Applying AI technology and rough set theory for mining association rules to support crime management and fire-fighting resources allocation

Show-Chin Lee
Department of Information Management, National Changhua University of Education, postgraduate
sarashow@ms38.hinet.net

Mu-Jung Huang
Department of Information Management, National Changhua University of Education, professor
mjhuang@cc.ncue.edu.tw

Abstract

The missions for the police and fire fighters are to protect for public safety and to fight and prevent from fires, respectively. In this dynamic environment, however, there are many potential dangers and uncertain factors that can't be predicted. In order to improve the total performance, some rules extracted from criminal and fire-fighting records are needed. The purpose of this paper is to mine association rules from a database to support crime management or fire-fighting resources allocation. The mining procedure consists of two essential modules. One is a clustering module based on a neural network, a Self-Organization Map (SOM), which performs grouping tasks on the tremendous number of database records. The another is a rule extraction module applying rough set theory that can extract association rules for each homogeneous cluster and the relationships between different clusters. An example is for illustration.

Keywords: Data mining, Self-Organizing Map（SOM）, Rough set theory, Rule-based knowledge, Crime management

1. Introduction

According to the investigations of the media, from the TV news, magazines, and questionnaires to the research, there have been more problems emerged in our society, such as the age of crime decreasing, economics recession, firing with uncertain reasons, and so on. The progress of technology has brought much convenience and change to our life. However, there has a big potential danger that the worthless fellows can use these new technologies to do the illegal activities with some amazing motivations. Although the police can introduce more efficient systems to figure out their commitment, it won't be last a long time if they don't have sufficient knowledge to create the suitable strategies.

Data mining is one of the most important techniques that can find potential useful knowledge, such as significant patterns and rules, from criminal and fire-fighting databases in supporting of making better decisions. The data mining tasks including association rules extracting, clustering, classifying, forecasting, and so on. An association rule is represented as $X \Rightarrow Y$, where X and Y are a set of items. The rule means that the records in database that contain X tend to contain Y. A good number of efficient algorithms for mining association rules have been proposed (Coenen and Leng, 2000; Mastsuzawa and Fukuda, 2000).

Clustering is a task of partitioning a population into a number of clusters or subgroups with similar features. There are numerous clustering algorithms such as the k-means method, the divisive methods, the grid-based method, the self-organizing map, and so on. Kiang, Kulkarni, and Tam used SOM to group similar parts into part families to be manufactured

（1995）; Vellido, lisboa, and Meehan（1999）used SOM to cluster data and find the potential interest for on-line marketers; Changchien and Lu（2001） used SOM and rough set theory to support on-line recommendation by customers and products fragmentation. While SOM networks have been successfully applied as a classification tool to various problem domains, including speech recognition, image data compression, image or character recognition, market segmentation, customer targeting, and medical diagnosis. In this paper, however, a neural network, a Self-Organizing Map (SOM), will be applied to cluster the criminal and fire-fighting records in order to discovery more knowledge and experiences to improve the quality and performance in the crime and hazard management. The SOM network serves as a clustering tool that can group criminals/fire accidents into appropriate categories based on the similarities in the characteristics of a given set of criminal/fire-fighting records. It performs unsupervised training for group technology so it doesn't require the knowledge of the corresponding output for comparison and learning. That is to say that the input patterns, during the process, adjust their weights based on the lateral feedback connections. The advantages of SOM network are good at transforming multi-dimensional inputs into a map of fewer dimensions and being easily used for classification of new entities into existing clusters without repeating the clustering process. Thus, it is very suitable for presenting the group distributions of the criminal records that are consisted of multiple attributes. But the drawback of SOM network is that it can't explain the clustering results specifically. This paper, for this reason, combines the rough set theory to achieve the goal—mining the association rules to explain the meaning of each cluster and the relationships between of them.

Rough set theory clarifies set-theoretic characteristics of the classes over combinatorial patterns of the attributes. This theory can be used to acquire some sets of attributes for classification and can also evaluate the degree of the attributes of database that are able to classify data. It uses the lower and upper sets that proposed by Pawlak in 1982（Lin and Cercone, 1997）. Several applications and extensions of the rough set theory have also been proposed. The reason for success in knowledge acquisition is that the rough set theory offers opportunities to discover useful information in training examples.

Coming with more social problems existed in our country, how to integrate the available resources（human knowledge and technologies） well to solve the exist problems and detect the potential problems is a challenging work. The police and fire fighters, hence, desire more available information and knowledge provided by the data mining techniques or knowledge extraction tools to assist them dealing with all kind of conditions. This paper, therefore, combine the strength of this two techniques to achieve the challenging task — more potential knowledge discovery in criminal and fire-fighting records.

The study is organized as follows. Section 2 explores some necessary background information while the mining approach is presented in Section 3. Section 4 uses a simple example to describe how self-organizing map and rough set theory applied in the crime management and Section 5 is the conclusion.

2. Literature Research

2.1 Self-Organizing Map（SOM）

The basic SOM network has an input layer and an output layer. When training data sets are fed into the network, SOM will compute a winning node. The winning node and neighborhood weights will be adjusted during learning (training) process until it converges to form clusters with similar properties. Figure 1 shows the flow chart of the SOM network for the training phases of the clustering module. The rules for clustering subgroups may be difficult to be defined, but SOM offers the flexibility to decide the composition and the number of clusters based on the selected objective and subjective criteria, such as criminal age, personality, type of commitment and so on. Beyond this, SOM network can be easily used for classification of new entities into existing clusters that the whole clustering process doesn't have to be repeated. The groupings can thus be continually adjusted in quick response to the change of environment.

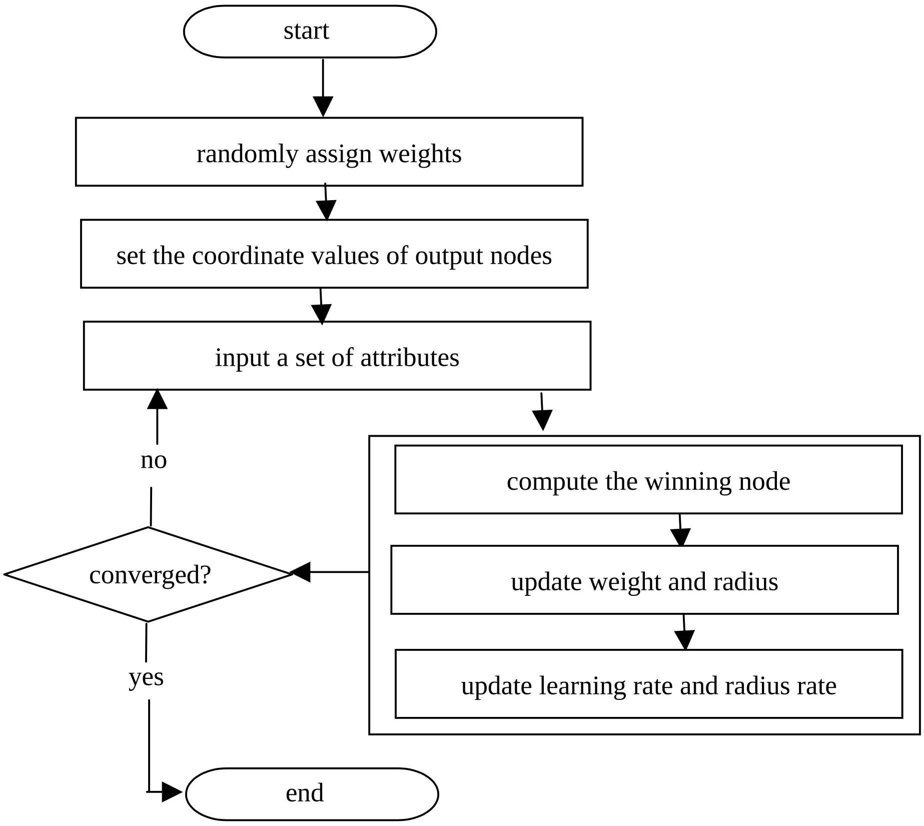


Figure1: The flow chart of SOM training procedure

2.2 Rough Set Theory

The rough set theory is a new mathematical approach to imprecision, vagueness, and uncertainty. The vague concepts were replaced by the precise concepts: the lower and the upper approximation of the vague concepts. Approximations are two basic operations in the rough set theory. The idea can be presented in the following manner. Let *U* be a universe, *A* be a finite set of attributes, *B* be a subset of *A*, *X* be a subset（a concept）of the universe, and *I* be an equivalence relation on *U*, called an indiscernible relation. *I(B)(x)* , in short *B(x)*, is an equivalence class containing an element *x*. Let us define the two basic operations, the B-lower and the B-upper approximations, on sets in the rough set theory:

$$B_*(X) = \{x \in U : B(x) \subseteq X\} \quad (1)$$

$$B^*(X) = \{x \in U : B(x) \cap X \neq \phi\} \quad (2)$$

We usually use the rough member ship function, called the confidence function to define approximations and the boundary region of a set. The confidence function is defined as:

$$CF(x) = \frac{Num(X \cap I(x))}{I(x)} \tag{3}$$

Where $CF(x) \in [0,1]$ denotes the degrees of how the element *x* belongs to the set *X* in view of the indiscernible relation *I*. The confidence function of the lower and the upper approximation can be redefined respectively as follows:

$$B_*(X) = \{x \in U : CF(x) = 1\} \tag{4}$$

$$B^*(X) = \{x \in U : CF(x) > 0\} \tag{5}$$

Applying the rough set theory to each cluster to find out some rules for association explanation, such as the characterization to the criminals.

The purpose of the paper is mining more detail and indiscernible relations to support the police and fire fighters making the best strategies in response to the changeable environment.

3. Knowledge discovery process

Knowledge discovery is an interactive and iterative process involving several steps. Data mining is one of the steps in it. Figure 2 shows the process of knowledge discovery while the data mining procedure is showed in Figure 3. In Figure 2, there are three procedures connected closely for achieving the goals.

First, managers select the target data set as the training data for the mining purpose and then to preprocess the selected data for the next procedure. The data mining procedure then starts to search for interesting patterns by the clustering module and rule extraction module. After the data mining of association rules, the work of rules evaluation is performed to eliminate weak rule under the initial values created by the knowledge workers. The useful rules are stored with different presentation styles in the knowledge bases that may be the case bases, rule bases and so on. They can be used as the strategies or some experiential laws in providing better recommendation and high-degree prediction. In this paper, we focus on clustering and association rules extraction in data mining tasks and following is the complete procedure of data mining for recommendation.

3.1 Step 1— data preprocessing

Before proceeding to data mining on data set, raw data must be preprocessed in order to decrease the mining time and facilitate the effectiveness. The data preprocessing process consists of two modules: clearing / selection module and transformation module.

1.) Data clearing / selection module: The users can retrieve data with constraints such as the range of an attribute for mining purposes. The procedure involves integration of multiple databases, removal of noises, and handling of missing data fields.

2.) Data transformation module: To have the same ranges, the target data should be normalized and/or scaled if it is necessary. Since this paper uses a SOM network for data clustering, the target data would be transformed to values ranged between 0 and 1. Here are two different formulas used to transform data.

    a.) If attribute = number then

$$response\_value_{jk} = \frac{(value_{jk} - min(att_j))}{(max(att_j) - min(att_j))} \qquad (6)$$

where response_value$_{jk}$ is the normalized value for the jth attribute of record k, min(att$_j$) is the minimum value of the jth attribute while max(att$_j$) is the maximum value of the jth attribute and value$_{jk}$ is the original value of the jth attribute of record k.

b.) If attribute = non-number then specific data transformation scaling should be assigned by the user. For instance, the data type of attribute job is character. The user has to assign a normalized numerical response value to it.

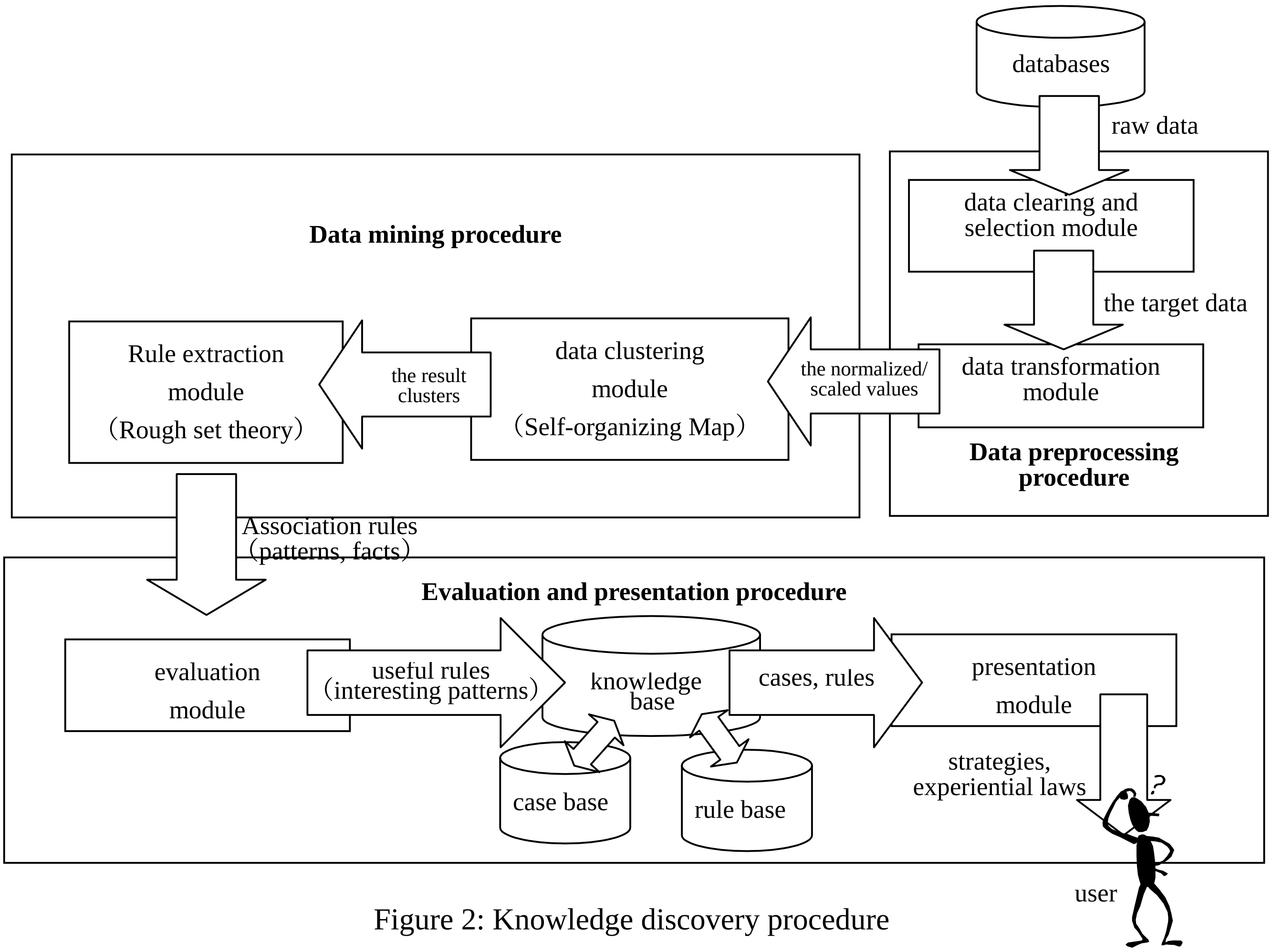


Figure 2: Knowledge discovery procedure

3.2 Step 2— data mining of association rules

When all attributes have the same measurement and ranges, then the association rules extraction process is proceeding. In this process, the SOM network based clustering module starts first and then a rough set theory based rule extraction module is employed to discovery association rules.

1.) Clustering module

The SOM network is an unsupervised two-layer network proposed by Kohonen (Changchien, and Lu, 2001) that can recognize a topological map from a random starting point. In SOM network, input nodes and output nodes are fully connected with each other. Each input node contributes to each output node with a weight. In Section 2, we show the flow chart for clustering. The SOM network can be assigned different numbers of cluster, learning rate, radius rate, converge error rate and so on. After the training phase, the weights are converged to the values that can be used to assign data items to their corresponding

$A_{ij^*}^{X_k}$

clusters. For instance, the fire managers can choose two attributes: damage degree and rescue need from the table fire accident as the SOM input nodes. Table 1 presents a simple example of clustering results of fire accidents. Two attributes, the degree of damage and the degree of necessary facilities for fire-fighting rescue, were normalized and clustered, where GID is a cluster number with special fire characteristics and ID is the accident's identification number.

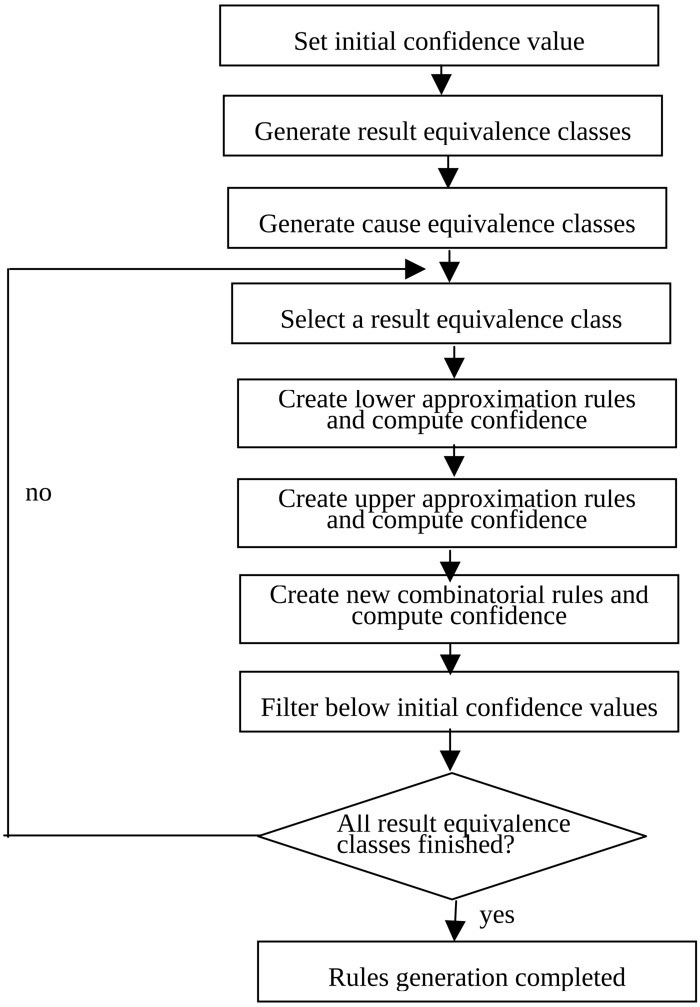


Figure 3: The flow chart of implementing rough set

2.) Rule extraction module

Rough set theory not only can extract association rules for each homogeneous cluster but also can find out the relationships for attributes among different clusters. Figure 3 shows the steps of rule extraction module and their descriptions are as below:

a.) Set initial confidence value. It is used to filter the lower and upper approximation rules unsatisfying the threshold value that is assigned by the users for holding the rules with high explainable abilities.

b.) Generate result equivalence classes. Let X_i denote the result equivalence class that consists of a set of objects for cluster *i*. Table 2 lists the clusters for a table fire accident. X_1 means that the fires in this cluster are belonged to the degree A.

c.) Generate case equivalence classes. Let Y_{jk} denote the cause equivalence class that consists of a set of objects having a specific attribute *j* with value of *k* (A_{jk}) . The Table 3 gives the case equivalence classes.

d.) Create lower approximation rules. Let $A_{jk*}(X_i)$ denote the set of objects that have the same attribute values（A_{jk}）all contained in cluster *i*. Here Eq. (1) is used to create lower approximation rules. For example, for $X_1 = \{2,5\}$ where GID = A （Table 2）, $A_{RN*}(X_1) = \{2\}$ satisfies the constraint so we have found a lower approximation rule, which is 'R1: If necessary rescue = N then GID = A'. Accordingly, all the lower approximation rules can be found and the confidence of every lower approximation rule is 100%.

e.) Create upper approximation rules and compute confidences. Let $A_{jk}^{*}(X_i)$ denote the set of objects that have the same attribute values（A_{jk}） only some but not all contained in cluster *i*. Here Eq. (2) and (3) are used to create upper approximation rules and to compute rules' confidences. Continue with the previous example, for GID = A, contained by only some of the objects in Y_{DH} and Y_{RH} that are contained in X_1. Hence there are two upper approximation rules for X_1, which are 'R2: If damage = H then GID = A' and 'R3: If necessary rescue = H then GID = A'. Using Eq. (3) to compute these two upper approximation rules' confidence. Here analysts can use the threshold value（minimum confidence） that are predetermined to decide which rules will be accepted. The confidence for R2 is 2/3 and for R3 is 1/3.

$$CF\left(A_{DH}{}^{*}(X_1)\right) = \left(\frac{2(ID(2,5))}{3(ID(2,5,6))}\right) = \frac{2}{3} \approx 0.67$$

$$CF\left(A_{RH}{}^{*}(X_1)\right) = \left(\frac{1(ID(5))}{3(ID(4,5,6))}\right) = \frac{1}{3} \approx 0.33$$

If the minimum confidence is set to 50%, only R2 will be accepted.

f.) Create new combinatorial rules. This step considers two or more attributes to generate association rules. First, we join each class in cause equivalence classes. In the above example, Figure 4 displays the results of joins of A_{RH} with the other attribute degree of damage to find the combinatorial rules. Only one node $A_{DH,RH}{}^{*}(X_1)$ that combines two classes $A_{DH}{}^{*}(X_1)$ and $A_{RH}{}^{*}(X_1)$ that is not empty.

$$\left(\left(X_1 \cap A_{DH,RH}{}^{*}\right) = \{5\}\right)$$

The confidence of $A_{DH,RH}{}^{*}(X_1)$ therefore is calculated as follows:

$$CF\left(A_{DH,RH}{}^{*}(X_1)\right) = Min\left(\frac{1}{3},\frac{1}{3}\right) = \frac{1}{3} \approx 0.33$$

Here we can define the confidence function of an n-attribute combinatorial rule:

$$CF(x) = Min\left(\frac{Num((X_1 \cap Y_{1k}))}{Num(y_{1k})}, \frac{Num((X_1 \cap Y_{2k}))}{Num(Y_{2k})}, \ldots, \frac{Num((X_1 \cap Y_{nk}))}{Num(Y_{nk})}\right) \quad (7)$$

So we create a new combinatorial rule 'R4: If damage = H and rescue = H then GID = A CF = 0.33（equals to the CF($A_{RH}*(X_1)$)）.

g.) Explain the characteristics of each cluster. Rules from R1 to R4 are used to explain the meaning of cluster A. In the above example, the fire accidents in cluster A have following characteristics: 100% accidents in cluster A need the fire-fighting facilities（including people, fire engines, etc) belonged to normal level, about 1/3 accidents in cluster A whose damage level is high and rescue need is high, and about 2/3 accidents in cluster A whose damage is high.

h.) Go to d.), to process the next equivalence class X_2 and repeat until all the result equivalence classes are completed.

Table 1: Records of fire accidents

N:normal H:high L:low

ID	Damage(D)	Rescue(R)	GID
1	N	L	B
2	H	N	A
3	L	L	C
4	N	H	B
5	H	H	A
6	H	H	B

Table 2: Results equivalence classes

i	GID	X_I
1	A	{2,5}
2	B	{1,4,6}
3	C	{3}

Table 3: Cause equivalence classes

A_{jk}		Y_{jk}
j = D	k = N	Y_{DN} = {1,4}
j = D	k = H	Y_{DH} = {2,5,6}
j = D	k = L	Y_{DL} = {3}
j = R	k = N	Y_{RN} = {2}
j = R	k = H	Y_{RH} = {4,5,6}
j = R	k = L	Y_{RL} = {1,3}

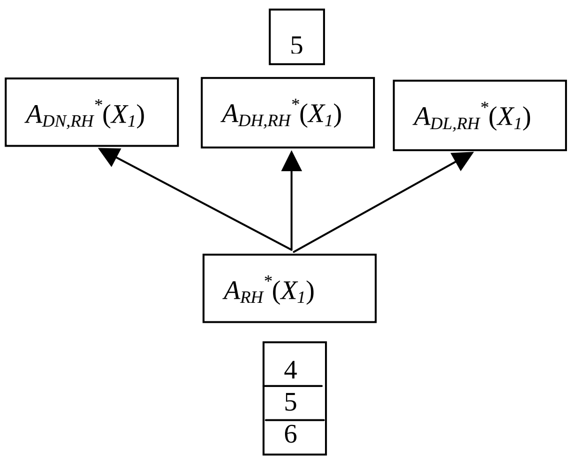


Figure 4: The combinatorial rules

3.3 Step 3— evaluation and presentation

After all related rules are created, analysts can use evaluation module to eliminate unused or redundant rules in order to keep the most valuable rules in the knowledge bases that may be the case bases or/and rule bases for future use.

According to the fire characteristics（size, features, and threat） in the fire table and other potential associations of interesting attributes（factors to catch fire, firing points, and intervention evaluation） in the fire-fighting accident table, we can create many rules with different purposes. The fire managers retrieve these rules stored in the knowledge bases to determine the best resource allocation（number of the fireman and fire engines needed to rescue）and to predict the locations （factories, hotels, markets）and the time periods（night, months, seasons）that may be easily caused fires for enhancing the fire prevention management.

4. Example for criminal management

We present an example to illustrate the application of data mining that may support crime management. The aim in this example is to mine the associations between the criminal personalities and the crimes. Since the criminal records and the personalities characterized by the psychologists can't be easy to retrieve, we assume all crimes and personal characteristics were pre-classified into seven clusters and five clusters, respectively. Here we select 100 samples from criminal table to test. To simplify, the five interesting attributes （crime type, criminal characterization, location, times to crime, and way adopted） were selected from the criminal table for clustering by its importance to crime prevention where two attributes, the crime type and criminal characterization, are the corresponding cluster numbers based on the predefined clusters. The approach for attribute selection used herein is to select the attributes with high weights manually and dynamically from the fact tables that are considered to correlate closely with the mining purpose by the manager. We use a clustering tool, J-Express （the SOM network）, and a statistical software, SPSS for windows, to perform group tasks and rules extraction, respectively. Table4 lists the criminal records. After the data preprocessing procedure, we fed these data into the SOM network for clustering. Figure 5(a) displays the training results of the SOM network while the clustering results of ten clusters and number of records in each cluster are shown in Figure 5(b). In Figure 5(c), for instance, there are 13 criminal records with similar features in the cluster 2, which is selected from the upper middle area in Figure 5(b). Each criminal identifier in the same cluster is mapped to the corresponding line at the right-hand side chart. We can use a rough set based rules extraction module to generate association rules describing the clustering results and relationships between different clusters. The result equivalence classes and cause equivalences classes are generated for creating association rules. They can be represented as the frequencies of criminals for each attribute with specific value emerged in the clusters and in the population, respectively. Figure 6(a), 6(b), and 6(c) give the examples of the cause equivalences classes that are the emerged frequencies of seven crime clusters, of five criminal personality clusters, and of the times to crime in the population, respectively.

Additionally, the examples of the result equivalences classes are summarized in Figure 7(a), 7(b), and 7(c). Figure 7(a) displays the emerged frequencies of personality cluster 2 and more than three times to crime in ten clusters, Figure 7(b) depicts the emerged frequencies of personality cluster 2 and crime-ridden way cluster 9 in ten clusters, and the emerged frequencies of two important locations in ten clusters, Taipei and Taichung, that are taken account into the locations with high-crime possibility in Taiwan are shown in Figure 7(c). Next we use Eq. (3) to compute the confidences for creating lower and upper approximation rules. Here the accepted confidence value is set to be 0.5 to filter unused associations. Table 5 summarizes the interesting facts in the clusters and the associations between different clusters.

Table 4: Criminal records

ID	Crime-type	Personality	Locations	Times-to-crime	Ways
1	Group 3	Group 2	Taipei	2	knife
2	Group 4	Group 3	Nanto	3	Gun
3	Group 6	Group 4	Taichung	1	trick
…	……..	……..	……..	…	……..
…	……..	……..	……..	…	……..
99	Group 3	Group 3	Kaochung	2	cord
100	Group 5	Group 2	Wufon	1	Brute force

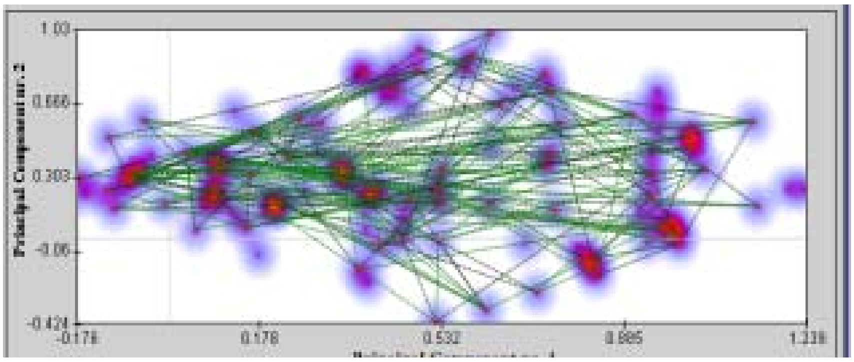


Figure 5(a): Training results of SOM network

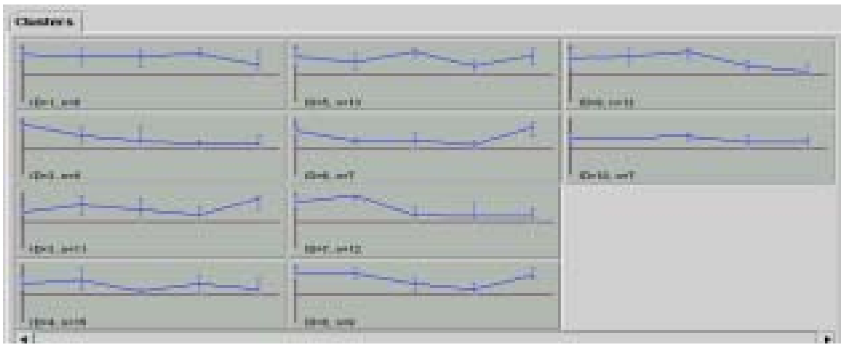


Figure 5(b): The number of clusters and records in each cluster

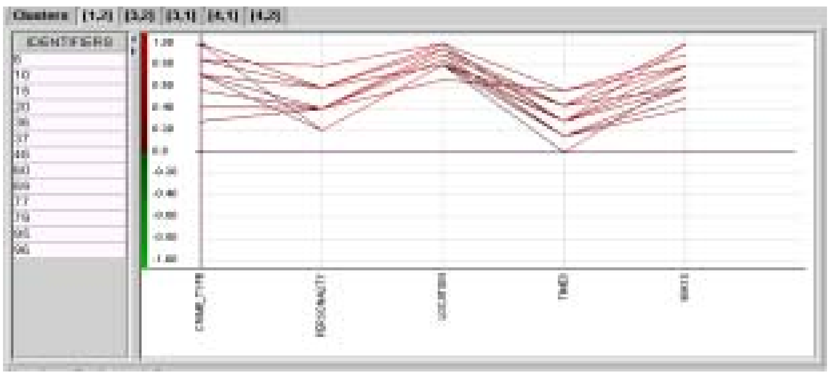
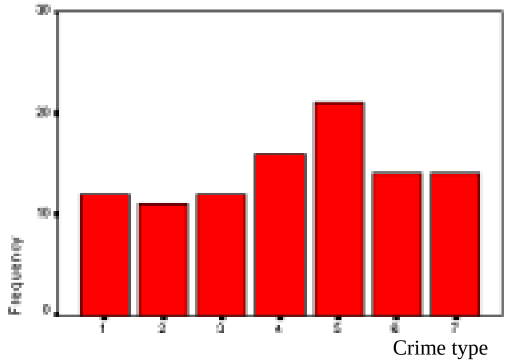
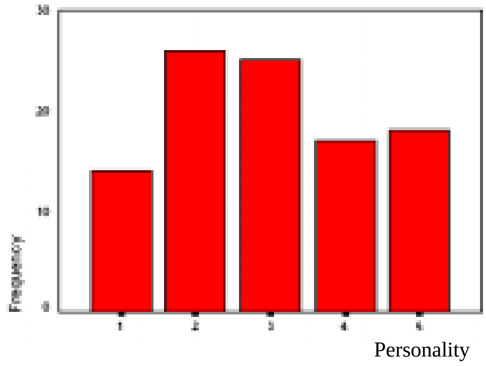


Figure 5(c): The clustering results in the cluster 1

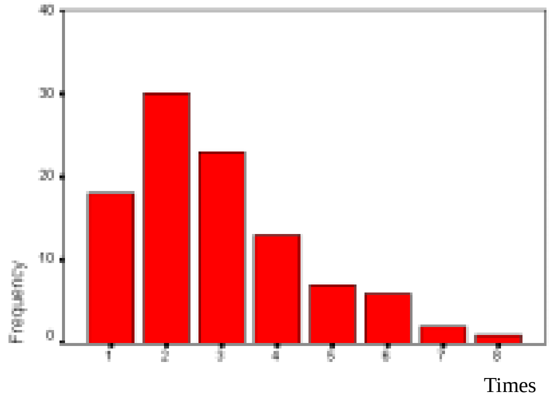
6 (a)



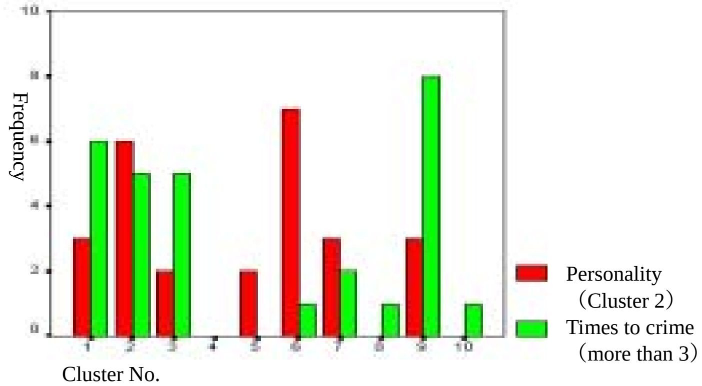
6(b)



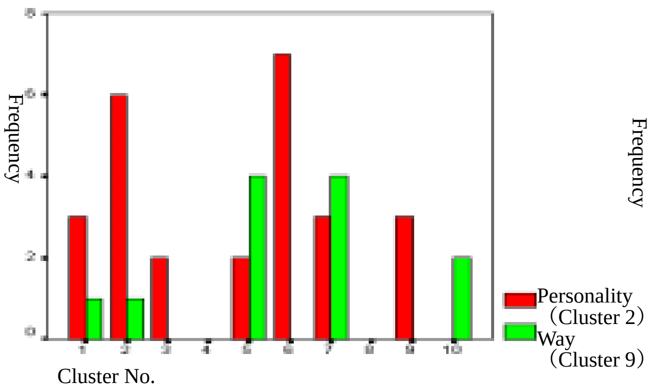
6 (c)



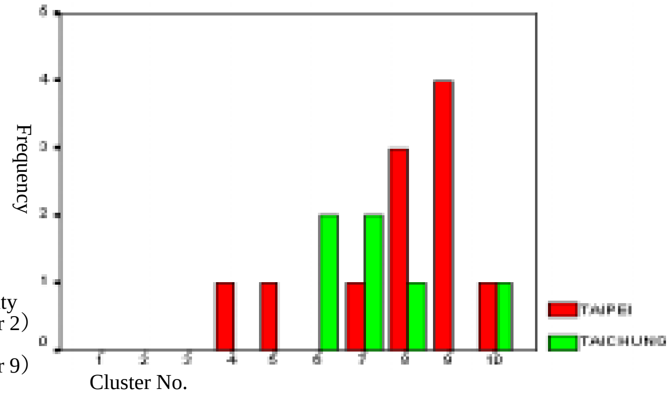
7 (a)



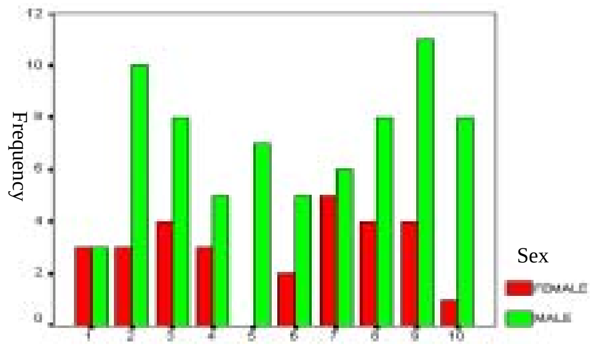
7 (b)



7(c)



8 (a)



8 (b)

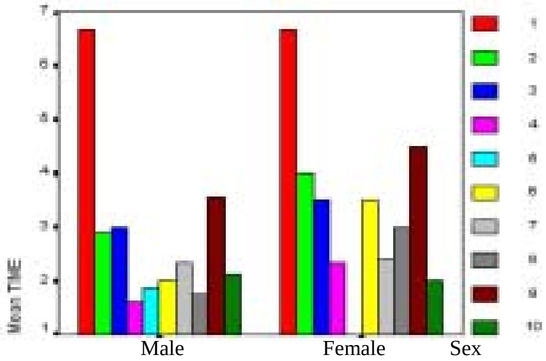


Table 5: the rules extracted from ten clusters

No.	Features	No.	Features
1	All criminals in this cluster possess more than six times to crime.	2	About 67% criminals emerge in Taipei and Kaochung.
3	About 67% criminals emerge in Tainan.	4	About 50% events are belonged to crime cluster 7.
5	About 50% criminals emerge in Nanto.	6	All criminals in this cluster with personality cluster 2.
7	About 75% criminals emerge in Taichung and favorite way is gun	8	About 56% criminals with personality cluster 5.
9	More than 60% criminals emerge in Taipei and Tainan	10	About 60% criminals tend to brute force

From the Table 5, we conclude the characteristics of each cluster and an implicit relationship between clusters 3 and 9 that Tainan is a favorite location for both criminals. Beyond to mine the associations between these items, many interesting attributes such as sex, age, and so on are also can be applied in our proposed mining procedure. For example, the diffused status of age and mean times to crime for the male and the female in each cluster are shown in Figure 8(a) and Figure 8(b), respectively. In a word, we can create as many useful rules with different interesting purposes as we can.

From the mining procedure, we can conclude many useful rules. These rules can be further applied to predict and to prevent from some events emerging in the future. For example, if a new case is grouped into cluster 7 then the police might assign more human resources around Taichung to against evildoers. Therefore, the discovered association of attributes in criminal table and relationships between different clusters can assist not only the further strategies making but also the better crime management.

5. Conclusion

This paper has briefly described how to integrate a neural network—SOM and rough set theory to discover association rules. Clustering module based on the SOM network separates the crimes and criminals into clusters and a rule extraction module is used to characterize each cluster and explain the relationships among different clusters. These rules may help the police offices to find out the crime and criminal fragmentation so the new suspect will be fast clustered into one of suitable groups according to his motivations on crime and personalities.

The data mining procedure offers not only the associations between criminal personalities and the crimes they made but also the suggestions of which areas the crimes prefer to skulk and possible commitments he will make next time. Furthermore, police can select the important attributes to cluster their criminals in order to enhance the performance on crime management and the suspects' behavior detection.

The data mining approach presented above can be used in the fraud detection, fire prevention, and many other fields. However, there is still space for the future research:

1.) A tool with intelligent ability may be developed to dynamically include feedback and attributes for analysis.

2.) The clustering module based on the SOM network would be combined with the fuzzy concept to group uncertain or non-numerical data in order to discover the associations more close to the real environment.

3.) The analysis and development of the prototype on the crime management or fire-fighting resources allocation may take into account more factors for future implementation.

Besides, the more efficient association rules mining and filtering algorithms may also be the future work.

Reference

[1] Changchien, S.W. and Lu, T.C., “Mining association rules procedure to support on-line recommendation by customers and products fragmentation,” Expert Systems with Applications (20) 2001, pp: 325-335.

[2] Kiang, M.Y., Kulkarni, U.R., and Tam, K.Y., “Self-organizing map network as an interactive clustering tool — An application to group technology,” Decision Support Systems (15) 1995, pp: 351-374.

[3] Kiang, M.Y., “Extending the Kohonen self-organizing map networks for clustering analysis,” Computational Statistics and Data Analysis (38) 2001, pp: 161-180.

[4] Kim, K.S. and Han, I., “The cluster-indexing method for case-based reasoning using self-organizing maps and learning vector quantization for bond rating cases,” Expert Systems with Applications (21) 2001, pp: 147-156.

[5] Kuo, R.J., Chi, S.C., and Teng, P.W., “Generalized part family formation through fuzzy self-organizing feature map neural network,” Computer and Industrial Engineering (40) 2001, pp: 79-100.

[6] Lin, T.Y. and Cercone, N., “Rough Sets and data mining — analysis of imprecise data,” Kluwer Academic Publishers, 1997.

[7] Mastsuzawa, H. and Fukuda, T., “Mining structured patterns from databases,” proceeding of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2000, pp: 233-244.

[8] Vellido, A., Lisboa, P.J., and Meehan, K., “Segmentation of the on-line shopping market using neural networks,” Journal of Information Science and Engineering (14) 1999, pp: 843-862.