A DISTRIBUTED APPROACH FOR SUPERVISED SOM AND APPLICATION TO FACIES CLASSIFICATION

Hoa Dinh Nguyen
Khoa Công nghệ thông tin, Học viện Công nghệ Bưu chính Viễn thông

Abstract: This study proposes a distributed classification framework, which adapts supervised Self-Organizing Maps (SOM) as base learners. The supervised SOM is the integration of the SOM algorithm with the Learning Vector Quantization (LVQ) algorithm, so called SOM-LVQ model. Multiple SOM-LVQ models are created using different feature subsets, each of which represents one different local information source. This approach aims at utilizing the information hidden in smaller feature subsets, that cannot be obtained if the data is processed as a whole original feature combination. The outputs from all different local supervised SOMs are fused together using some specific fusing rule to provide the final decision on the class label of the input data. This proposed distributed classification approach is applied on well-log data to determine the facies classes of the log samples. Experiments are conducted based on the well-log data-set collected from Cuu Long basin, which is an early Tertiary rift basin located off the southeast coast of Vietnam. The experimental results show that the newly proposed distributed supervised SOM-based classification approach outperforms not only the single supervised SOM model but also some other commonly used machine learning models in terms of accuracy rate. It is also shown that the distributed approach is more useful when the number of input features is high, and is a flexible solution for many real-life classification problems.

Keywords: facies classification, distributed supervised SOM, learning vector quantization, self-organizing map.

I. INTRODUCTION

Neural networks (NN) have been a well-known machine learning model recently. NN is a combination of computational nodes, known as neurons, connected to each other to model a relationship between input and output data. Each node contains many static parameters, known as weights and bias, and a transfer function to map input information to output variable. Changing the value of parameters of all neurons will lead to the change in the behavior of the network [1, 2]. There are various ways of training a neural network depending on its real applications or purposes.

Competitive learning is an unsupervised learning method, each iteration of which determines one winning neuron. The weights of the winning neuron are adjusted accordingly with the input data, which is known as winner-take-all procedure. In order to avoid the domination of a small number of neurons during learning process, conscience can be applied [3].

Self-organizing map (SOM), which was first introduced by Kohonen in 1982 [4], is an excellent example of competitive learning models. SOM is an efficient tool to visualize high dimensional data in a much lower dimensional representation [5]. SOM is applied in many fields of studies, including bio-informatics, textual document analysis, outlier detection, financial technology, robotics, pattern recognition, and much more [6]. Though it is well known as an unsupervised learning model, there have been many approaches to adopt SOM as a supervised learning algorithm to solve classification problems. Supervised SOM can be used to analyze textual documents in [7] and [8]. WEBSOM [9] is an extension of SOM developed by Kurasova to classify different textual document collections. Hierarchical Overlapped SOM [8] is another version of supervised SOM developed for handwritten character recognition. Stefanovič and Kurasova [10] proposed the travelling salesman approach to enable SOM to cover the outlier detection problem. KNN can be combined with SOM to create a supervised version of SOM [11]. K-means algorithm can also be adopted to form a simple version of supervised SOM [12]. Hoa [6] proposed the supervised SOM-LVQ model, which is the integration model between SOM and the modified learning vector quantization (LVQ) algorithm. He also improved its classification ability by using the adaptive boosting method.

In this paper, a new distributed approach is introduced to enhance the classification performance of SOM-LVQ algorithm. Different supervised SOM classifiers are deployed in different local data sources to produce multiple local classification outputs. Those local outputs are then fused to provide final classification decision on the input data. This new distributed supervised SOM method is then applied on real well-log data to determine the facies label for each individual log data sample. Experimental results illustrate the advancement of the new approach over many conventional methods.

This paper is organized as follows. Section 2 briefly describes the supervised SOM model. The proposed distributed approach for supervised SOM and its application for facies classification is presented in Section 3. Section 4 includes all experimental results and

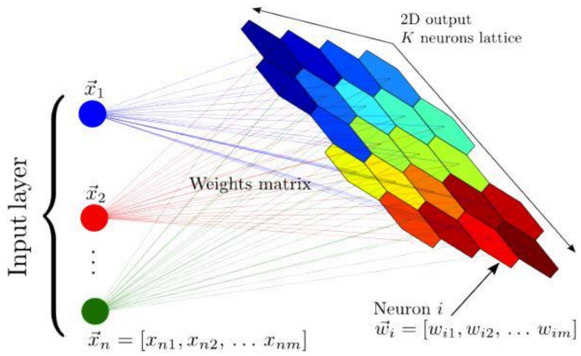
Contact author: Hoa Dinh Nguyen,
Email: hoand@ptit.edu.vn
Manuscript Received: 11/2021, revised: 12/2021, accepted: 12/2021.

discussion of the proposed approach. Section 5 concludes what have been accomplished in the research.

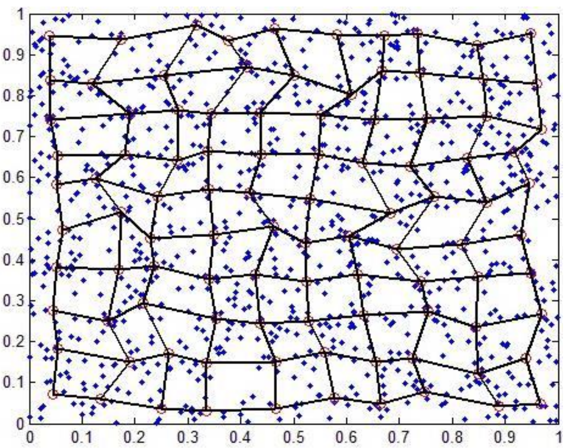
II. SUPERVISED SELF-ORGANIZING MAP

A. *Self-organizing map*

The self-organizing map (SOM) composes of a set of nodes (or neurons) connected to each other via the topology of rectangle or hexagon. Each neuron contains a vector of weights of the same dimension as the input data. There are usually to types of SOM representation [14], semantic representation and spatial positioning representation. Semantic representation includes neurons ordered in a network having 1 or 2 dimensions. The input data is mapped into a K neuron network. Each neuron can represent multiple input data samples. The network of neurons is painted with different colors representing different clusters of input data. Each color corresponds to a set of similar input samples.



a) Semantic representation



b) Spatial position representation

Figure 1. Representations of an SOM [10]

In spatial positioning representation, each neuron represents a point in data space. After training process, all neurons are allocated in different region in data space. The number of neurons together with the spatial distances among neighboring neurons in each spatial region represent the distribution of the real data within that location. If the number of neurons is much smaller than the number of data samples, each neuron can also be considered as a data cluster. Figure 1 illustrates the two representations of SOM.

The training process goes through many iterations to update the weights of neurons in the network. Each iteration is also called a competition. In the t^{th} iteration, a data vector x is randomly selected from the input data set X. The algorithm determines the "winning" neuron w_c, which is also known as the best matching unit (BMU), for the sample x. Euclidean distance, $D(x, w_i)$, is commonly used as the measure to determine how close a neuron i is to the input sample x. The BMU is defined as the neuron having the smallest distance to the sample. The weight vector, $w_i(t)$, of the BMU i is then updated by a learning rule:

$$w_i(t) = w_i(t-1) +$$
$$+\alpha(t-1).\, h_i^c(t-1).\, D\big(x(t-1), w_i(t-1)\big)$$

Where, t is the current iteration; $\alpha(t-1)$ is the learning rate at the previous iteration, which normally decreases during the training process, $\alpha(t) = \frac{\alpha_0}{1+decayrate*t}$; $h_i^c(t-1)$ is the neighborhood function at the previous iteration. The neighborhood function determines which nearby neurons are updated along with the BMU i. There are two popular neighborhood functions used in the literature, which are bubble and Gaussian functions [15]. In bubble function algorithm, all neurons within the neighborhood region are updated with the same rate, i.e.

$$h_i^c(t) = \begin{cases} 1 \text{ if neuron c is inside neighborhood} \\ 0 \text{ if neuron c is outside neighborhood} \end{cases}$$

In Gaussian function approach, the updating rate of one neighboring neuron depends on how close it is between that neuron and the BMU i. The Gaussian neighborhood function is defined as

$$h_i^c(t) = e^{-\frac{\left(D(w_c, w_i)\right)^2}{2\eta_i(t)^2}}$$

Where, $D(w_c, w_i)$ is the distance from neuron c to the BMU i; η_i is the neighboring radius around the BMU i. The neighboring radius is also used to determine the neighborhood region in bubble function above. In order to speed up the convergent speed of the training process, the neighboring radius is a declining function starting from an initial value η_0, as $\eta_i(t) = \frac{\eta_0}{1+\text{decayrate}*t}$.

During the SOM training process, in each iteration, only the BMU and its limited number of neighboring neurons are updated. If the SOM size is large, there is a great chance that many neurons are not updated if they are initialized far away from the data samples. Those are dead neurons. Conscience is a technique applied to prevent one neuron from winning in so many iterations, and help other neurons are more likely to win. This method simply adds or subtracts a small fraction (called bias, which can be negative) to the distance from the neurons to the input data. The more often the neuron wins, the larger its bias is, making it less likely to win. The other neurons will have their bias reduced in the loops where they don't win. The bias function for each neuron i can defined as

$$b_i^{new} = \begin{cases} 0.5 b_i^{old} \text{ if neuron i does not win} \\ 1.5 b_i^{old} \text{ if neuron i wins} \end{cases}.$$

SOM is good at visualizing the data, and is an unsupervised learning method mainly used in clustering problems. In order to tackle supervised classification problems, traditional SOM must be modified to increase the classification accuracy.

B. Supervised SOM-LVQ algorithm

Learning vector quantization (LVQ) is developed from SOM algorithm. LVQ is used for supervised learning applications. The training rule applied in the LVQ is an association method for training the SOM network in a supervised learning manner [2]. Specifically, each neuron in the competitive network will be assigned the label of the data cluster it represents. Many different neurons can have the same data class label. After training, a given the new input vector will be assigned the class label of the neuron closest to it.

In this research, the two-stage SOM-LVQ model for classification problem introduced in [13] is adopted. In this supervised model, the training data is first clustered by an SOM algorithm. The label for each neuron is then assigned according to the class of the nearest data sample from the training set. LVQ is finally applied to train the whole labeled network. After training process, the labeled neurons are moved closer to the regions dominated by the data having the same class label. In case the training data is not distributed in specific regions for each class label, the label neurons are located close to some local data samples.

III. DISTRIBUTED SUPERVISED SOM AND APPLICATION TO FACIES CLASSIFICATION

In this research, a distributed supervised SOM framework is proposed and applied on facies classification problem. The data-set is assumed to be collected from different types of sensors. Each sensor contributes one piece of information, and is presented by a sub set of features. One local supervised SOM is built based on the sub feature set from each local sensor. The outputs of all local supervised SOMs are then fused to produce the final decision on the class of the input data. The distributed supervised SOM method is depicted in Figure 2.

A. Facies classification

Facies are the overall characteristics of a rock unit that reflect its origin and differentiate the unit from others around it. Each facies class distinguishes itself from other classes based on mineralogy and sedimentary source, fossil content, sedimentary structures and texture. In reservoir characterization and simulation, the most important facies properties are the petrol-physical characteristics which control the fluid behavior in it [16]. Some certain facies classes exhibit characteristic measurement signatures that help facilitate the prediction of some important properties such as permeability, porosity, and liquid content. Hence, correct presentation of facies classes for well-log data is an important and challenging task for oil and gas engineers.

Deep-water reservoirs are deposited in a wide range of depositional environments, and exhibit a variety of temporal and spatial scales. Detailed core description of reservoir intervals allows identification of facies and strati-graphic units at multiple scales. At the small scale, lithofacies are rocks with similar lithology, sedimentary structures and rock properties. In many depositional environments these can be grouped together into depositional facies (or depofacies) that represent genetically related deposits with predictable dimensions and relationships with other depofacies. Characterization of depofacies is important for reservoir modeling and predicting reservoir continuity away from the borehole. For example, a channel axis depofacies may consist of conglomerate lithofacies and massive sandstone lithofacies, and have a predictable range of widths and predictable relationship with channel margin depofacies. Identification of such geologic units from core data and manual interpretation from wire-line logs has been established, but core sampling is typically limited, and manual interpretation of wire-line logs involves some degree of uncertainty and subjectivity [17].

Recently, most of the researches on facies classification are based on well-log data. It is desirable to find either the relationship between well-log measurements and facies classes or well-logs patterns corresponding to each class representation. There have been a lot of methods based on wire-line log measurements including statistical approaches, fuzzy methods, and artificial neural networks [17].

B. Facies classification based on distributed supervised SOM

In this study, supervised SOM is applied on facies classification in a distributed manner using well-log data. General speaking, instead of using the whole data features collected during the drilling process to generate a generalized supervised SOM model for predicting the class label of each well-log sample, multiple local SOMs is generated from many individual local feature subsets. The output of all local SOMs are then fused to provide a final decision on the facies label of the log data.

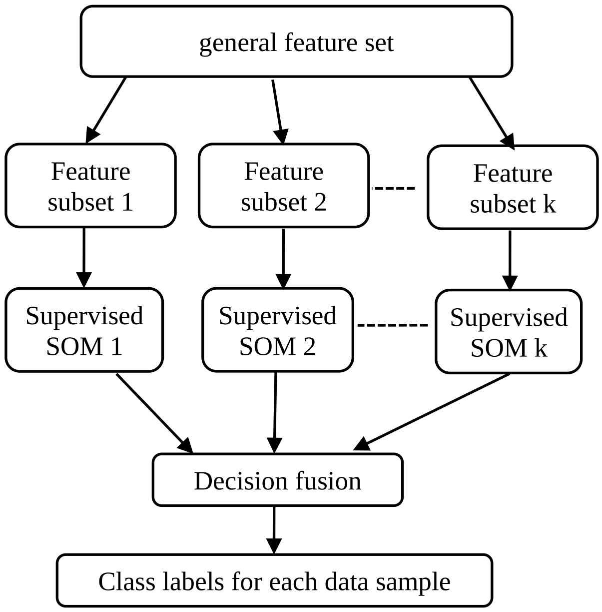


Figure 2. distributed supervised SOM framework

A data set with full attribute values often has many different distributions in the data space, each of which may have different and complex properties. In addition, one general SOM-LVQ network can only organize the neurons according to the distribution of general data clusters. This means that one single SOM-LVQ model can ignore the distribution of complex data clusters located in a larger data set. Therefore, with complex data, especially geological data, if only one SOM-LVQ network is used to model the entire data set, its efficiency in data interpretation is not high. The main advantages of a distributed approach for classification using SOM-LVQ model are two folds. First, the original data set will be examined in smaller dimensional sub-spaces to find such data distributions that would be difficult to discern if viewed in the general all-dimensional space of the original data. Second, there is always an overlap between the data of different facies classes. The majority part of missed classified well-log samples fall into the overlapping regions of different facies. By using local information from different feature subsets, the information of each well-log data is put under different angles and processed individually. This approach can help exploit all valuable information for the classification process. Each local supervised SOMs provides different probabilities that the data sample belongs to each of possible facies labels.

C. Decision fusion

Multiple supervised SOM models produce multiple local decisions from different data feature subsets. An efficient fusion process, which combines all these local decisions, may influence the classification performance of the whole system. There are normally two main rules for decision fusion process, majority voting and confidence score based fusion [18].

In majority voting approach, the final class label is selected as the one having the most number of local decisions on it. A modification version of majority voting is to select the class label having the highest weights among all local classifiers. There can be many ways to define the weight for each local decision. Regarding supervised SOM classifiers, where the label of the test data is decided based on the neuron closest to that data, the reasonable way to calculate the weight of the output is the reciprocal of the best matching distance.

In confidence score based approach, the final class label is assigned based on the highest average value of all confidence scores from local classifiers. In other words, each local output is associated with a confidence value. The confidence score of data sample x belonging to class c can be defined based on the distances from neurons and data sample as follows.

$$S_c(x) = \frac{\sum_{i=1, y_i=c}^{k} \frac{1}{D(w_i, x)}}{\sum_{i=1}^{k} \frac{1}{D(w_i, x)}}$$

Where, y_i is the label of neuron i; k is the number of neurons nearest to the data sample x; $D(w_i, x)$ is the distance between neuron i and sample x. The average of all confidence scores of local classifiers having the same output label is used to determine the final label of the data sample at the fusion stage.

IV. EXPERIMENTAL RESULTS

A. Dataset

The proposed distributed classification approach based on supervised SOM is evaluated using the well-log data collected from three wells in Cuu Long basin. There are 4 facies classes in the data-set. The distribution of data sample in each facies class in the data-set is presented in Table 1.

The data-set is divided into two subsets, training data subset contains all samples from well 1 and well 2, and testing data subset includes the data from well 3. It can be seen that the number of facies labels 1 and 3 is small, which means those two facies rarely present in those three wells. In this research, all samples of classes 1 and 3 are omitted, which means only facies labels 0 and 2 are used.

Table 1. The distribution of data samples in the dataset

Information	Data in well 1	Data in well 2	Data in well 3
Number of samples	438	457	232
Samples of class 0	249	194	146
Samples of class 1	5	13	0
Samples of class 2	169	211	80
Samples of class 3	15	39	6

B. Experimental scenarios

The accuracy score is used to evaluate the classification performance of the system. The accuracy is calculated as the ratio of correct decision number to total number of testing samples.

$$Accuracy = \frac{N_{true\ prediction}}{N_{total\ prediction}}$$

This evaluate metric works well in case there is a balance among the sample numbers of all classes, which is true for this data-set.

The experiments have been conducted to evaluate the classification performance of the proposed distributed classification approach, and compare it with the centralized supervised SOM model.

The distributed model adopts multiple small sized SOM-LVQ models working on multiple feature subsets. A total of 10 randomly selected feature sets are generated. In order to randomly create different feature sets, each feature is first assigned a weight based on its correlation with the output labels. Higher correlation means higher weights. The selection process is conducted in 10 iterations, in each of which one feature is selected. The features having higher weights tend to be selected more times than others. This means high correlated features are expected to present in many feature subsets. 10 local supervised SOM models are created based on 10 feature subsets. Majority voting based on maximum average weights are used in the decision fusion stage.

A general traditional supervised SOM is generated from the whole feature set. The training process for single supervised SOM model is the same as that for multiple distributed SOM models. The only difference is that the number of training iterations for multiple distributed SOM is smaller than the single SOM model due to their simpler structures. The training process is repeated multiple times and the best models for each scenario are recorded.

C. Results and discussion

Table 2 presents the training parameters of all supervised SOM models implemented in the experiments together with their classification accuracy scores.

The experimental results in table 2 show that the combination of multiple smaller sized SOM models works better than one single complicated SOM model. This result is more meaningful considering the training time as well as the processing time of the classification models. In each training iteration, the distributed models only have to work on 9 neurons compared with 100 neurons of the single models. Additionally, distributed models can work in parallel, which helps speed up the computational process significantly. The distributed supervised SOM classification system is more flexible than the single supervised SOM model since different fusion rules can be applied to adjust the classification performance of the whole system. This is meaningful in real practice where different real data-sets may require different system structure setups.

Table 2. Model parameters and accuracy scores.

Parameters	Single SOM-LVQ model	Multiple distributed SOM-LVQ models
Number of SOM models	1	10
SOM size	10x10	3x3
Weight initialize method	PCA	PCA
Number of training iterations	2000	2000 (200 for each local model)
Initial updating rate	1	1
Updating rate declining coefficient	1	1
Neighborhood function	Bubble	Bubble
Neighborhood radius	1.5	1.5
Accuracy score	0.9181	0.9397

The experiment is further conducted on different classification machine learning models, such as decision tree, Naïve Bayes, K nearest neighbor (K=5) and Random forest (n=10). The results are presented in Figure 3.

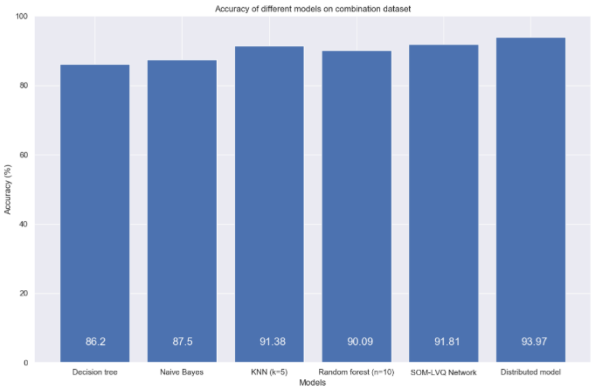


Figure 3. Accuracy scores of different classification models on the same data-set

The experimental results show that the distributed SOM-LVQ model works best in comparison with other machine learning models. The decision tree and Naïve Bayes models both have the lowest accuracy scores, while KNN and random forest have almost the same classification performance.

The distributed SOM-LVQ model also has some additional advantages over other traditional machine learning models. Compared with KNN model, the distributed SOM-LVQ model requires less memory resource. Specifically, KNN model needs to store all training sample to find the nearest data during the testing process, while the SOM-LVQ models only need to store the weights of their neurons after training process. The number of the neurons is normally much smaller than the number of training data samples.

Distributed SOM-LVQ model and random forest have some similar characteristics. Random forest may have an advantage over the distributed SOM-LVQ model when it can process the training data in multiple small data subsets, in which some special properties of the training data can be recognized. However, this distributed SOM-LVQ model can be further modified to cope with different smaller subsets of training data, from which more local supervised SOM classifiers can be created. As a results, the classification performance of distributed SOM-LVQ model can be adjusted accordingly.

V. CONCLUSIONS

In this paper, a new distributed classification approach based on supervised SOM is introduced. Multiple SOM-LVQ models are generated from different local feature sets and their decisions are fused to produce the final class label for the input data. Each local supervised SOM can exploit different pieces of information presented in its smaller feature subset. Fusion stages combines all local classification decisions from multiple supervised SOM models to output the final class label for the test sample. Multiple pieces of the information are combined using different fusion rules, which open up the flexibility of the proposed system to cope with different situations in real life practice. This newly distributed classification system is applied on facies classification problems, which is a well-known and difficult issue in geology. The experimental results show that the distributed classification approach for supervised SOM works better than the traditional single supervised SOM model as well as many other conventional machine learning models. In our further research, more experiments with bigger number of data features will be conducted to investigate the usefulness and flexibility of this distributed approach.

REFERENCES

[1] J.-S. R. Jang and C.-T. Sun, "Neuro-Fuzzy Modeling and Control," Proceedings of The IEEE, pp. 378-406, 1995.

[2] M.T. Hagan, H.B. Demuth, M.H. Beale, O.D. Jesus. “Neural network design (2nd edition)”, ISBN-10: 0-9717321-1-6, ISBN-13: 978-0-9717321-1-7, 1996.

[3] Mia Louise Westerlund, "Classification with Kohonen Self-Organizing Maps," Soft Computing, 2005.

[4] T. Kohonen. "Self-Organized Formation of Topologically Correct Feature Maps", Biological Cybernetics. 43 (1), pp. 59 - 69, 1982.

[5] N. Skific and J. Francis, “Self-Organizing Maps: A Powerful Tool for the Atmospheric Sciences”, 2012.

[6] S. Haykin, “Neural Network and Learning Machines”, 3rd ed., Pearson, 2008.

[7] A. Rauber and D. Merkl, "Automatic Labeling of Self-Organizing Maps: Making a Treasure-Map Reveal Its Secrets", Methodologies for Knowledge Discovery and Data Mining, p. 228 – 237, 1999.

[8] P.N. Suganthan. "Hierarchical overlapped SOM's for pattern classification", IEEE Transactions on Neural Networks, vol. 10, pp. 193-196, 1999.

[9] O. Kurasova, "Strategies for Big Data Clustering", 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, 2014.

[10] P. Stefanovič, O. Kurasova. "Outlier detection in self-organizing maps and their quality estimation", Neural Network World, 28 (2), pp. 105-117, 2018.

[11] L.A. Silva, E.D.M. Hernandez. "A SOM combined with KNN for classification task", Proceedings of the International Joint Conference on Neural Networks, pp. 2368-2373, 2011.

[12] M. Mishra, H.S. Behera. "Kohonen Self Organizing Map with Modified K-means clustering For High Dimensional Data Set", International Journal of Applied Information Systems, pp. 34-39, 2012.

[13] Hoa Dinh Nguyen, "A Boosting Classification Approach Based on SOM". Journal of Science and Technology on Information and Communications, Vol 1 No 2, 2020.

[14] M. C. Kind and R. J. Brunner, "SOMz: photometric redshift PDFs with self-organizing maps and random atlas," 2013.

[15] N. Wangsoh, W. Watthayu and D. Sukawat, "Appropriate Learning Rate and Neighborhood Function of Self-organizing Map (SOM) for Specific Humidity Pattern Classification over Southern Thailand," International Journal of Modeling and Optimization, pp. 61-65, 2016.

[16] O. Serra. "Fundamentals of well-log interpretation. Volume 2: the interpretation of log data", ENSPM, Etudes et Productions Schlumberger, Montrouge, France, 632p., 1985.

[17] C.M. Gifford, A. Agah." Collaborative multi-agent rock facies classification from wireline well-log data", Engineering Applications of Artificial Intelligence, 23, pp.1158–1172, 2010.

[18] A. Mandelbaum and D. Weinshall, "Distance-based Confidence Score for Neural Network Classifiers," 2017.W.D.Patrick, Natural sciences citations and references. Available online at ftp://ftp.freesoftware.com/pub/ tex/ctan/macros/contrib/supported/natbib, (1997) 2-10.

MỘT CÁCH TIẾP CẬN PHÂN TÁN CHO BẢN ĐỒ TỰ TỔ CHỨC CÓ GIÁM SÁT VÀ ỨNG DỤNG VÀO PHÂN LOẠI FACIES

Tóm tắt: Bản đồ tự tổ chức (SOM) là một mạng nơ-ron nổi tiếng với khả năng biểu diễn dữ liệu đa chiều. SOM còn được sử dụng cho các bài toán phân loại bằng cách kết hợp với một số thuật toán huấn luyện phù hợp khác. Nghiên cứu này đề xuất một cách tiếp cần phân toán cho bài toán phân loại sử dụng SOM có giám sát. SOM có giám sát được xây dựng từ sự tích hợp giữa thuật toán SOM với thuật toán huấn luyện lượng tử hóa vectơ (LVQ), mô hình SOM-LVQ. Các mô hình SOM-LVQ khác nhau được tạo ra từ các tập thuộc tính con khác nhau. Mỗi tập thuộc tính con đại diện cho một nguồn thông tin trong hệ thống. Cách tiếp cận này cho phép sử dụng thông tin tiềm ẩn trong các tập thuộc tính con. Những thông tin này không thể tận dụng được nếu dữ liệu được xử lý trên bộ thuộc tính nguyên gốc. Kết quả đầu ra của các SOM thành phần này sẽ được hợp nhất với nhau bằng một số quy tắc hợp nhất cụ thể để đưa ra quyết định cuối cùng về nhãn lớp của dữ liệu đầu vào. Cách tiếp cận phân tán này được áp dụng trên dữ liệu giếng khoan để xác định các nhãn facies cho các mẫu đất trong giếng. Các thí nghiệm được thực hiện dựa trên tập dữ liệu giếng khoan được thu thập từ thềm địa chất Cửu Long, đây là một lưu vực rạn nứt sơ khai bậc III nằm ngoài khơi bờ biển Đông Nam Việt Nam. Kết quả thực nghiệm cho thấy cách tiếp cận phân toán cho bài toán phân loại dựa trên SOM có giám sát được đề xuất có độ chính xác tốt hơn không chỉ so với mô hình SOM có giám sát đơn lẻ mà còn so với một số mô hình học máy thường được sử dụng khác. Nó cũng cho thấy rằng cách tiếp cận phân tán sẽ hữu ích hơn khi số lượng các tính năng đầu vào là lớn, và là một giải pháp linh hoạt cho nhiều bài toán phân loại trên thực tiễn.

Từ khóa: bản đồ tự tổ chức, huấn luyện lượng tử hoá vector, phân tán, phân loại facies

Hoa Dinh Nguyen earned bachelor and master of science degrees from Hanoi University of Technology in 2000 and 2002, respectively. He got his PhD. degree in electrical and computer engineering in 2013 from Oklahoma State University. He is now a lecturer in information technology at PTIT. His research fields of interest include dynamic systems, data science, and machine learning.

Email: hoand@ptit.edu.vn