

## **GDP Growth vs. Criminal Phenomena: Data Mining of Japan 1926-2013**

Xingan Li<sup>a</sup>, Henry Joutsijoki<sup>b</sup>, Jorma Laurikkala<sup>b</sup>, Martti Juhola<sup>b\*</sup>

<sup>a</sup> School of Governance, Law and Society, Tallinn University, Estonia

<sup>b</sup> Computer Science, School of Information Sciences, University of Tampere, Finland

**Abstract** The aim of this article is to inquire about potential relationship between change of crime rates and change of GDP growth rate, based on historical statistics of Japan. This national-level study used a dataset covering 88 years (1926-2013) and 13 attributes. The data were processed with the Self-Organizing Map (SOM), separation power checked by our ScatterCounter method, assisted by other clustering methods and statistical methods for obtaining comparable results. The article is an exploratory application of the SOM in research of criminal phenomena through processing of multivariate data. The research confirmed previous findings that SOM was able to cluster efficiently the present data and characterize these different clusters. Other machine learning methods were applied to ensure clusters computed with SOM. The correlations obtained between GDP and other attributes were mostly weak, with a few of them interesting.

**Keywords** Data mining; Self-organizing map; Classification methods; Japan; GDP growth rate; Crime rate; Development of criminal phenomena

\*Corresponding author. Tel: +358 40 1901716; fax: +358 3 2191001

E-mail address: [Martti.Juhola@sis.uta.fi](mailto:Martti.Juhola@sis.uta.fi)

### **1 Introduction**

Data mining methods are broadly used in recent decades in research of many disciplines. Criminal phenomena can be observed and studied from different perspectives, of which data mining methods are playing an increasingly significant role. These methods enable research with demographic, psychological, economic, socio-legal, and historical indicators. The self-organizing map (Kohonen 1979), which employs an unsupervised learning approach to cluster and visualize data in accordance with patterns identified in a dataset, is a proficient apparatus

designed for such data investigation. The interaction between artificial intelligence and research of criminal phenomena facilitates an innovative study.

On one hand, for years, the capacity of the SOM of identifying acts that are suspected of guilty or abnormal has been studied in a broad range. Here there are just some of the examples that were frequently mentioned: the detection of automobile bodily injury insurance fraud (Brockett, Xia, and Derrig, 1998), homicide (Kangas et al., 1999; Memon, and Mehboob, 2006), mobile communications fraud (Hollmén, Tresp, and Simula, 1999; Hollmén, 2000; Grosser, Britos, and García-Martínez, 2005), murder and rape (Kangas, 2001), burglary (Adderley and Musgrave, 2003; Adderley 2004), network intrusion (Axelsson, 2005; Lampinen, Koivisto, and Honkanen, 2005; Leufven, 2006), cybercrime (Fei et al., 2005; Fei et al., 2006), and credit card fraud (Zaslavsky, and Strizhak, 2006). These are primary fields where the application of the SOM has been emphasized in the research of criminal phenomena. These can be regarded as microscopic research.

On the other hand, research in criminal phenomena in general at international level and national level has also been developed. Li and Juhola (2013) and Li (2014) applied the SOM in the study of criminal phenomena based on international databases, assisted with some other data mining techniques. The research dealt with the relationship between crime and demographic factors (Li and Juhola, 2014a; Li et al. 2015a), economic factors (Li and Juhola, 2015), historical developments of criminal phenomena in the USA (Li and Juhola, 2014b), and that between a particular offence, homicide and its social context (Li et al., 2015b). These studies concluded that the SOM, in addition to its application in microscopic research, could be a helpful instrument for research in crime at international and national levels based on relevant statistical data. However, the conclusions also revealed that more numbers of more extensive experiments in the same field would be expected and necessary.

While Li and Juhola (2014b) was an innovative study of using the SOM in the research of criminal phenomena based in the USA on a multivariate historical data at the national level, interests also emerged in expanding the research to more jurisdictions, wherever historical data are available and feasible for process with such methodologies, in order to acquire comparative results. The present study applies the SOM to the field of macroscopically exploring into multi-dimensional data of development of criminal phenomena in Japan, over a span of time of 88 years, with emphasis on the interaction between GDP (Gross Domestic Product) growth rate and change of crime rates, aiming at seeking an innovative field in which artificial intelligence can play a role in simplifying the process of analysis. In a word, this article endeavors to make

inquisition into relationship between change of GDP growth rate and change of crime rates in Japan.

In sum, the article continues to explore the advantage of the SOM in research in criminal phenomena with assistance of other clustering and statistical methods, by processing available and feasible data sets. This national-level study uses a dataset covering 88 years and 13 attributes. The data will be processed with the Self-Organizing Map (SOM), refined by ScatterCounter (Juhola and Siemala, 2012), assisted with some other machine learning methods, and statistical techniques for verification.

Following this section, the next section of the article will briefly introduce the methods used in processing crime-related data. In the third section, a brief introduction to crime in modern Japan will be presented. In the fourth section, information will be given about how the experiments are designed. The fifth section will present results and discussions. The final section concludes the article with findings from the data mining of crime and its demographic factors. Due to the fact that the SOM has not frequently been used in the study of criminal phenomena in the similar way, and that such a study is more methodology-oriented in criminology than application-oriented in computer science, it is expected that the research will be a valuable experiment in exploiting of statistical data.

## **2 Methodology**

The SOM, developed by Kohonen (1979) to cluster and visualize data was used in this study. The SOM is an unsupervised learning mechanism that clusters objects with multi-dimensional attributes into a lower-dimensional space, in which the distance between every pair of objects captures the multi-attribute similarity between them. Upon processing the data, maps will be generated using software packages. By observing and comparing the clustering map and feature planes, there is the potential to explore into the correlation between crime and demographic indicators. These results, including clustering maps, feature planes as well as correlation tables constitute the fundamental ground for further analysis.

This study applies the SOM to historical development of crime in Japan during 88 years (1926–2013). Including an analysis based on available data, the results of the study will revolve around whether the SOM can be a feasible tool for mapping criminal phenomena through processing of large amounts of multidimensional historical data, and to what extent interaction between GDP growth rate and change of crime rates can be expected.

For the application of the SOM the method called ScatterCounter (Juhola and Siermala, 2012) was used for attribute selection by computing separation powers for attributes. In other words, it was determined which attributes are strong in classification and which are poor. The latter ones could be removed from the dataset and the reduced dataset will be used in final processing and analysis. If all attributes are good as to their separation powers, they can be all reserved for analysis. To apply ScatterCounter missing data in the original dataset have to be filled with estimated values. We filled them by the means of the available values of the attributes in the same clusters.

Apart from the SOM, discriminant analysis, *k*-nearest neighbor classifier, Naïve Bayes classification, decision trees, random forests and support vector machines (SVMs) will also be used to validate the clusters by computing how accurately these methods classify the same countries into the same clusters compared to those of the SOM.

### **3 Brief history of crime in Japan**

Generally, Japan has had low crime rates compared with other industrialized countries. According to the data set used in this paper, for example, homicide rate in Japan during the last 90 years never surpassed 5 per 100,000 people. Today, when the USA maintains a homicide rate of 5.8 per 100,000 people, Japan has less than 0.8 per 100,000 people. Other crime rates are also lower. Of course, the homicide rate in the USA can also be regarded as low, if it is compared with some figures from other countries, for example, the records of the highest homicide rate in the world were 101 per 100,000 people in Iraq in 2006, 89 in Iraq in 2007, and 88.61 in Swaziland in 2000. After looking at these figures, generally speaking, violence and homicide in developed countries are the lowest in the world, for example in Germany, Denmark, Norway, Japan, and Singapore, with homicide rate below 1 per 100,000 inhabitants.

Rapid development of Japan in modern history started long before 1926, known from the mid-1800s. From the record of our dataset, from 1926 to the mid-1930s, violent crime rates almost all stable, but property crime rates almost doubled, forming the first peak of crime in the 88-year period. During this period, Japanese GDP growth rates were not stable, as low as -7.27 in 1930, but as high as 9.82 in 1933.

From 1935, major crime rates obviously lowered up to the year 1945, when Japan surrendered due to defeat in the WWII. In fact, most crime rates in 1945 fell into a valley. During

this period, Japanese GDP growth rates were not stable either, as low as -4.30 in 1942, but as high as 15.75 in 1939. The year 1945 recorded -50% GDP decrease.

From 1946, however, a sharp increase of crime rates was seen, to reach a peak in 1948, followed with a quarter of century of decrease until the valley of 1973. During this period, Japanese economy enjoyed a high speed increase, with GDP growth rates ranging from 4.69% to 14.88%, but an average as high as 9.34%. It was these three decades of rapid development that facilitated Japan to become an industrialized modern economy.

From 1974, the crime rates increased year by year again, forming another peak in early 2000s (theft 2002, fraud 2005, embezzlement 2004, blackmail 2001, burglary 2003, homicide 2003, abduction 2004, rape 2003, indecent assault 2003, injury 2003, robbery 2003, arson 2004). For the first time in the past quarter of century, Japanese GDP growth rate fell to -1.22% in 1974. Thereafter, the rates were between -1.22% to 6.19%, and averaged 2.88%. Generally, Japanese economy was still growing. But after 2001, its growth stopped.

Thereafter, a new round of sharp fall began. Now the crime rates in Japan are roughly equal to the level of end of 1920s, end of 1930s and early 1940s, and that of 1945. GDP growth rates were between -2.3% to 2.4%, an average of -0.5%.

Here we have already had a clear illustration of the change of crime rates and GDP growth rate. In this article, the topic concerning the rise and fall of Japanese crime will be examined from a new stand using a new method, the SOM and some other clustering and statistical methods, so as to check whether there is a potential internal mechanism affecting the interaction between them. Traditionally, there did not lack such hypotheses and conclusions that demonstrate a close relationship between crime and economy, based on both qualitative and quantitative analysis. The aim of this research is to provide further understanding of the relationship between economic growth and crime by applying several clustering and statistical methods, taking Japanese historical data as an example.

## **4 Design of experiments**

### 4.1 Period covered

Modernization and Westernization of Japan started in the mid-1800s, which marked a new era of Japanese society. However, the data used in this study covers a period of 88 years. These years were selected based on the availability of data on their selected indicators. In fact, during the years of 1933-2013, data of all the attributes are available, while during the years of 1926-1932, only the data of rape and indecent assault are missing.

### 3.2 Attributes

A synopsis of all attributes that were used in this study is given in Table 1. One of them is GDP growth rate. The other 12 attributes are such crime rates that are usually recorded by Western countries as most important indicators to measure criminal phenomena in those countries. The selection of the contents of these indicators was principally based on availability of data. See also Fig. 1 and Fig. 2.

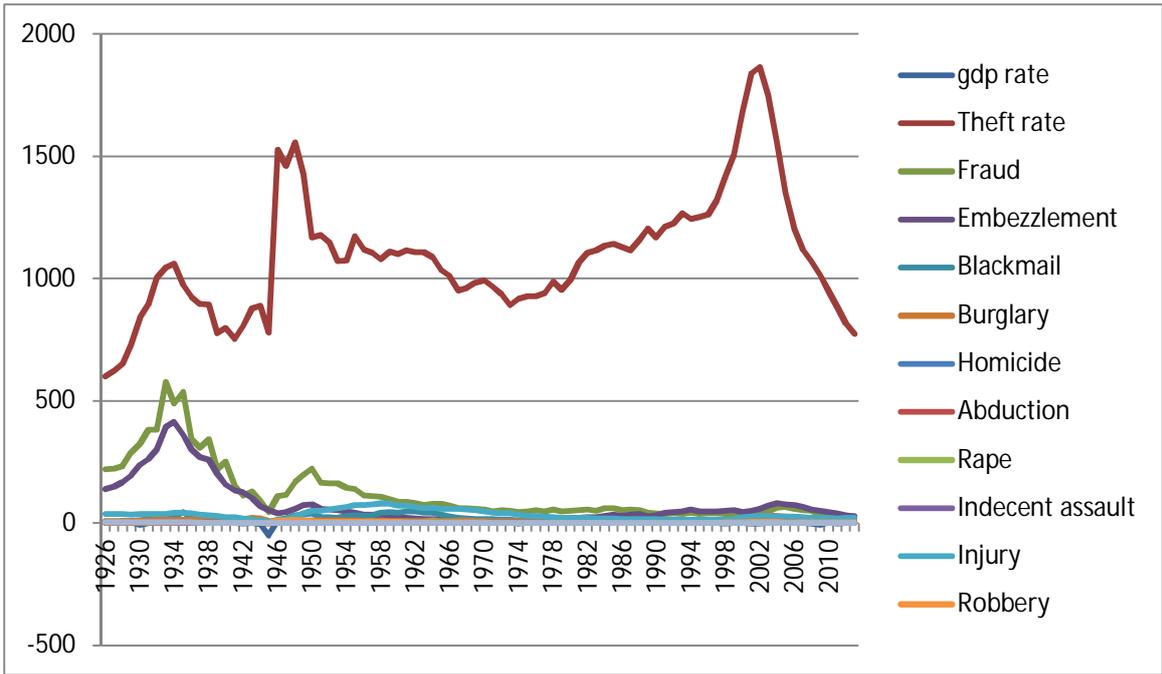
**Table 1** GDP growth rate and criminal phenomena indicated by 12 different attributes

Non-crime attributes	Name	Codification
1	GDP growth rate	GDP
Crime-related indicators	Name	Codification
2	Theft	THE
3	Fraud	FRA
4	Embezzlement	EMB
5	Blackmail	BLA
6	Burglary	BUR
7	Homicide	HOM
8	Abduction	ABD
9	Rape	RAP
10	Indecent assault	IND
11	Injury	INJ
12	Robbery	ROB
13	Arson	ARS

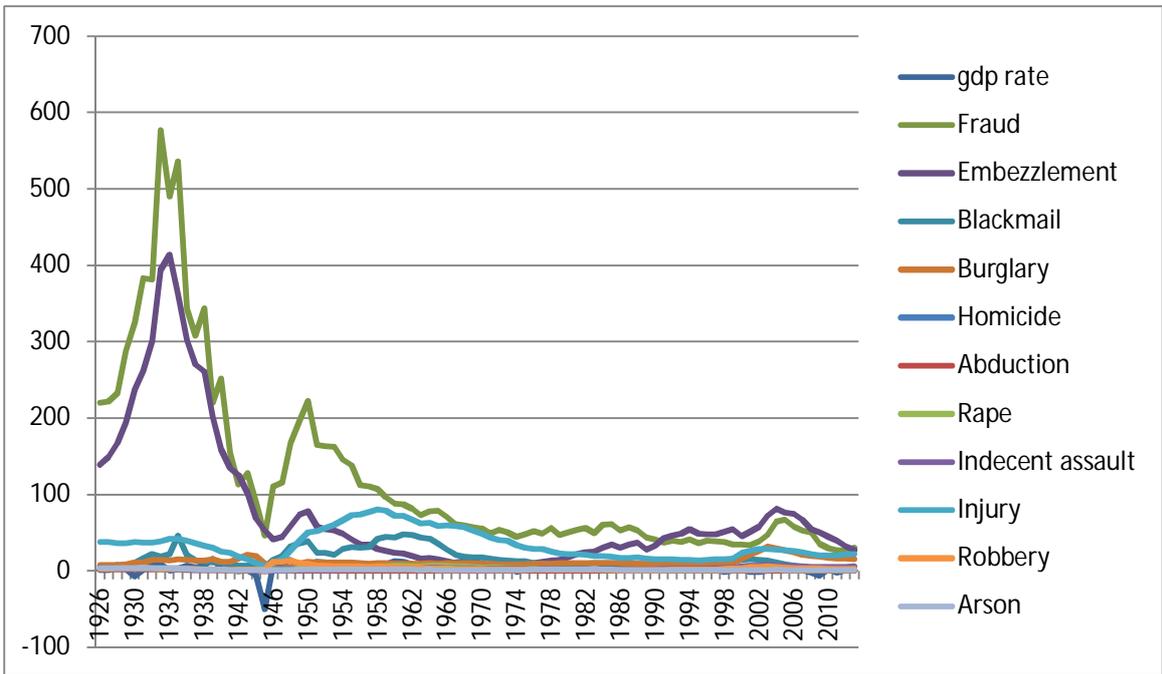
There have not been standard abbreviations in use for shortening attributes. Information about most items was derived from the database of Japanese Ministry of Justice. Unavailable items were imputed by attribute means. The sources of data are listed below in Table 2.

**Table 2** Sources of data

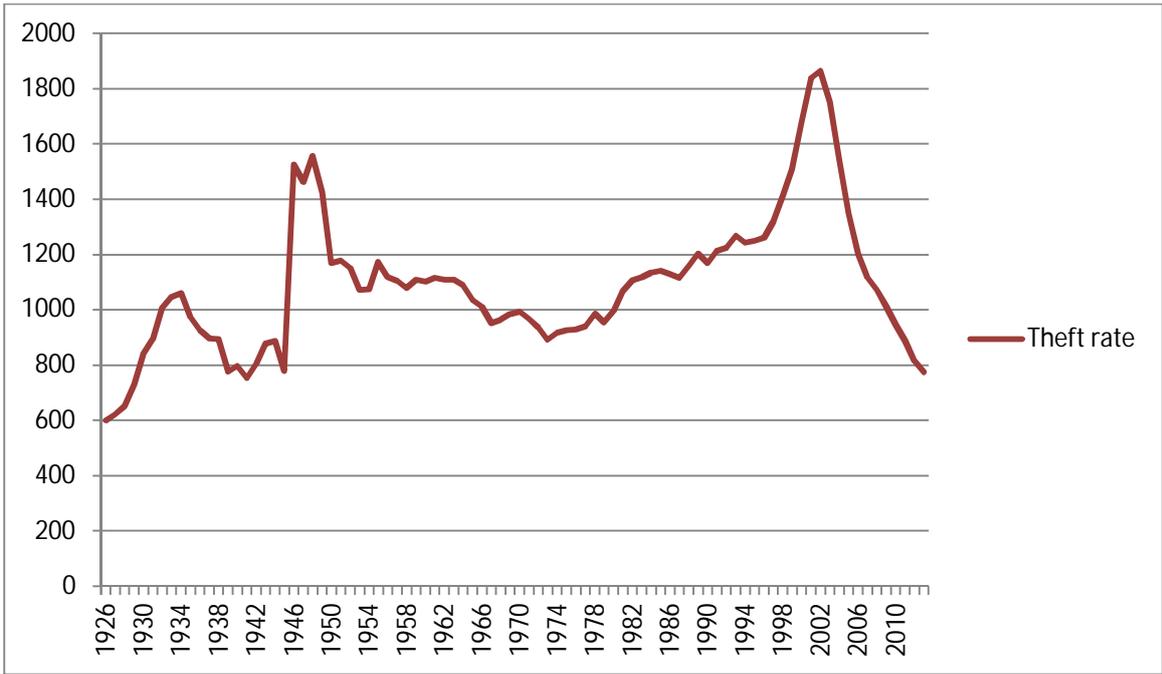
Institutions	Websites
Ministry of Internal Affairs and Communications	<a href="http://www.soumu.go.jp/menu_seisaku/toukei/">http://www.soumu.go.jp/menu_seisaku/toukei/</a>
National Police Agency	<a href="https://www.npa.go.jp/toukei/index.htm">https://www.npa.go.jp/toukei/index.htm</a>
Ministry of Justice	<a href="http://www.moj.go.jp/hakusyotokei_index.html">http://www.moj.go.jp/hakusyotokei_index.html</a>



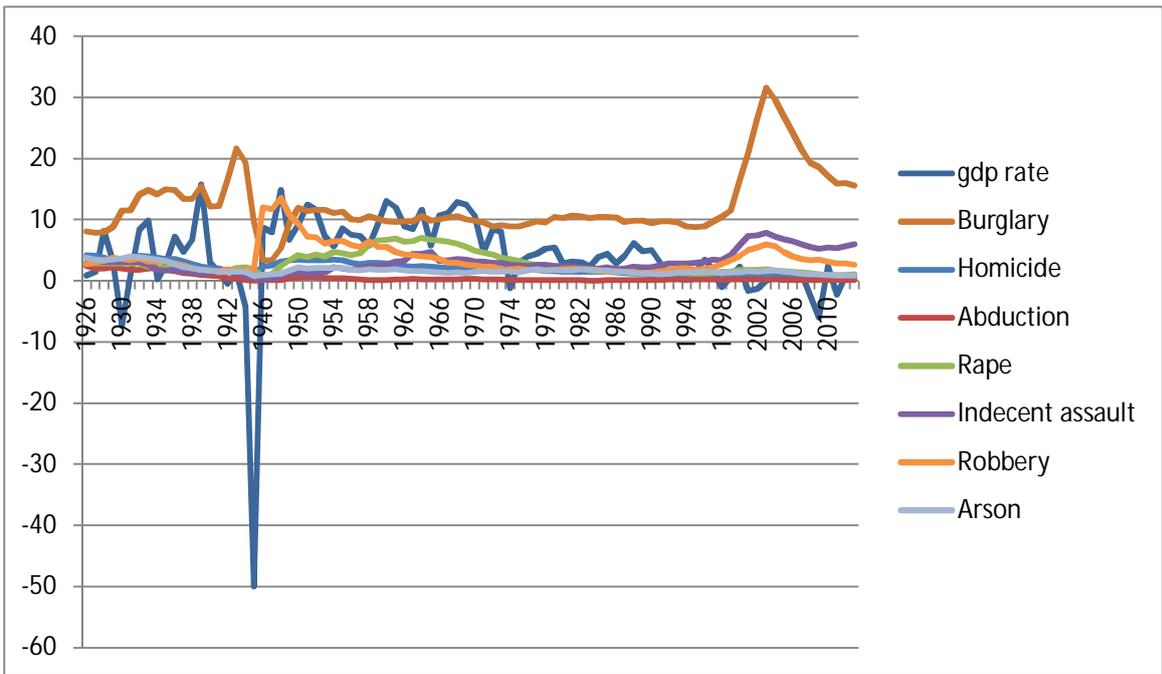
(a) GDP rate and crime rates



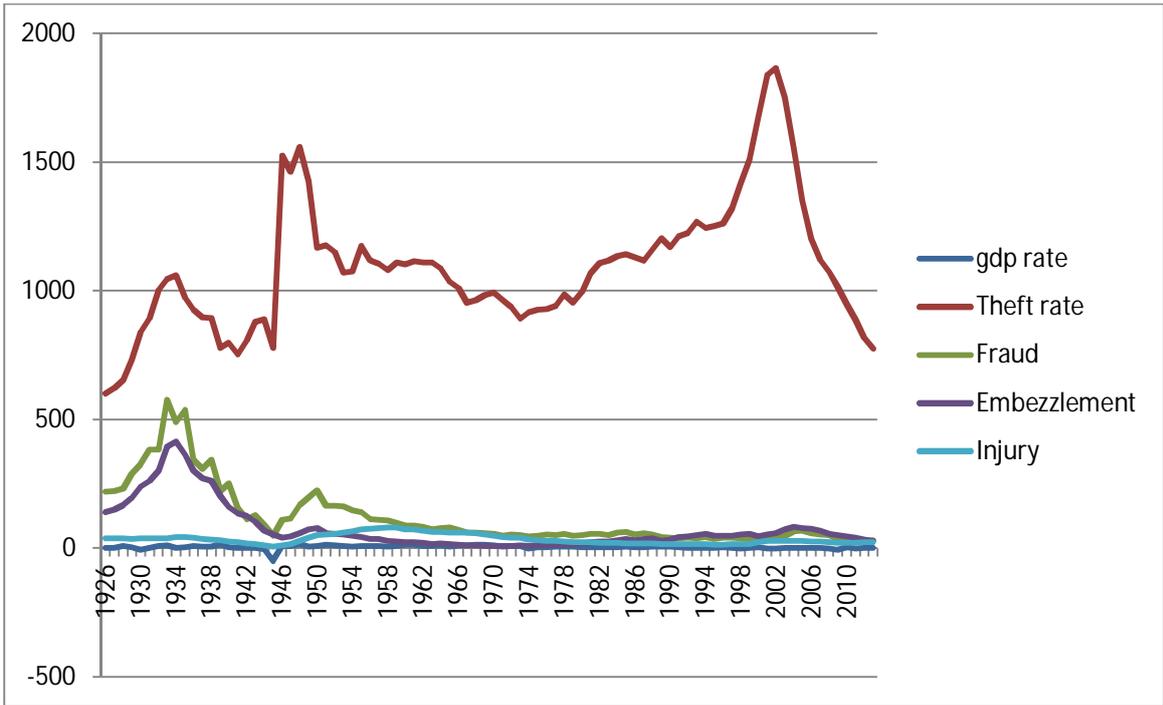
(a) GDP rate and crimes rates (extraordinary case Theft left out)



**(b) Theft rate**

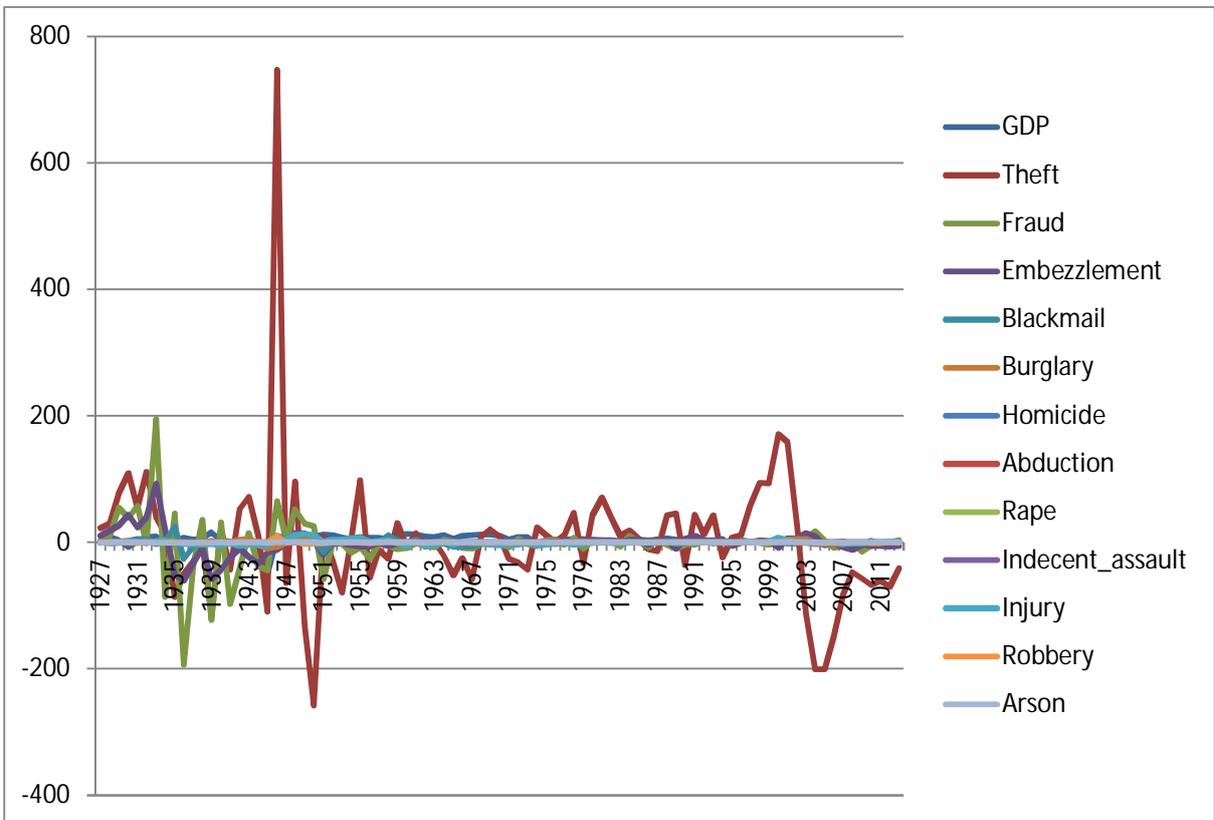


**(c) GDP rate and crime rates (theft, fraud, embezzlement, injury rates left out)**

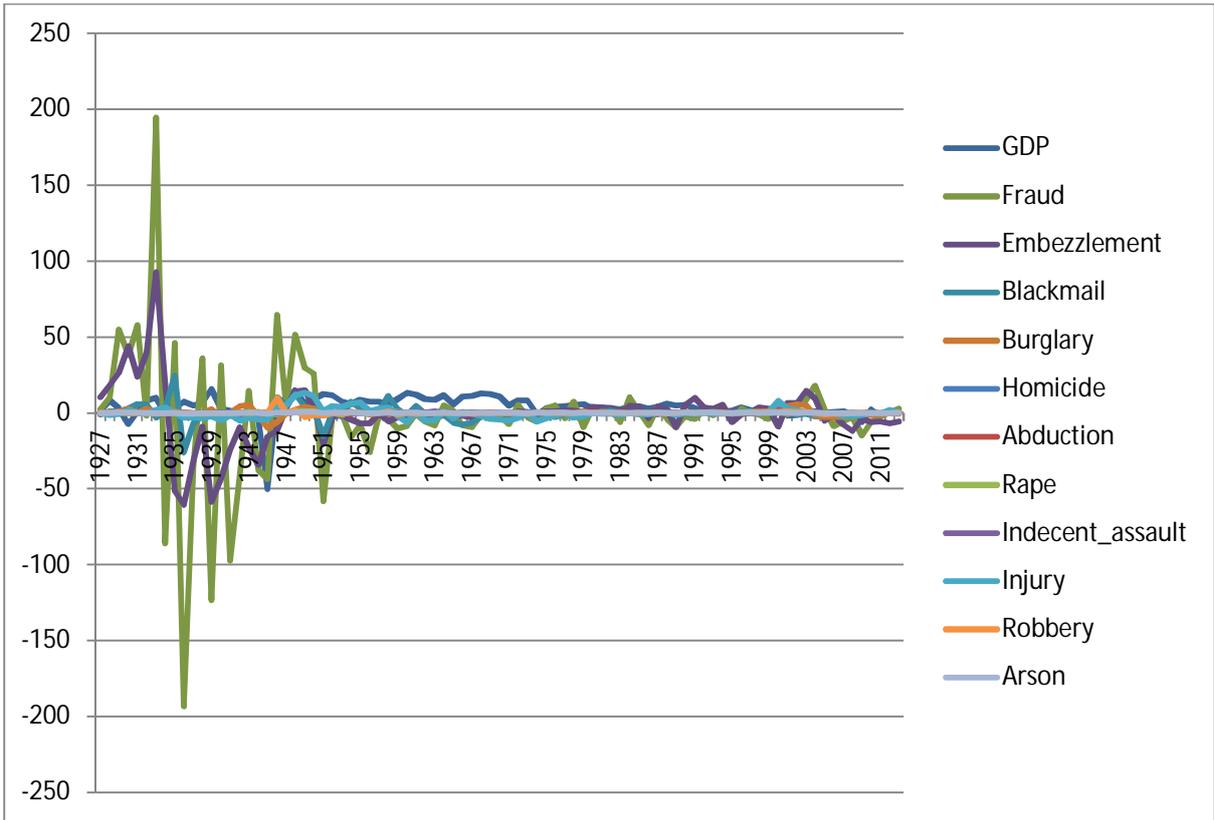


(d) GDP rate and rates of theft, fraud, embezzlement, and injury

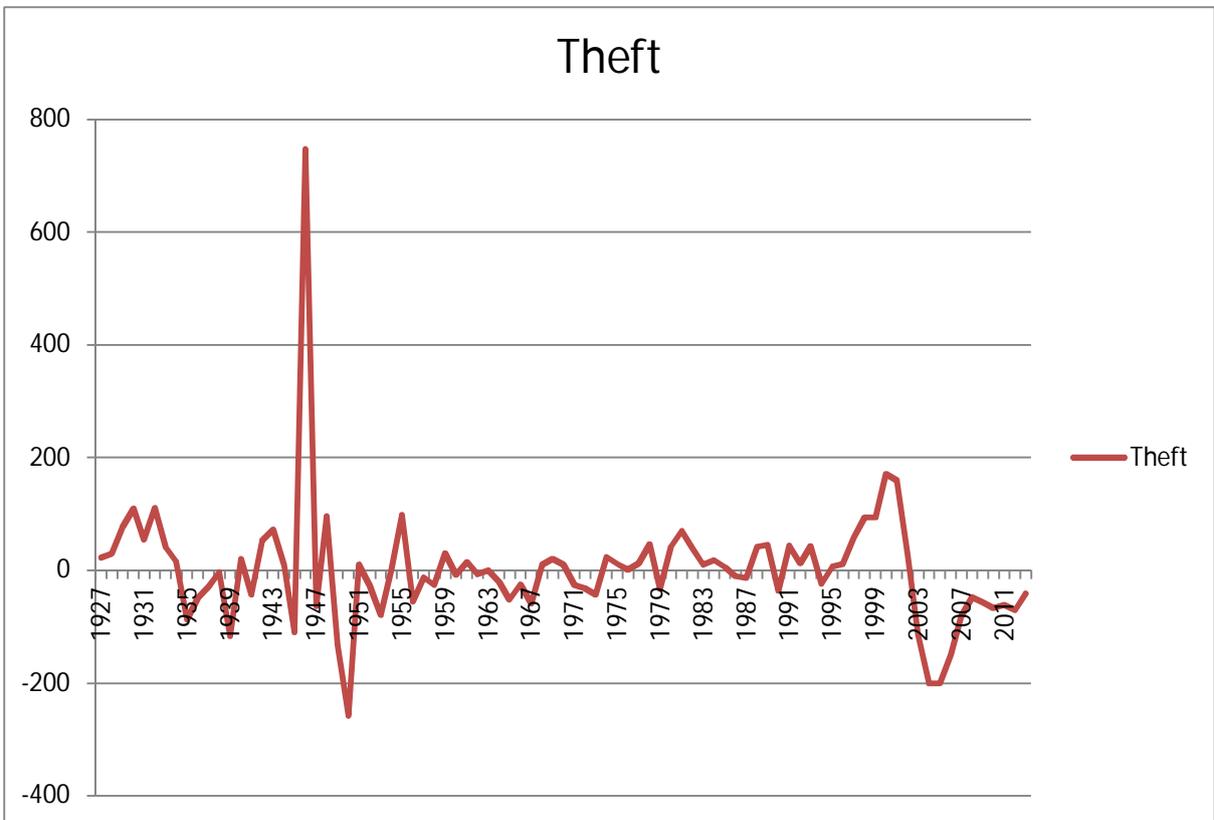
Fig. 1 GDP rate and crime rates



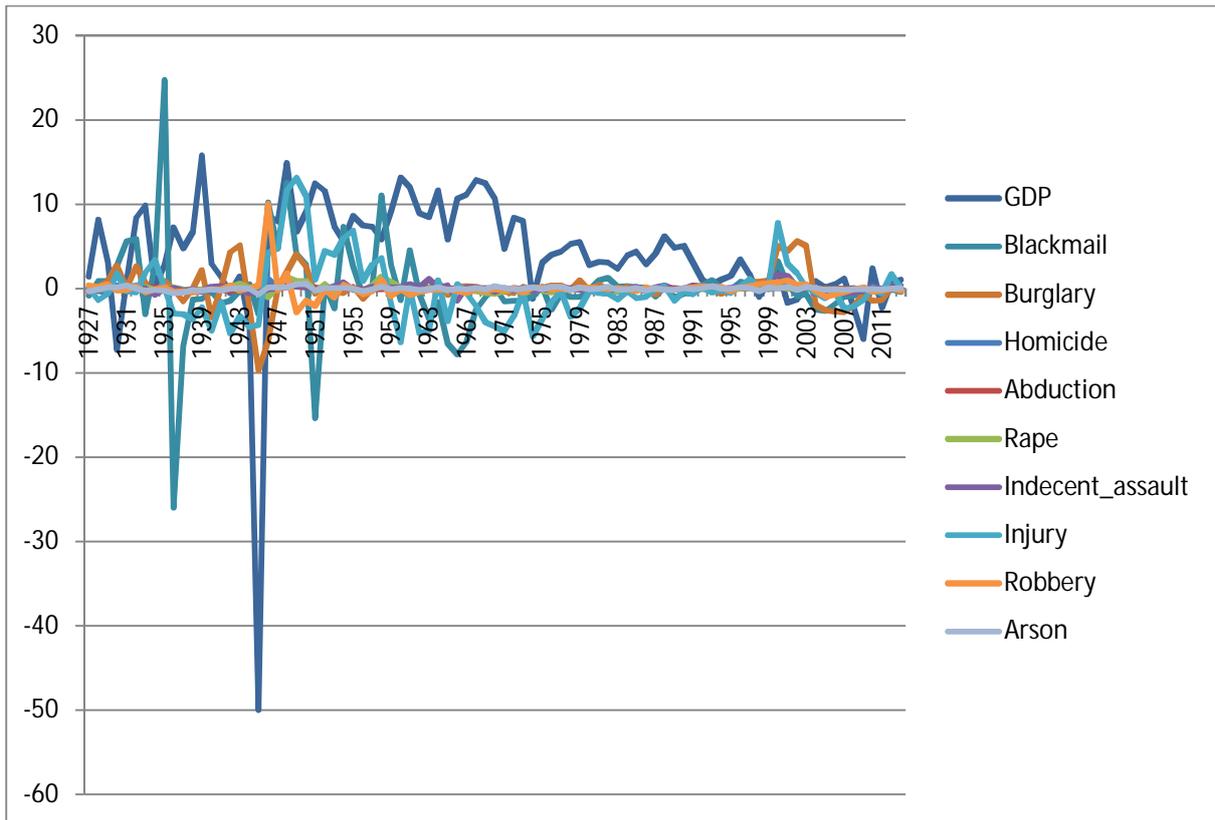
(a) GDP rate and change of crime rates



(b) GDP rate and change of crime rates (extraordinary theft rate left out)



(c) Change of theft rate



(d) GDP rate and change of crime rates (rates of theft, fraud, embezzlement and injury left out)

**Fig. 2** GDP rate and change of crime rates

### 3.3 Description of the experiments

The experiments are divided into two steps. The first step uses original values, i.e. GDP growth rate and crime rates. The second step uses GDP growth rate, and annual change of crime rates, e.g. value corresponding to 1927 equals to original value of 1927 minus original value of 1926. So there are only 87 years in the second step.

For example,

(1) Original values,  $GDP_{1927} = GDP \text{ annual growth rate of } 1927$ ,  $crime \text{ rates } 1927 = crime \text{ rates } 1927$

(2) Annual change,  $GDP_{1927} = GDP \text{ annual growth rate of } 1927$ ,  $change \text{ of crime rates } 1927 = crime \text{ rates } 1927 - crime \text{ rates } 1926$

The purpose of current study was to look at relationship between GDP growth rate and crime rates, change of crime rates over a period of 88 or 87 years. It is based on historical statistics of Japan, composed of 88 or 87 rows and 13 columns.

Although the SOM can process a dataset with missing data, as will be noted in certain stages of this study, missing values were imputed by mean of each attribute since most other methods require complete data. The total of missing values was 1.22% or 1.06% as to all data values when the size of the data matrix applied to all calculations was  $88 \times 13 = 1144$  or  $87 \times 13 = 1131$  elements. Besides missing values, descriptions presented in Table 3 are mean, standard deviation, minimum and maximum of each attribute.

**Table 3** Descriptions of the data used

(1) Data for the first step (GDP growth rate and original crime rates)

Attribute	Mean	Std. Deviation	Minimum	Maximum	Number of Missing Values (and %)
GDP	3.97	7.51	-50.00	15.75	0 (0.00%)
Theft	1089	252	600	1865	0 (0.00%)
Fraud	118.6	118.6	27.1	576.4	0 (0.00%)
Embezzlement	77.0	91.4	7.6	413.7	0 (0.00%)
Blackmail	16.60	12.04	2.85	48.05	0 (0.00%)
Burglary	12.37	5.17	3.38	31.60	0 (0.00%)
Homicide	2.066	1.008	0.740	4.140	0 (0.00%)
Abduction	0.481	0.611	0.030	2.460	0 (0.00%)
Rape	2.835	1.856	0.810	7.060	7 (7.95%)
Indecent_assault	3.019	1.783	0.290	7.850	7 (7.95%)
Injury	34.47	19.13	6.23	80.62	0 (0.00%)
Robbery	3.65	2.49	1.29	13.57	0 (0.00%)
Arson	1.776	0.753	0.770	3.990	0 (0.00%)

(2) Data for the second step (GDP growth rate and change of original crime rates)

Attribute	Mean	Std. Deviation	Minimum	Maximum	Number of Missing Values (and %)
GDP	3.97	7.51	-50.00	15.75	0 (0.00%)
Theft	2.0	108.6	-257.6	747.0	0 (0.00%)
Fraud	-2.2	40.4	-193.4	194.7	0 (0.00%)
Embezzlement	-1.28	19.24	-60.55	92.52	0 (0.00%)
Blackmail	-0.05	5.32	-26.02	24.69	0 (0.00%)
Burglary	0.087	2.137	-9.730	5.610	0 (0.00%)
Homicide	-0.039	0.209	-0.500	1.090	0 (0.00%)
Abduction	-0.0266	0.1061	-0.4500	0.2600	0 (0.00%)
Rape	-0.013	0.385	-1.020	1.400	6 (6.90%)
Indecent_assault	0.043	0.433	-1.450	1.620	6 (6.90%)
Injury	-0.18	3.62	-6.38	13.11	0 (0.00%)
Robbery	0.00	1.23	-2.83	10.00	0 (0.00%)
Arson	-0.0336	0.1988	-0.7100	0.5500	0 (0.00%)

#### 4.4 Evaluation of separation power of attributes in the dataset

After the dataset was established for processing, Viscovery SOMine was used for clustering. Upon initial clusters were identified, the structure of dataset was modified to be processed with ScatterCounter (Juhola and Siermala, 2012). The missing data values were replaced with medians computed from pertinent clusters so that the completed dataset could be processed by ScatterCounter. A main characteristic is that these years are labelled by cluster identifiers given by the preliminary SOM runs with the original 12 variables (attributes, as used in Viscovery SOMine).

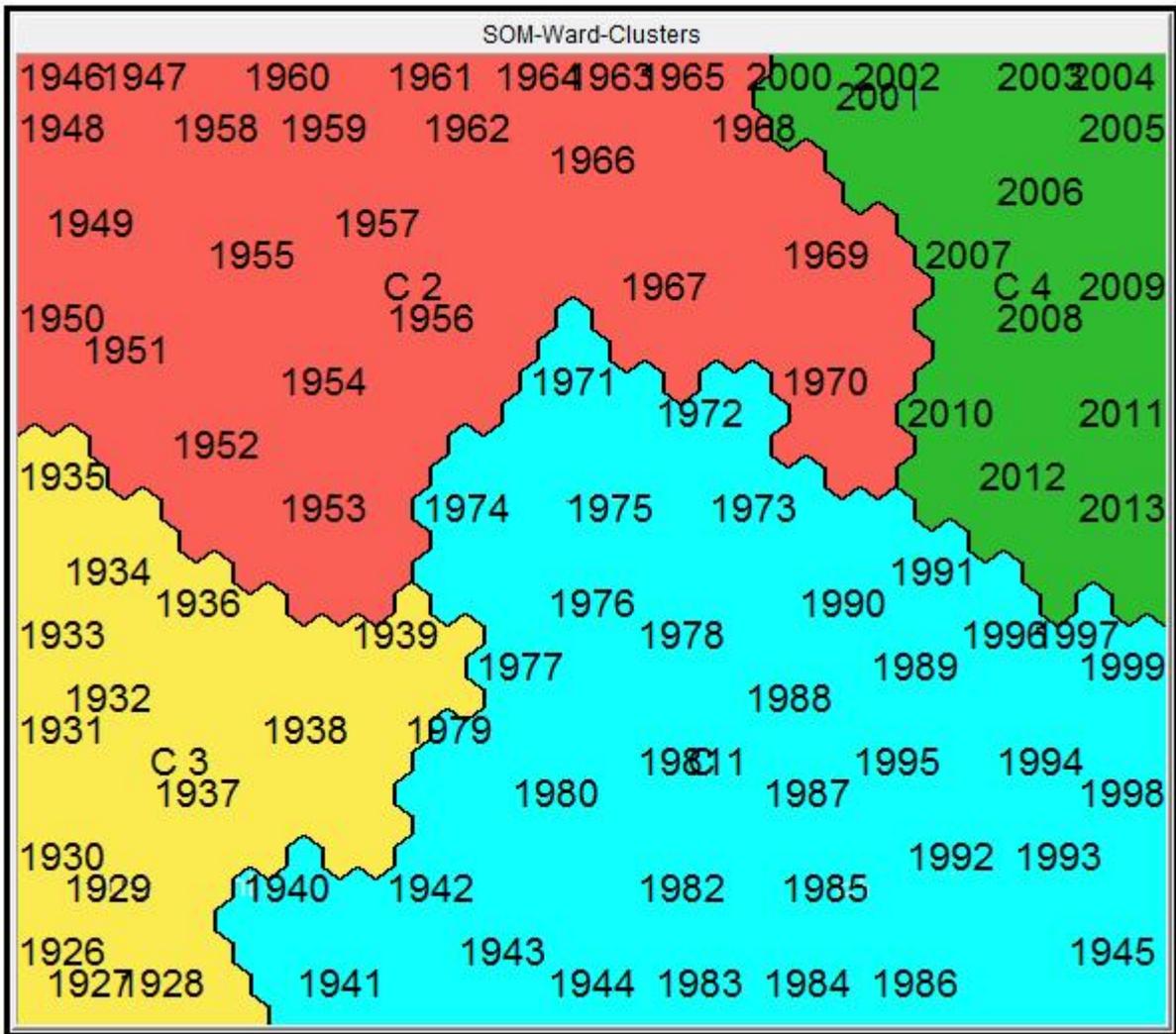
The objective of ScatterCounter is to evaluate how much subsets labelled as classes (clusters given by SOM) differ from each other in a dataset. Its principle is to start from a random instance of a dataset and to traverse all instances by searching for the nearest neighbour of the current instance, then to update the one found to be the current instance, and iterate the whole dataset this way. During searching process, every change from a class to some else class is counted. The more class changes, the more overlapped the classes of a dataset are.

To compute separation power, the number of changes between classes is divided by their maximum number and the result is subtracted from a value which was computed with random changes between classes but keeping the same sizes of classes as in an original dataset applied.

Since the process includes randomised steps, it is repeated from five to ten times to use an average for separation power.

Separation powers can be calculated for the whole data or separately for every class and for every attribute (Juhola and Siermala, 2012). Absolute values of separation powers are from  $[0, 1)$ . They are usually positive, but small negative values are also possible when an attribute does not separate virtually at all in some class. However, note that such an attribute may be useful for some other class. Thus, we typically need to find such attributes that are rather useless for all classes. Classes in our research are the clusters given by the SOM at the beginning before the current phase, attribute selection. With these results and observations, in this dataset, almost all have certain level (say, above 0.1) of positive separation powers and are kept in the dataset used in the following experiments and analysis. Unlike in some other experiments with different datasets where some attributes are due to be removed, this dataset reserves intact after evaluation of separation power.

(1) In the first step, i.e., data comprising of GDP growth rate and original crime rates, four clusters were generated, as shown in Fig. 3. The ScatterCounter gave the results in Table 4.



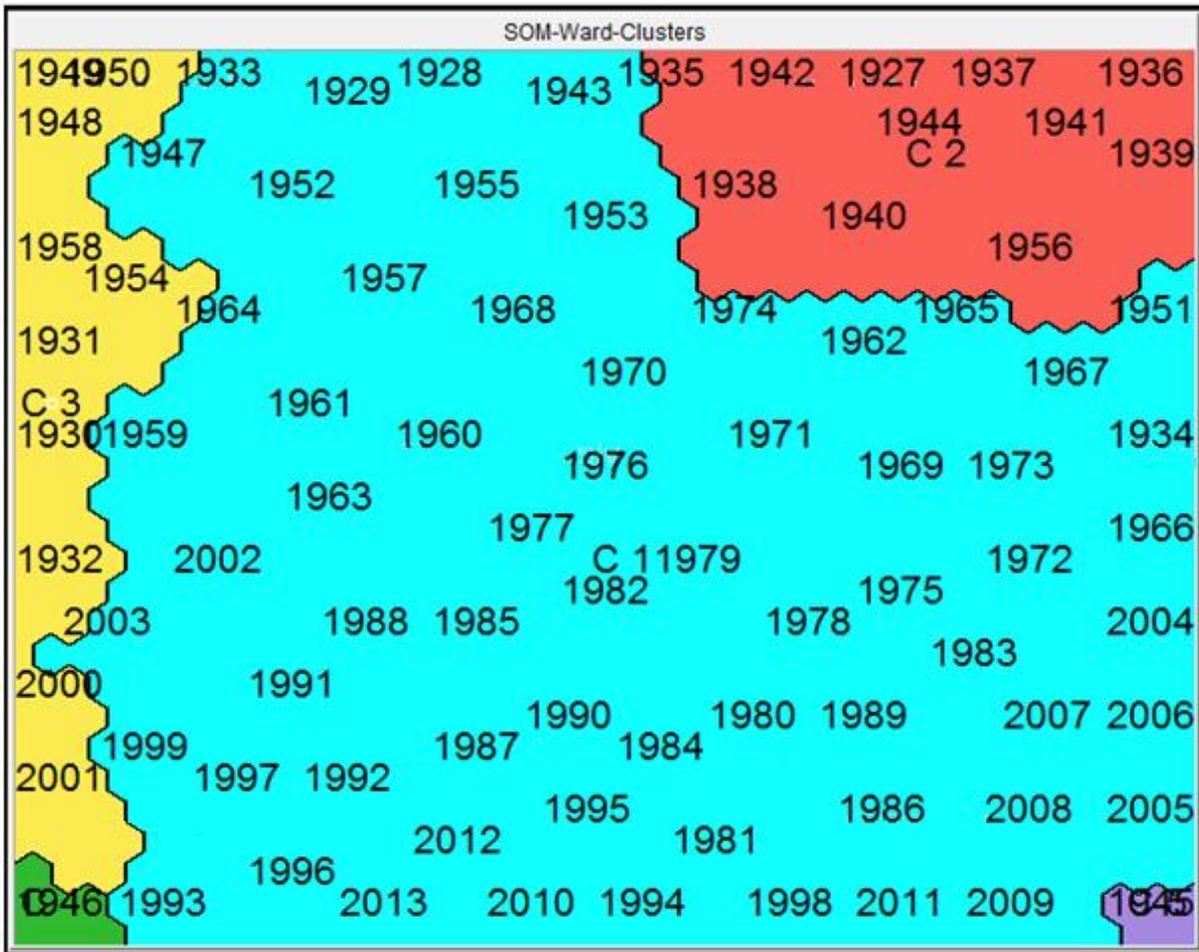
**Fig. 3** Four clusters given by SOM for GDP growth and original crime rates (not imputed, but the same for the impute data)

**Table 4** Separation power of attributes

Attribute	Cluster 1	Cluster 2	Cluster 3	Cluster 4
GDP	0.157	0.499	-0.035	0.107
Theft	0.057	0.059	0.214	0.035
Fraud	0.199	0.279	0.571	0.321
Embezzlement	0.042	-0.06	0.642	0.071
Blackmail	0.342	0.459	-0.071	0.285
Burglary	0.171	0.319	0.428	0.535
Homicide	0.385	0.439	0.5	0.464
Abduction	0.085	0.259	0.75	0.25
Rape	0.199	0.439	0.392	0.178
Indecent_assault	0.099	0.12	0.357	0.785
Injury	0.257	0.52	0.5	0.357
Robbery	0.457	0.259	0.142	0.107
Arson	0.099	0.099	0.535	0.107

According to the separation power of each attributes and overall of them in the four clusters, no attribute should be removed. So the further processing of the data will be the same as in this step.

(2) In the second step, i.e., data comprising of GDP growth rate and annual change of original crime rates, 5 clusters were generated, as shown in Fig. 4.



**Fig. 4** Five cluster given by SOM for data comprising of GDP growth rate and annual change of original crime rates

Among them, two clusters, cluster 4 and cluster 5 included only one year each, forming very small clusters. Therefore, such results with too small clusters for machine learning methods were left out in the further investigation.

#### 4.5 Construction of the map

In this study, the software package used is Viscovery SOMine 6. Compared with some other software packages of the SOM, Viscovery SOMine has almost the same requirements on the format of the dataset. At the same time, requiring less programming, it enables an easier and more operable data processing and visualization.

Missing values were marked with “NaN”. The SOMine software automatically generated maps from the dataset of 88 years and 13 attributes. The clustering map (Fig. 3) as well as some other detailed statistics, such as correlations as discussed below, can be used in further analysis.

## 5 Results

Upon processing of data, four clusters have been generated, each representing groups of years sharing similar characteristics. As a default practice in self-organizing maps, values are expressed in colors: warm colors denote high values, while cold colors denote low values.

### 5.1. Clusters

In order to give a full picture of these clusters, the following lists all the years in each cluster:

C1: 1940, 1941, 1942, 1943, 1944, 1945, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999

C2: 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970

C3: 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939

C4: 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013

Although the Viscovery SOMine software package provides the possibility for adjusting the number of clusters, usually automatically generated clusters represented the results that might occur the most naturally. In other experiments the same number of clusters could be set deliberately, years in these clusters were still re-grouped slightly one-way or the other. In this experiment, a more significant change of a cluster number was still tolerated, because this was expected to leave a new space where the similar issue could be speculated.

### 5.2 Validation of clusters

Total 88 years times 13 attributes with original around 1% missing values imputed with clusterwise medians, with new clusters (classes) given by the SOM clustering method.

After imputation, the results produced with the SOMs were compared to those given by several methods such as discriminant analysis,  $k$ -nearest neighbor classifier, Naïve Bayes classification, decision trees, support vector machines (SVMs) and random forests. For these the cluster labels found by the SOM were used as class labels in training and finally in tests to check whether the SOMs and classification results of the others agreed or disagreed. The tests were run on the basis of the leave-one-out principle. The classifications were programmed with Matlab.

**Table 5** Accuracy rates [%] when the imputed attribute values were scaled to interval [0,1] or standardized (for decision trees, parameter minparent means the minimum size of a node possibly to be divided into child nodes): discriminant analysis,  $k$ -nearest neighbor search, Naïve Bayes rule and decision trees

	Four clusters (original attribute values)		Three clusters (annual growth rates)	
	Scaled	Standardized	Scaled	Standardized
Discriminant analysis				
Linear	98.9	98.9	92.9	92.9
Logistic	98.9	98.9	89.4	89.4
$k$ -nearest neighbour searching				
$k=1$	97.7	100.0	87.1	87.1
$k=3$	97.7	97.7	84.7	80
$k=5$	96.6	95.5	80	80
$k=7$	96.6	98.9	80	78.8
Naïve Bayes with kernel density estimation	96.6	96.6	87.1	87.1
Naïve Bayes with Gaussian distribution	94.3	94.3	90.6	90.6
Decision trees				
minparent=8	89.8	89.8	88.2	89.4
minparent=6	89.8	89.8	88.2	89.4
minparent=3	95.5	95.5	88.2	88.2

Least-Squares Support Vector Machines (LSSVM) (Suykens and Vandewalle, 1999a, 1999b; Suykens et al., 2002) applied are a powerful machine learning method used for classification and regression problems. Origin of LSSVM lies in SVM research (Abe, 2010; Cortes and Vapnik, 1995; Vapnik 2000). However, there are several differences between SVM

and LSSVM. Firstly, inequality constraints in optimization have been changed to equalities. Secondly, LSSVM applies 2-norm cost function compared to 1-norm cost function introduced in the original SVM formulation. Thirdly, in LSSVM convex quadratic programming optimization is replaced with solving linear equation system. Several multi-class extensions have been proposed for SVM such as one-vs-one (Hsu and Lin, 2002), one-vs-all (Rifkin, 2004), DAGSVM (Platt et al., 2000) or tree-based solutions like in (Takahashi and Abe, 2002; Lei and Govindaraju, 2005) presented. Although the extensions in the references use traditional SVM approach, in all of them LSSVM can be used as well. We selected for our study a tree-based solution in which the basic idea is to separate one class in each layer of tree. This kind of idea was presented also in (Takahashi and Abe, 2002), but the difference to it is that we apply Scatter algorithm (Siermala and Juhola, 2006; Siermala et al., 2007; Juhola and Siermala, 2012) when finding the best separable class. The general idea of the method can be presented as follows

1. Assume that we have  $K$  classes in a dataset.
2. Search class having the highest separation power from the existing dataset using Scatter algorithm. Let that class be  $C_i$  and  $i$  in  $\{1, 2, \dots, K\}$ .
3. Construct a binary LSSVM classifier which separates class  $C_i$  from the remaining classes.
4. Exclude class  $C_i$  data from the dataset.
5. Repeat steps 2-4 until there are only two classes left in the dataset.

Following the given guidelines we construct a tree-based multi-class LSSVM architecture where one class is eliminated at each tree layer. Classifying new example begins at the root node and based on the classification result of LSSVM classifier we either get a predicted class label for the test example or move to the next layer in the tree. The best case scenario is that classification can end immediately in the root node and in the worst case scenario we need  $K-1$  comparisons before the predicted class label can be solved. In all cases the predicted class label is found from the leaf nodes of the tree construction. Figures 5-7 show the tree constructions used in this paper. Figures 5 and 6 show tree construction for the cases when dataset contained four clusters and dataset was standardized or normalized to  $[0,1]$  interval. Figure 7 instead shows tree construction when we examined difference dataset in which there are three clusters. The same tree construction given in Fig. 7 holds for both standardized and normalized ( $[0,1]$  interval) datasets. Since the tree constructions were built according to Scatter

algorithm results Tables 6 and 7 present the separation power values which defined the tree constructions.

An essential question when LSSVM is used for classification is the choice of kernel function. We selected for this study eight kernels to be used. These were the linear, polynomial kernels (degrees from 2 to 6), Radial Basis Function (RBF) and sigmoid. Furthermore, selected parameter values have significant impact on the LSSVM performance. Hence, we performed a thorough parameter value search. Let  $P=\{2^{-14}, 2^{-14}, 2^{-13}, \dots, 2^{13}, 2^{14}, 2^{15}\}$  and  $R=\{-2^{-14}, -2^{-14}, -2^{-13}, \dots, -2^{13}, -2^{14}, -2^{15}\}$ . For the linear and polynomial kernels there is only one parameter to be estimated (namely boxconstraint, i.e.  $C$ ) and for this parameter we tested all values  $C \in P$ . For RBF there are two parameters to be estimated (boxconstraint and the width of RBF function  $\sigma$ ). We chose that  $C$  and  $\sigma$  both have the same parameter values space and, hence, we performed grid search where we tested all  $(C, \sigma)$  combinations which are included to the Cartesian product  $P \times P$  (altogether 900 parameter value combinations). For the last kernel, sigmoid, the number of parameter estimated is three ( $C > 0$ ,  $\kappa > 0$  and  $\delta < 0$ ). We again performed grid search but now we tested all triplets  $(C, \kappa, \delta)$  which are included to the Cartesian triplet  $P \times P \times R$  (altogether 27000 parameter value combinations). Parameter value search was made using leave-one-out procedure and the selection criterion for parameter values was accuracy (trace of a confusion matrix divided by the sum of all elements in a confusion matrix). The same procedure was made for all datasets.

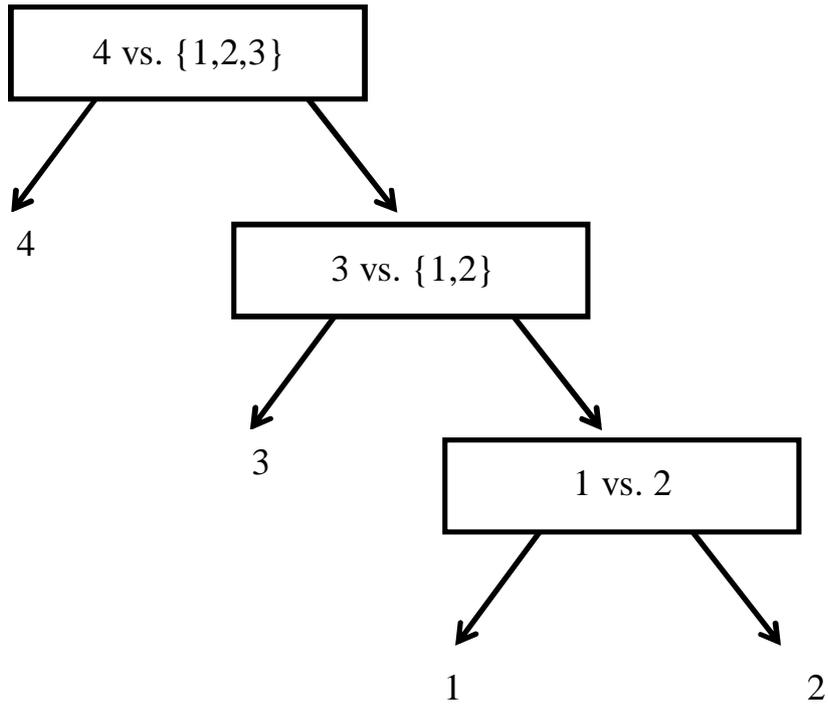
Random Forest (RF) is an ensemble learning method developed by Breiman (2001). RF has shown great performance in many applications and is widely used machine learning method. RF can be used both in classification and regression tasks. The basic idea behind RF is to extend the concept of decision tree learning. In RF several decision trees are collected together forming a forest. For each decision tree randomly selected feature subset is selected and, hence, RF uses the random subspace method in classification. Predicting a class label for the test example is made by giving the test example to all decision trees in a forest. Each one of the decision trees gives a predicted class label for the test example and the most frequent class is selected as a final predicted class label for the test example. An important parameter in RF is the number of trees. We varied the number of trees from 1 to 25 in the case of all datasets (datasets including three and four clusters). Classification was performed using leave-one-out procedure and accuracy was the performance measure likewise LSSVM classification.

**Table 6** Separation power values (the highest in Bold) for the classes which are used in building multi-class LSSVM tree constructions (dataset contains four clusters)

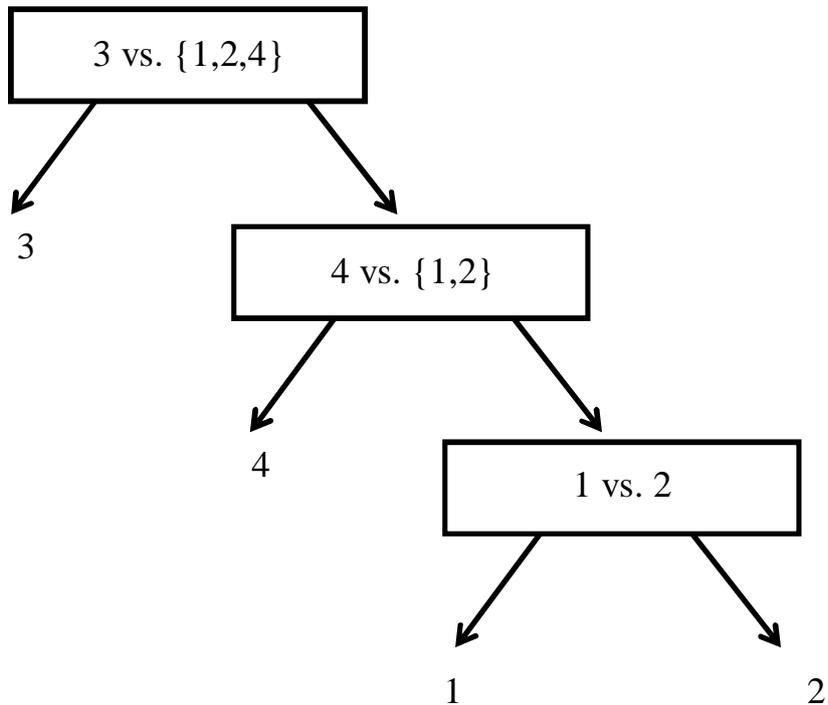
Dataset				Dataset			
Four clusters and dataset standardized				Four clusters and dataset normalized into [0,1] interval			
Separation power values		Separation power values		Separation power values		Separation power values	
First layer		Second layer		First layer		Second layer	
Class 1	0.515	Class 1	0.453	Class 1	0.501	Class 1	0.45
Class 2	0.634	Class 2	0.594	Class 2	0.66	Class 2	0.592
Class 3	0.743	Class 3	<b>0.7</b>	Class 3	<b>0.743</b>	Class 4	<b>0.721</b>
Class 4	<b>0.747</b>			Class 4	0.739		

**Table 7** Separation power values (the highest in Bold) for the classes which are used in building multi-class LSSVM tree constructions (dataset contains three clusters)

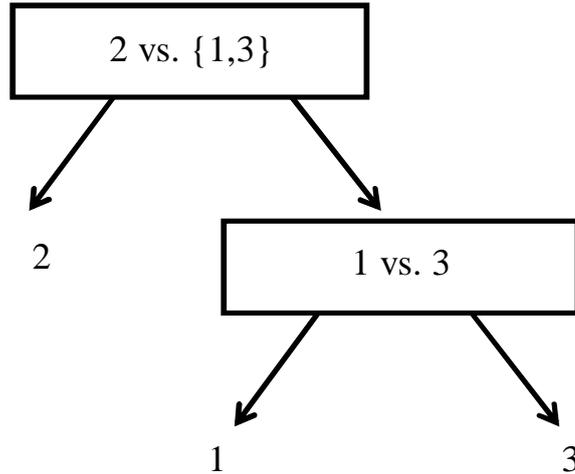
Dataset		Dataset	
Three clusters and dataset standardized		Three clusters and dataset normalized into [0,1] interval	
Separation power values		Separation power values	
First layer		First layer	
Class 1	0.434	Class 1	0.444
Class 2	<b>0.508</b>	Class 2	<b>0.55</b>
Class 3	0.455	Class 3	0.495



**Fig. 5** Tree construction for multi-class LSSVM built using Scatter algorithm. The dataset contains four clusters and is standardized.



**Fig. 6** Tree construction for multi-class LSSVM built using Scatter algorithm. The dataset contains four clusters and is normalized to [0,1] interval.



**Fig. 7** Tree construction for multi-class LSSVM which is built using Scatter algorithm. Dataset contains three clusters and the same construction holds for both standardized and normalized ([0,1] interval) dataset

**Table 8** Accuracy rates [%] when the imputed attribute values were scaled to interval [0,1] or standardized: support vector machines and random forests

	Four clusters (original attribute values) with parameter values in parentheses		Three clusters (annual growth rates) with parameter values in parentheses	
	Scaled	Standardized	Scaled	Standardized
SVM: kernel				
Linear	98.9 ( $2^{-1}$ )	98.9 ( $2^{-10}$ )	95.3 ( $2^{-6}$ )	96.5 ( $2^{-10}$ )
Polynomial degree 2	98.9 ( $2^{-3}$ )	98.9 ( $2^{-8}$ )	95.3 ( $2^{-11}$ )	91.8 ( $2^{-8}$ )
Polynomial degree 3	98.9 ( $2^{-4}$ )	96.6 ( $2^{-11}$ )	95.3 ( $2^{-2}$ )	88.2 ( $2^{-9}$ )
Polynomial degree 4	98.9 ( $2^{-6}$ )	95.5 ( $2^{-14}$ )	94.1 ( $2^{-6}$ )	85.9 ( $2^{-14}$ )
Polynomial degree 5	97.7 ( $2^{-7}$ )	94.3 ( $2^{-14}$ )	94.1 ( $2^{-7}$ )	83.5 ( $2^{-14}$ )
RBF	98.9 ( $2^{-1}, 1$ )	98.9 ( $2^{-14}, 2$ )	96.5 ( $2^{-1}$ )	97.6 ( $2^{-14}, 2^3$ )
Sigmoid	100.0 ( $2, 2^{-2}, -2^{-1}$ )	100.0 ( $2^{-6}, -2^4$ )	100.0 ( $2^{-14}, 2^{-4}, -2^{-14}$ )	100.0 ( $2^{-10}, 2^{-14}, -2^4$ )
Random forests: number of trees				
1	89.8	89.8	82.4	82.4
5	94.3	94.3	85.9	85.9

8	95.5	95.5	85.9	85.9
11	96.6	96.6	84.7	84.7
15	98.9	98.9	84.7	84.7
17	98.9	98.9	87.1	85.9
21	97.7	97.7	88.2	88.2
25	97.7	97.7	83.5	83.5

In Table 5 linear and logistic discriminant analysis produced the highest accuracies. For the situation of the original crime rates k-nearest neighbors were also very efficient. In Table 8 most results were even better than those in Table 5 and the best of all were the results generated by the SVMs with the sigmoid kernel. The accuracy of 100% is naturally exceptional and possible because of the small number of 88 years only in the present data, i.e., very slightly also because of random influence. Comparing the results between the original data (the first and second columns in Tables 5 and 8) and differences (the third and fourth columns), the former are almost always higher than the latter. This denotes that the latter formed a slightly more complicated classification task. In general, since there are very high accuracies greater than 90% and even close to 100%, these indicate that the two SOMs obtained present good mappings with high confidence for the present data.

### 5.3 Correlations

A detailed list of correlations was generated, based on which Table 7 was created. These correlations were computed from the original, not yet imputed dataset. Although even strong correlation between two attributes does not necessarily indicate causation, this will bring about materials for further analysis and reference. There are many opportunities that these results can be used to compare with previous studies on crime using other methods. We obtained 8 out of 24 comparisons ( $p < 0.05$ ) to be statistically significant.

**Table 9** Correlations between the GDP growth rate and the crime-related where the asterisk indicates the statistically significant ( $p < 0.05$ ) correlations. The  $p$ -values were adjusted for multiple testing with the Holm's method

Attribute	Original values	
	Non-imputed	Imputed
Theft	0.07	0.07
Fraud	0.13	0.13
Embezzlement	-0.04	-0.04
Blackmail	0.45 *	0.45 *
Burglary	-0.23	-0.23
Homicide	0.35 *	0.35 *
Abduction	0.05	0.05
Rape	0.44 *	0.44 *
Indecent assault	-0.12	-0.11
Injury	0.48 *	0.48 *
Robbery	0.28	0.28
Arson	0.13	0.13

From Table 9 one third of the correlation values were interesting, while others were very weak. Certainly, while currently such kind of research has been carried out in a small scale, extensive exploration is still necessary to conclude how socio-economic elements interconnect with criminal phenomena, either affecting their occurrence, or their increase or decrease.

## 6 Conclusions

This paper dealt with data from statistics at the national level for historical development of criminal phenomena in Japan, with reference to GDP growth rate. Conventionally, analysis in the study of crime did not handle large-scale multidimensional data due to technical or methodological limits. With the help of the self-organizing map, multidimensional comparison was realized. The research objects, in this paper, years, were grouped into different clusters with more convergent features.

By using discriminant analysis, *k*-nearest neighbor classifier, Naïve Bayes classification, decision trees, random forests and support vector machines (SVMs) to verify the SOM results, findings of the study gave additional proof that the self-organizing map was an interesting tool for assisting research on individual types of crime. The clustering results were easily visualized and convenient to interpret, facilitating practical comparison between different historical periods with GDP growth rate and criminal features. The article found that there were only weak correlations between GDP growth rate and crime rates. Nevertheless some of the correlations were still interesting. This was concluded according to data mining, in the process of which GDP growth rate and crime rates were placed in same years.

The relationship between crime and economic development has been considered complicated. Typically, stability of society can contribute to smaller quantity of crimes. Stability does not indicate wealth or poverty, but meaning swift or sluggish transformation. From the data set itself, we have already found that GDP growth could be accompanied by change of crime rate in an interesting way. For example, when velocity of Japanese economic development was prompt, crime rates increased as well; when velocity of economic development decelerated, crime rates decreased as well. In this sense, GDP growth rate could be a signal of societal stability.

It must be noted that one single economic indicator is definitely not in a position to represent the whole portrait of economy, predominantly, for example, changing weights of industries, such as primary, secondary and tertiary sectors. Economic situation can also be expressed in other indicators, as involved in many other studies including ours. It must also be noted that findings in this study cannot yet substantiate the whole image of relationship between crime and GDP growth rate, because such different social phenomena need not to be synchronous. GDP development is not unavoidably reflected concurrently in criminal phenomena. For instance, economic crisis, in which GDP growth rate would be plummeting, is likely to be translated into change of different crime rates after some months, or some years.

In addition, economic development can also be accompanied by change of crime rates in different patterns, for example, rates of some types of crimes ascending, while some others descending. Therefore, it is interesting in the future to probe the relationship between alteration of GDP growth rate and crime rates with a temporal lag, and segregate crime rates into different groups.

One of the limits of applying the SOM was found to be requirements for the well-framed data sets, in dealing with which high quality statistics were necessary and the acquiring and preparation for them might take same efforts as the activities of processing and analyzing

themselves. Another limit was that continuous official historical statistics, over centuries or millennia did not exist, and such kind of a situation was taken as granted in traditional research, which sought remedies in qualitative methods. Therefore, there has been usually a conflict of ideas between qualitative and quantitative methods when statistical data were involved. In this case, it can be expressed more as a conflict between pragmatic and technical approaches. This limit also revealed the fact that more future research in such a field, where it is more applicative in computer sciences, has more methodological sense in social sciences.

### **Acknowledgment**

The second author is thankful for the Finnish Cultural Foundation Pirkanmaa Regional Fund for the support.

### **Conflict of interest**

The authors declare that they have no conflict of interest.

### **References**

- Abe, S., 2010. Support vector machines for pattern classification. Springer-Verlag, London, UK, 2<sup>nd</sup> edition.
- Adderley, R., 2004. The use of data mining techniques in operational crime fighting. In: Proceedings of Symposium on Intelligence and Security Informatics, Tucson A.Z., ETATS-UNIS (10/06/2004) 3073 (2), 418-425.
- Adderley, R., and Musgrave, P., 2003. Modus operandi modelling of group offending: a data-mining case study. *International Journal of Police Science and Management* 5 (4), 265-276.
- Allwein, E. L., Schapire, R.E., and Singer, Y., 2000. Reducing multiclass to binary: A unifying approach for margin classifiers. *J Machine Learning Res* 1, 113-141.
- Axelsson, S., 2005. Understanding intrusion detection through visualization, PhD thesis, Chalmers University of Technology, Göteborg, Sweden.
- Breiman, L., 2001. Random forests. *Machine Learning* 45(1), 5-32.
- Brockett, P. L., Xia, X., Derrig, R. A., 1998. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. *J Risk and Insurance* 65 (2), 245-274.
- Burges, C. J. C., 1998. A tutorial on support vector machines for pattern recognition. *Data Mining Knowledge Discovery* 2 (2), 121-167.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learning* 20 (3), 273-297.

- Dietterich, T. G., and Bakiri, G., 1995. Solving multiclass learning problems via error-correcting output codes. *J Artif Intell Res* 2, 263-286.
- Escalera, S., Pujol, O., and Radeva, P., 2010. On the decoding process in ternary error-correcting output codes. *IEEE Trans Pattern Anal Machine Intell* 32(1), 120-134.
- Fei, B., Eloff, J., Olivier, M., and Venter, H. 2006. The use of self-organizing maps for anomalous behavior detection in a digital investigation, *Forensic Sci Int* 162 (1-3), 33-37.
- Fei, B., Eloff, J., Venter, H., and Olivier, M. 2005. Exploring data generated by computer forensic tools with self-organising maps. *Proceedings of the IFIP working group 11.9 on digital forensics*, 1-15.
- Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F., 2011. An overview of ensemble methods for binary in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn* 44 (8), 1761-1776.
- Grosser, H., Britos, P., García-Martínez, R., 2005. Detecting fraud in mobile telephony using neural networks. In: M. Ali, and F. Esposito (Eds.). *Lecture notes in artificial intelligence*, Springer-Verlag, Berlin, Germany, 3533, 613-615.
- Hollmén, J., 2000. User profiling and classification for fraud detection in mobile communications networks, PhD thesis, Helsinki University of Technology, Finland.
- Hollmén, J., Tresp, V., Simula, O., 1999. A self-organizing map for clustering probabilistic models. *Artif Neural Networks* 470, 946-951.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2013. A practical guide to support vector classification, Technical report. Retrieved June 11, 2016, from <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Hsu, C.-W., Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Networks* 13(2), 415-425.
- Juhola, M., Siermala, M. 2012. A scatter method for data and variable importance evaluation, *Integr Comp-Aided Eng* 19, 137-149.
- Kangas, L. J., 2001. Artificial neural network system for classification of offenders in murder and rape cases, The National Institute of Justice, Finland.
- Kangas, L. J., Terrones, K. M., Keppel, R. D., La Moria R. D., 1999. Computer-aided tracking and characterization of homicides and sexual assaults (CATCH). *Proc. SPIE* 3722, Applications and Science of Computational Intelligence II.
- Kohonen, T., 1979. *Self-organizing maps*. Springer-Verlag, New York, USA.
- Lampinen, T., Koivisto, H., Honkanen, T., 2005. Profiling network applications with fuzzy C-means and self-organizing maps. *Classification Clust Knowledge Discovery* 4, 15-27.

- Lei, H., Govindaraju, V., 2005. Half-Against-Half multi-class support vector machines, Proceedings of the 6<sup>th</sup> international workshop on multiple classifier systems, Lecture Notes Comp Sci, 3541, pp. 156-164.
- Leufven, C., 2006. Detecting SSH identity theft in HPC cluster environments using self-organizing maps, Master thesis, Linköping University, Sweden.
- Li, X., Juhola, M., 2013. Crime and its social context: analysis using the self-organizing map. In: Intelligence and security informatics conference (EISIC), 2013 European, 12-14 Aug. 2013, Uppsala, Sweden, IEEE, pp. 121 - 124.
- Li, X., 2014. Application of data mining methods in the study of crime based on international data sources, PhD thesis, University of Tampere, Tampere, Finland.
- Li, X., Juhola, M., 2014a. Country crime analysis using the self-organizing map, with special regard to demographic factors. *Artif Intell Society* 29(1), 53 - 68.
- Li, X., Juhola, M., 2014b. Application of the self-organising map to visualisation of and exploration into historical development of criminal phenomena of the USA, 1960-2007. *Int J Society Systems Sci* 6(2), 120 - 142.
- Li, X., Juhola, M., 2015. Country crime analysis using the self-organising map, with special regard to economic factors. *I J Data Mining, Modelling and Manag* 7(2), 130 - 153.
- Li, X., Joutsijoki, H., Laurikkala, J., Siemala, M., Juhola, M., 2015a. Crime vs. demographic factors revisited: application of data mining methods. *Webology* 12(1), Article 132. Retrieved June 11, 2016, from <http://www.webology.org/2015/v12n1/a132.pdf>
- Li, X., Joutsijoki, H., Laurikkala, J., Siemala, M., Juhola, M., 2015b. Homicide and its social context: analysis using the self-organizing map. *Applied Artif Intell* 29(4), 382 - 401.
- Mathworks Documentation Center, 2015. Retrieved June 11, 2016, from Available at <http://se.mathworks.com/help/>
- Memon, Q. A., Mehboob, S., 2006. Crime investigation and analysis using neural nets. In: Proceedings of international joint conference on neural networks, Washington, D.C., pp. 346-350.
- Platt, J.C., 1998. Sequential minimization optimization: A fast algorithm for support vector machines. Microsoft Res Techn Report MSR-TR-98-14.
- Platt, J.C., Christiani, N., Shawe-Taylor, J., 2000. Large margin DAGs for multiclass classification. *Adv Neural Inform Processing Systems* 12, 547-553.
- Rifkin, R., Klautau, A., 2004. In defense of one-vs-all classification. *J Machine Learning Res* 5, 101-141.

- Siermala, M., Juhola, M., Laurikkala, J., Iltanen, K., Kentala, E., Pyykkö, I., 2007. Evaluation and classification of otoneurological data with new data analysis methods based on machine learning. *Inf Sciences* 177(9), 1963-1976.
- Siermala, M., Juhola, M., 2006. Techniques for biased data distribution and variable classification with neural networks applied to otoneurological data. *Comp Meth Progr Biomed* 81(2), 128-136.
- South, S. J., Messner, S. F., 2000. Crime and demography: multi linkages, reciprocal relations. *Ann Rev Sociol* 26, 83-106.
- Suykens, J. A. K., van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. Least squares support vector machines. World Scientific, New Jersey, USA.
- Suykens, J. A. K., Vandewalle, J., 1999a. Least squares support vector machines, *Neural Processing Letters*, 9(3), 293-300.
- Suykens, J. A. K., Vandewalle, J., 1999b. Multiclass least squares support vector machines, *Proceedings of the International Joint Conferences on Neural Networks*, 2, pp. 900-903.
- Takahashi, F., Abe, S., 2002. Decision-tree-based multiclass support vector machines, *Proceedings of the 9<sup>th</sup> international conference on neural information processing*, 3, pp. 1418-1422.
- Vapnik, V. N., 2000. *The nature of statistical learning theory*. Springer-Verlag, New York, USA, 2<sup>nd</sup> edition.
- Viscovery Software GmbH. 2015. Viscovery SOMine, <http://www.viscovery.net/somine/>.
- Zaslavsky, V., Strizhak, A., 2006. Credit card fraud detection using self-organizing maps. *Inform Security* 18, 48-63.