A FACIES CLASSIFICATION APPROACH BASED ON CROSS RECURRENCE PLOTS

Hoa Dinh Nguyen
Học Viện Công Nghệ Bưu Chính Viễn Thông

Abstract— Facies classification for well log data is an important task that helps facilitate the estimation of some other properties such as permeability, porosity, and liquid content. This paper presents a new approach for facies classification based on cross recurrence plots from well log data. The proposed method is evaluated using real well log data collected in Cuu Long basin. The experimental results show that the approach is efficient for facies classification, especially when the data has a small number of well log curves. This is very meaningful in real world implementation where the collection of well log measurements is either difficult or expensive.

Keywords— facies classification, cross recurrence plots, well log curves.

I. INTRODUCTION

Facies are the overall characteristics of a rock unit that reflect its origin and differentiate the unit from others around it [5, 7]. According to the Dictionary of Geological Terms [11], facies are defined as "the aspect, appearance, and characteristics of a rock unit, usually reflecting the conditions of its origin; especially as differentiating it from adjacent or associated units". 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 petro-physical characteristics which control the fluid behavior in it [1]. 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 labeling of facies classes for well log data is an important and challenging task for oil and gas engineers. 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 wireline log measurements including statistical approaches, fuzzy methods, and artificial neural networks [2].

Author: Hoa Dinh Nguyen
Email: hoand@ptit.edu.vn
Received: 6/2019, revised: 7/2019, accepted: 8/2019.

In this study, a new facies classification approach based on cross recurrence plots (CRPs) [3] is investigated. CRPs, which are extension of recurrence plots (RPs) [9], are an efficient tool to visualize the relationship between two processes. They are built based on the construction of phase states from time series of different processes [3]. The proposed method is evaluated using real well log data collected from Cuu Long basin and the results show that the classification performance of this approach is very promising where the accuracy rate is almost 90%.

The structure of this paper is organized as follows. Section II provides all initial materials used in the research, including the description of the data as well as the background information in cross recurrence plots. The detailed method for facies classification based on CRPs is presented in Section III. Section IV includes all experimental results and discussion of the proposed approach. Section V concludes what have been accomplished in the research.

II. MATERIALS

Dataset

In this research, we investigate the possibility of detecting facies classes based on well log curve shapes. In other words, the relationship between the well log curve shapes with all facies classes is utilized for facies classification using CRPs. In general, well log data contains a lot of measurement curves. However, there is a limited number of log curves that have relationship with facies classes. Some well know published datasets, some commonly used log curves for facies classification problems include gamma ray, resistivity logging, photoelectric effect, neutron-density porosity difference and average neutron-density porosity [7]. Indeed, each published dataset may have different number of log curves available for the facies classification task. It is expected that each facies class creates its own trends in well log features, which are different from one class to the other classes. It also well stated in the literature that there is some correlation between facies classes and well log shapes [4]. Duboisa et al. [4] show that log curve shapes can be utilized as predictive tools for facies interpretation. Nazeer et al. [6] present five common shapes of gamma ray (GR) corresponding to different facies classes, which are cylindrical shape, funnel shape, bell shape, bow shape, and irregular shape. Based on their research, the first four types of curve shapes are useful for facies class identification. However, the fifth type of curve shape is unpredictable and can worsen the facies classification results. Besides, each facies classes can also cause different trends in many other log features, resulting in a lot of inconsistent trends of log curves caused by one particular facies class.

In some cases, using known data trends of one particular log curve may help identify some facies classes efficiently. However, in most cases, the combined trending information from different log curves is needed to completely display facies classification results for well log data. Finding an efficient tool for visualizing the data trending of well logs is our goal in this paper. This tool must be able to present the characteristics of natural geologic data, which is believed to be nonlinear and non-stationary.

Among many log curves available in well log data, there are empirically only 6 most significant curves useful for facies classification, which are compressional wave delay time (DTCO), gamma ray (GR), neutron porosity (NPHI), effective porosity (PHIE), bulk density (RHOB), and volume of clay (VCL). These six log curves are used in this research. 12 well log datasets collected from Cuu Long basin are used. All the data samples are classified by experts and divided into two subsets, each of which consists of 6 well logs. One subset is for training process and the other is for testing.

Following section will present technical tools to capture the trending behaviors of well log data useful for facies classification.

Cross Recurrence Plots

Recurrence is an important characteristic of a dynamic system, according to which the system tends to return to its current working state at some points in the future [8]. Recurrence plots are a tool proposed by Eckmann et al. [9] that help visualize the trends of time series from complex dynamic systems. Assuming that a working system is observed using a time series $x(t)$. The phase state of the system at the time t_i is defined as [10]:

$$\vec{u_i} = [x(t_i), x(t_i+\tau), \dots, x(t_i+(m-1)\tau)] \quad (1)$$

where τ is the delay and m is the dimension of the embedding phase space. Typically, τ is chosen such that all components in one phase state are not correlated, while m depends on the number of factors that directly influent the system states. A recurrence plot of $x(t)$ is an $N \times N$ matrix, each element of which is calculated by the following equation.

$$R_{i,j} = \ominus\left(\varepsilon_i - \|\vec{u_i} - \vec{u_j}\|\right),\ i,j = 1,2,\dots,N \quad (2)$$

Where $\ominus(\cdot)$ is the unit step function, ε_i is the cut-off distance, and $\|\cdot\|$ is the Euclidean norm. According to this, if a state vector $\vec{u_i}$ is within the range of ε_i from vector $\vec{u_j}$, then $R_{i,j} = 1$, otherwise, $R_{i,j} = 0$. The values one or zero in the matrix can be represented by colors black and white. The distance ε_i can be either a predefined value or iteratively chosen such that there is a fixed number of neighbors at every state $\vec{u_i}$. The predefined value of ε_i is greatly based on the characteristics of the time series $x(t)$ as well as the applications of its recurrence plots. Recurrence plot is a powerful tool to visualize the recurrence behavior of nonlinear and dynamic systems. It is noted that single recurrence point at $(t_i,\ t_j)$ does not contain much information about the current states at the time t_i and t_j. In general, the totality of recurrence points can be used to reconstruct the properties of the data [9].

Cross recurrence plots (CRPs) [3] is an extension of recurrence plots, which enables visualizing the dependent behavior of two processes using time series. CRPs is based on the comparison of the two trajectories in the same phase space of the two processes. It can be utilized to study the similarity between two different phase state trajectories. CRPs of the two time series $x(t)$ and $y(t)$ is an N-by-M matrix, of which each element is computed by the equation:

$$CR_{i,j} = \ominus\left(\varepsilon_i - \|\vec{u_i} - \vec{v_j}\|\right),$$
$$i = 1,\dots,N; j = 1,\dots,M \quad (3)$$

Where $\vec{u_i} = [x(t_i), x(t_i+\tau), \dots, x(t_i+(m-1)\tau)]$ and $\vec{v_j} = [y(t_j), y(t_j+\tau), \dots, y(t_j+(m-1)\tau)]$. Other notations are the same as in the definition of recurrence plot presented above. If the state at time t_j of the second process is close to the state at time t_i of the first process, then $CR_{i,j} = 1$, which is presented by a black dot, otherwise, $CR_{i,j} = 0$, which is presented by a white dot. In fact, this does not represent the recurrences of any state but the conjunctures of states of the two processes. In other words, the CRPs reveals all the time points when the phase space trajectory of the first system visits roughly the same area in the phase space as the second system is at a given point of time. The data length of both processes can be different resulting in a non-square CRP matrix.

The following session presents the detailed facies classification method based on CRPs.

III. METHODOLOGY

The general facies classification approach based on cross recurrence plots from log curves is depicted in figure 1. Training data is divided into smaller data groups. Each facies data group, which only contain the data sequences of one facies class, will be the input for the detection algorithm based on CRP to detect the appearance of that particular facies. The data samples are stored in nature sequences collected from wells. Fusion stage combines all individual labeling results from all facies class detection algorithm to provide the final sequence of facies classes for all test well log samples. For simplicity, majority voting is used in fusion stage. This research focuses on the first part of the whole facies classification system, which is to detect one particular facies class from one data group using CRPs. Figure 2 illustrates the name of all facies classes as well as their color codes. In general, there are 11 facies classes for well log data. However, in most cases, not all 11 classes present in one well log.

CRPs between testing well logs and training data of each facies class are calculated using the fixed cutoff thresholding method, i.e. ε is fixed. CRPs help visualize the closeness between each phase state in testing data and all available states in the training data corresponding to each facies class. If one phase state of testing data is close to a phase state of one particular facies class of training data, the data sequence constructing that state will be considered to be in that class. In this research, we only investigate the application of CRPs on labeling one individual class to the testing data. A binary facies classification algorithm is proposed based on CRPs to determine which portion of the testing data belonging to one specified facies class. Our future work will be proposing a data fusion mechanism to combine the results from all individual binary labeling process for the complete presentation of all facies classes for testing data.

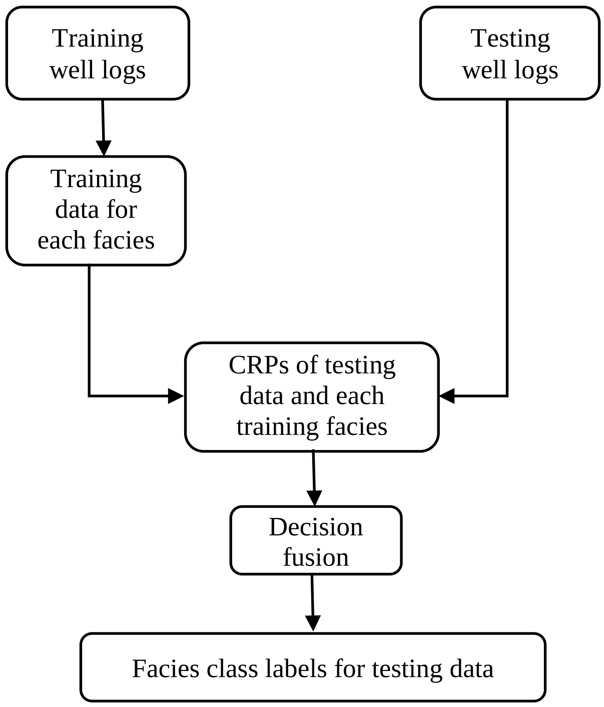
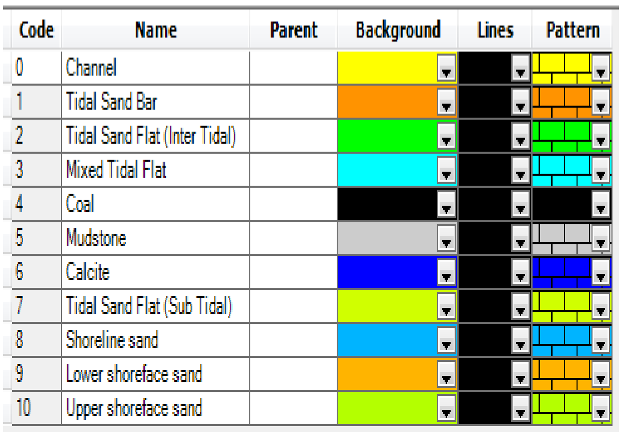


Figure 1: Flow chart of facies classification approach based on cross recurrence plots.

The motivation of the proposed algorithm is from the nature properties of geological facies. Each facies, for examples depositional facies, is created by one depositional process, which may take about thousands of years. Different facies contain different structure of rocks, sands, and soils. This results in different representation of well log curves, which may be recognized by CRPs properties. In other words, CRPs help differentiate the phase states of the well log data from different facies.



Code	Name	Parent	Background	Lines	Pattern
0	Channel				
1	Tidal Sand Bar				
2	Tidal Sand Flat (Inter Tidal)				
3	Mixed Tidal Flat				
4	Coal				
5	Mudstone				
6	Calcite				
7	Tidal Sand Flat (Sub Tidal)				
8	Shoreline sand				
9	Lower shoreface sand				
10	Upper shoreface sand				

Figure 2: Facies class names and their color codes.

Algorithm: label one facies class to testing data using CRPs

Input: training well log data of one facies class, testing data including well log cures of some unlabeled well points; select distinctive curves: GR, VCL, PHIE.

Step 1: Construct CRPs of testing well log curves and training well log curves. All parameters for CRP construction are determined empirically.

Step 2: Construct a histogram curve presenting number of training phase states neighboring to each testing phase state.

Step 3: If total number of neighboring states for one testing state exceeds a predefined threshold λ, the whole data sequence constructing that testing phase state will be labeled as targeted facies class.

Output: testing data labeled with investigating facies class

In fact, well log data may form different types of phase states. Those single states may arbitrary be almost the same between different facies classes. However, due to the geological properties of each particular facies class, one phase state of the well log data belonging to one facies class will be similar to a bigger number of other phase states within its facies class compared to the phase states of the other classes. Step 3 of the algorithm aims at discarding situations where one phase state is similar to a small number of random phase states of different classes.

This proposed method requires that training data must contain sequences of data with the length of greater or equal to the data length of a phase state. This is to ensure there are enough phase state data in the training set. In other words, the proposed method is expected not to work well with facies classes having too small training datasets, or the training data are so scattering that not enough phase states can be formed.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed approach, several scenarios of the experiments have been conducted. Data from first six wells are used for training, while data from remaining six wells are used for testing. Based on the nature of the data, class 5 has the biggest amount of training data. Only facies class 5 is concerned in the experiments. In other words, samples of class 5 are labeled as "1", while data of all other facies classes are labeled as "0". Since all six log curves are measured in different units and scales, they are normalized to the range of [0, 1] before any further processes. Based on different empirical trials, it is noted that three log curves (GR, PHIE, VCL) have the highest relationship with the facies labels. In this work, these three curves are investigated more often.

Several experimental scenarios have been conducted. First, each curve of GR, PHIE, or VCL is input to the facies classification algorithm. Next, each of the combination of two curves of GR, PHIE, and VCL is input to the algorithm. Then, all three curves are input to the algorithm. Finally, all six curves are input to the algorithm. All parameters of CRPs constructed in each are empirically selected with the highest correction scores. Figure 3 and 4 illustrate examples of CRPs constructed from different sets of parameters.

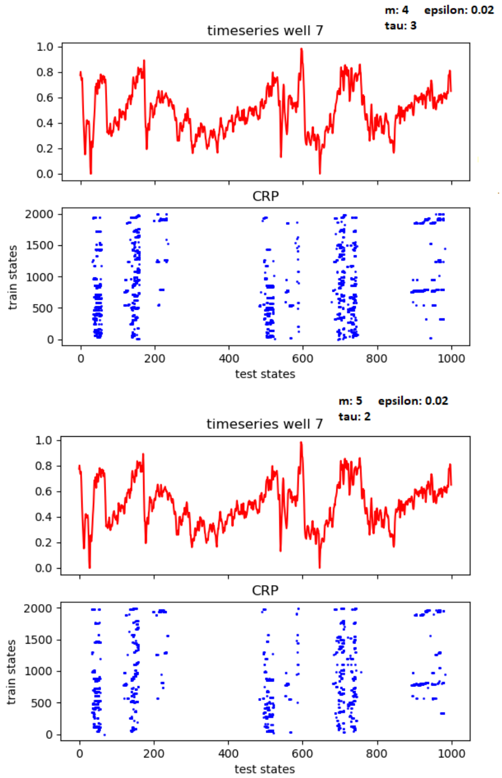


Figure 3: CRPs with different sets of parameters constructed for testing well 7

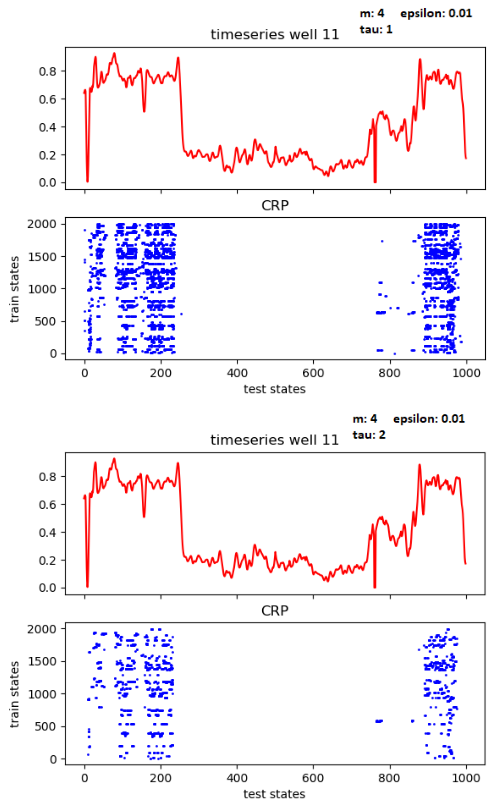
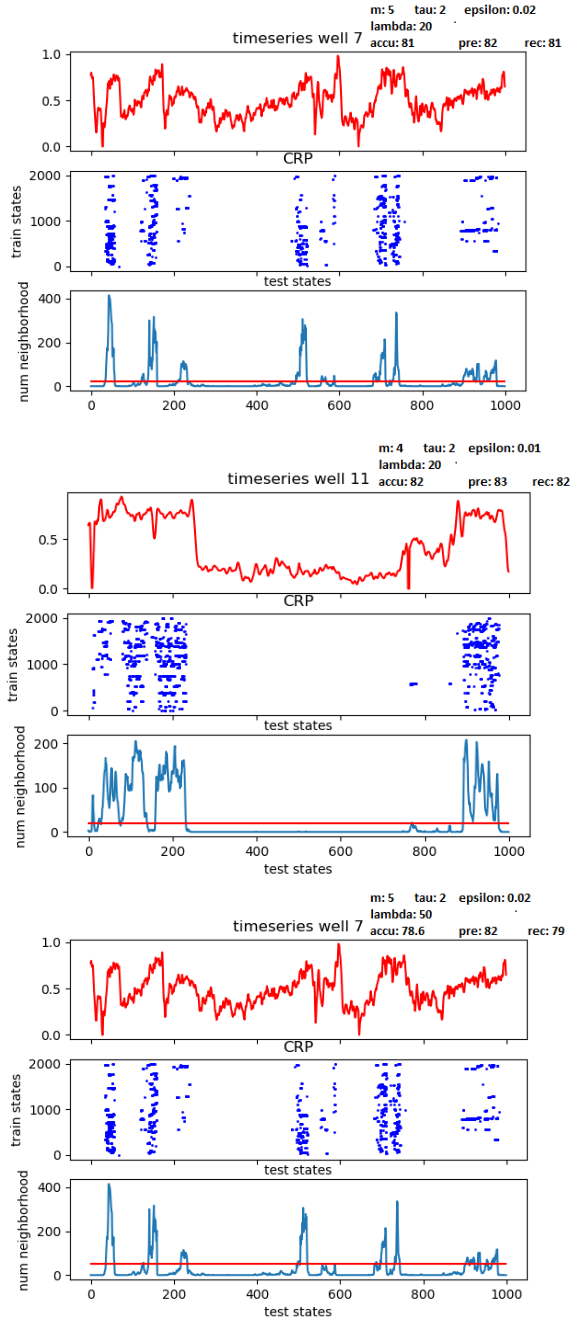


Figure 4: CRPs with different sets of parameters constructed for testing well 11

In order to identify which testing state is close to training state of the concerned facies class, a threshold λ is set on the histogram to avoid any confusion caused by random states that are similar to the investigated testing state. Different values of λ lead to different performance quality of the system. Figure 5 presents some examples of different λ values with their respective classification performances. After investigating different combinations of CRPs parameter sets and λ values, the set of $m = 4$, $\tau = 1$, $\varepsilon = 0.0001$, $\lambda = 50$ is selected for all scenarios since it can provide acceptable performance scores. The experimental results of all scenarios are summarized in Table 1. Classification performance is evaluated based on three indices: precision, recall, and accuracy.



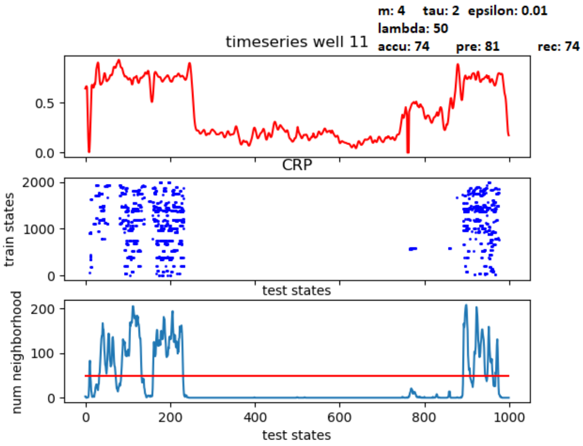


Figure 5: different values of λ correspond to different classification performances

Table 1: Performance information of the methods using different number of log curves

Scenario	Precision	Recall	Accuracy
Input: curve GR	0.82	0.89	0.847
Input: curve VCL	0.82	0.97	0.877
Input: curve PHIE	0.84	0.95	0.88
Input: curves GR and VCL	0.88	0.81	0.846
Input: curves GR and PHIE	0.88	0.84	0.858
Input: curves VCL and PHIE	0.82	0.97	0.873
Input: curve GR, VCL, and PHIE	0.89	0.82	0.855
Input: all 6 curves	0.86	0.94	0.89

Experimental results show that classification performance of the proposed method based on CRPs is very promising. There are some slight changes in the accuracy between different scenarios. The most important thing from those results is that there is not different in classification performance between using one curve and using many available curves. In other words, the proposed approach is very useful when the number of available log curves is limited. This is very meaningful for oil & gas industry, where measuring geophysical information at some well logs are very expensive or difficult.

As discussed in the previous session, the proposed approach is expected to work well with facies classes having long enough sequences of training and testing data, where phase states can be formed properly. For some facies classes with small number of log samples, especially with too short sequences of log data, this classification algorithm cannot work. In this case, the proposed method can be combined with some other machine learning techniques to fully classify all remaining facies classes. The main advantage of the classification method based on CRPs is the ease of implementation, and it requires only small number of well log measurements.

V. CONCLUSIONS

In this research, a new facies classification algorithm based on CRPs is introduced. CRPs are an efficient tool to visualize the relationship between two processes presented in time series. This helps recognize the data patterns of different facies classes, which facilitate the facies classification process based on pattern detection. Experimental results show that the proposed approach can work well with facies classes having long data sequences. The new method can be combined with traditional machine learning tools to efficiently provide the complete facies classification picture for well log data

REFERENCES

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

[2] 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.

[3] N. Marwan, J. Kurths. "Cross Recurrence Plots and Their Applications", Mathematical Physics Research at the Cutting Edge, pp. 101-139, 2004.

[4] M.K. Duboisa, G.C. Bohlinga, S. Chakrabarti. "Comparison of four approaches to a rock facies classification problem", Computers & Geosciences, 33, pp.599–617, 2007.

[5] T. Crampin, "Well log facies classification for improved regional exploration", Exploration Geophysics, 39:2, 115-123, 2008.

[6] A. Nazeera, S.A. Abbasib, S.H. Solangi. "Sedimentary facies interpretation of Gamma Ray (GR) log as basic well logs in Central and Lower Indus Basin of Pakistan", Geodesy and Geodynamics, 7(6), pp.432-443, 2016.

[7] V. Tschannen, M. Delescluse, M. Rodriguez, J. Keuper, "Facies classification from well logs using an inception convolutional network", Computer Vision and Pattern Recognition, 2017.

[8] H. Poincare, "Sur la probleme des trois corps et les quations de la dynamique", Acta Mathematica, 13, pp. 1-271, 1890.

[9] J.P. Eckmann, S.O. Kamphorst, D. Ruelle, "Recurrence Plots of Dynamical Systems", Europhysics Letters, 5, pp. 973–977, 1987.

[10] F. Takens, D.A. Rand, L.S. Young, "Detecting strange attractors in turbulence", Dynamical Systems and Turbulence, 898, pp. 366–381, 1981.

[11] Bates, R. L., and Jackson, J. A., 1984, Dictionary of geological terms: American Geological Institute.

MỘT PHƯƠNG PHÁP PHÂN LOẠI TƯỚNG ĐỊA CHẤT DỰA TRÊN ẢNH HỒI QUY CHÉO

Tóm tắt: Phân loại tướng địa chất cho dữ liệu giếng khoan là một nhiệm vụ quan trọng trong việc thúc đẩy khả năng đánh giá một số tính chất địa chất khác như độ thấm, độ xốp và hàm lượng chất lỏng. Bài báo này trình bày một cách tiếp cận mới trong việc phân loại phân loại tướng địa chất dựa trên các ảnh hồi quy chéo chéo từ các dữ liệu địa chất đã được minh giải rõ ràng với các dữ liệu mới lấy lên từ giếng khoan. Phương pháp đề xuất được đánh giá bằng cách sử dụng dữ liệu giếng khoan thực được thu thập trong lưu vực Cửu Long. Các kết quả thử nghiệm cho thấy phương pháp này có hiệu quả đối với việc phân loại tướng địa chất, đặc biệt là khi dữ liệu chỉ có một số ít đường thông tin giếng khoan. Điều này rất có ý nghĩa trên thực tiễn vì việc thu thập các thông số đo tại giếng khoan là rất khó khăn và tốn kém.

Từ khoá: phân loại tướng địa chất, ảnh hồi quy chéo, đường đo giếng khoan.

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 mining and machine learning.