

PRE-STACK TEXTURE-BASED SEMI-SUPERVISED SEISMIC FACIES ANALYSIS USING GLOBAL OPTIMIZATION

HANPENG CAI, QINGPING WU, HAIYANG REN, HUIQIANG LI and QING QIN

*School of Resources and Environment & Center of information geoscience, University of Electronic Science and Technology of China, Sichuan 611731, P.R. China.
wqpkita@foxmail.com*

(Received July 16, 2018; revised version accepted September 10, 2019)

ABSTRACT

Cai, H.P., Wu, Q.P., Ren, H.Y., Li, H.Q. and Qin, Q., 2019. Pre-stack texture-based semi-supervised seismic facies analysis using global optimization. *Journal of Seismic Exploration*, 28: 513-532.

There are some problems in conventional seismic facies analysis methods, such as easily plunge into local optimal solution, low sensitivity and without using prior knowledge. To solve the above-mentioned problems, we propose a pre-stack texture-based semi-supervised seismic facies analysis method with global optimization. Firstly, the pre-stack seismic texture attributes are introduced to highlighting the information of micro-spatial and amplitude variation with azimuth/offset in seismic reflection data. Then, the self-organizing map (SOM) neural network is used to compress a large amount of redundant information of the samples on the premise of maintaining the topology of the data. Finally, the artificial bee colony (ABC) algorithm is used to realize the global optimization of the clustering of neurons in the SOM output layer under the constraints of prior knowledge. Besides, according to the probability estimation results based on the probabilistic neural network (PNN), we define the confidence measures to quantitative analysis the classification results. The synthetic test and practical application results show that the proposed method can not only significantly improve the recognition ability of the seismic microfacies, but also improve the horizontal resolution and the accuracy of the seismic facies map. These satisfactory results illustrate the proposed method is an effective tool for seismic facies analysis.

KEY WORDS: seismic facies analysis, pre-stack seismic texture attributes, semi-supervised learning, self-organizing map (SOM), artificial bee colony (ABC), probabilistic neural network (PNN).

INTRODUCTION

Seismic facies analysis is based on the division of seismic sequences, the seismic waveforms in a particular sequence are divided into different seismic facies regions by using the parameters difference, and then the inferences of lithofacies and sedimentary environment can be made. Seismic facies analysis has become an indispensable method in the field of describing and detecting seismic reservoir characteristics (Nivlet, 2007). In recent years, referring to the principle of pattern recognition and basing on seismic-data-driven, seismic attributes and other auxiliary information to describe the spatial distribution of geological bodies can be obtained only from the massive seismic data, and there are gradually formed some unsupervised seismic facies analysis methods whose theory and method are more consummate (de Matos et al., 2007). Zhang et al. (2015) combine the SOM with particle swarm optimization (PSO) algorithm to implement an unsupervised seismic facies analysis, and improve the calculation efficiency and the accurate of the facies map. Roy et al. (2014) and Zhao et al. (2015) using the generative topographic mapping (GTM) classification algorithm to obtain a probability estimates of whether a given voxel is falls into a specific category, and implement the seismic facies estimation. Song et al. (2017) adapt the regularized fuzzy c-means (RegFCM) algorithm to implement the seismic facies analysis with spatial constraints, and improve the spatial continuity of the facies map. Song et al. (2018) introduce an adaptive phase K-means algorithm to alleviate the horizon error in waveform classification.

K-means clustering method is an important method for unsupervised seismic facies analysis. It divides N objects into K clusters by making higher similarity in each cluster and lower similarity between the different clusters, the measurement of its similarity can be calculated by the Euclidean distance from each object to their cluster center. Clustering analysis of seismic data by K-means requires a predefined number of clusters, in addition, it is also vulnerable to noise and outliers, and can not keep the topology of input data. Kohonen (1982) proposed the SOM neural network, which is an unsupervised pattern recognition method. By maintaining the topology of the original data through the output layer neurons, SOM has been widely used in seismic facies analysis (Marroquín, 2009). However, the mapping between output neurons and facies categories may lead to fuzziness or even confusion of facies boundaries when SOM is used to process strong redundant seismic data. In addition, this method does not use the valuable prior knowledge such as drilling, logging and geology.

Many existing seismic facies analysis methods are based on raw seismic waveforms. However, due to the effect of earth filtering, the frequency band of seismic data shows low frequency and narrowband characteristics, which makes it difficult for original seismic waveforms to describe the subtle changes in seismic reflection information. In the past decades, significant progress has been made in various seismic attributes calculation methods applied to underground exploration (Chopra and Marfurt, 2005). Many geophysicists combine time frequency analysis,

instantaneous attributes, elastic parameter inversion and texture analysis with SOM to identify seismic facies automatically (de Matos et al., 2007). Faced with numerous seismic attributes, Qi et al. (2016) introduce a degree for each attribute by computing the correlation coefficient between histogram for the attribute and histograms for a suit of user-defined facies, and make the choice of attributes, then they combine the GTM and Bhattacharyya distance to obtain the probability that a given voxel belongs to a specific facies. Gao (2006, 2011) defined the seismic texture as the spatial arrangement of the adjacent seismic amplitude, and pointed out that compared with the average amplitude in identifying the seismic characteristics, the seismic texture analysis can effectively emphasize the spatial variability of the seismic reflection amplitude. In recent years, seismic texture attributes have been applied to structure interpretation (Gao, 2006), detecting salt bodies (Tamir and Ghassan, 2014; Shafiq et al., 2015; Guillen, 2015), reservoir prediction and characterization (Chopra and Alexeev, 2006; Yenugu, 2010), sedimentary environment analysis (Ruffo, 2007; Angelo, 2009). Chopra (2006), Gao (2011), Chuai (2014), have successfully applied the seismic texture attributes to implement the seismic facies analysis, but all of them are based on post-stack seismic data, Marfurt (2014) pointed out that the development of pre-stack attributes is an important part of the future development of seismic attributes. Song et al. (2016) proposed a method of seismic facies analysis based on pre-stack seismic texture attributes and preliminarily used it to identify the development zones of faults and karst caves.

In many research fields, unlabeled data is relatively easy to obtain, but the acquisition of labeled data requires massive resources, in order to make full use of the valuable label information, semi-supervised learning methods have been widely used in these fields such as network intrusion detection (Erman et al., 2007), image classification (Hong and Zhu, 2015; Dornaika and Traboulsi, 2017), brain image segmentation (Portela et al., 2014; Saha et al., 2016), text classification (Zhang et al., 2015; Altinel and Ganiz, 2016), seismic facies analysis (Qi et al., 2016), and porosity estimation (Lima et al., 2017). This paper propose a pre-stack texture-based semi-supervised seismic facies analysis method with global optimization, this method can not only extracts more information of micro-spatial and amplitude variation with azimuth/offset from pre-stack seismic data, but also effectively utilizes the small amount of prior label information. This proposed method optimizes the mapping relationship between the SOM output layer and the seismic facies, and alleviates the confusion of facies boundary in the seismic facies map.

PRINCIPLE AND METHODS

In order to obtain a more accurate seismic facies map, this paper proposed a semi-supervised artificial bee colony based self-organizing map (SSABC-SOM) clustering method. Qi et al. (2016) use GTM to map the attributes onto a latent space, and by calculating the Bhattacharyya distance between the probability density function (PDF) of the testing voxel and the

PDFs of the supervised voxels on the latent space, the probability that this given voxel belongs to a specific facies is obtained. The proposed SSABC-SOM method uses the ABC algorithm and the existing labels information to optimize the clustering of association neurons in SOM output layer, and finally obtains the optimal mapping from the output layer neurons to the seismic facies. Meanwhile, we define the confidence measures to quantitative analysis the result. The workflow of our seismic facies analysis method is shown in Fig. 1, and the specific processes of the method are as follows:

1) Extract the texture attributes of each reflection bin from the pre-stack seismic data.

2) Train the SOM network based on the training set which contains the labeled and unlabeled samples.

3) Label the response neurons in the SOM output layer of the labeled samples and clarify their Must-link and Cannot-link relationships.

4) Make an optimizing K-means clustering of SOM output layer neurons by using the ABC algorithm and the clarified connection relationships that obtained from step 3.

5) Repeat step 4 until the stop condition of ABC is satisfied, and the optimal mapping relationship between the SOM output neurons and the seismic facies is obtained.

6) Input the samples into the SOM network, recording their BMUs, and their corresponding seismic facies are obtained according to the optimized mapping relationship.

7) Evaluate the results by calculating the confidence indicators defined in this paper.

Pre-stack texture attributes

Compared with the characteristics of a single seismic waveform, texture features consider not only the response waveform of a reflection itself, but also the spatial information of multiple directions, which is an integration of seismic waveforms, spatial information and reservoir characteristics. Under the assumption that the spatial distribution of each pixel in the image contains the image texture information, Haralick et al. (1973) proposed a grayscale co-occurrence matrix (GLCM) statistical method and defined 14 texture attributes, in research of Ulaby, et al. (1986) they pointed out that in 14 GLCM texture attributes, only energy, contrast, homogeneity and correlation are irrelevant. Fig. 2 shows the process of building the GLCMs of a 2D grayscale image. The pre-stack texture elements and GLCMs statistic directions of a reflection bin are shown in Fig. 3 and Table 1, respectively. N_x , N_y , N_z shown in Fig. 3 are the length of each gather volume in the inline, crossline and time directions. M represents the number of traces from one reflection point. Typically, N_z is related to the sampling rate of the waveform, N_x and N_y are range from 3 to 9 (Gao, 2007).

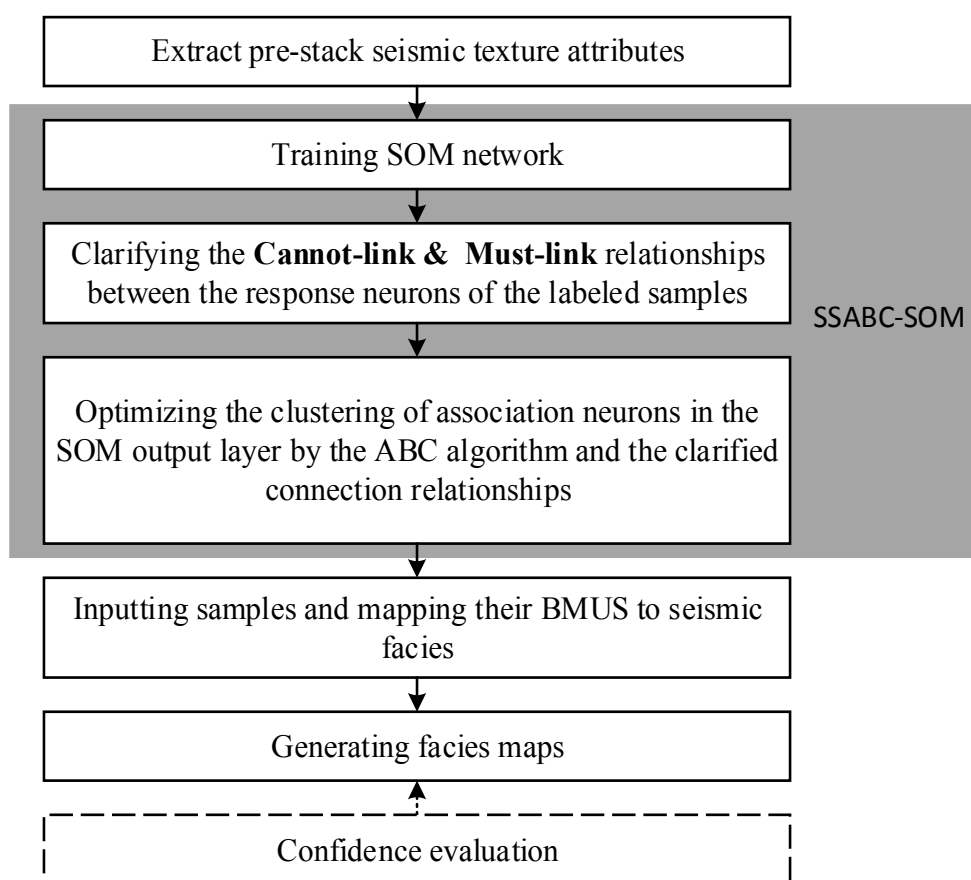


Fig. 1. Workflow for seismic facies analysis.

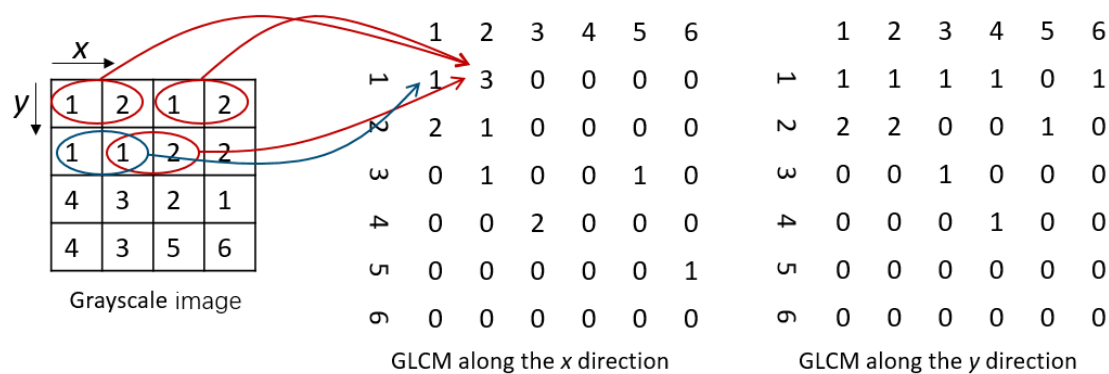


Fig. 2. The process of building GLCMs.

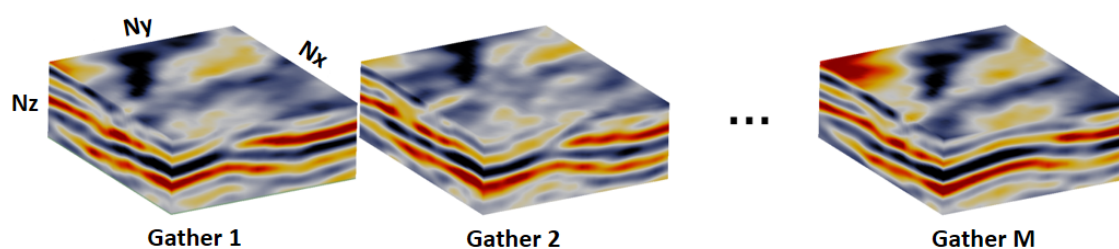


Fig. 3. Pre-stack texture elements of a reflection bin.

Table 1. GLCMs statistic directions of each reflection bin.

Directions	Orientation			
	Inline	Crossline	Time	Gather
Direction 1	1	0	0	0
Direction 2	0	1	0	0
Direction 3	0	0	1	0
Direction 4	1	1	0	0
Direction 5	-1	1	0	0
Direction 6	0	0	0	1

After extracting the pre-stack texture elements of a reflection bin, we normalize the seismic data according to formulas (de Matos et al., 2011)

$$D_{(x,y,z,m)} = \text{Round}(0.5 * (Ng - 1) * (\frac{K_{(x,y,z,m)}}{K_{\max}} + 1) + 1) , \quad (1)$$

$$K_{\max} = \max_{x,y,z,m} K_{(x,y,z,m)}$$

$$0 \leq x \leq Nx, 0 \leq y \leq Ny, 0 \leq z \leq Nz, 0 \leq m \leq M \quad (2)$$

where $D(x,y,z,m)$ represents the normalized element value, Ng is the matrix size of GLCM, $K(x,y,z,m)$ is the amplitude value of the sample located in (x,y,z,m) in the pre-stack texture elements and $\text{Round}()$ is the integer function. In a direction \vec{w} from Table 1, the GLCM element $E(i,j)$ and its corresponding probability value p_{ij} of this direction can be calculated by formulas

$$E(i, j) = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \sum_{z=1}^{N_z} \sum_{m=1}^M \{D_{(x,y,z,m)} = i, D_{(x,y,z,m)+\vec{w}} = j\} , \quad (3)$$

$$p_{ij} = E(i, j) * (1 / \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} E(i, j)) \quad (4)$$

After building the GLCMs, four irrelevant texture attributes of energy, contrast, homogeneity and correlation are extracted from each GLCM by formulas

$$\text{Energy} = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p_{ij}^2 , \quad (5)$$

$$Contrast = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p_{ij} * (i - j)^2 \quad , \quad (6)$$

$$Homogeneity = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p_{ij} * \frac{1}{1 + (i - j)^2} \quad , \quad (7)$$

$$Correlation = \frac{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i * j) * p_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y} \quad , \quad (8)$$

where μ_x , μ_y , σ_x and σ_y are the means and standard deviations of $p_x(i) = \sum_{j=1}^{Ng} p_{ij}$ and $p_y(j) = \sum_{i=1}^{Ng} p_{ij}$.

Self-organizing map

SOM is a kind of unsupervised competitive learning method, which is composed of 2-layer neurons. The network architecture is shown in Fig. 4, the first layer is the input layer whose dimensionality is consistent with the input sample vector, and the second layer is the output layer, in which an output layer node represents a neuron and these neurons are arranged in rectangular or hexagonal rules. A full connection is taken between the input and output layer while the neurons in the output layer are lateral inhibitory connections. Suppose the training set $\{X_1, \dots, X_N\}$ containing N vectors and each vector represents the texture attributes extracted from pre-stack seismic data. In each training cycle $i (i \in [1, N])$, the best math unit (BMU) whose weight vector is most similar to the sample can be found by identify the shortest distance between the sample X_i and each neurons in the output layer. Once the BMU is determined, the weight vectors of the BMU and the neurons close to it can be updated by formula

$$W_j(t+1) = W_j(t) + (\eta(t) * h(dist(BMU, v_j), t) * (X_i - W_j(t))) \quad , \quad (9)$$

where W_j represents the weight vector of BMU and its neighborhood neurons, $\eta(t)$ is a learning rate that decreases with time, $h(\)$ is the neighborhood function whose independent variable $dist(BMU, v_j)$ is the distance between the neuron v_j and the BMU, generally, $h(\)$ set as a Gaussian function and is shrinks with time. After training, the adjacent output layer nodes can be considered to correlate the similar samples in the input data.

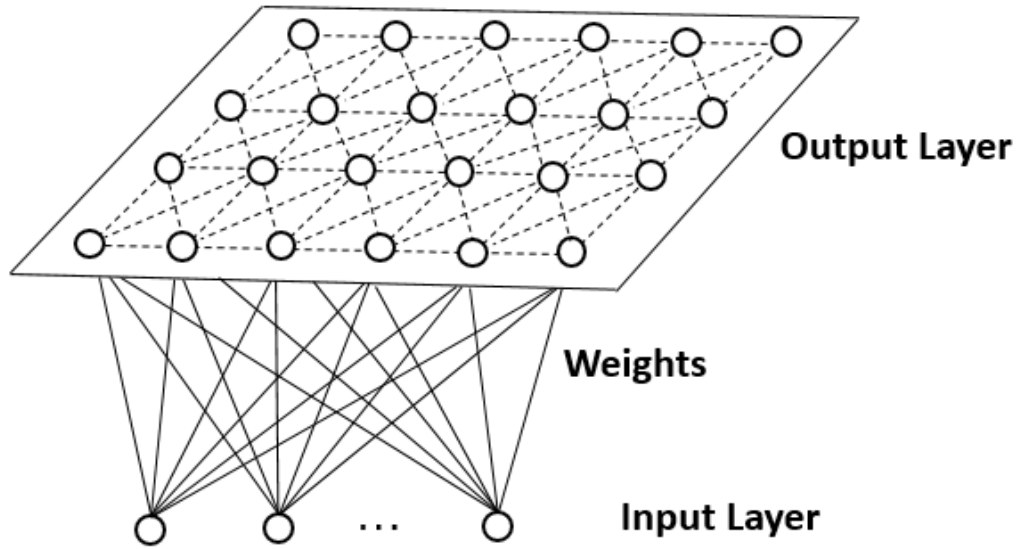


Fig. 4. Network architecture of SOM.

Artificial bee colony algorithm

ABC is a swarm intelligent optimization algorithm proposed by Karaboga(2005) by imitating the intelligent foraging behavior of bees. This algorithm can be divided into 4 parts: initiating food sources, employed bees search food, onlooker bees follow and scout bees replace food sources. A food source indicates a potentially feasible solution to the problem, and its initialization method is

$$x_{mj} = l_j + r * (u_j - l_j) , \quad (10)$$

where x_{mj} represents the j -th dimension element of the m -th food source, note that m and j are the numbers selected from set $\{1, 2, \dots, Ne\}$ and set $\{1, 2, \dots, D\}$, respectively. Ne is the size of the employed bees and D is the dimension size of the solution, r is a random number within $[-1, 1]$, and l_j, u_j represent the lower and upper bound of the j -th dimension of the solution, respectively. After initializing Ne food sources, employed bees exploit these food sources and new sources around them. The new food sources are produced in the following ways

$$v_{mj} = x_{mj} + r * (x_{kj} - x_{mj}) , \quad (11)$$

where v_{mj} represents the j -th dimension element of a new food source around the food source x_m , x_{kj} is the j -th dimension element of an existing source x_k , k is a random number within $[1, Ne]$. Once the new food source v_{mj} is generated, the corresponding employed bee makes a greedy selection between x_m and v_m based on the fitness value of these two food sources. The fitness value of food source x_m is calculated by formula

$$fitness_{x_m} = \begin{cases} 1 / (1 + f(x_m)), & \text{if } f(x_m) \geq 0 \\ 1 + abs(f(x_m)), & \text{if } f(x_m) < 0 \end{cases} \quad (12)$$

where $f(x_m)$ is the objective function value of the food source x_m . After all employed bees completing their search, onlooker bees selectively exploit foods around the exploited food sources base on the information returned by employed bees. The probability of the m -th food source be chosen by an onlooker bee is

$$p_m = \frac{fitness_m}{\sum_{i=1}^{Ne} fitness_i} \quad (13)$$

Every onlooker bee first selects the food source to follow according to the roulette wheel mechanism, and uses the formula (11) to find a new food source around the chosen one. Then the onlooker bee makes a greedy selection between the new food source and the chosen one base on the formula (12). Assuming that the maximum allowable number of foraging is *limit*, if a food source has not been replaced after *limit* times of foraging, it means that this food source may trap in local optimum, the corresponding employed bee will give up this food source and put it into a taboo table. Then this employed bee becomes a scout bee and uses the formula (10) to find a new food source outside the taboo list. The best food source will be updated after each round of foraging. The algorithm loops through the operation of employed bees, onlooker bees, and scout bees until the stop condition is met and the optimal food source is obtained.

Time complexity analysis

The SSABC-SOM process consists of three steps, (1) training SOM network, (2) introducing label information, (3) optimal clustering of neurons in SOM output layer. Let N be the number of samples in the training set, D be the dimension of each sample, X be the number of labeled samples, K be the number of facies and M be the number of neurons in the output layer of SOM network. Assuming that the partial time complexity of the first step is $\mathcal{O}(A)$, we define N in terms of S as $N = \mathcal{O}(S)$, then the $\mathcal{O}(A)$ is equal to $\mathcal{O}(S^2)$ (Roussinov, 1998). The time complexity of K-means is $\mathcal{O}(nkl)$, where n is the number of elements, k is the number of cluster centers and l is the number of iterations (Jassar, 2016), therefore, the time complexity of clustering SOM output layer neurons using K-means is $\mathcal{O}(MKL)$, where L is the maximal iteration. We can conclude that the overall time complexity of the traditional unsupervised method based on SOM and K-means is $\mathcal{O}(A) + \mathcal{O}(MKL)$. In step 2, the time complexity of finding the corresponding neurons with labeled samples is $\mathcal{O}(XM)$, and we can obtain the constraint relationship in $\mathcal{O}(1)$, therefore the time complexity of this step is $\mathcal{O}(XM) +$

$O(1)$. The upper bound of the complexity of ABC algorithm is $O(BDZ)$, where B is the total number of bees and Z is the maximal iteration (Banharnsakun, 2011), therefore, the time complexity of the steps 3 is approximated to $O((M-Y)KL*BDZ)$, where Y is the number of neurons corresponding to the labeled samples. We can conclude that the overall time complexity of the proposed SSABC-SOM method is approximated to $O(A) + O(XM) + O((M-Y)KL*BDZ)$. In general, the maximal M should be equal to $5*\sqrt{N}$ (Estévez, 2012). In practical seismic facies analysis, N is usually large enough to make $N \gg \sqrt{N}$, which makes $O(A) \gg O(MKL)$ and $O(A) \gg O(XM) + O((M-Y)KL*BDZ)$, therefore, compared with the traditional method, the SSABC-SOM method does not increase significant time complexity, the overall time complexity mainly depends on $O(A)$.

Confidence analysis

The seismic facies maps obtain from clustering methods such as SOM used in this study only show the label, which means it can not make a definite quantitative analysis on the classification result of a certain sample, and can not get the indicators that evaluate the overall clustering results. On the basis of SSABC-SOM method and combining with the PNN, this paper calculates the probabilities of each sample that belongs to each category, and the indicators for measuring the classification effect is defined according to the calculated probabilities.

PNN is a non-parametric probability density estimation method that based on Bayes classification rule and originated from Parzen probability density estimation method, it is consist of four layers of neurons, including input layer, pattern layer, summation layer and output layer. The network structure of the network is shown in Fig. 5.

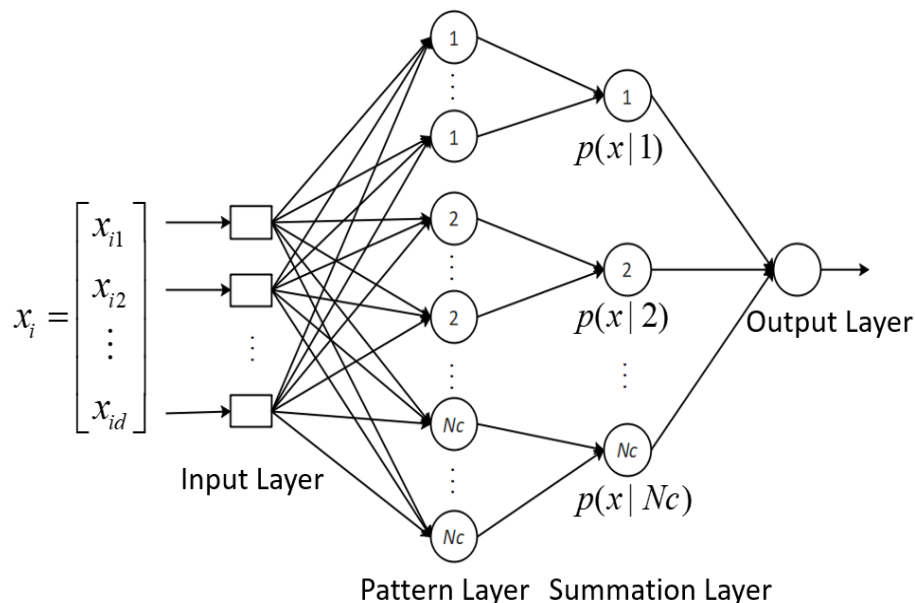


Fig. 5. Network structure of PNN.

In this paper, the optimal clustered neurons in the SSABC-SOM output layer are set as the pattern layer of the PNN network. For a sample x , the probability that this sample belongs to a class is calculated by formula

$$p(x | k) = \frac{1}{N_k \sigma^d} \sum_{i=1}^{N_k} W\left(\frac{\|x - w_i\|}{\sigma}\right), \quad (14)$$

where $p(x|k)$ is the probability that the sample x belongs to the k -th class, N_k is the number of neurons belonging to the k -th class in the SSABC-SOM output layer, d is the dimensionality of x , w_i is the weight vector of the i th neuron within the k -th class in the output layer of SSABC-SOM. σ is the scale factor, $W(\)$ is Kernel function and the Gaussian function is taken in this paper.

$$W(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x^2}{2}}, \quad (15)$$

By the formulas (14) and (15), the probabilities of the sample belonging to each class under the SSABC-SOM network is obtained. For a sample x , three indicators *distinction* (D), *affiliation* (A) and *confidence* (C) for quantitative evaluation the clustering effect are defined as follows

$$D = \frac{Large(p,1) - Large(p,2)}{Large(p,1)}, \quad (16)$$

$$A = \frac{Large(p,1) - Large(p,2)}{Large(p,1) + Large(p,2)}, \quad (17)$$

$$C = \frac{D + A}{2}, \quad (18)$$

where $Large(p,1)$ and $Large(p,2)$ represents the maximum and second-largest probability of $p(x|k), k \in [1, N_c]$, respectively. In other methods such as FCM and LIBSVM, $Large(p,1)$ and $Large(p,2)$ represent the same meaning, that means we can use these indicators to evaluate the results obtained by these methods.

Distinction describes the degree of separation of the identified facies from the rest facies, the higher the measurement, the more obvious the separation from other facies, and the less likely it is to be confused. *Affiliation* describes the fusion between the identified facies and the facies representative, the higher the measurement, the more similar it belongs to this category. *Confidence* combined *distinction* and *affiliation*, the higher *confidence* indicates that the classification result of the sample is more reliable.

SYNTHETIC TEST

To verify the superiority of pre-stack texture features and the effectiveness of the proposed method, we use 30 Hz Ricker wavelet and the Aki-Richards approximation (Aki and Richards, 1980) to generate 4D synthetic pre-stack seismic data based on the stratigraphic structure model shown in Fig. 6a. This model contains four layers construct media whose S-wave propagation velocities, P-wave propagation velocities and densities are shown in Fig. 6, the second layer contains three different large seismic facies and a micro fault. In this study, 40 pre-stack seismic traces collected at different angles of incidence are simulated. In order to increase the complexity of the signals, we add Gaussian noise filtered in the seismic bandwidth (10-15 Hz 50-60 Hz) (Song, 2017) to the 4D synthetic traces with $S/N = 8$ dB. After adding noise, the pre-stack seismic traces at the locations of A_1 , A_2 , A_3 and A_4 are shown in Fig. 7.

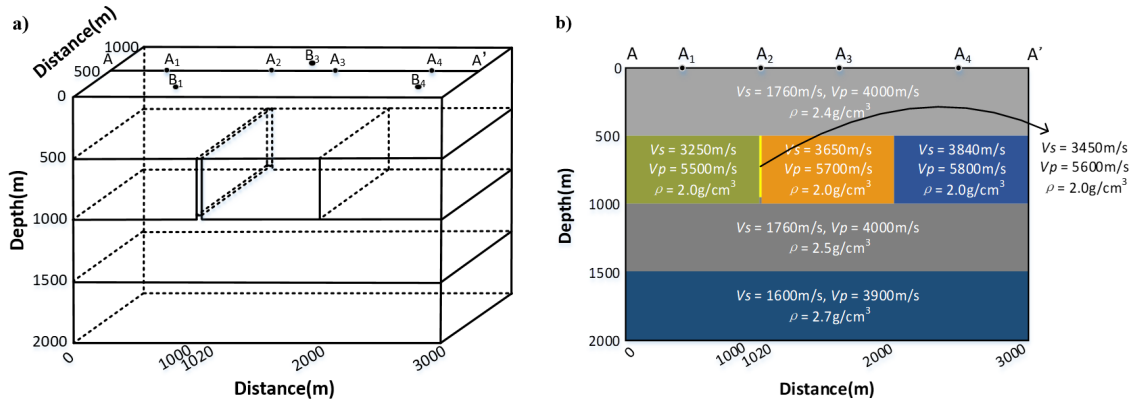


Fig. 6. (a) 3D stratigraphic structure model. (b) Geologic model.

After the stacking processing, the partial post-stack seismic data section along AA' line are shown in Fig. 8a, and the post-stack seismic data at the locations of A_1 , A_2 , A_3 and A_4 are shown Fig. 8b. This study chose the data set from the 40 ms time window around the bottom of second layer to analyze seismic facies. Comparing Figs. 8 and 7, the slight differences between pre-stack traces were disappeared after stacking, which results in a high similarity of their post-stack waveforms on the target interval. We obtained the facies maps using traditional LIBSVM, SOM, FCM and the SSABC-SOM method, the results are shown in Fig. 9, comparing the facies maps obtained from the same method, pre-stack texture features obtained relatively good facies map while post-stack waveforms are completely indistinguishable, which proves the superiority of pre-stack texture features in describing seismic facies. Comparing the facies map obtained from the traditional SOM method (Fig. 9f) and the facies map obtained from the SSABC-SOM method (Fig. 9h), it proves the advantages of the global optimization. Comparing Figs. 9e, f, g and h, these facies maps confirm the effectiveness of the proposed SSABC-SOM method, it can not only obtain the correct facies distribution with a few errors, but also depict the microfacies well.

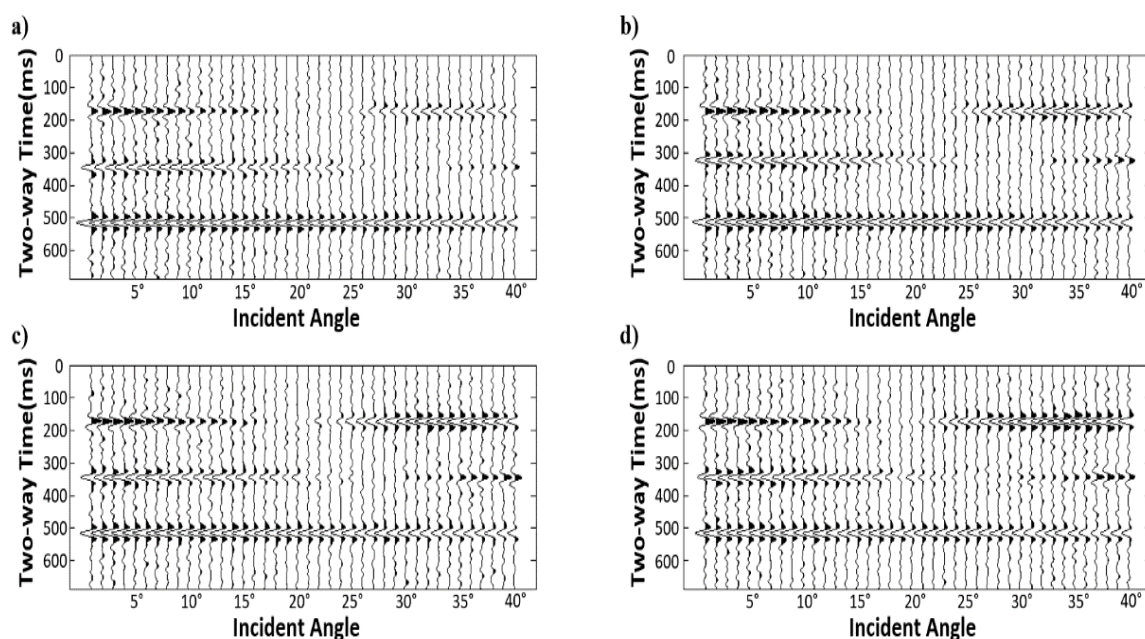


Fig. 7. (a), (b), (c) and (d) are the pre-stack seismic traces of location A_1 , A_2 , A_3 and A_4 , respectively.

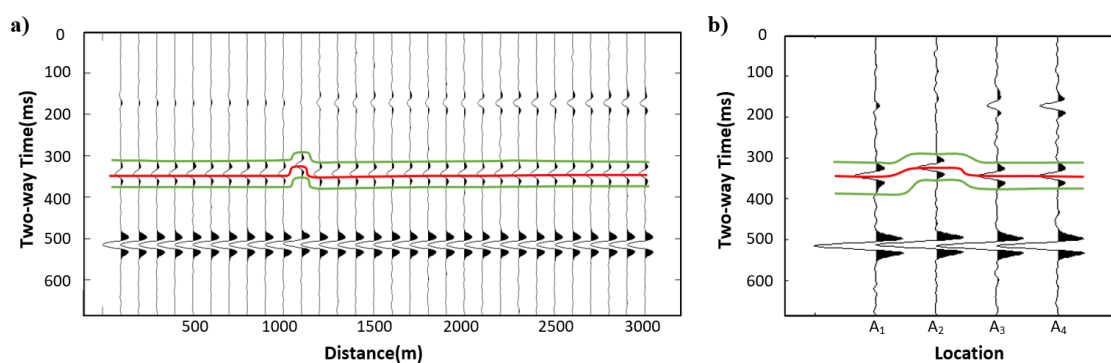


Fig. 8. (a) Partial post-stack seismic data section along AA' line. (b) Post-stack seismic waveform at location A_1 , A_2 , A_3 and A_4 . The red line denotes the target horizon, the green lines represent the top and bottom of the windowed interval.

These classification results are also quantitatively analyzed by the confidence analysis method proposed in this paper. We compare the indicators of *distinction*, *affiliation* and *confidence* of each sample, and the results are shown in Fig. 10. We also calculate the average values for different indicators of each class, and the results are shown in Table 2, this table shows that the three indicators for different methods based on post-stack waveform are very close to 0, which proves that these classification result are completely undesirable, and in those results based on pre-stack texture features, the proposed SSABC-SOM method achieves the best indicators in general. Figs. 10a, b and c show the classification indicators based on LIBSVM method, due to insufficient supervision

information, indicators in the second, third and fourth class are relatively low, and the traditional SOM method still does not perform well in these classes. The indicators based on FCM method are shown in Figs. 10g, h and i, although they have been improved in the fourth class, they remain poorly represented in the second and third classes. In Figs. 10j, k and l, the indicators of all classes are excellent, even on the microfacies and boundaries, there still can get large values, which quantitatively confirms the effectiveness of the proposed SSABC-SOM method.

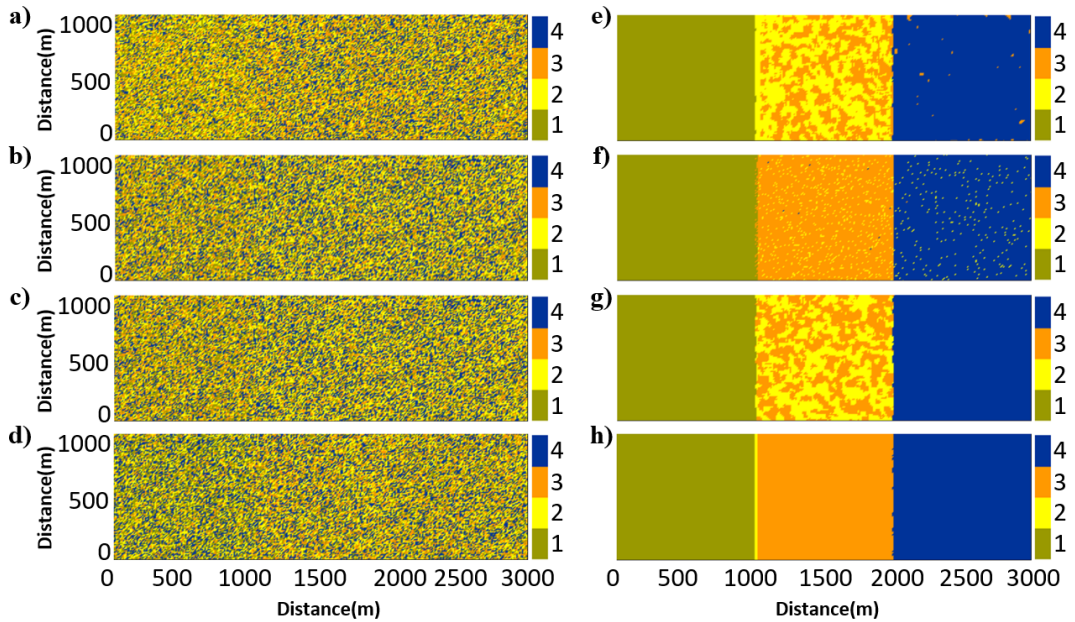


Fig. 9. Facies map obtain from (a) LIBSVM, (b) SOM, (c) FCM and (d) SSABC-SOM based on post-stack waveform. Facies map obtain from (e) LIBSVM, (f) SOM, (g) FCM and (h) SSABC-SOM based on pre-stack texture features.

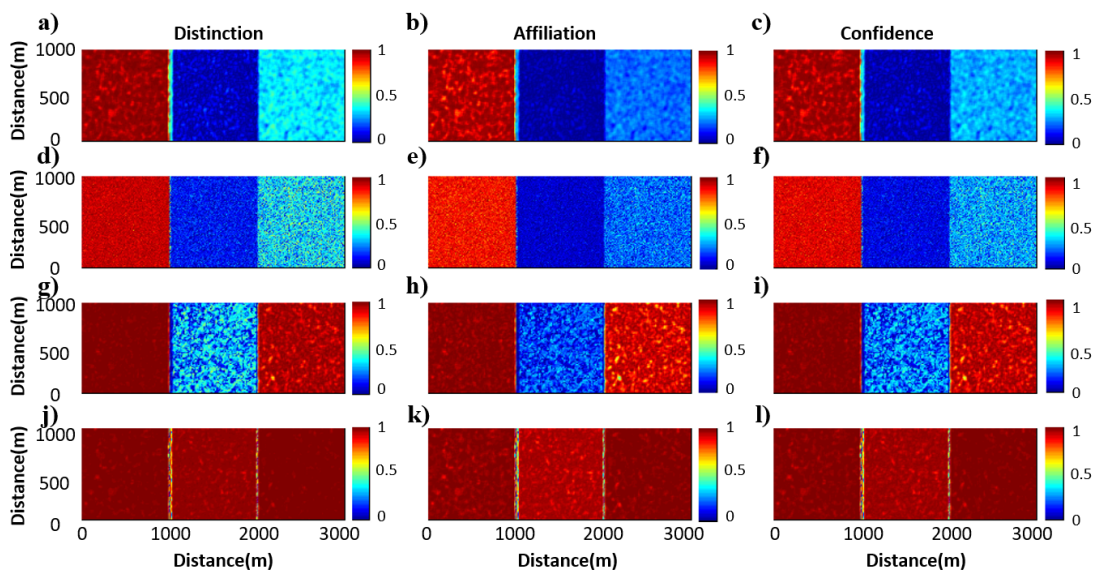


Fig. 10. *Distinction*, *affiliation* and *confidence* maps. (a), (b), and (c) corresponding to the result in Figs. 9e. (d), (e), and (f) corresponding to the result in Figs. 9f. (g), (h), and (i) corresponding to the result in Fig. 9g. (j), (k) and (l) corresponding to the result in Fig. 9h.

Table 2. Confidence measures of different methods with post-stack waveform and pre-stack texture features

CLASS	Indicators Methods	Post-stack waveform based			Pre-stack texture attributes based		
		Distinction	Affiliation	Confidence	Distinction	Affiliation	Confidence
1	LIBSVM	0.00313	0.00177	0.00245	0.90270	0.82581	0.86425
2		0.00162	0.00063	0.00113	0.05252	0.02777	0.04015
3		0.00201	0.00104	0.00153	0.07084	0.03900	0.05492
4		0.00348	0.00196	0.00272	0.32407	0.19405	0.25906
Overall		0.00256	0.00135	0.00196	0.33753	0.27166	0.30460
1	SOM	0.00351	0.00198	0.00270	0.90350	0.83068	0.86709
2		0.00120	0.00060	0.00089	0.04853	0.02532	0.03692
3		0.00282	0.00134	0.00211	0.12711	0.06908	0.09810
4		0.00358	0.00190	0.00273	0.33961	0.21085	0.27523
Overall		0.00280	0.00151	0.00212	0.35470	0.28398	0.31933
1	FCM	0.00423	0.00212	0.00318	0.98125	0.96757	0.97441
2		0.00122	0.00061	0.00092	0.26746	0.16155	0.21450
3		0.00302	0.00152	0.00227	0.26301	0.15865	0.21083
4		0.00371	0.00186	0.00279	0.93749	0.88762	0.91256
Overall		0.00305	0.00153	0.00229	0.61230	0.54385	0.57808
1	SSABC-SOM	0.00472	0.00237	0.00354	0.98062	0.96481	0.97272
2		0.00488	0.00245	0.00366	0.95889	0.92606	0.94247
3		0.00462	0.00232	0.00347	0.75449	0.65946	0.70698
4		0.00436	0.00219	0.00328	0.98254	0.96925	0.97589
Overall		0.00465	0.00233	0.00349	0.91914	0.87990	0.89952

APPLICATION TO FIELD DATA

We now apply the proposed method to a data set collected from a known area in southwestern of China, on the interest horizon, the development thickness of karst body in each well position is obtained according to the log data. A crossline section of this post-stack data set is show in Fig. 11a. We extract the data set according to a time gate of [-15 ms, +35 ms] around the interest horizon to analyze the facies, the continuity property slice along the interest horizon is shown in Fig. 11b. According to logging data, well M11 and M19 are located in strong karst environment, well M8, M18 and M204 are located in karst area, and well M205 is located in non-karst environment. According to these information, this study divides this work area into three facies: strong karst, karst and non-karst. This pre-stack seismic data of this work area contain six wide azimuth gathers with different azimuth, we extract the pre-stack texture features of this data set, and then choose the FCM method that performed better than LIBSVM and SOM in synthetic test to compare with the proposed SSABC-SOM

method. Fig. 12 shows the final horizontal distribution maps of karst development thickness. In term of the qualitative analysis, comparing the result obtained by FCM method, the result obtained by SSABC-SOM is clearer and neater, and the facies map is more consistent with the continuity property slice shown in Fig. 11b. In term of correct classification of well locations, in the map obtained by FCM method, only half of the wells are correctly classified, the locations of well M19 and M8 are classified as non-karst area, and the location of well M204 is classified as strong karst area, which is not consistent with the logging information. In the result obtain by SSABC-SOM method, the environment of each well are correctly classified.

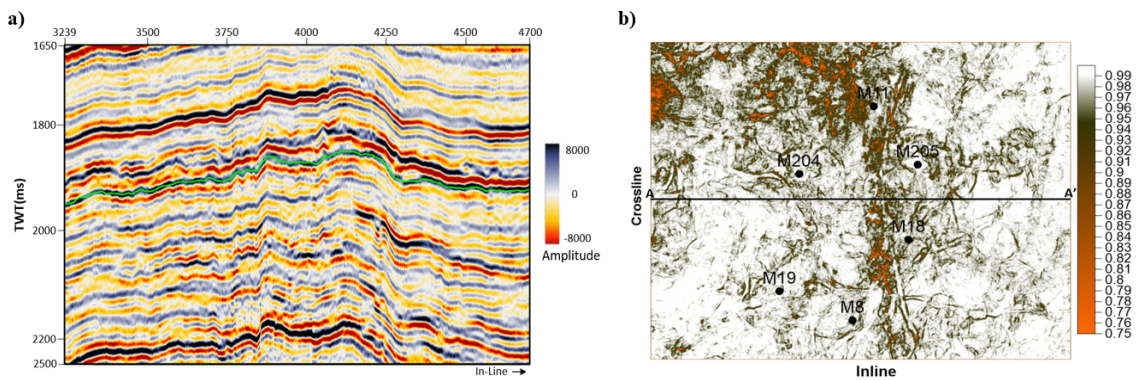


Fig. 11. (a) The crossline section of the post-stack data set along the line AA' shown in (b), the used data is the interpretation horizon marked in green line and the data within a time gate of [-10ms, +30ms] around the horizon. (b). The continuity property slice along the interest horizon.

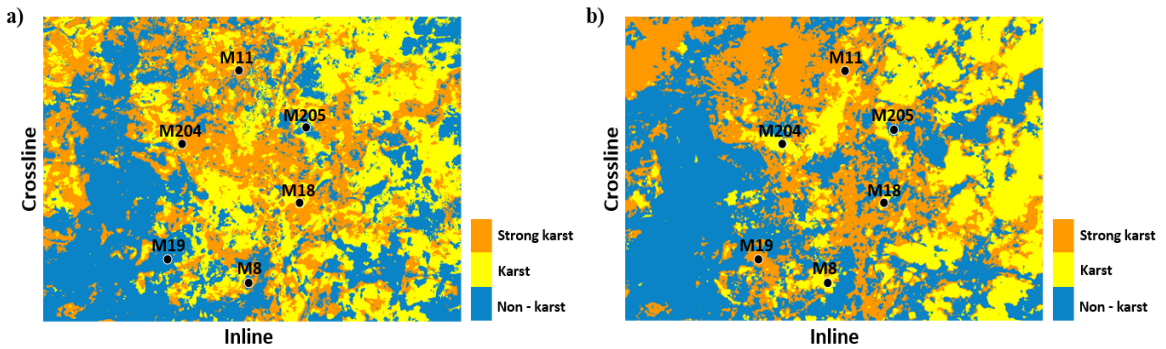


Fig.12. Horizontal distribution maps of karst development thickness, (a) is obtained by FCM method, (b) is obtained by SSABC-SOM method.

For quantitative analysis, both the results shown in Fig. 12 are analyzed by the confidence analysis method presented in this paper, and the results are shown in Fig. 13. In most locations, indicators of Fig. 12b are larger than indicators of Fig. 12a. We calculate the percent statistical distribution of *distinction*, *affiliation* and *confidence* for two results by formula

$$p(i) = \frac{N_i}{N} * 100\% , \quad (19)$$

where N_i represents the number of samples whose indicator value is i , N is the total number of samples. The percent statistical distribution results are shown in Fig. 14. The percentages for result of SSABC-SOM method are more in high indicators section and less in low indicators section than that for result of FCM method, which proves the result of the proposed SSABC-SOM method is more reliable.

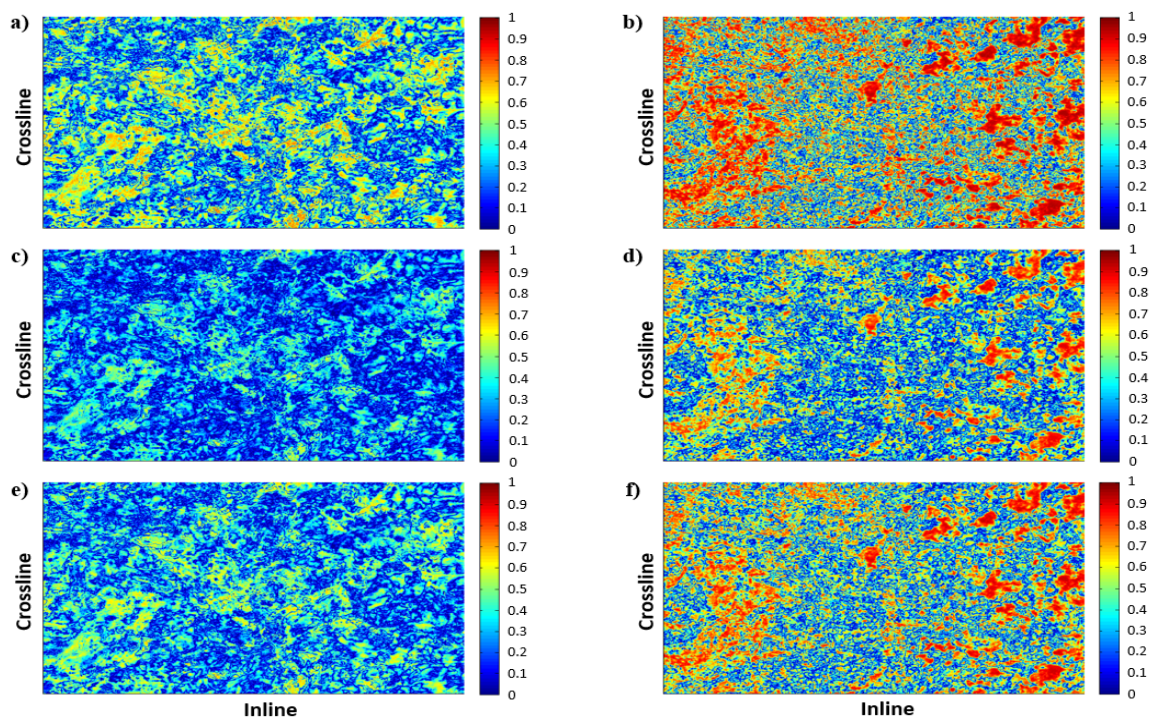


Fig. 13. Indicator maps of the results shown in Fig. 12. (a), (c) and (e) is the *distinction*, *affiliation* and *confidence* map of the result obtained by FCM method. (b), (d) and (f) is the *distinction*, *affiliation* and *confidence* map of the result obtained by SSABC-SOM method.

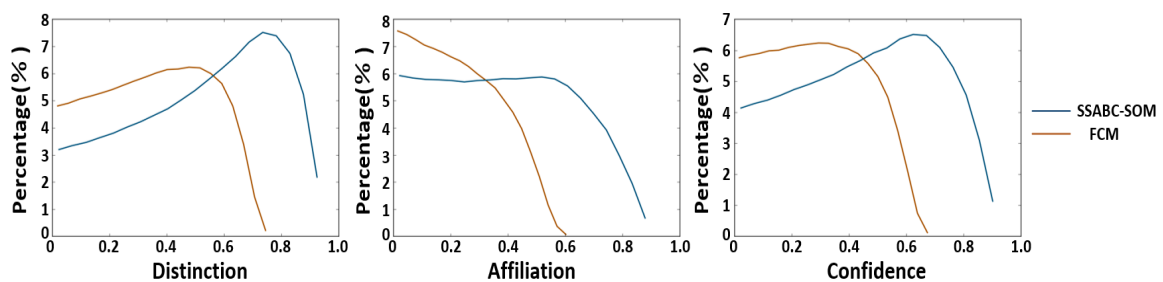


Fig. 14. Indicators statistics of the results shown in Fig. 12.

CONCLUSION

Seismic facies analysis is an important method for regional stratigraphic interpretation, sedimentary system determination, and sedimentary development history interpretation, it can be used to predict the oil source area and the reservoir belts. There are some problems in seismic facies analysis through the traditional waveform classification methods such as some seismic reflect information of micro geologic is disappears in post-stack data, and the small amount of log data is not used effectively. Here, we extracted the texture features from pre-stack data and preserved the tiny information of stratum. Then the logging data and ABC algorithm were employed to the training of classifier, which reduced the multiple solvability of target recognition, and the result that is more consistent with the actual geological conditions is obtained. Finally, combined with PNN, we quantitatively analyzed the classification results.

In the synthetic test, a 3D stratigraphic structure model that contain different facies was designed to generate the pre-stack seismic data, and the corresponding post-stack data set is obtained by stacking processing. By comparing the processing results of these data sets with different methods, the superiority of pre-stack texture and the effectiveness of the proposed SSABC-SOM method are demonstrated.

Finally, the method proposed in this paper is applied to field data. Compared with the traditional method, the SSABC-SOM method obtained a better facies map, meanwhile, the confidence analysis results also show the effectiveness of the method.

In our study, the processing of pre-stack data such as extract the pre-stack texture features takes most of the time of the project due to the big size of the pre-stack seismic data, how to compress the pre-stack data under the premise of keeping fine information remains a question for the future.

ACKNOWLEDGMENTS

We acknowledge financial support from the National Natural Science Foundation of China (Grant No. 41604107 & No. U1562218), Scientific Research and Technology Development Project of CNPC (2018D-4407, 2017D-3503) and Sinopec's Science and Technology Major Project "Hydrocarbon accumulation law and target evaluation in carboniferous area" (Grant No. P17001-4), and thanks to the related company for providing the data.

REFERENCES

- Aki, K. and Richards, P., 1980. *Quantitative Seismology, Theory and Methods*. W.H. Freeman and Co., San Francisco.
- Altinel, B. and Ganiz, M.C., 2016. A new hybrid semi-supervised algorithm for text classification with class-based semantics. *Knowledge-Based Syst.*, 108: 50-64.
- Banharnsakun, A., Achalakul, T. and Sirinaovakul, B., 2011. The best-so-far selection in artificial bee colony algorithm. *Appl. Soft Comput. J.*, 11(2):2888-2901.
- Angelo, M.S., Matos, M.C. and Marfurt, K.J., 2009. Integrated seismic texture segmentation and clustering analysis to delineate reservoir geometry. *Expanded Abstr.*, 79th Ann. Internat. SEG Mtg., Houston: 1107-1111.
- Chopra, S. and Marfurt, K.J., 2005. Seismic attributes - A historical perspective. *Geophysics*, 70(5): 3S0-28S0.
- Chopra, S. and Alexeev, V., 2006. Applications of texture attribute analysis to 3D seismic data. *The Leading Edge*, 25(8): 934-940.
- Chuai, X.Y., Wang, S.X., Shi, P.D., Wei, X. and Chen, W., 2014. Applications of texture attribute analysis to seismic interpretation. *J. Cent. South Univ.*, 21: 3617-3626.
- de Matos, M.C., Osorio, P.L. and Johann, P.R., 2007. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *Geophysics*, 72(1): P9-P21.
- Dornaika, F. and Traboulsi, Y.E., 2017. Matrix exponential based semi-supervised discriminant embedding for image classification. *Pattern Recognit.*, 61: 92-103.
- Erman, J., Mahanti, A., Arlitt, M., Cohen, I. and Williamson, C., 2007. Offline/real time traffic classification using semi-supervised learning. *Perform. Evaluat.*, 64: 1194-1213.
- Estévez, P., Príncipe, J. and Zegers, P., 2012. *Advances in Self-Organizing Maps*. 9th Internat. Workshop, WSOM 2012, Santiago, Chile, December 12-14. *Proc., Advances in Intelligent Systems and Computing*.
- Gao, D.L., 2006. Structure-oriented texture model regression for seismic structure visualization and interpretation. *Gulf Coast Assoc. Geol. Soc. Transact.*, 56(3): 207-216.
- Gao, D.L., 2011. Latest developments in seismic texture analysis for subsurface structure, facies, and reservoir characterization: A review. *Geophysics*, 76(2): W1-W13.
- Guillen, P., Larrazabal, G., Gonzalez, G. and Sineva, D., 2015. Detecting salt body using texture classification. 14th Internat. Congr. Brazil. Geophys. Soc., Rio de Janeiro: 1155-1158.
- Haralick, R., Shanmugam K. and Dinstein, I., 1973. Textural features for image classification. *IEEE Transact. Systems, Man., Cybernet.*, 23: 610-621.
- Hong, Y. and Zhu, W.P., 2015. Spatial co-training for semi-supervised image classification. *Patt. Recognit. Lett.*, 63: 59-65.
- Jassar, K.K. and Dhindsa, K.S., 2016. Comparative study and performance analysis of clustering algorithms. *IJCA Proc. Internat. Conf. ICT for Healthcare, ICTHC 2015*: 1-6.
- Lima, L.A., Görnitz, N., Varella, L.E., Vellasco, M., Müller, K. and Nakajima, S., 2017. Porosity estimation by semi-supervised learning with sparsely available labeled samples. *Comput. Geosci.*, 106: 33-48.
- Marfurt, K.J., 2014. Seismic attributes and the road ahead. *Expanded Abstr.*, 84th Ann. Internat. SEG Mtg., Denver: 4421-4426.
- Marroquín, I.D., Brault, J.J. and Hart, B.S., 2009. A visual data-mining methodology for seismic facies analysis. Part 1 - Testing and comparison with other unsupervised clustering methods. *Geophysics*, 74(1): P1-P11.

- Nivlet, P., 2007. Uncertainties in seismic facies analysis for reservoir characterisation or monitoring. Causes and consequences. *Oil Gas Sci. Technol.*, 62: 225-235.
- Portela, N.M., Cavalcanti, G.D.C. and Ren, I.T., 2014. Semi-supervised clustering for MR brain image segmentation. *Expert Syst. Applicat.*, 41: 1492-1497.
- Qi, J., Lin, T.F., Zhao, T., Li, F.Y. and Marfurt, K.J., 2016. Semisupervised multiattribute seismic facies analysis. *Interpretation*, 4(1): SB91-SB106.
- Roussinov, D.G. and Chen, H., 1998. A scalable self-organizing map algorithm for textual classification: a neural network approach to thesaurus generation. *Communicat. Cognit. Artific. Intellig. Spring*, 15: 81-112.
- Roy, A., Romero-Peláez, A.S., Kwiatkowski, T.J. and Marfurt, K.J., 2014. Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico. *Interpretation*, 2(1): SA31-SA47.
- Ruffo, P., Corradi, A., Corrao, A., Corrao, C.A. and Visentin, C., 2007. 3D hydrocarbon migration in alternate sand-shale environment through percolation technique. AAPG Hedberg Conf., The Hague, The Netherlands.
- Saha, S., Alok, A.K. and Ekbal, A., 2016. Brain image segmentation using semi-supervised clustering. *Expert Syst. Applicat.*, 52: 50-63.
- Shafiq, M.A., Wang, Z., Amin, A., Hegazy, T., Deriche, M. and Al Regib, G., 2015. Detection of salt-dome boundary surfaces in migrated seismic volumes using gradient of textures. *Expanded Abstr.*, 85th Ann. Internat. SEG Mtg., New Orleans: 1811-1815.
- Song, C.Y., Liu, Z.N., Cai, H.P., Qian, F. and Hu, G.M., 2016. Pre-stack-texture-based reservoir characteristics and seismic facies analysis. *Appl. Geophys.*, 13: 69-79.
- Song, C.Y., Liu, Z.N., Cai, H.P., Wang, Y.J., Li, X.M. and Hu, G.M., 2017. Unsupervised seismic facies analysis with spatial constraints using regularized fuzzy c-means. *J. Geophys. Engineer.*, 14:1535-1543.
- Song, C.Y., Liu, Z.N., Wang, Y.J., Li, X.M. and Hu, G.M., 2017. Multi-waveform classification for seismic facies analysis. *Comput. Geosci.*, 101: 1-9.
- Song, C.Y., Liu, Z.N., Wang, Y.J., Xu, F., Li, X.M. and Hu, G.M., 2018. Adaptive phase k-means algorithm for waveform classification. *Explor. Geophys.*, 49: 213-219.
- Tamir, H. and Ghassan, A., 2014. Texture attributes for detecting salt bodies in seismic data. *Expanded Abstr.*, 84th Ann. Internat. SEG Mtg., Denver: 1455-1458.
- Ulaby, R.T., Kouyate, F., Brisco, B. and Williams, T.H.L., 1986. Textural information in SAR images. *IEEE Transact. Geosci. Remote Sens.*, 24: 235-245.
- Yenugu, M., Marfurt, K.J. and Matson, S., 2010. Seismic texture analysis for reservoir prediction and characterization. *The Leading Edge*, 29: 1116-1121.
- Zhao, T., Jayaram, V., Roy, A. and Marfurt, K.J., 2015. A comparison of classification techniques for seismic facies recognition. *Interpretation*, 3(4): SAE29-SAE58.
- Zhang, W., Tang, X.J. and Yoshida, T., 2015. TESC: An approach to TExt classification using semi-supervised clustering. *Knowledge-Based Syst.*, 75: 152-160.
- Zhang, Y., Zheng, X.D., Li, J.S., Lu, J.T., Cao, C.Y. and Sui, J.K., 2015. Unsupervised seismic facies analysis technology based on SOM and PSO. *Chin. J. Geophys.*, 58: 3412-3423.