International Journal of Engineering & Technology, 7 (2.2) (2018) 80-83

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Artificial Neural Network Optimized Approach for Improving Spatial Cluster Quality of Land Value Zone

Haviluddin[1*], Fahrul Agus[1], Muhamad Azhari[1], Ansari Saleh Ahmar[2]

[1]Faculty of Computer Science and Information Technology, Mulawarman University, Indonesia
[2]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Indonesia
**Corresponding author E-mail:haviluddin@unmul.ac.id*

Abstract

A geostatistics practical approach is divided data sample into several groups with certain rules. Then, the data groups are used for spatial interpolation. Furthermore, clustering technique is quite commonly used in order to get distance function between sample data. In this study, Self-Organizing Maps (SOM) optimized by using Learning Vector Quantization (LVQ) especially in distance variance have been implemented. The land value zone datasets in Samarinda, East Kalimantan, Indonesia have been used. This study shows that the SOM optimized by LVQ technique have a good distance variance value in the same cluster than SOM technique. In other words, SOM-LVQ can be alternative clustering technique especially centroid position in clusters.

Keywords: *SOM; LVQ; clustering; optimized; centroid; land value zone.*

1. Introduction

A geostatistics is a subset of statistics that specialized in analysis and interpretation of geographically referenced data. Geostatistics provides a set of statistical tools for analyzing spatial variability and spatial interpolation. This technique generates not only prediction surfaces but also error or uncertainty surfaces [1].
Currently, geostatistics is widely applied in analyzing data points, remote sensing, image measuring and filtering (such as dimension DEMs), optimization of spatial partition retrieval, spatial data etc [2]. One of the major uses of geostatistics is to predict a sample of the entire area of interest, called spatial interpolation. The geographical position of a location is expressed in geographical coordinate system in the form of latitude, longitude, elevation or X, Y, Z. Where, variables [lat, lon or X, Y] are state the location position, then the variable [elevation or Z] represents the location value. In principle, geostatistics is dividing the sample data into several groups with certain rules. Next step is to predict or spatial interpolation in each data set. Where, clustering technique such as K-Means, Fuzzy C-Means, Fuzzy Gustafson-Kessel (FGK) and SOM, etc. are widely used in the grouping of datasets based on distance function between the sample data [3-6].
The purpose of this research is optimization of final weight of intra-layer as centroid on SOM method by using LVQ method. This paper is organized as follows. Experimental is summarized in Section 2. Results and discussions are given in Section 3, and Section 4 draws conclusions.

2. Experimental Details

2.1. Self-Organizing Maps (SOM)

SOM is one type of NN, which is classified as unsupervised learning. The SOM architecture consists of an input layer with n training vector units, output layer with k category / cluster and intra-layer unit that connects between input layer and output layer [7, 8], Figure 1. Each neuron in the input layer is directly connected to each neuron at the input layer where each relationship has a weighted vector of length n^3, Figure 2.

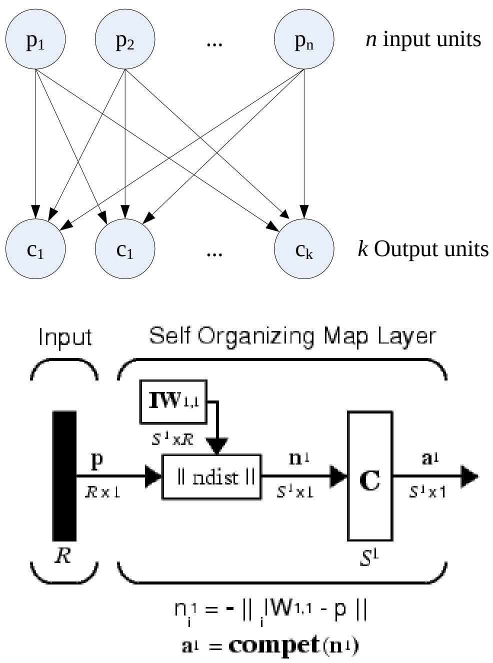

Fig. 1: SOM architecture.

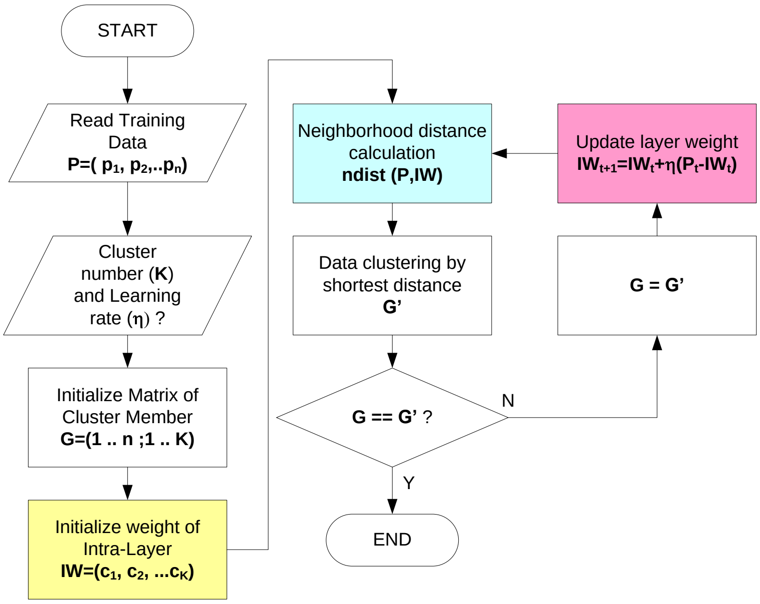

Fig. 2: SOM algorithm.

2.2. Learning Vector Quantization (LVQ)

LVQ is one type of NN, which is classified as supervised learning based on vector quantization. LVQ is also called a supervised version of SOM that reconstructs intra-layer weights with quantization techniques based on the distance function on the data [9, 10]. The LVQ architecture is shown in Figure 3.

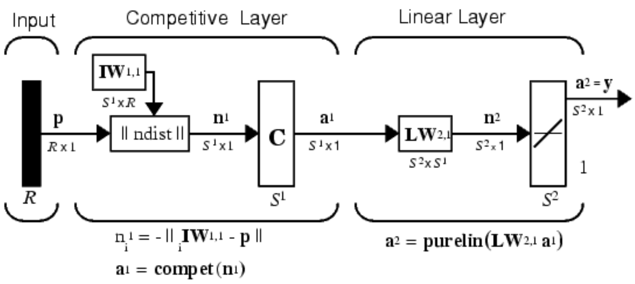

Fig. 3: LVQ architecture.

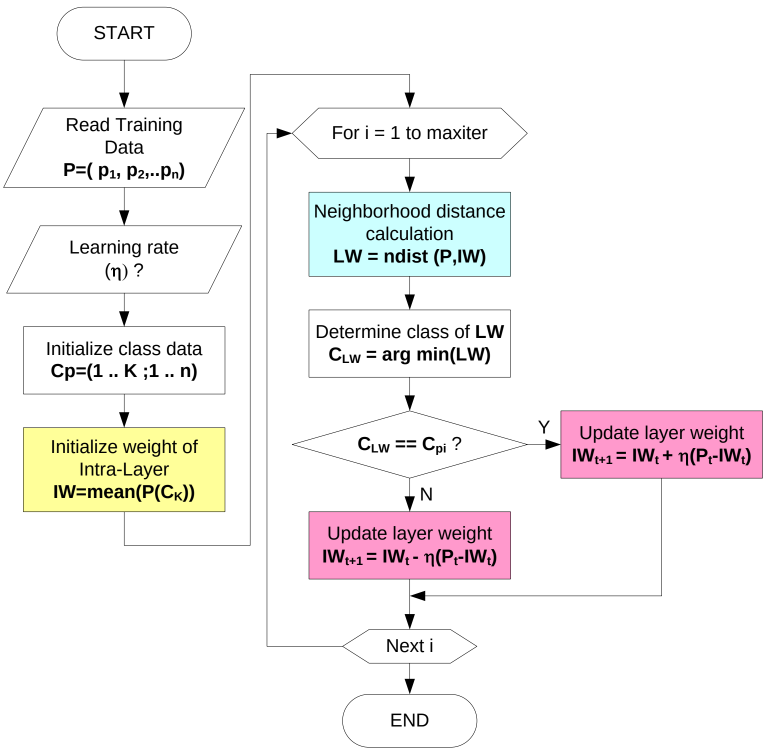

Fig. 4: LVQ algorithm.

2.3. Data Sample

Samarinda region is in geographical position, Latitude is 0.31230 - 0.70710S and Longitude is 117.04710 - 117.30390E. In the Universal Transverse Mercator (UTM) system, Samarinda's geographical position is in the 50M zone that projected into X-Y, Easting (East) is 5.0524 x 105 - 5.3382 x 105 and Northing (North) is 9.9219 x 106 - 9.9655 x 106 coordinate systems. Meanwhile, 310 samples of land price data were obtained through field observations in several locations representing all sub-districts in Samarinda. Where, land prices are in the range Rp.45.000 - Rp.8.212.198 with UTM system in Easting (East) is 5.0863 x 105 - 5.3050 x 105 and Northing (North) is 9.9349 x 106 - 9.9596 x 106 regions.

2.4. Performance measurement

2.4.1. Variance

One measure of the distribution of data which is the square of the standard deviation is variance, Equation 1.

$$Var(X) = \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})^2 \qquad (1)$$

Where x_i is the data to i, $\bar{x}$ is the average of data, n is the amount of data. In this study, the variance will be used to measure the spread of distance between cluster members to the centroid and the spacing between centroid clusters.

2.4.1. Variance

In the land price mapping, centroid is a spatial interpolation of the cluster member's entire land price. Ideally, the land price is the average land price of all cluster members expressed in Equation 2.

$$Z_{im}^{*} = \frac{1}{n}\sum\nolimits_{i=1}^{n} Z_i \qquad (2)$$

Where Z_i is the price of land from the location of members to i cluster; Z_{im}^{*} is the ideal land price; n is a lot of members of a cluster.

$$\text{MAPE} = \frac{abs\,(Z_{im}^{*} - Z_{im})}{Z_{im}^{*}}\, x\ 100\% \qquad (3)$$

3. Results and Discussion

In this experiment, clustering for several different clusters by using learning rate (η) = 0.001 has been tested. After several experiment, the best result is on 7 clusters with MAPE of 6.12% have been presented. In other words, the SOM-LVQ is better than SOM without optimization. The performance values of both methods can be seen in Table 1.

Table 1: Font Specifications for A4 Papers

Learning rate (η) = 0.001						
Cluster	SOM			SOM-LVQ		
	Variance distance between members	Variance distance between members	MAPE (%)	Variance distance between members	Variance distance between members	MAPE (%)
3	4.70E+10	4.54E+07	24.90	4.45E+10	2.61E+07	17.67
4	1.07E+13	5.79E+07	20.65	1.18E+12	3.61E+07	20.68
5	1.91E+12	2.56E+07	30.62	2.29E+11	2.75E+07	16.85
6	1.80E+12	2.10E+07	28.10	2.11E+11	2.12E+07	20.12
7*	2.74E+12	3.02E+07	17.99	3.38E+11	2.70E+07	6.12
8	4.50E+11	1.12E+07	13.75	3.39E+11	1.46E+07	21.98
9	2.98E+12	1.59E+07	19.87	3.08E+11	1.38E+07	24.86
10	1.18E+12	1.16E+07	38.36	3.22E+11	1.12E+07	12.00
11	1.94E+12	1.10E+07	24.44	4.71E+11	1.19E+07	24.64
12	6.14E+11	1.45E+07	43.68	6.87E+11	1.44E+07	13.90
Average	2.44E+12	2.44E+07	26.24	4.13E+11	2.04E+07	17.88

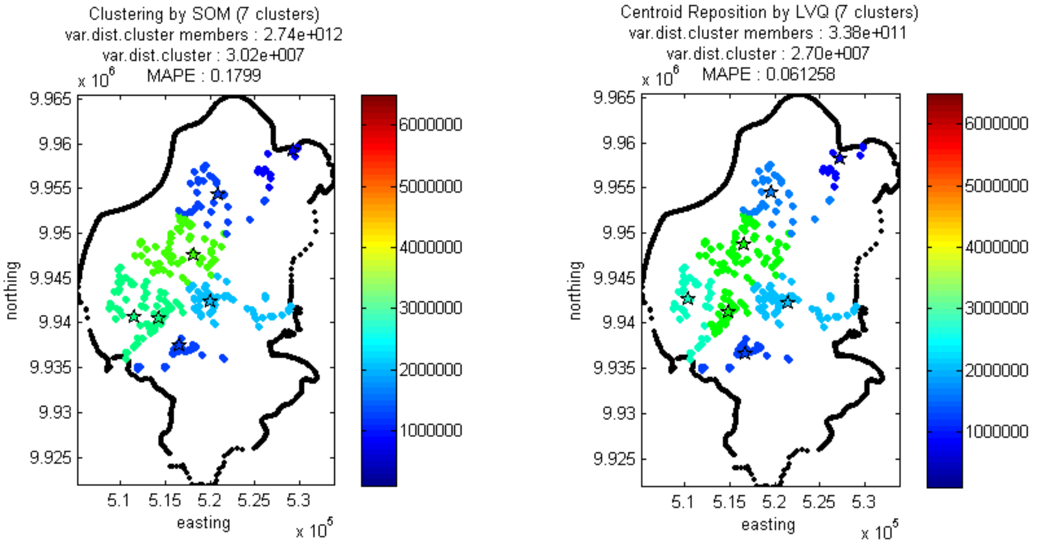

Fig. 5: Clustering results using SOM and SOM-LVQ.

4. Conclusion

This paper has presented SOM optimized with LVQ technique. Where, the intra-layer final weights generated by SOM, can be considered as centroid clusters that represent the characteristics of all cluster members. Therefore, the average value of land prices of all cluster members has been implemented as a reference in order to test the performance of SOM and SOM-LVQ. The results show that the average SOM-LVQ has a smaller MAPE average compared to SOM.

Acknowledgement

The authors would like to acknowledge the support to this project by the IDB Project Grant and Faculty of Computer Science and Information Technology (CSIT) Mulawarman University, Samarinda, Indonesia for providing financial support.

References

[1] M. Ahmadi and B. Shamoradi, "Application of geostatistical methods for mapping groundwater phosphate concentration in Eyvan plain, Ilam, Iran," Journal of Advances in Environmental Health Research, vol. 4, No.2., 2016.
[2] T. Hengi. (2007). A Practical Guide to Geostatistical Mapping of Environmental Variables.
[3] Haviluddin, A. Yunianta, A. H. Kridalaksana, Z. Arifin, B. Kresnapati, F. Rahman, A. F. O. Gaffar, H. Y. Irawan, M. Mulyo, and A. Pranolo, "Modelling of Network Traffic Usage Using Self-Organizing Maps Techniques," in 2016 2nd International Conference on Science in Information Technology (ICSITech), 2016, pp. 334-338.
[4] S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, "An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives," Expert Systems with Applications, vol. 85, pp. 169-181, 2017.
[5] Purnawansyah and Haviluddin, "K-Means Clustering Implementation in Network Traffic Activities," in 2016 International Conference on Computational Intelligence and Cybernetics, Makassar, Indonesia, 2016, pp. 51-54.
[6] M. Mahdi and I. Genc, "Defensive Islanding Using Self-Organizing Maps Neural Networks and Hierarchical Clustering," in PowerTech, Eindhoven, 2015, pp. 1 - 5.
[7] V. Golmah, "A Case Study of Applying SOM in Market Segmentation of Automobile Insurance Customers," International Journal of Database Theory and Application, vol. 7, No.1 (2014), pp. 25-36, 2014.
[8] J. Rodríguez, M. A. Medina-Pérez, A. E. Gutierrez-Rodríguez, R. Monroy, and H. Terashima-Marín, "Cluster validation using an ensemble of supervised classifiers," Knowledge-Based Systems, vol. 145 (2018), pp. 134–144, 2018.
[9] B. Mokbel, B. Paassen, F.-M. Schleif, and B. Hammer, "Metric learning for sequences in relational LVQ," Neurocomputing, vol. 169 (2015), pp. 306–322, 2015.
[10] B. J. Jain and D. Schultz, "Asymmetric learning vector quantization for efficient nearest neighbor classification in dynamic time warping spaces," Pattern Recognition, vol. 76 (2018), pp. 349–366, 2018.