mining techniques applied in crime data analysis. The developed framework helps investigators to find associations and identifies patterns to make predictions.

They implemented some case studies and showed their developed framework can help increase efficiency and reduce errors. Liu et al., [15] proposed a STT (spatio-temporal-textual) search engine for extracting, indexing, querying and visualizing crime information. They give scores to each of the data factors (spatial, temporal, and textual) and use the score to rank the data. This tool helps crime detection for investigators, identification of crime trends and patterns for decision makers and researchers, and security of city life for residents and journalists. Shah et al., [16] proposed CROWDSAFE, a novel convergence of Internet crowd sourcing and portable smart devices to enable real time, location based crime incident searching and reporting. It is targeted to users who are interested in crime information. The system leverages crowd sourced data to provide novel features such as a Safety Router and value added crime analytics. In addition to the demonstration of hotspots, CROWDSAGE also provides crime clusters, historical crime statistics to the users in a dashboard interface. The methodological approach will be discussed in the next section below.

III. Methodological Framework

A. *Preliminary discussion and experiment preparation*

Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown[6]. The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is unknown).There are many methods for constructing classification models, such as Artificial Neural Network, Decision tree (which is a flow-chart-like tree structure), Naïve Bayesian classification, Support Vector Machines (SVM), k-Nearest Neighbors, weighted voting. All these techniques can be applied to a dataset for discovering set of models to predict the unknown class label. When dealing with classification, the training and test phase comes into play i.e., the dataset is divided into two phase, namely training (dependent) phase and testing (independent) phase. The data mining algorithm originally runs on the training set, and then later the predicting model is applied on the test set [2], [7].

This dataset manifest different groups, each having certain attributes associated with it. All of this groups having different social and economic factors related with any community. After studying and performing some analysis, three distinct groups were chosen from this dataset. These three groups are related to race, education and marital status in a community. For every chosen group some important attributes are selected. Different methods for attribute or feature selection exist. For this experiment, attribute are selected using the manual method [8] based on human perceptive and intellect. It is sensible, especially when dealing with a large number of attributes. In the case of choosing the attributes of interest, it is noted that only attributes which do not have missing values are taken into consideration. Artificial neural network and Naïve Bayesian classifier are used to perform classification. During the first phase of the experiment, the model is built on the training set with known class label and in the second phase of the experiment, the proposed model is applied to predict the class labels on the test set. To show the robustness and performance of the developed models, after performing the experiment using the above mentioned classification algorithms; the accuracy, precision and recall of the algorithms are evaluated, for predicting the attribute 'Cat'.

B. Crime Dataset Collection

The dataset used to perform this experiment is real and authentic. It is obtained from UCI machine learning repository website. The label of the dataset is 'Crime and Communities'. It is structured using real data from Socio-economic data from 1990 US Census, law enforcement data from the 1990 US LEMAS survey and crime data from the 1995 FBI bUCR [5]. This dataset contains a total number of 128 attributes and about 2000 instances and all data provided in this dataset is numeric and normalized. The data in each instance belong to different states of the US. The states are represented in the form of numbers, every number representing its respective US state [9]. The complete details of all the 128 attributes can be obtained from the UCI machine learning repository website [5]. The first group represents the race to which the people are classified. In the race group there are four attributes with their respective data type. The second group represents the marital status of various people, also having four attributes with their respective data type. For each group, the attributes State, Population and ViolentCrimePerPop are included to obtain optimum result in relation to the various groups. The two classification algorithms were applied on the two different groups and the results obtained are analyzed.

C. Crime Dataset Pre-processing

Data preprocessing plays an important role in the knowledge discovery process as it gives rise to development of quality and efficient data before data mining algorithms are been applied to it. The dataset used for the experiment consists of about 2000 instances, which also contain some missing values. In order to obtain optimum results, quality data should be obtained. Therefore, the need for data pre-processing comes into play. Some of the few techniques in practice, employed for the purpose of data pre-processing are data cleaning, data integration, data transformation and data reduction [2], [4]. Prior to application of classification algorithm usually some pre-processing is performed on the dataset. Here in this experiment the first step is to perform data reduction. In this step the most informative attributes in a dataset are selected, while attempting not to lose any important information. There are different methods available for attribute or feature selection [8]. For this experiment, manual method was chosen for attribute selection [8] based on human understanding and intellect. Here in this experiment, the three attributes (state, population and Violent Crime per Pop) are added to both groups before classification is performed. It is practical, especially when dealing with a large number of attributes. It was also noted that only those attributes which do not contain any missing values are chosen. In the second step, in other to perform prediction a new attribute is added labeled 'Cat'. The values of the 'Violent Crimes Per Pop' attribute, which depicts the total number of violent crimes per 100K population stands as the benchmark on which the added attribute base. The rationale behind the addition of this new attribute is that, in order to perform prediction, the class (Cat) attribute should be nominal in nature. In this original dataset all values are numerical so there is a need to provide a nominal class to enable prediction. As stated earlier, this attribute is based on the data values in 'Violent Crimes per Pop'; this dependency also retains the integrity of the dataset. The new attribute is just acts, as a means for providing different nominal labels for the values in 'Violent Crimes per Pop', for prediction purposes. The new added nominal attribute have three values, which are 'Low', 'Medium', and 'High'. If the value in 'Violent Crimes Per Pop' is less than 25 percent than the value of 'Cat' is 'Low', If the value in 'Violent Crimes Per Pop' is equal to or greater than 25 percent and less than 40 percent, than the value of 'Cat' is 'Medium', If the value in 'Violent Crimes Per Pop' is equal to or greater than 40 percent than the value of 'Cat' is 'High'. All the values were added in the newly created attribute carefully for all instances, and cross checked multiple times by all authors, to eradicate any chances of errors.

D. Building Classifiers (Models) and Measurements for Performance Evaluation

Here, the algorithms which are used to carry out the experiment are explained and the performance measures are also described. A Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumption. Bayesian classifiers adopt a supervised learning approach. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. They have the ability to predict the probability that a given tuple belongs to a particular class [10]. The strength of Naïve Bayesian classifier, as a powerful probabilistic model has been proven for solving classification tasks effectively [11]. For any given instance, $X= (X_1,X_2\ldots\ldots X_n ,)$ where,
X_1 is the value of attribute X_1,
$P(C|X)$ is calculated by Bayesian classifier for all possible class values C and predicts:

$$\mathbf{C^*} = \text{argmax}_c\, p(x|c)$$

As the class value for instance X.

Hence, estimating *P(X|C)* which is proportional to *P(X|C) P(C)* is the main step of a Bayesian classifier.

Back propagation algorithm (BP), can be defined as a model of reasoning based on the human brain. A typical BP is made up of a hierarchy of layers, and the neurons in the networks are arranged along these layers [1]. The weights are modified to bring the network input/output behavior into line with that of the environment. The equation for the model is:

$$Y(p) = step\left[\sum_{i=1}^{n} x_i(p)w_i(p) - \theta\right],$$

Where n is the function of the perceptron
Step is a step activation function
w_i is the weights of the perceptron and x_i is the inputs.
Y (p) is the desired output.

The two algorithms used in this experiment a have a great ability of giving a correct prediction. Both algorithms result will be evaluated based on the following performance measurements which are listed below:

- Accuracy which defines or outline the percentage of correctly classified instances by the classifiers [2].
- Precision which refers to the proportion of data correctly classified using classification algorithm.
- Recall which is also a significant performance measure that refers to the percentage of information which is relevant to the class and is correctly classified.

IV. Experiment Results, Analysis And Performance Evaluation

In the experiment, a comparison between Naïve Bayesian and Artificial Neural Network was performed, over the crime dataset [5]. When carrying out the experiment, the preprocessed dataset was converted to .ARFF file, which is the standard file type for WEKA input [8]. Also, during the experiment 10 fold cross validation was applied to the input dataset of the two different groups, separately for both Naïve Bayesian and Back Propagation algorithms. The Accuracy for 10 fold cross validation for Naïve Bayesian and Back Propagation for the different groups is shown in table 3 below.

Table 3. Accuracy of NB and BP

	Naïve bayes	BP
Group1	90.2207%	65.9478%
Group2	94.0822%	65.9469%

From the result obtained from the experiment Naïve Bayesian out performed BP and manifested higher performance. However the confusion matrix for both algorithms on each group is shown in table 4, 5, 6 & 7 below.

Table 4. Confusion matrix using NB for group1

	Low	High	Medium	
Low	1223	43	49	1315
High	0	363	23	386
Medium	54	26	213	293

Table 5. Confusion matrix using BP for group1

	Low	High	Medium	
Low	1315	0	0	1315
High	386	0	0	386
Medium	293	0	0	293

Table 6. Confusion matrix using NB for group2

	Low	High	Medium	
Low	1246	20	49	1315
High	0	370	16	386
Medium	21	12	260	293

Table 7. Confusion matrix using BP for group2

	Low	High	Medium	
Low	1315	0	0	1315
High	386	0	0	386
Medium	293	0	0	293

Table 4 & 6 illustrates the classification of low, high and medium classes using Naïve Bayesian algorithm for each group. Similarly, table 5 & 7 illustrates the classification of low, high and medium classes using BP algorithm.

The results from the confusion matrixes and its explanation above, it can be seen that Naïve Bayesian performed better than the BP in both groups. The classification based on group 1 which is related to race is showing the highest accuracy. Moreover, from the analysis it has been seen that precision and recall remain the same when BP is used as a classifier for both groups. BP as a classifier, it can be clearly seen that precision is 0.659 for both groups and recall remains 1 also for both groups. Naïve Bayesian performed better in predicting all the classes, namely Low, Medium and High and for BP as seen in the confusion matrix predicts only the class label Low. Table 8 & 9, respectively illustrates the Accuracy (correctly classified instances), incorrectly classified instances, Precision and Recall for both the algorithms used for the two groups in the experiment. The comparison between different measures of both groups is illustrated below. Here, Fig.2 shows the different measures comparison based on group 1. Similarly Fig.3 shows the different measures comparison based on group 2. The percentage comparison for correctly and incorrectly classified instances for both algorithms for each group is demonstrated in Figure 4 and 5 respectively.

TABLE 8. ACCURACY, INCORRECTLY CLASSIfiED INSTANCES, PRECISION, RECALL AND F-MEASURE FOR BOTH ALGORITHMS OF GROUP1

Group1	Method	Accuracy(correctly classified instances)	Incorrectly classified instances	Precision	Recall
	Naïve bayes	90.2207%	9.7793%	0.958	0.93
	BP	65.9478%	34.0522%	0.659	1

TABLE 9. ACCURACY, INCORRECTLY CLASSIfiED INSTANCES, PRECISION, RECALL AND F-MEASURE FOR BOTH ALGORITHMS OF GROUP2

Group2	Method	Accuracy(correctly classified instances)	Incorrectly classified instances	Precision	Recall
	Naïve bayes	94.0822%	5.9178%	0.983	0.948
	BP	65.9469%	34.0531%	0.659	1

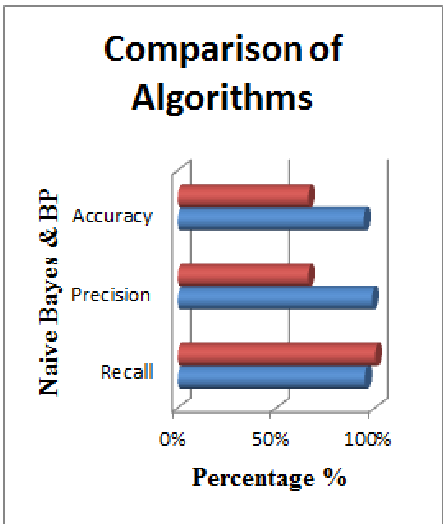

Fig. 2.Comparison between different measures of both algorithms of group 1.

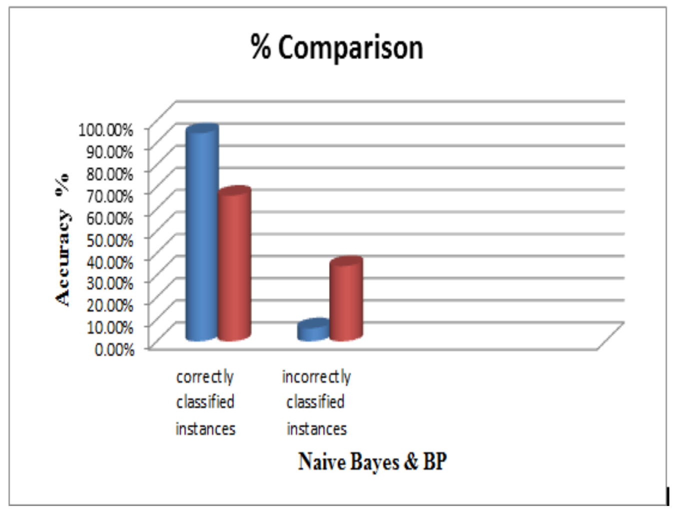

Fig. 4. Percentage comparison of correctly and incorrectly classified instances of group1.

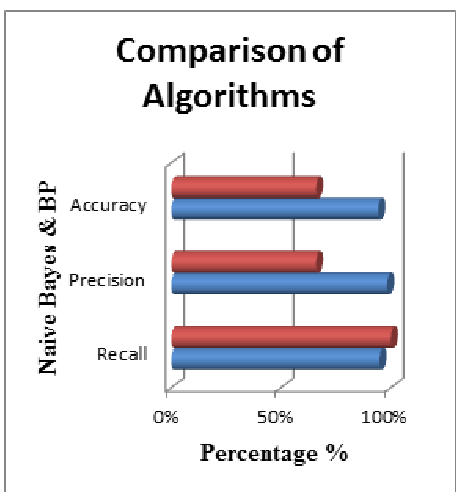

Fig. 3. Comparison between different measures of both algorithms of group 2.

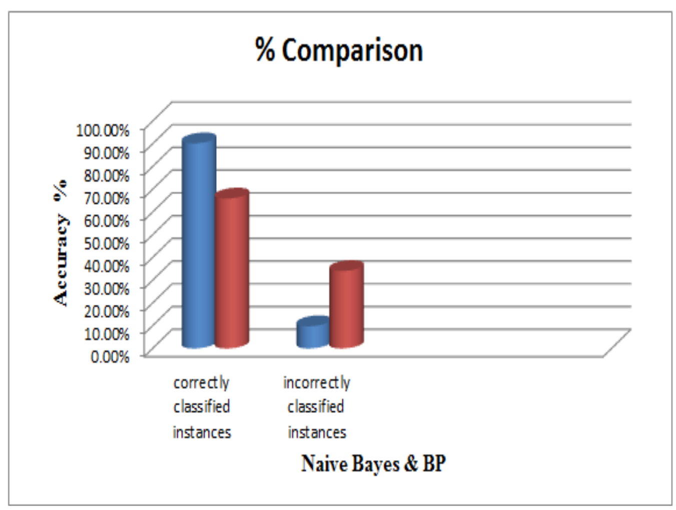

Fig. 5. Percentage comparison of correctly and incorrectly classified instances of group 2.

V. CONCLUSION AND FUTURE WORK

This paper presents a comparison between two classification algorithms namely, Naïve Bayesian and Back Propagation for predicting the 'Crime Category' attribute, having labels, namely 'Low', 'Medium', and 'High'. For Naïve Bayesian, the Accuracy, Precision and Recall of group 1 are 90.2207%, 95.8% and 93% and for group 2 are 94.0822%, 98.3% and 94.8%. On the other hand, Accuracy, Precision and Recall values for BP for group1 are 65.9478%, 65.9% and 100% and for group 2 are 65.9469%, 65.9% and 100% respectively. Experimental results for both the algorithms manifest that, Naïve Bayesian performed better than the BP for the crime dataset using WEKA. This experiment was performed using 10-fold cross-validation. It is evident that law enforcing agencies can take great advantage, using machine learning algorithms like Naïve Bayesian to effectively fight crime and war against terrorism. For future research, there is a plan to further apply other classification algorithms on the crime dataset and evaluate their prediction performance. Another direction for future work is to use other techniques for feature selection, and study their effect on the prediction performance of different algorithms.

REFERENCES

[1] Negnevitsky, M.: Artificial Intelligence: A Guide to intelligent systems, second edition, pearson education limited, pp. 167-172. (2005)

[2] Han, J.., and Kamber, M.: Data mining: concepts and techniques, Morgan Kaufmann Publishers, San Francisco, CA. (2006)

[3] Li, G., and Wang, Y.: A privacy-preserving classification method based on singular value decomposition, Arab Journal of Information Technology, vol 9(6), 529–534. (2012)

[4] Santhi, P., and Bhaskaran, V. M.: Performance of clustering algorithms in healthcare database, International Journal for Advances in Computer Science, vol 2(1), 26–31. (2010)

[5] UCI Machine Learning Repository.: Available from: http://archive.ics.uci.edu/ml/datasets.html (2012)

[6] Ngai, E. W. T., Xiu, L. et al.,: Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications, vol 36(2), 2592–2602. (2009)

[7] Selvaraj, S., and Natarajan,, J.: Microarray data analysis and mining tools, Bioinformation, vol 6(3), 95–99. (2011)

[8] Witten, I., Frank, E. et al.,: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann. (2011)

[9] WikiPedia.: WIKIPEDIA, Available from: http://en.wikipedia.org/wiki/List_of_U.S._states_by_area. (2012)

[10] Batchu, V., Aravindhar, D. J. et al.,: A classification based dependent approach for suppressing data, IJCA Proceedings on Wireless Information Networks & Business Information System (WINBIS 2012), Foundation of Computer Science (FCS). 2011

[11] Geenen, P. L.., van der Gaag, L. C. et al.,: Constructing naive Bayesian classifiers for veterinary medicine: A case study in the clinical diagnosis of classical swine fever, Research in Veterinary Science, vol 91(1), 64–70. (2011)

[12] Aziz, N., Ihm, S., Park, Y.: A Decision Tree-based Classification Model for Crime Prediction. Information Technology Convergence, lecture notes in Electrical Engineering 253. (2013)

[13] Rizwan, I., Masrah, A. A. M. et al.,: An experimental study of classification algorithms for crime prediction. International journal of science and technology vol 6 (3). (2013)

[14] Chen, H., Chung, W., Xu, JJ., Wang, G., Qin, Y., Chau, M.: Crime data mining: a general framework and some examples. The Comp 37(4):50–56. (2004)

[15] Liu, X., Jian, C., Lu, CT.: A spatio-temporal-texture crime search engine. In: Proceeding of the 18th SIGSPATIAL international conference on advances in geographical information systems, pp 528–529. (2010)

[16] Shah, S., Boa, F., Lu, CT., Chen, IR.: CROWDSAFE: Crowd sourcing of crime incidents and safe routing on mobile devices. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographical information systems, pp 521–524. (2011)

[17] Baumgartner, K., Ferrari, S., Palermo, G.: Constructing Bayesian networks for criminal profiling from limited data. Knowl Based Syst 21(7):563–57. (2008)