XINGAN LI

Application of Data Mining Methods in the Study of Crime Based on International Data Sources

ACADEMIC DISSERTATION
To be presented, with the permission of the Board of the School of Information Sciences of the University of Tampere, for public discussion in the Auditorium Pinni B 1097, Kanslerinrinne 1, Tampere, on April 25th, 2014, at 12 o'clock.

UNIVERSITY OF TAMPERE

XINGAN LI

Application of Data Mining Methods in the Study of Crime Based on International Data Sources

Acta Universitatis Tamperensis 1923
Tampere University Press
Tampere 2014

ACADEMIC DISSERTATION
University of Tampere
School of Information Sciences
Finland

Cover design by
Mikko Reinikka

Distributor:
kirjamyynti@juvenes.fi
http://granum.uta.fi

Acta Universitatis Tamperensis 1923
ISBN 978-951-44-9418-5 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1407
ISBN 978-951-44-9419-2 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2014

Abstract

The objective of this dissertation is to apply data mining methods in the comparative study of crime based on international data sources. Crime control is fundamental to the welfare, stability and development of modern society. Crime occurs in a composite of surrounding variables, typically uncontrollable by government, society or citizens. These environmental variables can roughly be classified into demographic, economic as well as historical factors, playing visible or mostly invisible roles in shaping geographic distribution of criminal phenomena on the international level, affecting occurrence and features of particular offences within particular jurisdictions, and providing a foundation for clustering relevant countries where there are comparable interaction of these variables in certain internal mechanism.

To reveal this internal mechanism, performing crime analysis using data mining and visualization techniques proves to be an intimidating assignment. They have been shown to be functional in a variety of domains but have not been extensively studied for applications in the macroscopic study of crime. The purpose of this dissertation is to apply the data mining methods, centred on the Self-Organizing Map (SOM), in mapping, clustering and comparing criminal phenomena among countries, and in identifying correlations between crime and demographic, economic and other social factors through processing of large amounts of crime data around the world and over history. During this process, the study is aimed at revealing to what extent the SOM, with assistance of other data mining techniques, can be a qualified tool in the study of crime.

Studies included in this dissertation, covered different sets of data and adopted different methods. The data sets covered countries from 1 to 181, and variables from 22 to 68. In one study, data were about historical development of 48 successive years, while data in other studies can be thought as static. These data were all processed by the SOM, and these data in four of the studies underwent a process of selecting variables by using a method called ScatterCounter. In validating the final clusters, different methods were employed, selected among *k*-means clustering, discriminant

analysis, *k*-nearest neighbour classifier, Naive Bayes classification, support vector machines, Kruskal-Wallis test, and Wilcoxon-Mann-Whitney U test.

In conclusion, the SOM can be a satisfactory candidate for the macroscopic study of crime, through processing multidimensional data. Incorporating other methods to improve variable selection and classification validation, the results generated by the SOM can provide broad potential for criminological and sociological exploration into social phenomena. In research on multiple countries and multiple crimes, findings have been found partially coincident with conventional study. Roughly defined patterns of crime situation have been found in some countries with some traditionally similar socio-economic conditions. In different groups of countries, different factors may work in different ways. Long-term development patterns of countries affected occurrence of crime. In research on historical development of crime in one single country, the USA, successive years are clustered together in one way or the other with few exceptions. In research on one single type of crime, homicide, it is clear that a string of thinking concerning potential research on what socio-economic factors cause homicide, affecting its occurrence, or its increase or decrease. Consequently, the applicative layer of data mining methods in information sciences has influential prospect in the methodological layer in other disciplines.

Keywords: machine learning, data mining, clustering, Self-Organizing Map, macroscopic study of crime

Acknowledgements

The willingness of supervising one who sought studies for a second doctorate interested and disinterested many professors. But finally, the completion of this dissertation was made spiritually and intellectually possible by Professor Martti Juhola, Ph. D., my supervisor, whose mentoring, encouragement, and help have been the important sources of my inspiration and efforts. I am grateful to him for all the supports that always make me feel that it is natural to transfer my thinking from one discipline (law) to another (information sciences).

I thank all those who co-operated in my research, especially in Publication V, Jorma Laurikkala, Ph.D., Henry Joutsijoki, Ph.D., and Markku Siermala, Ph.D. Together with Professor Juhola, their work forms an integral basis of such a subject matter. I would like to extend my appreciations to Kati Iltanen, Ph. D, Kirsi Varpa, M. Sc., Jyri Saarikoski, M. Sc., and all other members of the Data Analysis Research Group (DARG), in which I had the privilege to work and to dialogue with great pleasure.

I would also like to thank Professor Natacha Gueorguieva, Ph.D., College of Staten Island, the USA, and Adjunct Professor Tapio Grönfors, Ph.D., University of Eastern Finland, for their time and efforts acting as my pre-examiners, and Professor Timo Honkela, Ph.D., University of Helsinki, for acting as my opponent. In Finnish academia, their tasks have been theoretically defined and respected as indispensable elements of a doctoral dissertation.

The School of Information Sciences, University of Tampere provided a sufficient organizational environment for my research. I would like to express my appreciations to all the members of the administrative, technical, and academic team.

Financial supports from School of Information Sciences, University of Tampere, and from Tampere Doctoral Programme in Information Science and Engineering (TISE) are gratefully appreciated. I thank Antti Niinistö, Ph.D., the Coordinator, and Professor Markku Renfors, Ph.D., the Head, of TISE for their

organizational, and administrative efforts, which, to some extent, put my work into a fast track.

The occurrence of crime depicts a dark world, into which the academic exploration can also gloom the process of research. For years, I continuously got rid of the darkness with many other people's interests in my research on crime from different disciplines. I thank all those people for inspiring me of rationalizing the process. I thank again Emeritus Professor Ahti Laitinen, LL.D., University of Turku, for supervising my first dissertation in sociology of law and criminology, and for hosting my postdoctoral research in that field.

Taking this opportunity, I would like also to appreciate Professor Zhang Zaibo, who supervised my bachelor's thesis on human rights at Inner Mongolia University in 1989 when that topic was politically sensitive and risky. I would like to appreciate Professor He Bingsong, who supervised my master's thesis on computer-related crime at China University of Political Science and Law. I would like to appreciate Professor Uchida Hirofumi, LL.D., former Director of College of Law, Kyushu University, Japan, who hosted me as a visiting scholar co-operatively arranged by China Ministry of Education and Japan Monbukagakusho (Ministry of Education, Culture, Sports, Science and Technology) in 2000-2001, during which I did research on Japanese economic crime, a part of which was computer-related crime. Each of them, in fact, represented a group of experts and professors in their academic field. I extend my appreciation to all those experts with them.

I am also grateful to teachers of University of Turku and those of Åbo Akademi where I took many courses in information systems, using my "flexible study rights" when I was in the process of waiting for pre-examination and for defence of my first doctoral dissertation. As a result of the wait and the use of these flexible study rights, I am able to have access to the palace of information sciences today.

I would like to thank my parents and parents-in-law, brothers and sisters, my wife Helen, and my daughters Peilin and Peiyun for their long-term support for my prolonged, tedious and cloistered research in different fields of sciences, the processes of which bring them little pleasure, but the results of which always pride them.

Xingan Li

Turku, March, 2014

Table of Contents

Abstract i

Acknowledgements iii

Chapter 1 Introduction 1

Chapter 2 Construction of the Study 7

2.1 Measurement of Criminal Phenomena 7

2.2 Environmental Variables of Crime 12

2.2.1 Demographic Factors 13

2.2.2 Socio-Economic Factors 16

2.2.3 Historical Development 19

2.3 The data sets 20

Chapter 3 Methods Applied in the Study 23

3.1 Attribute Selection 23

3.2 Clustering 24

3.3 Classification Validation 31

3.4 Correlation 33

3.5 Summary of data processing 33

Chapter 4 Results 35

4.1 Publication I Crime and social context, a general examination 35

4.2 Publication II Crime and demographic factors 37

4.3 Publication III Crime and Socio-economic factors 39

4.4 Publication IV Crime and historical development, the case of the United States 40

4.5 Publication V Crime and social context, the case of homicide 42

Chapter 5 Conclusions 45

Chapter 6 Personal Contributions 49

Bibliography 51

Publication I 63
Publication II 69
Publication III 87
Publication IV 107
Publication V 127

Publications

I. X. Li and M. Juhola. Crime and its social context: analysis using the self-organizing map. In *Proceedings of European Intelligence & Security Informatics Conference (EISIC 2013)*, IEEE, pp. 121-124, 2013. DOI 10.1109/EISIC.2013.26.

II. X. Li and M. Juhola. Country crime analysis using the self-organizing map, with special regard to demographic factors. *Artificial Intelligence and Society*, 2013. DOI 10.1007/s00146-013-0441-7.

III. X. Li and M. Juhola. Country crime analysis using the self-organising map, with special regard to economic factors, *International Journal of Data Mining, Modelling and Management, Vol. X, No. Y, xxxx*2013. Accepted.

IV. X. Li and M. Juhola. Application of the self-organising map to visualisation of and exploration into historical development of criminal phenomena in the USA, 1960–2007, *International Journal of Society Systems Science, Vol. X, No. Y, xxxx,* 2013. Accepted.

V. X. Li, H. Joutsijoki, J. Laurikkala, M. Siermala and M. Juhola. Homicide and Its Social Context: Analysis Using the Self-Organizing Map, submitted to *Applied Artificial Intelligence* (AAI).

Chapter 1 Introduction

There has been long-lasting interaction between development of science and technology and the study of crime (for example, Li 2008). This dissertation locates itself at applying the Self-Organizing Map and other data mining methods in the study of crime based on international data sources, in order to map global distribution of criminal phenomena by referring to multiple variables, to identify correlations between crime and its environmental factors, and to verify the applicability of the SOM and other data mining methods in the study of crime. Due to the special nature of this study, this chapter discusses the necessity, value and potentiality of application of these methods.

In the current era when a large volume of crimes occur in society, prevention of crime has become one of the most imperative global issues, along with the great concern of strengthening public security. Crime has negatively influenced the societies of both developed and developing countries through threatening the quality of life, intimidating human rights and fundamental freedom, and causing a severe challenge to the society. No country has remained untouched, even though the intensity and seriousness of the problem might be different from country to country. Crime control is fundamental to the welfare of people, stability of countries, and development of societies all around the world.

The study of crime is expected not only to control present crime but also to analyse the criminal phenomena so that future occurrences of similar incidents can be overcome. Government and community officials are making a thoroughgoing effort to improve the effectiveness of prevention of crime. Abundant investigations addressing this problem have generally employed disciplines of behaviour science and statistics. Studies and research in criminal justice and criminology have long sought assistance from, and have always been promoted by implementation of discoveries, invention and innovations in many other disciplines, such as sociology, biology, psychology, chemistry, mathematics, statistics, physiology, medicine, genetics, and information technology, to name some. At this point, disciplinary borders become ambiguous, and

multidisciplinary approaches are prevailing. As previous researchers pointed out that, the study of crime has been marked by a diversity of theoretical perspectives, due to the fact that criminology emphasizes the interest of a number of different intellectual and professional fields (Wheeler 1962, p. 14; Quinney 1971, p. 228). The study of social problem has provided scientists with the scrupulous advantage of allowing them to examine the application of their understanding to the resolution of human problems.

In academic history, scientists and technicians have infrequently been exclusively devoting themselves to the study of crime. The 2000s witness an increased interest of computer scientists in humanities and social sciences (for example, Niemelä and Honkela 2009; Honkela 2010). On the other hand, legal and criminological tradition has been characterised by delving into science and technology to search for answers to questions of crime. Increased necessity and interests have been promoting the interaction between jurists, lawyers and criminologists on one side and scientists and technicians on the other. Criminologists desire new tools, new techniques, new methods, and new theories from natural science. Scholars from different backgrounds might develop a common academic career in the field of studying the issues of crime. The reason why the study of crime can be practiced from multidisciplinary points of view exists in the facts that crime occurs in compositional environmental variables, let alone its means, tools and detection involving a variety of techniques. To some extent, amount of offences and its relationship with these variables can be measured by different methods. The purpose of this study is to apply the self-organizing map (SOM), with other data mining techniques, to the study of crime, clustering criminal phenomena according to spatial and temporal criteria, investigating correlation between crime and demographic and economic factors, studying its historical development, and considering applicability of the SOM as a useful computational method in the study of crime.

The first consideration for doing such a study is to map global distribution of criminal phenomena by referring to multiple variables. An analysis based on the available information in judicial statistics, academic literature and media reports, results of the study revolved around whether the SOM can be a feasible tool for mapping criminal phenomena through processing of large amounts of crime data involving a number of variables. By comparing countries within each cluster and between all clusters, social patterns on which crime situation is based can be

described according to the maps generated. The question here is to what extent these maps drawn and these patterns identified can reflect crime reality of these countries.

The second consideration of this study is to answer questions such as what factors have positive correlation on crime, what factors have negative correlation on crime, and what kind of country profiles are expressed in terms of the level of crime situation. The research findings can play a role in shaping social policy for diminishing factors that lead to crime and increase factors that restrain crime. Or lest these factors cannot be basically diminished or increased, their side-products or side effects should be controlled by predefined measures. For example, urbanization may cause rates of certain offences to increase; but it is difficult to intervene the process of urbanization. During this process, transformative social structure can merely be regularized through such ways as to carrying out anti-crime campaign, improving social security, or enhancing supervision and monitor of public places, as practised in the UK, the USA and other countries.

Crime is primarily the outcome of multiple adverse internal and external causes and conditions, such as biological, psychological, physiological, social, economic, educational, ethnical, environmental, seasonal, political, cultural and family conditions, etc. Regardless of its complexity, to prevent crime it is vital to have an understanding of its roots (The Community Safety and Crime Prevention Council, 1996). The study of crime has been situated in stretched historical settings with plenteous studies making attempts to reveal causes of crime and seek solutions, from classical theories, positivist theories, critical theory, and feminist theory to post-modern theory. No exclusively practicable theory has hitherto been invented to provide clear-cut response for tackling crime, despite the fact that numerous theorists presented countless persuasive suggestions (Rock 1994). The study of crime is dealing with a social phenomenon that hardly has a faultless solution. Crime is such a phenomenon that no one can supply conditions that can definitely create a crime, but once a crime is committed there must be certain reasons that can be identified. This study is not to search for a new solution but to test a new method for identifying factors that are important in seeking potential solutions, either in a short-term or in a long-term foundation.

Cause of crime is a theme of continuing significance and concern to various parties. Many perspectives have been offered in the academic literature in the study of crime to identify positive or negative correlation factors of crime, either demographic

factors or economic factors; compare geographical distribution of crime in different countries, and recognize (including but not limited to predicting) crime tendencies. For example, criminologists have long wanted to reveal causative and correlation factors of crime with abundant hypotheses, observations, comparison and conclusion, if not in vain. Law enforcement has the motivation for recognizing developmental tendencies of crime, such as its characteristic change in either microscopic or macroscopic aspects. Legislators have acquired the power from people to make effective law to eliminate, prevent or reduce crime. Governments are in need of making feasible policy to combat crime and assist victims. Victims make up their mind to get rid of criminal effects and affects. The general public are curious of creating, enjoying, and maintaining a society free from crime. The international society is committed to coordinating and cooperating in reckoning with crimes crossing borders. All these tasks are to be realized through various activities, in which the study of crime occupies a significant position. Processing crime data has been a basis for knowledge-detection and decision-making in this field.

The third consideration of this study is to extend the emerging interest of information systems scientists in application of computational methods in the study of crime and their empowered contribution to this field in general (recent examples are Mena 2003; Ollikainen and Juhola 2008), verify applicability of the SOM in the study of crime in particular in the above-mentioned fields. Since the SOM algorithm was introduced, during more than two early decades, very few publications were recorded dealing with crime-related topics (see Kaski, Kangas and Kohonen 1998; Oja, Kaski and Kohonen 2003; Pöllä, Honkela and Kohonen 2009). Only from the mid-2000s, we can uncover that the SOM has been fairly widely used in crime detection - a field relevant to this current topic, but not exactly the same. The application of the SOM to the study of crime in the sense similar to the present research has so far not been found. While the application of the SOM in crime detection has acquired much confirmation from previous research, its applicability in the crime research, which is focused on mapping distribution of criminal phenomena, identifying correlations between crime and environmental variables, and depicting historical process of crime trends, is to be examined in this study. In addition, the results generated by the SOM are validated by other classification methods (see Chapter 3 Section 3.3 for a brief explanation), both supervised and unsupervised. The mechanism of the validation is to take classification results of these methods to compare with the results of the SOM.

The comparison is expressed in the degree to what extent the similarity is, that is to say, what percentage of the SOM results can be reflected in the results of each of these validation methods separately.

In sum, this dissertation deals primarily with the usefulness and applicability of the SOM in the study of crime: mapping distribution, identifying correlation, and depicting trends of crime. If the study of crime can be taken as a grand multidisciplinary architecture, this dissertation constitutes a piece of the material, used to improve the efficiency in the process of the construction.

It is noteworthy that upon identifying correlations between crime and demographic or economic factors cannot simply indicate that measures such as change one factor or the other will change crime situation. The government will not interfere with civil society in a way to change its routine development track. The government will not intrude much into citizens' personal life. Rather, the government can only adjust its policy in some macroscopic aspects so as to harvest some long-term effects. Studies as the present dissertation is doing are only a start point for providing evidences for policymakers. Findings in some disciplines (if not all) of natural sciences can usually be directly applied. Presently, the application of new theories, new inventions and new products are increasingly accelerated. Comparatively, social problems are due to abundant possibilities that there are simply no precise answers to present. A well-administered democratic society is undergoing a strict inertia caused by the mechanism majority-decision. If we put an equality sign between science and democracy, it would cause great problems in social and political lives. Therefore, even if the methods used in this study can produce some tendentious policy opinions, it is a remote target and beyond the scope of this dissertation to have these policies in hand.

Following Chapter 1, the introductory part of the dissertation first proceeds to Chapter 2, demonstrating the construction of the study, including attributes and their categorisation as well as the feature of the five datasets. Chapter 3 briefly introduces the methods applied in the study, covering clustering, attribute selection, and classification validation methods. Chapter 4 presents the results of the five studies. Chapter 5 summarises and extends discussions and conclusions drawn from the five studies. Chapter 6 summarizes personal contribution in the five publications.

Chapter 2 Construction of the Study

Conventional measurement of criminal phenomena divides data into two categories: one category covers data measuring scale of crime elements, directly reflecting levels of a part number of offences; and the other covers data measuring scale of anti-crime elements, indirectly reflecting level of crime and anti-crime efforts invested by the government or required by the public order. Chapter 2 demonstrates data in three aspects: demographic factors, economic factors and historical factors.

2.1 Measurement of Criminal Phenomena

Nowhere in the world is free from the victimization of crime, but the levels of crime in different geographic areas and societal communities are apparently different from each other. In order to establish comparison according to one factor or two, the level of criminal phenomena must be measured by more or less a common criterion. The level of criminal phenomena can be measured by many different rules, such as household surveys, hospital or insurance records, and compilations by police and similar law enforcement agencies. With reference to the reliability of data, official statistics are typically competitive in their access to first-hand figures, their institutional scales and capacity, and continuity of their operations.

Criminal statistics collected by numerous agencies for different purposes have traditionally served as principal forms of data for the study of crime. The employment of these statistics has been a source of considerable controversy among scholars. Much of the controversy has revolved around the issue of the collection of criminal statistics. But the assumption that official statistics can serve as indexes of the actual amount of crime has been accepted (Quinney 1971, pp. 229-230). Habitually, violent crimes and property crimes have been well recorded and therefore worldwide data or historical data in countries like the USA, are more available for this study. But if data for new types of offences and new factors of crime are available, they are taken into account with priority.

The impact of several changes in the world today on the level of crime has been both negative and positive. Particularly, the risks posed by advancements in demography and economy may severely affect the situation of crime if not handled carefully. Correlation factors of crime are extensively present in societies. Conventionally, scholars and law enforcement have made their attempts to find the root causes of crime. A well-accepted formulation is the good generates the good, while the bad causes the bad. Several negative individual and social elements will unquestionably lead to crime. Considering that there are no consistently accepted theories on root causes of crime and that artificial intelligence has not been designed to identify causations, the term "correlation factors" is used in this dissertation as an expression of these factors.

In some other studies, considered factors can cover, for example, the following aspects: economic factors such as lack of financial resources, lack of educational opportunities, lack of meaningful employment options, poor housing, lack of hope, and prejudice against persons living in poverty; social environment such as inequality, not sharing power, lack of support to families and neighbourhoods, real or perceived inaccessibility to services, lack of leadership in communities, low value placed on children and individual well-being, and the overexposure to television as a means of recreation; and family structure--dysfunctional family conditions, such as parental inadequacy, parental conflict, parental criminality, lack of communication (both in quality and quantity), lack of respect and responsibility, abuse and neglect of children, and family violence (The Community Safety and Crime Prevention Council 1996, pp. 2-3). South African Human Rights Commission suggested community factors, such as social disorganisation, low household income; sparse social networks, family disruption. Link between exposure to violence and development of anti-social tendencies (South African Human Rights Commission 2007, p. 18); family influences such as coercive and hostile parenting styles, poor supervision can lead to bad behaviour and turning to deviant peer groups, substance abuse, early parenthood, broken homes, neglect, breakdown of traditional values (South African Human Rights Commission 2007, p. 18); and individual factors such as anti-social behaviour, and hyperactivity (South African Human Rights Commission 2007, p. 18).

These are only a few examples of the typical conclusions drawn from relevant researches in recent years. It seems to the academic field that everything existing

among human being, in the society, and in the history can be studied in relation to the phenomena of crime.

In addition to such comprehensive reports, some specialised studies also revealed different correlation between crime and some environment factors. The following are some selected documented positive correlation: urbanization (Stucky 2005, p. 8); unemployment (Eide, Rubin, and Shepherd 2006, pp. 27-28; McGuire 2005); inequality (İmrohoroğlu, Merlo and Rupert 2000, pp. 1-25); expenditures on police and redistribution (İmrohoroğlu, Merlo and Rupert 2000, pp. 1-25); alcohol (Gyimah-Brempong, 2001); the volume of illegal immigration (Orrenius and Coronado, 2005); the volume of apprehensions of illegal immigrants (Orrenius and Coronado, 2005); high levels of GDP (gross domestic product) per capita and greater income inequality (Fajnzylber, Lederman, and Loayza, 1998), air pollution (Levinson 2002, p. 600), and apprehension and treatment (Levinson 2002, p. 1093).

Negative correlation relationships between crime and some factors have also been revealed, for example, economic theory implies a negative correlation between educational attainment and most types of crime (Lochner 2007). Other negative correlated factors include wage rates (Freeman 1996, Gould, Mustard and Weinberg 2000, Grogger 1998, Machin and Meghir 2000, and Viscusi 1986), time spent in school (Gottfredson 1985, Farrington 1986, Witte and Tauchen 1994), certainty of punishment (Chiricos and Waldo 1970, Gibbs 1968, Logan 1971, 1975, Tittle 1969, Tittle and Rowe 1974), intelligence (Hama 2002), percentage of whites and percentage of Asians (Hama 2002), rate of church membership (Stark, Doyle, and Kent 1980), religiosity (Butts, Stefano, Fricchione and Salamon 2003), and house prices (Brehon 2007).

However, it should be noted that correlation between crime and some factors may be conditional on different other factors. For example, Yang, Phillips and Howard (1974) found that in rural areas, the positive correlation between preventive efforts and crime rate, and negative correlation between poverty and crime rate were contrary to findings from some urban studies. This indicates that some factors can be positively or negatively correlated with crime in studies of differently-orientations. Each study should be prepared to generate different results, accepting of which will help to understand the diversification of criminal phenomena based on diversified socio-economic foundations.

These examples provide us with a starting point to identify what kinds of factors should be considered to be included in the study. In such studies, data of crime and related factors can be collected, using official publications as the primary source, covering studies by international organizations, national statistical and judicial agencies, and other official documents. It should be emphasized that, any studies can only research on correlation factors, to the extent that data of those variables are available, that is, accessible to the researchers.

Mapping horizontal geographical distribution and longitudinal historical development can be more easily realized than research on correlation factors. There are high possibilities for the new techniques to be more efficient substitutes for the traditional ones. In visualising high-dimensional data, the self-organizing map may be irreplaceable by previous tools.

This dissertation processes crime-related data in two aspects:

(1) Data measuring scale of crime elements, including several rates of crimes, such as assault, burglary, fraud, murder, rape, robbery, software piracy, and total crime per 100,000 people that were reported to the police. These are direct measurement of criminal phenomena. First six crimes are selected because they are usually regarded as the most serious offences in all countries. They together usually represent nation-level situation of crime. Software piracy is added to the analysis because it is relatively a new type of offence and its number is available.

It is noteworthy that, types of regulatory crime in each country often reach a few hundred. It is theoretically possible to have figures of every country to cover every type of crime, but it is practically improbable to have them ready. Obstacles for such figures include, for example, different criteria for what constitutes a crime (particularly a small or petit crime), too numerous items for statistical purpose, more dark figure in petit offences, ambiguous limit between some crimes, and so on. So, some collective crime-related categories were used, such as property crime rate, violent crime rate and total crime rate in Publications I-IV. In Publication V, only one crime-related variable, homicide, was taken as a case study.

Traditionally, to have a look at national crime level, never was a complete set of statistics used, but a panorama of a few serious types of crimes. Even within this small scope, offences unreported to the police have traditionally been neglected in statistical research methods. The study of dark figures is usually supplemented by crime victimization survey, which is not touched in this study.

In brief, this group of data can directly reflect a part level of a part number of offences. But typically, this is formal practice in the study of crime to represent overall level criminal phenomena.

(2) Data measuring scale of anti-crime elements, including convicted, jails, police, and prisoners per 100,000 people, and share of prison capacity filled. Compared with the other group of data, these are indirect measurement of criminal phenomena. Police and jails per 100,000 people reflect the scale of professional anti-crime forces, and total prison capacity available. The convicted and prisoners per 100,000 people reflect the level of crime measured by rate of persons whose responsibilities are established. Share of prison capacity filled reflects the relationship between state's anti-crime readiness and actual anti-crime demand.

These figures are indirect measurement of criminal phenomena, because they do not reflect actual numbers of crimes or criminals. A certain number of offences occurred, but a part of them were neither detected nor reported to the police (or reported but the police did not record them and did not count them). A part of offences were reported to the police and the police recorded them, but investigation led to no finding of evidence or the criminal, so that these offences would have been charged. A part of offences were charged, but trials at courts led to no conviction. A part of suspects were tried, convicted but they were not necessarily sent to jails. A small part of the convicted may have to be cured in a medical institution due to serious illness. Yet a smaller part of the convicted may be in a death row and then be executed. These possibilities rendered these figures indirect reflection of overall criminal phenomena.

To put it briefly, this group of data can indirectly reflect level of crime and anti-crime efforts. The study of crime cannot be done without using statistical results, yet statistics of crime are not always reliable. As Sutherland and Cressey noted that, "the general statistics of crime and criminals are probably the most unreliable and most difficult of all social statistics" (Sutherland and Cressey 1966, p. 27). However, it is a reasonable process to commence with data formally documented in them and then to think about studies that has made efforts to "go behind" the recorded numbers (Baldwin and Bottoms 1976, p. 4). During this process, a variety of instruments are used to support investigation and exploration for differently oriented studies.

2.2 Environmental Variables of Crime

Crime, either as an individual incident or as a phenomenon, does not occur or exist by itself. Rather, crime occurs or exists in a composition of environmental variables. We cannot establish a theory such as determinism to conclude what exacts factors generate or produce what offences. Precise causation as those in natural science seems irrelevant here in identifying causes and effects relationship between an offence and certain factors. However, crime must be related to something else. The study of crime has not only targeted at crime alone, but also has explored into its surrounding variables.

Unlike natural sciences in which a condition causes some consequences, the consequences may consistently follow the condition. This can scarcely be correct in the study of crime, where the conditions, which can be regarded as causing criminal conduct merely, increase the possibility (risks). Thus crime control is more like risk control but not control of chemical reaction.

To ask what causes crime can have diversified meanings. From the extent of involved people, it can mean the individual level, group level (for example, juvenile delinquency), community level, regional level, national level (for example, crime in the USA), and international level. From the extent of relevant crime, it can mean the individual case of offence, a type of offence (for example robbery), a category (for example, property crime), and criminal phenomena as a whole. Under such circumstances, a single theory explaining causation of risk condition and crime cannot be proved true. Many theorists posed their theories based on their own viewpoints and claimed particular strings of causation. As Becker's economic analysis of crime supposed that the number of crime might be "determined" by both potential perpetrator's internal factors and external socio-economic factors (Becker 1968).

Crime rates differ enormously from each other among countries, and their difference in this aspect is orders of magnitude greater than their difference through time in a given country (Soares 2004, p. 155). The possible explanations for these cross-country differences are diverse, ranging from distinct definitions of crimes and different reporting rates (percentage of the total number of crimes actually reported to the police), to actual difference in the incidence of crimes, due to different culture, religion, and level of economic development or natural conditions (Soares 2004, p. 156). Political system and criminal policy may also have close relationship with crime

levels. Autocracy and "strike hard" are usually effective in maintaining low crime levels, of course sometimes in sacrifice of democracy and human rights.

Environmental variables of crime cannot be straightforwardly measured by a certain number or a certain category. Strict measurement or categorization is supposedly impossible. People usually take into account a pair or a dozen of factors in their research. That way, research is easy to formulate. This dissertation has to deal with more factors, some of which may have closer relationship with each other than with others. In order to make it more convenient and operable to process data and make analysis, the dissertation categorises surrounding variables of crime into demographic, economic as well as time-series factors.

2.2.1 Demographic Factors

Demographic factors, including static, dynamic, and structural factors, have been studied since the eighteenth century (South and Messner 2000, p. 83; for a recent research, see Juhola and Juhola 1996). Demographic factors play an important role in understanding variation in crime rates across time and place. Demographic features of the population effect crime rates in two distinct ways. First, characteristics of population structure have compositional effects: crime rates are higher when demographic groups that have greater levels of involvement in crime constitute a larger share of the population. Second, aspects of population structure may have contextual effects on crime when they exert causal influences on criminal motivations and opportunities for crime independent of individual level for criminal tendencies. Demographic factors have been considered in relation to crime for centuries. However, the use of demographic variable, as determinant of the aggregate level of crime, is still little explored in the crime literature. The discussion of some elements such correlation between race, sex and criminal phenomena has been challenged. This dissertation includes fifteen demographic factors, which can be roughly divided into three categories: population structure, population quality, and population dynamics.

(1) Regarding population structure, three rates are selected, including population older than 64, unemployment rate, and urban population. Crime has been believed a youth's cause. People under the age of 64 committed absolute majority of crimes. With the increase of age, crime decreases.

A consensus has been reached that unemployment causes crime, though the explanations on the reason why unemployment causes crime differs one theory from others. Many studies show a strong relationship between unemployment and crime and giving explanation based on the debilitating effects of powerlessness, alienation, absence of stake in conformity, lower class pathology, culture poverty, relative deprivation, wasted human capital, the negative effects of labelling, bad schools, blocked legitimate opportunities, illegitimate opportunity structures in areas with high unemployment (Braithwaite, Chapman, and Kapuscinski 1992). However, high unemployment rate will reduce some offences such as burglary. Increased unemployed population means decreased vacant houses during routine work time, improved home deterrence, and enhanced neighbourhood supervision, etc. In worse economic countries, people also have decreased presence in public places such as supermarkets, bars, transportation centres, and entertainment places. In families with unemployed members, potential monetary losses in crime will also be decreased.

City living has characterized some areas for centuries, but has spread with such acceleration over the past century as to encompass hundreds of millions of people (Clinard 1958, p. 54). For centuries writers have been concerned about the debauchery and moral conditions of the cities and have generally praised rural life. Delinquency and crime rates today are generally much higher in urban areas than in rural (Clinard 1958, p. 68). Urbanism with its mobility, impersonality, individualism, materialism, norm and role conflict, and rapid social change, appears to increase the incidence of deviant behaviour (Clinard 1958, p. 89). Crime is largely an urban phenomenon (Bottoms 1976, p. 1). Statistics from many countries, and in many periods of time, indicate that urban areas have higher crime rates than rural areas (Cressey 1964, p. 61). Wirth (1938) in his classic article on urbanization took the three concepts of size, density and heterogeneity as key features from which one could analyse social action and organization in cities. The rates for certain forms of deviant behaviour generally increase with the size of the city (p. 90).

Urbanization and labour mobility leads to increased numbers of strangers. Traditional intimate relationship between neighbourhoods has been superseded. It is so that criminologists found that less severe the bodily harm inflicted on the victim, the greater the likelihood of the crime being committed by a stranger (Thio 1978, p. 99). In other words, there are possibly more numbers of crimes in most urbanized

countries, but these crimes are possibly less severe; while in less urbanized countries, there is less number of crimes, but these crimes are possibly more severe.

McGuire (2005) pointed out that, "Emile Durkheim and other functional theorists" hypothesize that densely populated areas and individuals there undergo a collapse or transformation in social order as a result of great density, thus there occur more conflict between individuals and higher crime rates (p. 1). However, McGuire found that population density has a very small correlation to crime rate, contradictory to Durkheimian/functional theory (p. 11).

(2) Regarding population quality, taken into account are such factors as adult illiteracy, health expenditure per capita, infant mortality rate, life expectancy, population growth rate, population undernourished, and under-five mortality rate. These cover both typical physical and intellectual conditions of the population facilitating international comparison.

(3) Regarding population dynamics, factors such as birth rate, death rate, fertility rate, net migration, and population density are selected. These factors reflect a dynamic process of population change. Net increase of population increases population density and thus affects occurrence of crime and control of over it (Harries, 2006).

Crime of immigrants has been an attractive subject matter for centuries. The immense labour relocation, globalisation of labour markets, and growth of tourism pose severe questions regarding the validity or applicability of the national or moral foundation of laws and blur the dissimilarity between crime and rights, deviance, and cultural diversity (Sumner 2005, p. 8). In America, it has long been found that there is no definite race factor involved in crimes committed by immigrants or by their children. Immigrants overall are no more criminal than natives overall (Taft p. 118). On the contrary, some positive effects have been identified in previous studies: the coming of people with different cultures has kept American culture fluid. It has compelled people to rethink their mores. Studies on the immigration to the United States have proven that people of different cultures can live together and make joint contribution to human welfare (Taft, p. 119).

Immigrants, from the time they have begun to arrive in considerable number, have been blamed for all sorts of social ills, not least of which is crime. While some investigations showed a great predominance of crime and vice among immigrants, others drew the conclusion that there was no proof for the conjecture that immigration

brought about an increase in crime unbalanced to the increase in the adult population (Koenig 1962, p. 140). Criminologists have found that distant from bringing them criminalistic behaviour, most immigrants do not lack a respect for law and authority which they acquired in their home countries, they come principally from established, homogeneous societies which extend strict control over the activities of individuals (Koenig 1962, p. 142).

The association between immigration and crime, while remaining strong in the public perception for over a century, has by no means obtained steady empirical support (Wadsworth 2010, p. 532). The findings offer insights into the multifaceted relationship between immigration and crime and propose that increase in immigration may have been responsible for part of the steep crime plummet of the 1990s (*ibid*. p. 531).

These are only a limited number of examples frequently investigated by previous and contemporary researcher. Other factors that have been inquired also led to conflicting conclusions. This phenomenon is common in scientific research, but more remarkable in research related to factors in society. It shows the difficulty to draw consistent conclusions and to reveal the reality. Thus more efficient data mining methods are necessary in this field of research.

2.2.2 Socio-Economic Factors

Economics has been employing ever-changing concept about human beings and their activities. Adam Smith's economics implied that man is a rational animal who seeks material pleasure or utilities, in competition with his fellows, and this selfish, competitive search for personal gain was socially favourable, and should be left unimpeded by government. Later study has shown that Adam Smith's view misinterpreted the real nature of human motivation, and underestimated the social ills resulting from unregulated individualism (Taft 1950, p. 123).

Economic distress has long been considered as the basic cause of society's ills (Clinard 1958, p. 92). However, poverty is by no means the only factor accounting for the deviant behaviour (*ibid.* p. 98). General delinquency and criminal trends are not directly sensitive to the downward and upward movements of economic conditions (Mowrer 1942, pp. 190-191). Widespread absolute poverty does not necessarily lead

to crime, but relative poverty, that is, intensified difference of living standard between each other, motivates people with low level of living towards high level of living, and motivates people with high level of living towards yet higher and higher level of living, up to luxury living.

The importance of economic conditions as causes of crime grows largely out of the fact that materialism is approved in our culture (Taft 1950, p. 124). In such a culture, men have positive ambitions even when not suffering actual discomfort. People reduce the difference of level of living through raising their plane of living (*ibid.* p. 124). Like standard of living, other economic factors may also affect the level of crime one way or the other.

A revived version of economic analysis of crime insists that criminals respond to economic incentives in the same way that legal workers do (Becker, 1968). Economic theories of crime relate the likelihood that an individual engages in criminal activities to the costs and benefits of these activities, when compared to legal occupations. At the aggregate level, the more prevalent the conditions which make crime attractive, the higher the crime rates (Soares 2004, p. 157).

This dissertation contains seventeen economic factors, which are divided into four categories, including economic and consumption level, economic structure, development of new economic phenomena, and extent of research and development.

(1) Regarding economic and consumption level, factors such as electricity consumption per capita, electrification rate, GDP per capita annual growth rate, GDP per capita, and GDP per capita PPP (purchasing power parity) are covered.

The positive link between crime and development—usually cited in the criminology literature but regarded with suspicion by economists—does not exist. Reporting rates of crimes are strongly related to development, mainly income per capita. Therefore, the positive correlation between crime and development sometimes reported is entirely caused by the use of official records. Development is not criminogenic (Soares 2004, p. 156). In fact, this kind of recognition has been common in the academic field of the study of crime. At the same time, some correlated pair wises can have also been very common, such as the conclusion drawn by Soares: income inequality affects crime rates positively, while education and growth reduce crime (*ibid.* p. 156). Data processing in this study provides further insight into relationship between crime and selected economic factors.

(2) Regarding economic structure, factors such as employment in agriculture, employment in industry, employment in services, exports of goods and services, foreign direct investment net inflows, forest area, and imports of goods and services are included.

In the present world, some countries have a predominantly agricultural economy. Many other countries have been undergoing transformation from agricultural economy to industry and services. During last two centuries, industry developed rapidly and the global economy tended increasingly to be characterized by industrialization, specialization, and urbanization.

Traditionally, theorists from Durkheim-Modernization perspective insist that rapid social-economic change creates a social platform where those factors leading to deviant behaviour reside, such factors are industrialization, urbanization, the division of the labour, social disorganization, anomie, modern values, and cultural heterogeneity (Masahiro 2002, p. 497). However, there has never been unanimously accepted conclusion concerning the properties of correlations between crime and such economic factors. This situation renders present study possibility to reconsider the previous conclusions from a different point of view.

(3) Regarding new economic phenomena, considered are factors such as cellular subscribers per 100,000 people, Internet users per 100,000 people, and telephone mainlines per 100,000 people.

Contemporary scientific progression and marvellous advancement in communications have facilitated criminals of every part of the globe to perpetrate an offence by means of complicated apparatus in one location and afterwards run away to a different location. We are confronted with a historic process, starting with the invention of the computer in the 1940s, accelerating through a variety of forces, and causing profound changes in the life of people all around the world. It is a clearly distinguishable force, or rather a complex of intimately connected forces. The ubiquitous use of telephone, mobile phone and the Internet create new opportunities both for crime and anti-crime.

Sixty years ago, when discussing the influence of mass media on crime, Taft wrote that newspapers might teach the technique of crime, make crime seem common, make crime seem attractive and exciting to the boy, make crimes seem unduly profitable, give prestige to the criminal, attract sympathy or hero worship for criminals, appeal to “lower” impulses and by sensationalism, reflect crime-producing

elements in our culture, make escape from justice seem easy and by hindering the apprehension of criminals, fail to stress the punishment of crime, ridicule the machinery of justice, or through "trial by newspaper," and advocate types of treatment of criminals which tend to increase crime (Taft 1950, pp. 206-211).

Based on these considerations, numbers of telephone and network users are factors used in this research. Even though investigation in any one of them alone requires many resources, this study initiated a valuable attempt to consider them together with other socio-economic factors.

(4) Regarding research and development, two factors, namely RD (research and development) expenditure and numbers of researchers in RD per 100,000 people are involved in the analysis. These factors are considered to reflect a country's long-term development policy and strategy and have long-term influence on a country's social development, and thus are taken into account jointly with other factors.

2.2.3 Historical Development

Collecting historical statistical data proved to be a difficult task. In this dissertation, data on thirteen factors are collected, including average size of consumer unit (number of persons), civilian labour force percent of population, employed civilian labour percent of population, unemployed civilian labour percent of population, and total fertility rate.

The United States has for decades been perplexed by violence. However, it completely depends on what the reference groups are. The records of the highest homicide rate in recent history 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 Germany, Denmark, Norway, Japan, and Singapore, with homicide rate below 1 per 100,000 inhabitants. Compared with the figures of 50 in Sierra Leone, and more than 45 in El Salvador, Jamaica, Venezuela, Guatemala, and Honduras, it is also true that the U. S. also has a low level of homicide rate, 5.8 per 100,000 people.

In fact, criminal phenomena in the United States have a dramatic rise and fall in the latter part of the 20th century. According to the United States Department of

Justice, Bureau of Justice Statistics, the overall crime rate in the United States began its rise as early as in 1962, since when the 1961 low point has never been reached again. The 1970s and 1980s witnessed an interminable increase of crime rate. After it reached the 1991 peak, the crime rate began to turn down from 1992. In 2007, the United States' crime rate fell back to the 1974 level (Bureau of Justice Statistics, 2012). Both violent crime and property crime have the similar tendency. Whenever violent crime rises property crime rises at the same rhythm, and vice versa.

During the last quarter of the 20th century, the US was one of countries with higher crime rates within the economically developed world. However, after decades of exploration into the paradox of sharp rise of crime accompanying the rapid increase of economy, people began to enjoy sharp fall of crime while suffering from economic declination. According to Wadsworth, many studies (such as Blumstein and Wallman, 2000; Conklin, 2003; Zimring, 2007) have explored a variety of explanations for the sharp and continuous drop in crime. The most outstanding proposition has focused on the increased use of imprisonment, changes in the age distribution, changing drug markets, the availability of weapons, economic development, new security strategies, and the legalization of abortion (Wadsworth 2010, p. 533). However, as other research on social phenomena, no precise conclusion can be drawn nor confirmative reasons can be given to explain either the rise or the fall of crime. In Publication V, the topic concerning the rise and fall of American crime was examined using the self-organizing map.

2.3 The data sets

No known laboratory can imitate a country or a world. No known controllable experiment can imitate the process of social phenomena. This determined that all data are from statistics. Availability of data is the primary selection criterion. In datasets with fewer countries and fewer variables, missing values of each country and each variable were easy to control to low level. As Table 1 demonstrates, with more countries and variables in some datasets, such as the one in Publication V, overall missing values and missing values in some individual variables or countries were relatively high.

For Publications I-III and V, data were primarily from United Nations Development Program (UNDP), United Nations Office on Drugs and Crime (UNODC), World Bank, and Statistics Finland. Data for Publication IV were mainly from several institutions of the United States, including Department of Labour, Census Bureau, and the Disaster Center.

TABLE 1 OVERALL DESCRIPTION OF DATASETS USED IN THIS STUDY

Studies	Countries	Variables in original data set	Variables removed by using ScatterCounter	Missing values before attribute selection	Missing values after attribute selection
Publication I	50 countries	44 variables: 15 crime-related, 29 socio-economic	5 removed	5.0%	5.3%
Publication II	56 countries	28 variables: 13 crime-related, 15 demographic	0 removed (Scatter-Counter not used).	1.2%	Not used
Publication III	50 countries	30 variables: 13 crime-related, 17 economic	4 removed	5.7%	6.0%
Publication IV	1 country, USA	22 variables: 9 crime-related, 13 socio-economic variables over 48 years	0 removed	0.2% (2 missing values)	0.2% (2 missing values)
Publication V	181 countries and territories	69 variables: 1 crime-related (homicide), 68 socio-economic	7 removed	6.8%	7.3%

As it will be explained in Chapter 3 Section 2, original datasets were undergoing a selection process according to the separation powers of their variables by applying the method called ScatterCounter (Juhola and Siermala, 2012a, 2012b). The following

table also gives such information. Three studies that applied ScatterCounter method to select variables and had several variable removed had slightly more missing values after the selection than before.

Chapter 3 Methods Applied in the Study

The chapter describes methods used in clustering, attribute selection and cluster validation. The first section briefly introduces the methodological features of the self-organizing map and presents the application of the SOM in the study of social phenomena and crime. The second section briefly introduces the mechanism of ScatterCounter used to identify separation powers of attributes, assisting selection of attributes. The last section gives information on methods used in validating the classifications of the SOM.

3.1 Attribute Selection

Upon initial clusters were identified through preprocessing of data by the SOM, the structure of dataset was modified to be processed with ScatterCounter (Juhola and Siermala, 2012a, 2012b). The missing data values were replaced with the medians of the attributes (variables) computed from pertinent clusters so that the completed dataset could be processed by ScatterCounter. A main characteristic is that, these countries are labelled by cluster identifiers given by the preliminary SOM running with the original attributes. Note that in Publication IV countries were not labelled, because there was only one country, but successive years of the data applied.

The objective of ScatterCounter is to evaluate how much subsets labelled as classes (here 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 other 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 that 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 5 to 10 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, 2012a, 2012b). 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, 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 in order to be able to leave them out. 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, variables that have poor separation powers were removed from the dataset used in the subsequent processing and analysis.

The results of attribute selection were shown in Chapter 2, Table 1. In Publication II, ScatterCounter was not yet used. In other four studies where this method was used, all attributes in Publication IV were kept due to their usefulness of clustering. In Publications I, III, and V, the numbers of removed attributes were 5, 4 and 7, with removed and reserved attributes ratios 11.4%, 13.3% and 10.1% separately.

3.2 Clustering

As “the most popular artificial neural algorithm for use in unsupervised learning, clustering, classification and data visualization” (Cottrell and Verleysen 2006), the SOM has spread into numerous fields of science and technology as an analysis method (Kohonen et al. 2002, p. 111). Although nothing specialized on the study of crime has been published before, some literature provided some preliminary exploration into explanation for thinking Self-Organizing methods as feasible to do research on society as a whole.

Some literature has been aware of the necessity, possibility and feasibility for application of the SOM to the study of crime. They recognised that criminal justice is confronted with increasingly tremendous amount of data (for instance, in mobile communications fraud, Abidogun 2005). Crime data mining techniques become indispensable (Chung et al. 2005). They can support police activities by profiling

single and series of crimes or offenders, and matching and predicting crimes (Oatley et al. 2006).

The difference between new techniques and old ones has been revealed in some literature (Dittenbach 2000). For example, they pointed out that unlike traditional data mining techniques that only identify patterns in structured data, newer techniques work both structured and unstructured data. Researchers have developed various automated data mining techniques, depending heavily on suitable unsupervised learning methods (Dittenbach 2000). Cluster analysis helps the user to build a cognitive model of the data, thus fostering the detection of the inherent structure and the interrelationship of data (Dittenbach 2000).

The previous literature has almost provided with a cohort description on the SOM. Developed by Kohonen (Kohonen 1997) to cluster and visualize data, the SOM is an unsupervised learning mechanism that clusters objects having multi-dimensional attributes into a lower-dimension space, in which the distance between every pair of objects captures the multi-attribute similarity between them. Some applications based on the concept of the SOM were developed particular to meet the demand of law enforcement (for example, Fei et al. 2005, demonstrating that SOMs are quite efficient at aiding computer forensic investigators who are conducting a digital investigation to determine anomalous behaviours among the Internet browsing behaviour of individuals within an organisation; Fei et al. 2006, Lemaire and Clérot 2005). In particular, even though the data on the storage media may contain implicit knowledge that could improve the quality of decisions in an investigation, when large volumes of data are processed, it consumes an enormous amount of time (Fei et al. 2005). The SOM may play a positive role in exploratory data analysis (Lemaire and Clérot 2005).

Three categories of the network architectures and signal processes have been in use to model nervous systems. The first category is feed-forward networks, which transform sets of input signals into sets of output signals, usually determined by external, supervised adjustment of the system parameters. The second category is feedback networks, in which the input information defines the initial activity state of a feedback system, and after state transitions the asymptotic final state is identified as the outcome of the computation. The third category is self-organizing networks, in which neighbouring cells in a neural network compete in their activities by means of

mutual lateral interactions, and develop adaptively into specific detectors of different signal patterns (Kohonen 1990, p. 1464).

The SOM has attracted substantial research interests in a wide range of applications. The SOM can be sketched as an input layer and an output layer constituting two-layer neural networks. The unsupervised learning method is used in SOM. The network freely organizes itself according to similarities in the data, resulting in a map containing the input data.

The SOM algorithm operates in two steps, which are initiated for each sample in the data set. The first step is designed to find the best-matching node to the input vector, which is determined using the smallest value of some distance function, for example, the Euclidean distance function. Upon finding the best match, the second step, the "learning step" is initiated, in which the network surrounding node c is adjusted towards the input data vector. Let index i denote a model in node i. Nodes within a specified geometric distance, h_{ci}, will activate each other and learn something from the same n-dimensional input vector $\mathbf{x}(\mathbf{t})$ wheretdenotes the iteration of learning process. The number of nodes affected depends upon the type of lattice and the neighborhood function. This learning process can be defined as (Kohonen 1997, p. 87) with n-dimensional vector:

$$\mathbf{m}_i(t+1)=\mathbf{m}_i(t)+h_{c(x),i}(\mathbf{x}(t)-\mathbf{m}_i(t)). \quad (1)$$

The function $h_{ci}(t)$ is the neighbourhood of the winning neuron (node) c, and acts as the neighbourhood function, a smoothing kernel defined over the lattice points. The function $h_{ci}(t)$ can be defined in two ways, either as a neighbourhood set of arrays around node c or as a Gaussian function (Kohonen 1997, p. 87). In the training process, the weight vectors are mapped randomly onto a two-dimensional, hexagonal lattice. A fully trained network facilitates a number of groups.

The SOM algorithm results in a map exhibiting the clusters of data, using dark shades to demonstrate large distances and light shades to demonstrate small distances (U-matrix method) (Kohonen, 1997). Feature planes, which are single vector level maps, can additionally be generated to discover the characteristics of the clusters on the U-matrix map. They present the distribution of individual columns of data.

In applying the SOM, some recommendations should be followed so as to generate stable, well-oriented, and topologically correct maps (Kohonen and Honkela, 2007: 1), for example, in the form of the array, a hexagonal grid of nodes is to be preferred for visual inspection. In scaling of the vector components, usually

normalizing all input variables is used to make them to use the equal scale. For the purpose of quality of learning, an appreciable number of random initialisations of the $m_i(1)$ may be tried, and the map with the least error selected. While these recommendations are useful as a starting point for constructing the SOM, alternatives should also be tried in order for different datasets and their processing to attain best results, which may still be achievable with different strategies.

Crime is one of the social problems attracting the most attention, research of which can borrow ideas from generic or neighbouring subjects. A couple of applications of the SOM to social research can help frame the study of crime. In practice, the SOM is one of the models of neural networks that acquire growing application in social research. Deboeck (2000) clusters world poverty into convergence and divergence in poverty structures based on multi-dimensions of poverty using the SOM, which reveal how new knowledge can be explored through artificial neural networks for implementing strategies for poverty reduction. Crime-related social phenomena have also been studied with this method. For example, Huysmans et al. (2006) apply the SOM to process a cross-country database linking macro-economical variables to perceived levels of corruption with an expectation of forecasting corruption for countries. Li et al. (2006) develop a linguistic cluster model aimed at meeting the demand of public security index and extracting relational rules of crime in time series. Lee and Huang (2002) make an attempt to extract associative rules from a database to support allocation of resources for crime management and fire fighting. Findings of many such studies prove that artificial neural network is a useful tool in social research, particularly, in research of topics about international comparison (for example, Mehmood et al 2011).

Criminological research in detailed offences from micro viewpoints has also been acquiring more assistance from application of artificial intelligence. Hitherto, a great many of researchers focus on application of artificial neural networks to law enforcement, in particular, detection of specific abnormal or criminal behaviours. Adderley et al. (2007) examine how data-mining techniques can support the monitoring of crime scene investigator performance. Oatley et al. (2006) present a discussion of data mining and decision support technologies for police, considering the range of computer science technologies that are available to assist police activities. Dahmane et al. (2005) have presented the SOM for detecting suspicious events in a

scene. These practical usages opened the door for artificial intelligence to play a part in the study of crime.

Some continuing and consistent studies have been done in detection of particular offences. The SOM has been, for instance, applied in (research on) detection of credit card fraud (Zaslavsky and Strizhak 2006), automobile bodily injury insurance fraud (Brockett et al. 1998), burglary (Adderley and Musgrove 2003, Adderley 2004), murder and rape (Kangas 2001), homicide (Memon and Mehboob 2006), network intrusion (Rhodes et al. 2000, Leufven 2006, Lampinen et al. 2005, Axelsson 2005), cybercrime (Fei et al. 2005, Fei et al. 2006), mobile communications fraud (Hollmén et al. 1999, Hollmén 2000, Grosser et al. 2005). Literature in this aspect has been abundant. And this is the primary field where the SOM has found application to research related to criminal justice.

Besides crime detection, neural networks are also found useful in research specialized in victimization detection in mobile communications fraud (Hollmén et al. 1999).

From present literature, the SOM has been applied in detection and identification of crimes. Application of the SOM to the study of crime, that is, in visualizing geographic distribution and historical development of criminal phenomena, in identifying correlation factors or recognizing preventive or deterrent factors, few have been published. Upon recognizing the current situation, there is a necessity for designing experiments exploiting this approach, in comparison with other methods.

In data processing and map visualization, software tools must be used. There are a handful of tools available. In this dissertation, two primary software tools are in use:

1) SOM Toolbox for Matlab. SOM Toolbox is a function package for Matlab 5 implementing the self-organizing map (SOM) algorithm and more. It can be used to train SOM with different network topologies and learning parameters, compute different error, quality and measures for the SOM, visualize SOM using U-matrices, component planes, cluster colour coding and colour linking between the SOM and other visualization methods, and do correlation and cluster analysis with SOM (SOM Toolbox Homepage. Retrieved 27 April 2011 from http://www.cis.hut.fi/projects/somtoolbox/). In Publication II, SOM Toolbox was used in processing data and clustering. Clustering is defined as the process of classifying a large group of data items into smaller groups that share the same or similar properties (Suh 2012, p. 280).

(2) Viscovery SOMine 5.2. “Viscovery SOMine is a desktop application for explorative data mining, visual cluster analysis, statistical profiling, segmentation and classification based on self-organizing maps and classical statistics in an intuitive workflow environment.” (Viscovery 2013) In Publications I, III, IV, and V, Viscovery SOMine was used to processing data, clustering and identifying correlations.

Figure 1 is a map generated by Viscovery SOMine from Publication V, consisting of 7 clusters formed by 181 countries. Clusters 1-7 covered 23, 26, 33, 24, 30, 15, and 30 countries separately.

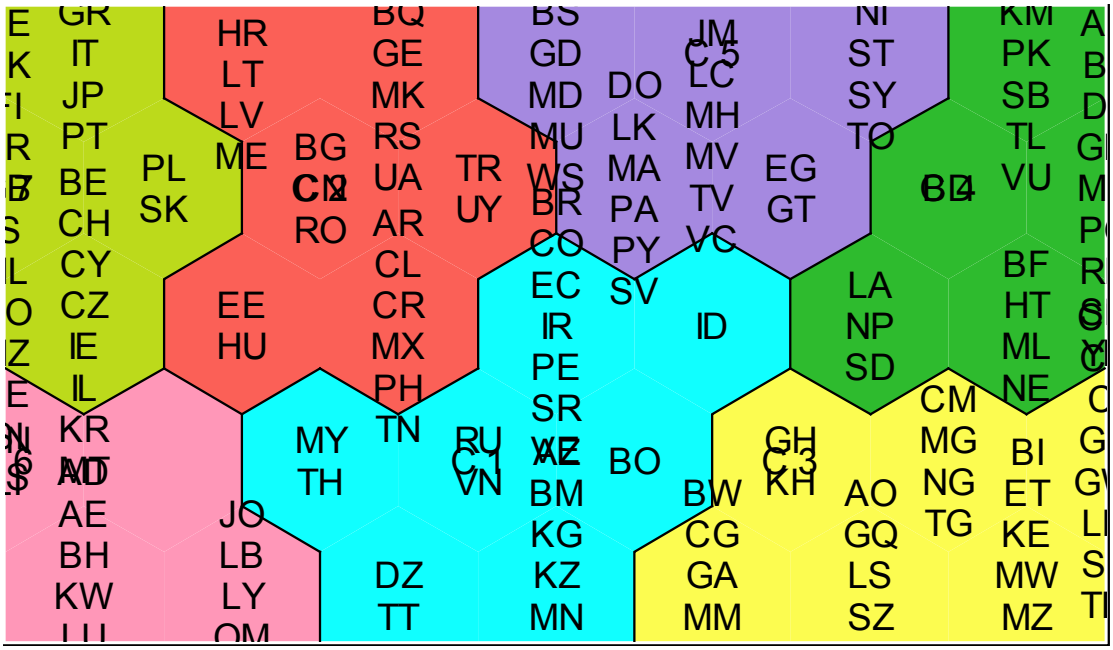


FIGURE 1 Clustering map with cluster names C1–C7 and labels of countries. (See the labels from Table 1 of Publication V.)

Because the unsupervised clustering map and feature maps were generated based on 62 attributes, description of these clusters became more complicated. Particularly, when special information about one attribute is needed, countries and territories in these seven clusters may be better regarded as components in fewer numbers of super-clusters. For example, according to the feature map of homicide rate (Figure 2), these seven clusters can be seen as components in three super-clusters:

The first one consists of C3 (34 countries) and C5 (30 countries). They have higher level of homicide rate.

The second one consists of C1 (23 countries) and C4 (24 countries). They have medium level of homicide rate.

The third one consists of C2 (26 countries), C6 (14 countries) and C7 (30 countries). They have lower level of homicide rate.

Certainly, according to other attributes, there were more possibilities to form different super-clusters, which would find their use in different research interests.

On the other hand, where necessary, within the frameworks of each of these seven clusters, several sub-clusters could also be identified. For a random example, in cluster 5, five countries, Dominican Republic (DO), Sri Lanka (LK), Morocco (MA), Panama (PA), Paraguay (PY), and El Salvador (SV) form a sub-cluster. It implied that they have closer common properties than those members in the same cluster. Because they were closely grouped with each other, their clustering would not differ in feature maps of different attributes.

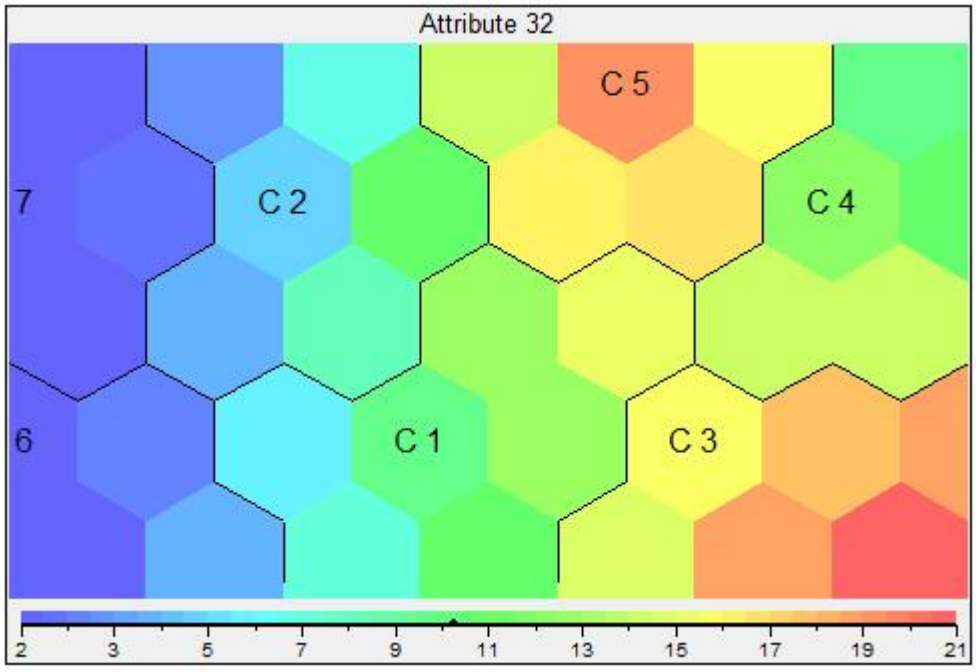


FIGURE 2 Feature map of homicide rate

While most countries were assembled in large or small groups, a few countries were isolated. They stayed separately far away from other countries, such as Bangladesh (BD), Bolivia (BO), and Indonesia (ID). Although they have much in common with other countries in the same clusters, the map can still be used in a way of establishing the elaboration of diversity.

3.3 Classification Validation

The results by the SOM were tested several methods for classification validation. In different studies, different methods were selected among by *k*-means clustering (Publications III, IV, V), discriminant analysis (I, III, IV, V), *k*-nearest neighbour search (I, III, IV, V), decision trees (I, III, IV, V), Naive Bayes classification (V), support vector machines (V), Kruskal-Wallis test (V), and Wilcoxon-Mann-Whitney U test (V).

For validation tests, the leave-one-out method was used where one country (or year in Publication IV) formed a test instance, one by one, and the other *n*-1 countries (or years) formed the corresponding training set used for building a classification model. Finally, all *n* test results were used jointly to give their accuracy result. The leave-one-out method is appropriate for small data sets as to numbers of instances.

The method of *k*-means clustering is "a process for partitioning an *N*-dimensional population into *k* sets on the basis of a sample", during which partitions are given efficiently as regard to the variance within the classes (MacQueen 1967, p. 281). It can be applied in similarity grouping and other fields (*ibid.* p. 281). During this research, it was used as a validation method to test and verify the results generated by the SOM method.

Discriminant analysis was proposed by R. A. Fisher and thereafter became a classic method of classification (Fisher 1936). It is the proper statistical technique when the dependent variable is categorical and the independent variables are quantitative (Burns and Burns 2008, pp. 589-590). It can be used in classifying cases into groups, in which sense it can produce comparable results as those from the SOM and thus is useful in acting as a validating method.

Based on the concept of similarity, the technique of *k*-nearest neighbour search is used in constructing a classification method exclusive of building hypothesis regarding the shape of the function that establishes relationship between the dependent attribute and the independent attributes. The aim of this technique is to dynamically identify *k* observations in the training data that are comparable to a fresh observation in which classifications are expected (Palmer, Jiménez and Gervilla 2011,

p. 380). In this dissertation, it was used as a validation method against the result of the SOM.

Decision trees are also a frequently used data mining tool, indicating a predictive model that is used, among others, in classification (Rokach and Maimon 2008, p. 5). As a result, they become sequential partitions of a set of data maximising the differences of a dependent variable Palmer, Jiménez and Gervilla 2011, p. 377). In this sense, it can also be called classification trees (Rokach and Maimon 2008, p. 5).

Based on Bayes theorem, Naive Bayes classification can predict the probability of a given case belonging to a certain class (Palmer, Jiménez and Gervilla 2011, p. 380). It articulates a commanding structure to join information from the sample with prior expert opinion so as to generate an up-to-date posterior expert opinion (Giudici, 2003, as cited in Palmer, Jiménez and Gervilla 2011, p. 382). In this way, it is feasible in validating results from the SOM taken as prior opinion.

Support Vector Machine (SVM) (Vapnik 2000; Burges 1998; Cortes, and Vapnik 1995) is a supervised classification method developed for two-class classification problems. The key idea in SVM is to construct a classes separating hyperplane in the input space such that the margin (distance between the closest examples of both classes) is maximized.

Kruskal-Wallis test is developed for the purpose of deciding "whether samples should be regarded as coming from the same population." (Kruskal and Wallis 1952, p. 584). The results got from this test can also be used in validation (Cieslak and Chawla 2007, p. 127).

Wilcoxon-Mann-Whitney U test was developed by Mann and Whitney in 1947, when they extended the Wilcoxon's equivalent test (Mann and Whitney 1947, pp. 50-60). As their paper's title, this method was designed to test "whether one of two random variables is stochastically larger than the other" (*ibid.*) It is one of the appropriate methods to compare a control group with a group receiving treatment (*ibid.*).

All of them are frequently used classification or statistical testing methods. By comparing the results generated by these tools and those generated by the SOM, accuracy of the SOM can be expressed in percentages, which denote how similar these results are.

3.4 Correlation

Viscovery SOMine could generate a detailed list of correlations. In Publication V, for example, the correlations between socio-economic attributes *A* and homicide rate (attribute 32) were generated based on data on 62 attributes from 181 countries. This provided 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. Traditionally, single research on crime did not include so many attributes (or named correlation factors). So it shall be highly expected to have such data mining methods to be able to process several dozens of attributes and to provide immediate reference for further analysis.

3.5 Summary of data processing

The overall background of the research and the data processing can be summarised by the following diagram (Figure 3).

In this figure, the light cloud represents the conventional criminological research that can be used in comparison with further investigation using the current methods. The dark cloud denotes to the data sources that can be utilised by different ways, including the SOM, for the purpose of knowledge mining. Upper parts composed of components without shadow describe the pre-processing of original data, with an action of data refinement using ScatterCounter. Processing of data is based on the refined data, and facilitates further inquiry. Primarily, clustering maps, feature maps and correlations are more relevant in the research (particularly for future and on a larger scale). Among them, clusters are validated by using other methods as listed in Section 3.3.

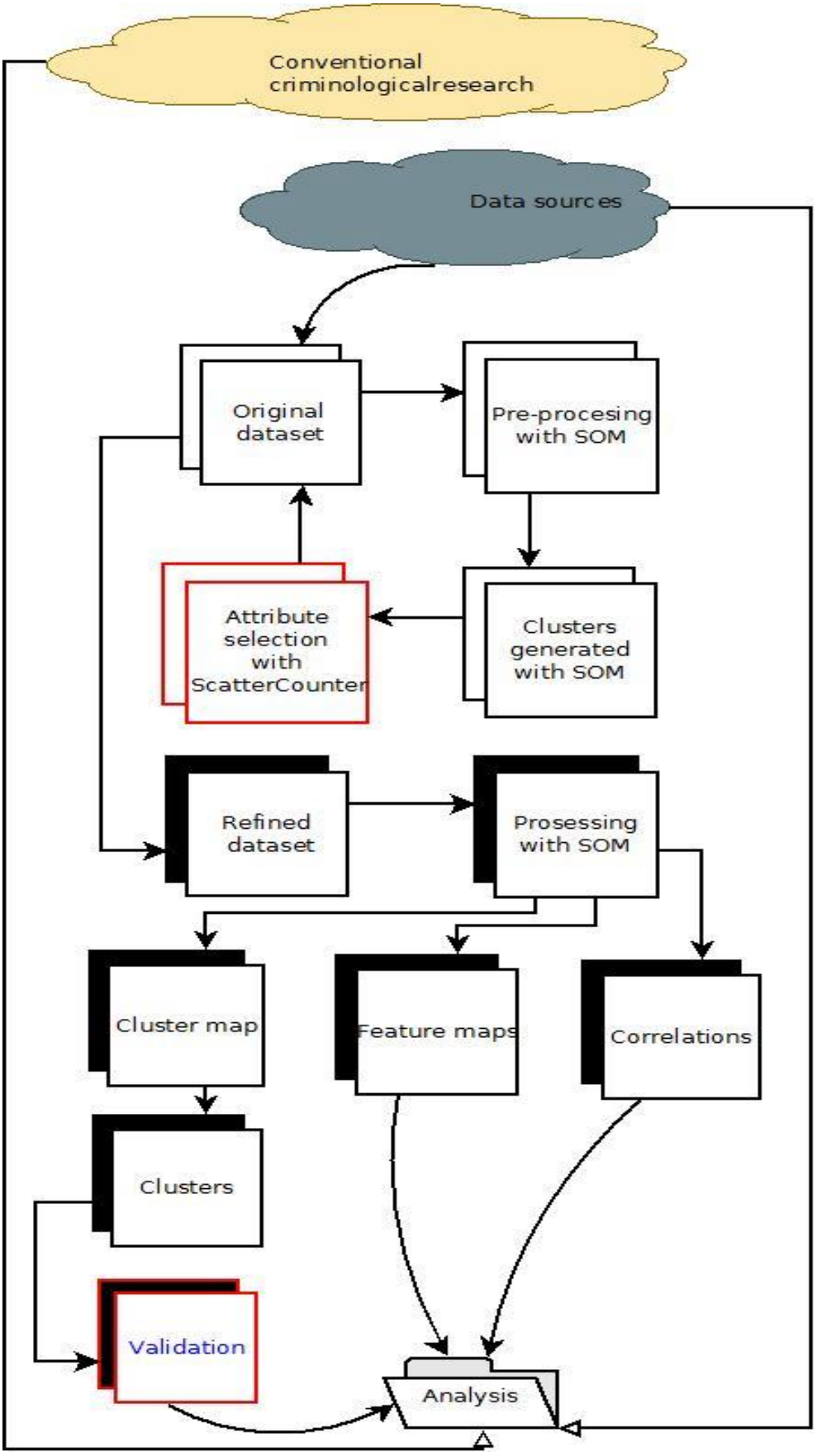


FIGURE 3 Diagram of data processing

Chapter 4 Results

4.1 Publication I Crime and social context, a general examination

The purpose of this study was to apply the SOM in mapping countries with different situations of socio-economic development. This study represented the efforts to innovatively apply the SOM to the study of crime, previously untouched by many other studies. Almost all of the previous studies were microscopic, while this study was macroscopically oriented. It did not look into individual offences in reality taking place, or perpetrators in reality acting. Rather, it revolved around a broad range of crimes, and socio-economic factors accompanying these crimes. Besides the SOM, nearest neighbour search, discriminant analysis and decision trees were used to validate the clusters and analysis by calculating how accurately these methods put the same countries into the same clusters as the SOM does.

There were 50 countries included in the experiment. This study contained 44 crime and socio-economic factors. Twenty-nine of them were socio-economic factors, while the rest fifteen were crime-related indicators. After the dataset was established for processing, Viscovery SOMine was used to generate clusters. Upon initial clustering, ScatterCounter (Juhola and Siermala 2012a, 2012b) was used to identify separation powers of different variables. With these results and observations, five variables had poor separation powers and were removed from the dataset used in the following experiments and analysis. As a result, all the left 39 variables preserved in the dataset had stronger separation power and supposedly enable valid clustering. In this reduced dataset, total missing values accounted for fewer than 5.3%. Of total 39 attributes, 22 had no missing values. The highest frequency of missing values in one attribute reached 38%.

With the SOM, clustering maps, feature maps and correlation table were generated from the dataset of all 50 countries and all 39 variables. Cluster 1 consisted of 13 countries with high of total crime rate. They were characterized by the high rape rate and high convicted rate per 100,000 people, but low level of homicide rate. According to the figures, most socio-economic indicators were located in the upper lever among the sample. In these countries, there were more prisons and at the same

time the prison capacity was highly filled. Cluster 2 consisted of 15 countries with the highest level of total crime rate. In these countries, according to the figures, there was high level of prisons per 100,000 people, but lower level of prisoner rate and prison capacity was less filled. These countries were characterized by highest level of socio-economic indicators. Cluster 3 consisted of 12 countries with the low level of total crime rate. However, they were characterized by the highest level of murders per 100,000 people. Some countries had high level of some socio-economic indicators, but others not. They had high level of burglaries per 100,000 people. In these countries, there were more prisons on average, and these prisons were highly filled. Yet, in these countries, according to the figures, there were lowest level of the convicted per 100,000 population. Cluster 4 consisted of 10 countries with low level of total crime rate. In these countries, there was highest level of prison capacity, which was the least filled. According to the figured, rates of rape and homicide per 100,000 population were the highest among the sample.

The results by the SOM were tested by nearest neighbour search method also applied with Euclidean distance. To look at the closest neighbours using k=3 and k=5 were used. When k=3 nearest neighbours, this method gave the probability of 80% correct compared with the results of the SOM when data were not scaled, and yet higher value, 94% when data were scaled. When k=5, this method gave the probability of 74% correct when data were not scaled, and higher value, 92% when data were scaled. Apparently, scaling of data significantly improved the probability of correct results in accordance with the SOM results.

In addition, logistic discriminant analysis was applied to compare with the SOM results. Whether the data were scaled or not scaled, the probability of putting into the same classes as the clusters generated by the SOM was 72%. Besides, linear discriminant analysis was also carried out, obtaining a result of 70%.

Finally, further tests were carried out by applying decision trees. When data were not scaled, the value was 72%. When data were scaled, the result was slightly improved, with a value of 74%. In these two cases, scaling of data got somewhat better result.

Using Viscovery SOMine, a detailed list of correlations was generated, showing the roles of selected socio-economic factors against different crime-related factors. Correlations from interval (-0.3,0.3) were left out seen as insignificant, absolute values from [0.3,0.6] were seen as interesting, and from [0.6,1] as significant.

Although even strong correlation between two attributes did not indicate one to be caused by another, this rendered materials for further analysis and reference. There were many potentials that these results could be used to compare with previous studies using other methods.

The results of the study provided further evidence that the self-organizing map is a feasible and effective tool for crime situation benchmarking. The results were easy to visualize and interpret, and provide a very practical way to compare the demographic factors of countries with different crime situation. This study showed that crime research is one application area that can benefit from better visualization and data mining techniques.

4.2 Publication II Crime and demographic factors

The emphasis of this study was rendered on the application of the SOM to explore into the relationship between demographic factors and crime based on 56 countries as study objects.

In this study, demographic factors were roughly divided into three categories: population structure, population quality, and population dynamics. Concerning the relationship between crime and some main aspects of demographic factors, there have been abundant sociological studies using different methods. Consequently, this study as well as all other studies included in this dissertation, has special significance because of its methodological starting point and its transdisciplinary feature.

Five clusters were identified by the SOM. Cluster **A** consisted of countries with very high level of total crime rate, including 12 countries, which had high GDP level, and satisfactory demographic indicators. They were characterized by, for example, high health expenditure, high life expectancy, aging, and high urbanization. Except murder, robbery and software piracy, other severe crimes, such as rape, fraud, assault and burglary were all at higher level than in most countries in other clusters. Cluster B consisted of countries with a high level of total crime rate, including 12 countries, which were characterized by having the highest death rate, the biggest rate of population older than 64, the highest murder rate and highest share of prison capacity filled. Otherwise, most demographic indicators were at satisfactory levels. Cluster C consisted of countries with a medium level of total crime rate, including 13 countries,

which have most indicators at medium levels, including demographic factors and crimes. Instead, they were characterized by the lowest level of net immigration rate, and the highest police per 100,000 population and rape rate. Cluster D consisted of countries with a low level of total crime rate, including 6 countries, which had the highest adult illiteracy rate, birth rate and death rate, fertility rate, population growth rate, population undernourished, under five mortality rate and unemployment rate. Prisoner per 100,000 people was also the highest. However, except rape rate and jails per 100,000 people, all other severe crimes were at a low level. Cluster E consisted of countries with a very low level of total crime rate, including 13 countries. Not all the indicators in countries of this cluster were lower than any other countries. Usually, in these countries, demographic indicators fall between two extremes. It also deserves to mention that, for example, they had the highest net immigration rate. Software piracy rate was the highest compared with other countries included in this study.

In this publication, a correlation table was generated. Furthermore, it also did some specialized case studies. One example was the correlation between prisoners per 100,000 people and other 27 attributes. Stronger positive correlation could be found with adult illiteracy, software piracy per 100,000 people and police per 100,000 people, possibly with infant mortality rate, under-five mortality rate, unemployment rate total, and share of prison capacity filled. Stronger negative correlation could be found with health expenditure per capita, life expectancy, and possibly urban population. Correlation between prisoners per capita and all other attributes, regardless of positive or negative, has shown to be weak.

Roughly defined patterns of crime situation have been found in some countries with some similar characteristics. On the other hand, in different groups of countries, different factors might work in different ways. They might have positive correlation with crime in some countries, but had negative correlation with crime in other countries. They might have weak correlation in some countries, but had strong correlation in some other countries. These could not be solved by the SOM alone and should be supplemented through other research routines.

4.3 Publication III Crime and Socio-economic factors

Another aspect deserved particular attention was the relationship between economic background where the criminal phenomena are located. This publication was designed to explore the possibility of applying the SOM to study the crime based on a set of economic data covering 50 countries and 30 variables. During the application of the SOM, in order to select variables, ScatterCounter (Juhola and Siermala 2012a, 2012b) was used to identify strong and weak variables in clustering, four variables have poor separation powers and were removed from the dataset used in the following experiments and analysis. As a result, all the 26 variables preserved in the dataset have strong separation power and supposedly enable valid clustering. In this reduced dataset, total missing values account for 6%.

Based on the data set after variable selection, the SOM generated 4 clusters. Cluster 1 consisted of 19 countries both with the low level and medium level of total crime rate. Most of them were economically developed with the high level of GDP per capita, highest rate of employment in services. They had high level of fraud, assault and burglary rates. Certainly, there were also countries with different characteristics in some fields such as economy, etc. Cluster 2 consisted of 15 countries with high level of total crime rate. They were characterized by the lowest level of murders per 100,000 people. Some countries had high level of some economic indicators, including employment in services, Internet and phone mainlines users. They had high level of burglaries per 100,000 people. Cluster 3 consisted of 13 countries with medium level of total crime rate. They were characterized by the low rape rate and police per 100,000 people. According to the figures, these countries had high level of cellular subscribers per 1000 people, employment rate in industry, and imports of goods and services per GDP. According to figures, in these countries, there were more jails and more people in these jails on average. Cluster 4 consisted of 3 countries with the low level of total crime rate. These countries were characterized by high employment rate in agriculture and software piracy. Many other aspects seemed to be low, both economic indicators and crime-related statistics.

In order to validate the SOM clustering, *k*-means clustering and nearest neighbour search were used. The method of *k*-means clustering with Euclidean distance got 49.5±0.9% correct in accordance with the SOM results when data were

not scaled, and 48.3±0.7% when data were scaled. Leave-one-out testing with k-means clustering run in the supervised manner got 43.5±2.0% (not scaled) and 45.4 ±1.7% (scaled). The method of *k*-nearest neighbour search got 50% (not scaled) and 60% (scaled) separately.

A detailed table showed absolute correlations from interval [0.3,0.6] were seen as interesting, and from [0.6,1] as significant. Traditionally, economically poor condition has been regarded to lead to crime. In this experiment, it was not clear. Both murder and prison population rates were negatively relevant with the GDP per capita. In this experiment, however, the results demonstrated that there were only few strong links between any pairwise attributes.

The results of the study provided further evidence that the self-organizing map could be a feasible and effective tool for the study of crime. The results were easy to visualize and interpret, and provide a very practical way to compare the economic factors of countries with different crime situation. This publication showed that the study of crime is an application area that can benefit from efficient data analysis and visualization techniques.

4.4 Publication IV Crime and historical development, the case of the United States

This study applied the SOM in the study of historical development of crime in the United States during 48 years (1960-2007). Including an analysis based on available data, the results of the study revolved around whether the SOM could also be a feasible tool for mapping criminal phenomena through processing of historical crime data.

Crime in the United States had a remarkable increase and decrease in the latter part of the 20th century. In this publication, the topic concerning the rise and fall of American crime was examined using the self-organizing map.

Twenty-two variables covering demographic and economic situation were selected to model the state of the socio-economic system for the period of 48 years (1960-2007). The emphasis in the selection of variables in this study was on demographic and economic data and rates of different crimes. In total, the data set used consisted of 48 years and 22 variables. Some other possibly important variables

had to be discarded due to the substantial amount of data missing. As a result, missing values constituted only under 0.2% in this data set. With ScatterCounter method, almost all had certain level of positive separation powers and were kept in the data set used in the subsequent experiments and analysis.

By the SOM, forty-eight years were grouped into 6 clusters, each of which representing different crime, demographic and economic circumstances. Cluster 1 contained the late 1990s and early 2000s. The crime rate in the US had been decreasing steadily. Cluster 2 dealt with some years of the early 1970s, the end of the 1970s, and the early 1980s. The crime rate in the US kept rising sharply in the 1970s. Cluster 3 included some other years of the early 1970s, late 1970s and mid-1980s. During these years, the crime rate was still in its increasing route, but in some years such as 1975, 1980, and 1982, there were slight falls. Cluster 4 began from the mid-1980s, through the late 1980s and entered the early 1990s. The crime rate rose throughout the 1980s, reached its peak in 1993 and then began to decrease throughout the 1990s. Cluster 5 covered the mid 1960s through the late 1960s. The crime rate in the US had risen sharply since the late 1960s. Cluster 6 included the early 1960s. The crime rate in the US was at its lowest point during the last five decades.

Using *k*-means clustering, when data were not scaled, 65.07±4.19 % valid results could be got, when scaled, 79±6.32%. The use of *k*-nearest neighbour search got 81.3% and 91.7% separately. Decision trees got 83.3% and 81.3%, while logistic discriminant analysis got 77.1% correctness in accordance with the SOM results.

Using Viscovery SOMine, correlations between every pair of attributes were identified. In this experiment, many of the results demonstrate that there were strong links between some attributes. The US has been a politically stable country over years. Every attribute was on a growing track, both favourable indicators and unfavourable indicators. There has been no sharp increase or sharp decrease during the past 50 years. For example, unemployment rate fluctuated between 3.5% and 9.7%. These figures could be explained as forming a large range. However, in contrast with many other countries, these figures were in relatively low values. So we have to look for, from the slight change, the correlation between unemployment rate and crime. It turned out that it has no strong level of correlation on crime rate.

Due to above reason, successive years were clustered together in one way or the other. The exceptions were rare. This demonstrated a one-direction developmental tendency of most variables. In such a case, traditional statistical methods could also

provide a good depiction for each single variable with tables or charts. However, comprehensive visualization could be achieved by applying the SOM as in this experiment.

4.5 Publication V Crime and social context, the case of homicide

Publication V was designed to provide a specialized study on one of the most serious types of offences: homicide. The purpose of current study was to find groups of countries and territories according to their homicide rate, and to explore correlation between homicide and its socio-economic context. This international-level comparative study used a dataset covering 181 countries and 69 attributes.

One attribute, homicide per 100,000 people was a crime-related indicator, while 68 others were socio-economic factors. The selection of the contents of these indicators was primarily based on availability of data. Less consideration was put on the traditional concept on what might in actual fact cause the occurrence of offences of homicide, because in this research pre-determined and presumed correlations were temporarily ignored. Accordingly, in this research, some of these factors might traditionally be considered closely related to homicide, but some others might be considered quite irrelevant. Both of these categories of factors were re-considered in this research with a view to search potential clues for a new explanation. With ScatterCounter, seven attributes (6, 8, 17, 18, 41, 63, 64) have poor separation powers and are removed from the dataset used in the following experiments and analysis.

Upon processing of data, seven clusters were generated, each representing groups of countries sharing similar characteristics. In analysing country homicide situation to fulfil different demands, when there were many objects involved, other two levels of clustering concepts: super-clusters and sub-clusters, could also be used. Within the frameworks of each level of clusters, members in each cluster have their common properties, based on which they were grouped, and based on which they could be further analysed.

Compared with clusters generated by the software before several attributes were removed by applying ScatterCounter to select more useful attributes based on their separation powers, clusters here were to some extent re-grouped. Principally, clusters before and after removing the attributes that had weak separation powers should be

with high similarity, including cluster numbers, and countries in each cluster. That could be taken as the initial purpose for applying separation power.

Although the Viscovery SOMine software package provided the possibility for adjusting the number of clusters, and this could be used to set the same number of clusters for experiments before and after the application of separation power, 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, countries in these clusters were still re-grouped slightly one-way or the other. In this experiment, a more significant change of cluster number was still tolerated, because this was expected to leave a new space where the similar issue could be speculated.

By emphasizing difference between clusters before and after removing weak attributes by applying the separation power, it did not ignore the actual fact that a majority of countries that were originally in the same small groups (sub-clusters in clusters) were subsequently still in the same sub-clusters. That is to say, clusters changed, but the change took place primarily at the sub-cluster level, not the individual level. Single countries did not move from here to there separately. Rather, closely joined small groups of countries migrated from one cluster to another.

This phenomenon enabled research on small groups of countries on the background of the whole world, by subtracting information from the self-organizing map established by processing the data depicting the panoramic view.

Because the unsupervised clustering map and feature maps were generated based on 62 attributes, description of these clusters became more complicated than in the other publications of the thesis. Particularly, when special information about one attribute was needed, countries and territories in these seven clusters might be better regarded as components in fewer numbers of super-clusters. For example, according to feature map of homicide rate, these seven clusters could be seen as components in three super-clusters: The first one consisted of Cluster C3 (34 countries) and Cluster C5 (30 countries). They had higher level of homicide rate. The second one consisted of C1 (23 countries) and C4 (24 countries). They had medium level of homicide rate. The third one consisted of C2 (26 countries), C6 (14 countries) and C7 (30 countries). They had lower level of homicide rate. Certainly, according to other attributes, there were more possibilities to form different super-clusters, which would find their use in different research interests.

On the other hand, where necessary, within the frameworks of each of these seven clusters, a number of sub-clusters could also be identified. For a random example, in Cluster 5, six countries, Dominican Republic, Sri Lanka, Morocco, Panama, Paraguay, and El Salvador form a sub-cluster. It implicated that they had closer common properties than those members in the same regular cluster. For the reason that they were closely grouped with each other, their clustering would not differ in feature maps of different attributes.

While most countries were assembled in large or small groups, a small number of countries were isolated. They stayed unconnectedly distant from other countries, such as Bangladesh, Bolivia, and Indonesia. Even though they had a great deal in common with other countries in the same clusters, the map could still be used in a way to establish the elaboration of diversity.

The results by the SOM were tested by several methods, including *k*-means clustering, discriminant analysis, *k*-nearest neighbour classifier, Naïve Bayes classification, decision trees, support vector machines (SVMs), Kruskal-Wallis test, and Wilcoxon-Mann-Whitney U test. The best classification results produced accuracies of 50% which are fairly good taking into account the dimension of the data set and that the largest cluster included 34 countries out of 181. In other words, its a priori probably was (34/181) 100%~18.8% meaning that a pure guess would probably give a correct result on the average of around 19%. Usually, scaling of data into [0,1] could get higher positive rates, standardization (all values of each variable subtracted with its mean and this difference divided by its standard deviation) improved the rates further. Exceptions also existed when lower rates emerged in the tests where scaling and standardization were applied. The results of the Kruskal-Wallis test showed that there were significant ($p < 0.05$) differences among the groups defined by the clusters in 60 out of the total 62 variables. On average, six out of the 21 pairwise test results obtained with the Wilcoxon-Mann-Whitney U test were significant after the p values were corrected with the Holm's method.

As for correlations, homicide was positively correlated with 23 attributes, while negatively correlated with rest 38 attributes. Some correlation values were interesting, while others were very weak. Besides its premature, this study provided a string of thinking concerning potential research on what socio-economic factors cause homicide, affecting its occurrence, or its increase or decrease, especially when more factors than that the traditional research had coverage could now be involved.

Chapter 5 Conclusions

The innovative feature of this research was rooted in the reality that there has neither been comparable inquiries in applying the SOM to macroscopically mapping international criminal phenomena and identifying correlation factors, nor a design as such. In addition to examining the feasibility of applying the SOM to the study of crime, this dissertation also detailed the potentialities of application of the SOM in this field.

Assisted by ScatterCounter to select attributes and refine the dataset, and by nearest neighbour search, discriminant analysis, decision trees, and other statistical methods to test the correctness of clustering, the results of the study provide further evidence that the self-organizing map is a feasible and effective tool for the study of crime. The results are easy to visualize and interpret, and provide an especially practical approach to compare the socio-economic factors of countries with diverse crime situation. This research has shown that the study of crime is an application area that can benefit from efficient data analysis and visualization techniques.

The self-organizing map has its advantage in visualizing criminal phenomena as a whole and their interactive relationship with demographic factors or socio-economic factors; and for comparing geographical distribution of crime in different groups of jurisdictions. Depicting crime tendencies has been of immense value for criminologists and international law enforcement. It is potentially constructive for stakeholders and decision-makers in legislature and law enforcement to adjust policies in combating crimes based on analysis of geographic and historical factors.

With the self-organizing map, multidimensional comparison could be realized. Countries and territories, as the research objects could be grouped into clusters of different levels: super-clusters, clusters, and sub-clusters. Super-clusters were useful to reveal common features of one or more attributes of research objects (countries and territories). They could be used in different grouping ways in connection with different attributes. Normal clusters provided primary foundation for the investigation of distribution of countries and territories with all attributes in the dataset. Sub-

clusters were used to investigate smaller groups of countries and territories contained in a cluster. In principle, countries and territories within a sub-cluster could be seen as having the most features in common in socio-economic context.

In processing of data, when attributes are numerous, results of clustering may become difficult since these might show a discrepancy. Specifically, a number of the countries might be clustered into "wrong" groups. So that it creates complexity to provide descriptions closer to the reality. Another aspect in the SOM is that, alteration of parameters in processing the data can cause change in clusters and maps as well. To this point, the SOM might be inferior to traditional processing methods.

Frequently, correlation between every pair of attributes demonstrated that there were merely a small number of strong links between any attributes, but numerous less-strong links were interesting. It must be pointed out that, correlation is an outcome that can regularly be generated by means of statistical techniques. It has some significance for additional consideration in identifying the causes. Nevertheless, correlated factors might contain but are not equivalent to causes. Correlation expressed in numbers can happen to be erroneous and it does not correspond to concluding analysis of the problem. In addition, given that correlation (Pearson product-moment correlation coefficient) is of linear type, it cannot divulge into relations of more complicated feature between attributes.

Contrasting the natural phenomena where there is not so much value judgment, correlation's extra inconsistency occurs when the public in the society presume something beneficial, but statistically it positively correlates with total crime rate or a certain type of exceptionally serious offence. Or vice versa, can crime take place on a background where some especially positive, favourable and good factors are located? The answer is positive. It is also correct that some negative and unfavourable socio-economic factors in actual fact bring about just less crime. Accordingly, correlation produced by statistics ought to be accompanied by supplementary analysis by adopting other methods including traditional ones.

Situation of crime in a country does not necessarily correspond to a country's socio-economic image, neither to the quantity of a country's total crime. The majority of the highly developed countries have high total crime rates. A universal misunderstanding is that each indicator should be satisfactory in developed countries. This study proves again that developed countries have higher total crime rates. This does not imply more than the fact. But behind the fact, there are always possible

statistical backgrounds: more transparent justice system, well established statistical institution, low crime standard, and zero tolerance approach to even minor crimes.

Correlation at macroscopic level cannot be directly applied at microscopic level. The former does not make sense in the latter. For example, countries with highly developed economy usually have high level of crime. This must be explained with the assistance of the techniques but also beyond the techniques. One of the tricks is that, most developed countries, such as the USA, have far stricter laws than these are in other countries. What are crimes in the USA might be so tiny, so trivial an act that they are not punishable in other parts of the world. Some developing countries have in actuality fewer offences, because of their merits in cultural tradition (so that smaller quantity of offences were committed), or shortage of investment in public security (so that smaller quantity of offences were detected), or law being frequently ignored (so that larger quantity of offences were not regarded as crime), or merely due to statistical defects (so that smaller quantity of offences were recorded). All these should be studied by using different methods, computational as well experts' consideration, but beyond the simple number of correlation. It is unworkable to formulate a clear-cut judgment.

Summarily, this study advised that the SOM could be used to improve the efficiency of implementing the mapping, clustering and correlation identifying during the study of crime. However, in order to build a bridge across these problems, intellectuals and official stakeholders must take action in applying these methods into additional analysis and solution. Simultaneously, reliability of data and accuracy of the data mining results must also be taken into account in an attempt to elevate the effectiveness of policy making.

Chapter 6 Personal Contributions

In the following the contribution of the present author is described. The author (referred to as XL hereafter) of the dissertation worked together with Martti Juhola (ML), Henry Joutsijoki (HJ), Jorma Laurikkala (JL), and Markku Siermala (MS).

I, III, IV. XL designed the framework of the study, implemented all the data collection, explored variables from the databases of the United Nations, Statistics Finland, and other sources evaluating (according to his previous knowledge, such as Li 2008) which of them would be useful and interesting for the current studies, and planned processing related to the Self-Organizing Map, did some validation tests and organized the writing-up. MJ provided overall guidance and did most validation tests.

II. XL designed the framework of the study, implemented the data collection and processing related to the Self-Organizing Map, and organized the writing-up. MJ provided overall guidance, gave comments, and corrected or added some technical explanations.

V. XL designed the framework of the study, implemented the data collection and processing related to the Self-Organizing Map, and organized the writing-up. MJ did not only provided overall guidance to the implementation, organized the validation tests, but also did some validation tests. HJ did support vector machines tests and provided detailed explanation on this test. JL did Kruskal-Wallis test and Wilcoxon-Mann-Whitney U test. MS aided by MJ developed and shared a new version of ScatterCounter and it was in use in this study (also, an old version was used in other studies).

Bibliography

1. Abidogun, O. A. 2005. *Data mining, fraud detection and mobile telecommunications: call pattern analysis with unsupervised neural networks*, PhD thesis, University of the Western Cape, South Africa.
2. Adderley, R. 2004. The use of data mining techniques in operational crime fighting, in *Proceedings of Symposium on Intelligence and Security Informatics*, no. 2, Tucson A.Z., ETATS-UNIS (10/06/2004), vol. 3073, pp. 418-425.
3. Adderley, R. and Musgrave, P. 2003. Modus operandi modelling of group offending: a data-mining case study. *International Journal of Police Science and Management*, vol. 5, no. 4, pp. 265-276.
4. Adderley, R., Townsley, M. and Bond, J. 2007. Use of data mining techniques to model crime scene investigator performance, in *Proceedings of the 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 170-176, Peterhouse College, Cambridge, UK.
5. Axelsson, S. 2005. *Understanding intrusion detection through visualization*, PhD thesis, Chalmers University of Technology, Göteborg, Sweden.
6. Baldwin. J. and Bottoms A. E. (eds.). 1976. *The Urban Criminal: A Study in Sheffield*. London, UK: Tavistock Publications.
7. Becker, G. 1968. Crime and punishment: an economic approach. *The Journal of Political Economy*, vol. 76, pp. 169–217.
8. Blumstein, A. and Wallman, J. 2000. *The Crime Drop in America,* Cambridge: Cambridge University Press.
9. Bottoms, A. E. 1976. Criminology and urban sociology, in J. Baldwin and A. E. Bottoms (eds.), *The Urban Criminal: A Study in Sheffield.*

Tavistock Publications, pp. 1-35.

10. Braithwaite J, Chapman B, Kapuscinski C. A. 1992. *Unemployment and crime: towards resolving the paradox*. Canberra, Australia: Australian National University.
11. Brehon, D. J. 2007. *Essays on the Economics and Econometrics of Urban Crime and House Price Prediction*, Department of Economics PhD thesis, Columbia University, US.
12. Brockett, P. L., Xia, X. and Derrig, R. A. 1998. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. *The Journal of Risk and Insurance*, vol. 65, no. 2, pp. 245-274.
13. Bureau of Justice Statistics. 2012. Reported crime in the United States 1960-2007. Retrieved 10 August, 2012, from http://bjsdata.ojp.usdoj.gov/dataonline/Search/Crime/State/StatebyState.cfm?NoVariables=Y&CFID=350216&CFTOKEN=91023531
14. Burges, C. J. C. 1998. A tutorial on support vector machiens for pattern recognition, *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167.
15. Burns, R. and Burns, R. 2008. Business Research Methods and Statistics using SPSS, UK, London: Sage.
16. Butts, C. O., Stefano, G. B., Fricchione, G., and Salamon, E. 2003. Religion and its effects on crime and delinquency. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, vol. 9, no. 8, pp. 79-82.
17. Chiricos, T., and Waldo, G. 1970. Punishment and crime: an examination of some empirical evidence. *Social Problems*, vol. 18, pp. 200-217.
18. Chung, W., Chen, H., Chaboya, L. G., O'Toole, C. D. and Atabakhsh, H. 2005. Evaluating event visualization: a usability study of COPLINK spatio-temporal visualizer. *International Journal of Human-Computer Studies*, vol. 62, no. 1, pp. 127-157.
19. Cieslak, D. A. and Chawla, N. C. 2007. Detecting fractures in

classifier performance, in *Proceedings of the Seventh IEEE International Conference on Data Mining*, IEEE Computer Society, pp. 123-132.

20. Clinard M. B. 1958. *Sociology of Deviant Behaviour*. New York, USA: Rinehart & Company.
21. Conklin, J. E. 2003. *Why Crime Rates Fell*, Boston, MA: Pearson Education.
22. Cortes , C., and Vapnik, V. 1995. Support-vector networks, *Machine Learning*, vol. 20, no. 3, pp. 273-297.
23. Cottrell, M., and Verleysen, M. 2006. Advances in self-organizing map, in M. Cottrell and M. Verleysen (eds.), Neural Networks 2006 Special Issue "*Advances in Self-Organizing Maps-WSOM 05*", Elsevier, doi:10.1016/j.neunet.2006.05.011, pp. 721–722.
24. Cressey D. R. 1964. *Delinquency, crime and differential association*. Hague, the Netherlands: Nijhoff.
25. Dahmane, M. and Meunier, J. 2005. Real-time video surveillance with self-organizing maps, in *Proceedings of the Second Canadian Conference on Computer and Robot Vision* (CRV'05), Washington, DC., pp. 136-143.
26. Deboeck, G. 2000. Self-organizing patterns in world poverty using multiple indicators of poverty repression and corruption. *Neural Network World*, vol. 10, pp. 239-254.
27. Dittenbach, M., Merkl, D. and Rauber, A. 2000. The growing hierarchical self-organizing map, in *Proceedings of International Joint Conference on Neural Networks* (IJCNN 2000), Como, Italy, vol. 6, pp. 15-19.
28. Eide, R., Rubin, P. H. and Shepherd, J. M. 2006. *Economics of Crime*, Hanover, US: Now Publishers, Inc.
29. Fajnzylber, P., Lederman, D., and Loayza, N. 1998. *Determinants of Crime Rates in Latin America and the World* (World Bank Latin American and Caribbean Studies Working Paper Series).

30. Farrington, D. P. 1986. Age and crime, in Tonry, M. and Morris, N. (Eds) *Crime and Justice*, vol. 7, pp. 189–250. Chicago: University of Chicago.

31. Fei, B. K., Eloff, J. H., Olivier, M. S. and Venter, H. S. 2006. The use of self-organizing maps for anomalous behavior detection in a digital investigation. *Forensic Science International*, vol. 162, no. 1-3, pp. 33-37.

32. Fei, B., Eloff, J., Venter, H. and Olivier, M. 2005. Exploring data generated by computer forensic tools with self-organising maps, in *Proceedings of the IFIP Working Group 11.9 on Digital Forensics,* pp. 1-15.

33. Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, vol. 7, pp. 179–188.

34. Freeman, R. B. 1996. Why do so many young American men commit crimes and what might we do about it? Journal of Economic Perspectives, vol. 10, no. 1, pp. 25-42.

35. Gibbs, J. P. 1968. Crime, punishment and deterrence. *South-Western Social Science Quarterly*, vol. 48, pp. 515-530.

36. Giudici, P. 2003. *Applied Data Mining - Statistical Methods for Business and Industry.* England: John Wiley & Sons.

37. Gottfredson, D. C. 1985. *School Size and School Disorder*. Baltimore, MD: Centre for Social Organization of Schools, Johns Hopkins University, US.

38. Gould, E. D., Weinberg, B. A. and Mustard, D. B., 2000. *Crime Rates and Local Labour Market Opportunities in the United States: 1979-1997*. University of Georgia Working Paper, US.

39. Grogger, J. 1998. Market wages and youth crime. *Journal of Labour Economics*, vol. 16, issue 4, pp. 756-791.

40. Grosser, H., Britos, P. and García-Martínez, R. 2005. Detecting fraud in mobile telephony using neural networks, in Ali, M. and Esposito F. (Eds.): *IEA/AIE 2005*, *Lecture Notes in Artificial Intelligence*, Berlin,

Germany: Springer-Verlag, vol. 3533, pp. 613–615.

41. Gyimah-Brempong, K. 2001. Alcohol availability and crime: evidence from census tract data. *Southern Economic Journal,* vol. 68, No. 1, pp. 2-21.

42. Hama, A. 2002. Demographic change and social breakdown: the role of intelligence. *Mankind Quarterly*, vol. 42, no. 3, pp. 267-282.

43. Harries K. 2006. Property crimes and violence in United States: an analysis of the influence of population density. *International Journal of Criminal Justice Siences*, vol. 1, no. 2, pp. 24-34.

44. Hirschi, T., and Gottfredson, M. (eds.). 1980. *Understanding Crime*, Beverly Hills: Sage.

45. Hollmén, J. 2000. *User Profiling and Classification for Fraud Detection in Mobile Communications Networks*, PhD thesis, Helsinki University of Technology, Finland.

46. Hollmen, J. 1996. Process modeling using the self-organizing map. Retrieved 28 April, 2011 from http://users.ics.tkk.fi/jhollmen/dippa/node26.html#SECTION00524300000000000000

47. Hollmén, J., Tresp, V. and Simula, O. 1999. A self-organizing map for clustering probabilistic models. *Artificial Neural Networks*, vol. 470, pp. 946-951.

48. Honkela, T. 2010. Directions for e-science and science 2.0 in human and social sciences, in Proceedings of MASHS 2010, Computational Methods for Modeling and Learning in Social and Human Sciences, pages 119–134. Multiprint.

49. Huysmans, J. et al. 2006. Country corruption analysis with self-organizing maps and support vector machines, in H. Chen et al. (Eds.): Intelligence and Security Informatics, International Workshop (*WISI) 2006, Lecture Notes in Computer Science 3917*, Singapore, pp. 104-114.

50. İmrohoroğlu, A., Merlo, A. and Rupert, P. 2000. On the political

economy of income redistribution and crime. *International Economic Review*, vol. 41, no. 1, pp. 1-25.

51. Joutsijoki, H., and Juhola, M. 2011. Comparing the one-vs-one and one-vs-all methods in benthic macroinvertebrate image classification, in P. Perner (ed.), *Lecture Notes in Artificial Intelligence*, Berlin, Germany: Springer-Verlag, vol. 6871, pp. 399-413.

52. Joutsijoki, H., and Juhola, M. 2013. Kernel selection in multi-class support vector machines and its consequence to the number of ties in majority voting method. Accepted to *Artificial Intelligence Review*, vol. 40, pp. 213-230. DOI: 10.1007/s10462-011-9281-3.

53. Juhola, K., and Juhola, M. 1996. Malthusian parameter on the Finnish population in the 20th century. *International Journal of Bio-Medical Computing*, vol. 41 (1996), pp. 5-11.

54. Juhola, M., and Siermala, M. 2012a. A scatter method for data and variable importance evaluation. *Integrated Computer-Aided Engineering,* vol. 19, no. 2, pp. 137-149.

55. Juhola, M., and Siermala, M. 2012b. ScatterCounter software via link: http://www.uta.fi/sis/cis/research_groups/darg/publications.html.

56. Kangas, L. J. 2001. *Artificial neural network system for classification of offenders in murder and rape cases*, The National Institute of Justice, Finland.

57. Kaski, S., Kangas, J., and Kohonen, T. 1998. Bibliography of Self-Organizing Map (SOM) Papers: 1981-1997, Neural Computing Surveys, vol. 1, pp. 102-350.

58. Koenig S 1962. The immigrants and crime, in Roucek J S ed. *Sociology of Crime*. London, UK: Peter Owen Ltd. pp. 138-159

59. Kohonen, T. 1990. The Self Organizing Map, in *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464-1480.

60. Kohonen, T. 1997. *Self-Organizing Maps*, Berlin, Heidelberg, New York: Springer-Verlag.

61. Kohonen, T. and Honkela, T. 2007. Kohonen network. *Scholarpedia*, vol. 2, no. 1, p. 1568.

62. Kohonen, T., Oja, M., Kaski, S., and Somervuo, P. Self-organizing map, in K. Puolamäki and L. Koivisto (eds) 2002. *Laboratory of Computer and Information Science Neural Network Research Centre Biennial Report*, pp. 111-116.

63. Kruskal, W. H. and Wallis, W. A. 1952. Use of ranks in one-criterion variance analysis, Journal of the American Statistical Association, vol. 47, no. 260, pp. 583-621.

64. Lampinen, T., Koivisto, H., and Honkanen, T. 2005. Profiling network applications with fuzzy c-means and self-organizing ,aps. *Classification and Clustering for Knowledge Discovery*, vol. 4, pp. 15-27.

65. Lee, S.-C., and Huang, M.-J. 2002. Applying AI technology and rough set theory for mining association rules to support crime management and fire-fighting resources allocation. *Journal of Information, Technology and Society*, no. 2, p. 65.

66. Lemaire, V. and Clérot, F. 2005. The many faces of a Kohonen map - a case study: SOM-based clustering for on-line fraud behaviour classification. *Classification and Clustering for Knowledge Discovery*, vol. 4, pp.1-13.

67. Leufven, C. 2006. *Detecting SSH identity theft in HPC cluster environments using self-organizing maps*, Master's thesis, Linköping University, Sweden.

68. Levinson, D. (ed.) 2002. *Encyclopaedia of Crime and Punishment*, Thousand Oaks, CA: SAGE Publications, Inc.

69. Li S.-T., Tsai F.-C., Kuo S.-C., Cheng Y.-C. 2006. A knowledge discovery approach to supporting crime prevention, in *Proceedings of the Joint Conference on Information Sciences 2006*, Taiwan, doi:10.2991/jcis.2006.146.

70. Li, X. 2008. *Cybercrime and Deterrence: Networking Legal Systems*

in the Networked Information Society, doctoral dissertation. Finland, Turku: Faculty of Law, University of Turku.

71. Lochner, L. 2007. Education and crime. Retrieved June 6, 2009 from http://economics.uwo.ca/faculty/lochner/papers/educationandcrime.pdf
72. Logan, C. H. 1971, On punishment and crime (Chiricos and Waldo, 1970): some methodological commentary. *Social Problems*, vol. 19, pp. 280 - 289.
73. Logan, C. H. 1975. Arrest rates and deterrence. *Social Science Quarterly*, vol. 56, pp. 376-389.
74. Machin S., and Meghir, C. 2000. *Crime and Economic Incentives*, Institute for Fiscal Studies Working Papers W00/17, London, UK: Institute for Fiscal Studies, London School of Economics.
75. MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations, in *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, University of California Press, pp. 281-297.
76. Mann, H. B. and Whitney, D. R. 1947. On a test of whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60.
77. Masahiro, T. 1996. Economic structure and crime: the case of Japan. *Journal of Social-Economics*, vol. 5, no. 4, pp. 497-515.
78. McGuire M. 2005. Effects of density on crime rates in U.S. cities: a modern test of classic Durkheimian theory. Retrieved July 10, 2012, from http://www.sociology.uiowa.edu/mbmcguir/capstone/TRSresearch.doc
79. Mehmood, Y., Abbas, M., Chen, X., and Honkela, T. 2011. Self-organizing maps of nutrition, lifestyle and health situation in the world, in *Advances in Self-Organizing Maps - Proceedings of WSOM 2011, 8th International Workshop*, Springer, pp. 160–167.
80. Memon Q. A, Mehboob S. 2006. Crime investigation and analysis

using neural nets, in *Proceedings of International Joint Conference on Neural Networks 2006,* pp 346-350.

81. Mena, J. 2003. *Investigative Data Mining for Security and Criminal Detection,* UK: Butterworth-Heinemann.

82. Mower, E. R. 1942. *Disorganization, Personal and Social*, Philadelphia: J. B. Lippincott Company.

83. Niemelä, P. and Honkela, T. 2009. Analysis of parliamentary election results and socio-economic situation using self-organizing map, in *Proceedings of 7th International Workshop on Self-Organizing Maps (WSOM 2009)* June 8-10, 2009, St. Augustine, Florida, USA, pp. 209–218.

84. Oatley, G. C., Ewart, B. W., Zeleznikow, J. 2006. Decision support systems for police: lessons from the application of data mining techniques to "soft" forensic evidence. *Artificial Intelligence and Law*, vol. 14, no. 1, pp. 35-100.

85. Oja, M., Kaski, S., and Kohonen, T. 2003. Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum, *Neural Computing Surveys*, vol. 3, pp. 1-156.

86. Ollikainen, J. and Juhola, M. 2008. On comparison methods of identifiers for DNA investigations in the context of crimes and accidents, *Intelligent Data Analysis*, vol. 12, no. 4, pp. 409-423.

87. Orrenius P. M., and Coronado, R. 2005. *The Effect of Illegal Immigration and Border Enforcement on Crime Rates along the U.S.-Mexico Border* (Working Paper 131), Federal Reserve Bank of Dallas.

88. Palmer, A., Jiménez, R., and Gervilla, E. 2011. Data mining: machine learning and statistical techniques, in Kimito Funatsu (ed.), *Knowledge-Oriented Applications in Data Mining*, Rijeka, Crotia – Shanghai, China: InTech, pp. 373-396.

89. Pöllä, M., Honkela, T., and Kohonen, T. 2009. *Bibliography of Self-Organizing Map (SOM) Papers: 2002-2005 Addendum*. TKK Reports in Information and Computer Science, Helsinki University of

Technology, Report TKK-ICS-R23.

90. Puolamäki, K. and Koivisto, L. (eds). 2002. *Biennial Report 2002-2003*. Laboratory of Computer and Information Science Neural Network Research Centre. Helsinki, Finland: Helsinki University of Technology.

91. Quinney, R. 1971. Crime: phenomenon, problem and subject of study, in Erwin O. Smiegel (ed.) *Handbook on the Study of Social Problems*, Rand McNally, pp. 209-246.

92. Rhodes, B., Mahaffey, J., and Cannady, J. 2000. Multiple self-organizing maps for intrusion detection, in *Proceedings of the 23rd National Information Systems Security Conference*, October 16-19, 2000, Baltimore, Maryland, USA. Accessed November 12, 2013 from http://csrc.nist.gov/nissc/2000/proceedings/papers/045.pdf

93. Rock, R. 1994. *History of Criminology*. Aldershot, UK: Dartmouth Publishing.

94. Rokach, L. and Maimon, O. 2008. Data Mining with Decision Trees - Theory and Applications, Singapore: World Scientific.

95. Roucek, J. S. (ed). 1962. *Sociology of Crime*, London, UK: Peter Owen Ltd.

96. Soares, R. R. 2004. Development, crime and punishment: Accounting for the international differences in crime rates. *Journal of Development Economics, vol.* 73, pp. 155– 184.

97. South African Human Rights Commission. 2007. *Crime and Its Impact on Human Rights: Ten Years of the Bill of Rights* (Crime Conference Report (Conference held on 22 to 23 March 2007), South African Human Rights Commission.

98. South S. J. and Messner, S. F. 2000. Crime and demography: multi linkages, reciprocal relations. *Annual Review of Sociology*, vol. 26, pp. 83-106.

99. Stark, R., Doyle D. P. and Kent, L. 1980. Rediscovering moral communities: church membership and crime, in T. Hirschi and M.

Gottfredson (eds), *Understanding Crime*, pp. 43-52, Beverly Hills: Sage.

100. Stucky, T. D. 2005. *Urban Politics, Crime Rates, and Police Strength*. El Paso, Texas: LFB Scholarly Publishing.

101. Suh, Sang C. 2012. *Practical Applications of Data Mining*, Burlington, MA: Jones & Bartlett Learning, LLC.

102. Sumner, C. 2005. The social nature of crime and deviance, in Colin Sumner (ed.), *The Blackwell Companion to Criminology*, Hoboken, New Jersey, US: Wiley-Blackwell, pp. 3-31.

103. Sutherland, E. and Cressey, D. R. 1966. *Principles of Criminology*. Seventh Edition. Philadelphia: Lippincott.

104. Taft, D. R. 1950. *Criminology: A Cultural Interpretation*, revised edition, London, UK: The MacMillan Company.

105. The Community Safety and Crime Prevention Council. 1996. *The root causes of crime - The Community Safety and Crime Prevention Council statement on the root causes of crime*. Waterloo, Canada: The Community Safety and Crime Prevention Council.

106. The United States Department of Justice, Bureau of Justice Statistics, Homicide trends in the U.S., 11 July, 2007, http://www.ojp.usdoj.gov/bjs/homicide/hmrt.htm

107. Thio, S. 1978. *Deviant Behaviour*, Boston, US: Houghton Mifflin Company.

108. Tittle, C. R. 1969. Crime rates and legal sanctions. *Social Problems*, vol. 16, pp. 409-423.

109. Tittle, C. R., and A. R. Rowe. 1974. Certainty of arrest and crime rates: a further test of the deterrence hypothesis. *Social Forces*, vol. 52, pp. 455-462.

110. Vapnik, V. N. 2000. *The Nature of Statistical Learning Theory*, New York, USA: Springer-Verlag.

111. Viscovery Software GmbH. 2013. Viscovery SOMine, retrieved

January 29, 2013, http://www.viscovery.net/somine/

112. Viscusi, K. 1986. Market incentives for criminal behaviour, Chapter 8, in R. Freeman, and H. Holzer (eds.), *The Black Youth Employment Crisis*. Chicago, US: University of Chicago Press.

113. Wadsworth, T. 2010. Is immigration responsible for the crime drop? An assessment of the influence of immigration on changes in violent crime between 1990 and 2000. *Social Science Quarterly*, vol. 91, no. 2, pp. 531-553.

114. Wirth, L. 1938. Urbanism as a way of life. *The American Journal of Sociology*, vol. 44, no. 1, pp. 1-24.

115. Witte, A.D., Tauchen, H. 1994. *Work and Crime: An Exploration Using Panel Data*, NBER Working Paper 4797, Cambridge, MA, US: National Bureau of Economic Research.

116. Yang, Shu-O W.; Phillips, G. Howard. 1974. *An Ecological Study of Crime in Rural Ohio*. Ohio, US: Ohio Farm Bureau Federation.

117. Zaslavsky, V. and Strizhak, A. 2006. Credit card fraud detection using self-organizing maps. *Information and Security: An International Journal*, vol.18, pp. 48-63.

118. Zimring F. E. 2007. *The Great American Crime Decline,* New York: Oxford University Press.

PUBLICATION I

Crime and its social context: analysis using the self-organizing map

Xingan Li and Martti Juhola

Crime and Its Social Context: Analysis Using the Self-Organizing Map

Xingan Li* and Martti Juhola
Computer Science, School of Information Sciences
University of Tampere
33014 Tampere, Finland
E-mail: Xingan.Li@uta.fi E-mail: Martti.Juhola@sis.uta.fi
*Corresponding author; tel. +358 45 8651215, fax +358 3 2191001

***Abstract*—Data mining and visualization techniques show their value in various domains but have not been broadly applied to the study of crime, which is in demand of an instrument to efficiently and effectively analyze available data. The purpose of this study is to apply the Self-Orgamizing Map (SOM) to mapping countries with different situations of socio-economic development. Supplemented by other methods, including ScatterCounter for attribute selection, and nearest neighbor search, discriminant analysis and decision trees for obtaining comparable results, the SOM is found to be a useful tool for mapping criminal phenomena through processing of multivariate data.**

Keywords—self-organizing map; nearest neighbor search; discriminant analysis; decision trees; crime situation

I. INTRODUCTION

The study of crime is expected to exercise control over existing crime as well as to overcome future occurrences. On the background of criminal phenomena there are multiple socio-economic conditions. Regardless of its complexity in preventing crime, it is important to have an understanding of its roots [1]. However, no solely workable theory has thus far been invented to provide any precise answer for tackling crime, though many theorists presented many persuasive suggestions [2]. What is special is that the study of crime deals with a social phenomenon that hardly has a perfect solution, which is not sought in this study either. Instead, it is to test a new method for identifying factors that are important in seeking potential ways out.

Data mining and visualizing techniques have shown their practical value in various domains but have not been extensively studied for application in crime analysis. One of such tools, the self-organizing map, uses an unsupervised learning method to group data according to patterns found in a dataset, making it an ideal tool for data exploration. This is an area in which innovative studies can be carried out.

Currently, criminological research in detailed offences from micro viewpoints has also been acquiring more assistance from application of artificial intelligence. Hitherto, many researchers focus on applications of artificial neural networks to law enforcement, in particular, the detection of specific abnormal or criminal behaviors.

There is a general lack of research on macroscopic aspects of criminal phenomena as related to other social factors. The current situation created a motivation for designing experiments exploiting this approach, compared with and supplemented by other methods.

This paper represents efforts to innovatively apply the SOM to the study of crime. It revolves around some significant categories of offences and their social context.

II. METHODOLOGY

Developed by Kohonen [4] to cluster and visualize data, the SOM is an unsupervised learning mechanism that clusters objects having 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 can be generated using software packages. By observing and comparing the clustering map and feature planes, rough correlation between different indicators (attributes) can be identified. A detailed correlation table can also be realized automatically with Viscovery SOMine [5], which adopts the correlation coefficient scale ranging from -1.0 to +1.0. These clustering maps, feature planes and correlation tables provide basis for further analysis.

During the application of the SOM, in order to select variables, ScatterCounter [6] [7] will be used to identify the strength of attributes. The weak ones will be removed from the dataset and the reduced dataset will be used in final processing and analysis.

Besides the SOM, nearest neighbor search, discriminant analysis and decision trees will be used to validate the clusters and analysis by calculating how accurately these methods put the same countries into the same clusters as the SOM does.

III. DESIGN OF EXPERIMENTS

A. Countries included

Fifty countries were included in the experiment, coded in Table I. These codes will be shown in the maps as "labels". These countries were selected based on the availability of data on selected attributes. Generally, missing values of different attributes of one single country were kept to fewer than 10%.

978-0-7695-5062-6/13 $26.00 © 2013 IEEE
DOI 10.1109/EISIC.2013.26

B. Crime and socio-economic factor

This study contains 44 attributes (crime and socio-economic factors). An overview of all attributes that were used in this study is given in Table II. Thirty-one of them are socio-economic factors, while the rest thirteen are crime-related indicators.

The purpose of current study was to map the contemporary crime situation of countries. It required proper information to be current. However, statistics of crime at international level prove to be a long-term task. If only concurrent data were used, the number of missing data would be too great, and the current study is simply impossible. Fortunately, many of such factors, especially when they are compared internationally, have relative stability over years. As a result, in this original dataset, total missing values account for 5.0% (see Table II).

C. Pre-processing and attribute selection

The dataset was processed by Viscovery SOMine to generate clusters. Upon initial clustering, the dataset was processed with ScatterCounter [6] [7]. In using ScatterCounter, missing values were replaced with medians of pertinent clusters, because this software package will not deal with missing values. In addition, countries were labeled by cluster identifiers given by the SOM.

The objective of ScatterCounter is to evaluate how many classes (named clusters in the SOM) differ from each other. Its principle is to start from a random instance of a dataset and to traverse all instances by searching for the nearest neighbor of the current instance, then to update the one found to be the current instance, and iterate the whole dataset. During the process of the search, every change from one class to another is counted. The more the class changes, the more the classes of a dataset are overlapped.

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 randomized steps, it is repeated from 5 to 10 times, and the average is used for separation power.

Separation powers can be calculated for the whole data, or for every class and for every attribute [6] [7]. Absolute values of separation powers are from [0,1]. They are usually positive, but small negative values are also possible when in some classes an attribute does not virtually separate at all.

TABLE I. COUNTRIES INCLUDED

Australia	AU	Latvia	LV	Ireland	IE
Azerbaijan	AZ	Moldova	MD	India	IN
Bulgaria	BG	Mauritius	MU	Iceland	IS
Belarus	BY	Mexico	MX	Italy	IT
Canada	CA	Netherlands	NL	Jamaica	JM
Switzerland	CH	Norway	NO	Romania	RO
Chile	CL	New Zealand	NZ	Russia	RU
Colombia	CO	Poland	PL	Slovenia	SI
Costa Rica	CR	Portugal	PT	Slovakia	SK
Czech	CZ	Spain	ES	Thailand	TH
Germany	DE	Finland	FI	Turkey	TR
Denmark	DK	France	FR	Ukraine	UA
Estonia	EE	UK	GB	US	US
Japan	JP	Georgia	GE	Uruguay	UY
Korea	KR	Greece	GR	South Africa	ZA
Kazakhstan	KZ	Hungary	HU	Zambia	ZM
Lithuania	LT	Indonesia	ID		

Although an attribute may have separation power around zero for some clusters, it may have larger separation power for at least one cluster and thus thought to be useful. With these results and observations, five variables (9, 21, 25, 33, 34) have poor separation powers and are removed from the dataset used in the following experiments and analysis.

TABLE II. COUNTRY SOCIO-ECONOMIC SITUATION BY 44 ATTRIBUTES WITH THEIR MEANS, STANDARD DEVIATIONS AND MISSING VALUES IN PERCENT

1. Cellular subscribers per 1000 people	16. Researcher in RD per 100,000 people	31. Expenditure on health, public (% of GDP) (%)
2. Electricity consumption per capita	17. Phone mainlines per 1000 people	32. Mean years of schooling (of adults) (years)
3. Electrification rate %	18. Prisoners per 100,000 people	33. Military expenditure (% of GDP) (dropped)
4. Employment in agriculture %	19. Prison capacity filled per cent	34. Net migration rate (per 1,000 people) (dropped)
5. Employment in industry %	20. Rapes per 100,000 people	35. Public expenditure on education (% of GDP) (%)
6. Employment in services %	21. Robberies per 100,000 people (dropped)	36. Employment to population ratio, population 25+ (% aged 25 and above)
7. Exports of goods and services % of GDP	22. Software piracy rate %	37. Gallup: Trust in other people (% answering "'yes'" to having the element)
8. Foreign direct investment net inflows % of GDP	23. Total crimes per 100,000 people	38. Homicide rate (per 100,000)
9. Forest area % (dropped)	24. Police per 100,000 people	39. Human Development Index (HDI) value
10. GDP per capita annual growth rate %	25. Jails per 100,000 people (dropped)	40. Labour force participation rate, female-male ratio
11. GDP per capita USD	26. Frauds per 100,000 people	41. Suicide rate: female (per 100,000)
12.GDP per capita PPP USD	27. Convicted per 100,000 people	42. Suicide rate: male (per 100,000)
13. Imports of goods and services % GDP	28. Assaults per 100,000 people	43. Total dependency ratio (per 100 people aged 15-64 years)
14. Internet users per 1000 people	29. Burglaries per 100,000 people	44. Youth Unemployment (% aged 15-24)
15. RD expenditure % of GDP	30. Consumer Price Index	

Sources for Table II: 15, 16: World Bank. http://data.worldbank.org/; 18, 19: International Centre for Prison Studies. http://www.prisonstudies.org/; 20, 21, 30-44: United Nations Office on Drugs and Crime (UNODC). https://www.unodc.org/unodc/en/data-and-analysis/; 22: Fifth Annual BSA and IDC Global Software Piracy Study, 2007; other: United Nations Development Program (UNDP). http://hdr.undp.org/en/statistics/data/

As a result, all the 39 variables preserved in the dataset have stronger separation power and supposedly enable valid clustering. In this reduced dataset, total missing values account for fewer than 5.3%. Of total 39 attributes, 22 had no missing values. The highest frequency of missing values in one attribute reached 38%.

D. Construction of the map

In this study, the software package used is Viscovery SOMine 5.2.2 Build 4241. Compared with SOM Toolbox, 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 visualisation.

The SOMine software automatically generated maps from the dataset of 50 countries and 39 attributes. The clustering map (Fig. 1) as well as some other detailed statistics, such as correlations as discussed below, can be used in further analysis.

IV. RESULTS

Upon processing of data, four clusters have been generated, each representing groups of countries sharing similar characteristics. In the SOM, values are expressed in colors: warm colors denote high values, cold colors low values.

A. Clusters

Clusters were given in Fig. 1 Cluster 1 consists of 13 countries with high total crime rates. They are characterized by a high rape rate and high convicted rate, but a low level of homicide rate. Most socio-economic indicators are located on an upper level among the sample. In these countries, there are more prisons and prison capacity is highly filled.

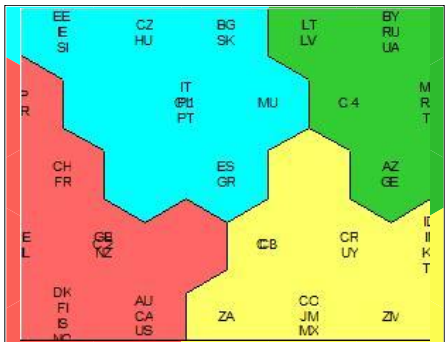


Fig. 1. Clustering map with cluster names C1 – C4 and labels of countries (C1: CZ, EE, HU, BG, LT, LV, SK, MU, IE, SI, PL, PT, ES, GR, IT; C2: RO, TR, AZ GE, CL, CO, CR, UY, ID, IN, TH, ZA, JM, MX, ZM; C3: KR, FR, JP, CH, DE, NL, GB, NZ, DK, FI, IS, NO, AU, CA, US; C4: BY, RU, UA, KZ, MD)

Cluster 2 consists of 15 countries with the highest level of the total crime rate. In these countries, according to the figures, there is a high level of prisons, but a lower level of the prisoner rate and prison capacity is less filled. These countries are characterized the by highest level of socio-economic indicators.

Cluster 3 consists of 12 countries with a low level of total crime rate, however, characterized by the highest level of the homicide rate. Some countries have a high level of some socio-economic indicators, but others not. They have a high level of the burglary rate. In these countries, there are more prisons on average, and these prisons are highly filled. Yet, in these countries, there is the lowest level of the convicted rate.

Cluster 4 consists of 10 countries both with a low level of the total crime rate. In these countries, there is the highest level of the prison capacity, which is the least filled. The rates of rape and homicide are the highest among the sample.

B. Validation of clusters

The results by the SOM were tested by nearest neighbor search also applied with Euclidean distances. All validations were performed by using the clusters in Fig. 1 as the classes of the countries. Then classifications were run based on leave-one-out, where, one by one, each single country formed a test set and other countries its training set. When k=1, 68% of the classifications were correct, i.e., agreed with the cluster labels, when the data were not scaled, and 92% when data were scaled. When k=3, the results were 56% and 92%, respectively. When k=5, they were 52% and 92%, when k=7, 62% and 92%, and when k=9, 62% and 92%. Apparently, scaling of the data significantly improved the probability of correct classifications.

In addition, logistic discriminant analysis was applied to test the SOM results, using both scaled and not scaled data. When the data were not scaled, the probability of putting into same clusters as generated by the SOM was 68%. When the data were scaled, the probability was nearly the same. In these two cases, scaling of the data made no obvious differences. Besides, linear discriminant analysis was also carried out obtaining a result of 82%, with scaling made no difference either. Correspondingly, decision trees classified 62% correctly.

C. Correlations

Viscovery SOMine generated a detailed list of correlations, based on which Table III was created. 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 using other methods.

The results showed few strong links between any pairwise attributes, but many less strong links that are interesting. Certain attributes were found to have strong negative correlation with other attributes, while most others not.

Further consideration into correlation has some importance in identifying the influential factors. Nevertheless, correlated factors might contain but are not equivalent to causes. Correlation as numbers can happen to be incorrect and it does not express the final examination of the problem. Furthermore, correlation (Pearson) is of linear type, it plays little part in revealing more complicated relations between attributes.

In a society, value judgment exists everywhere. Another paradox is that something favorable is positively associated with a criminal indicator, or vice versa. For that reason, correlation produced by statistics must be examined through supplementary analysis by adopting other methods.

TABLE III. CORRELATIONS BETWEEN SOCIO-ECONOMIC ATTRIBUTES *A* AND CRIME-RELATED ATTRIBUTES: CORRELATIONS FROM INTERVAL (-0.3,0.3) WERE LEFT OUT SEEN AS INSIGNIFICANT, FROM [0.3,0.6] WERE SEEN AS INTERESTING, AND FROM [0.6,1] MARKED IN BOLD FACE AS SIGNIFICANT.

A	Crime-Related Attributes										
	A18	A19	A20	A22	A23	A24	A26	A27	A28	A29	A38
A1	–	−0.32	–	**−0.64**	0.46	–	0.34	0.40	–	0.40	–
A2	–	−0.31	0.30	−0.56	0.51	–	0.32	0.32	0.40	0.35	−0.31
A3	–	−0.51	–	−0.45	–	–	–	0.33	–	–	−0.54
A4	–	0.35	–	**0.68**	−0.56	–	−0.38	−0.36	−0.37	−0.43	0.32
A5	–	–	–	–	–	0.46	–	–	–	–	−0.35
A6	–	−0.34	0.34	**−0.69**	**0.63**	–	0.36	0.40	0.47	0.44	–
A8	–	–	–	0.39	–	–	–	–	–	–	–
A11	–	–	–	**−0.76**	**0.60**	–	0.38	0.34	0.35	0.40	−0.39
A12	–	–	–	**−0.82**	**0.64**	–	0.42	0.39	0.39	0.44	−0.47
A13	–	–	–	–	–	0.31	–	–	–	–	–
A14	–	−0.38	–	**−0.73**	0.55	–	0.39	0.31	–	0.46	−0.34
A15	–	−0.32	–	**−0.69**	0.59	–	0.46	0.43	–	0.42	−0.41
A16	–	−0.48	–	**−0.60**	0.50	–	0.34	0.44	–	0.50	−0.34
A17	–	−0.35	–	**−0.71**	0.51	–	0.47	0.34	–	0.42	−0.53
A30	0.44	–	–	0.58	−0.52	–	−0.33	–	–	–	0.39
A31	–	–	–	**−0.71**	**0.61**	–	0.46	0.35	–	0.40	–
A32	–	−0.56	–	−0.52	0.36	–	0.33	–	–	0.36	−0.37
A35	–	−0.32	–	–	0.38	–	–	–	–	0.31	–
A37	–	–	–	−0.49	0.49	–	–	0.32	–	0.46	–
A39	–	−0.51	–	**−0.74**	0.52	–	0.41	0.36	–	0.42	−0.57
A40	–	–	–	–	–	–	–	–	–	0.37	–
A42	0.35	–	–	–	–	–	–	–	–	–	–
A43	–	0.48	–	–	–	−0.30	–	–	–	–	0.46

V. CONCLUSIONS

The self-organizing map deserves further investigation into the potentiality of establishing its status in the study of crime, particularly in clustering and visualization, as well as identifying correlation.

These tasks can all be assisted by the SOM by processing large amounts of data. Through analyzing results generated by the self-organizing map, an express sketch can be prepared for depicting socio-economic patterns of criminal phenomena of different regions automatically grouped. Some of the findings in this research were coincident with those in a conventional study, particularly, on traditionally homogenous countries. This denoted that, long-term development patterns of countries or regions could possibly affect many aspects of social life, including the occurrence of crime.

Consequently, this study was well based on the innovative application of the SOM in mapping criminal phenomena and identifying correlation factors, which have been studied significantly by conventional means but have not been effectively assisted by such method as the SOM.

By applying ScatterCounter to select attributes and refine the dataset and nearest neighbor search, discriminant analysis and decision trees to test the accuracy of clustering, findings of the study provide additional proof that the self-organizing map is an appropriate and effective instrument for research on crime. The clustering results are easily visualized and convenient to interpret, facilitating practical comparison of socio-economic factors between countries with diversified criminal phenomena.

ACKNOWLEDGMENT

The first author is grateful to Tampere Doctoral Program in Information Science and Engineering (TISE) for financial support.

REFERENCES

[1] The Community Safety and Crime Prevention Council, The root causes of crime - CS&CPC statement on the root causes of crime, 1996.

[2] P. Rock, History of Criminology, Dartmouth Publishing, Aldershot, UK, 1994.

[3] V. Zaslavsky, and A. Strizhak, "Credit card fraud detection using self-organizing maps," Information and Security: An International Journal, Vol. 18, pp. 48-63, 2006.

[4] T. Kohonen, Self-Organizing Maps, Springer-Verlag, New York, USA, 1979.

[5] Viscovery Software GmbH, Viscovery SOMine, http://www.viscovery.net/somine/, 2013.

[6] M. Juhola and M. Siermala, "A scatter method for data and variable importance evaluation," Integrated Computer-Aided Engineering, Vol. 19, pp. 137-149, 2012a.

[7] M. Juhola and M Siermala. ScatterCounter software via link: http://www.uta.fi/sis/cis/research_groups/darg/publications.html, 2012b.

PUBLICATION II

Country Crime Analysis Using the Self-Organizing Map, with Special Regard to Demographic Factors

Xingan Li and Martti Juhola

AI & Soc
DOI 10.1007/s00146-013-0441-7

ORIGINAL ARTICLE

Country crime analysis using the self-organizing map, with special regard to demographic factors

Xingan Li · Martti Juhola

Received: 4 August 2012 / Accepted: 8 January 2013

Abstract Modern research on criminal phenomena has been revolving not only around preventing existing offenses, but also around analyzing the criminal phenomena as a whole so as to overcome potential happenings of similar incidents. Criminologists and international law enforcement have been attracted to the cause of examining demographic context on which a crime is likely to arise. Traditionally, little has been explored in using demographic variables as determinants of the aggregate level of crime in the crime literature. Rapid development and ubiquitous application of information technology enables academic field to perform crime analysis using visualization techniques. Automation and networking make it available to access massive amounts of crime data, typically in the form of crime statistics. In numerous fields, studies and research have shown that visualization techniques are valuable; in crime research, nevertheless, there is a general lack of its application. In order to efficiently and effectively process crime data, criminologists and law enforcement are in demand of a more powerful tool. The self-organizing map (SOM), one of the widely used neural network algorithms, may be an appropriate technique for this application. The purpose of this study is to apply the SOM to mapping countries with different situations of crime. A total of 56 countries and 28 variables are included in the study. We found that some roughly definite patterns of crime situation can be identified in traditionally homogeneous countries. In different countries, positive correlation on crime in some countries may have negative correlation in other countries. Overall, correlation of some factors on crime can still be concluded in most groups. Results of the study prove that the SOM can be a new tool for mapping criminal phenomena through processing of large amounts of crime data.

Keywords Data mining · Self-organizing map · Crime situation

1 Introduction

Modern society has long suffered from large volume of crimes almost everywhere in the globe. Deterrence of crime has become one of the most significant global tasks, along with the critical concern for reinforcing public security. Studies and research on criminal phenomena take the responsibility not only for control of existing crime, but also for exploring the criminal phenomena as a whole so as to prevent future occurrences of similar event. Both public and private sectors attempt to improve the effectiveness of crime prevention. Abundant exploration addressing this problem has by and large made use of behavioral sciences and statistics. Finding causation between crime and other phenomena has been a pursued and difficult research objective for long in natural and social domains. Crime can be perceived as an outcome of manifold undesirable personal, social, economic, cultural, and family conditions. Regardless of its complexity, to prevent crime, it is important to have an understanding of its causes (The Community Safety and Crime Prevention Council 1996). Studies and research on criminal phenomena have been situated in a long historical background with numerous studies endeavoring to divulge reasons of crime and search for final solutions, unfortunately, only in vain. There has

X. Li · M. Juhola (✉)
Computer Science, School of Information Sciences, University of Tampere, 33014 Tampere, Finland
e-mail: Martti.Juhola@sis.uta.fi

X. Li
e-mail: Xingan.Li@uta.fi

Published online: 20 January 2013

not been one exclusively feasible theory thus far invented to supply any clear-cut response for tackling crime. This has encouraged more theorists to look for novel suggestions for the current research issue (Rock 1994). A perfect solution can hardly be invented to deal with social phenomena. The nature of crime research decides that this study cannot be designed to hunt for a new way out but to check a fresh technique for discovering factors that are important in seeking possible solutions.

Traditionally, little has been explored in using demographic variables as the determinant of the aggregate level of crime in the crime literature (Hartung and Pessoa 2007, p. 1). At the present time, identifying correlation factors of crime, comparing geographical distribution of crime in different countries, and recognizing (including but not limited to predicting) criminal tendencies are attracting more players than ever. The highest ideal for criminologists is to reveal causational and correlation factors of crime. Law enforcement wants to recognize developmental tendencies of crime. Legislators want to enact effective law to eliminate, prevent, or reduce crime. Government wants to make feasible policy to combat crime and assist victims. Victims want to be socially rehabilitated to a peaceful, safe, and harmonious life. The general public wants to create, enjoy, and maintain a society on the principle of rule by law. The international society wants to coordinate and cooperate in reckoning with transnational crime. Processing crime data has been a basis for knowledge-detection and decision-making. The difficulty in analyzing large volume of crime data posed great challenge. Comparison and analysis in traditional ways become complicated and time-consuming. Advanced analytical methods are required to extract useful information from large amount of crime data.

The traditional modes for scholars to realize their goals are through either qualitative or quantitative or combined methods. Quantitative analysis of criminal phenomena has been made possible by publicly available national statistics, with a variety of analytical instruments being employed in different fields of issues. With the construction of transparent governance, the information required for the research can generally be found in databases of official publications and the Internet, including websites of international organizations, national statistical and judicial agencies, and other official documents.

Demand for user interactive interfaces based on current technologies has been in existence to meet and accomplish the emerging responsibilities and tasks. Many scientific fields have utilized visualization techniques, except studies and research of criminal phenomena, which lack broad applications. The data mining approach has been shown to be a proactive decision support tool in predicting and preventing crime. Data mining tools make it possible to find hidden relationships in data. One of such tools, the self-organizing map, uses the unsupervised learning method to group data according to patterns found in the dataset, making it an ideal tool for data exploration. The SOM has attracted substantial research interests in a wide range of applications. Nonetheless, while many papers on the SOM have been published, very rare studies have dealt with the use of the SOM in research of criminal phenomena.

This study applies the SOM to investigate relations between demographic factors and criminal phenomena. After this brief introduction, Sect. 2 reviews applications of the SOM in social and criminal research. Section 3 deals with methodological issues of the SOM. Section 4 demonstrates the design of experiments in applying the SOM in research on demographic factors and crime, including concerned countries, demographic factors, construction of the map. Section 5 presents the results of the experiments, including the clusters generated and the correlation identified. Section 6 discusses the findings, and Sect. 7 concludes the research.

2 Application of the SOM in social and criminal research

Social problems have been an eternal topic in modern society. An attractive problem was crime appeared in many countries suffering from insecurity and unstable. Research on crime has thus been one of the most popular themes, which borrow ideas from generic or neighboring subjects. Existing endeavor for applying the SOM to social research can help frame crime research. It proved that the SOM is one of the models of neural networks that acquire growing application research on social problems. For example, in Deboeck (2000), world poverty was clustered into convergence and divergence in poverty structures based on multi-dimensions of poverty using the SOM, which reveals how new knowledge can be explored through artificial neural networks for implementing strategies for poverty reduction. Criminal phenomena have also been explored with the SOM. Huysmans et al. (2006) applied the SOM to process a cross-country database linking macro-economical variables to perceived levels of corruption with an expectation of forecasting corruption for countries. In another study, Li et al. (2006), a linguistic cluster model was developed to meet the demand of public security index and extracting relational rules of crime in time series. Compared with the previous studies in crime theory which mostly rely on traditional behavior science, they turned to a hybrid approach to overcome the obstacle of linguistic clustering in original SOM model. They analyzed the trend uncovered so as to support decision-making for planning police human resources. Another interesting research was done by Lee and Huang (2002), who made efforts to extract

association rules from a database to support allocation of resources for crime management and fire-fighting. Many of such studies found that artificial neural networks are a valuable instrument for improving studies and research of social problems.

Studies in individual offenses from micro-viewpoints have been a field gained additional support from application of artificial intelligence. Previously, numerous researchers devoted themselves to the application of artificial neural networks to law enforcement, particularly detection of specific irregular activities. Dahmane and Meunier (2005) presented an SOM application for detecting suspicious events in a scene. Oatley et al. (2006) discussed data mining and decision support technologies for police, considering the range of computer science technologies that are available to assist police activities. Adderley et al. (2006) examined how the monitoring of crime scene investigator performance can benefit from data mining techniques.

Some of the examples of researches revolved around the application of the SOM to the detection of offenses, including automobile bodily injury insurance fraud (Brockett et al. 1998), mobile communications fraud (Hollmén et al. 1999; Hollmén 2000; Grosser et al. 2005), network intrusion (Rhodes et al. 2000; Leufven 2006; Lampinen et al. 2005; Axelsson 2005), murder and rape (Kangas 2001), burglary (Adderley and Musgrove 2003; Adderley 2004), cybercrime (Fei et al. 2005, 2006), credit card fraud (Zaslavsky and Strizhak 2006), homicide (Memon and Mehboob 2006), and discovering serial criminal patterns in crime databases (Dahbur and Muscarello 2003). Abundant achievements in this specific area make it a primary field where the SOM has found applications to research related to criminal justice before.

Besides crime detection, neural networks are also found useful in researches that are specialized in victimization detection in mobile communications fraud (Hollmén et al. 1999).

Application of artificial intelligence in research of criminal phenomena can adopt abundant methods. Some can be based solely on one method, while others combined several methods in each research. Adderley and Musgrove (2003) applied three data mining techniques—the multilayer perception (MLP), radial basis function (RBF), and the SOM—to the building descriptions, modus operandi (MO), and temporal and spatial attributes of domestic and commercial burglaries attributed to a network of offenders. Lampinen et al. (2005) introduced two clustering methods, the SOM and the fuzzy c-means clustering (FCM) algorithm to be used in the analysis of network traffic. Axelsson (2005) tried four different visualization approaches, including two direct approaches and two indirect approaches, to the problem of intrusion detection. Abidogun (2005) provided a comparative analysis and application of the SOM and long short-term memory (LSTM) recurrent neural networks algorithms to user call data records in order to conduct a descriptive data mining task on users call patterns. The discussion and experimentation of Oatley et al. (2006) are even wider, including decision support techniques based on spatial statistics, and a wide range of data mining technologies.

From present literature, the SOM has been applied to the detection and identification of crimes. Applications of the SOM to crime research, that is, to identifying causative or correlative factors or to recognizing preventive or deterrent factors, have rarely been published. The current situation created a motivation for designing experiments exploiting this approach, in comparison with other methods.

3 Methodology

Until today, application of the Self-Organizing Maps to research in criminal phenomena has only limited presence. This facilitates an extensive exploration along the strings of limited achievements in existing efforts for thinking Self-Organizing methods as feasible to do research on society as a whole. Situngkir (2003) pointed that the enormous information technology provides the potentiality to establish the sociology to cope with any sociological emergence phenomena.

Recognizing that criminal justice is confronted with increasingly tremendous amount of data (for instance, in mobile communications fraud, Abidogun 2005), researchers have been aware of the necessity, possibility, and feasibility for application of the SOM to crime research. Data mining techniques in research of crime become indispensable (Chung et al. 2005). These techniques can support police activities by profiling single and series of crimes or offenders, and matching and predicting crimes (Oatley et al. 2006).

Some literature has revealed the difference between new techniques and old ones. Dittenbach et al. (2000) pointed out that, unlike traditional data mining techniques that only identify patterns in structured data, newer techniques work with both structured and unstructured data. Researchers have developed various automated data mining techniques, depending heavily on suitable unsupervised learning methods. He also concluded that cluster analysis helps a user to build a cognitive model of data, thus fostering the detection of an inherent structure and the interrelationship of data (Dittenbach et al. 2000).

Developed by Kohonen (1997) to cluster and visualize data, the SOM is an unsupervised learning mechanism that clusters objects having multi-dimension attributes into a lower-dimensional space, in which the distance between

every pair of objects captures the multi-attribute similarity between them. Based on the concept of the SOM, some applications particularly to meet the demand of law enforcement have been developed, such as Fei et al. (2005, 2006), and Lemaire and Clérot (2005). Specifically, although the data on the storage media may contain implicit knowledge that could improve the quality of decisions in an analysis, when huge volumes of data are processed, it consumes an enormous amount of time (Fei et al. 2005). The SOM's constructive role in exploratory data analysis has been confirmed in subsequent research (Lemaire and Clérot 2005).

In order to model nervous systems, in use there have been three categories of the network architectures and signal processes (Kohonen 1990, p. 1464). The first category is feed-forward networks, which transform sets of input signals into sets of output signals, usually determined by external, supervised adjustment of the system parameters. The second category is feedback networks, in which the input information defines the initial activity state of a feedback system, and after state transitions the asymptotic final state is identified as the outcome of the computation. The third category is self-organizing networks, in which neighboring cells in a neural network compete in their activities by means of mutual lateral interactions and develop adaptively into specific detectors of different signal patterns (ibid).

In the SOM, an input layer and an output layer constitute two-layer neural networks. Used in the SOM is the unsupervised learning method. The network freely organizes itself according to similarities in the data, resulting in a map containing the input data.

The SOM algorithm operates in two steps, which are initiated for each sample in the dataset. The first step is designed to find the best matching node to the input vector, which is determined using some distance function, for example, the Euclidean distance function. The least distance determines which node c is the closest of all, i.e., the best matching. Upon finding the best match, the second step is initiated, the "learning step", in which the network surrounding node c is adjusted toward the input data vector. Let index i denote a model in node i. Nodes within a specified geometric distance, h_{ci}, will activate each other and learn something from the same n-dimensional input vector $\mathbf{x}(t)$ where t denotes the iteration of learning process. The number of nodes affected depends upon the type of lattice and the neighborhood function. This learning process can be defined as (Kohonen 1997, p. 87) with n-dimensional vector:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \mathrm{h_{ci}}(\mathbf{x}(t) - \mathbf{m}_i(t)). \quad (1)$$

The function $h_{ci}(t)$ is the neighborhood of the winning neuron c and acts as the neighborhood function, a smoothing kernel defined over the lattice points. The function $h_{ci}(t)$ can be defined in two ways, either as a neighborhood set of arrays around node c or as a Gaussian function (Kohonen 1997, p. 87). In the training process, the weight vectors are mapped randomly onto a two-dimensional, hexagonal lattice. A fully trained network facilitates a number of groups.

In order to produce stable, well-oriented and topologically correct maps, some recommendations should be followed (Kohonen and Honkala 2007; 1): (1) A hexagonal grid of nodes is to be preferred for visual inspection as for the form of the array. (2) It is a useful strategy to normalize all input variables so that their variances become equal in scaling of the vector components. (3) Perform learning with a number of available training samples. (4) In order to achieve higher quality of learning, an appreciable number of random initializations of weight vector $\mathbf{m}_i(1)$ may be tried, and the map with the least quantization error selected. While these recommendations are useful as a starting point for constructing the SOM, alternatives should also be tried in different datasets and their processing to attain best results, which may still be achievable with different strategies.

In this study, in preparation for the training, programs were designed in advance, based on predefined parameters. For example, using sequential algorithms to train the SOM, the training type can be "epochs". In map structure, the reference vectors are randomly initialized. The training is based on either a bubble or a Gaussian mixture model. Either way, the data can be normalized through methods of variance or histogram. Variance is normalized to 1. Approximate histogram equalization is sometimes applied. Values are scaled for [0, 1]. Vector neighborhood radius in this study, in different experiments, is defined as 12 or 11, which generated better results than with other values, for example, smaller than 10 or bigger than 13. The map size was designed either as 8×9 or as 7×9. These map sizes can well meet the demand for demonstrating the attributes in this sample and have good visualizing effects. Training length is calculated according to the map size. Training rate is selected in the experiments as 0.5 or 0.05, separately. Certainly, these combinations of different parameters are not the only feasible values for the experiments, but they can well demonstrate the scale of the sampling and the number of attributes.

To ensure the efficiency of work, programs for the experiments were compiled, for sequential training using epochs as training type, random initiation, including both bubble and Gaussian training methods. The compilation was prepared according to the general recommendations (Kohonen and Honkala 2007: 1), considering the detailed situation of dataset used in this research.

The SOM algorithm results in a map exhibiting the clusters of data, using dark shades to demonstrate large

distances and light shades to demonstrate small distances (*U*-matrix method) (Kohonen 1997). Feature planes, which are single vector level maps, can in addition be generated to discover the characteristics of the clusters on the *U*-matrix map. They present the distribution of individual columns of data. Four examples of feature planes are demonstrated below (Fig. 1a–f), in which high values are presented by light shades, while low values are presented by dark shades.

Figure 2 demonstrates examples of clustering maps generated by different software. Figure 2a was generated by SOM toolbox for Matlab, while Fig. 2b was generated by Viscovery SOMine for the data of 28 attributes and 56 countries described below.

By observing and comparing the clustering map and feature planes, rough correlation between different indicators (attributes) can be artificially identified. However, accurate calculation can also be realized automatically with Viscovery SOMine. For example, Table 1 shows a part of correlations generated between Attribute 28 and other attributes (1–27).

Both positive and negative correlations from interval [−1, 1] are shown in Table 1. From this example, both positive and negative correlations, both strong and weak correlations can be identified. In principle, subject to positive correlation we can assume that its increase would be connected to increase of burglary, whereas subject to negative correlation its increase would be connected to decrease of burglary attribute. However, without further analysis, correlation can hardly be an indicator to what extent the relationship between two factors is significant in solving problems in reality. The self-organizing map can be a powerful tool to process data, but cannot be expected to generate further analysis.

This study applies the SOM to explore demographic factors of crime. Based on the analysis of available data, the results of the study will revolve around whether the SOM can be a feasible tool for visualizing criminal phenomena through processing of large amounts of crime data. Analysis will also be focused on the interaction between demographic factors and criminal phenomena.

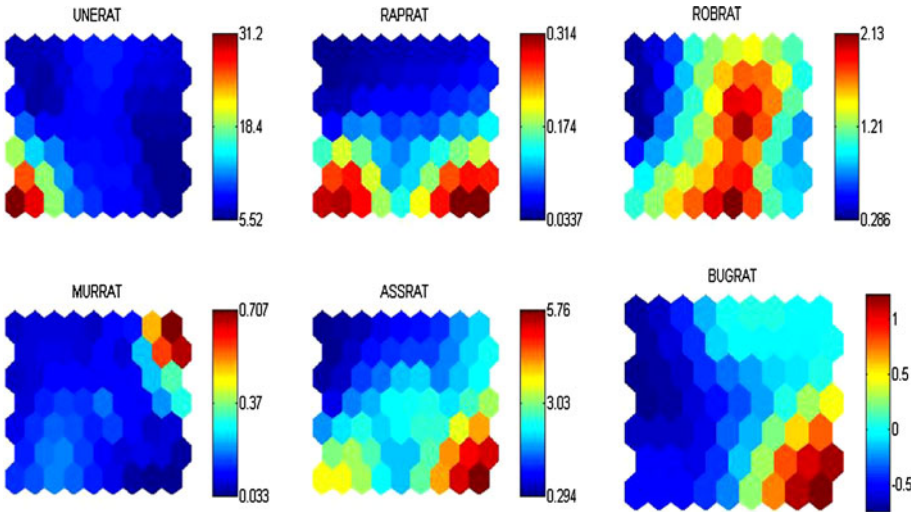


Fig. 1 **a** Unemployment rate; **b** Rape per 100,000 people; **c** Robbery per 100,000 people; **d** Murder per 100,000 people; **e** Assault per 100,000 people; and **f** Burglar per 100,000 people planes

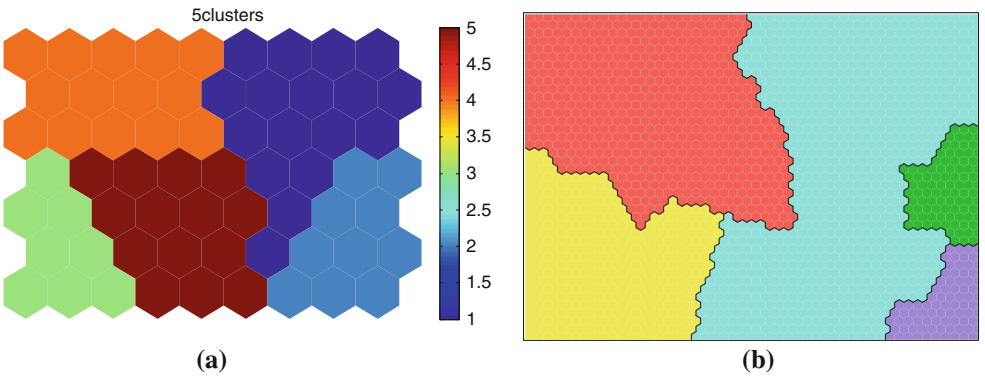


Fig. 2 Clustering maps: **a** Clustering map generated by SOM toolbox for Matlab and **b** clustering map generated by Viscovery SOMine

Table 1 An example of correlations between attribute 28, burglary per 100,000 people, and the other attributes

Other Attributes	Correlation
Adult illiteracy	−0.33
Birth rate	−0.36
Death rate	0.28
Fertility rate	−0.41
Health expenditure per capita	−0.04
Infant mortality rate	−0.15
Life expectancy	−0.04
Net migration	−0.03
Population density	−0.14
Population growth rate	−0.41
Population older than 64	0.38
Population undernourished	−0.40
Under-five mortality rate	−0.15
Unemployment rate total	−0.31
Urban population	0.20
Prisoners per Capita	0.02
Share of prison capacity filled	0.39
Rape per 100,000 people	0.06
Robbery per 100,000 people	0.05
Software piracy per 100,000 people	0.18
Total crime per 100,000 people	0.25
Police per 100,000 people	−0.13
Murder per 100,000 people	−0.10
Jails per 100,000 people	0.13
Fraud per 100,000 people	0.77
Convicted per 100,000 people	0.08
Assault per 100,000 people	0.19

4 Design of experiments

4.1 Countries included

The number of countries included in the experiment was 56. They are illustrated in Table 2. Countries were encoded according to ISO 3166-1-alpha-2 code elements, that is, each country was referred to by two letters, for example, Finland was denoted by FI, Sweden by SE, Norway by NO, and Denmark by DK, etc. These codes will be shown in the maps as labels.

4.2 Demographic factors

Demographic factors, including static, dynamic, and structural factors, have been studied since the eighteenth century (South and Messner 2000, p. 83).

Demographic factors such as age, sex, and race play an important role in understanding variation in crime rates across time and place. Demographic features of the population effect crime rates in two distinct ways. First, characteristics of population structure have compositional effects: crime rates are higher when demographic groups that have greater levels of involvement in crime constitute a larger share of the population. Second, aspects of population structure may have contextual effects on crime when they exert causal influences on criminal motivations and opportunities for crime independent of individual level for criminal tendencies. Demographic factors have been considered in relation to crime for centuries. However, demographic variables are still infrequently studied as determinants for crime in the crime literature. Some ele-

Table 2 Countries included

Country	Code	Country	Code	Country	Code	Country	Code
Australia	AU	Spain	ES	Republic of Korea	KR	Romania	RO
Azerbaijan	AZ	Finland	FI	Kazakhstan	KZ	Russian Federation	RU
Bulgaria	BG	France	FR	Lithuania	LT	Saudi Arabia	SA
Belarus	BY	United Kingdom	GB	Latvia	LV	Slovenia	SI
Canada	CA	Georgia	GE	Republic of Moldova	MD	Slovakia	SK
Switzerland	CH	Greece	GR	Mauritius	MU	Thailand	TH
Chile	CL	Hungary	HU	Mexico	MX	Turkey	TR
Colombia	CO	Indonesia	ID	The Netherlands	NL	Ukraine	UA
Costa Rica	CR	Ireland	IE	Norway	NO	United States	US
Czech Republic	CZ	India	IN	New Zealand	NZ	Uruguay	UY
Germany	DE	Iceland	IS	Papua New Guinea	PG	Yemen	YE
Denmark	DK	Italy	IT	Poland	PL	South Africa	ZA
Dominica	DM	Jamaica	JM	Portugal	PT	Zambia	ZM
Estonia	EE	Japan	JP	Qatar	QA	Zimbabwe	ZW

ments that correlate between race, sex and criminal phenomena have been challenged.

In this article, demographic factors are roughly divided into three categories: population structure, population quality, and population dynamics.

(1) Regarding population structure, three rates are selected, including population older than 64 (years old), unemployment rate, and urban population. Crime has been believed a youth's cause. People under the age of 64 committed absolute majority of crimes. With the increase of age, crime decreases.

A consensus has been reached that unemployment causes crime, though the explanations on the reason why unemployment causes crime differ one theory from another. Many studies show a strong relationship between unemployment and crime and giving explanation based on the debilitating effects of powerlessness, alienation, absence of stake in conformity, lower class pathology, culture poverty, relative deprivation, wasted human capital, the negative effects of labeling, bad schools, blocked legitimate opportunities, and illegitimate opportunity structures in areas with high unemployment (Braithwaite et al. 1992). However, high unemployment rate will reduce some offenses such as burglary. Increased unemployed population means decreased vacant houses during routine work time, improved home deterrence, and enhanced neighborhood supervision, etc. In worse economic countries, people also have decreased presence in public places such as supermarkets, bars, transportation centers, and entertainment places. In families with unemployed members, potential monetary losses in crime will also be decreased.

City living has characterized some areas for centuries, but has spread with such acceleration over the past century as to encompass hundreds of millions of people (Clinard 1958, p. 54). For centuries, writers have been concerned about the debauchery and moral conditions of the cities and have generally praised rural life. Delinquency and crime rates are generally much higher in urban areas than in rural (Clinard 1958, p. 68). Urbanism with its mobility, impersonality, individualism, materialism, norm and role conflict, and rapid social change appears to increase the incidence of deviant behavior (Clinard 1958, p. 89). Crime is largely an urban phenomenon (Bottoms 1976, p. 1). Statistics from many countries, and in many periods of time, indicate that urban areas have higher crime rates than rural areas (Cressey 1964, p. 61). Louis Wirth (1938) in his classic article on urbanization took the three concepts of size, density, and heterogeneity as key features from which one could analyze social action and organization in cities. The rates for certain forms of deviant behavior generally increase with the size of the city (p. 90).

Urbanization and labor mobility leads to increased numbers of strangers. Traditional intimate relationship between neighborhoods has been superseded. It is so that criminologists found that less severe the bodily harm inflicted on the victim, the greater the likelihood of the crime being committed by a stranger (Thio 1978, p. 99). In other words, there are possibly more numbers of crimes in most urbanized countries, but these crimes are possibly less severe, while in less urbanized countries, there is a lower number of crimes, but these crimes are possibly more severe.

Classical sociologists, such as Emile Durkheim and other functional theorists, hypothesize that areas of high density and the individuals there will experience disorganization or transformation in social order as a result of high density, resulting in increased clash between individuals and increased crime rates in these areas (McGuire 2005, p. 1). Comparatively, population density has a very small correlation to crime rate in the present study, and the correlation is not significant. The lack of a strong (or even moderate) density association to crime is contradictory to Durkheimian/functional theory (McGuire 2005, p. 11).

(2) Regarding population quality, the following factors are taken into account: adult illiteracy, health expenditure per capita, infant mortality rate, life expectancy, population growth rate, population undernourished, and under-five mortality rate.
(3) Regarding population dynamics, factors such as birth rate, death rate, fertility rate, net migration, and population density are selected.

Birth rate, death rate, fertility rate, net migration, and population density reflect a dynamic process of population change. Net increase of population increases population density. "The role of population density as a generator or inhibitor of crime has been the subject of research and debate for decades. Some studies asserted that crime is promoted by high densities; while others suggested that there is a significant surveillance effect inhibiting crime" (Harries 2006).

In many countries, people have habitually been quick to blame the influx of immigrants for mounting crime rates. Historically, among the most popular reasons given for the wide-spread existence of crime and delinquency in the United States is the settling of large numbers of immigrants (Koenig 1962, p. 138). However, criminologists have found a fact contrary to the hypothesis was that far from bringing with them criminal behavior, most immigrants come with a respect for law and authority, which they acquired in their home countries; they come mostly from stable, homogeneous societies which exerted strict control over the behavior of individuals (ibid., p. 142).

In PPIC (2008), findings revealed that high immigrant population does not appeal to be associated with high crime

rates. National studies have examined crime rates in jurisdictions with large and/or increasing immigrant populations and have found either no discernible link or a slightly negative one. A study of California cities with large populations of recently arrived immigrants showed no significant relationship between immigrant inflows and property crimes, and even a negative relationship with violent crime rates (PPIC 2008, p. 1).

Crime of immigrants has been an attractive subject matter for centuries. The immense labor relocation, globalization of labor markets, and growth of tourism pose severe questions regarding the validity or applicability of the national or moral foundation of laws and blur the dissimilarity between crime and rights, deviance, and cultural diversity (Sumner 2005, p. 8). In America, it has long been found that there is no definite race factor involved in crimes committed by immigrants or by their children. Immigrants overall are no more criminal than natives overall (Taft 1950, p. 118). On the contrary, some positive effects have been identified in previous studies: the coming of people with different cultures has kept American culture fluid. It has compelled people to rethink their mores. Studies on the immigration to the United States have proven that people of different cultures can live together and make joint contribution to human welfare (Taft 1950, p. 119).

Immigrants, from the time they have begun to arrive in a considerable number, have been blamed for all sorts of social ills, not least of which is crime. While some investigations showed a great predominance of crime and vice among immigrants, others drew the conclusion that there was no proof for the conjecture that immigration brought about an increase in crime unbalanced to the increase in the adult population (Koenig 1962, p. 140). Criminologists have found that distant from bringing them criminal behavior, most immigrants do not lack a respect for law and authority which they acquired in their home countries, and they come principally from established, homogeneous societies which extend strict control over the activities of individuals (Koenig 1962, p. 142).

As a result although the idea that immigration boosts crime rates has traditionally occupied a significant position in criminological theory and has been fundamental to the public and political discourses and disputes on immigration policy, in contradiction of the widespread response, some scholars have questioned whether the growth in immigration between 1990 and 2000 may have in reality been responsible for part of the national decrease in crime during the 1990s (Wadsworth 2010, p. 531). The association between immigration and crime, while remaining strong in the public perception for over a century, has by no means obtained steady empirical support (Wadsworth 2010, p. 532). The findings offer insights into the multifaceted relationship between immigration and crime and propose that increase in immigration may have been responsible for part of the steep crime plummet of the 1990s (Wadsworth 2010, p. 531).

An overview of all variables that were used in this study is given in Table 3. Fifteen of these variables are demographic factors and the rest thirteen are crime-related indicators.

There have not been standard abbreviations in use for shortening variables. For this study, data from three different sources were combined: two official sources and one unofficial source. Information about most items was derived from the database of United Nations Development Program (UNDP). Information about some items, which was missing in UNDP database, was derived from the World in Figures of the Statistics Finland (Tilastokeskus). In case information about some items of some countries was unavailable in UNDP database, but it was available in Statistics Finland database, information about such items was supplemented by Statistics Finland data. Such items include: birth rate, death rate, net migration, marriage rate, divorce rate, and population density. Unavailable items still appeared in the last datasheet and were labeled "NaN", (not a number) as required by SOM program. The sources of data are listed below in Table 4.

The purpose of current study was to map the contemporary crime situation of countries. It required information to be current. Information about most items was from UNDP's Human Development Report 2007/2008, with information dated to the year 2005. Information about some items requires a time span. In such cases, the time span ranges from 2 to 10 years. Some items were depicted with information dated 2004, 2006, 2007, or 2008. These items were seen as most relevant data in UNDP database in the sense of time (even though other sources have quite up-to-date information).

The dataset was retrieved from different online sources primarily of 2005, but information of some items was dated 2004, 2006, 2007 or 2008. Twenty-eight variables covering demographic situation were selected on the basis of usual statistical items available on international online platforms. However, figures of sex (gender) and race were excluded because their relationship with crime requires large-scale in-depth study and much has been done by other researchers. Thus, the dataset was composed of 56 rows and 28 columns.

Although the SOM can process a dataset with missing data, in this study, the dataset avoided attributes (in columns) and countries (in rows) with five or more data values unavailable. That is to say, not all attributes with available data are included in this study. All of these countries have no more than five missing values, and most of these attributes have less than ten missing data values. Three of attributes have more missing values, for example, police

Table 3 The country demographic situation measured by 28 different attributes

Demographic attributes	Name	Codification	Crime-related indicators	Name	Codification
1	Adult illiteracy (%)	ADUILL	16	Prisoners per Capita per 100,000 people	PRIPER
2	Birth rate per 1,000	BIRRAT	17	Share of prison capacity filled (%)	PRIFIL
3	Death rate per 1,000	DEARAT	18	Rape per 100,000 people	RAPPER
4	Fertility rate (children born per woman)	FERRAT	19	Robbery per 100,000 people	ROBPER
5	Health expenditure per capita (USD)	HEAEXP	20	Software piracy per 100,000 people	SOFPIR
6	Infant mortality rate per 1,000	INFMOR	21	Total crime per 100,000 people	TOTCRI
7	Life expectancy in years	LIFEXP	22	Police per 100,000 people	POLPER
8	Net migration per 1,000	NETMIG	23	Murder per 100,000 people	MURPER
9	Population density per km^2	POPDEN	24	Jails per 100,000 people	JAIPER
10	Population growth rate (%)	POPGRO	25	Fraud per 100,000 people	FRAPER
11	Population older than 64 (%)	POPOLD	26	Convicted per 100,000 people	CONPER
12	Population undernourished (%)	POPUND	27	Assault per 100,000 people	ASSPER
13	Under-five mortality rate per 1,000	UNDFIV	28	Burglary per 100,000 people	BURPER
14	Unemployment rate total (%)	UNERAT			
15	Urban population (%)	URBPOP			

Table 4 Sources of data

Name of sources	Websites
United Nations Development Program (UNDP), Statistics of the Human Development Reports, statistical update 2008	http://hdr.undp.org/en/statistics/
The Statistics Finland (Tilastokeskus), World in Figures, updated January 22, 2009	http://www.stat.fi/tup/maanum/index_en.html

per 100,000 people. By so doing, missing values have been deliberately controlled to a low rate. The total of missing values was 5.3 % as to all data values when the size of the data matrix applied to all calculations was $56 \times 28 = 1{,}568$ elements. Besides missing values, descriptions presented in Table 5 are mean, standard deviation, minimum and maximum of each attribute.

4.3 Construction of the map

In constructing the map, recommendations suggested by Kohonen and Honkala (2007) have been followed, but some flexibility has also been practiced in order to get better visualization. Several hundred maps were trained during the course of the experiments. The final selected network size was 7×9 nodes.

The SOM automatically generated maps including data of all countries and all variables. The *U*-matrix map (Fig. 3), clustering map (Fig. 4) and feature planes (Fig. 5) can be used in further analysis.

We formed many other maps in addition to those presented in Fig. 4. Larger maps than 7×9 in Fig. 4 resulted in greater numbers of clusters such as 12 for the size of 10×15 (Fig. 6) and 11 for that of 15×15 (Fig. 7). Smaller maps would have tended toward a lower number of clusters.

5 Results

Upon processing of data, five clusters of the map from Fig. 4 have been identified, each representing a group of countries sharing similar characteristics.

5.1 Clusters

The characteristics of each cluster can be demonstrated according to one of attributes so as to provide consistent explanation. In the following, the clusters are regrouped by total crime rate from high to low.

Cluster A consists of countries with very high level of total crime rate, including Switzerland, Germany, Norway, Denmark, France, Iceland, Great Britain, Canada, Australia, Finland, New Zealand, and the United States. From the

Table 5 Descriptions

Attribute	Mean	SD	Minimum	Maximum	Missing values
1	6.6	10.7	0.2	45.9	0 (0 %)
2	14.9	7.2	8.3	40.3	0 (0 %)
3	9.3	3.9	1.9	21.4	0 (0 %)
4	2.0	1.0	1.2	6	1 (1.8 %)
5	1,381	1,313	63	6,096	0 (0 %)
6	18.6	23.3	2	102	0 (0 %)
7	72.8	8.9	40.5	82.3	0 (0 %)
8	3.1	25.4	−47	157.9	0 (0 %)
9	108.8	123.8	2.8	614	0 (0 %)
10	1.0	1.1	−0.4	5.1	2 (3.6 %)
11	11.5	5.2	1.3	19.7	1 (1.8 %)
12	6.8	9.9	2.5	47	2 (3.6 %)
13	24.3	35.0	3	182	0 (0 %)
14	10.0	12.5	0.6	80	0 (0 %)
15	65.3	18.0	13.4	95.4	0 (0 %)
16	116.7	34.8	62.8	245.9	8 (14.3 %)
17	194.8	152.1	29	715	11 (19.6 %)
18	0.14	0.22	0	1.2	0 (0 %)
19	1.1	2.0	0	12.3	0 (0 %)
20	52.2	21.0	20	92	6 (10.7 %)
21	35.2	32.4	1.2	113.8	4 (7.14 %)
22	2.7	1.4	0.4	7.28	14 (25 %)
23	0.06	0.11	0	0.62	1 (1.8 %)
24	0.08	0.38	0	2.08	8 (14.3 %)
25	1.2	1.8	0	10.9	2 (3.6 %)
26	6.7	6.7	0.2	33.2	11 (19.6 %)
27	2.3	2.8	0.03	12.1	3 (5.4 %)
28	5.6	5.9	0	21.8	9 (16.1 %)

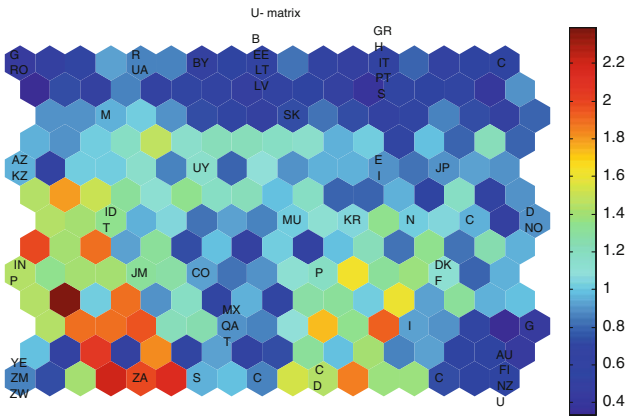


Fig. 3 U-matrix map for different countries

maps generated by SOM Toolbox, these countries have high GDP level, and satisfactory demographic indicators. They are characterized by, for example, high health expenditure, high life expectancy, aging, and high

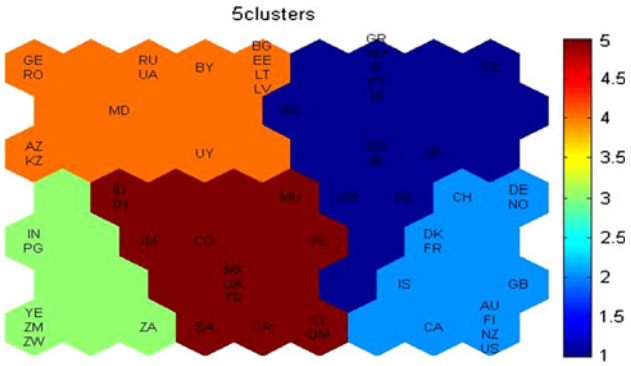


Fig. 4 Clustering map for countries

urbanization. Except murder, robbery and software piracy, other severe crimes, such as rape, fraud, assault and burglary are all at higher level than in most countries in other clusters.

Urbanization leads to increased numbers of strangers. Traditional intimate relationship between neighborhoods has been superseded. It is so that criminologists found that the less severe the bodily harm inflicted on the victim, the greater the likelihood of the crime being committed by a stranger (Thio 1978, p. 99). In other words, there are possibly more numbers of crimes in most urbanized countries, but these crimes are possibly less severe, while in less urbanized countries, there are a lower number of crimes, but these crimes are possibly more severe.

Cluster B consists of countries with a high level of total crime rate, including Slovakia, Greece, Hungary, Italia, Portugal, Slovenia, Czech Republic, Spain, Ireland, Japan, Korea, and the Netherlands. These countries are characterized by having the highest death rate, the biggest rate of population older than 64, the highest murder rate and highest share of prison capacity filled. Otherwise, most demographic indicators are at satisfactory levels.

Cluster C consists of countries with a medium level of total crime rate, including Indonesia, Thailand, Jamaica, Colombia, Mauritius, Poland, Mexico, Qatar, Turkey, Saudi Arabia, Costa Rica, Chile, and Dominica. It can be seen that most indicators are at medium levels, including demographic factors and crimes. Instead, they are characterized by the lowest level of net immigration rate, and the highest police per 100,000 population and rape rate.

In many countries, people have habitually been quick to blame the influx of immigrants for mounting crime rates. Historically, among the most popular reasons given for the wide-spread existence of crime and delinquency in the United States is the settling of large numbers of immigrants (Koenig 1962, p. 138). However, criminologists have found that far from bringing with them criminalistic behavior, most immigrants come with a respect for law and authority, which they acquired in their home countries;

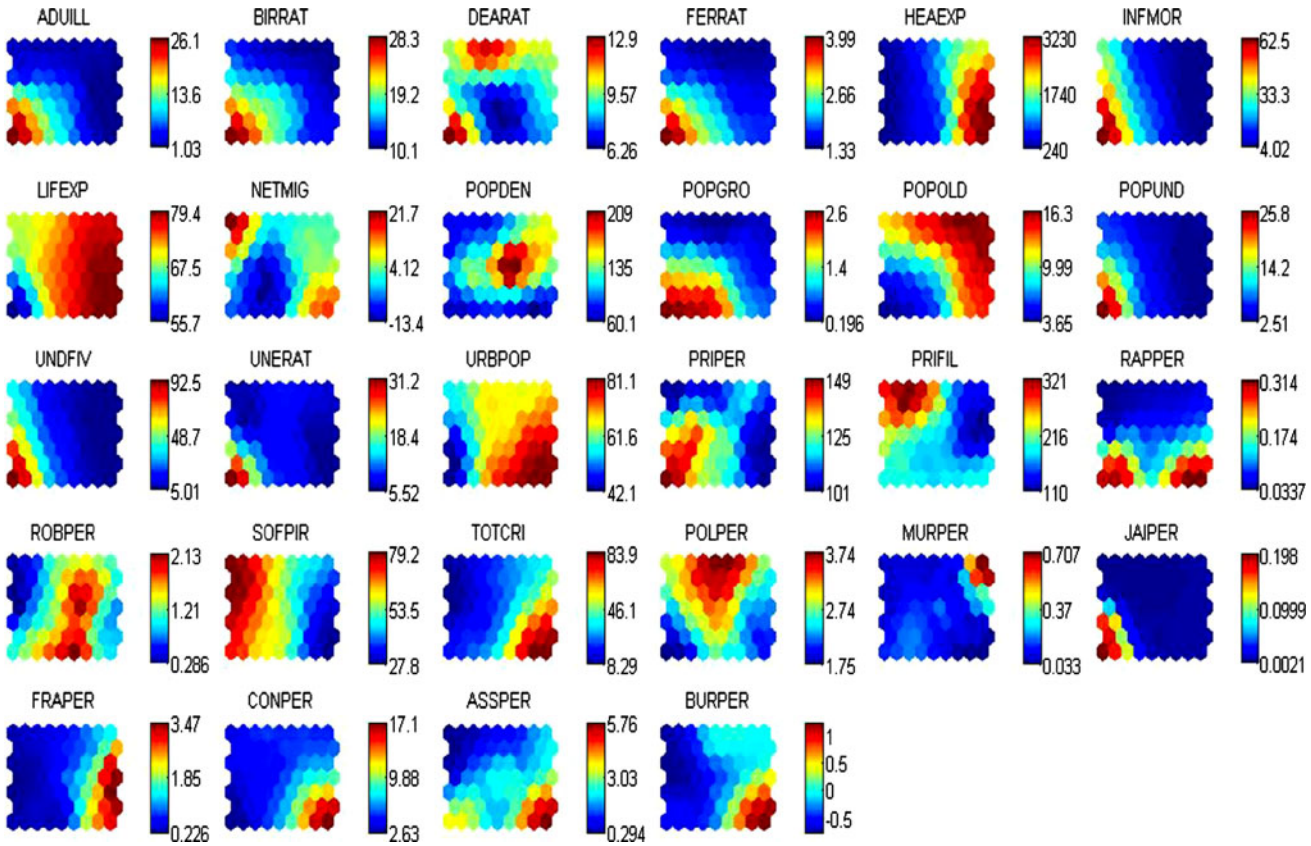


Fig. 5 Feature planes

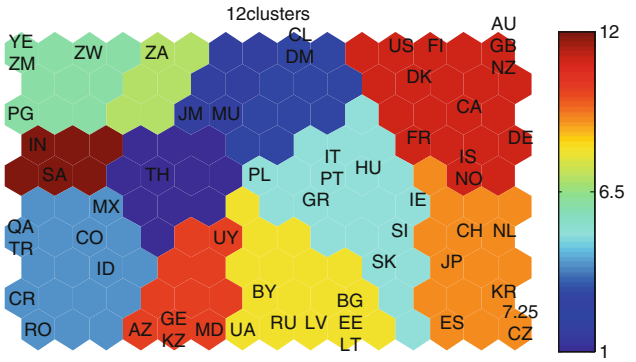


Fig. 6 Clustering map of size 10 × 15 nodes for 56 countries

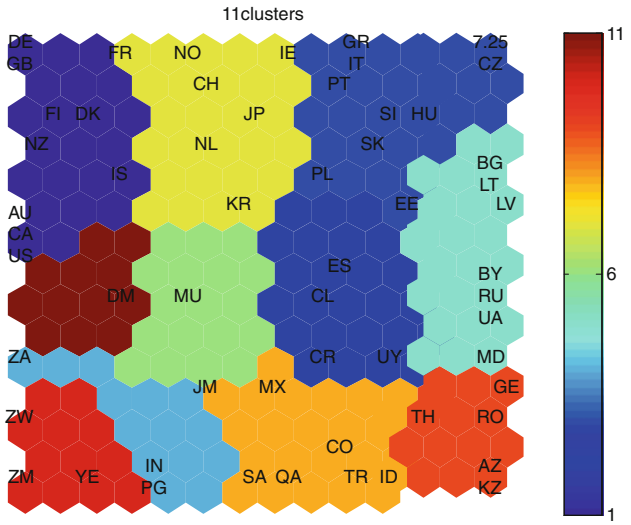


Fig. 7 Clustering map of size 15 × 15 nodes for 56 countries

they come mostly from stable, homogeneous societies which exerted strict control over the behavior of individuals (ibid., p. 142).

Cluster D consists of countries with a low level of total crime rate, including India, Papua New Guinea, Yemen, Zambia, Zimbabwe, and South Africa. These countries have the highest adult illiteracy rate, birth rate and death rate, fertility rate, population growth rate, population undernourished, under-five mortality rate and unemployment rate. Prisoner per 100,000 people is also the highest. However, except rape rate and jails per 100,000 people, all other severe crimes are at a low level.

Cluster E consists of countries with a very low level of total crime rate, including Georgia, Romania, Russia, Ukraine, Belorussia, Bulgaria, Estonia, Latvia, Lithuania, Moldova, Azerbaijan, Kazakhstan, and Uruguay. Not all

the indicators in countries of this cluster are lower than any other countries. Usually, in these countries, demographic indicators fall between two extremes. It also deserves to mention that, for example, they have the highest net immigration rate. Software piracy rate is the highest compared with other countries included in this study.

In PPIC (2008, p. 1) study, findings revealed that high immigrant population does not appeal to be associated with high crime rates. National studies have examined crime rates in jurisdictions with large and/or increasing immigrant populations and have found either no discernible link or a slightly negative one. A study of California cities with large populations of recently arrived immigrants showed no significant relationship between immigrant inflows and property crimes, and a negative relationship with violent crime rates (PPIC 2008, p. 1).

Although countries included in each cluster have similar level of total crime rate, other factors may vary from one to another. In this case, it is impossible to have a clear-cut division among countries merely on colors of final clusters.

The general distribution of countries included in this experiment can be found in Fig. 8.

It must be pointed out that total crime rate of a country does not necessarily correspond to a country's socio-economic image, neither to the number of a country's total crime. Most of the highly developed countries have high total crime rates, while in countries with large total crime numbers can only have a low total crime rate due to their large population base. A universal misunderstanding is that every indicator should be satisfactory in developed countries. This study proves once again that developed countries have higher total crime rates. This does not imply more than the fact. But behind the fact, there are always possible statistical backgrounds: more transparent justice system, well-established statistical institution, low crime standard, and zero tolerance approach to even minor crimes.

5.2 Correlation between attributes

Using Viscovery SOMine, correlation between every pair of attributes could also be identified. In this experiment, however, the results demonstrate that there are hardly any strong links between any attribute pairs. A part of the automatically generated correlation is shown in Table 6.

Table 6 shows correlation between prisoners per 100,000 people and other 27 attributes. Stronger positive correlation can be found with adult illiteracy, software piracy per 100,000 people and police per 100,000 people, possibly with infant mortality rate, under-five mortality rate, unemployment rate total, and share of prison capacity filled. Stronger negative correlation can be found with health expenditure per capita, life expectancy, and possibly urban population. Correlation between prisoners per capita

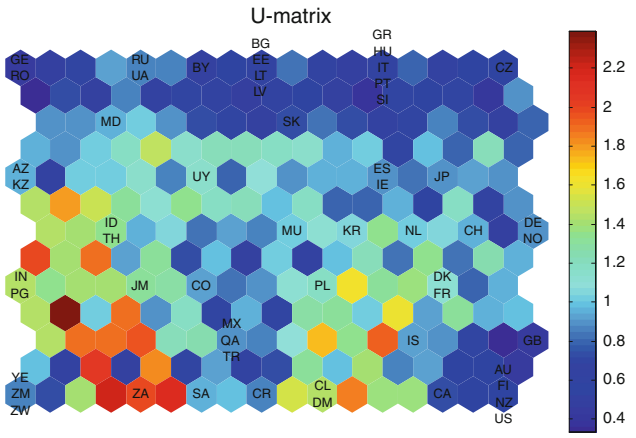


Fig. 8 World distribution between countries

and all other attributes, regardless of positive or negative, has shown to be weak.

Murder is a relatively rare form of criminal act. The murder rate is the lowest in comparison with other major crimes. Aside from its being rare, murder is the most serious offense—its cost to the victim is forever irreparable and its cost to the victim's loved ones is incalculably high (Thio 1978, p. 97). Strong positive correlation between murder per 100,000 people and birth rate, share of prison capacity filled, software piracy per 100,000 people, unemployment rate total, and even infant mortality rate was identified. On the other hand, strong negative correlation was found with health expenditure per capita, life expectancy and death rate. Weak correlation with other attributes was shown in Table 7, whether positive or negative.

Based on correlation pairs, each demographic factor has correlation with every indicators of crime, either positive or negative, either strong or weak.

(1) Adult illiteracy rate

Adult illiteracy rate is positively correlated with share of prison capacity filled, prisoners per capita, robbery rate, software piracy rate, murder rate and jails per 100,000 people. In these variables, adult illiteracy rate has strong correlation with murder, software piracy, prisoners per capita, and share of prison capacity filled.

It is negatively correlated with rape rate, total crime rate, police per 100,000 people, fraud rate, convicted rate, assault rate and burglary rate. Strong negative correlations exist between adult illiteracy rate and fraud and police per 100,000 people.

(2) Birth rate

Birth rate is positively correlated with prisoners per capita, share of prison capacity filled, rape rate, robbery rate, software piracy rate, murder rate, assault rate, and burglary rate. Strong correlations are those between birth

Table 6 Correlation between attribute 16, prisoners per capita, and other attributes

Other attributes	Correlation
1 Adult illiteracy	0.56
2 Birth rate	0.12
3 Death rate	0.08
4 Fertility rate	0.14
5 Health expenditure per capita	−0.48
6 Infant mortality rate	0.36
7 Life expectancy	−0.36
8 Net migration	−0.08
9 Population density	0.08
10 Population growth rate	0.10
11 Population older than 64	−0.07
12 Population undernourished	0.15
13 Under-five mortality rate	0.34
14 Unemployment rate total	0.31
15 Urban population	−0.28
17 Share of prison capacity filled	0.27
18 Rape per 100,000 people	0.17
19 Robbery per 100,000 people	0.02
20 Software piracy per 100,000 people	0.53
21 Total crime per 100,000 people	0.07
22 Police per 100,000 people	0.58
23 Murder per 100,000 people	0.03
24 Jails per 100,000 people	−0.02
25 Fraud per 100,000 people	−0.06
26 Convicted per 100,000 people	0.17
27 Assault per 100,000 people	0.22
28 Burglary per 100,000 people	−0.06

Table 7 Correlation between attribute 23, murder per capita, and other attributes

Other attributes	Correlation
1 Adult illiteracy	0.10
2 Birth rate	0.80
3 Death rate	−0.40
4 Fertility rate	−0.23
5 Health expenditure per capita	−0.65
6 Infant mortality rate	0.40
7 Life expectancy	−0.63
8 Net migration	0.09
9 Population density	−0.23
10 Population growth rate	−0.19
11 Population older than 64	−0.33
12 Population undernourished	0.28
13 Under-five mortality rate	0.39
14 Unemployment rate total	0.59
15 Urban population	−0.19
16 Prisoners per capita	0.04
17 Share of prison capacity filled	0.69
18 Rape per 100,000 people	0.22
19 Robbery per 100,000 people	0.40
20 Software piracy per 100,000 people	0.57
21 Total crime per 100,000 people	0.11
22 Police per 100,000 people	0.09
24 Jails per 100,000 people	0.04
25 Fraud per 100,000 people	0.30
26 Convicted per 100,000 people	0.10
27 Assault per 100,000 people	−0.07
28 Burglary per 100,000 people	0.37

rate and murder, software piracy rate, and prisoners per capita.

It is negatively correlated with total crime rate, police per 100,000 people, jails per 100,000 people, fraud rate and convicted rate. Only strong negative correlation is that between birth rate and fraud.

(3) Death rate

Death rate is positively correlated with prisoners per capita, total crime rate, police per 100,000 rate, jails per 100,000 people, fraud rate, convicted rate, assault rate, and burglary rate. Police per capita is strongly correlated with death rate.

It is negatively correlated with share of prison capacity filled, rape rate, robbery rate, software piracy rate, and murder rate. Death rate has strong negative correlation with software piracy rate and murder rate.

(4) Fertility rate

Fertility rate is positively correlated with prisoners per capita, share of prison capacity filled, software piracy rate, murder rate, jails per 100,000 people, assault rate, and burglary rate. Strong correlations are those between birth rate and murder, and software piracy rate.

It is negatively correlated with robbery rate, total crime rate, police rate, fraud rate, and convicted rate. Only strong negative correlation is that between birth rate and fraud.

(5) Health expenditure per capita

Health expenditure per capita is positively correlated with rape rate, total crime rate, police rate, jails per 100,000 people, fraud rate, convicted rate and burglary rate. Strong correlations are those between health expenditure per capita and fraud and burglary.

It is negatively correlated with prisoners per capita, share of prison capacity filled, robbery rate, software piracy rate, and murder rate. Only strong negative correlation is that between health expenditure per capita and prisoners per capita and software piracy rate.

(6) Infant mortality rate

Infant mortality rate is positively correlated with prisoners per capita, share of prison capacity filled, robbery rate, software piracy rate, police rate, and murder rate.

Strong correlations are those between birth rate and murder, and software piracy rate.

It is negatively correlated with rape rate, total crime rate, jails per 100,000 people, fraud rate, convicted rate, assault rate, and burglary rate. Strong negative correlation can be identified between infant mortality rate and fraud and total crime rate.

(7) Life expectancy

Life expectancy is positively correlated with rape rate, total crime rate, police per 100,000 people, jails per 100,000 people, fraud rate, convicted rate, assault rate, and burglary rate. Strong positive correlations can be identified between life expectancy and jails per capita, burglary and theft.

It is negatively correlated with prisoners per capita, share of prison capacity filled, software piracy rate, robbery rate, and murder rate. Life expectancy is strong correlate with the first three variables.

(8) Net migration

Net migration is positively correlated with rape rate, robbery rate, total crime rate, murder rate, jails per 100,000 people, fraud rate, convicted rate, assault rate and burglary rate. Mostly they are only weak correlative. Only stronger correlation is that between net migration and theft, jails per capita and convicted rate.

It is negatively correlated with prisoners per capita, share of prison capacity filled, software piracy rate, and police per 100,000 people. A stronger correlation exists between net migration and share of prison capacity filled.

(9) Population density

Population density is positively correlated with total crime rate, and fraud rate, but only weakly.

It is negatively correlated with prisoners per capita, share of prison capacity filled, rape rate, robbery rate, software piracy rate, murder rate, jails per 100,000 people, convicted rate, assault rate, and burglary rate. These correlations are also weak.

(10) Population growth rate

Population growth rate is strongly positively correlated with prisoners per capita, software piracy rate, and murder rate. It is also positively correlated with share of prison capacity filled, rape rate, and robbery rate,

It is negatively correlated with total crime rate, police rate, jails per 100,000 people, fraud rate, convicted rate, assault rate and burglary rate. Population growth rate is strongly negatively correlated with police per capita.

(11) Population older than 64

Population older than 64 is positively correlated with total crime rate, police rate, jails per 100,000 people, fraud rate, convicted rate, and burglary rate. Only strong correlation exists between population older than 64 and police per capita.

It is negatively correlated with prisoners per capita, share of prison capacity filled, rape rate, robbery rate, software piracy rate, murder rate and assault rate. Among these, it is strongly correlated with prisoners per capita and murder rate.

(12) Population undernourished

Population undernourished is positively correlated with prisoners per capita, share of prison capacity filled, rape rate, robbery rate, software piracy rate, and murder rate. It has strong correlation with murder rate.

It is negatively correlated with total crime rate, police rate, jails per 100,000 people, fraud rate, convicted rate, assault rate, and burglary rate. The strongest correlation is between population undernourished and police per capita.

(13) Under-five mortality rate

Under-five mortality rate is positively correlated with prisoners per capita, share of prison capacity filled, robbery rate, software piracy rate, and murder rate, in which it is strongly correlated with the last two.

It is negatively correlated with rape rate, total crime rate, police per 100,000 people, jails per 100,000 people, fraud rate, convicted rate, assault rate, and burglary rate. Under-five mortality rate is strongly negatively correlated with fraud rate.

(14) Unemployment rate total

Unemployment rate is positively correlated with prisoners per capita, share of prison capacity filled, robbery rate, software piracy rate, total crime rate, police per 100,000 people, jails per 100,000 people, and convicted rate. Strong correlation can be identified between unemployment rate total and share of prison capacity filled, robbery and convicted rate.

It is negatively correlated with rape rate, murder rate, fraud rate, assault rate, and burglary rate. But all correlations are weak.

(15) Urban population

Urban population is positively correlated with prisoners per capita, rape rate, robbery rate, software piracy rate, total crime rate, police per 100,000 people, jails per 100,000 people, fraud rate, convicted rate and assault rate. Urban population is strongly correlated with rape rate, total crime rate, and jails per capita.

It is negatively correlated with share of prison capacity filled, murder rate, and burglary rate. But all correlations are weak.

6 Discussion

Analysis is always subject to availability of data, particularly data on measurement of crime situation. Data used in this article covered most of the important demographic indicators and criminal phenomena.

When indicators are numerous, the result of clustering may become difficult. That is to say, some of the countries might be clustered into wrong groups. Exceptions occur now and then, here and there. So that it creates difficulties to give description more approximate to the fact. The SOM may lose to manual processing in this aspect.

However, it has the advantage that manual method cannot easily be compared with. With manual method, although an expert jurisprudent would make this, it is just impossible to put 56 countries into 5 clusters as in this experiment based on 28 different variables in a reasonable time limit. For example, calculating the number of combinations into five clusters would yield a huge number of different alternatives.

With the SOM, much human labor was superseded by automated computer processing. Even what human labor cannot be superseded, can be simplified or enhanced by the SOM. Certainly, much analytical work still requires human intervention and efforts.

Roughly defining patterns of crime situation have been found in some countries with some similar characteristics. It is not a miracle that the SOM Toolbox grouped Russia and some Eastern European countries into the same cluster (Cluster E), which included Georgia, Romania, Russia, Ukraine, Belorussia, Bulgaria, Estonia, Latvia, Lithuania, Moldova, Azerbaijan, Kazakhstan, and Uruguay. We cannot easily get such a conclusion, even if we deliberately want to do so. Nevertheless, the artificial intelligence method achieved this by disclosing certain implied mechanisms. In fact, when we look at other clusters as well, we can still be strongly shaken by the analytical power of the SOM over such demographic factors and criminal phenomena.

Using Viscovery SOMine, correlation between every pair of attributes can be identified. In this experiment, the results demonstrated that there are some strong positive or negative links between attribute pairs. Yet many other pairs are only weakly correlated.

In different countries, different factors may work in different ways. They may have positive correlation with crime in some countries, but have negative correlation with crime in other countries. They may have weak correlation in some countries, but have strong correlation in some other countries. These cannot be solved by the SOM alone and need to be supplemented through other research routines.

7 Conclusions

In this study, demographic information for 56 countries has been collected using the Internet as a source of information and a demographic database has been created. A number of demographic factors have been selected. Then, a data mining tool, the self-organizing map, has been used to perform a benchmarking of crime situation in these countries. The results of the study provide further evidence that the self-organizing map is a feasible and effective tool for crime situation benchmarking. The results are easy to visualize and interpret and provide a very practical way to compare the demographic factors of countries with different crime situation. This study has shown that crime research is one application area that can benefit from better visualization and data mining techniques.

References

Abidogun OA (2005) Data mining, fraud detection and mobile telecommunications: call pattern analysis with unsupervised neural networks. PhD thesis, University of the Western Cape, South Africa

Adderley R (2004) The use of data mining techniques in operational crime fighting. In: Proceedings of symposium on intelligence and security informatics no. 2. Tucson A. Z.: ETATS-UNIS (10/06/2004) 3073:418-425

Adderley R, Musgrave P (2003) Modus operandi modelling of group offending: a data-mining case study. Int J Police Sci Manag 5(4):265–276

Adderley R, Townsley M, Bond J (2006) Use of data mining techniques to model crime scene investigator performance. Knowl-Based Syst 20(2):170–176

Axelsson S (2005) Understanding intrusion detection through visualization. PhD thesis, Chalmers University of Technology, Göteborg

Bottoms AE (1976) Criminology and urban sociology. In: Baldwin J, Bottoms AE (eds) The urban criminal: a study in Sheffield. Tavistock Publications, UK, pp 1–35

Braithwaite J, Chapman B, Kapuscinski CA (1992) Unemployment and crime: towards resolving the paradox. Australian National University, Canberra

Brockett PL, Xia X, Derrig RA (1998) Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. J Risk Insur 65(2):245–274

Chung W, Chen H, Chaboya LG, O'Toole CD, Atabakhsh H (2005) Evaluating event visualization: a usability study of COPLINK spatio-temporal visualizer. Int J Hum-Comput Stud 62(1):127–157

Clinard MB (1958) Sociology of deviant behaviour. Rinehart & Company, New York

Cressey DR (1964) Delinquency, crime and differential association. Nijhoff, Hague

Dahbur K, Muscarello T (2003) Classification system for serial criminal patterns. Artif Intell Law 11:251–269

Dahmane M, Meunier J (2005) Real-time video surveillance with self-organizing maps. In: Proceedings of the second Canadian conference on computer and robot vision. The University of Victoria, Canada, pp 136–143

Deboeck G (2000) Self-organizing patterns in world poverty using multiple indicators of poverty repression and corruption. Neural Netw World 10:239–254

Dittenbach M, Merkl D, Rauber A (2000) The growing hierarchical self-organizing map. In: Proceedings of the international joint conference on neural networks (IJCNN'2000), Como, Italy, July 24–27, 2000. IEEE Computer Society Press, Los Alamitos, pp VI-15–VI-19

Fei B, Eloff J, Venter H, Olivier M (2005) Exploring data generated by computer forensic tools with self-organising maps. In:

Proceedings of the IFIP working group 11.9 on digital forensics, pp 1–15
Fei BK, Eloff JH, Olivier MS, Venter HS (2006) The use of self-organizing maps for anomalous behavior detection in a digital investigation. Forensic Sci Int 162(1–3):33–37
Grosser H, Britos P, García-Martínez R (2005) Detecting fraud in mobile telephony using neural networks. In: Ali M, Esposito F (eds) IEA/AIE 2005, LNAI 3533. Springer, Berlin, pp 613–615
Harries K (2006) Property crimes and violence in United States: an analysis of the influence of population density. Int J Crim Justice Sci 1(2):24–34
Hartung G, Pessoa S (2007) Demographic factors as determinants of crime rates. Retrieved July 10, 2012, from http://www.abep.nepo.unicamp.br/SeminarioPopulacaoPobrezaDesigualdade2007/docs/SemPopPob07_1062.pdf
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 Netw 470:946–951
Huysmans J, et al (2006) Country corruption analysis with self-organizing maps and support vector machines. In: Chen H et al (eds): WISI 2006, LNCS 3917, pp 104–114
Kangas LJ (2001) Artificial neural network system for classification of offenders in murder and rape cases. The National Institute of Justice, Finland
Koenig S (1962) The immigrants and crime. In: Roucek JS (ed) Sociology of crime. Peter Owen Ltd, London, pp 138–159
Kohonen T (1997) Self-organizing maps. Springer, Berlin
Kohonen T, Honkela T (2007) Kohonen network. Scholarpedia 2(1):1568
Lampinen T, Koivisto H, Honkanen T (2005) Profiling network applications with fuzzy c-means and self-organizing maps. Classif Clust Knowl Discov 4:15–27
Lee S-C, Huang M-J (2002) Applying AI technology and rough set theory for mining association rules to support crime management and fire-fighting resources allocation. J Inf Technol Soc 2:65
Lemaire V, Clérot F (2005) The many faces of a Kohonen map a case study: SOM-based clustering for on-line fraud behavior classification. Classif Clust Knowl Discov 4:1–13
Leufven C (2006) Detecting SSH identity theft in HPC cluster environments using self-organizing maps. Master's thesis, Linköping University, Sweden
Li S-T, Tsai F-C, Kuo S-C, Cheng Y-C (2006) A knowledge discovery approach to supporting crime prevention. In: Proceedings of the joint conference on information sciences, Taiwan
McGuire M (2005) Effects of density on crime rates in U.S. cities: a modern test of classic Durkheimian theory. Retrieved July 10, 2012, from http://www.sociology.uiowa.edu/mbmcguir/capstone/TRSresearch.doc
Memon Q A, Mehboob S (2006) Crime investigation and analysis using neural nets. In: Proceedings of international joint conference on neural networks, pp 346–350
Oatley GC, Ewart BW, Zeleznikow J (2006) Decision support systems for police: lessons from the application of data mining techniques to "soft" forensic evidence. Artif Intell Law 14(1): 35–100
PPIC (Public Policy Institute of California) (2008) Immigrants and crime. Retrieved July 10, 2012, from http://www.ppic.org/content/pubs/jtf/JTF_ImmigrantsCrimeJTF.pdf
Rhodes B, Mahaffey J, Cannady J (2000) Multiple self-organizing maps for intrusion detection. In: Proceedings of the 23rd National information systems security conference. October 16–19, 2000, Baltimore, Maryland, USA
Rock P (1994) History of criminology. Dartmouth, Aldershot
Situngkir H (2003) Merging the emergence sociology: the philosophical framework of agent-based social studies. J Soc Complex. Retrieved July 10, 2012, from http://cogprints.org/3519/1/Soc.pdf
South SJ, Messner SF (2000) Crime and demography: multi linkages, reciprocal relations. Ann Rev Sociol 26:83–106
Sumner C (2005) The social nature of crime and deviance. In: Sumner C (ed) The blackwell companion to criminology. Blackwell, Oxford, pp 3–31
Taft DR (1950) Criminology: a cultural interpretation. The MacMillan Company, New York
The Community Safety and Crime Prevention Council (1996) The root causes of crime—the community safety and crime prevention council statement on the root causes of crime. Waterloo, Canada
Thio S (1978) Deviant behavior. Houghton Mifflin Company, Boston
Wadsworth T (2010) Is immigration responsible for the crime drop? An assessment of the influence of immigration on changes in violent crime between 1990 and 2000. Soc Sci Quart 91(2): 531–553
Wirth L (1938) Urbanism as a way of life. Am J Sociol 44(1):1–24
Zaslavsky V, Strizhak A (2006) Credit card fraud detection using self-organizing maps. Inf Secur Int J 18:48–63

PUBLICATION III

Country crime analysis using the self-organizing map, with special regard to economic factors

Xingan Li and Martti Juhola

Country crime analysis using the self-organizing map, with special regard to economic factors

Xingan Li and Martti Juhola*

School of Information Sciences, 33014 University of Tampere, Finland

E-mail: Xingan.Li@uta.fi E-mail: Martti.Juhola@sis.uta.fi

*Corresponding author; tel. +358 40 1901716, fax +358 3 2191001

Abstract: In addition to analyzing existing criminal phenomena, the study of crime has witnessed an increasing demand for deterring potential occurrences of similar incidents. For this purpose, correlations between crime and economic factors became the focus of concern of criminologists and international law enforcement. Present information society proves performing crime analysis using data mining and visualization techniques to be an intimidating task. With the development of information systems, ubiquitous accessible statistics contain massive amounts of data on crime. Data mining and visualization techniques show their value in various domains but have not been broadly applied in the study of crime. Criminologists and law enforcement are in demand of an instrument allowing them to efficiently and effectively analyze these data. The self-organizing map (SOM), one of the widely used neural network algorithms, may be an appropriate technique for this application. The purpose of this study is to apply the SOM to mapping countries with different economic situations of crime. Besides, the SOM is also supplemented by other methods, including ScatterCounter used for attribute selection, and *k*-means clustering and nearest neighbor searching to obtain comparable results from those of the SOM. The dataset is comprised of a total of 50 countries and 30 variables. After initial processing of the data with the SOM, 4 clusters of countries were identified. Then the dataset was re-processed by ScatterCounter and four weak variables were removed reducing the final dataset to include 26 variables. It was found that some roughly defined patterns of crime situation can be identified in traditionally economically homogeneous countries. Among different countries, positive correlation on crime in some countries may have negative correlation in other countries. Overall, correlation of some factors on crime can still be found interesting. Results of the study proved that, after the validation of ScatterCounter's separation power function, *k*-means clustering and nearest neighbor searching, the SOM can be a new tool for mapping criminal phenomena through processing of multivariate data.

Keywords: data mining; self-organizing map; ScatterCounter; *k*-means clustering; nearest neighbor searching, crime situation; economic factors

Bibliographic notes: Xingan Li took his B. and M. Laws in China in 1989 and 1994. In 2008 he took his Dr. Laws degree at the University of Turku, Finland. He has been lecturer and researcher in Chinese and Finnish Universities. His research topics are criminology and artificial intelligence.

Martti Juhola took his M.Sc. and Ph.D. degrees at the University of Turku, Finland, in the 1980s. From 1992 he was a professor of computer science at the University of Kuopio and since 1997 at the University of Tampere, Finland. His research interests cover data mining and signal analysis.

1 Introduction

In modern society, justice is prevalent. So is unfortunately crime. With large volume of offences, prevention of crime has been one of the most important global concerns, along with the great efforts to strengthen public security. The study of crime cannot only be expected to control crime but also to analyze the social background of criminal phenomena so that potential occurrences of similar incidents can be overcome. Government and community officials are making an all-out effort to improve the effectiveness of crime prevention. Numerous investigations addressing this problem have generally employed disciplines of behavior science and statistics. To establish causation between crime and other phenomena in natural and social domain has perplexed generations of human beings.

Crime can be perceived as an outcome of multiple adverse personal, social, economic, cultural, and family conditions. Regardless of its complexity, to prevent crime, it is important to have an understanding of its roots (The Community Safety and Crime Prevention Council 1996). The study of crime has been situated in a long historical background with numerous studies making attempts to reveal causes of crime and seek ultimate solutions. No solely workable theory has thus far been invented to provide any precise answer for tackling crime, though many theorists presented many persuasive suggestions (Rock 1994). What is so special is that the study of

crime deals with a social phenomenon that hardly has a perfect solution. This study is not to search a new solution but to test a new method for identifying factors that are important in seeking potential solutions.

Currently, identifying correlation factors of crime, comparing geographical distribution of crime in different countries, and recognizing (including but not limited to predicting) criminal tendencies are attracting more than ever players. Criminologists want to reveal causation and correlation of crime. Law enforcement wants to recognize developmental tendencies of crime. Legislators want to enact effective laws to eliminate, prevent or reduce crime. Government wants to make feasible policy to combat crime and assist victims. Victims want to be socially rehabilitated to a peaceful, safe and harmonious life. The general public wants to create, enjoy, and maintain a society ruled by law, and the international society wants to coordinate and cooperate in reckoning with transnational crime. Processing crime data has been a basis for knowledge-detection and decision-making. The difficulty in analyzing large volume of crime date posed great challenge. Comparison and analysis in traditional ways become complicated and time-consuming. Advanced analytical methods are required to extract useful information from large amount of crime data.

The traditional modes for scholars to realize their goals are through either qualitative or quantitative or both methods. Publicly available national statistics facilitate quantitative analysis of criminal phenomena, even though a variety of analytical instruments have been employed in different fields of issues. The information required for the research can normally be found in databases of official publications and the Internet, including websites of international organizations, national statistical and judicial agencies, and other official documents.

Demand emerges for user interactive interfaces based on current technologies to meet and fulfill the new emerging responsibilities and tasks. Data mining and visualization techniques have shown their practical value in various domains but have not been extensively studied for applications in crime analysis. Today, these tools are becoming more and more sophisticated and extending their applications in more and more areas (such as Jing et al. 2009; Kumar et al. 2009; Chandra et al. 2010; Rahman et al 2011; Wang et al. 2007). One of such tools, the self-organizing map, uses an unsupervised learning method to group data according to patterns found in a dataset, making them ideal tools for data exploration. The SOM has attracted substantial research interests in a wide range of applications. Nonetheless, while many papers on the SOM have been published, very few studies have dealt with the use of the SOM in macroscopic research of criminal phenomena. This is an area in which innovative studies can be carried out. This study applied the SOM to investigate geographical distribution of criminal phenomena as regard to their economic background. While the scale of criminal phenomena can be expressed in many different aspects, particular attention here was paid to some important indicators, such as total crime rate.

2 Application of SOM in social and criminal research

Crime is one of the social problems attracting the most attention, research of which borrows ideas from generic or neighboring subjects. A couple of applications of the SOM to social research can help frame the study of crime. In practice, the SOM is one of the models of neural networks that acquire growing application in social research. Deboeck (2000) clustered world poverty into convergence and divergence in poverty structures based on multi-dimensions of poverty using the SOM, which reveal how new knowledge can be explored through artificial neural networks for implementing strategies for poverty reduction.

Crime-related social phenomena have also been studied with this method. For example, Lee and Huang (2002) made an attempt to extract associative rules from a database to support allocation of resources for crime management and fire-fighting; Huysmans et al. (2006) applied the SOM to process a cross-country database linking macro-economical variables to perceived levels of corruption with an expectation of forecasting corruption for countries; and Li et al. (2006) developed a linguistic cluster model aimed at meeting the demand of public security index and extracting relational rules of crime in time series. Findings of many such studies prove that artificial neural networks are a useful tool in social research.

Criminological research in detailed offences from micro viewpoints has also been acquiring more assistance from application of artificial intelligence. Hitherto, many researchers focus on applications of artificial neural networks to law enforcement, in particular, the detection of specific abnormal or criminal behaviors. Adderley et al. (2007) examined how data-mining techniques can support the monitoring of crime scene investigator performance. Oatley et al. (2006) presented a discussion of data mining and decision support technologies for police, considering the range of computer science technologies that are available to assist police activities. Dahmane et al. (2005) have presented an SOM application for detecting suspicious events in a particular scene.

The SOM has been, for instance, applied in the detection of credit card fraud (Zaslavsky and Strizhak 2006), automobile bodily injury insurance fraud (Brockett et al. 1998), burglary (Adderley and Musgrove 2003, Adderley 2004), murder and rape (Kangas 2001), homicide (Memon and Mehboob 2006), network intrusion (Rhodes et al. 2000, Leufven 2006, Lampinen et al. 2005, Axelsson 2005), cybercrime (Fei et al. 2005, Fei et al. 2006), and mobile communications fraud (Hollmén et al. 1999, Hollmén 2000, Grosser et al. 2005). Literature in

this aspect is abundant. This is the primary field where the SOM has found applications to research related to criminal justice before. From these studies, a field that has not been touched can also be identified, that is, a general lack of research on macroscopic aspects of criminal phenomena as related to other social factors, such as demographic characteristics, economic situation as well as historical development.

Most of these studies are based solely on the SOM, but in some cases, several methods are combined in a research. Lampinen et al. (2005) introduced two clustering methods, the SOM and the fuzzy *c*-means clustering (FCM) algorithm to be used in the analysis of network traffic. Abidogun (2005) provided a comparative analysis and application of the SOM and long short-term memory (LSTM) recurrent neural networks algorithms to user call data records in order to conduct a descriptive data mining on users call patterns. Adderley and Musgrove (2003) applied three data-mining techniques - the multi-layer perception (MLP), radial basis function (RBF), and the SOM - to the building descriptions, modus operandi (MO), and temporal and spatial attributes of domestic and commercial burglaries attributed to a network of offenders. Axelsson (2005) tried four different visualization approaches, including two direct approaches and two indirect approaches, to the problem of intrusion detection. Oatley's et al. (2006) discussion and experimentation are even wider, including decision support techniques based around spatial statistics, and a wide range of data mining technologies.

Besides crime detection, neural networks are also found useful in research specialized in victimization detection in mobile communications fraud (Hollmén et al. 1999).

From present literature, the SOM has been applied in the detection and identification of crimes. Applications of the SOM to the study of crime, that is, to identifying causative or correlative factors or to recognizing preventive or deterrent factors, have rarely been published. The current situation created a motivation for designing experiments exploiting this approach, in comparison with other methods. This article presents such an effort in applying the SOM to map geographic distribution of criminal phenomena as related to economic factors. What is in common with studies mentioned above is that this study can be categorised as one of the applications of the SOM in the study of crime, in broad sense. What is different is that this study is macroscopic, while others are microscopic. In this sense, this study is innovative, unique and different.

3 Methodology

Although nothing specialized on the study of macroscopic criminal phenomena has been published before, some literature provided some preliminary exploration into explanation for thinking self-organizing methods as feasible to do research on society as a whole. The vast computational technology provides us with the possibility to establish the sociology to cope with any sociological emergence phenomena (Situngkir 2003).

Some literature has been aware of the necessity, possibility and feasibility for applications of the SOM to the study of crime. They recognised that criminal justice is confronted with increasingly tremendous amount of data (for instance, in mobile communications fraud, Abidogun 2005). Crime data mining techniques become indispensable (Chung et al. 2005). They can support police activities by profiling single and series of crimes or offenders, and matching and predicting crimes (Oatley et al. 2006).

The difference between new techniques and old ones has been revealed in some literature. For example, they pointed out that unlike traditional data mining techniques that only identify patterns in structured data, newer techniques work both on structured and unstructured data. Researchers have developed various automated data mining techniques, depending heavily on suitable unsupervised learning methods (Dittenbach 2000). Cluster analysis helps the user to build a cognitive model of the data, thus fostering the detection of the inherent structure and the interrelationship of data (Dittenbach 2000).

Developed by Kohonen (Kohonen 1997) to cluster and visualize data, the SOM is an unsupervised learning mechanism that clusters objects having multi-dimensional attributes into a lower-dimensional space, in which the distance between every pair of objects captures the multi-attribute similarity between them. Some applications based on the concept of the SOM were developed particularly to meet the demand of law enforcement (for example, Fei et al. 2005, Fei et al. 2006, Lemaire and Clérot 2005). In particular, even though the data on the storage media may contain implicit knowledge that could improve the quality of decisions in an investigation, when large volumes of data are processed, it consumes an enormous amount of time (Fei et al. 2005). The SOM may play a positive role in exploratory data analysis (Lemaire and Clérot 2005).

Three categories of the network architectures and signal processes have been in use to model nervous systems. The first category is feed-forward networks, which transform sets of input signals into sets of output signals, usually determined by external, supervised adjustment of the system parameters. The second category is feed-back networks, in which the input information defines the initial activity state of a feedback system, and after state transitions the asymptotic final state is identified as the outcome of the computation, and the third category is self-organizing networks, in which neighboring cells in a neural network compete in their activities by means of

mutual lateral interactions, and develop adaptively into specific detectors of different signal patterns (Kohonen 1990, p. 1464).

The SOM can be sketched as an input layer and an output layer constituting two-layer neural networks. An unsupervised learning method is used in the SOM. The network freely organizes itself according to similarities in the data, resulting in a map representing the data input.

The SOM algorithm operates in two steps, which are initiated for each sample in the dataset. The first step is designed to find the best-matching node to the input vector, which is determined using some form of distance function, for example, the smallest Euclidian distance function. Upon finding the best match, the second step is initiated, the "learning step", in which the network surrounding node c is adjusted towards the input data vector. Nodes within a specified geometric distance, h_{ci}, will activate each other, and learn something from the same input vector $\mathbf{x}$. The number of nodes affected depends upon the type of lattice and the neighborhood function. This learning process can be defined as (Kohonen 1997, p.87):

$$\mathbf{m}_i(t+1)=\mathbf{m}_i(t)+h_{ci}(\mathbf{x}(t)-\mathbf{m}_i(t)). \quad (1)$$

The function $h_{ci}(t)$ is the neighborhood of the winning neuron c, and acts as a smoothing kernel defined over the lattice points. The function $h_{ci}(t)$ can be defined in two ways, either as a neighborhood set of arrays around node c or as a Gaussian function (Kohonen 1997, p. 87). In the training process, weight vectors are mapped randomly onto a two-dimensional, hexagonal lattice. A fully trained network facilitates a number of groups.

The SOM algorithm results in a map exhibiting the clusters of data, using dark shades to demonstrate large distances and light shades to demonstrate small distances (U-matrix method) (Kohonen, 1997). Feature planes, which are single vector level maps, can additionally be generated to discover the characteristics of the clusters on the U-matrix map. They present the distribution of individual columns of data.

Upon processing the data, maps can be generated using software packages. By observing and comparing the clustering map and feature planes, rough correlation between different indicators (attributes) can be identified. Detailed correlation table can also be realized automatically with Viscovery SOMine (Viscovery Software GmbH, 2012), which adopts the correlation coefficient scale ranging from -1.0 to +1.0. These clustering maps, feature planes and correlation tables provide major basis for further analysis.

This study applied the SOM to explore the economic background on which the criminal phenomena are located. Based analysis on available data, the results of the study will revolve around the feasibility of the SOM as a tool for mapping criminal phenomena, identifying correlations, with special regard to economic factors, through processing of relevant data.

During the application of the SOM, in order to select variables, ScatterCounter (Juhola and Siermala 2012a, 2012b) will be used to identify which variables are strong in clustering, and which are weak. The weak ones will be removed from the dataset and the selected dataset will be used in final processing and analysis. In using ScatterCounter, missing data in the original dataset have to be filled with real values. In this research, missing values are filled by the medians of the available values of the variables in the same clusters.

Besides the SOM, k-means clustering and nearest neighbor searching will be used to validate the clusters and analysis by calculating how accurately k-means clustering and nearest neighbor searching methods put the same countries into the same clusters as the SOM does.

4 Design of experiments

4.1 Countries included

There were 50 countries included in the experiment as illustrated in Table 1. The countries were codified according to ISO 3166-1-alpha-2 code elements, that is, each country was referred to by two letters, for example, Finland was denoted by FI, Norway by NO, and Denmark by DK, etc. These codes will be showed in the maps as labels.

4.2 Crime and economic factors

Economics has been employing ever-changing concept about human beings and their activities. Adam Smith's economics implied that man is a rational animal who seeks material pleasure or utilities, in competition with his fellows, and this selfish, competitive search for personal gain was socially favorable, and should be left unimpeded by government. Later study has shown that Adam Smith's view misinterpreted the real nature of human motivation, and underestimated the social ills resulting from unregulated individualism (Taft 1950, p. 123).

Table 1 Countries included with their codes

Australia	AU	Spain	ES	Japan	JP	Romania	RO
Azerbaijan	AZ	Finland	FI	Korea, Republic of	KR	Russian Federation	RU
Bulgaria	BG	France	FR	Kazakhstan	KZ	Slovenia	SI
Belarus	BY	United Kingdom	GB	Lithuania	LT	Slovakia	SK
Canada	CA	Georgia	GE	Latvia	LV	Thailand	TH
Switzer-land	CH	Greece	GR	Moldova, Republic of	MD	Turkey	TR
Chile	CL	Hungary	HU	Mauritius	MU	Ukraine	UA
Colombia	CO	Indonesia	ID	Mexico	MX	United States	US
Costa Rica	CR	Ireland	IE	Netherlands	NL	Uruguay	UY
Czech Republic	CZ	India	IN	Norway	NO	South Africa	ZA
Germany	DE	Iceland	IS	New Zealand	NZ	Zambia	ZM
Denmark	DK	Italy	IT	Poland	PL		
Estonia	EE	Jamaica	JM	Portugal	PT		

Economic distress has long been considered as the basic cause of society's ills (Clinard 1958, p. 92). However, poverty is by no means the only factor accounting for the deviant behavior (Ibid, p. 98). General delinquency and criminal trends are not directly sensitive to the downward and upward movements of economic conditions (Mowrer 1942, pp. 190-191). Widespread absolute poverty does not necessarily lead to crime, but relative poverty, that is, intensified difference of living standard between each other, motivates people with low level of living towards high level of living, and motivates people with high level of living towards yet higher and higher level of living, up to luxury living.

The importance of economic conditions as causes of crime grows largely out of the fact that materialism is approved in our culture (Taft 1950, p. 124). In such a culture, men have positive ambitions even when not suffering actual discomfort. People reduce the difference of level of living through raising their level of living (Ibid., p. 124). Like standard of living, other economic factors may also affect the level of crime in one way or the other.

A revived version of economic analysis of crime insists that criminals respond to economic incentives in the same way that legal workers do (Becker, 1968). Economic theories of crime relate the likelihood that an individual engages in criminal activities to the costs and benefits of these activities, when compared to legal occupations. At the aggregate level, the more prevalent the conditions which make crime attractive, the higher the crime rates (Soares 2004, p. 157).

This study contains seventeen economic factors, which are divided into four categories, including economic and consumption level, economic structure, development of new economic phenomena, and extent of research and development.

(1) Regarding economic and consumption level, factors such as electricity consumption per capita, electrification rate, GDP per capita annual growth rate, GDP per capita, and GDP per capita purchasing power parity (PPP), an indicator concerning the change of consumption, are covered.

The positive link between crime and development—usually cited in the criminology literature but regarded with suspicion by economists—does not exist. Reporting rates of crimes are strongly related to development, mainly income per capita. Therefore, the positive correlation between crime and development sometimes reported is entirely caused by the use of official records. Development is not criminogenic (Soares 2004, p. 156). In fact, this kind of recognition has been common in the academic field of the study of crime. At the same time, some correlated pair wises can have also been very common, such as the conclusion drawn by Soares: income inequality affects crime rates positively (increasing crime), while education and growth reduce crime (Ibid., p. 156). Data processing in this study will provide further insight into relationship between crime and selected economic factors.

(2) Regarding economic structure, factors such as employment in agriculture, employment in industry, employment in services, exports of goods and services, foreign direct investment net inflows, forest area, and imports of goods and services are included.

In today's world, some countries have a predominantly agricultural economy. Many other countries have been undergoing transformation from agricultural economy to industry and services. During the last two centuries, industry developed rapidly and the global economy tended increasingly to be characterized by industrialization, specialization, and urbanization.

Traditionally, theorists from Durkheim-Modernization perspective insist that rapid social-economic change creates a social platform where those factors leading to deviant behavior reside. Such factors are industrialization, urbanization, the division of the labor, social disorganization, anomie, modern values, and cultural heterogeneity (Masahiro 2002, p. 497). However, there has never been unanimously accepted conclusion concerning the properties of correlations between crime and such economic factors. This situation renders present study possibility to reconsider the previous conclusions from a different point of view.

(3) Regarding new economic phenomena, considered are factors such as cellular subscribers per 1000 people, Internet users per 1000 people, and telephone mainlines per 1000 people.

Contemporary scientific progression and marvelous advancement in communications have facilitated criminals of every part of the globe to perpetrate an offence by means of complicated apparatus in one location and afterwards run away to a different location. We are confronted with an historic process, starting with the invention of the computer in the 1940s, accelerating through a variety of forces, and causing profound changes in the life of people all around the world. It is a clearly distinguishable force, or rather a complex of intimately connected forces. The ubiquitous use of telephone, mobile phone and the Internet create new opportunities both for crime and anti-crime. They can play different roles in crime.

In addition, sixty years ago, when discussing the influence of mass media on crime, Taft wrote that newspapers might spread the skills of crime, make offences seem to normal activities, make crime seem to be attractive and exciting to vulnerable groups of people, make crimes seem to be excessively profitable, give reputation to the wrong-doers, attract sympathy or admiration for criminals, appeal to "lower" impulses and by sensationalism, reflect crime-producing elements in society, make escape from justice seem to be simple and by hindering the apprehension of criminals, fail to stress the punishment of crime, ridicule the mechanism of justice, or through "trial by newspaper," and advocate types of treatment of criminals which tend to increase crime (Taft 1950, pp. 206-211). Later on until today, the same can also be found with radio, television, telephone, and the Internet. Mass media usually have the function of maneuvering the popular cultural trends, and the function of broadcasting negative ideas at the same rate as broadcasting those positive ones.

(4) Regarding research and development, two factors, namely Research & Development (RD) expenditure and numbers of researchers in RD per 100,000 people are involved in the analysis. These factors are considered to reflect a country's long-term development policy and strategy and have long-term influence on a country's social development. They are taken into account together with other factors.

An overview of all variables that were used in this study is as given in Table 2 and 3. Seventeen of them are socioeconomic factors, while the rest thirteen are crime-related indicators.

There has not been a standard codification method in use for shortening variables. In this, codifications were realized by capitalizing the combined six letters comprised of the first three letters of each of the first two words in the names of the items. For example, "cellular subscribers per 1000 people" was codified as CELSUB (CEL from cellular, and SUB from subscribers). Exception occurred in codifying the item named population in poverty, which was codified as EMPINA ("employment in agriculture") (EMP from employment, IN from in, and the last A from agriculture), software piracy rate, which was codified as SFTPIR. For this study, data from different sources was combined. Except sources noted otherwise, information about most items, was derived from the database of United Nations Development Program (UNDP).

The purpose of current study was to map the contemporary crime situation of countries. It required proper information to be current. Information about most items was from UNDP's Human Development Report 2007/2008, with information dated to the year 2005. Information about some items required a time span. In such cases, the time span ranges from 2 to 10 years. Some items were depicted with information dated 2004, 2007, or 2008, as shown in Table 3. These items were seen as most relevant data in UNDP database in the sense of time (even though other sources have quite up-to-date information).

Unlike experiments in natural sciences that can give updated data, statistics of crime at international level prove to be a long-term and gradual task. If we select only concurrent data, the number of missing data is too great, and the current study simply cannot be designed. Fortunately, many of such factors, especially when they are compared internationally, have relative stability over years. The study in this field must make a choice between taking a statistical risk and giving up. In this original dataset, total missing values account for 7.7%.

Because this is not a study to seek final explanation and final solution to social problem, it takes a reasonable statistical risk for the purpose of testing this method in a new application field.

4.3 Attribute selection of the dataset

After the dataset was established for processing, Viscovery SOMine was used to identify clusters. Upon initial clusters were identified, the structure of dataset was modified to be processed with ScatterCounter (Juhola and Siermala 2012a, 2012b). 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 countries are labeled by cluster identifiers given by the preliminary SOM runs with the original 30 attributes.

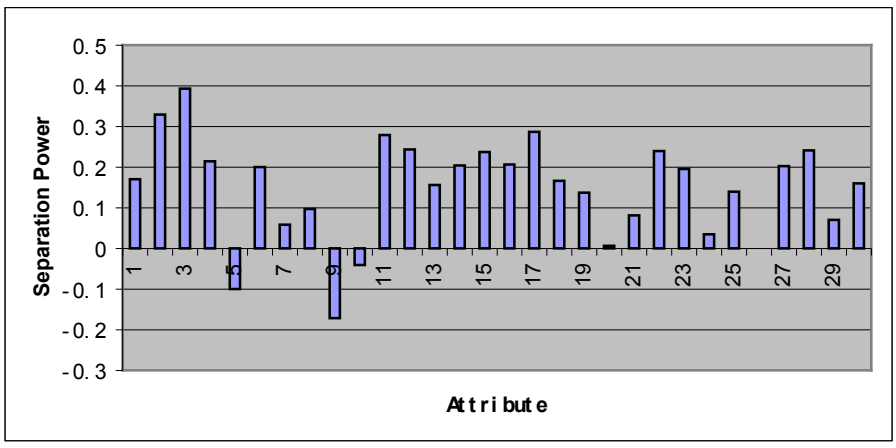
The objective of ScatterCounter is to evaluate how much subsets labeled as classes (clusters given by SOM) differ from each other in a dataset. Its principle is to start from a random instance of a dataset and to traverse all instances by searching for the nearest neighbor of the current instance, then 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 randomized steps, it is repeated from 5 to 10 times to use an average for separation power.

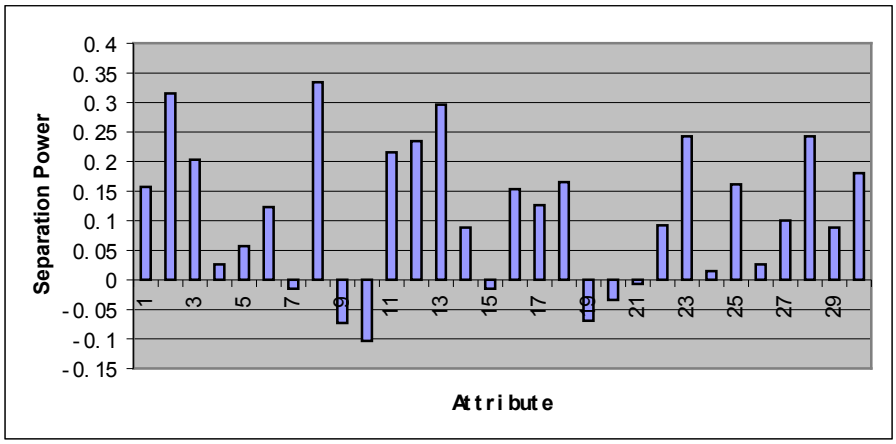
Separation powers can be calculated for the whole data or separately for every class and for every attribute (Juhola and Siermala 2012a, 2012b). Absolute values of separation powers are from [0,1). They are usually positive, but small negative values are also possible when an attribute does not separate 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, four variables have poor separation powers (Figure 1) and are removed from the dataset used in the following experiments and analysis: attribute (7) exports of goods and services, (9) forest area, (10) GDP per capita, and Jails per 100,000 people. Attribute (21) (robberies per 100,000 people) could also be dropped on the basis of slight separation power. Notwithstanding this result, we kept it as a plausible attribute. Attribute (5) was useful for clusters 2 and 3 although otherwise poor. In principle, our analysis with ScatterCounter might suffer from the small size of the dataset used and especially from the small size of cluster 4, merely three countries.

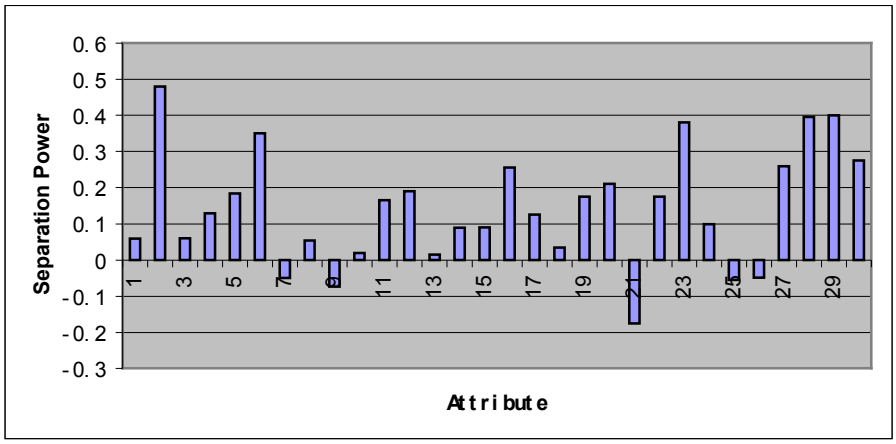
Figure 1 Separation powers of each variable in: (a) cluster 1, (b) cluster 2, (c) cluster 3, (d) cluster 4, and (e) whole dataset



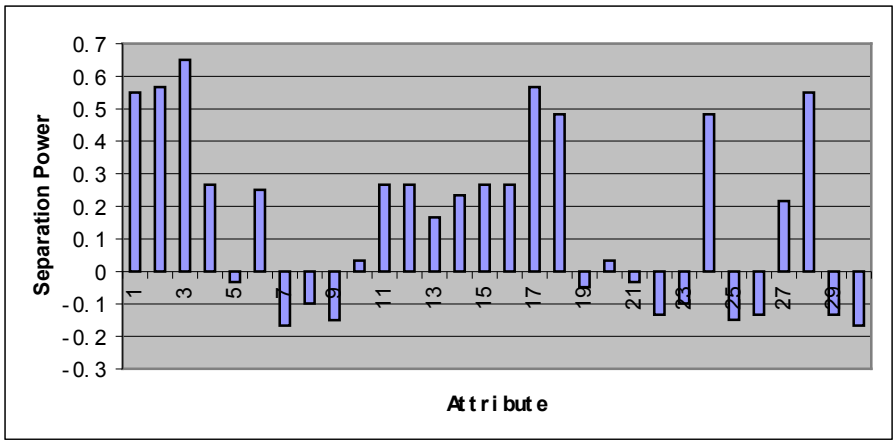
(a)


Separation Power
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
Attribute

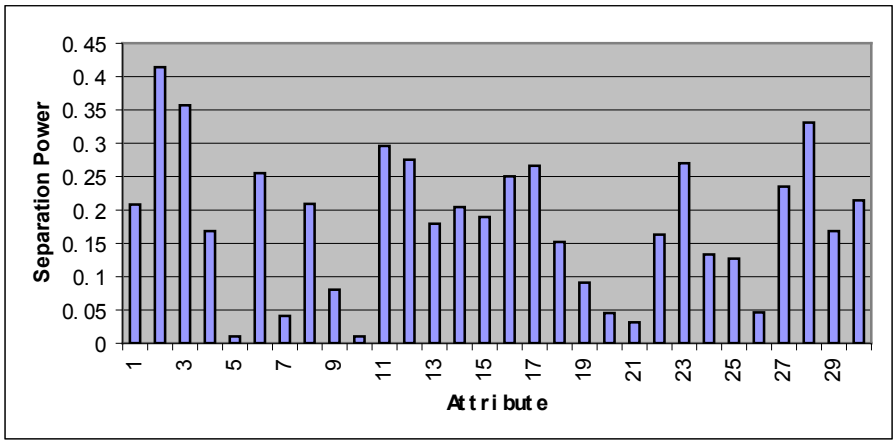
(b)


Separation Power
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
Attribute

(c)


Separation Power
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
Attribute

(d)



(e)

As a result, all the 26 variables preserved in the dataset have strong separation power and supposedly enable valid clustering. In this reduced dataset, total missing values account for 6%. Of total 26 attributes, 13 had no missing values. The highest frequency of missing values in certain attribute reached 38%.

4.4 Construction of the map

In this study, software packages used is Viscovery SOMine 5.2.2 Build 4241. Compared with SOM Toolbox, Viscovery SOMine has almost the same requirements on the format of the dataset. (A dataset designed for SOM Toolbox has a header, which must be removed in Viscovery SOMine. Or else, each blank in the header will be calculated as one missing value in each variable.) At the same time, requiring less programming, it enables an easier and more operable data processing and visualisation.

The SOMine software automatically generated maps including data of all 50 countries and all 26 variables. The clustering map (Figure 2) and feature planes (Figure 3) as well as some other detailed statistics, such as correlations as discussed below, can be used in further analysis.

Table 2 The country economic situation measured by 30 different attributes or factors

Items		Code in the SOM	Year of the Data
1.	Cellular subscribers per 1000 people	CELSUB	UNDP 2005
2.	Electricity consumption per capita	ELECON	UNDP 2004
3.	Electrification rate %	ELERAT	UNDP 1996-2005
4.	Employment in agriculture %	EMPINA	UNDP 1996-2005
5.	Employment in industry %	EMPINI	UNDP 1996-2005
6.	Employment in services %	EMPINS	UNDP 1996-2005
7.	Exports of goods and services % of GDP (dropped after selection by ScatterCounter)	EXPOFG	UNDP 2005
8.	Foreign direct investment net inflows % of GDP	FORDIR	UNDP 2005
9.	Forest area % (dropped after selection by ScatterCounter)	FORARE	UNDP 2005
10.	GDP per capita annual growth rate % (dropped after selection by ScatterCounter)	GDPANN	UNDP 1975-2005
11.	GDP per capita USD	GDPPER	UNDP 2005
12.	GDP per capita PPP USD	GDPPPP	UNDP 2005
13.	Imports of goods and services % GDP	IMPOFG	UNDP 2005
14.	Internet users per 1000 people	INTUSE	UNDP 2005
15.	RD expenditure % of GDP	RDEXPE	World Bank 2007
16.	Researcher in RD per 100,000 people	RESINR	Ibid 2007

17.	Phone mainlines per 1000 people	TELMAI	UNDP 2005
18.	Prisoners per 100,000 people	PRIPER	Int. Centre for Prison Studies 2003
19.	Prison capacity filled per cent	PRIFIL	Ibid. 2003
20.	Rapes per 100,000 people	RAPPER	UNODC 1998-2000
21.	Robberies per 100,000 people	ROBPER	UNODC 1998-2000
22.	Software piracy rate %	SFTPIR	Annual Software Piracy Study 2007
23.	Total crimes per 100,000 people	TOTCRI	UNODC 1998-2000
24.	Police per 100,000 people	POLPER	UNODC 1998-2000
25.	Murders per 100,000 people	MURPER	UNODC 1998-2000
26.	Jails per 100,000 people (dropped after selection by ScatterCounter)	JAIPER	UNODC 1998-2000
27.	Frauds per 100,000 people	FRAPER	UNODC 1998-2000
28.	Convicted per 100,000 people	CONPER	UNODC 1998-2000
29.	Assaults per 100,000 people	ASSPER	UNODC 1998-2000
30.	Burglaries per 100,000 people	BURPER	UNODC 1998-2000

Table 3 Descriptive statistics

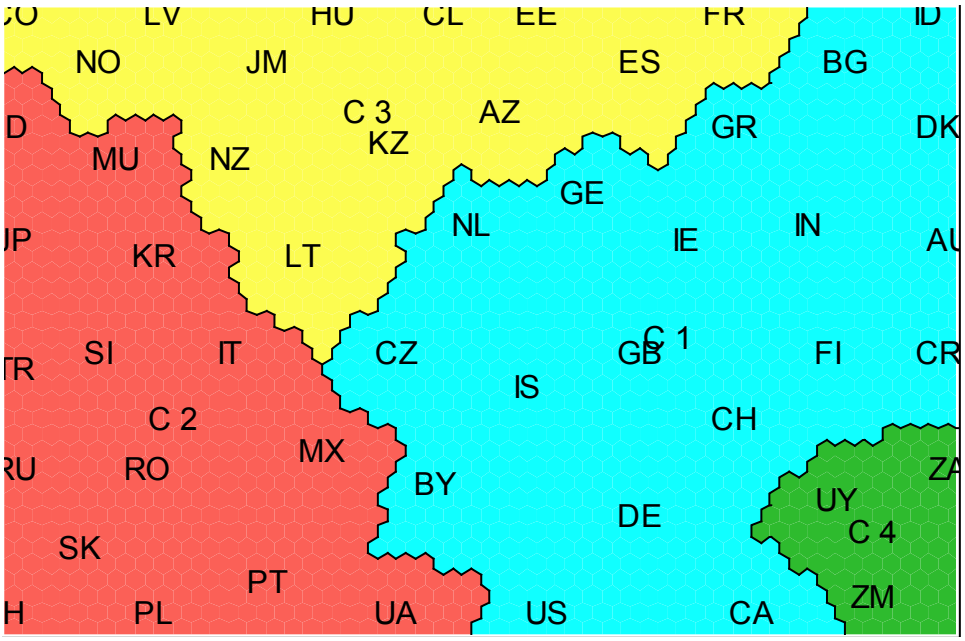
Attribute	Mean	Std. Deviation	Minimum	Maximum	Missing Values
Cellular subscribers per 1000 people	726	312	81	1275	0 (0%)
Electricity consumption per capita	6325	5944	476	29430	0 (0%)
Electrification rate %	92.2	18.3	19	100	19 (38%)
Employment in agriculture %	15.84	16.87	1	70	0 (0%)
Employment in industry %	24.82	7.15	7	40	0 (0%)
Employment in services %	59.06	13.57	20	78	0 (0%)
Exports of goods and services % of GDP (dropped after selection by ScatterCounter)	42.72	18.93	10	84	0 (0%)
Foreign direct investment net inflows % of GDP	36.78	25.02	4	91	0 (0%)
Forest area % (dropped after selection by ScatterCounter)	31.52	17.56	0.5	73.9	0 (0%)
GDP per capita annual growth rate % (dropped after selection by ScatterCounter)	1.59	2.036	-4.4	6	1 (2%)
GDP per capita USD	17700	17418	623	63918	0 (0%)
GDP per capita PPP USD	18572	11947	1023	41890	0 (0%)
Imports of goods and services % GDP	44.68	20.15	11	91	0 (0%)
Internet users per 1000 people	342.9	224.8	20	869	0 (0%)
RD expenditure % of GDP	1.143	0.919	0.01	3.46	0 (0%)
Researcher in RD per 100,000 people	2271	1866	51	7832	7 (14%)
Phone mainlines per 1000 people	344.1	177.3	8	689	0 (0%)
Prisoners per 100,000 people	194.8	152.1	29	715	5 (10%)
Prison capacity filled per cent	116.4	35.4	62.8	245.9	5 (10%)
Rapes per 100,000 people	0.128	0.221	0	1.2	0 (0%)
Robberies per 100,000 people (dropped after selection by ScatterCounter)	1.2	2.08	0.01	12.33	0 (0%)
Software piracy rate %	50.54	20.29	20	92	4 (8%)
Total crimes per 100,000 people	35.8	30.6	1.6	105.9	3 (6%)
Police per 100,000 people	2.715	1.276	0.37	7.28	11 (22%)
Murders per 100,000 people	0.0684	0.1176	0	0.62	1 (2%)
Jails per 100,000 people (dropped after selection by ScatterCounter)	0.052	0.317	0	2.08	7 (14%)
Frauds per 100,000 people	1.33	1.88	0.03	10.87	2 (4%)

Convicted per 100,000 people	6.89	6.94	0.17	33.15	9 (18%)
Assaults per 100,000 people	2.38	2.85	0.03	12.08	3 (6%)
Burglaries per 100,000 people	5.82	5.77	0.06	21.75	9 (18%)

5 Results

Upon processing of data, four clusters have been identified, each representing groups of countries sharing similar characteristics. As a default rule in self-organizing maps, values are expressed in colors: warm colors denote high values, while cold colors denote low values. In each of the feature planes (Figure 3), there is a color bar indicator depicting cold colors from the left and warm colors to the right. In the clustering map (Figure 2), there is not a color bar indicator, but in default, it applies the same as in the feature planes. However, order of clusters is not made according to colors, but according to the number of countries included in each cluster. That is to say, cluster with the most countries is numbered as 1, cluster with the second most countries is numbered as 2, and so on.

Figure 2 Clustering map with cluster names C1–C4 and labels of countries (added by the software)



5.1 Clusters

Clusters were given in Figures 2 and 3. Cluster 1 consists of 19 countries both with the low level and medium level of total crime rate. Most of them are economically developed with the high level of GDP per capita, highest rate of employment in services. They have high level of fraud, assault and burglary rates. Certainly, there are also countries with different characteristics in some fields such as economy, etc.

Table 4 Countries in clusters C1-C4

C 1: ID, BG, GR, DK, GE, NL, IN, IE, AU, CZ, CR, FI, CH, GB, DE, CA, IS, BY, US
C 2: MD, MU, JP, KR, SI, IT, TR, MX, RU, RO, SK, TH, PL, PT, UA
C 3: CO, LV, HU, CL, EE, FR, NO, JM, ES, AZ, KZ, NZ, LT
C 4: ZA, UY, ZM

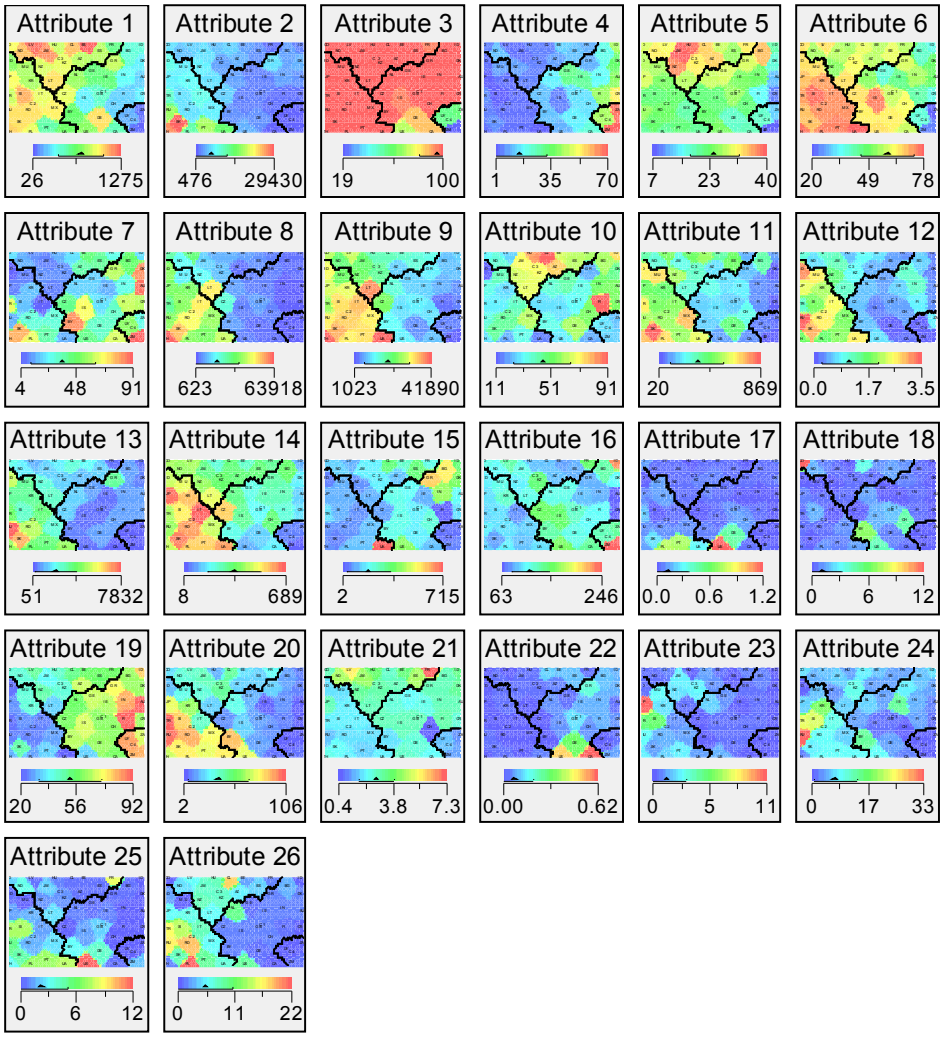
Cluster 2 consists of 15 countries with high level of total crime rate. They are characterized by the lowest level of murders per 100,000 people. Some countries have high level of some economic indicators, including

employment in services, Internet and phone mainlines users. They have high level of burglaries per 100,000 people.

Cluster 3 consists of 13 countries with medium level of total crime rate. They are characterized by the low rape rate and police per 100,000 people. According to the figures, these countries have high level of cellular subscribers per 1000 people, employment rate in industry, and imports of goods and services per GDP. According to figures, in these countries, there are more jails and more people in these jails on average.

Cluster 4 consists of 3 countries with the low level of total crime rate. These countries are characterized by high employment rate in agriculture and software piracy. Many other aspects seem to be low, both economic indicators and crime-related statistics.

Figure 3 Feature planes, showing, through warm or cold colors and numerical scale, differences in 26 attributes of those countries in four clusters



5.2 Validation of clusters with k*-means clustering and nearest neighbor searching*

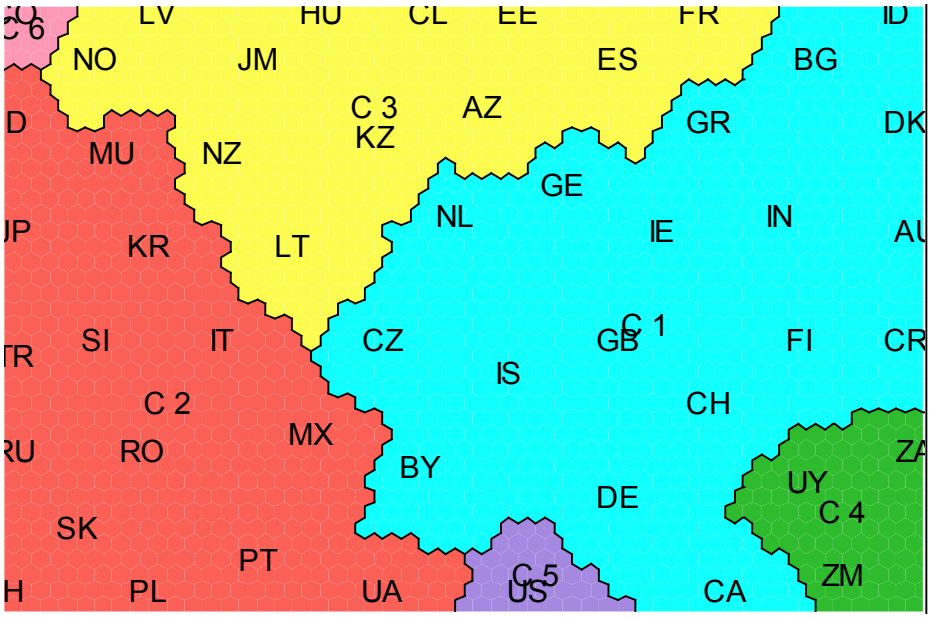
The results by the SOM were tested by *k*-means clustering methods with Euclidean distances. First, we used *k*-means to compute $k \geq 4$ clusters in the unsupervised manner and compared groups of the SOM results to the clusters formed. Here *k*-means clustering methods were used both with data not scaled and scaled. With *k*-means clustering, when data were not scaled, the probability of putting into the same clusters as generated by the SOM was more than 49.5±0.9%. When data were scaled, the probability was 48.3±0.7%. In these two cases, scaling of data made no obvious differences.

In addition, leave-one-out testing was used, in supervised manner, to test countries one by one against the result generated by the SOM. In leave-one-out testing, n-1 countries or cases (n=50) are used to build a training

set and one single country was the only test case for that model. All n=50 tests were repeated 26 times because random initialization was used by k-means clustering, and the mean and standard deviation were computed. Again, in this supervised clustering, both not scaled and scaled data were used. With not scaled data, the probability of putting into the same clusters as generated by the SOM is more than 43.5±2.0%; when data were scaled, the probability is 45.4±1.7%. In these two cases, scaling of data got somewhat better result.

Finally, k nearest neighbour searching function was also applied with Euclidean distances. Because the smallest SOM cluster have only 3 countries, it is reasonable to look at the closest neighbor using k=1. When tests are made in the leave-one-out way, each country is once a test case. For the smallest cluster from which a test case is taken, there are 2 countries left. This function gave the probability of 66% correct clustering when data were not scaled, and yet higher value, 84% when data were scaled, both of which are better than the result by k-means clustering, either supervised or unsupervised. Typically, larger (odd) k values are used since they give somewhat better results. Although k=1 is known to be unstable, and when every decision is made based on 1-nearest neighbour only, it may easily get wrong result. But for this case, the result is effective.

Figure 4 Fragmented map with 6 clusters



In these comparable tests, cluster 1 in the SOM results is the largest, including the most countries (19) and can be grouped into two sub-clusters. Cluster 2 of the SOM and cluster 3 can also form separate clusters in k-means clustering. Cluster 4, with the fewest countries, cannot form an independent cluster in k-means clustering with which the three countries were incorporated into cluster 1. After all, they all were within the same cluster built by k-mean clustering indicating to be of similar type. This also shows that subject to the present data the SOM was able to form their own cluster for these three countries, but k-means clustering failed in this respect. The small number of countries compared to those other clusters, only three, in this situation is a possible cause.

Considering these test results, it is possible to adjust clusters in the SOM with ViscoverySOMine. However, with this dataset, when adjusting the number of clusters, to 6, 7, or more, it did not produce a better clustering of those countries in original cluster 1. Instead, other clusters emerged from original clusters, and the map only became more and more fragmented. An example map of 6 clusters is given in Figure 4. From it, newly created clusters C5 and C6 are separated from other clusters and include one country each.

5.3 Correlations

Using Viscovery SOMine, a detailed table of correlations can be generated. The following presented are the roles of selected economic factors against different crime-related factors.

Even strong correlation between two attributes does not, of course, indicate necessarily that one is caused by the other. We have to remember this basic property of correlation.

Table 5 Correlations between economic attributes *A* and crime-related attributes: correlations from interval (-0.3,0.3) were left out seen as insignificant, from (-0.6,-0.3] and [0.3,0.6) were seen as interesting, and from [-1,-0.6] and [0.6,1] marked in Bold face as significant.

A	Crime-related attributes											
	15	16	17	18	19	20	21	22	23	24	25	26
1	-	-	-	-	-0.35	0.31	-	-	-	0.38	-0.49	-
2	-0.48	-	-	-	-0.56	-	-0.50	-0.30	0.37	-	-0.44	0.57
3	-0.31	-0.35	0.45	-	-0.46	-	**-0.67**	-	-	0.30	**-0.78**	0.34
4	-	-0.34	-	-	0.46	-0.38	-	0.38	-0.42	-0.45	-	-0.56
5	-0.30	-	-	-	-	-	-	-0.33	0.54	-	-	0.38
6	-	-	-	-	-	-	-	-	-	0.51	-	-
7	0.56	-0.49	-	-	**0.82**	-0.48	-	**0.71**	-0.48	-0.35	-	-0.34
8	-0.49	-	-	-	**-0.65**	-	-	-0.46	-	-	-	-
9	-0.51	-	-	-	**-0.75**	-	-	-0.50	-	-	-	-
10	0.38	-0.39	-	-	-	-	-0.56	0.40	-	-	-	0.50
11	-	-0.41	-	-	-	-	-0.33	0.35	-	-	**-0.64**	0.57
12	-0.52	-	-	-	-0.58	0.44	-0.33	-0.35	0.58	-	-	0.33
13	-	-0.54	-	-	-0.46	-	**-0.76**	-	-	-	-0.51	0.53
14	-0.51	-	-	-	-0.45	-	-	-0.47	-	-	-	-

6 Discussion

First of all, dataset in this paper underwent a process of selection by using ScatterCounter to identify strong and weak variables. Within the scope of this dataset and this research, results of the SOM can be similar before and after the attribute selection with ScatrerCounter. The ScatterCounter played an important role of validating the processing of the SOM, and gave a proof for valuable use of the SOM.

In processing of data, when attributes are numerous, results of clustering may become difficult since these might vary. That is to say, some of the countries might be clustered into wrong groups. So that it creates difficulties to give descriptions closer to the fact. Another aspect in the SOM is that, changes of parameters in processing the data, clusters and maps can also change. The SOM might be inferior to tranditional processing in this aspect in principle.

However, it has the advantage that straightforward computational comparisons, let alone a manual method cannot be used here. With them, it is just impossible to put 50 countries into 4 clusters as in the initial stage of this experiment according to 30 attributes in any reasonable time limit because of the huge number of different cluster combinations. Viz., according to (Truss, 1999, p. 204) for n elements set into k (non-empty) clusters there are $P(n,k)$ partitions or possibilities to make different collections of clusters here:

$$P(n,k) = \frac{1}{k!}\sum_{r=0}^{k}(-1)^r\binom{k}{r}(k-r)^n$$

Using our values n=50 and k=4, we obtain

$$\frac{1}{4!}\left(1\cdot 1\cdot 4^{50} - 1\cdot 4\cdot 3^{50} + 1\cdot 6\cdot 2^{50} - 1\cdot 4\cdot 1^{50} + 1\cdot 1\cdot 0^{50}\right) = \frac{1}{4\cdot 3!}\left(4^{50} - 4\cdot 3^{50} + 6\cdot 2^{50} - 4\right)$$

The right-hand side of the equation can be simplified and reordered as follows

$$= \frac{1}{3!}\left(4^{49} - 3^{50} - 1 + 3 \cdot 2^{49}\right)$$

and by removing the first three terms, the sum of which is positive since

$$4^{49} > 3^{50} + 1,$$

we can estimate with the last term that the entire sum is

$$> 2^{48} = 1024^4 \cdot 256 > 256 \cdot 10^{12}.$$

Traditionally, economically poor condition has been regarded to lead to crime. In this experiment, it was not clear. Both murder and prison population rates were negatively relevant with the GDP per capita.

Using Viscovery SOMine, correlation between every pair of attributes can also be identified. In this experiment, however, the results demonstrated that there are only few strong links between any pairwise attributes.

It must be pointed out that, correlation is a result that can usually be generated by using statistical methods. It has some importance for further consideration in identifying the causes. However, correlated factors might contain but are not equal to causes. Correlation expressed in numbers can happen to be wrong and it does not represent final analysis of the problem. In addition, since correlation (Pearson product-moment correlation coefficient) is of linear type, it cannot reveal more complicated relations between attributes.

Correlation's another paradox occurs when people in the society think something good, but statistically it positively correlates with total crime rate or a certain type of very serious offence. Or vice versa, can crime take place on a background where some very positive, favorable and good factors are located? The answer is positive. It is also positive that some negative, unfavorable and bad factors in actual fact bring about just less crime. Therefore, correlation produced by statistics must be assembled again in human mind.

Correlation at macroscopic level cannot be directly applied at microscopic level. The former does not make sense in the latter. For example, countries with highly developed economy usually have high level of crime. This must be explained with the assistance of the techniques but also beyond the techniques. One of the tricks is that, in most developed countries, such as the USA, have far stricter laws than in other countries. What are crimes in the USA might be so tiny, so trivial an act that they are not punishable in other parts of the world. Some developing countries have really fewer offences, because of their merits in cultural tradition (so that fewer offences were committed), or lack of investment in public security, or law being frequently ignored, or simply because of statistical defects (so that fewer offences were detected). All these should be studied by using various methods, computational as well experts' consideration, but beyond the simple number of correlation. It is impractical to make a clear-cut judgment.

7 Conclusions

In this study, economic information for 50 countries has been collected using the Internet as a source of information and an economic database has been created. A number of economic attributes or factors have been selected. Then, a data-mining tool, the self-organizing map, has been used to perform a benchmarking of crime situation in these countries. During this process, ScatterCounter was used to select the variables. By comparing the results generated by the SOM using these two datasets, it proved that the SOM get similar results. In addition, clusters generated by the SOM were further tested by *k*-means clustering and nearest neighbor searching. The results of the study provide further evidence that the self-organizing map is a feasible and effective tool for the study of crime. The results are easy to visualize and interpret, and provide a very practical way to compare the economic factors of countries with different crime situation. This paper has shown that the study of crime is an application area that can benefit from efficient data analysis and visualization techniques.

Acknowledgments

The first author is grateful to Tampere Doctoral Program in Information Science and Engineering (TISE) for support.

References

Abidogun, O. A (2005) 'Data mining, fraud detection and mobile telecommunications: call pattern analysis with unsupervised neural networks', PhD thesis, University of the Western Cape, South Africa.

Adderley, R. (2004) 'The use of data mining techniques in operational crime fighting', in *Proceedings of symposium on intelligence and security informatics*, No. 2, Tucson A.Z., ETATS-UNIS (10/06/2004), Vol. 3073, pp. 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*, Vol. 5, No. 4, pp. 265-276.

Adderley, R., Townsley, M. and Bond, J. (2006) 'Use of data mining techniques to model crime scene investigator performance', *Knowledge-Based Systems*, Vo. 20, No. 2, pp. 170-176.

Axelsson, S. (2005) 'Understanding intrusion detection through visualization', PhD thesis, Chalmers University of Technology, Göteborg, Sweden.

Brockett, P. L., Xia, X. and Derrig, R. A. (1998) 'Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud', *The Journal of Risk and Insurance*, Vol. 65, No. 2, pp. 245-274.

Chandra, D. K., Ravi, V., and Ravisankar, P. (2010) 'Support vector machine and wavelet neural network hybrid: application to bankruptcy prediction in banks', *Int. J. of Data Mining, Modelling and Management*, Vol.2, No.1, pp.1 – 21.

Chung, W., Chen, H., Chaboya, L. G., O'Toole, C. D. and Atabakhsh, H. (2005) 'Evaluating event visualization: a usability study of COPLINK spatio-temporal visualizer', *International Journal of Human-Computer Studies*, Vol. 62, No. 1, pp. 127-157.

Clinard, M. B. (1958) *Sociology of Deviant Behaviour*, Rinehart & Company.

Dahmane, M. and Meunier, J. (2005) 'Real-time video surveillance with self-organizing maps', in *Proceedings of the Second Canadian Conference on Computer and Robot Vision* (CRV'05), Washington, DC., pp. 136-143.

Deboeck, G. (2000) 'Self-organizing patterns in world poverty using multiple indicators of poverty repression and corruption', *Neural Network World*, Vol. 10, pp. 39-254.

Dittenbach, M., Merkl, D. and Rauber, A. (2000) 'The growing hierarchical self-organizing map', in *Proceedings of International Joint Conference on Neural Networks* (IJCNN 2000), Como, Italy, Vol. 6, pp. 15-19 .

Fei, B. K., Eloff, J. H., Olivier, M. S. and Venter, H. S. (2006) 'The use of self-organizing maps for anomalous behavior detection in a digital investigation', *Forensic Science International*, Vol. 162, No. 1-3, pp. 33-37.

Fei, B., Eloff, J., Venter, H. and Olivier, M. (2005) 'Exploring data generated by computer forensic tools with self-organising maps', in: *Proceedings of the IFIP Working Group 11.9 on Digital Forensics,* pp. 1-15.

Grosser, H., Britos, P. and García-Martínez, R. (2005) 'Detecting fraud in mobile telephony using neural networks', in Ali, M. and Esposito F. (Eds.): *IEA/AIE 2005*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Berlin, Germany, Vol. 3533, pp. 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. and Simula, O. (1999) 'A self-organizing map for clustering probabilistic models', *Artificial Neural Networks*, Vol. 470, pp. 946-951.

Juhola, M. and Siermala, M. (2012a) 'A scatter method for data and variable importance evaluation', *Integrated Computer-Aided Engineering*, Vol. 19, pp. 137-149.

Juhola, M. and Siermala, M. (2012b) ScatterCounter software via link: http://www.uta.fi/sis/cis/research_groups/darg/publications.html

Kangas, L. J. (2001) 'Artificial neural network system for classification of offenders in murder and rape cases', The National Institute of Justice, Finland.

Kohonen, T. (1997) *Self-Organizing Maps*. Springer-Verlag, New York, USA.

Kohonen, T. and Honkela, T. (2007) 'Kohonen network', *Scholarpedia*, Vol. 2, No. 1, p. 1568.

Koskela, M. (2003) 'Interactive image retrieval using self-organizing maps', PhD thesis, Helsinki University of Technology, Finland.

Kumar, S., Sural, S., Watve, A., and Pramanik, S. (2009) 'CNODE: clustering of set-valued non-ordered discrete data', *Int. J. of Data Mining, Modelling and Management*, Vol.1, No.3, pp.310 - 334.

Lampinen, T., Koivisto, H and Honkanen, T. (2005) 'Profiling network applications with fuzzy C-means and self-organizing maps', *Classification and Clustering for Knowledge Discovery*, Vol. 4, pp. 15-27.

Lee, S.-C. and Huang, M.-J. (2002) 'Applying AI technology and rough set theory for mining association rules to support crime management and fire-fighting resources allocation', *Journal of Information, Technology and Society*, Vol. 2, p. 65.

Lemaire, V. and Clérot, F. (2005) 'The many faces of a Kohonen map a case study: SOM-based clustering for on-line fraud behavior classification', *Classification and Clustering for Knowledge Discovery*, Vol. 4, pp. 1-13.

Leufven, C. (2006) 'Detecting SSH identity theft in HPC cluster environments using self-organizing maps', Master's thesis, Linköping University, Sweden.

Li, S.-T., Tsai, F.-C., Kuo, S.-C., Cheng, Y.-C. (2006) 'A knowledge discovery approach to supporting crime prevention', in *Proceedings of the Joint Conference on Information Sciences*, Taiwan, retrieved January 29, 2013, http://www.atlantis-press.com/php/download_paper.php?id=146

Ling, L., Li, JJ., Ng, M. K., Cheung, Y., and Huang, J. (2009) 'SMART: a subspace clustering algorithm that automatically identifies the appropriate number of clusters', *Int. J. of Data Mining, Modelling and Management*, Vol.1, No.2, pp.149 - 177.

Masahiro, T. (1996) 'Economic structure and crime: the case of Japan', *Journal of Social-Economics*, Vol. 5, No. 4, pp. 497-515.

Memon, Q. A. and Mehboob, S. (2006) 'Crime investigation and analysis using neural nets'. in *Proceedings of International Joint Conference on Neural Networks*,. Washington, DC., pp. 346-350.

Mower, E. R. (1942) *Disorganization, personal and social,* J. B. Lippincott Company, Philadelphia, USA.

Oatley, G. C., Ewart, B. W. and Zeleznikow, J. (2006) 'Decision support systems for police: lessons from the application of data mining techniques to "soft" forensic evidence', *Artificial Intelligence and Law*, Vol. 14, No. 1, pp. 35-100.

Rahman, S. M. M., Yu, X., and Siddiky, F. A. (2011) 'An unsupervised neural network approach to predictive data mining', *Int. J. of Data Mining, Modelling and Management*, Vol.3, No.1, pp.1 – 17.

Rock, P. (1994) *History of Criminology*, Dartmouth Publishing, Aldershot, UK.

Scime, A., Murray, G. R., and Hunter, L. Y. (2010) 'Testing terrorism theory with data mining', Int. J. of Data Analysis Techniques and Strategies, 2010 Vol.2, No.2, pp.122 - 139.

Situngkir, H. 2003. 'Merging the emergence sociology: the philosophical framework of agent-based social studies', *Journal of Social Complexity,* retrieved January 29, 2013, http://cogprints.org/3519/1/Soc.pdf

Soares, R. R. (2004) 'Development, crime and punishment: accounting for the international differences in crime rates', *Journal of Development Economics*, Vol. 73, pp. 155– 184.

Taft, D. R. (1950) *Criminology: a Cultural Interpretation,* The MacMillan Company, New York, USA.

The Community Safety and Crime Prevention Council 1996, The root causes of crime - CS&CPC statement on the root causes of crime.

Truss, J. K. (1999) *Discrete Mathematics for Computer Scientists*, Pearson Education Limited, Harlow, England.

Viscovery Software GmbH (2012) Viscovery SOMine, retrieved January 29, 2013, http://www.viscovery.net/somine/

Wang, J., Hu, X., and Zhu D. (2007) 'Diminishing downsides of Data Mining', International Journal of Business Intelligence and Data Mining, Vol. 2, No. 2, pp. 177-196.

Zaslavsky, V. and Strizhak, A. (2006) 'Credit card fraud detection using self-organizing maps', *Information and Security: An International Journal*, Vol. 18, pp. 48-63.

PUBLICATION IV

Application of the Self-Organizing Map to Visualization of and Exploration into Historical Development of Criminal Phenomena in the United States, 1960-2007

Xingan Li and Martti Juhola

Application of the Self-Organizing Map to Visualization of and Exploration into Historical Development of Criminal Phenomena in the United States, 1960-2007

Xingan Li and Martti Juhola*

Computer Science, School of Information Sciences, 33014 University of Tampere, Finland

E-mail: Xingan.Li@uta.fi, Martti.Juhola@sis.uta.fi

Abstract: Underneath the prevalence of criminal phenomena, many variables can be used to describe the background data such as the historical development of crime against socio-economic development. With large amount of data and evolution of data processing, multi-dimensional analysis becomes possible. Based on longitudinal (1960-2007), crime and socio-economic data (22 variables), we used the self-organizing map (SOM) for development of criminal phenomena in the United States. Classification power of variables was evaluated and, e.g., *k*-means clustering were used for obtaining comparable results. After initial processing of the data with the SOM, 6 clusters of years were identified. We show how the SOM is applied to analyzing criminal phenomena over a span of several decades. Results proved that, after the evaluation of variables for classification, and validation with *k*-means clustering, nearest neighbor searching, decision trees, and logistic discriminant analysis, SOM can be a new tool for mapping criminal phenomena processing multivariate data.

Keywords: development of criminal phenomena; data mining; self-organizing map; variable evaluation; *k*-means clustering; nearest neighbor searching; decision trees; logistic discriminant analysis; the United States

Bibliographic notes: Xingan Li took his B. and M. Laws in China in 1989 and 1994. In 2008 he took his Dr.Laws degree at the University of Turku, Finland. He has been lecturer and researcher in Chinese and Finnish Universities. His research topics are criminology and artificial intelligence.

Martti Juhola took his B.Sc., M.Sc. and Ph.D. degrees at the University of Turku, Finland, in the 1980s. From 1992 he was professor of computer science at the University of Kuopio and from 1997 he is at the University of Tampere, Finland. His research interests cover data mining, machine learning, signal analysis and artificial intelligence, particularly applied to medical and other areas.

1 Introduction

Exploration into criminal phenomena must be relied on examining extracurricular factors, such as multiple adverse personal, social, economic, cultural and family conditions. The prerequisite for crime prevention is a good understanding of its reasons (The Community Safety and Crime Prevention Council, 1996). For centuries, attempts have been made in the study of crime to reveal causes of crime and seek ultimate solutions. However, no solely workable theory has thus far been invented to provide any precise answer for tackling crime, though many theorists presented many persuasive suggestions (Rock, 1994). It turned out that the study of crime, which is dealing with a social phenomenon, hardly has a perfect solution. This study is not to search for a new solution but to test a new method for identifying factors that are important in seeking potential solutions, either in a short-term or in a long-term basis.

Causational factors of crime have puzzled societies over years. Criminologists have long wanted to divulge causational and correlation factors of crime with abundant hypotheses, observations, comparison and conclusion, if not in vain. Law enforcement has the motivation for recognizing developmental tendencies of crime, such as its characteristic change in either microscopic or macroscopic aspects. Legislators have acquired the power from people to make effective law to eliminate, prevent or reduce crime. Governments are in need of making feasible policy to combat crime and assist victims. Victims make up their mind to be socially rehabilitated to a peaceful, safe and harmonious life. The general public is curious of creating, enjoying, and maintaining a society ruled by

* Corresponding author. Fax +358 3 2191001, tel. +358 40 1901716

law. The international society is committed to coordinating and cooperating in reckoning with transnational crime. All these tasks are to be realized through various activities, in which the study of crime occupies a significant position. Processing crime data has been a basis for knowledge-detection and decision-making in this field.

Conventionally, the study of crime relied heavily on both qualitative and quantitative methods. Publicly available national statistics facilitate quantitative analysis of criminal phenomena, even though a variety of analytical instruments have been employed in different fields of issues. The information required for the research can normally be found in databases of official publications, including those of international organizations, national statistical and judicial agencies, and other official documents. With the pervasive use of the computing facilities, the large amount of data available for analysis poses great challenges to analytical capacity. Comparison and analysis in traditional ways become complicated and time-consuming.

As a development in research methods, data-mining tools, either supervised or unsupervised, make it possible to discover hidden correlation factors in data, either structured or unstructured. Applied in many fields, the self-organizing map is one of such feasible tools. In fact, the SOM has been applied in, though only a small number of, studies of crime, among which, a high frequency of application of the SOM can be perceived concentrated on the detection of some types of crimes. These can be seen as microscopically oriented studies, compared with which the domain of macroscopically mapping crime has still been absent from thousands of pieces of publicly available literature.

This study applies the SOM to the field of macroscopically exploring into multi-dimensional data of development of criminal phenomena, aiming at seeking an innovative field in which artificial intelligence can play a role in simplifying the analysis.

Following this introduction, Section 2 presents the methodology used in this study. Section 3 gives a brief retrospect over development of criminal phenomena in the US history over the latter part of 20th century. Section 4 presents data set used in the experiment. Subsequently, the construction of the map including clustering map and feature planes are demonstrated in Section 5. Section 6 describes experimental results by showing identified clusters on the map, and describing correlations between crime and environmental factors. The paper proceeds to discussion based on the experiments and findings and concludes with feasibility of application of the SOM in time-series analysis of correlates of crime. Section 7 concludes the whole article.

2 Methodology

Data analysis has been applied in a broad range of social research (for instance, Lozano and Gutierrez 2008, and many other studies). Knowledge discovery and Decision-making in different fields can benefit from different data mining methods (Priya, Vadivel and Thakur 2012). The idea of applying the SOM to the study of macroscopic criminal phenomena is innovative, even though some literature provided some preliminary exploration into explanation for thinking self-organizing methods as feasible to do research on society as a whole. The vast computational technology provides us with the possibility to establish the sociology to cope with any sociological emergence phenomena (Situngkir, 2003).

The study of crime has been in need of an innovative method to dealing with increasingly large amount of data on one hand, and rapid development of data mining techniques showed their potentiality on the other hand (for instance, Abidogun 2005; Chung et al. 2005). They can support police activities by profiling single and series of crimes or offenders, and matching and predicting crimes (Oatley et al., 2006).

Developed by Kohonen (Kohonen, 1997) to cluster and visualize data, the SOM is an unsupervised learning mechanism that clusters objects having multi-dimensional attributes into a lower-dimensional space, in which the distance between every pair of objects captures the multi-attribute similarity between them. Some applications based on the concept of the SOM were developed particularly to meet the demand of law enforcement (for example, Fei et al., 2005; Fei et al., 2006; Lemaire and Clérot, 2005). In particular, even though the data on the storage media may contain implicit knowledge that could improve the quality of decisions in an investigation, when large volumes of data are processed, it consumes an enormous amount of time (Fei et al., 2005). The SOM may play a positive role in exploratory data analysis (Lemaire and Clérot, 2005).

Three categories of the network architectures and signal processes have been in use to model nervous systems. The first category is feed-forward networks, which transform sets of input signals into sets of output signals, usually determined by external, supervised adjustment of the system parameters. The second category is feed-back networks, in which the input information defines the initial activity state of a feedback system, and after state transitions the asymptotic final state is identified as the outcome of the computation, and the third category is self-organizing networks, in which neighboring cells in a neural network compete in their activities by means of mutual lateral interactions, and develop adaptively into specific detectors of different signal patterns (Kohonen 1990, p. 1464).

The SOM can be sketched as an input layer and an output layer constituting two-layer neural networks. An unsupervised learning method is used in the SOM. The network freely organizes itself according to similarities in the data, resulting in a map representing the data input.

The SOM algorithm operates in two steps, which are initiated for each sample in the data set. The first step is designed to find the best matching node to the input vector, which is determined using some form of distance function, for example, the smallest Euclidian distance function. Upon finding the best match, the second step is initiated, the "learning step", in which the network surrounding node c is adjusted towards the input data vector. Nodes within a specified geometric distance, h_{ci}, will activate each other, and learn something from the same input vector $\mathbf{x}$. The number of nodes affected depends upon the type of lattice and the neighborhood function. This learning process of vector $\mathbf{m}$ of node i can be defined as (Kohonen 1997, p.87):

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)(\mathbf{x}(t) - \mathbf{m}_i(t)). \quad (1)$$

The function $h_{ci}(t)$ is the neighborhood of the winning neuron c, and acts as a smoothing kernel defined over the lattice points. The function $h_{ci}(t)$ can be defined in two ways, either as a neighborhood set of arrays around node c or as a Gaussian function (Kohonen 1997, p. 87). In the training process, weight vectors are mapped randomly onto a two-dimensional, hexagonal lattice. A fully trained network facilitates a number of groups.

The SOM algorithm results in a map exhibiting the clusters of data, using dark shades to demonstrate large distances and light shades to demonstrate small distances (U-matrix method) (Kohonen, 1997). Feature planes, which are single vector level maps, can additionally be generated to discover the characteristics of the clusters on the U-matrix map. They present the distribution of individual columns of data. (Features are also called attributes or variables.)

Upon processing the data, maps can be generated using software packages. By observing and comparing the clustering map and feature planes, rough correlation between different attributes can be identified. Detailed correlation table can also be realized automatically, e.g., with Viscovery SOMine (Viscovery Software GmbH, 2012), which adopts the correlation coefficient scale ranging from -1.0 to +1.0. These clustering maps, feature planes and correlation tables provide major basis for further analysis.

Crime is one of the social problems attracting the most attention, research of which can borrow ideas from generic or neighboring subjects. A couple of applications of the SOM to social research can help frame the study of crime. In practice, the SOM is one of the models of neural networks that acquire growing interests in social research. Deboeck (2000) clustered world poverty into convergence and divergence in poverty structures using the SOM, with a view to implement poverty reduction strategies through artificial neural networks. Crime-related social phenomena have also been studied with this method, such as corruption (Huysmans et al., 2006), public security (Li et al., 2006), and crime management and fire-fighting (Lee and Huang, 2002). Findings of many such studies proved that artificial neural networks are useful tools in social research.

Naturally, criminological research in detailed offences from micro viewpoints has also been acquiring more assistance from application of artificial intelligence. Hitherto, a great many researchers focus on applications of artificial neural networks to law enforcement, in particular, the detection of specific abnormal or criminal behaviours. Adderley et al. (2007) examined how data-mining techniques can support the monitoring of crime scene investigator performance. Oatley et al. (2006) presented a discussion of data mining and decision support technologies for police, considering the range of computer science technologies that are available to assist police activities. Dahmane et al. (2005) have presented the SOM for detecting suspicious events in a scene.

The SOM has acquired more attention, for instance, in detection of credit card fraud (Zaslavsky and Strizhak, 2006), automobile bodily injury insurance fraud (Brockett et al., 1998), burglary (Adderley and Musgrove, 2003; Adderley, 2004), murder and rape (Kangas, 2001), homicide (Memon and Mehboob, 2006), network intrusion (Rhodes et al., 2000; Leufven, 2006; Lampinen et al., 2005; Axelsson 2005), cybercrime (Fei et al., 2005; Fei et al., 2006), mobile communications fraud (Hollmén et al., 1999; Hollmén 2000; Grosser et al., 2005). Literature in this aspect can be stated as abundant, but this is the only field where the SOM has found application to crime-related research before.

Besides crime detection, neural networks are also used in research specialized in victimization detection in mobile communications fraud (Hollmén et al., 1999).

From present literature, much can be found on applications of the SOM to crime detection, a field relating to law enforcement. In real sense of application of the SOM to the study of crime, that is, to identifying causational or correlation factors or to recognizing preventive or deterrent factors, rare has been published. The current

situation created a motivation for designing experiments exploiting this approach, in comparison with other methods.

This study applies the SOM to historical development of crime in the United States during 48 years (1960-2007). Including an analysis based on available data, the results of the study will revolve around whether the SOM can be feasible a tool for mapping criminal phenomena through processing of large amounts of historical crime data.

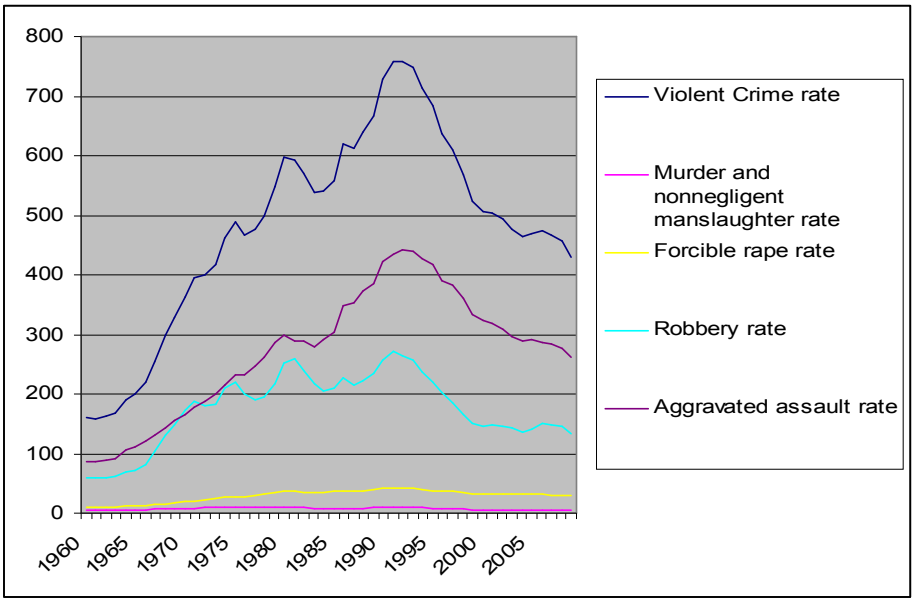
During the application of the SOM, in order to select variables, the method called ScatterCounter (Juhola and Siermala 2012a, 2012b) was used to generate separation powers of variables. It means to identify which variables are strong in clustering or classification, and which are weak. The weak ones will be removed from the data set and the selected data set will be used in final processing and analysis. Certainly, if all the variables have satisfactory separation powers, they can be all reserved in the final test and analysis. In using ScatterCounter, missing data in the original data set have to be filled with real values. In this research, missing values were filled by the medians of the available values of the variables in the same clusters.

Besides the SOM, *k*-means clustering, nearest neighbor searching, decision trees, and logistic discriminant analysis were used to validate the clusters and analysis by calculating how accurately *k*-means clustering, nearest neighbor searching and decision trees methods put the same years into the same clusters as the SOM does.

3 Development of criminal phenomena in the United States

The United States has for decades been described as a country full of violence and killing among the most industrialized countries. It completely depends on what the reference groups are. 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 Germany, Denmark, Norway, Japan, and Singapore, with homicide rate below 1 per 100,000 inhabitants. Compared with the figures of 50 in Sierra Leone, and more than 45 in El Salvador, Jamaica, Venezuela, Guatemala, and Honduras, it is also true that the U. S. also has a low level of homicide rate, 5.8 per 100,000 people.

Figure 1 Violent crime rate per 100,000 people in the US 1960-2009. Sources: Chart drawn according to FBI, Uniform Crime Reports as prepared by the National Archive of Criminal Justice Data

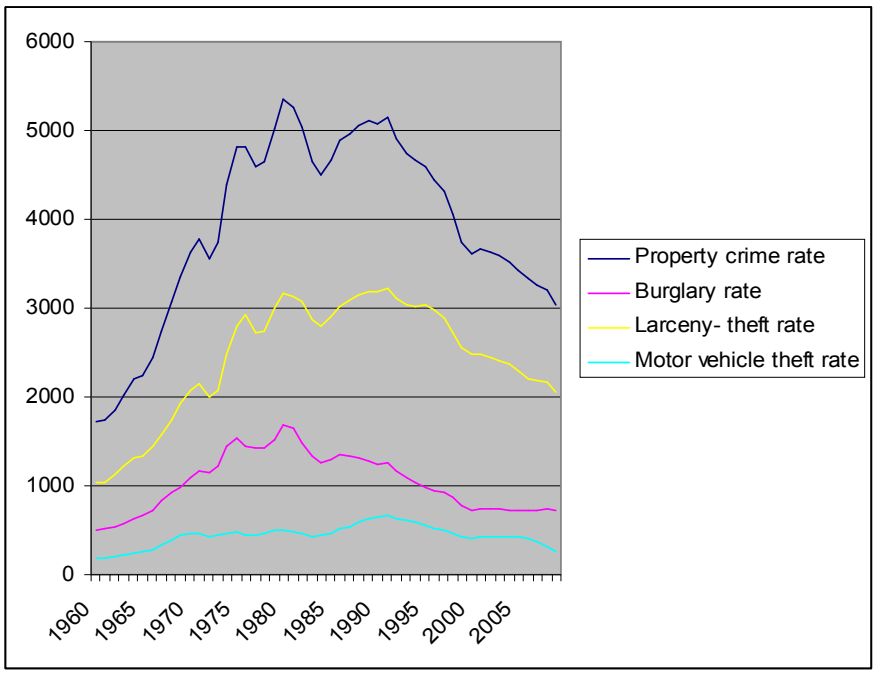


In fact, criminal phenomena in the United States have a dramatic rise and fall in the latter part of the 20th century. According to the United States Department of Justice, Bureau of Justice Statistics, the overall crime rate in the United States began its rise as early as in 1962, since when the 1961 low point has never been reached again. The 1970s and 1980s witnessed an interminable increase of crime rate. After it reached the 1991 peak, the crime rate began to turn down from 1992. In 2007, the United States' crime rate fell back to the 1974 level (Bureau of Justice Statistics, 2012). Both violent crime and property crime have the similar tendency, as that are

shown in Figure 1 and Figure 2 separately. Whenever violent crime rises property crime rises at the same rhythm, and vice versa.

During the last quarter of the 20th century, the US was (imaged by her perplex criminal situation as) one of countries with higher crime rates within the economically developed world. However, after decades of exploration into the paradox of sharp rise of crime accompanying the rapid increase of economy, people began to enjoy sharp fall of crime while suffering from economic declination. According to Wadsworth (2010), many studies (such as Blumstein and Wallman, 2000; Conklin, 2003; Zimring, 2007) have explored a variety of explanations for the sharp and continuous drop in crime. The most outstanding proposition has focused on the increased use of imprisonment, changes in the age distribution, changing drug markets, the availability of weapons, economic development, new security strategies, and the legalization of abortion (Wadsworth 2010, p. 533). However, as other research on social phenomena, no precise conclusion can be drawn nor confirmative reasons can be given to explain either the rise or the fall of crime. In this article, the topic concerning the rise and fall of American crime will be examined from a new stand using a new method, the self-organizing map.

Figure 2 Property crime rate per 100,000 people in the US 1960-2009. Sources: Chart drawn according to FBI, Uniform Crime Reports as prepared by the National Archive of Criminal Justice Data



4 The data set

4.1 Selection and codification of variables

The data set was retrieved from different online sources. Twenty-two variables covering demographic and economic situation were selected to model the state of the socio-economic system for the period of 48 years (1960-2007). The emphasis in the selection of variables in this study is on demographic and economic data and rates of different crimes. The selected variables are listed in Table 1, in which all the rates of crime were calculated on the base of 100,000 people. All the other rates were expressed as percentage.

There has not been a standard codification method in use for shortening names of variables. This study employs its own coding system. Codifications were realized by capitalizing the combined six letters comprised of each of the first three letters of the first two words in the names of the items. For example, “Property crime rate” was codified as PROCRI (PRO from property, and CRI from crime).

In addition, the years are codified as one letter plus two digits, such as U60 (U from US, the two-digit code of the United States, and 60 from the year 1960). Because there are only 48 different years, three-letter codes can well be shown in maps of reasonable sizes.

In total, the data set used consisted of 48 rows (years) and 22 columns (variables). Some other possibly important variables had to be discarded due to the substantial amount of data missing. As a result, missing values constitute only under 0.2% in this data set.

Table 1 Variables used in the creation of the SOM

Variables	*Code*	*Sources*
1. Aggravated assault rate	AGGASS	2
2. Burglary rate	BUGRAT	2
3. Population growth rate	POPGRO	4
4. GDP growth rate	GDPGRO	4
5. Federal Government Deficit	FEDGOV	4
6. Real GDP growth rate	RGDPGR	4
7. Nominal GDP growth rate	NGDPGR	4
8. Exports of goods and services	EXPGOO	4
9. Imports of goods and services	IMPGOO	4
10. Government consumption expenditures and investment	GOVCON	4
11. Personal consumption expenditures	PERCON	4
12. Employed civilian labour percent of population	EMPCIV	1
13. Forcible rape rate	FORRAP	2
14. Civilian labour force percent of population	CIVLAB	1
15. Larceny-theft rate	LARRAT	2
16. Motor vehicle theft rate	MOTVEH	2
17. Murder and non-negligent manslaughter rate	MURAND	2
18. Property crime rate	PROCRI	2
19. Robbery rate	ROBRAT	2
20. Total fertility rate	TOTFER	3
21. Unemployed civilian labour percent of population	UNECIV	1
22. Violent crime rate	VIOCRI	2

Sources: 1. United States Department of Labour, Bureau of Labour Statistics, http://www.bls.gov/cps/cpsaat1.pdf
2. United States Crime Rates 1960 – 2007, http://www.disastercenter.com/crime/uscrime.htm
3. Martin, J. A., Hamilton, B. E., Sutton, P.D., Ventura S.J., Menacker, F., Kirmeyer, S. and Mathews, T. J. (2006) Births: Final Data for 2006. National Vital Statistics Reports 57(7):29.
4. United States Census Bureau (2012) The 2012 Statistical Abstract: Historical Statistics, http://www.census.gov/compendia/statab/hist_stats.html

4.2 Evaluation of separation power of variables in the data set

After the data set was established for processing, Viscovery SOMine was used for clustering. Upon initial clusters were identified, the structure of data set was modified to be processed with ScatterCounter (Juhola and Siermala 2012a, 2012b). The missing data values were replaced with medians computed from pertinent clusters so that the completed data set could be processed by ScatterCounter. A main characteristic is that these years are labeled by cluster identifiers given by the preliminary SOM runs with the original 22 variables (attributes, as used in Viscovery SOMine).

The objective of ScatterCounter is to evaluate how much subsets labeled as classes (clusters given by SOM) differ from each other in a data set. Its principle is to start from a random instance of a data set and to traverse all instances by searching for the nearest neighbor of the current instance, then to update the one found to be the current instance, and iterate the whole data set 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 data set 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 data set applied. Since the process includes randomized steps, it is repeated from 5 to 10 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 2012a, 2012b). 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.

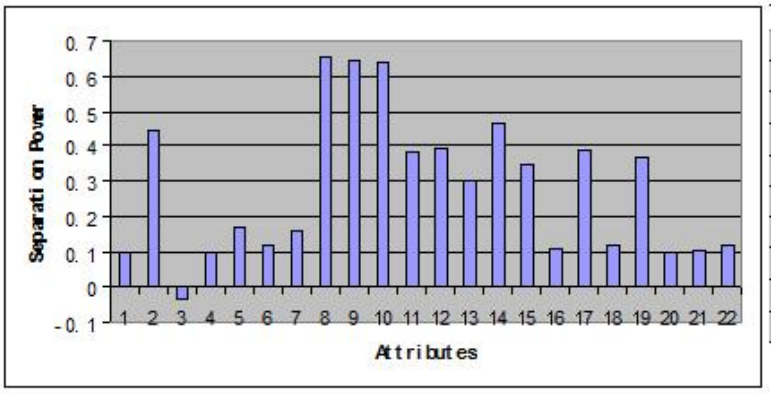
With these results and observations, in this data set, almost all have certain level of positive separation powers (Figure 3) and are kept in the data set used in the following experiments and analysis. Unlike in some other experiments with different data sets where some attributes are due to be removed, this data set reserves intact after evaluation of separation power.

Table 2 Descriptive statistics

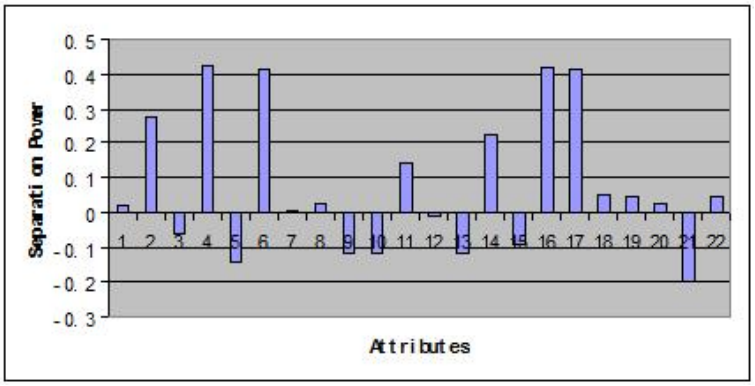
Attribute	Mean	Std. Deviation	Minimum	Maximum	Missing Values
Attribute 1	270.4	106.7	85.7	441.8	0 (0.00%)
Attribute 2	1050	330	509	1684	0 (0.00%)
Attribute 3	1.063	0.245	1	2	0 (0.00%)
Attribute 4	3.383	1.929	-2	7	1 (2.08%)
Attribute 5	-2.035	1.854	-5.876	2.406	0 (0.00%)
Attribute 6	3.305	1.997	-1.935	7.187	0 (0.00%)
Attribute 7	7.16	2.58	3.17	12.99	0 (0.00%)
Attribute 8	0.0683	0.0278	0.0352	0.1237	0 (0.00%)
Attribute 9	0.0854	0.0401	0.0401	0.1712	0 (0.00%)
Attribute 10	0.2236	0.0376	0.1745	0.2938	0 (0.00%)
Attribute 11	0.6658	0.0244	0.6242	0.7161	0 (0.00%)
Attribute 12	59.88	2.92	55.4	64.4	0 (0.00%)
Attribute 13	28.86	10.11	9.4	42.8	0 (0.00%)
Attribute 14	63.61	3.03	58.7	67.1	0 (0.00%)
Attribute 15	2453	663	1035	3229	0 (0.00%)
Attribute 16	443.1	117.5	183	658.9	0 (0.00%)
Attribute 17	7.48	1.78	4.6	10.2	0 (0.00%)
Attribute 18	3946	1037	1726	5353	0 (0.00%)
Attribute 19	175.8	60.9	58.3	272.7	0 (0.00%)
Attribute 20	2.186	0.509	1.74	3.65	1 (2.08%)
Attribute 21	5.842	1.424	3.5	9.7	0 (0.00%)
Attribute 22	482.6	169.9	158.1	758.1	0 (0.00%)

Figure 3 Separation powers of all 22 attributes in: (a) Cluster 1, (b) Cluster 2, (c) Cluster 3, (d) Cluster 4, (e) Cluster 5, (f) Cluster 6, and (g) whole data set

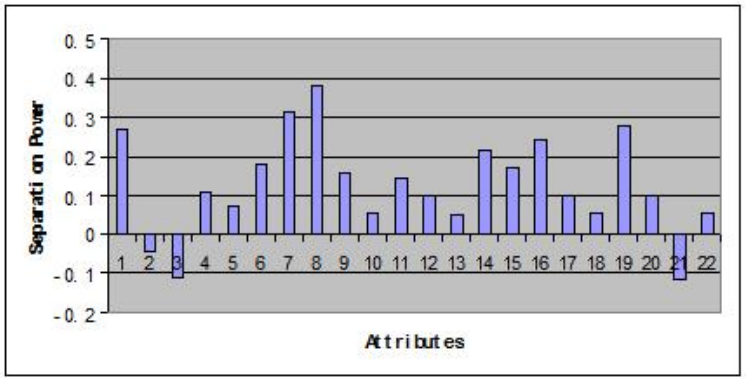
(a) Cluster 1



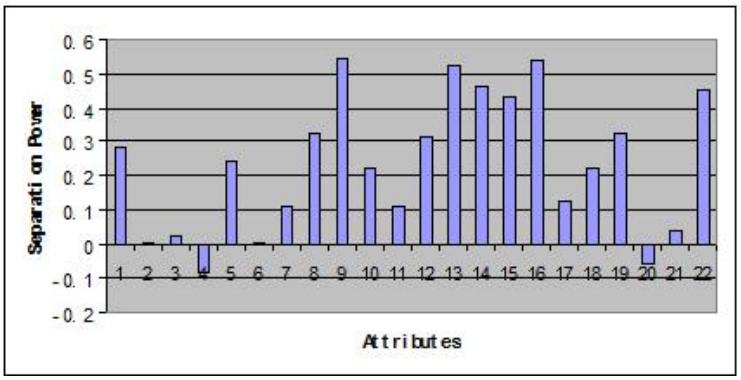
(b) Cluster 2



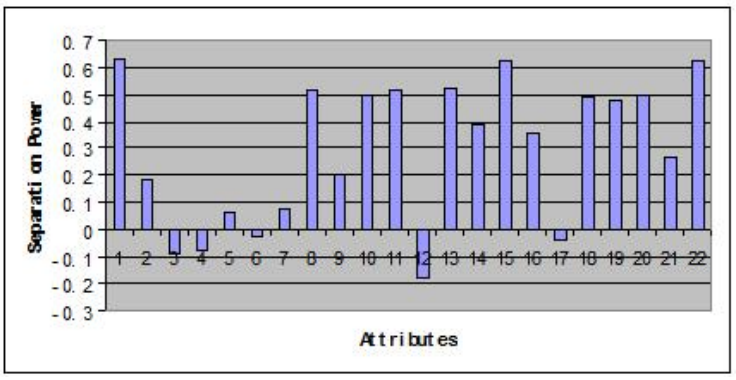
(c) Cluster 3



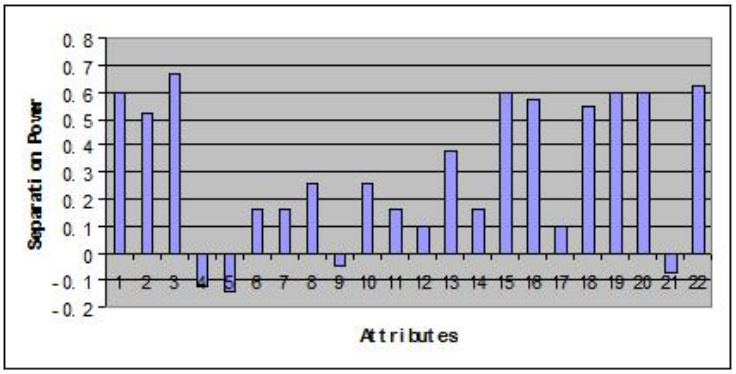
(d) Cluster 4



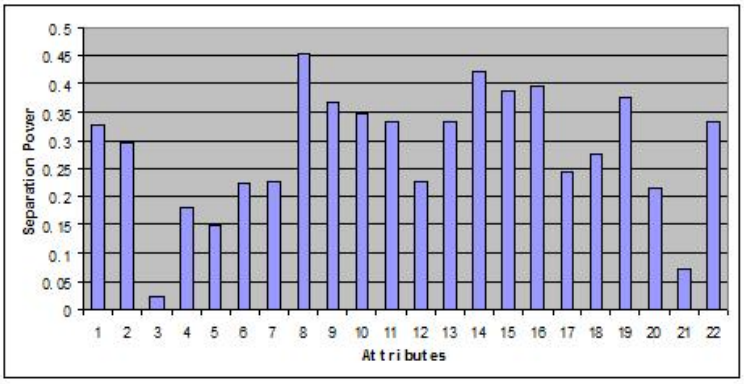
(e) Cluster 5



(f) Cluster 6



(g) Whole data set



5 Construction of the SOM

There are several different SOM software packages available for academic use. The software used for creating the SOM in this experiment was Viscovery SOMine Ver. 5.2 (Viscovery 2012). When creating the map (Figure 4), a network size of 2000 nodes and a tensile value of 0.5 were applied. Six clusters were determined to provide sufficient detail for our purposes and labeled as C1, C2, …C6, in each of which years are also labeled with the predefined codes. The distribution of the years into six clusters was fairly insensitive to changes of the numbers of nodes. While using 500, 1000, 10,000 or 20,000 nodes there were no alteration compared with Figure 4. Using 150 or 200 nodes, a few of these were located in different clusters. However, for 80 or 100 nodes giving the same distribution, one year (U95) only was in a different cluster (C1) compared with Figure 4.

Feature plane representation is used to visualize the relative component distributions of the input data. In feature representation, cold values represent relatively small values while warm values represent relatively large values. Similar outlook of planes typically indicates their correlation (Hollmen, 1996). In Figure 4, each attribute, the name used by Viscovery SOMine, represents one variable listed in Table 1. Each of them contains codes of all the 48 years as distributed in 6 clusters in Figure 3.

As a default rule in self-organizing maps, values are expressed in colors: warm colors (light in gray levels) denote high values, while cold colors (dark in gray levels) denote low values. In each of the feature planes (Figure 4), there is a color bar indicator depicting cold colors from the left and warm colors to the right. In the clustering map (Figure 3), there is not a color bar indicator, but in default, it applies the same as in the feature planes. However, order of clusters is not made according to colors, but according to the size of clusters (measured by the number of years included in each cluster). That is to say, the cluster with the most years is numbered as C1, cluster with the second most years is numbered as C2, and so on.

Figure 4 The clustering map of six clusters

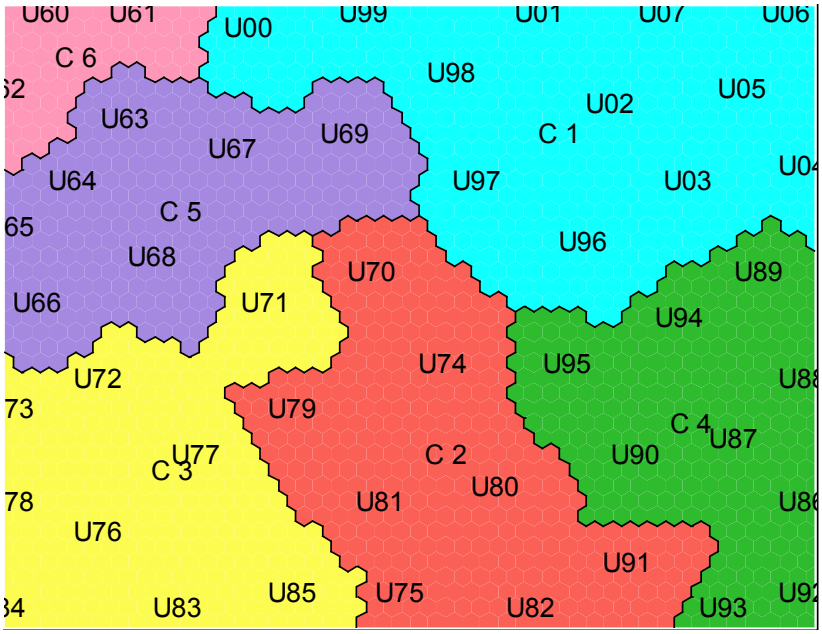
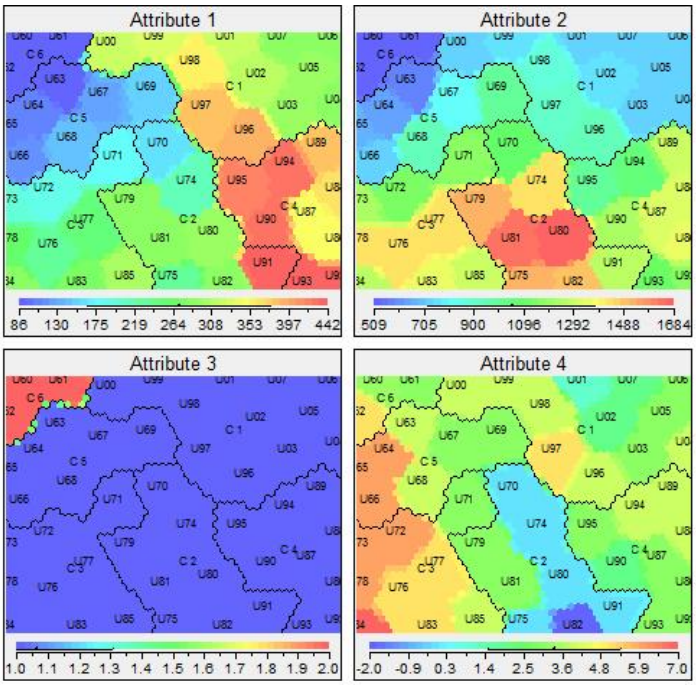
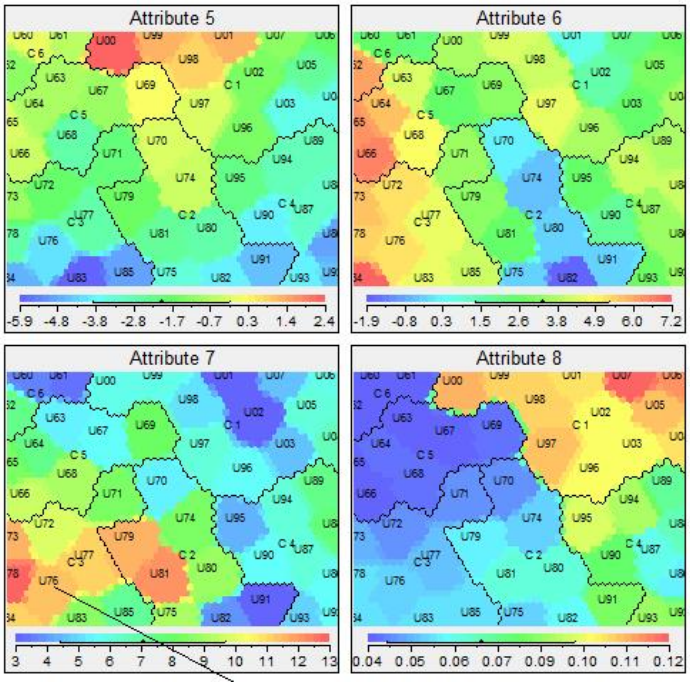


Figure 5 (a)-(e) The feature planes of the map

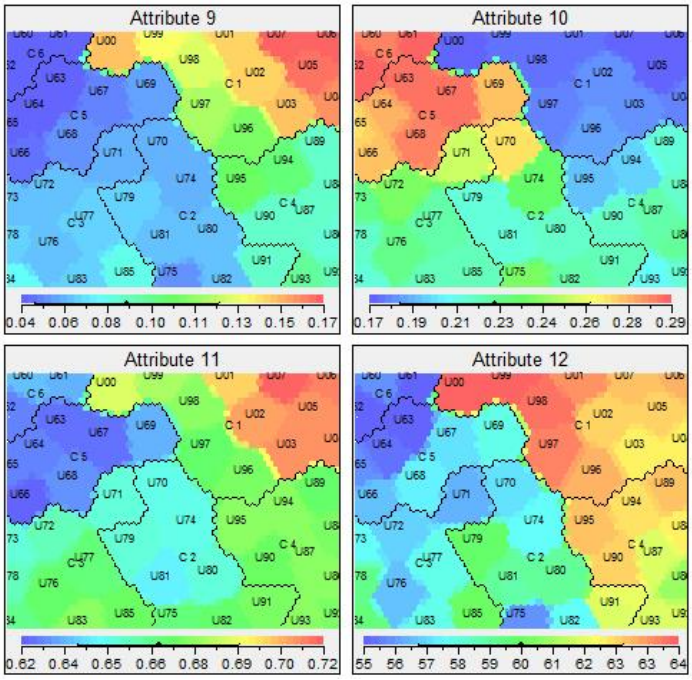
(a)



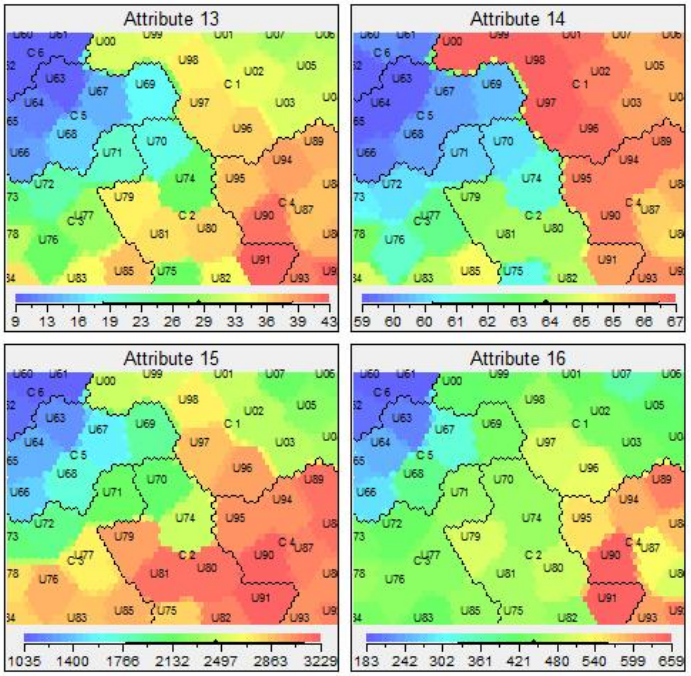
(b)


Attribute 5
-5.9 -4.8 -3.8 -2.8 -1.7 -0.7 0.3 1.4 2.4
Attribute 6
-1.9 -0.8 0.3 1.5 2.6 3.8 4.9 6.0 7.2
Attribute 7
3 4 5 6 7 8 9 10 11 12 13
Attribute 8
0.04 0.05 0.06 0.07 0.09 0.10 0.11 0.12

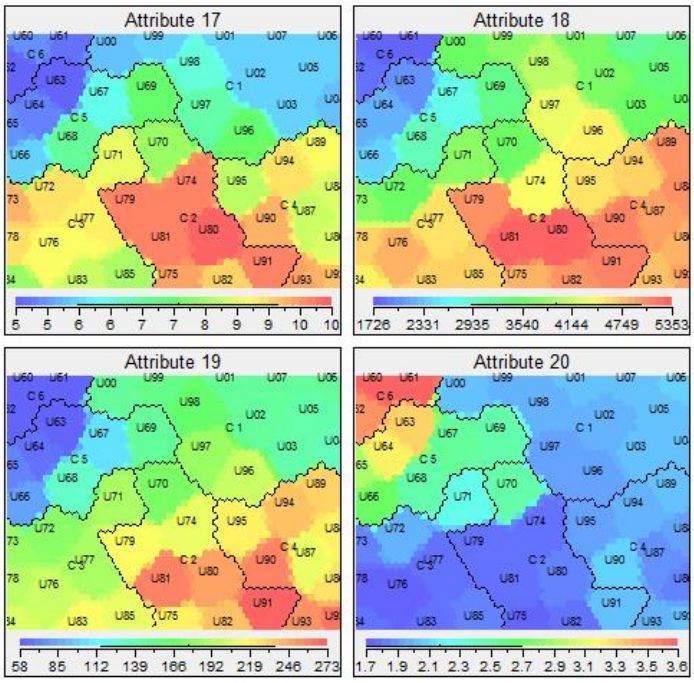
(c)


Attribute 9
0.04 0.06 0.08 0.10 0.11 0.13 0.15 0.17
Attribute 10
0.17 0.19 0.21 0.23 0.24 0.26 0.28 0.29
Attribute 11
0.62 0.64 0.65 0.66 0.68 0.69 0.70 0.72
Attribute 12
55 56 57 58 59 60 61 62 63 63 64

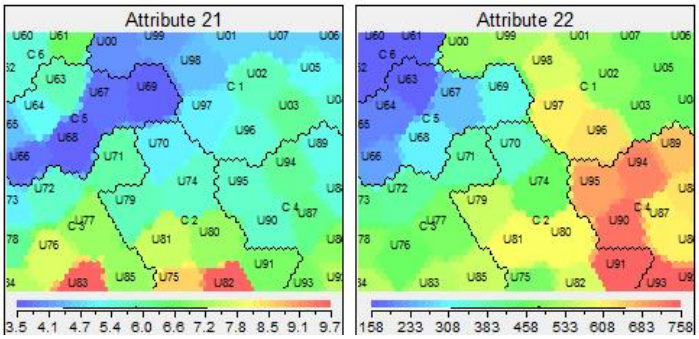
(d)


Attribute 13
9 13 16 19 23 26 29 33 36 39 43
Attribute 14
59 60 60 61 62 63 64 65 65 66 67
Attribute 15
1035 1400 1766 2132 2497 2863 3229
Attribute 16
183 242 302 361 421 480 540 599 659

(e)


Attribute 17
5 5 6 6 7 7 8 9 9 10 10
Attribute 18
1726 2331 2935 3540 4144 4749 5353
Attribute 19
58 85 112 139 166 192 219 246 273
Attribute 20
1.7 1.9 2.1 2.3 2.5 2.7 2.9 3.1 3.3 3.5 3.6

(f)


Attribute 21
3.5 4.1 4.7 5.4 6.0 6.6 7.2 7.8 8.5 9.1 9.7
Attribute 22
158 233 308 383 458 533 608 683 758

6 Clusters and correlations

6.1 Clusters

Using Viscovery SOMine, the clustering map can be tuned as in Figure 3. The number of clusters can be freely adjusted according to the needs of the study. In this study, considering the scale of the data set, six clusters were selected. The final map and the component planes are both generated based on six clusters, so that they are consistent when analyzing the characteristics of each cluster, or each attribute. Combining the two figures, broad and deep analysis can be carried out. In these figures, values of variables are depicted with colours: warm colours represent high values, while cool colour low values.

Table 3 Clusters C1-C6 including years 1960-2007 encoded as U60-U07

C 1: U99 , U01 , U07 , U06 , U00 , U98 , U05, U02, U97, U03, U04, U96
C 2: U70 , U74 , U79 , U80 , U81 , U91 , U75, U82
C 3: U71 , U72 , U73 , U77 , U78 , U76 , U85, U83, U84
C 4: U89 , U94 , U95 , U88 , U87 , U90 , U86, U93, U92
C 5: U63 , U69 , U67 , U64 , U65 , U68 , U66
C 6: U60 , U61 , U62

Forty-eight years, each coded with three letters, are grouped into 6 clusters, each of which representing different crime, demographic and economic. Based on Table 3, with reference to distribution of colors in feature planes in Figure 3, the characteristics of each cluster can be summarized.

Cluster 1 contains the late 1990s and early 2000s. The crime rate in the US has been decreasing steadily.

Cluster 2 deals with some years of the early 1970s, the end of the 1970s, and the early 1980s. The crime rate in the US kept rising sharply in the 1970s.

Cluster 3 includes the early 1970s, late 1970s and mid-1980s. During these years, the crime rate was still in its increasing route, but in some years such as 1975, 1980, and 1982, there were slight falls.

Cluster 4 begins from the mid-1980s, through the late 1980s and enters the early 1990s. The crime rate rose throughout the 1980s, reached its peak in 1993 and then began to decrease throughout the 1990s.

Cluster 5 covers the mid 1960s through the late 1960s. The crime rate in the US had risen sharply since the late 1960s.

Cluster 6 includes the early 1960s. The crime rate in the US was at its lowest point during the last 5 decades.

6.2 Validation of clusters with k-*means clustering, nearest neighbor searching and decision trees*

The results by the SOM were tested by *k*-means clustering methods with Euclidean distances. First, we used *k*-means to compute $k \geq 6$ clusters in the unsupervised manner and compared groups of the SOM results to the clusters formed. Here *k*-means clustering methods were used both with data not scaled and scaled. With *k*-means clustering, when data were not scaled, the probability of putting into the same clusters as generated by the SOM was 65.07±4.19 % for k equal to 6. When data were scaled, the probability was 79±6.32%. Here 30 runs were executed because *k*-means clustering uses random initializations.

In addition, leave-one-out testing was used, in supervised manner, to test years one by one against the result generated by the SOM. In leave-one-out testing, n-1 years or cases (n=48) were used to build a training set and one single year was the only test case for that model. All n=48 tests were again repeated several times. Nevertheless, with not scaled and scaled data clustering also created empty clusters thus failing. This is fairly expected since there were only 48 years in total that were attempted to cluster 6 or more clusters. Consequently, we employed the following classification techniques.

Furthermore, k nearest neighbor searching was applied with Euclidean distances. Because the smallest SOM cluster had only 3 years, it was reasonable to look at the closest neighbor only using k=1. When tests were made in the leave-one-out way, each year was once a test case. For the smallest cluster from which a test case was taken, there were 2 years left. This method gave the probability of 81.3% for the 48 years to be in the same clusters as those of the SOM when data were not scaled, and yet higher value, 91.7% when data were scaled, both of which

are better than the results by *k*-means clustering. Typically, larger (odd) *k* values are used since they may give somewhat better results. Although *k*=1 is known to be unstable, and when every decision is made based on 1-nearest neighbour only, it may get wrong result. Nonetheless, for this data set, the result was effective.

Again, decision tree tests with pruning were also used to validate the SOM clusters. When data was not scaled, the result was 83.3%. When data was scaled, the result was 81.3%. These two cases made no obvious differences.

Finally, using logistic discriminant analysis, 77.1% of correct classifications were given.

Through these tests, the SOM clusters proved to be satisfactory for further analysis.

6.3 Correlations

Using Viscovery SOMine, correlations between every pair of attributes can also be identified. A part of the generated correlations is shown in Table 4. Columns are attributes of crime. Rows are demographic and economic attributes. This table selects correlations between crime and socio-economic attributes.

In this experiment, many of the results demonstrate that there are strong links between some attributes. They can be summarized as in Table 4.

Table 4 Correlations between non-crime attributes *A* and crime-related attributes: correlations from interval (-0.3,0.3) were left out seen as insignificant, from (-0.6,-0.3] and [0.3,0.6) were seen as interesting, and from [-1,-0.6] and [0.6,1] marked in Bold face as significant.

Attribute	*Violent crime rate*	*Murder and non-negligent man-slaughter rate*	*Forcible rape rate*	*Robbery rate*	*Aggravated assault rate*	*Property crime rate*	*Burglary rate*	*Larceny-theft rate*	*Motor vehicle theft rate*
Civilian labour force percent of population	**0.84**	-	**0.90**	0.57	**0.93**	**0.61**	-	**0.76**	**0.68**
Employed civilian labour percent of population	**0.71**	-	**0.77**	0.39	**0.84**	0.43	-	**0.60**	0.59
Total fertility rate	**-0.79**	**-0.67**	**-0.82**	**-0.81**	**-0.71**	**-0.87**	**-0.71**	**-0.87**	**-0.77**
Personal consumption expenditures	0.58	-	**0.68**	0.34	**0.66**	0.37	-	0.51	0.43
Government consumption expenditures and investment	**-0.74**	-	**-0.82**	-0.50	**-0.81**	-0.56	-	**-0.70**	-0.57
Exports of goods and services	0.57	-	**0.64**	**0.82**	-	-	-	0.45	0.41
Imports of goods and services	0.41	-	0.52	-	0.56	-	-0.30	0.30	-
Real GDP growth rate	-	-	-	-0.35	-	-0.30	-	-	-
GDP growth rate	-	-	-	-0.34	-	-	-	-	-

The US has been a politically stable country over years. Every attribute is on a growing track, both favorable indicators and unfavorable indicators. There has been no sharp increase or sharp decrease during the past 50 years. For example, unemployment rate fluctuated between 3.5% and 9.7%. These figures can be explained as forming a

large range. However, in contrast with many other countries, these figures are in relatively low values. So we have to look for, from the slight change, the correlation between unemployment rate and crime. It turned out that it has no strong level of correlation on crime rate.

Due to above reason, successive years are clustered together in one way or the other. The exceptions are rare. This demonstrates a one-direction developmental tendency of most variables. In such a case, traditional statistical methods can also provide a good depiction for each single variable with tables or charts. However, comprehensive visualization can be achieved by applying the SOM as in this experiment.

It must be pointed out that, correlation is a result that can usually be generated by using various statistical methods. It has some importance for further consideration in identifying the causes. However, correlated attributes might contain but are not equal to causes themselves. Correlation expressed in numbers can happen to be irrelevant and it does not necessarily represent final analysis of the social problem in question.

The historical situation of crime in the US proved the correlation's another paradox, that is, when people in the society think something good, but statistically it positively correlates with total crime rate or a certain type of very serious offence. Or vice versa, crime takes place on a background where some very positive, favorable and good attributes are located. It is also positive that some negative, unfavorable and bad attributes in actual fact bring about just less crime. Hence the rise and fall of crime in the US, as reflected in correlation, as in this study, must be assembled again in human mind.

7 Discussion and conclusions

A society is a phenomenal compound, in which beneficial and unbeneficial elements are intertwined with each other. Criminal policy starts from the study of crime itself and inevitably its socio-economic context. Today, the study of crime can benefit from a variety of methodological developments facilitated by information systems. Analysis of crime in both a certain point of time of a certain span of time adds to the difficulties of research, and deserves special attention from academics.

Coping with such demands, this study applied the SOM in visualization and analysis of the historical development of criminal phenomena in the US over several decades. It showed that the SOM can be a feasible tool in clustering factors with similar characteristics, and it can be an alternative way to enhance analysis by visualization based on large amount of multi-dimensional data, which traditional methods have difficulties in dealing with.

During this process, ScatterCounter (Juhola and Siermala, 2012a) was used to validate the variables. Separation power, an indicator used to denote the strength of each attribute in their pre-defined classifications (clusters), is generated by this software kit. It proved that the dataset used in this test gives satisfactory results. It means that every variable had good level of separation power and is reserved in the latter stage of analysis. Unlike in other cases where some variables had to be removed from the dataset due to their weak and useless separation power, here all variables seemed to be strong and useful. Therefore, the results generated by the SOM using the present data set were reasonable.

In order to examine the results mentioned above, clusters generated by the SOM were further tested by *k*-means clustering, nearest neighbor searching, decision trees and logistic discriminant analysis, with data scaled or not scaled. Similarity between the results of the SOM and those others was high, ranging from 65% to 92% depending on the methods.

These provide innovative possibilities for the study of crime in the aspect of historical comparison and a simplified version of data processing and knowledge discovery for more sophisticated exploration.

Acknowledgments

The first author is grateful to Tampere Doctoral Programme in Information Science and Engineering for support.

References

Adderley, R. (2004) 'The use of data mining techniques in operational crime fighting', in *Proceedings of Symposium on intelligence and security informatics,* No. 2, Tucson A.Z., ETATS-UNIS (10/06/2004), Vol. 3073, pp. 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,* Vol. 5, No. 4, pp. 265-276.

Adderley, R., Townsley, M. and Bond, J. (2007) 'Use of data mining techniques to model crime scene investigator performance', in *Proceedings of the 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 170-176.

Axelsson, S. (2005) 'Understanding Intrusion Detection through Visualization', PhD thesis, Chalmers University of Technology, Goteborg, Sweden.

Blumstein, A. and Wallman, J. (2000) *The Crime Drop in America,* Cambridge University Press.

Brockett, P. L., Xia, X. and Derrig, R. A. (1998) 'Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud', *The Journal of Risk and Insurance,* Vol. 65, No. 2, pp. 245-274.

Bureau of Justice Statistics (2012), 'Reported crime in the United States 1960-2007'. Retrieved 10 August, 2012, from http://bjsdata.ojp.usdoj.gov/dataonline/Search/Crime/State/StatebyState.cfm?NoVariables=Y&CFID=350216&CFTOKEN=91023531

Conklin, J. E. (2003) *Why Crime Rates Fell*, Boston, MA: Pearson Education.

Dahmane, M. and Meunier, J. (2005) 'Real-Time Video Surveillance with Self-Organizing Maps', in *Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV'05),* Washington, DC.

Deboeck, G. (2000) 'Self-organizing patterns in world poverty using multiple indicators of poverty repression and corruption', *Neural Network World,* Vol. 10, pp. 239-254.

Fei, B. K., Eloff, J. H., Olivier, M. S. and Venter, H. S. (2006) 'The use of self-organising maps for anomalous behaviour detection in a digital investigation'. *Forensic Science International,* Vol. 162, Nos. 1-3, pp. 33-37.

Fei, B. K., Eloff, J., Venter, H. and Olivier, M. (2005) 'Exploring data generated by computer forensic tools with self-organising maps', in *Proceedings of the IFIP Working Group 11.9 on Digital Forensics*.

Grosser, H., Britos, P. and García-Martínez, R. (2005) 'Detecting Fraud in Mobile Telephony Using Neural Networks', in M. Ali and F. Esposito (Eds.): *IEA/AIE 2005*, LNAI 3533, pp. 613–615, Springer-Verlag Berlin Heidelberg.

Haykin, S. (1999) *Neural Networks - A Comprehensive Foundation*, Prentice Hall International, Upper Saddle River, N.J.

Higgs, N. (2002) 'Measuring Socio-Economic Status: A Discussion and Comparison of Methods', University of the Witwatersrand, Johannesburg, South Africa.

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. and Simula, O. (1999) 'A Self-Organizing Map for Clustering Probabilistic Models', in *Proceedings of Conference on Artificial Neural Networks*, Conference Publication No. 470, IEE 1999, pp. 946-951.

Huysmans, J. (2006) 'Country Corruption Analysis with Self-Organizing Maps and Support Vector Machines', in H. Chen et al. eds: *WISI 2006*, LNCS 3917, pp. 104-114.

Juhola, M. and Siermala, M. (2012a) 'A scatter method for data and variable importance evaluation', *Integrated Computer-Aided Engineering*, Vol. 19, pp. 137-149.

Juhola, M. and Siermala, M. (2012b) ScatterCounter software via link: http://www.uta.fi/sis/cis/research_groups/darg/publications.html

Kangas, L. J. (2001) 'Artificial Neural Network System for Classification of Offenders in Murder and Rape Cases', The National Institute of Justice, Finland.

Kohonen, T. (1990) The Self Organizing Map, in *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1464-1480.

Kohonen, T. (1997) *Self-organizing maps*. Springer-Verlag, New York, USA.

Lampinen, T., Koivisto, H. and Honkanen, T. (2005) 'Profiling network applications with fuzzy C-means and self-organizing maps', *Classification and Clustering for Knowledge Discovery*, Vol. 4, pp. 15-27.

Lee, S.-C. and Huang, M.-J. (2002) 'Applying AI technology and rough set theory for mining association rules to support crime management and fire-fighting resources allocation', *Journal of Information, Technology and Society*, Vol. 2, p. 65.

Leufven, C. (2006) 'Detecting SSH identity theft in HPC cluster environments using Self-organizing maps', Master's thesis, Linköping University, Sweden.

Li, S.-T., Tsai, F.-C., Kuo, S.-C. and Cheng, Y.-C. A. (2006) 'Knowledge Discovery Approach to Supporting Crime Prevention', in *Proceedings of the Joint Conference on Information Sciences 2006*, Taiwan.

Lozano, S., Gutierrez, E. (2008) 'Data envelopment analysis of the human development index', *Internationa Journal of Society Systems Science*, Vol. 1, No. 2, pp. 132 - 150.

Memon, Q. A. and Mehboob, S. (2006) 'Crime investigation and analysis using neural nets', in *Proceedings of International Joint Conference on Neural Networks,* Washington, DC, pp. 346-350.

Oatley, G. C., Ewart, B. W. and Zeleznikow, J. (2006) 'Decision support systems for police: lessons from the application of data mining techniques to "soft" forensic evidence', *Artificial Intelligence and Law*, Vol. 14, No. 1, pp. 35-100.

Priya, R. V., Vadivel, A., & Thakur, R. S. (2012). 'Maximal Pattern Mining Using Fast CP-Tree for Knowledge Discovery', *International Journal of Information Systems and Social Change*, Vol. 3, No. 1, pp. 56-74.

Rhodes, B., Mahaffey, J. and Cannady, J. (2000) 'Multiple Self-Organizing Maps for Intrusion Detection', in *Proceedings of the 23rd National Information Systems Security Conference,* October 16-19, 2000, Baltimore, Maryland, USA.

Viscovery (2012) Viscovery SOMine 5.2. Retrieved 10 August, 2012, from http://www.viscovery.net/somine/

Wadsworth, T. (2010) 'Is Immigration Responsible for the Crime Drop? An Assessment of the Influence of Immigration on Changes in Violent Crime between 1990 and 2000', *Social Science Quarterly,* Vol. 91, No. 2, pp. 531-553.

Zaslavsky, V. and Strizhak, A. (2006) 'Credit Card Fraud Detection Using Self-organizing Maps', *Information and Security: An International Journal,* Vol. 18, pp. 48-63.

Zimring F. E. (2007) *The Great American Crime Decline,* Oxford University Press, New York.

PUBLICATION V

Homicide and Its Social Context: Analysis Using the Self-Organizing Map

Xingan Li, Henry Joutsijoki, Jorma Laurikkala, Markku Siermala and Martti Juhola

Submitted to *Applied Artificial Intelligence*

HOMICIDE AND ITS SOCIAL CONTEXT:

ANALYSIS USING THE SELF-ORGANIZING MAP

Xingan Li, Henry Joutsijoki, Jorma Laurikkala, Markku Siermala and Martti Juhola

School of Information Sciences, University of Tampere, Tampere, Finland

Abbreviated title: Homicide Analysis Using the SOM

Abstract—Homicide has been one of the most serious kinds of offences. Research on causes of homicide has never reached definite conclusion. The purpose of this paper is to put homicide in its broad range of social context to seek correlation between this offence and other microscopic socio-economic factors. This international-level comparative study used a dataset covering 181 countries and 69 attributes. The data were processed by the Self-Organizing Map (SOM), assisted other clustering methods, including ScatterCounter for attribute selection, and several statistical methods for obtaining comparable results. The SOM is found to be a useful tool for mapping criminal phenomena through processing of multivariate data, and correlation can be identified between homicide and socio-economic factor.

Keywords—data mining; self-organizing map; k-means clustering; discriminant analysis; k-means nearest neighbor classifier; Naïve Bayes classification; Decision trees; Support vector machines (SVMs); Kruskal-Wallis test; Wilcoxon-Mann-Whitney U test; homicide; socio-economic factors

The first author is grateful to Tampere Doctoral Program in Information Science and Engineering (TISE) for financial support.

Address correspondence to Martti Juhola, School of Information Sciences, 33014 University of Tampere, Tampere, Finland. E-mail: Martti.Juhola@sis.uta.fi

INTRODUCTION

Homicide is an offence of one person illegally depriving the life of another person. Among all offences punished by law, homicide is one of the most serious kinds that attract global attention (United Nations Office on Drugs and Crime 2011). Exploration of fundamental causes of crime proved to be a nearly impossible task, because crime is always complicated, usually secret, concealed, and underreported. Distribution of crime differs between geographical units, between demographical groups, and between socio-economic combinations. Different combinations of demographical or socio-economic factors have played a significant role in the study of crime (Rock 1994).

In dealing with a broad range of socio-economic factors, the study of crime needs data mining and visualizing techniques, which have broadly shown their practical value in various domains. The self-organizing map, using an unsupervised learning method to group data according to patterns found in a dataset, is a qualified tool for data exploration. The interconnection between artificial intelligence and the study of crime makes an innovative study possible.

Currently, the SOM has been used in identifying individual offences, for instance, applied in the detection of credit card fraud (Zaslavsky, and Strizhak 2006), automobile bodily injury insurance fraud (Brockett, Xia, and Derrig 1998), burglary (Adderley 2004; Adderley, and Musgrave 2005), murder and rape (Kangas 2001), homicide (Kangas et al 1999; Memon, and Mehboob 2006), network intrusion (Leufven 2006; Lampinen, Koivisto, and Honkanen 2005; Axelsson 2005), cybercrime (Fei et al. 206; Fei et al. 2005), and mobile communications fraud (Hollmén 2000; Hollmén, Tresp, and Simula 1999; Grosser, Britos, and García-Martínez 2005). This is the main area where the application of the SOM has previously been emphasized in the research related to criminal justice.

A general lack of research can be found on macroscopic aspects of criminal phenomena, in this paper, homicide in particular, as related to socio-economic factors. The current situation created a motivation for designing experiments using the SOM, compared with and supplemented by other methods. Applying the SOM to investigate correlation between homicide and a broad range of socio-economic factors, this paper represents an effort to innovatively apply the SOM to the study of homicide, previously untouched by many other studies.

METHODOLOGY

This study applies the SOM, developed by Kohonen (Kohonen 1979) to cluster and visualize data. It is an unsupervised learning mechanism that clusters objects having 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 can be generated using software packages. By observing and comparing the clustering map and feature planes, it is possible to identify distribution of homicide among different socio-economic factors, and correlation between homicide and socio-economic factors. Detailed correlation table can also be realized automatically with Viscovery SOMine (Viscovery Software GmbH 2013), which adopts the correlation coefficient scale ranging from -1.0 to +1.0. These results, including clustering maps, feature planes and correlation tables constitute the fundamental ground for further analysis.

Although every attribute can be used by the SOM in clustering, their roles in the clustering are not evaluated. In order to select attributes, ScatterCounter (Juhola and Siermala 2012a; 2012b) will be used to measure the separation powers (between clusters) of all attributes. Those have weak powers will be dropped from the dataset and the reduced dataset will be used in final processing and analysis. For concrete format of data, a difference between using the SOM and using ScatterCounter is that, the SOM can process missing data by marking as "NaN" (Not-a-Number), while the ScatterCounter can only be used when the missing data in the original dataset are substituted with real values. In this research, missing values are imputed by the medians of other available values of the same attributes in the same clusters.

Besides the SOM, *k*-means clustering, discriminant analysis, k-means nearest neighbor classifier, Naïve Bayes classification, Decision trees, Support vector machines (SVMs), Kruskal-Wallis test, and Wilcoxon-Mann-Whitney U test will be used to validate the clusters and analysis by calculating how accurately these methods put the same countries into the same clusters as the SOM does.

DESIGN OF EXPERIMENTS

Countries included

Included in the experiment are 181 countries and territories, coded in Table 1. These codes will be shown in the maps as "labels". These countries were selected based on the availability of data on selected attributes. Most of these countries are members of the United Nations,

which, however, maintains the database containing information from non-member countries and territories. Usually, statistics of members are more available than non-members. But in exceptional cases, this is not true. Therefore, some members were dropped due to unavailability of data on a significant amount of indicators, while some non-members were kept because of their well-maintained statistical systems. Generally, the ratio of available data on indicators of individual countries or territories was controlled above 70%, and mostly above 80%.

Crime and socio-economic factor

This study contains 69 "attributes" (crime and socio-economic factors). An overview of all attributes that were used in this study is given in Table 2. One attribute, homicide per 100,000 people is a crime-related indicator, while 68 others are socio-economic factors. The selection of the contents of these indicators was primarily based on availability of data. Less consideration was put on the traditional concept on what might in actual fact cause the occurrence of offences of homicide, because in this research pre-determined and presumed correlations were temporarily ignored. Accordingly, in this research, some of these factors might traditionally be considered closely related to homicide, but some others might be considered quite irrelevant. Both of these categories of factors were re-considered in this research with a view to search potential clues for a new explanation.

The purpose of current study was to find groups of countries and territories according to their homicide rate, and to explore correlation between homicide and its socio-economic context. In this study, data source is United Nations Development Program (UNDP) online database. In this original dataset, total missing values accounted for 6.80%. Ten of the attributes had no missing values, while missing values in other attributes ranging from 0.55% to 21.55% (see Table 2). This criterion was similarly set as that of selecting countries and territories. As a result, in the final dataset, most columns and most rows were with the ratios of available data above 80%, with a few exceptions slightly below 80%.

Pre-processing and attribute selection

In the experiment, an important component is to pre-process the data with the SOM software and select attributes with ScatterCounter. The starting point was to generate clusters using the Viscovery SOMine, with which missing values did not need to be imputed as real number, instead they were marked as "NaN" only. Since four clusters were produced, the clusters were numbered as 1, 2, 3, and 4. These numbers were taken as cluster identifiers, which were used

to mark countries and territories in a separate column. The dataset with countries and territories bearing the marks of cluster identifiers would be processed with ScatterCounter (Juhola and Siermala 2012a; 2012b). In this dataset, missing values were imputed by substituting them with medians of pertinent clusters, because this software package will not deal with missing values.

TABLE I. COUNTRIES AND TERRITOORIIES INCLUDED

Afghanistan	AF	Denmark	DK	Lesotho	LS	Samoa	WS
Albania	AL	Djibouti	DJ	Liberia	LR	Sao Tome and Principe	ST
Algeria	DZ	Dominican Republic	DO	Libya	LY	Saudi Arabia	SA
Andorra	AD	Ecuador	EC	Liechtenstein	LI	Senegal	SN
Angola	AO	Egypt	EG	Lithuania	LT	Serbia	RS
Antigua and Barbuda	AG	El Salvador	SV	Luxembourg	LU	Sierra Leone	SL
Argentina	AR	Equatorial Guinea	GQ	Macedonia	MK	Singapore	SG
Armenia	AM	Eritrea	RE	Madagascar	MG	Slovakia	SK
Australia	AU	Estonia	EE	Malawi	MW	Slovenia	SI
Austria	AT	Ethiopia	ET	Malaysia	MY	Solomon Islands	SB
Azerbaijan	AZ	Fiji	FJ	Maldives	MV	South Africa	ZA
Bahamas	BS	Finland	FI	Mali	ML	Spain	ES
Bahrain	BH	France	FR	Malta	MT	Sri Lanka	LK
Bangladesh	BD	Gabon	GA	Marshall Islands	MH	Sudan	SD
Barbados	BB	Gambia	GM	Mauritania	MR	Suriname	SR
Belarus	BY	Georgia	GE	Mauritius	MU	Swaziland	SZ
Belgium	BE	Germany	DE	Mexico	MX	Sweden	SE
Belize	BZ	Ghana	GH	Moldova	MD	Switzerland	CH
Benin	BJ	Greece	GR	Mongolia	MN	Syrian Arab Republic	SY
Bhutan	BM	Grenada	GD	Montenegro	ME	Tajikistan	TJ
Bolivia	BO	Guatemala	GT	Morocco	MA	Tanzania	TZ
Bosnia and Herzegovina	BQ	Guinea	GN	Mozambique	MZ	Thailand	TH
Botswana	BW	Guinea-Bissau	GW	Myanmar	MM	Timor-Leste	TL
Brazil	BR	Guyana	GY	Namibia	NA	Togo	TG
Brunei Darussalam	BN	Haiti	HT	Nepal	NP	Tonga	TO
Bulgaria	BG	Honduras	HN	Netherlands	NL	Trinidad and Tobago	TT
Burkina Faso	BF	Hungary	HU	New Zealand	NZ	Tunisia	TN
Burundi	BI	Iceland	IS	Nicaragua	NI	Turkey	TR
Cambodia	KH	India	IN	Niger	NE	Turkmenistan	TM

Cameroon	CM	Indonesia	ID	Nigeria	NG	Tuvalu	TV
Canada	CA	Iran	IR	Norway	NO	Uganda	UG
Cape Verde	CV	Iraq	IQ	Oman	OM	Ukraine	UA
Central African Republic	CF	Ireland	IE	Pakistan	PK	United Arab Emirates	AE
Chad	TD	Israel	IL	Panama	PA	United Kingdom	GB
Chile	CL	Italy	IT	Papua New Guinea	PG	United States	US
China	CN	Jamaica	JM	Paraguay	PY	Uruguay	UY
Colombia	CO	Japan	JP	Peru	PE	Uzbekistan	UZ
Comoros	KM	Jordan	JO	Philippines	PH	Vanuatu	VU
Congo	CG	Kazakhstan	KZ	Poland	PL	Venezuela	VE
Congo (Democratic Republic of)	CD	Kenya	KE	Portugal	PT	Viet Nam	VN
Costa Rica	CR	Korea (Republic of)	KR	Qatar	QA	Yemen	YE
Côte d'Ivoire	CI	Kuwait	KW	Romania	RO	Zambia	ZM
Croatia	HR	Kyrgyzstan	KG	Russian Federation	RU	Zimbabwe	ZW
Cuba	CU	Lao People's Democratic Republic	LA	Rwanda	RW		
Cyprus	CY	Latvia	LV	Saint Lucia	LU		
Czech Republic	CZ	Lebanon	LB	Saint Vincent and the Grenadines	VC		

ScatterCounter is a software tool designed to evaluate how many “classes” (named “clusters” in the SOM) differ from each other in a dataset. The process starts from a random instance of a dataset and to traverse all instances by searching for the nearest neighbor of the current instance, then to update the one found to be the current instance, and to iterate the whole dataset. During the process of the search, every change from one class to another is counted. The more the class changes, the more the classes of a dataset are overlapped.

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 randomized steps, it is repeated for 5 to 10 times, and the average is used for separation power.

Separation powers can be calculated for the whole data, or for every class and for every attribute (Juhola and Siermala 2012a; 2012b). Absolute values of separation powers are from [0,1]. They are usually positive, but small negative values are also possible when in some classes an attribute does not virtually separate at all. Considering that such kind of attributes may be useful for some other classes, typically it needs to find those attributes that are rather useless for all classes so that these can be left out from continuation.

TABLE II. COUNTRY SOCIO-ECONOMIC SITUATION BY 69 ATTRIBUTES WITH THEIR MEANS, STANDARD DEVIATIONS AND MISSING VALUES IN PERCENT (ATTRIBUTES IN BOLD WERE FINALLY REMOVED ACCORDING TO THEIR POOR SEPARATION POWER)

Attributes	Year of data	Mean	Standard deviation	Missing values %
1. Adolescent fertility rate (women aged 15-19 years) (births per 1,000 women aged 15-19)	2008	53	43.6	2.76
2. Carbon dioxide emissions per capita (tonnes)	2008	4.9	6.8	1.66
3. Adult literacy rate, both sexes (% aged 15 and above)	2010	82.2	18.6	21.55
4. Adult mortality rate, female (per 1,000 people)	2009	165	123.1	0.55
5. Adult mortality rate, male (per 1,000 people)	2009	242.4	136.9	0.55
6. Agricultural land as a percentage of total land area (%)	**2009**	**40.78**	**22.06**	**0.00**
7. Average annual population growth rate (%)	2010	1.57	1.82	0.00
8. Change in forest area, 1990/2010 (%)	**1990-2010**	**1.4**	**29.5**	**0.00**
9. Combined gross enrolment in education (both sexes) (%)	2011	74.7	15.7	9.39
10. Consumer Price Index	2011	145.8	35.7	5.52
11. Deaths due to cardiovascular diseases and diabetes (per 1,000 people)	2008	348.5	141.8	2.76
12. Employment to population ratio, population 25+ (% aged 25 and above)	2011	65.88	12.31	8.29
13. Expenditure on health, public (% of GDP) (%)	2010	4.09	2.48	1.10
14. Export of merchandize goods as % of GDP (% of GDP)	2010	29.2	23.1	19.89
15. Export of services as % of GDP (% of GDP)	2010	12	14.6	18.23
16. Fixed and mobile telephone subscribers per 100 people (per100 people)	2010	108.5	54.1	0.00
17. Foreign direct investment, net inflows (% of GDP)	**2010**	**6.7**	**30.2**	**3.87**
18. Forest area (% of total land area)	**2010**	**30.63**	**22.37**	**0.00**
19. Fresh water withdrawals (% of actual total renewable	2012	58	264	8.29

	water resources)				
20.	GII: Gender Inequality Index, value	2012	0.38	0.19	18.23
21.	Gross primary enrolment ratio (% of primary school-age population)	2011	105.1	15.2	2.21
22.	Gross secondary enrolment ratio (% of secondary school-age population)	2011	77.6	27	5.52
23.	Gross tertiary enrolment ratio (% of tertiary school-age population)	2011	32.4	26.8	7.18
24.	Education index	2012	0.64	0.20	1.10
25.	Expected Years of Schooling (of children) (years)	2012	12.46	3	0.00
26.	GDP per capita (2005 PPP $)	2011	11669	13750	3.87
27.	GNI per capita in PPP terms (constant 2005 international $) (Constant 2005 international $)	2010	11542	14157	1.10
28.	Health index	2012	0.78	0.15	0.00
29.	Human Development Index (HDI) value	2012	0.67	0.17	1.10
30.	Income index	2012	0.62	0.19	1.10
31.	Life expectancy at birth (years)	2012	69.73	9.59	0.00
32.	Mean years of schooling (of adults) (years)	2012	7.63	3.02	1.10
33.	Non-income HDI value	2012	0.71	0.17	1.10
34.	HIV prevalence, Youth (% aged 15–24), female (% aged 15-24)	2009	1.23	2.69	20.99
35.	HIV prevalence, Youth (% aged 15–24), male (% aged 15-24)	2009	0.60	1.01	20.99
36.	Homicide rate (per 100,000)	2010	10.41	13.49	1.10
37.	Inequality-adjusted life expectancy index	2012	0.66	0.21	2.76
38.	Loss due to inequality in life expectancy (%)	2012	18.49	13.45	2.76
39.	Immunization coverage among 1-year-olds, DTP 1 (%)	2010	93.83	7.76	0.55
40.	Immunization coverage among 1-year-olds, measles (%)	2010	87.62	13.18	0.55
41.	**Impact of natural disasters: number of deaths (average per year per million people)**	**2011**	**5.8**	**23.6**	**8.29**
42.	Import of merchandize goods as % of GDP (% of GDP)	2010	40.7	21.3	20.44
43.	Import of services as % of GDP (% of GDP)	2010	12.3	12.6	18.78

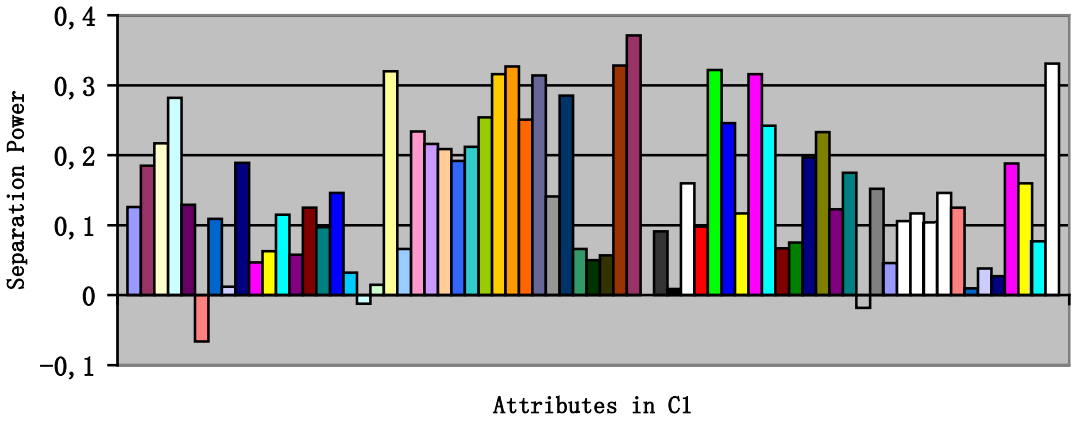
44.	Infant mortality rate (per 1,000 live births)	2010	30.1	28.7	0.00
45.	Internet users (per 100 people)	2010	34.06	27.83	2.76
46.	Labour force participation rate, female-male ratio (Ratio of female to male shares)	2011	0.71	0.20	4.42
47.	Maternal mortality ratio (deaths of women per100,000 live births)	2010	160.9	206.8	3.31
48.	Median age of the total population (years)	2010	27.71	8.3	2.76
49.	Natural resource depletion (% of GNI)	2010	6.15	10.34	16.57
50.	Net migration rate (per 1,000 people)	2010	1.8	15.5	2.76
51.	Net ODA received (% of GNI)	2010	7.2	17.1	17.68
52.	Personal computers (per 100 people)	2009	16.87	22.24	5.52
53.	Physicians per 1000 population (per 1,000 people)	2009	1.58	1.44	9.39
54.	Population with at least secondary education, female/male ratio (Ratio of female to male rates)	2010	0.84	0.24	21.55
55.	Population, urban (%) (% of population)	2012	56.1	23.2	0.00
56.	Remittances outflows (Workers' remittances and compensation of employees, total paid), (% of GDP)	2010	1.42	2.64	19.34
57.	Sex ratio at birth (Male births per100 female births)	2010	1.05	0.02	2.76
58.	Share of agricultural exports in total merchandize exports (%)	2010	25.87	24.56	19.34
59.	Share of agricultural imports in merchandize imports (%)	2010	14.66	6.64	19.89
60.	Share of manufactured exports in merchandize exports (%)	2010	40.94	30.24	19.34
61.	Share of manufactured imports in total merchanize imports (%)	2010	63.34	11.72	19.89
62.	Share of parts and components exports in total manufactured exports (%)	2010	15.97	15.74	19.34
63.	**Shares in parliament, female-male ratio**	**2011**	**0.28**	**0.27**	**2.21**
64.	**Stock of emigrants as a percentage of population (% of population)**	**2010**	**10.03**	**11.95**	**0.55**
65.	Stock of immigrants as a percentage of population (% of population)	2010	8.39	13.23	0.55

66. Total dependency ratio (per 100 people aged 15-64 years)	2012	58.7	17.4	2.76
67. Total fertility rate (births per woman)	2012	2.79	1.37	3.31
68. Total reserves minus gold (% of GDP)	2010	20.6	17.3	7.73
69. Under-five mortality (per 1,000 live births)	2010	42.8	46.7	0.00

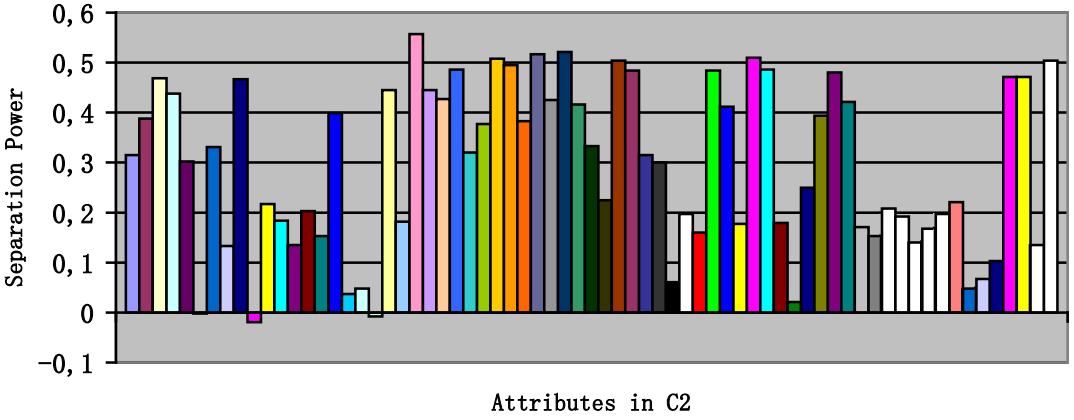
Sources: United Nations Development Program (UNDP). Available at http://hdr.undp.org/en/statistics/data/

Separation powers of attributes in each cluster and in the whole dataset are presented in Figure 1. Although an attribute may have separation powers around zero for some clusters, if for one cluster its separation power is larger, it can be useful to separate this cluster from the other. Therefore, separation powers of each attribute were computed both cluster by cluster and for all the data. With these results and observations, seven attributes (6, 8, 17, 18, 41, 63, 64) have poor separation powers and are removed from the dataset used in the following experiments and analysis.

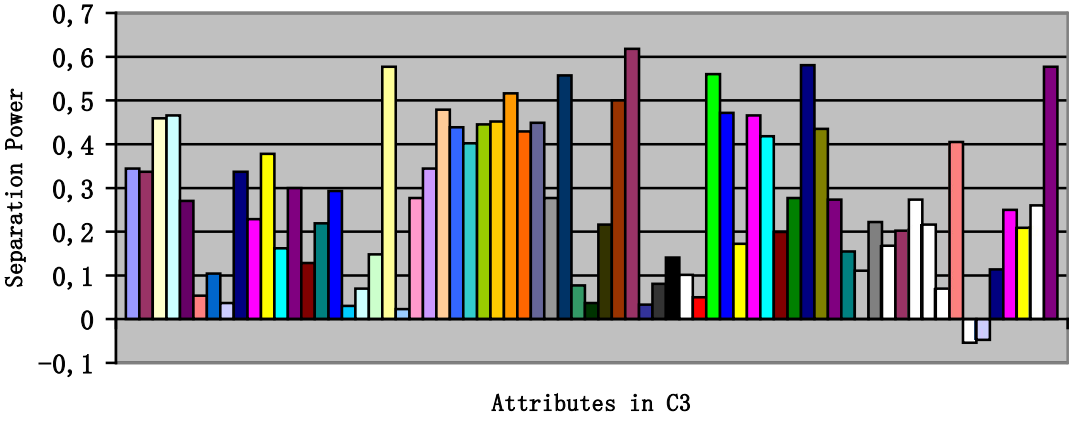
As a result, all the 62 attributes preserved in the dataset had stronger separation power and supposedly enable valid clustering. In this reduced dataset, missing values accounted for fewer than 7.32%. The rate of missing value was increased due to the fact that seven attributes were identified as having weak separation power and discarded: three attributes with no missing value, three other attributes with the ratio of missing values below initial average, and the other attribute with the ratio of missing value below the attribute with the most missing value. Because the number of countries had no change, the ratios of missing values of reserved single attributes were the same as in the original dataset.


0, 4
0, 3
0, 2
0, 1
0
−0, 1
Separation Power
Attributes in C1

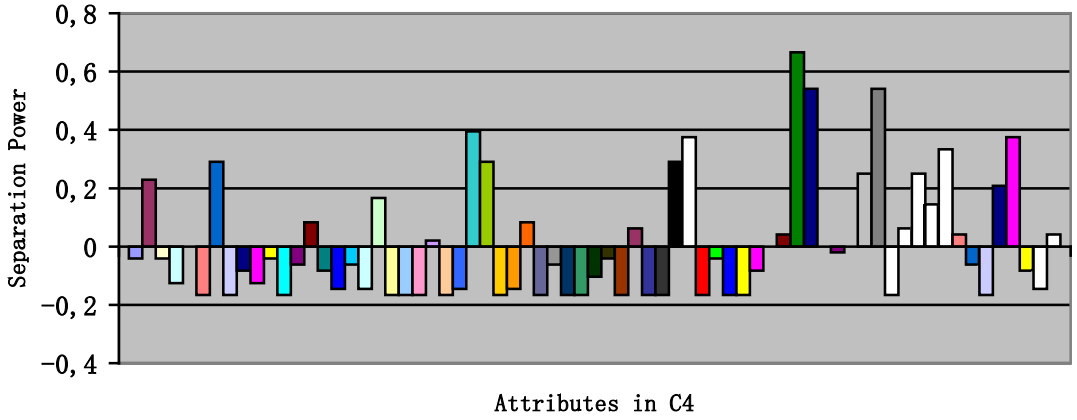
(a)


0, 6
0, 5
0, 4
0, 3
0, 2
0, 1
0
−0, 1
Separation Power
Attributes in C2

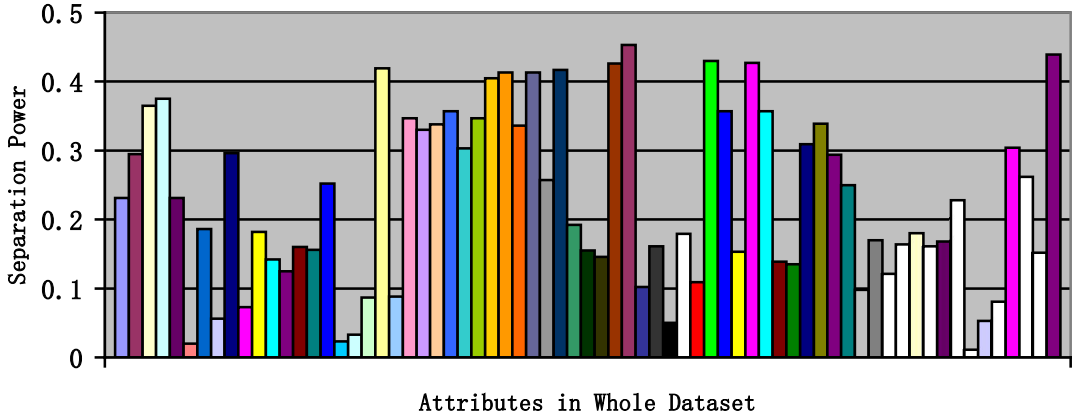
(b)


0, 7
0, 6
0, 5
0, 4
0, 3
0, 2
0, 1
0
−0, 1
Separation Power
Attributes in C3

(c)



(d)



(e)

Fig. 1. Separation powers of each attibute in: (a) cluster C1, (b) cluster C2, (c) cluster C3, (d) cluster C4, and (e) the whole dataset. The order of the 69 attributes from the left to the right along with the horizontal axis is that given in Table 2. Finally, seven attributes (6, 8, 17, 18, 41, 63, 64) with poor separation powers were discarded.

Construction of the map

In this study, the software package used is Viscovery SOMine 5.2.2 Build 4241. Compared with some other software packages, 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.

Here again, the dataset was formed from the original one by deleting attributes with weakest separation power. Missing values were marked with "NaN". The SOMine software automatically generated maps from the dataset of 181 countries and 62 attributes. The clustering map (Figure 2) as well as some other detailed statistics, such as correlations as discussed below, can be used in further analysis.

Support Vector Machine, One-vs-One Method and Parameter Estimation

Support Vector Machine (SVM) (Vapnik 2000; Burges 1998; Cortes, and Vapnik 1995) is a supervised classification method developed for two-class classification problems. The key idea in SVM is to construct a classes separating hyperplane in the input space such that the margin (distance between the closest examples of both classes) is maximized. By this means the generalization ability (in other words, the ability to predict the unseen test samples correctly) of an SVM classifier is the highest. When the classes in a dataset cannot be directly separated by a hyperplane in the input space, we need to use the kernel functions (Vapnik 2000; Burges 1998). The main point in kernel functions is that the training set, which is located in the input space, is mapped by a nonlinear transformation to (possibly) an infinite dimensional space where the classes can be separated by a hyperplane. The construction of a maximum margin SVM classifier is based on optimization theory and since the basic theory of an SVM classifier is well known from the literature, a reader can find the detailed mathematical derivation from (Vapnik 2000) for instance.

Due to the two-class restriction of SVM, different kind of approach for multi-class cases (the number of classes is greater than 2) is needed. One-vs.-One (OVO) method (Galar et al. 2011; Joutsijoki, and Juhola 2013; Joutsijoki, and Juhola 2011) is a commonly used multi-class extension for SVM. In OVO method one classifier is constructed for each class pair. Thus, altogether $M(M-1)/2$ classifiers are needed for $M>2$ class classification problem. Each one of the classifiers gives a predicted class label (vote) for test examples. The final class label for a test example is the label that occurred mostly. If a tie occurs among the classes, the final class label is solved by 1-Nearest Neighbor as in Joutsijoki and Juhola (2013) and Joutsijoki and Juhola (2011).

The use of SVM requires estimation of parameters. In this study we applied seven kernel functions. These were: linear, polynomial kernels (degrees from 2 to 5), Radial Basis Function (RBF) and Sigmoid (see formulas Hsu et al. 2013). For the linear and polynomial kernels only one parameter (C which is a penalty parameter) is to be estimated and for the RBF there are two parameters (C and γ, which is the width of Gaussian basis function) and for Sigmoid there are three parameters to be estimated (C, $\kappa>0$ and $\delta<0$). We performed the parameter value estimation in the following manner. Firstly, a dataset was divided into training and test sets with leave-one-out method. Secondly, every training set was divided into training and test sets such that 1/3 was left to validation and 2/3 for training. Thirdly, SVMs were trained by using a smaller training set and the accuracy of validation set was evaluated. Fourthly, the average

accuracy of validation sets was determined. Fifthly, when the best parameter value combination was found (the highest mean accuracy of validation sets determined the best parameter values), SVMs were trained again by using the full training set and they were tested with the test set. Since the OVO method contains several binary SVM classifiers, we decided to simplify the test arrangements so that we used for every SVM classifier the same parameter values. We tested the linear and polynomial kernels with 26 parameter values ($C\epsilon\{2^{-15}, 2^{-14},\ldots,2^{10}\}$) and both RBF and Sigmoid kernels with 676 parameter value combinations ($C,\gamma,\kappa \epsilon\{2^{-15}, 2^{-14},\ldots,2^{10}\}$ and $\delta\epsilon\{-2^{-10}, -2^{-9},\ldots,-2^{-15}\}$). All the tests and implementation of OVO method were made with Matlab 2010b and the Bioinformatics Toolbox of Matlab.

RESULTS

Upon processing of data, seven clusters have been generated, each representing groups of countries 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.

In analyzing country homicide situation to fulfill different demands, when there were many objects involved, other two levels of clustering concepts: super-clusters and sub-clusters, can also be used. Within the frameworks of each level of clusters, members in each cluster have their common properties, based on which they were grouped, and based on which they could be further analyzed.

Super-cluaters, clusters, and sub-clusters

Clusters were given in Figure 2. Due to the feature of the software package, countries and territories were not completely shown in the map. In order to give a full picture of these clusters, the following lists all the countries and territories in each cluster:

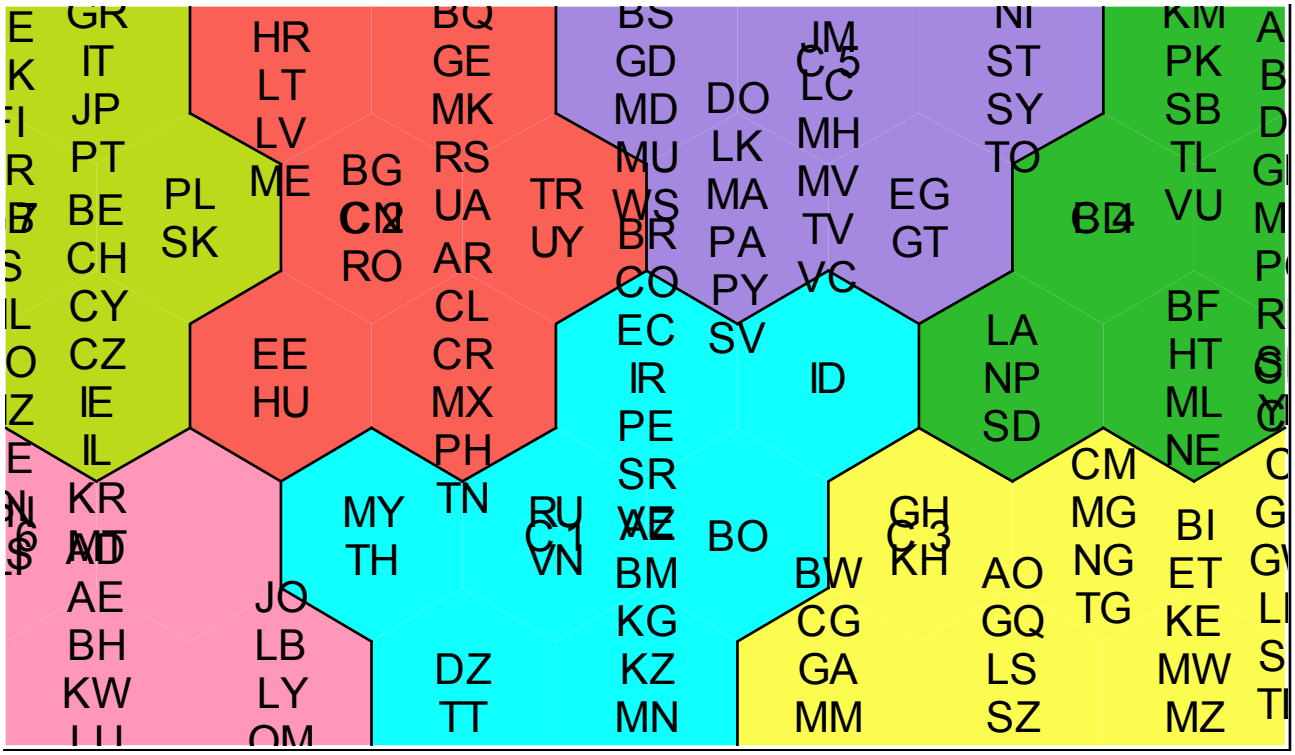


Fig. 2. Clustering map with cluster names C1–C7 and labels of countries

a) Cluster C1 consists of 23 countries: BR, CO, EC, IR, PE, SR, VE, ID, MY, TH, DZ, TT, RU, VN, AZ, BM, KG, KZ, MN, TJ, TM, UZ, BO

b) Cluster C2 consists of 26 countries: BY, CU, HR, LT, LV, ME, EE, HU, BG, CN, RO, AL, AM, BQ, GE, MK, RS, UA, AR, CL, CR, MX, PH, TN, TR, UY

c) Cluster C3 consists of 33 countries: GH, KH, CM, MG, NG, TG, BI, ET, KE, MW, MZ, RW, TZ, UG, CD, CF, CI, GN, GW, LR, SL, TD, BW, CG, GA, MM, NA, AO, GO, LS, SZ, ZM, ZW

d) Cluster C4 consists of 24 countries: IN, IQ, KM, PK, SB, TL, VU, BD, AF, BJ, DJ, GM, MR, PG, RE, SN, YE, LA, NP, SD, BF, HT, ML, NE

e) Cluster C5 consists of 30 countries: AG, BB, BS, GD, MD, MU, WS, BZ, FJ, GY, HN, JM, LC, MH, MV, TV, VC, CV, NI, ST, SY, TO, DO, LK, MA, PA, PY, SV, EG, GT,

f) Cluster 6 consists of 15 countries: BN, LI, AD, AE, BH, KW, LU, QA, SG, JO, LB, LY, OM, ZA, SA

g) Cluster C7 consists of 30 countries: ES, GR, IT, JP, PT, AT, AU, CA, DE, DK, FI, FR, GB, IS, NL, NO, NZ, SE, SI, US, BE, CH, CY, CZ, IE, IL, KR, MT, PL, SK

Compared with clusters generated by the software before several attributes were removed by applying ScatterCounter to select more useful attributes based on their separation powers, clusters here were to some extent re-grouped. Principally, clusters before and after removing the attributes that had weak separation powers should be with high similarity, including cluster numbers, and countries in each cluster. That could be taken as the initial purpose for applying separation power.

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

By emphasizing difference between clusters before and after removing weak attributes by applying the separation power, it did not ignore the actual fact that a majority of countries that were originally in the same small groups (sub-clusters in clusters) were subsequently still in the same sub-clusters. That is to say, clusters changed, but the change took place primarily at the sub-cluster level, not the individual level. Single countries did not move from here to there separately. Rather, closely joined small groups of countries migrated from one cluster to another.

This phenomenon enabled research on small groups of countries on the background of the whole world, by subtracting information from the self-organizing map established by processing the data depicting the panoramic view.

Because the unsupervised clustering map and feature maps were generated based on 62 attributes, description of these clusters became more complicated. Particularly, when special information about one attribute is needed, countries and territories in these seven clusters may be better regarded as components in fewer numbers of super-clusters. For example, according to feature map of homicide rate (Figure 3), these seven clusters can be seen as components in three super-clusters:

The first one consists of cluster C3 (34 countries) and cluster C5 (30 countries). They have higher level of homicide rate.

The second one consists of C1 (23 countries) and C4 (24 countries). They have medium level of homicide rate.

The third one consists of C2 (26 countries), C6 (14 countries) and C7 (30 countries). They have lower level of homicide rate.

Certainly, according to other attributes, there were more possibilities to form different super-clusters, which would find their use in different research interests.

On the other hand, where necessary, within the frameworks of each of these seven clusters, several sub-clusters could also be identified. For a random example, in cluster 5, five countries, DO, LK, MA, PA, PY, and SV form a sub-cluster. It implicated that they have closer common properties than those members in the same cluster. Because they were closely grouped with each other, their clustering would not differ in feature maps of different attributes.

While most countries were assembled in big or small groups, a few countries were isolated. They stayed separately far away from other countries, such as Bangladesh, Bolivia, and Indonesia. Although they have much in common with other countries in the same clusters, the map can still be used in a way of establishing the elaboration of diversity.

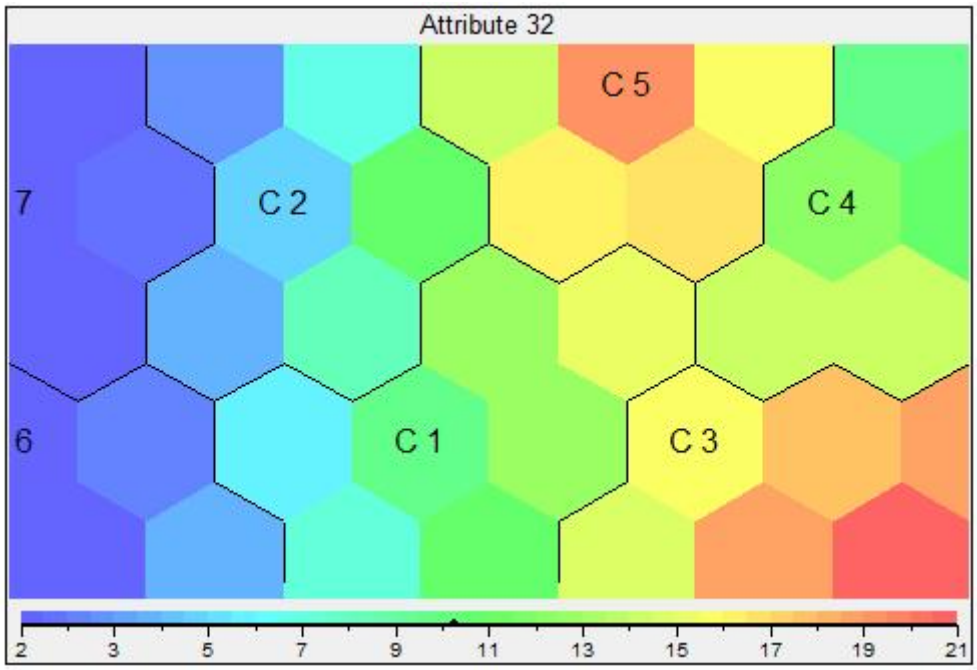


Fig. 3. Feature map of homicide rate

Validation of clusters

Total 181 countries times 62 attributes (after removing the poorest seven attributes identified by applying the ScatterCounter method) with original 7.32% missing values were imputed

once again with clusterwise medians, with new clusters (classes) given by the SOM clustering method.

TABLE III. TRUE POSITIVE RATES [%]

	Not scaled	**Scaled**	**Standardized**
Unsupervised, 30 iterations			
k-means *k*=7	36.6±1.6	45.5±2.3	44.3±2.8
k=8	37.8±1.3	45.6±1.8	45.9±2.1
k=9	39.0±1.4	46.6±1.9	46.8±2.5
k=10	40.5±0.8	46.2±1.9	48.1±2.0
k=11	41.6±0.7	46.8±1.7	48.0±2.0
Supervised, 30 iterations			
k-means *k*=7	34.6±1.3	41.8±1.5	39.9±2.3
k=8	- (empty cluster)	42.8±2.0	41.9±2.3
k=9	- (empty cluster)	41.6±1.9	42.6±2.0
k=10	- (empty cluster)	42.2±2.0	43.2±2.3
k=11	- (empty cluster)	42.9±1.8	- (empty cluster)
Discriminant analysis			
Linear	35.9	35.9	35.9
Quadratic		- (not positive definite covariance matrix)	
Mahalanobis		- (not positive definite covariance matrix)	
Logistic			26.5
k-nn, *k*=1	23.2	31.5	28.7
k=3	20.4	31.5	36.5
k=5	25.4	31.5	40.9
k=7	32.0	31.5	45.9
k=9	29.8	31.5	48.1
k=11	30.9	31.5	48.6
k=13	33.7	31.5	47.5
Naïve Bayes with ‘normal’ for ‘dist’	38.7	38.7	38.7
Naïve Bayes with ‘kernel’ for ‘dist’	44.2	44.2	44.2
Decision trees	45.3	45.3	45.3
SVM			
Linear	38.1%	50.3%	48.6%
Polynomial degree 2	32.0%	50.8%	37.6%
Polynomial degree 3	21.5%	49.2%	37.6%
Polynomial degree 4	24.3%	48.1%	33.1%
Polynomial degree 5	18.8%	43.6%	33.7%
RBF	29.8%	49.2%	48.1%
Sigmoid	0.0%	50.3%	48.6%

After imputation, the results by the SOM were tested by several methods, including *k*-means clustering, discriminant analysis, *k*-nearest neighbor classifier, Naïve Bayes classification, Decision trees, Support vector machines (SVMs), Kruskal-Wallis test, and Wilcoxon-Mann-Whitney U test (Table 3). True positive rates, in other words, percentual ratios of correctly classified countries and all countries were computed.

Unsupervised *k*-means clustering gave true positive rates between 36%-41% when data were not scaled. Scaling of data made the results slightly better, between 45%-46%. When data were standardized (attribute by attribute, by subtracting with the mean and dividing with the standard deviation of each attribute), overall results still bettered off, between 44%-48%. Supervised *k*-means behaved worse than unsupervised, true positive rates are generally 4-5% lower. In some cases, empty clusters also occurred.

Different methods of discriminant analysis were tested. Linear discriminant analysis got the same rate of 35.9% regardless of the date being scaled, unscaled, or standardized. Quadratic and Mahalanobis analysis got no positive definite covariance matrix, while logistic analysis got only 26.5% when data were standardized.

Furthermore, *k*-nearest neighbor classifiers (*k*-nn) gave results between 23%-33% when data were not scaled, and 31.5% (*k*=1, 3, 5, 7, 9, 11, 13) when data were scaled. When the data were standardized, results had a broader range, from 28.7% (*k*=1) to 47.8% (*k*=13).

Naïve Bayes with 'normal' for distance function 'dist' got a result of 38.7%, Naïve Bayes with 'kernel' for 'dist' got a result of 44.2%, decision trees got a result of 45.3%, regardless of data being scaled, not scaled or standardized.

SVMs obtained the lowest rates when data were not scaled, but the highest rates when date was scaled, and medium rates when data were standardized (detailed in Table 3).

The results of the Kruskal-Wallis test showed that there were significant ($p < 0.05$) differences among the groups defined by the clusters in 60 out of the total 62 variables. On average, six out of the 21 pairwise test results obtained with the Wilcoxon-Mann-Whitney U test were significant after the p values were corrected with the Holm's method.

Correlations

Viscovery SOMine could generate a detailed list of correlations, based on which Table 4 was created. 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. Traditionally, single research on crime did not include so many attributes (or named correlation factors or causes). Even in textbooks, only a dozen or two were introduced. So it shall be highly expected to have such data mining methods to be able to process several dozens of attributes and to provide immediate reference for further analysis.

TABLE IV. CORRELATIONS BETWEEN SOCIO-ECONOMIC ATTRIBUTES *A* AND HOMICIDE RATE (A32). THE ORDER OF THE ATRRIBUTES WAS THAT GIVEM IN TABLE 2 AFTER REMOVING THE 7 ATTRIBUTES.

A1	0.55	A22	-0.29	A43	-0.43
A2	-0.23	A23	-0.31	A44	0.13
A3	-0.19	A24	-0.39	A45	-0.16
A4	0.44	A25	-0.37	A46	0.02
A5	0.47	A26	-0.35	A47	-0.34
A6	0.06	A27	-0.39	A48	-0.35
A7	-0.24	A28	-0.31	A49	-0.08
A8	0.17	A29	-0.37	A50	-0.24
A9	0.16	A30	0.31	A51	-0.09
A10	0.31	A31	0.35	A52	-0.32
A11	-0.19	A33	-0.40	A53	0.25
A12	-0.17	A34	0.39	A54	0.15
A13	-0.17	A35	-0.11	A55	-0.29
A14	-0.27	A36	-0.16	A56	0.04
A15	-0.12	A37	-0.10	A57	-0.26
A16	0.41	A38	-0.07	A58	-0.25
A17	0.17	A39	0.34	A59	0.37
A18	-0.35	A40	-0.40	A60	0.33
A19	-0.33	A41	0.13	A61	-0.14
A20	-0.33	A42	0.35	A62	0.32
A21	-0.35				

From Table 4, homicide was positively correlated with 23 attributes, while negatively correlated with rest 38 attributes. Some correlation values were interesting, while others were very weak. Certainly, it is still too early to conclude what socio-economic factors cause homicide, affecting its occurrence, or its increase or decrease, especially when than ever before this study included more factors that traditional research did have full coverage.

CONCLUSIONS

As one of the most serious offences, the occurrence and distribution of homicide in the world can be analyzed though accessible statistical data. This paper dealt with macroscopic data for international comparison. Conventionally, analysis in the study of crime, either on general

issues or on particular issues, did not handle large-scale of multidimensional data. Specifically, when international comparison was carried out, discussion was much abstract and theoretical, lack of systematic data processing. By applying the Self-Organizing Map, this task was made possible.

With the self-organizing map, multidimensional comparison was realized. The research objects, countries and territories, could be grouped into clusters of different levels: super-clusters, clusters, and sub-clusters. Super-clusters were useful to reveal common features of one or more attributes of research objects (countries and territories). They could be used in different grouping ways with regard to different attributes. Clusters provided primary basis for the analysis of distribution of countries and territories with all attributes in the dataset. Sub-clusters were used to investigate smaller groups of countries and territories within a cluster. In principle, countries and territories within a sub-cluster could be seen as having the most features in common in socio-economic context.

In fact, in the sense of self-organizing map, countries in sub-clusters were more closely joined together, forming a more stable structure than clusters. It was found that, before and after removing weak attributes by applying the separation power, the structures of clusters could be varied one way or the other. Countries in sub-clusters, however, could stay in a rather stable framework. In research on small groups of countries on the background of the whole world, it is a useful way to subtract comparative information from the self-organizing map. It is also possible to situate individual countries, particularly those isolated ones in the map, on the international context.

By applying ScatterCounter to select attributes and refine the dataset, and by using *k*-means clustering, discriminant analysis, *k*-means nearest neighbor classifier, Naïve Bayes classification, Decision trees, Support vector machines (SVMs), Kruskal-Wallis test, and Wilcoxon-Mann-Whitney U test 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 more easily visualized and more convenient to interpret, facilitating practical comparison between countries with diversified socio-economic and criminal features. The paper provided broad potential for applying data analysis and visualization methods in the field of the study of crime, where in turn would find significant methodological value of this application.

REFERENCES

Adderley, R. 2004. The use of data mining techniques in operational crime fighting. In *Proceedings of symposium on intelligence and security informatics*, No. 2, Tucson A.Z., ETATS-UNIS (10/06/2004) 3073: 418-425.

Adderley, R., and Musgrave, P. 2005. Modus operandi modelling of group offending: a data-mining case study, *International Journal of Police Science and Management* 5 (4) 265-276.

Axelsson, S. 2005. *Understanding intrusion detection through visualization*, PhD thesis, Chalmers University of Technology, Göteborg, Sweden.

Brockett, P. L., Xia, X., and Derrig, R. A. 1998. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud, *The Journal of Risk and Insurance* 65 (2): 245-274.

Burges, C. J. C. 1998. A tutorial on support vector machiens for pattern recognition, *Data Mining and Knowledge Discovery* 2 (2): 121-167.

Cortes , C., and Vapnik, V. 1995. Support-vector networks, *Machine Learning* 20 (3): 273-297.

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 Science International* 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. In *Proceedings of the IFIP Working Group 11.9 on Digital Forensics* 1-15.

Galar, M., Fernández, A., Barrenechea, E., Bustince, H., and 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 Recognition* 44 (8): 1761-1776.

Grosser, H., Britos, P., and 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., and Simula, O. 1999. A self-organizing map for clustering probabilistic models, *Artificial Neural Networks* 470: 946-951.

Hsu, C.-W., Chang, C.-C., and Lin, C.-J. 2013. A practical guide to support vector classification, Technical report. Available at: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Joutsijoki, H., and Juhola, M. 2011. Comparing the one-vs-one and one-vs-all methods in benthic macroinvertebrate image classification, In P. Perner (ed.), *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Berlin, Germany 6871: 399-413.

Joutsijoki, H., and Juhola, M. 2013. Kernel selection in multi-class support vector machines and its consequence to the number of ties in majority voting method, Accepted to *Artificial Intelligence Review*. DOI: 10.1007/s10462-011-9281-3.

Juhola, M., and Siermala, M. 2012a. A scatter method for data and variable importance evaluation, *Integrated Computer-Aided Engineering* (19)2: 137-149.

Juhola, M., and Siermala, M. 2012b. ScatterCounter software via link: http://www.uta.fi/sis/cis/research_groups/darg/publications.html.

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 (March 22, 1999).

Kohonen, T. 1979. *Self-Organizing Maps*, Springer-Verlag, New York, USA

Lampinen, T., Koivisto, H., and Honkanen, T. 2005. Profiling network applications with fuzzy C-means and self-organizing maps, *Classification and Clustering for Knowledge Discovery* 4: 15-27.

Leufven, C. 2006. *Detecting SSH identity theft in HPC cluster environments using self-organizing maps*, Master's thesis, Linköping University, Sweden.

Memon, Q. A., and Mehboob, S. 2006. Crime investigation and analysis using neural nets. In *Proceedings of International Joint Conference on Neural Networks*, Washington, DC. 346-350.

Rock, R. 1994. *History of Criminology*. Dartmouth Publishing, Aldershot, UK.

United Nations Office on Drugs and Crime (UNODC). 2011. *Global study on homicide – trends, contexts, data.* Vienna: United Nations Office of Grugs and Crime. Available at http://www.unodc.org/.../Homicide/Globa_study_on_homicide_2011_web.pdf

Vapnik, V. N. 2000. *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, USA, 2nd edition.

Viscovery Software GmbH. 2013. Viscovery SOMine, http://www.viscovery.net/somine/.

Zaslavsky, V., and Strizhak, A. 2006. Credit card fraud detection using self-organizing maps, *Information and Security: An International Journal* 18: 48-63.