A Knowledge Discovery Approach to Supporting Crime Prevention

Sheng-Tun Li[,1, 2]   Fu-Ching Tsai[1]   Shu-Ching Kuo[2, 3]   Yi-Chung Cheng[2,4]

[1]Institute of Information Management, National Cheng Kung University, Taiwan, R.O.C.

[2]Department of Industrial and Information Management, National Cheng Kung University, Taiwan, R.O.C.

[3] Department of Information Management, Diwan College of Management, Taiwan, R.O.C.

[4]Department of International Trade, Tainan Woman's College of Arts & Technology, Taiwan, R.O.C.

Abstract

Recently, the increasing volume crimes have been one of the most serious issues in Taiwan. The Ministry of Internal Affairs has been working out a project to examine the public security index with red, yellow, purple, green, blue five lights in order to strengthen the public security. For analyzing and predicting huge linguistic data which evolve with time, we propose a novel fuzzy self-organization map network to uncover crime trend and use association rule to discover the hidden causal effects between two different criminal time series data. The fuzzy self-organization model integrates the features of dealing with clustering and linguistic data of SOM and fuzzy logic, respectively. We analyze the clustering results on distinguishing different trends of each criminal category and find rules relating patterns in a time series to other patterns, for the purpose of specifying the instructions of police human resources planning. The resulting findings can facilitate the development of a useful decision-support tool for helping decision makers determine appropriate law enforcement strategies.

Keywords: Trend discovery, crime prevention, self-organizing map, fuzzy theory.

1. Introduction

In recent years, volumes of crime have brought serious problems to many countries in the world. Take in Taiwan for example, the crime volumes have increased more than 71% in a decade, which may not only bring physical harms but also serious mental injuries for the victims. In order to avoid the emergence of the calamity, the governmental police department has to invest more human resources in criminal investigation and increase the law enforcement duties for preventing all kind of hazards and promoting the welfare of citizens. Traditional law enforcement strategies for crime prevention focus on preventive police patrol which is the most general duty for the police. Since 2002, the Ministry of Internal Affairs in Taiwan has been worked out a project to examine the public security index with red, yellow, purple, green, blue five lights that represent five kinds of status of public security which are "very bad", "bad", "intermediate", "good", "very good" in order to strengthen the degree of people's impression of public security. However, studies in recent years show that traditional law enforcement policy has less effectiveness on crime prevention due to failing in analyzing characteristics of each crime type and uncovering the criminal pattern. Therefore, it is of great necessity to develop a modern law enforcement policy, of which the police human resource planning is the most essential to the strategies for accommodating crime prevention. Efficient human resource strategies allow police department to deploy the appropriate duties in the specific time and place [1, 5]. Different strategies should be applied to different situations depend on the time and place. On the basis of above-mentioned reasons, we will use data mining to discover interesting crime pattern and to extract efficient crime rules from large amounts of time series database.

In discovering crime pattern, we consider the self-organizing map (SOM) network which is one of the most famous models attributed to its salient capability in clustering and visualization techniques [4, 6, 9]. The SOM network has been shown an effective model to solve problems with nonlinear data. However, the general SOM model can only manipulate quantitative numeric data but has a deficiency in linguistic data. Since the data set containing five light indexes under investigation in this study collected from National Police Agency (NPA) in Taiwan is linguistic in nature, the classic SOM network has to be enhanced to incorporating the capability of processing linguistic data. To solve this problem, we propose a Fuzzy-SOM model by applying Zadeh's fuzzy theory in 1965. The fuzzy-SOM network can handle linguistic data based on the traditional SOM scheme. In extraction efficient crime rules, we use rule extraction method that is similar Das et al. (1998) [3] and Cotofrei & Stoffel (2002)[2] methods to obtain crime rules from the time series that help police understand crime patterns.

2. The Fuzzy Self-organizing map neural networks

The SOM neural network is one of the most popular unsupervised neural network models, which quantize the data space and simultaneously performs a topology-preserving projection from the data space onto a regular two-dimensional grid [6]. The SOM network can be used for clustering, classification and visualization. However, the Kohonen SOM has not capabilities in linguistic data. To remedy the shortcoming, we propose a novel fuzzy self-organizing maps (FSOM) network.

The input to the model is a triangular fuzzy numbers sequence of k-dimensional vector $F = (F_1, F_2, ..., F_k)$. The input layer contains neurons for each element (F_i) in the input vector. Each neuron i in the map is represented by a n-dimensional time sequence, $F_i(x) = (\mu_{i1}(x), \mu_{i2}(x), ..., \mu_{in}(x))$ where the components of each $\mu_{in}(x)$ are triangular fuzzy numbers and n is equal to the number of neurons in the

input layer.

The neurons in the map are connected to adjacent ones by a neighborhood relation of the neurons. When an input vector $F_i(x) \in R^n$ is presented to the network, the neurons in the map compete to each other to be the winner (or the best-matching unit, BMU) $F_j(x)$, which is the closest to the input vector in terms of some kind of dissimilarity measure. Determining an appropriate similarity measure is usually the first step for clustering linguistic data. There are various choices and a specific choice for any given application should depend on the application environment in which the observed data set is generated. By taking the computation efficiency into consideration, the fuzzy similarity based on the fuzzy t-norm and s-norm is used in this study. That is, for two fuzzy time series $F_i(x) = (\mu_{i1}(x), \mu_{i2}(x), ..., \mu_{in}(x))$ and $F_j(x) = (\mu_{j1}(x), \mu_{j2}(x), ..., \mu_{jn}(x))$, we define the membership function between $F_i(x)$ and $F_j(x)$ as follows:

$$\mu_{(ih,jh)}(x) = \frac{\mu_{ih \cap jh}(x)}{\mu_{ih \cup jh}(x)} \tag{1}$$

$$SF_{ij}(x) = \frac{1}{n} \sum_{h=1}^{n} \mu_{(ih,jh)}(x) \tag{2}$$

where the membership function $\mu_{ih \cap jh}(x)$ and $\mu_{ih \cup jh}(x)$ of the fuzzy t-norms and s-norms define the degree of memberships to which an element x belongs to $ih \cap jh$ and $ih \cup jh$, respectively. The degree will be any value within the unit interval [0, 1]. These operations can be defined in various ways; we restrict ourselves to the well-known operations *min* and *max*.

During the session of training SOM networks, weights of neurons that are topographically close in the map are moved toward a given input vector. It is usually to use certain geometric distance to update weight in crisp data. However, since the input vector is linguistic in this study, one needs to use fuzzy bias number to replace the geometric distance. For a given fuzzy set $\mu_{ih}(x)$, the fuzzy-SOM algorithm finds the closest cluster center of fuzzy set $\mu_{jh}(x)$ and moved toward the input $\mu_{ih}(x)$ using the so-called "self-organization" learning rule. We use the triangular membership function which is specified by three parameters fuzzy set $\mu_{jh}(x) = tri(\mu_{jh}^{left}, \mu_{jh}^{middle}, \mu_{jh}^{right})$ and fuzzy set $\mu_{ih}(x) = tri(\mu_{ih}^{left}, \mu_{ih}^{middle}, \mu_{ih}^{right})$ for the reduction of the computational cost. The new cluster center after updating will be $\mu_{jh}'(x) = tri(\mu_{jh}^{left'}, \mu_{jh}^{middle'}, \mu_{jh}^{right'})$,

$$\begin{cases} \mu_{jh}^{left'} = \mu_{jh}^{left} + \eta \cdot (\mu_{ih}^{left} - \mu_{jh}^{left}) \\ \mu_{jh}^{middle'} = \mu_{jh}^{middle} + \eta \cdot (\mu_{ih}^{middle} - \mu_{jh}^{middle}) \\ \mu_{jh}^{right'} = \mu_{jh}^{right} + \eta \cdot (\mu_{ih}^{right} - \mu_{jh}^{right}) \end{cases} \tag{3}$$

The notation η is a small positive learning rate which can be defined as

$$\eta(g) = \eta_0 \exp(-\frac{g}{g_{total}}), \quad g = 0, 1, 2, ... \tag{4}$$

Where: η_0 is the constant learning rate, g is the number of generation time and g_{total} is the total

generation time.

Therefore, similar input vectors are grouped into a single neuron or neighboring ones in the map when learning is accomplished.

3. Rule discovery from time series

This research another focal point is that discover crime rules from two different temporal linguist set of time series. In extracting crime rules, we use the similar approach that [2, 3, 8] suggest in their rule discovery from time series. The rule format and rule extraction algorithm are illustrated as follow.

The rule format is: " If the precinct P_1 is red (or very bad or L_1) light, then after T month the other precinct P_2 will also become yellow (or bad or L_2) light in the status of crime c_i. We write the rule notion as

$$\{P_1 \text{ is } L_1\} \overset{T}{\Rightarrow} \{P_2 \text{ is } L_2\} \quad \text{or} \quad P_1(L_1) \overset{T}{\Rightarrow} P_2(L_2)$$

Where: P_i is a set of precinct which have 20 precincts, L_i is assumed to be linguistic variable {*red, yellow, purple, green, blue*}, C_i is a set of criminal categories which means 14 criminal categories in this research. The temporal record of $P_i(L_i)$ represents the public security status in precinct P_i.

The crime rule extraction algorithm in time series is a two-step process.
Step 1: Transform two different temporal sequences into a relation matrix. The linguistic elements which change over time are recorded in $S_i = (L_2(1), L_3(2), ..., L_1(t), ..., L_3(n))$. Let two different temporal records s_1 and s_2. In the process to transform into a relation matrix (R) from s_1 and s_2, the linguistic elements (L_i) in s_1 would sequentially make a pair to the element (L_j) that shift T time unit in S_2. $R = [r_{ij}] =$ The number of occurrences of L_i that are followed by a L_j within T.

Step 2: Set the confidence and support threshold value. The confidence and support threshold value can be set by domain experts in rule creation. They are two widely used to measure the interestingness of association rules. The support refers to the percentage of task-relevant data transactions for which the pattern is true, it is defined as

$$Sup(P_1(L_1) \overset{T}{\Rightarrow} P_2(L_2)) = p(P_2(L_2)_T \cap P_1(L_1)) = \frac{n(P_1(L_1), P_2(L_2), T)}{n(P_1(L_1))} \tag{5}$$

Where: $n(P_1(L_1), P_2(L_2), T)$ is the number of occurrences of $P_1(L_1)$ that are followed by a $P_2(L_2)$ within T.
The confidence is the conditional probability and is defined as

$$Con(P_1(L_1) \overset{T}{\Rightarrow} P_2(L_2)) = p(P_2(L_2)_T | P_1(L_1)) = \frac{Sup(P_1(L_1) \overset{T}{\Rightarrow} P_2(L_2))}{p(P_1(L_1))} \tag{6}$$

The main advantage of this crime rule extraction algorithm in time series is capable of extracting rules that are easy to understand for police to describe criminal patterns and deploy the human resources more effectively.

4. Experimental design

The data under investigation for this study is a set of crime volume of 20 county police bureaus in Taiwan provided by NPA, including 24 months the period from 2003 to 2004. There are three counties been excluded in this study because the crime volume is not enough to represent the criminal trend. Each county has 14 criminal categories describing the information of various criminal types, intimidation, drugs, automobile theft, sexual offences, anti-social behavior, theft, motorcycle theft, firearms crime, damage, force taking, robbery, gambling and injury. These categories are selected based on the statistic data from NPA. The data set is formed by 14 offenses each county and 24 months per crime summed 280 linguistic time series data. In order to avoid the influence of crime volumes and extract the trend of each criminal category, these crime volumes are standardized to be a number between 0 to1. After the data set has been standardized, it is categorized to five light indexes by fuzzy triangle membership function.

At the mining stage, we apply the fuzzy SOM network to cluster the fuzzied crime data. In the meantime, the cluster number and learning rate need to be decided in advance. The selection of an appropriate number of clusters is problem dependent, and is, in general, based on the principle of minimizing the inter-distance and maximizing the intra-distance among clusters. In this study we use the so-called silhouette value to determine the cluster numbers and evaluate the cluster numbers ranging from 2 to 10 using the 280 time series data. The parameter of learning rate that can affect the generalization performance of fuzzy SOM network is further discussed in this study.

The stopping criterion for training is based on whether the training model converges or not. There are some reasonable criteria with their own practical merits. However, we can simply use the error-distance diagram to estimate the converge status. In this study, Euclidean distance is used to measure the difference between the cluster center and the input data set. As the data set and the cluster centers are fuzzy sets, a defuzzification procedure is needed to estimate the Euclidean distance by converting the fuzzy set to a single real number. The error-distance can be defined in the following:

$$E(S_i, C_j) = \sqrt{\sum_{k=1}^{n} dist(S_{ik}, dfuz(C_{jk}))} \qquad (7)$$

where the 'dfuz' function is the gravity method to defuzzify the fuzzy set of center vectors.

Before generating association rules, the three parameters: support, confidence and time interval T are determined by the police for generating the appropriate and reasonable rules. From the viewpoint of empirical practices, the senior police officer, claims that it is hard to explain the relations if time interval is larger than 2 month. Therefore, the parameter T is set from 0 to 2 in this research. Another, the threshold value support and confidence is set 0.2 and 0.8 respectively.

In addition to the analysis of generalization performance, the effectiveness of fuzzy SOM model can be compared with the researches which been proposed using traditional behavior science. Likewise, the modern police human resource strategies could be compared with traditional ones. In contrast to most work in crime prevention, we conduct practical human resource strategies on analyzing the characteristic of each cluster and their impact on the police duty strategies.

5. Experimental results and analysis

Fig. 1 exhibits the different rates of training performance when the model is trained using learning rates of 0.001, 0.005 and 0.01. The figure indicates that these three different learning rates can converge to the same status, but the stability is remarkable difference. The model with learning rate of 0.001 achieves the best stability and learning performance comparatively.
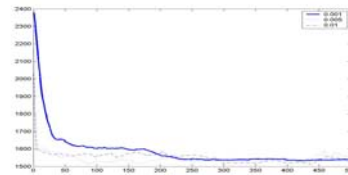

Fig 1. Learning curve of the fuzzy SOM network

To evaluate the number of clusters, we tried to explore the range from 2 to 10 using the silhouette value [7]. Fig. 2 illustrates experimental results of evaluating cluster numbers. The highest value is reached around 0.19 when two clusters are identified. However, it is meaningless to uncover the trend of all kinds of crime categories. Therefore, we chose four clusters as the target, which achieved the second highest silhouette value.
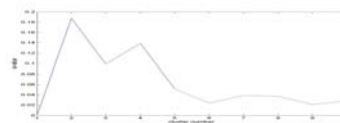

Fig 2. The silhouette values for different clusters

5.1 The cluster analysis of crime precincts

After using Fuzzy-SOM model to cluster crime offenses into 4 clusters. Fig. 3 displays the criminal trends which are represented by the four cluster centers identified. According to each cluster trend we are labeled cluster 1, 2, 3 and 4 as 'typical ', 'initial rising', 'keeping rising', and 'winter' pattern, respective. Among these clusters, the initial rising and keep rising patterns are the unusual and dangerous crime tendencies with shift-around effect. For reducing and preventing organized crime, we have to analyze more detailed about the initial rising and keep rising trend in each precinct. We expect to discover any similar crime patterns by shift-around effect and use visual analysis to robust human explain and analysis that mark the same color in precincts that possess the same trend of criminality. Fig. 4 shows the keep rising pattern in green and initial rising in purple.
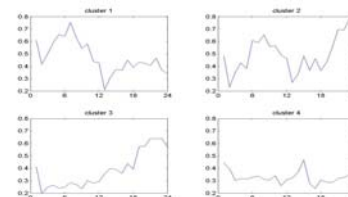

Fig 3. The Fuzzy-SOM cluster center map

Theft	Drugs	Damage	Injury	Robbery	Fraud	Gambling

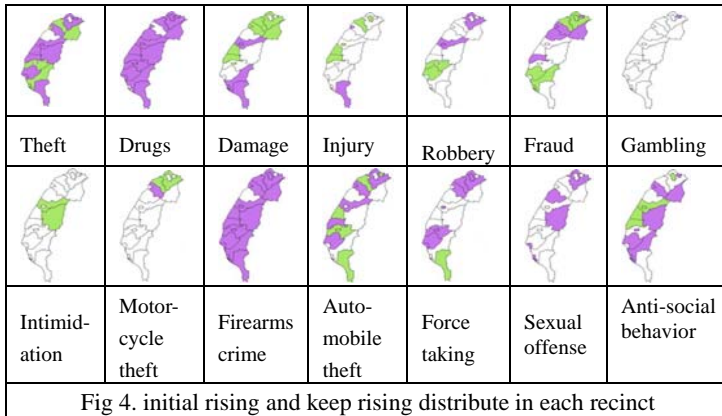| Intimid-ation	Motor-cycle theft	Firearms crime	Auto-mobile theft	Force taking	Sexual offense	Anti-social behavior

Fig 4. initial rising and keep rising distribute in each recinct

As shown in fig. 4, there is an obvious shift-around effect among nearby precincts that possess the same pattern in the same time. According to the visual analysis, we can specifically confirm the precincts that possess the shift-around effect and help the decision maker in planning the special inter-precinct project to prevent from recidivists keep shifting around crime locations. In general, precincts around the metropolises should integrate the law enforce duties with metropolitan ones in the same time. Moreover, police can decide which precinct around the metropolises should enforce the same duties more precisely via figure and save more human resources.

5.2 Crime rule analysis

In order to discover the relational rules with shift-around effect among these precincts and criminal categories, we generate the rules from initial rising and keeping rising groups. Our experiments focused on finding out whether these rules can help the police to describe criminal patterns more specifically. In addition, the police may investigate the reasons which exist in causal rules between two time series data. Therefore, we set a criterion to omit the causal rules that contain purple, green and blue lights which means the pubic security status is still in the safe range. Our study finally extracts nine crime causal rules. Table 1 exhibits the causal rules with shift-around effect.

Table 1. crime rules with shift-around effect

(*Automobile theft*)
The Pingtung county is yellow $\overset{T=1}{\Rightarrow}$ *The Changhua county is yellow.* $(0.22, 0.83)$
(*Motorcycle theft*)
The Taipei county is red $\overset{T=0}{\Rightarrow}$ *The H* $\sin chu$ *county is yellow.* $(0.21, 0.83)$
(*Fraud*)
The Kaohsiung county is red $\overset{T=0}{\Rightarrow}$ *The Changhua county is yellow.* $(0.21, 1)$
The Changhua county is yellow $\overset{T=0}{\Rightarrow}$ *The Kaohsiung county is red.* $(0.21, 0.83)$
The Changhua county is yellow $\overset{T=0}{\Rightarrow}$ *The Taipei City is yellow.* $(0.21, 0.83)$
The Taipei City is yellow $\overset{T=2}{\Rightarrow}$ *The Taipei County is yellow.* $(0.23, 0.83)$
The YiLan County is yellow $\overset{T=2}{\Rightarrow}$ *The Taipei County is yellow.* $(0.23, 0.83)$
(*Anti - social behavior*)
The Taichung city is yellow $\overset{T=1}{\Rightarrow}$ *The Changhua county is yellow.* $(0.26, 0.86)$
The Taipei city is red $\overset{T=0}{\Rightarrow}$ *The Taichung county is red.* $(0.21, 0.83)$

Obviously, the shift-around effect is remarkable in the criminal categories of drugs, automobile theft, motorcycle theft, etc. Table 1 summarize that causal rules are usually composed by the nearby precincts which represent causal rules are helpful for the police to extract shift-around effect that always happen in the neighbor precincts. Moreover, the two criminal categories of every causal rule are the same, which indicate that criminals are rarely change their offenses while shift around to other precincts. However, the shift-around effect is not remarkable for causal rules of fraud that the precinct of a fraud rule is not usually neighbors to the other one. The distinctive situation is because of the most common pattern of fraud in Taiwan is to commit the crime by telephone. Therefore, the shift-around effect of fraud is not remarkable in neighbor precincts.

6. Conclusions and future works

The main objective of this study is developing a linguistic cluster model in order to meet the public security index requirement and extract crime rule in time series. In contrast to the current studies in crime theory which mostly rely on traditional behavior science, we turned to a hybrid approach to overcome the hurdle of linguistic clustering in original SOM model. In order to support decision making for planning police human resources, we analyze the trend uncovered. One limitation of the current study is the difficulties of evaluating the real effectiveness of the proposed model on the police human resources planning. Therefore, the future work is to enhance the measurement of crime reduction and its impact on the human resources.

References

[1] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin & M. Chau, *"Crime Data Mining: A General Framework and Some Examples,"* IEEE Computer Society, 2004, pp. 50 –56.

[2] P. Cotofrei & K. Stoffel, "Rule Extraction from Time Series Databases using Classification Trees," In Proceedings of IASTED International Conference on Applied Informatics, Innsbruck, 2002, pp. 327 –332.

[3] G. Das, K. Lin, H. Mannila, G. Renganathan & P. Smyth, *"Rule discovery from time series,"* The Fourth International Conference on Knowledge Discovery and Data Mining, 1998, pp. 16 –22.

[4] A. Flexer, *"On the ues of self-organizing maps for clustering and visualization,"* Intelligent Data Analysis, vol(5), 2001, pp. 373 –384.

[5] W. Gorr & R. Harries, *"Introduce to crime forecasting,"* International Journal of Forecasting, vol. 19, 2003, pp. 551 –555.

[6] T. Kohonen, *"Self-Organizing Maps,"* Springer-Verlag Berlin, Heideberg, 1997.

[7] P. J. Rousseeuw, *"Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,"* Journal of Computational and Applied Mathematics, vol. 20,1987, pp. 53 –65.

[8] J. Vesanto & E. Alhoniemi, *"Clustering of the self-organizing map,"* IEEE Transactions on Neural Networks, vol. 11(3), 2000, pp. 586 –600.

[9] L. A. Zadeh., *"Fuzzy sets,"* Information and Control, 8, 1965, pp. 338 – 353.