



Indexed Crime Data Visualization Utilizing Self-Organizing Map Algorithm

Jan Carlo T. Arroyo¹, Allemar Jhone P. Delima², Markdy Y. Orong³

¹College of Computing Education, University of Mindanao, Davao City, Davao del Sur, Philippines

²College of Engineering, Technology and Management, Cebu Technological University-Barili Campus, Cebu, Philippines

³Misamis University, Ozamiz City, Philippines

jancarlo_arroyo@umindanao.edu.ph¹, allemarjpdjca@yahoo.com², markdy.orong@mu.edu.ph³

ABSTRACT

Visualization is the transformation of symbolic representation to geometric representation. The goal of visualization is to analyze, explore, discover, illustrate, and communicate information in well understandable form. Visualization of crime data provides sufficient input to the authorities on the trends of crimes in the locality. This study clustered the indexed crime data of Misamis Occidental Philippines. Further, the study utilized Self-Organizing Map (SOM) to visualize the data. Results showed that five indexed crime data were clustered together, and three crime data were on the same cluster.

Key words: Data mining, neural networks, neurons, visualization

1. INTRODUCTION

Crime is an offense against the society that often prosecuted and punishable by the law [1]. Further, it is a major issue where the top priority has given by our government [2]. It has been noted that criminals commit crimes at any place and in any form [3]. The term vulnerability used as the intangible focus for conversation of the opportunities provided by the legal environment to planned crime, although with some reservation [4]. Moreover, it depends on the intentions of those criminals involved in the execution of their acts, upon the state's regulations and the standards of the sector itself, and of course, the intentions of those participating in the area [3].

Nowadays, the criminals are becoming technologically sophisticated in committing crimes [5]. Therefore, police need crime analysis tool to catch offenders and to remain advance in the battle between the criminals and the law enforcement unit. The police should use the current technologies [6] to provide themselves the needed boundary. Accessibility of significant and updated information is the paramount requirement in conducting of daily business and operation by the police, particularly in crime investigation and detection of criminals. Police organizations everywhere have been handling a large amount of such information and a

huge volume of records. Thus, crime data visualization is needed for the authority.

Visualization of crime data is very important in order to identify the crime pattern and support decision making in crime prevention [7]. Moreover, this technique is important as it can represent crime data into more understandable presentation and can be used to support decision making processes.

There are many visualization techniques that can be found in the literature. For an example, [8] compared several different ways of visualizing high-volume crime data and discussed a few possibly unforeseen, implications of mapping techniques. Moreover, [9] focused on comparison of three visualization techniques (map animation, the isosurface and the comap) in visualizing time and space in crime patterns.

With the availability of different technology-based techniques of data visualization, most researchers may find it easy to visualize data for any purposes. One of the most popular techniques in data visualization is artificial neural network (ANN) inspired by biological processes in the brain that typically recognized as non-linear distributive adaptive systems [10]. Application areas of ANN include the system identification and control, visualization, data mining, pattern recognition, medical diagnosis, sequence recognition, e-mail spam filtering and financial applications.

The Self-Organizing Map of [11] is an unsupervised learning neural network. It provides a nonlinear mapping from a high-dimensional input space to a lower-dimensional, often two-dimensional, and output space. In the process of mapping, the topology of the input space is mostly preserved. Input vectors that are close to each other in the input space are mapped to units that are close to each other in the output space. SOMs can be used for analysis and visualization, vector quantization, and clustering. In the context of GIScience, SOMs are primarily employed for spatial pattern analysis.

There are numerous studies that utilized SOM for visualization of unsupervised data in different field of specialization [12]–[15]. The study of [16] applied SOM on ecological data that included the presence or absence of 283 species at the 252 sites in the AdoureGaronne drainage basin

in South-Western France and classified the sampling sites as well as visualized the spatial distribution of each of the 283 considered species. In addition, [17] utilized SOM for clustering sample units (in species abundance database which consists of the species and the sample units) and visualization of species abundance by utilizing the component planes visualization technique. The use of 4x4 neurons in the Kohonen layer was observed and claimed that the selected map size is larger than the amount of sample units with 10 sample units.

2. METHODOLOGY

The self-organizing map is a learning algorithm that was originally proposed by [18]. The SOM is a fascinating neural network method that has found increasing interest in water resources applications. SOM networks learn to cluster groups of similar input patterns from a high dimensional input space in a non-linear fashion onto a low dimensional (most commonly two-dimensional) discrete lattice of neurons in an output layer [11]. In this study, the self-organizing map algorithm was utilized where flowchart is shown in Figure 1.

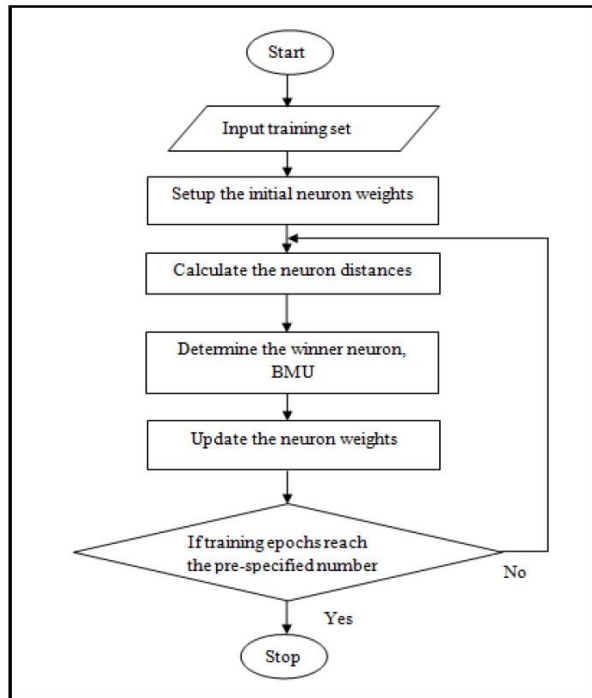


Figure 1: The flowchart of the SOM algorithm [19]

The algorithm accepts the indexed crime incidence that serves as the input training data. Based on the data input, the algorithm sets up the initial neurons weights that serve as the basis for calculating the neuron distances on the next step. Given the neuron distances, the algorithm then determines the winning neuron and updates the weights of the neuron. Further, the algorithms check if the training epochs reach the pre-specified number. If yes, the process will stop. Otherwise, the algorithm will calculate the neuron distance and will continue the process until the training epochs reach the specified number.

3. SIMULATIONS

In this study, the use of eight (8) indexed crime data was observed. The data were obtained from the different police stations of the province of Misamis Occidental, Philippines. In the study, there were 100 neurons specified and 200 epochs that can complete the 10x10 self-organizing maps.

Figure 2 presents the self-organizing map topology. The study applied the hexagonal SOM topology which is in two-dimensional array of neurons represented as:

$$M = \{m_1, \dots, m_{pq}\}$$

The neurons are connected to adjacent neurons by a neighborhood relation. This dictates the topology, or the structure, of the map. Usually, the neurons are connected to each other via rectangular or hexagonal topology.

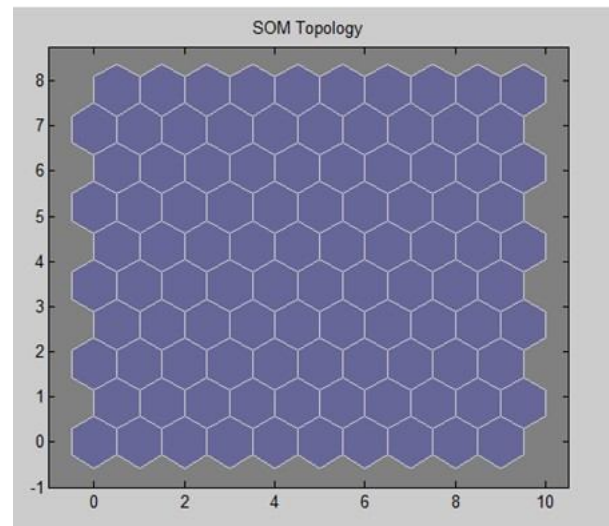


Figure 2: SOM Topology

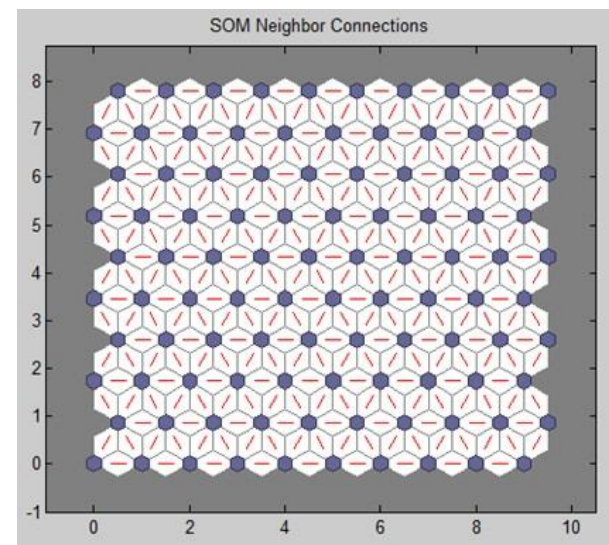


Figure 3: SOM neighbor connections

Figure 3 presents the establishment of the SOM neighbor connections. Immediate neighbors (the neurons that are adjacent) belong to the neighborhood N_c of the neuron mc . The neighborhood function should be a decreasing function of time: $N_c = N_c(t)$.

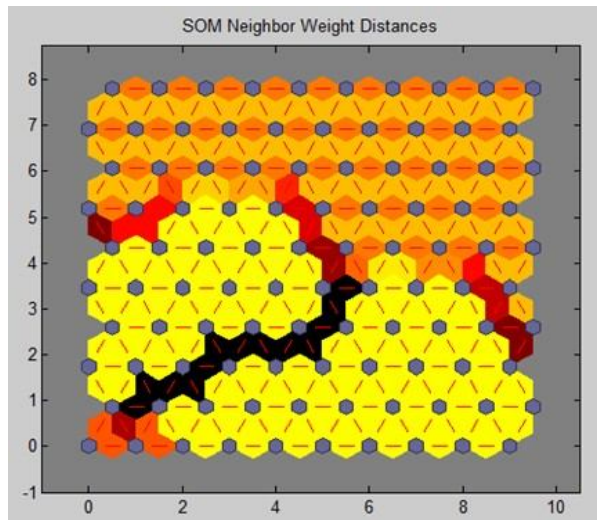


Figure 4: SOM neighbor weight distances

Figure 4 presents the SOM Neighbor Weight Distance. The blue hexagons represent the neurons. The red lines connect neighboring neurons. The colors in the regions containing the red lines indicate the distances between neurons. The darker colors represent larger distances, and the lighter colors represent smaller distances. A band of dark segments crosses from the lower-center region to the upper-right region. The SOM network appears to have clustered the flowers into two distinct groups.

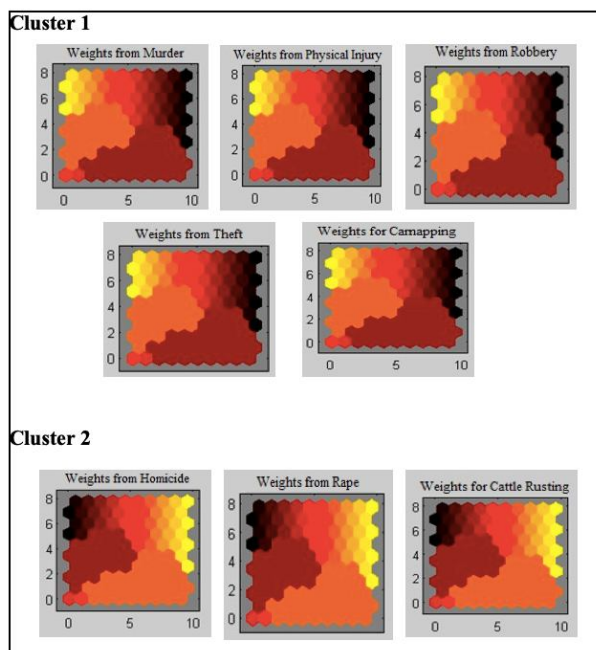


Figure 5: Indexed crime data visualization results

Figure 5 presents the indexed crime data of the province of Misamis Occidental clustered into two groups. As shown, there were five indexed crime data on cluster 1 such as murder, physical injury, robbery, theft and carnapping. On the other hand, homicide, rape and cattle rusting were on the second cluster.

4. CONCLUSION AND RECOMMENDATION

Self-Organizing Map successfully clustered the indexed crime data of Misamis Occidental. Murder, physical injury, robbery, theft and carnapping were clustered together. On the other hand, homicide, rape and cattle rusting were on the other cluster. Authorities may utilize the results of the study to gain insights pertaining to the common crimes in the province of Misamis Occidental.

REFERENCES

- [1] R. F. Backer, *Criminal investigation*, 3rd Editio. Burlington, MA: Jones and Bartlett Learning Publishers, 2010.
- [2] A. Malathi and S. S. Baboo, "Enhanced Algorithms to Identify Change in Crime Patterns," *Int. J. Comb. Optim. Probl. Informatics*, vol. 2, no. 3, pp. 32–38, 2011.
- [3] B. T. Vander, "Risky business: a risk-based methodology to measure organized crime," *Crime, Law Soc. Chang.*, vol. 41, no. 5, pp. 471–516, 2004.
- [4] V. B. Tom and V. D. Stijin, "Legitimate businesses and crime vulnerabilities," *Int. J. Soc. Econ.*, vol. 35, no. 10, pp. 739–750, 2008, doi: 10.1108/03068290810898954.
- [5] T. Abraham and O. de Vel, "Investigative profiling with computer forensic log data and association rules," in *IEEE International Conference on Data Mining*, 2002, pp. 11–18.
- [6] J. J. Corcoran, I. D. Wilson, and J. A. Ware, "Predicting the geo-temporal variations of crime and disorder," *Int. J. Forecast.*, vol. 19, no. 4, pp. 623–634, 2003.
- [7] N. M. Mohamad Noor and S. H. Ab Hamid, "Visualization of Crime Data Using Self-organizing Map (SOM) and Improvement in SOM: A Review and Available Tools," *J. Comput. Sci. Comput. Math.*, vol. 6, no. 2, pp. 37–43, 2016, doi: 10.20967/jcscm.2016.02.003.
- [8] J. H. Ratcliffe, "Visualising crime hotspots and making sense of high volume crime Visualising crime hotspots," in *Crime Mapping: Adding Value to Crime Prevention*, 2000, pp. 1–8.
- [9] C. Brunson, J. Corcoran, and G. Higgs, "Visualising space and time in crime patterns: A comparison of methods," *Comput. Environ. Urban Syst.*, vol. 31, no. 1, pp. 52–75, 2007, doi: 10.1016/j.compenvurbsys.2005.07.009.
- [10] Z. Kurd and T. P. Kelly, "Using fuzzy self-organising maps for safety critical systems," *Reliab. Eng. Syst. Saf.*, vol. 92, no. 11, pp. 1563–1583, 2007.
- [11] T. Kohonen, *Self-Organizing Maps*. Berlin, 2001.
- [12] C. Budayan, I. Dikmen, and M. . T. Birgonul, "Comparing the performance of traditional cluster analysis, self organizing maps and fuzzy C-means

- method for strategic grouping,” *Expert Syst. with Appl.*, vol. 36, no. 9, pp. 11772–11781, 2009, doi: 10.15957/j.cnki.ijdl.2009.07.004.
- [13] X. Li and M. Juhola, “Country crime analysis using the self-organizing map with special regard to demographic factors,” *J. Knowledge, Cult. Commun.*, vol. 29, no. 1, 2014.
- [14] Y. Nam, S. H. Koh, S. J. Jeon, Y. S. Park, and W. Choi, “Hazard rating of coastal pine forests for a black pine bark scale using self-organizing map (SOM) and random forest approaches,” *Ecol. Inform.*, vol. 29, pp. 206–213, 2015.
- [15] P. Sarlin, “Exploiting the self-organizing financial stability map,” *Eng. Appl. Artif. Intell.*, vol. 26, no. 5–6, pp. 1532–1539, 2013.
- [16] R. Céréghino, J. L. Giraudel, and A. Compin, “Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps,” *Ecol. Modell.*, vol. 146, no. 1–3, pp. 167–180, 2001, doi: 10.1016/S0304-3800(01)00304-0.
- [17] J. L. Giraudel and S. Lek, “A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination,” *Ecol. Modell.*, vol. 146, pp. 329–339, 2001, doi: 10.1016/S0304-3800(01)00324-6.
- [18] T. Kohonen, “Self-Organized Formation of Topologically Correct Feature Maps,” *Biol. Cybern.*, vol. 43, pp. 59–69, 1982.
- [19] J. J. Fuentes, M. Dominguez, P. Reguera, M. A. Prada, I. Diaz, and A. A. Cuadrado, “Visual dynamic model based on self-organizing maps for supervision and fault detection in industrial processes,” *Eng. Appl. Artif. Intell.*, vol. 23, no. 1, pp. 8–17, 2010.