

# Application of self organizing map in construction, geology and petroleum industry

Pham Son Tung, Truong Minh Huy, Pham Ba Tuan

**Abstract**— In recent years, Artificial Intelligence (AI) has become an emerging subject and been recognized as the flagship of the Fourth Industrial Revolution. AI is subtly growing and becoming vital in our daily life. Particularly, Self-Organizing Map (SOM), one of the major branches of AI, is a useful tool for clustering data and has been applied successfully and widespread in various aspects of human life such as psychology, economic, medical and technical fields like mechanical, construction and geology. In this paper, the primary purpose of the authors is to introduce SOM algorithm and its practical applications in geology and construction. The results are classification of rock facies versus depth in geology and clustering two sets of construction prices indices and building material costs indice.

**Index Terms**— Self Organizing Map, Hierarchical Clustering, Geology, Well logging, Construction Economics.

## 1 INTRODUCTION

SOM (Self-Organizing Map) is a result of Unsupervised learning algorithm – this algorithm bases on the structure of input data in order to reduce the quantity of data dimensions or to cluster this data into different sections without precise result at the output [1, 2]. SOM's result is a clustering map including a number of nodes (cells) with similar characteristic into the same field or section [2, 3]. The biggest advantage of SOM is that it illustrates visually multi-

dimension (multi-characteristic) input data on two-dimension map but still retains the essence of original data. SOM prevails over traditional algorithms on clustering functions [4]. Specifically, kinds of data having similar tendency or symptoms are grouped together by SOM while traditional algorithms just determine average values, variance, standard deviation and data frequency. Thus, thanks to SOM, readers will have more visual assessments in order to give more suitable conclusions. The primary purpose of the paper is to introduce SOM algorithm and its practical applications in geology and construction. In geology, geophysical surveys are conducted to determine rock formation characteristics versus depth such as density, sonic travel time, and gamma rays. After well logging, a substantially large data sheet containing these rock properties with respect to every depth is obtained. From this logging data sheet, the rocks at specific depth will be categorized into different facies. The conventional interpretation process would take a large amount of time and efforts. SOM is definitely a solution to this problem because it eliminates the time-consuming manual interpretation. SOM will effortlessly characterize rock intervals into particular facies simultaneously.

In construction, catching up with and being able to forecast the change in construction as well as major materials cost are considered not only huge advantages but also key factors affecting decisions of contractors and investors. SOM algorithm brings us visual look at the changing tendency of one factor if other factors in the same section change.

## 2 METHODOLOGY

### 2.1 Self-Organizing Maps (SOM) – Kohonen network

Self-Organizing map or SOM, initialized by Professor T. Kohonen since the early of 1982s, is undeniably useful clustering tool. Since it was introduced, SOM has been applied widely in

*Manuscript Received on August 7th, 2017. Manuscript Revised December 25th, 2017*

Pham Son Tung - Faculty of Geology and Petroleum Engineering, Ho Chi Minh city University of Technology – VNU-HCM (e-mail: phamsontung@hcmut.edu.vn).

Truong Minh Huy- Faculty of Geology and Petroleum Engineering, Ho Chi Minh city University of Technology – VNU-HCM.

Pham Ba Tuan - Faculty of Geology and Petroleum Engineering, Ho Chi Minh city University of Technology – VNU-HCM.

\* Corresponding author: Email: [phamsontung@hcmut.edu.vn](mailto:phamsontung@hcmut.edu.vn).

various fields such as psychology, economy, medical care, engineering and a vast majority of other professionals [1, 3].

SOM algorithm helps simplifying data, reducing dimensions (properties) of input data and thus SOM results in a map with lesser dimension than the former, usually 2-D map. SOM is built on a foundation of Unsupervised Learning algorithm on input data. An example pertaining to a simple Kohonen network with the size of 4x4 (16 nodes) is shown in figure 1. Each node from the map represent a vector with as many dimensions as those of the input vectors (data); i.e., if the input vector has n dimensions  $\vec{V}(V_1; V_2; \dots; V_n)$ , then the weighted vector of a node would contain n dimensions  $\vec{W}(W_1; W_2; \dots; W_n)$  [2]. At the beginning of SOM algorithm, weighted vectors in the Kohonen network have random values associated with different properties ranging from 0 to 1. After each iteration, these random values will be adjusted to a random input vector chosen from the normalized input data [3]. The number of iterations is usually 500 times bigger than the quantity of network nodes [1]. The following section describes how SOM algorithm works in more detail.

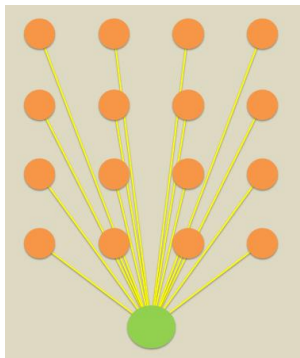


Figure 1. 4x4 Kohonen network

### 2.2 Algorithm sequence

*Step 1:* Build Kohonen network with configuration of  $n \times n$  (nodes) with random values for each property ranging from 0 to 1.

*Step 2:* Normalize input data in order to determine relative effect level between properties. In other words, eliminating the unit of each property with the following formula:

$$b_i = \frac{v_i - v_{min}}{v_{max} - v_{min}} \quad (1)$$

*Step 3:* Choose randomly one input vector from normalized input vector series. Determine the distance from chosen input vector to each node on

Kohonen network according to Euclidean distance formula:

$$Distance = \sqrt{\sum_{i=0}^n (b_i - W_i)^2} \quad (2)$$

*Step 4:* Obtain the node with the smallest distance to chosen input vector and name this node Best Matching Unit (BMU)

*Step 5:* From BMU, determine neighborhood radius with formula  $\sigma_o^2 = n$  (3)

*Step 6:* Calibrate all the nodes within neighborhood radius according to below formula:

$$W' = W_o + \theta_o \cdot L_o \cdot (V_o - W_o) \quad (4)$$

With:  $L_o$  – initial learning rate (usually 0.5)

$$\theta_o - \text{initial effect level} = \exp\left(-\frac{distance^2}{2\sigma_o^2}\right) \quad (5)$$

*Step 7:* After calibration, continue to conduct iteration from step 3, with t increase by 1. This step is repeated over again until reaching the desire number of iterations N. With increasing t, these below parameters change concurrently:

$$\sigma(t) = \sigma_o e^{-\frac{t}{\lambda}} \quad (6)$$

$$\text{With: } \lambda - \text{constant} = \frac{N}{\log(\sigma_o)} \quad (7)$$

$N$  – number of iterations

$$\theta(t) = e^{-\frac{distance^2}{2\sigma^2(t)}} \quad (8)$$

$$L(t) = L_o e^{-\frac{t}{N}} \quad (9)$$

$$W(t+1) = W(t) + \theta(t) \cdot L(t) \cdot (V(t) - W(t)) \quad (10)$$

### 2.3 SOM algorithm's results

From a network with nodes containing random values of different parameters, after experiencing a training progress with determined number of iterations, SOM algorithm results in a network with the same number of nodes but these nodes have been modified so that the closer the nodes are, the more similarity in characteristics they would be [2]. Based on Euclidean formula about distance calculated from a random input vector to a node on Kohonen network, a specific node containing the interested elements could be identified. We eventually know how each node in the resultant map represents which input vectors from initial input data. Thus, SOM algorithm not only illustrates huge input data series into self-organizing map with much lesser nodes but also visually highlights similar behaviors or characteristics of input data. Besides, SOM results also facilitate Hierarchical clustering method in reducing the amount of time for algorithm calculation and number of iterations as well. In the next section, the authors will present Hierarchical clustering method and its application via SOM

algorithm.

2.4 Hierarchical clustering algorithm

Hierarchical clustering is an algorithm in which normalized input data is clustered according to Euclidean distance formula [5]. For instance, if initial input data has 5 elements then the number of segments created from 5 element is  $(5 - 1)! = 4! = 24$ , we then have to calculate the distance for 24 times. The result of Hierarchical algorithm is a chart as shown in figure 2. More branches represent more detail in clustering or more groups we could obtain. Specifically, in figure 2, if an interpreter just wants to divide the data into two groups then he just needs to look in position 1 with group (A, B, C) and group (D, E, F, G). On the other hand, if the interpreter decides to divide the data into five groups instead of two, then the results will be (A), (B, C), (D, E), (F) and (G) groups. Therefore, it is obvious that the advantages of Hierarchical algorithm are simple theoretical basis, easy computation and direct visual results. However, if the size of input data is relatively large, for example 100.000 elements, the number of times to calculate Euclidean distances would increase substantially (up to 99.999! times) leading to unrealistic computation time.

In order to solve the above problems of Hierarchical algorithm, then authors have decided to use SOM algorithm. First, SOM algorithm will simplify the input data into a map (network) with substantially lesser nodes, and then Hierarchical algorithm will be applied on this map to cluster these representative nodes. In this paper, we will apply both algorithms in two fields: geology and construction.

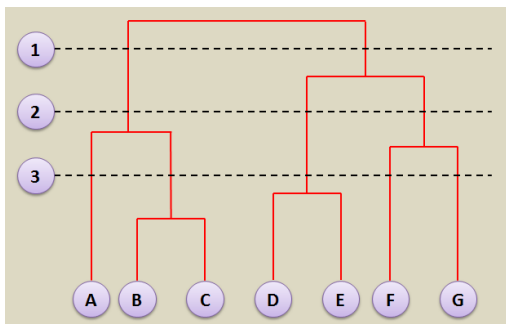


Figure 2. Example Hierarchical algorithm result

3 APPLICATIONS

3.1 Application of SOM algorithm in geology

In geology, SOM algorithm is applied to

determine rock facies. From initial input data containing different properties acquired from well logging processes such as neutron log, sonic log, density log, gamma ray, etc... Groups of depth associated with similar log characteristics are located into the same node of a Kohonen network using SOM. Combining with Hierarchical algorithm, the ultimate result will be a self-organizing map with nodes divided into separate sections representing different kinds of rock facies (the number of rock facies will be determined based on experience of the interpreter).

3.2 Input data

Figure 3 represents well geo-physical data of four main properties with start depth at 1843.278 ft and end depth at 4248.302 ft with an increment of 0.154 ft. The values in column DT, NPHI, RHOZ and GR are the results of logging activity namely sonic log, neutron log, density log and gamma log (gamma ray).

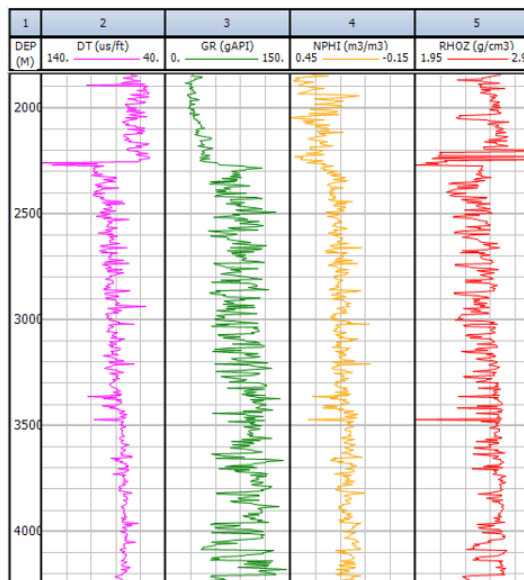


Figure 3. Well logging data versus depth

3.3 Result

From initial input data, SOM algorithm has clustered into nodes in Kohonen network. Here, authors have chosen a 30x30 (900 nodes) network to represent the whole input data series (Figure 4). Basic parameters that need to be set from the beginning are iteration number N (60,000), initial learning rate  $L_0$  (default 0.5). After Kohonen network has been built, Hierarchical algorithm will be applied on top of this map with a size of 30x30,

and then there will be just distance calculation on 900 elements instead of over 15,000 elements.

Each node on the network contains four different properties namely DT, GR, NPHI and RHOZ which are the characteristics of input data series. In the beginning period ( $t = 0$ ), Kohonen network contains 900 nodes having random values. After lots of iterations, node's value has been calibrated and nodes have been arranged more discipline. SOM algorithm has successfully reduced not only the number of dimensions but also the amount of input data; this will reduce the amount of work which Hierarchical algorithm has to deal with. Depending on purposes and experience of geological interpreter, the number of rock facies will be decided. In this case, authors choose to divide into six facies (Figure 5).

From the result in Figure 4 (below), at a specific depth with corresponding properties, we could determine whether that depth belongs to which specific node in Kohonen network. From the clustering result in Figure 5, we could obtain a random node and know which facies it belongs to. Thus, when combining two algorithms, we could know what kind of facies there are in a specific depth interval (Table 1). The result of this application will facilitate geological mapping which has critical effect on construction (determining compaction of a particular section), petroleum industry (determining depth that has potential of oil preserve) and environment (determining underground water field in order to avoid contaminating activities) and other fields.

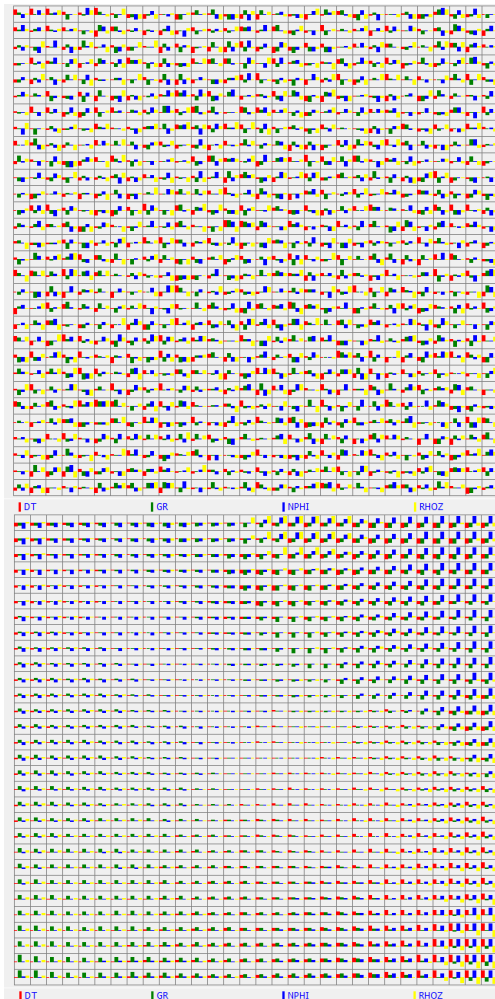


Figure 4. Self – organizing maps before (above) and after iterations

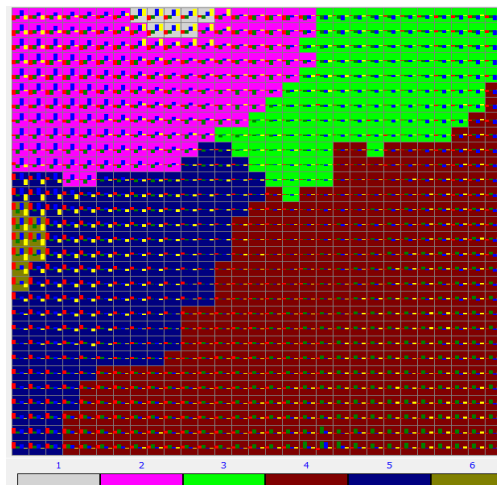


Figure 5. Six facies resulted from Hierarchical clustering algorithm

TABLE 1

AN EXTRACTION OF RESULTS OF FACIES WITH  
RESPECT TO DEPTH

Depth (ft)	Facies	Depth (ft)	Facies
2253.3865	2	2255.3677	5
2253.5388	2	2255.5200	5
2253.6912	2	2255.6724	5
2253.8435	2	2255.8247	5
2253.9961	1	2255.9773	5
2254.1484	1	2256.1296	5
2254.3008	1	2256.2820	5
2254.4531	1	2256.4343	6
2254.6057	1	2256.5869	6
2254.7581	1	2256.7393	6
2254.9104	1	2256.8916	6
2255.0627	1	2257.0439	6
2255.2151	2	2257.1963	6

### 3.4 SOM application in construction professional

Another application shown in here is in economic evaluation in construction, SOM will help clustering level change in cost of construction and building material, facilitating contractors as well as investors in choosing optimum way to invest their money on. Here, SOM algorithm helps analyzing data of construction price indices and core materials price indices.

Table 2 and Table 3 list the data pertaining to construction price indices and core materials price indices, respectively. Price index is a parameter describing how much the price of a specific period has changed comparing to a chosen standard period [7]. The chosen time here is year 2015.

TABLE 2

CORE MATERIALS PRICE INDICES IN THREE FIRST QUARTERS OF 2016 [7-10]

	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16
Cement	100	100	100	99	96.5	94.5	96.67	96.67	96.67
Construction sand	100	100	100	100.5	101.5	102	103.36	102.35	102.85
Construction rock	100	100	100	100.6	101	101	101.88	101.37	101.67
Construction brick	100	100	100	100	100	100	100	100	100
Construction wood	100	100	100	100	100	100	100	100	100
Structural steel	100	100	100	100	100	100	100	100	100
Asphaltene	100	100	100	76.13	72.9	74.84	77.59	78.71	78.71
Enameled tile	96	94	91	93.67	93.67	93.67	93.67	93.67	93.67
Roof materials	100	100	100	99.5	98.5	97	98.33	98.33	98.33
Construction glass and aluminum frame	100	100	100	100	100	100	100	100	100
Paint and architecture materials	100	100	100	100	100	100	100	100	100
VND/USD	100.16	100.16	100.16	99.8	100.13	99.98	99.98	99.98	100.23
Electrical materials	97	95	93.5	94.5	93.74	93.26	93.83	93.83	93.83
Water pipeline materials	100	100	100	100	100	100	100	100	100
Gasoline	91.53	83.28	80.65	85.78	89.87	94.11	91	83.94	92.44
Diesel 0.05S	81.31	72.14	72.36	75.08	82.39	89.39	91.93	85.68	91.69

TABLE 3  
CONSTRUCTION PRICE INDICES OF 2016 [7-10]

	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16
Less than 8F building	99.57	99.33	99.19	99.32	99.3	99.27	99.52	99.4	99.5
9F to 15F building	99.61	99.4	99.29	99.33	99.22	99.12	99.39	99.29	99.39
16F to 19F building	99.54	99.31	99.21	99.25	99.19	99.12	99.4	99.27	99.39
20F to 25F building	99.51	99.24	99.12	99.15	99.03	98.92	99.23	99.11	99.25
1F building with brick wall, aluminum roof	99.82	99.72	99.64	99.66	99.52	99.41	99.6	99.56	99.59
Closed 1F building, baring brick, reinforced concrete roof	99.68	99.52	99.36	99.46	99.33	99.22	99.45	99.19	99.43
2F and 3F building, reinforced concrete frame and roof	99.52	99.25	99.02	99.18	99.05	98.96	99.16	99.11	99.15
2F and 3F villas, reinforced concrete frame and roof	99.62	99.41	99.29	99.38	99.32	99.28	99.46	99.38	99.45
Educational infrastructure	99.37	99.04	98.98	99.46	99.49	99.45	99.63	99.53	99.65
Cultural infrastructure	99.42	99.11	98.93	99.05	99.05	99.02	99.24	99.12	99.25
Office	99.63	99.43	99.3	99.37	99.33	99.29	99.47	99.38	99.46
Medical infrastructure	99.78	99.63	99.54	99.46	99.64	99.59	99.7	99.62	99.8
Urban technical network infrastructure	99.56	99.32	99.16	96.71	96.41	96.62	97.22	97.17	97.3
Water resources infrastructure	99.28	98.92	98.92	98.97	99	99.03	99.56	99.23	99.51
Factory infrastructure	99.73	99.59	99.56	99.6	99.57	99.5	99.92	99.72	99.86
Concrete road	98.43	97.65	97.65	94.85	94.92	95.57	96.58	96.07	96.64
Bridges and roads infrastructure	98.28	97.44	97.45	97.57	98.15	98.72	99.14	98.52	99.1

3.5 Result

With the table containing data about construction price indices, because the size of input data is relatively small, authors will adjust parameters of SOM algorithm to fit the input data. Specifically, Kohonen network chosen has the size of 4x4 (16 nodes), iteration number N of 8,000 and initial learning rate  $L_0$  of 0.5. After running SOM, we obtain a map containing 16 nodes divided into 4 sections representing 17 objects.

TABLE 4  
CLUSTERING CONSTRUCTION PRICE INDICES  
RESULT OF FIRST THREE QUARTERS IN 2016

<p><b>GROUP 1</b> Bridges and roads Cultural infrastructure Water resources infrastructure</p>	<p><b>GROUP 2</b> Urban technical infrastructures Concrete roads</p>
<p><b>GROUP 3</b> Educational infrastructure Medical infrastructure Industrial building</p>	<p><b>GROUP 4</b> The others</p>

TABLE 5  
CLUSTERING CONSTRUCTION PRICE INDICES  
RESULT OF 2016

<p><b>GROUP 1</b> Bridges and roads Cultural infrastructure 9F to 15F building</p>	<p><b>GROUP 2</b> Urban technical infrastructures Concrete roads</p>
<p><b>GROUP 3</b> Educational infrastructure Medical infrastructure Factory infrastructure Water resources infrastructure</p>	<p><b>GROUP 4</b> The others</p>

From Table 4, it is obvious that there are four major groups and objects from the same group have the similarities in price indices. For instance, bridges and roads constructions, irrigation construction and cultural construction have similar tendency in price indices in the first three quarter of 2016 compared to 2015. SOM is undoubtedly useful tool for contractors because base on the map provided, they could choose projects that have similarities in cost change. Furthermore, assuming this trend will continue in quarter IV of 2016,

clustering will help contractors choose kinds of construction with price indices lower than 100%, facilitate them in optimizing economical problem. In the upcoming table, authors will conduct clustering on quarter IV of 2016 to check how efficient the forecast could be. Result of this procedure is presented in Table 5.

Table 5 shows the differences between forecasting and real data in quarter IV of 2016, there are two objects which have changed their positions in four sections namely water resources infrastructure and apartment building from 9 to 15 floors. The accuracy of the IV quarter price indices prediction procedure is up to 89.47%. With this result, forecasting price indices has proved to be reliable and could be used for further calculation in the future. In the similar vein, applying the two algorithms on the data relating to core material price indices has an outcome of a characterized table of core materials grouping together. Table 6 is the result obtained from the input data series.

TABLE 6  
CLUSTERING CORE MATERIAL PRICE INDICES IN THE FIRST THREE QUARTER IN 2016

<b>GROUP 1</b> Asphaltene	<b>GROUP 2</b> Enameled tile Electrical materials Cement Roof materials
<b>GROUP 3</b> Diesel 0.05S Gasoline	<b>GROUP 4</b> The others

The results presented in Table 6 provide a visual look into the change in core material price, facilitating the determination of total price falsity of construction. Specifically, we will conduct a forecast of quarter IV depending on the three first quarters. The result will be shown in Table 7.

TABLE 7  
CLUSTERING RESULT IN CORE MATERIAL PRICE INDICES IN QUARTER IV YEAR 2016

<b>GROUP 1</b> Asphaltene Enameled tile Electrical materials	<b>GROUP 2</b> Roof materials Gasoline
<b>GROUP 3</b> Diesel 0.05S Cement	<b>GROUP 4</b> The others

In Table 7 resulting from the analyze of core

material price indices in quarter IV, there are four object having the change in their positions namely enameled tile, electrical materials, cement and gasoline. The accuracy of the IV quarter price indices prediction procedure is up to 75%, less than in case of construction price indices but could still be reliable on.

#### 4 CONCLUSION

Since its inception, SOM has had applications in various majors, facilitating artificial intelligent to become key factor in the fourth technology revolution. To take the advantages of SOM, this study focuses on applying this technique to construction and geology engineering.

In geological aspect, SOM aids in clustering different facies based on logging data, allowing the form of geology maps for construction major (determine compaction of areas), petroleum major (determine which layer contains oil and gas) and environment major (determine which layer contains water in order to avoid contaminating that layer) and a majority of other professionals. Details level of the result depends upon the number of cells (nodes) in Kohonen network as well as iteration number N of algorithm. The higher this number can be, the more exact SOM algorithm can be, but the longer it takes for calculation. Thus, the problem relating to optimization the figure for nodes and iterations is the one that bears much consideration.

Turning to the issue regarding economical construction, SOM helps obtaining groups of core materials that have similar tendency in price changing. From that analyzer could predict the change in price of one object if the properties of others are known. This will facilitate contractors and investors in making different decisions in order to optimize their investment.

#### REFERENCES

- [1] Dr. Saed Sayad, "Self - Organizing Maps", University of Toronto, 2010.
- [2] Sasinee Pruekprasert, Thatchaphol Saranurak, Tarat Diloksawatdikul, "Self - organizing map (SOM)", Kasetsart University, 2009.
- [3] Pavel Stefanovic, Olga Kurasova, "Visual analysis of self - organizing maps", *Nonlinear Analysis: Modelling and Control*, Vol. 16, No. 4, 488-504, 2011.
- [4] L.V. Fausett, "Fundamentals of Neural Networks architectures, algorithms, and applications", Prentice Hall, 1994.
- [5] Orange Software tutorials on SOM.

- [6] "Publication of construction price indices month 01, 02, 03 and quarter I in 2016", Hanoi Construction Department, 2016.
- [7] "Publication of construction price indices month 04, 05, 06 and quarter II in 2016", Hanoi Construction Department, 2016.
- [8] "Publication of construction price indices month 07, 08, 09 and quarter III in 2016", Hanoi Construction Department, 2016.
- [9] "Publication of construction price indices month 10, 11, 12 and quarter IV in 2016", Hanoi Construction Department, 2016.

**Truong Minh Huy** is a student in Department of Drilling & Production Engineering, Faculty of Geology & Petroleum Engineering, Ho Chi Minh City University of Technology – VNU-HCM.

**Pham Ba Tuan** is a student in Department of Drilling & Production Engineering, Faculty of Geology & Petroleum Engineering, Ho Chi Minh City University of Technology – VNU-HCM.

**Pham Son Tung** received the Master in Civil Engineering in 2007 from UniverstéLibre de Bruxelles, Belgium, the Master in Business Administration in 2008 from UniversitéCatholique de Louvain, the Ph.D degree in Civil Engineering in 2014 from National Institut of Applied Sciences of Rennes, France. He is Lecturer in Department of Drilling & Production Engineering, Faculty of Geology & Petroleum Engineering, Ho Chi Minh City University of Technology – VNU-HCM.



# Ứng dụng thuật toán bản đồ tự tổ chức (self organizing map-SOM) trong các lĩnh vực xây dựng, địa chất và dầu khí

Phạm Sơn Tùng, Trương Minh Huy, Phạm Bá Tuân

*Tóm tắt*—Trong những năm gần đây, Trí tuệ nhân tạo đang ngày càng thịnh hành và từng bước khẳng định vị trí đầu tàu cho cuộc cách mạng công nghệ lần thứ tư. Trí tuệ nhân tạo đang lên ngôi và trở thành một phần không thể thiếu trong cuộc sống hằng ngày của con người. Trong đó, khái niệm Bản đồ tự tổ chức, một phân nhánh của lĩnh vực Trí tuệ nhân tạo, là một công cụ phân vùng dữ liệu hữu ích được ứng dụng rất rộng rãi và đã thành công trong các lĩnh vực xã hội như tâm lý, kinh doanh, y tế và ngay cả trong các lĩnh vực kỹ thuật như cơ khí, xây dựng và địa chất. Trong bài viết này, mục đích của tác giả nhằm giới thiệu thuật toán bản đồ tự tổ chức và các ứng dụng thực tiễn trong lĩnh vực địa chất và lĩnh vực xây dựng. Kết quả của nghiên cứu là phân loại tương đá theo độ sâu và phân cụm chỉ số giá xây dựng và chỉ số giá vật liệu xây dựng.

*Từ khóa*— Bản đồ tự tổ chức, phân vùng dữ liệu, địa chất, địa vật lý giếng khoan, kinh tế xây dựng.