

The Improvement of Neural Network Cascade-Correlation Algorithm and Its Application in Picking Seismic First Break

LI Zhe^{[a],*}; SHENG Rong^[b]; XU Weixiu^[c]; YANG Jing^[c]

^[a]College of information and control engineering, China University of Petroleum (East China), China.

^[b]Xinjiang Oilfield Karamay Power Plant, Xinjiang.

^[c]Sinopec Oil Engineering Geophysical Limited Company Shengli Branch, Dongying, China.

*Corresponding author.

Received 12 April 2013; accepted 17 June 2013

Abstract

Neural Network is a kind of widely used seismic wave travel time auto-picking method. Most commercial software such as Promax often uses Back Propagation (BP) neural network. Here we introduce a cascade-correlation algorithm for constructing neural network. The algorithm's convergence is faster than BP algorithm and can determine its own network architecture according to training samples, in addition, it can be able to expand network topology to learn new samples. The cascaded-correlation algorithm is improved. Different from the standard cascade-correlation algorithm, improved algorithm starts at an appropriate BP network architecture (exits hidden units), but the standard one's initial network only includes input layer and output layer. In addition, in order to prevent weight-illgrowth, adding regularization term to the objective function when training candidate hidden units can decay weights. The simulation experiment demonstrates that improved cascade-correlation algorithm is faster convergence speed and stronger generalization ability. Analytically study five attributes, including instantaneous intensity ratio, amplitude, frequency, curve length ratio, adjacent seismic channel correlation. Intersection figure shows that these five attributes have distinctiveness of first break and stability. The neural network first break picking method of this paper has achieved good effect in testing actual seismic data.

Keywords: Neural network; Cascade-correlation algorithm; Picking seismic first break

Li, Z., Sheng, R., Xu, W. X., & Yang, J. (2013). The Improvement of Neural Network Cascade-Correlation Algorithm and Its Application in Picking Seismic First Break. *Advances in Petroleum Exploration and Development*, 5(2), 41-51. Available from: URL: <http://www.cscanada.net/index.php/aped/article/view/j.aped.1925543820130502.1503> DOI: <http://dx.doi.org/10.3968/j.aped.1925543820130502.1503>

PREFACE

First break picking in seismic data is a premise for some techniques, such as refraction static correction, tomography and vertical seismic profile (VSP). Whether first break is picked correctly or not affects the accuracy of subsequent processes. At present, in spite of its development in the past decades, it is still difficult to achieve good picking effect in complex surface and low SNR.

People usually detect seismic wave's first break time according to the variation of amplitude, energy, frequency or phase of seismic signal, Besides, adjacent seismic channel correlation is also used to detect first break time. Up to now, people have already proposed many first break picking methods, such as energy ratio method^[1], instantaneous intensity ratio method^[2], neural network technology, graphic processing technology^[3], fractal method^[4]. Energy ratio method, instantaneous intensity ratio method and fractal dimension method are used to detect first break based on the time window characteristic of single seismic channel, without considering the interrelation of many seismic channels. In addition, these methods only consider single characteristic and don't consider comprehensively other characteristics. As a result, these algorithms don't have high stability, and picking effect is not ideal in low SNR. So in order to enhance the stability and accuracy of first break picking algorithm, we need to consider various attributes and the

interrelation of adjacent seismic traces. Obviously, neural network first break picking method makes it.

At present, because BP network architecture is simple and easy to implement, BP neural network has been mostly studied in the field of first break picking. But BP algorithm has some problems, such as low convergence speed, prone to getting trapped in local extremum and difficult to determine network architecture. In addition, because BP network is static, it is unable to learn extra assembly of samples to expand network knowledge. Obviously this fails to satisfy first break picking requirement in the region with complex topography. Therefore, BP algorithm isn't widely applied in the commercial field^[13]. However, the cascaded-correlation algorithm can well overcome the problems which BP algorithm has, and has stronger potential in the filed of first break picking.

By means of neural network's learning samples, neural network first break picking method creates a classifying rule, by which we take a pattern recognition for first arrival. Whether the classifying rule is right or not depends on two aspects. One is neural network's training quality, another is proper selection of attributes. Because the cascaded-correlation algorithm starts from a minimal network (without hidden units), the initial stage of training will cost a lot of processes and time. Thus this will increase amount of calculation and simultaneously reduce convergence speed of network^[5,6]. The improvement of the cascaded-correlation algorithm is discussed in this paper. The improved algorithm begins to learn from an appropriate network architecture, and has more efficient calculation and faster convergence speed. At the same time, in order to guarantee generalization ability of neural network, a regularization term is added to the object function when training candidate hidden units using CC algorithm. This method can effectively prevent weight-illgrowth of candidate units, improve the generalization ability of network, and can also accelerate the convergence speed of neural network.

The effect of first break picking simultaneously depends on the selection of attributes. The selected attributes must have certain stability^[7]. The combination of various attributes must be able to distinguish first arrival from no first arrival.

1. THE CASCADE-CORRELATION ALGORITHM

The cascade-correlation algorithm^[8,9] (Cascade-Correlation, short for CC algorithm), which constructs neural network from bottom to top, was proposed by Fahlman in order to solve the problem of low convergence speed of traditional BP algorithm. The cascaded-correlation algorithm includes two meanings. First of all, its network architecture is cascade-correlative. Every time

a hidden unit is added, its position is cascaded with pre-existing hidden units' one. Once a hidden unit is added, its position and input weight will no longer vary. In addition, learning algorithm adopts correlation learning algorithm. Every time a new hidden unit is added, maximize the correlation between its output and original network output residual by adjusting its input weight in order to reduce error to the greatest extent.

CC algorithm has several advantages over BP algorithm. 1) It can determine its own network architecture by itself. 2) Its learning speed is quite fast. 3) It can still retain network architecture and knowledge it has built when assembly of samples vary. 4) when training extra assembly of samples, It can expand the size of network and thus combine new samples with original knowledge system.

Implementation process of CC algorithm follows:

(1) Initialize network. Construct a two-layer network only including input layer and output layer, and initialize weight with random value.

(2) Train output layer. Adjust weight of output units using arbitrary algorithm (gradient descent method or Quickprop algorithm). After a certain number of training cycles, obvious error improvement doesn't appear, and evaluate the error. If the error is enough small, stop training. Otherwise switch to 3.

(3) Train candidate hidden units. Create a new hidden unit which will connect input hidden units and pre-existing hidden units, and adjust weights so as to maximize the correlation between new hidden units' output and network output residual error.

(4) Freeze and connect new hidden unit. Fix weight of new hidden unit and connect its output to network output layer. Adjust the output layer again, and switch to 2.

Although CC algorithm has some advantages compared with BP network, some problems is still needed to solve when applying it in first break picking. 1) Because CC algorithm begins with a minimal network architecture only including input layer and output layer, and automatically adds hidden units one by one until object network is convergent, this means to need a lot of processes and time. So this increases amount of calculation and gets convergence speed slow^[5,6]. 2) CC algorithm constructs a multilayer neural network. Adding a hidden unit amounts to adding a new hidden layer in effect. Newly added hidden units are more, the neural network constructed by CC algorithm is more complex. This tends to lead to over fitting of neural network for training samples^[9]. 3) Maximize the correlation between candidate hidden units and original network output residual error. This often gets weights of candidate hidden units to overcompensate for error, which is called weight-illgrowth^[10]. The phenomenon severely affects generalization ability of network.

1.1 The Improvement of CC Algorithm

Because original CC algorithm starts from a minimal network (without hidden units), this must increase the amount of calculation and simultaneously reduce network convergence speed. If training begins with an appropriate network, it must enhance network convergence speed^[5]. The improved CC algorithm also includes three stages: initial network training, candidate hidden units training and output layer units training.

The improved algorithm starts from an appropriate BP network. Figure 1 gives three-layer BP network

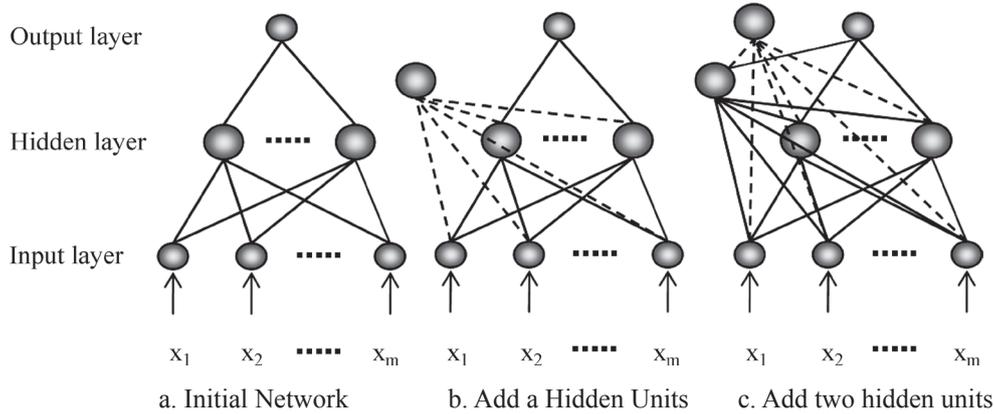


Figure 1
Three-Layer BP Network Structure

In candidate hidden units training stage, train simultaneously many candidate hidden units, and then add the best candidate unit to network, see Figure 1(b) and Figure 1(c). The candidate hidden units' input connects both input units of network and all hidden units' output. The candidate units' output keeps open, without being connected to output units' input. In the process of training, the input weights of each candidate hidden unit are adjusted by Quickprop algorithm, so as to maximize correlation C between hidden units' output V_p and network output residual error $E_{p,o}$, correlation C is as follows:

$$C = \sum_o \left| \sum_p (V_p - \bar{V})(E_{p,o} - \bar{E}_o) \right| \quad (1)$$

in this formula, o represents the network's output units, p represents training samples, \bar{V} is the average of candidate hidden units' output V_p for all samples, \bar{E}_o is the average of original network's output units' residual error $E_{p,o}$ for all samples.

In order to prevent weight-illgrowth, when maximizing the correlation C, we add regularization item to formula (1). Then formula (1) is modified by:

$$C = \sum_o \left| \sum_p (V_p - \bar{V})(E_{p,o} - \bar{E}_o) \right| - \frac{\lambda}{2} \sum w_i^2 \quad (2)$$

in the formula, λ is regularization coefficient. Adding regularization item to object function of candidate hidden units training can effectively decay weight,

structure, and the number of hidden units is calculated by $N = 2^{m-1}/m + 1$, where m is the number of input units, the number of output units is one, in practical application, we determine the number of output units according to actual situation. In initial network training stage, we adopt Quickprop algorithm to minimize object error (including candidate units training and output units training) instead of gradient descent method among classical BP algorithm. When initial network's global mean square error keeps stable, stop training and freeze weights of hidden input units.

prevent weight-illgrowth and simultaneously reduce the complexity of network. Therefore this enhances generalization ability of network.

When the correlation of between all of candidate hidden units and original network no longer varies after many times training, stop training and then select candidate hidden unit with the biggest correlation from several hidden units (generally from four to eight) as new hidden unit of network. Connect just selected unit's output to network output unit's input, freeze the weight of new hidden unit's input, and delete others. Next begin output layer units training.

In the output layer training stage, minimize global mean square error by adjusting weights of connecting output layer units. Global mean square error is as follow:

$$E_A = \frac{1}{2} \sum_p \sum_o (y_{p,o} - t_{p,o})^2 \quad (3)$$

in this formula, o represents network output layer units, p represents training samples, m is the number of output units, $t_{p,o}$ is the expectation of samples p among output layer units o , $y_{p,o}$ is network output value of samples p among output layer units o .

Train algorithm can adopt gradient descent method or Quickprop algorithm, but Quickprop algorithm has some advantage in convergence speed.

Training algorithm may adopts gradient descent method or Quickprop algorithm, but the latter has faster convergence speed. Suppose that original hidden units still

is effective for object function in the process of training. So keep fixing weights of all original hidden units' input and only allow to change weights of output units' input. When no significant global mean square error reduction occurs, we stop training and then begin a new round of candidate hidden units training. Throughout the whole training process, the candidate units training stage and output layer training stage are repeating alternately, until global mean square error is smaller than object value.

1.2 Quickprop Algorithm

Quickprop algorithm, which is proposed by Fahlman who also proposes CC algorithm, brings advantages of many optimal back propagation algorithms together. Quickprop algorithm utilizes curved part of error plane to accelerate learning, which needs to calculate error's second-order derivative. It has two basic hypotheses^[11]. First of all, error function curve for each weight is two-order curve; in addition, slope variation of error curve for each weight is not affected by other weights. Quickprop algorithm calculates directional derivative of each weight. Thus, according to the slopes and weights of current time and previous time, we can obtain the best adjustment of current time by calculating the parabola's minimum.

$$\Delta w_{ij}(t+1) = \frac{S(t+1)}{S(t) - S(t+1)} \Delta w_{ij}(t) \quad (4)$$

in this formula, Δw_{ij} is connecting weight adjustment of unit i and j , $S(t+1)$ and $S(t)$ are partial derivative of current time and previous time's error function for w_{ij} .

In addition, When the direction of $S(t+1)$ and $S(t)$ is identical, with the value of $S(t)$ almost the same as $S(t+1)$, or with the value of $S(t)$ bigger than $S(t+1)$, the adjustment calculated by formula (4) will be very big, or adjust in reverse direction to the maximum point of parabola. If we only use formula (4) to adjust weights, the algorithm will be instable and even difficult to be convergent. The adjusting weight formula this paper adopts is as follows:

$$\Delta w_{ij}(t+1) = \mu\beta\Delta w_{ij}(t) - (1-\mu)\eta S(t) \quad (5)$$

in this formula, η is learning speed, μ is adjustment varying from 0 to 1, especially, when μ equals to 0, the formula (5) will degenerate into gradient descent formula,

$$\beta = \begin{cases} \delta & \text{if } \left| \frac{S(t+1)}{S(t) - S(t+1)} \right| > \delta \text{ or } S(t+1) \cdot S(t) > 0 \\ \frac{S(t+1)}{S(t) - S(t+1)} & \text{others} \end{cases} \quad (6)$$

in this formula, $S(t) = \frac{\partial E_A}{\partial w_{ij}}$ or $-\frac{\partial C}{\partial w_{ij}}$ is error's first-order partial derivative or correlation's one respectively.

Selecting proper parameters can get the algorithm to be stable and accelerate convergence speed. When the algorithm begins, Δw_{ij} equals to 0 and isn't adjusted by formula (5), we calculate the weight adjustment of next

moment using gradient descent method, that is

$$S(t) = \eta \frac{\partial E_A}{\partial w_{ij}} \quad (7)$$

1.3 Simulation Experiment

In order to test the convergence performance and generalization ability of the improved cascade-correlation neural network algorithm, we have done a simulation experiment using original CC algorithm and the improved one respectively. The object function of neural network training is as follows:

$$y_i = \frac{2}{5} \sin(2\pi x_i) + \frac{1}{2} \quad (8)$$

equally spaced values are used as training points with x_i varying from 0 to 1, that is, input x_i is as follows:

$$x_i = \frac{i}{50} \quad (i = 0, 1, 2, \dots, 50) \quad (9)$$

Thus we can obtain 50 sample pairs $\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_{50}, y_{50})\}$ as neural network's training samples. 500 testing samples whose x_i values are random from 0 to 1 are input into trained neural network to get the approximation values of function. Actual function values are calculated by the formula (8).

Initial network is 1:2:1 three-layer BP network. Sigmoidal function is used as activation function, and initial weights are random from -1.0 to 1.0. In the training process of candidate units, 8 candidate units are trained simultaneously. The mean square error contrast diagram of original CC algorithm and the improved one is shown below.

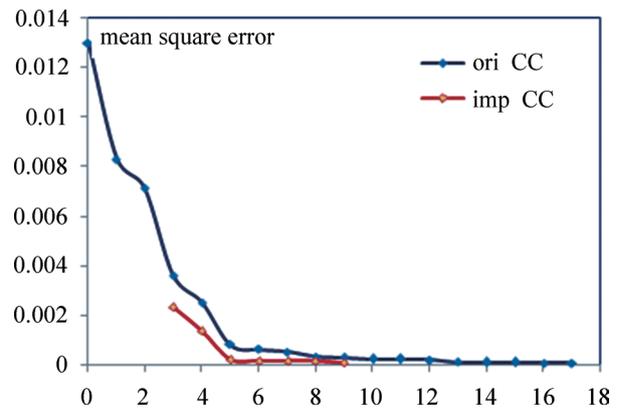


Figure 2 The Mean Square Error Contrast Diagram

As we can see from the diagram above, the improved CC algorithm whose mean square error is 0.0001 only needs to install 7 new hidden units to be completely convergent, but original one needs 17 new hidden units to converge to object function.

500 testing samples whose values are random from 0 to 1 are input into well trained neural network to test neural network's function approximation effect. Network output

results are shown in Figure 3. Original CC algorithm's function approximation has obvious zigzag phenomenon, and weight-illgrowth affects network generalization ability, which is shown in Figure 3 red curve. The green

curve in figure 6 represents function approximation effect of the improved one, has no zigzag phenomenon, and fits the actual curve very well.

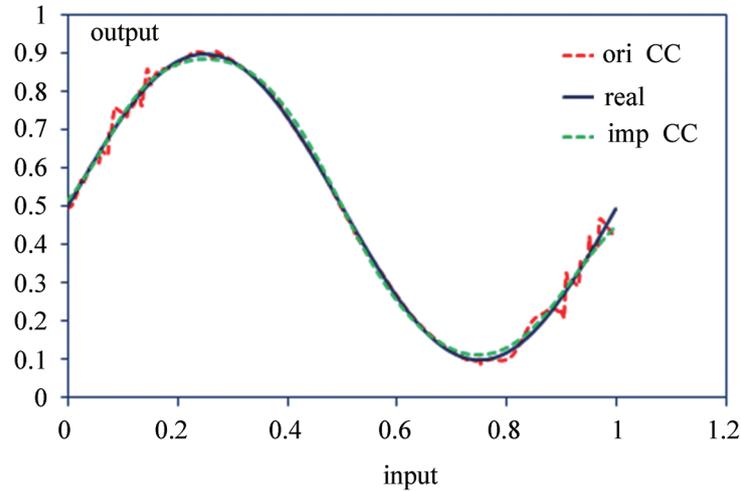


Figure 3
The Function Approximation Figure

In the process of simulation, we take 30 times training and testing. Table 1 is contrast performance table of original CC algorithm and the improved one. As we can

see from the table below, the improved CC algorithm not only accelerates convergence speed, but also enhances the generalization ability of neural network to some degree.

Table 1
Algorithm Performance Contrast Table

Algorithm	Average of new hidden units	Average training error	Average training time	Average testing error
Original CC algorithm	16.5	7.82×10^{-4}	43.6s	1.63×10^{-3}
Improved CC algorithm	7	4.15×10^{-4}	5.7s	5.19×10^{-4}

2. ATTRIBUTE EXTRACTING AND INTERSECTION ANALYSIS

The selection of attributes is an important link in the process of neural network first break picking, and the stability of attributes affects the accuracy of first break picking. Different attributes in different regions and for different seismic sources will bring out different picking effects. So the attribute which is stable and can correctly distinguish first break from no first break is a premise for neural network to detect first arrival.

By studying and analyzing various attributes and testing different kinds of seismic data via programming, we choose instantaneous intensity ratio, amplitude, frequency, curve length ratio and adjacent channel correlation to comprehensively detect first break. Instantaneous intensity ratio is that we can obtain instantaneous amplitude by doing Hilbert transform of seismic channels and complex seismic channels analysis technique. Then we calculate mean square root ratio of the sum of squared instantaneous amplitudes in adjacent time windows as instantaneous intensity ratio. Frequency is used to investigate the main frequency of local wave shape

of seismic channel. Generally we obtain main frequency by doing short time Fourier transform of seismic data in a Gauss time window. Figure 4 is the 3D intersection figure of amplitude, frequency and instantaneous intensity ratio. From the figure below, we can see obvious classifying boundary between first arrival and no first arrival.

In order to investigate the similarity of local wave shape of between current channel and n adjacent channels before and after it, adjacent channel correlation is proposed with aim to detect the event of first arrival. The calculation method is that we can obtain the maximum of the average of correlation between current channel and n adjacent channels before and after it using dip scanning correlation method^[14]. The calculated maximum ranges from 0 to 1. The first arrival of both current channel and n adjacent channels before and after it is generally strongly similar and the attribute value is higher, see Figure 5. The random noise exists before first arrival, and the attribute value is lower. If the event of effective waves after first arrival exists, the attribute value is still higher. Thus the attribute is semi-stable and is used to detect first break time by combining other attributes.

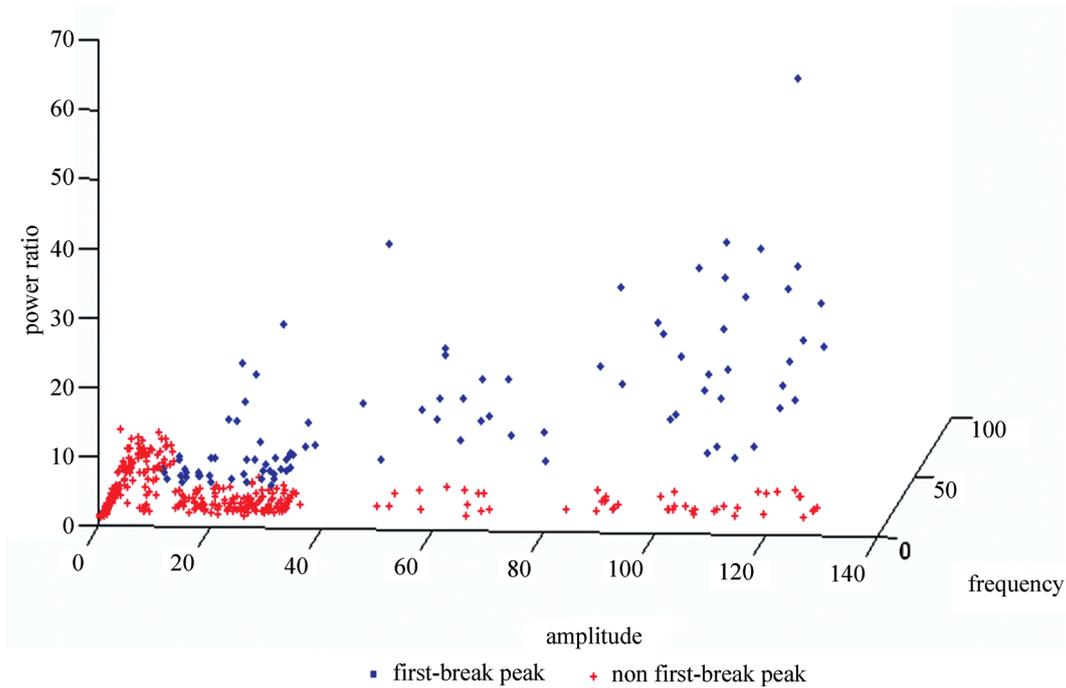


Figure 4
The 3D Intersection Figure of Amplitude, Frequency and Instantaneous Intensity Ratio

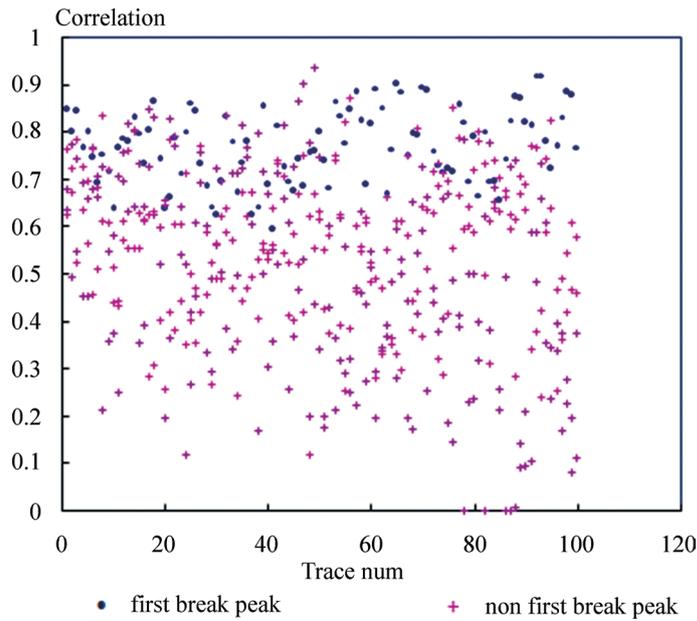


Figure 5
The Distribution Adjacent Seismic Channel Correlation with channel

Curve length ratio uses the length variation of seismic wave's envelop in a time window as an attribute, by which to detect first break. Curve length ratio is defined as the ratio of the line integral of seismic wave in two adjacent time windows. It represents the seismic wave's characteristics of amplitude and frequency, and its

variation also can reflect the variation of seismic wave shape^[19]. Figure 6 is the intersection figure of curve length ratio and adjacent seismic channel correlation. Figure 7 is the 3D intersection figure of curve length ratio, amplitude and frequency. From the two figures below, we can also see obvious classifying boundary between first arrival and no first arrival.

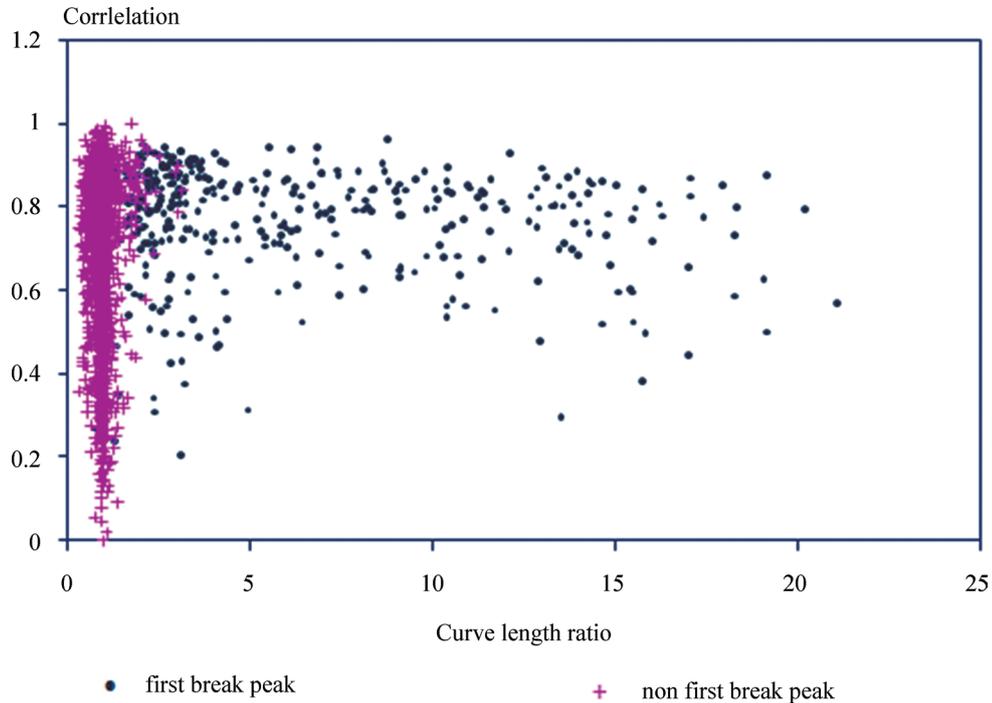


Figure 6
The Intersection Figure of Adjacent Seismic Channel Correlation and Curve Length Ratio

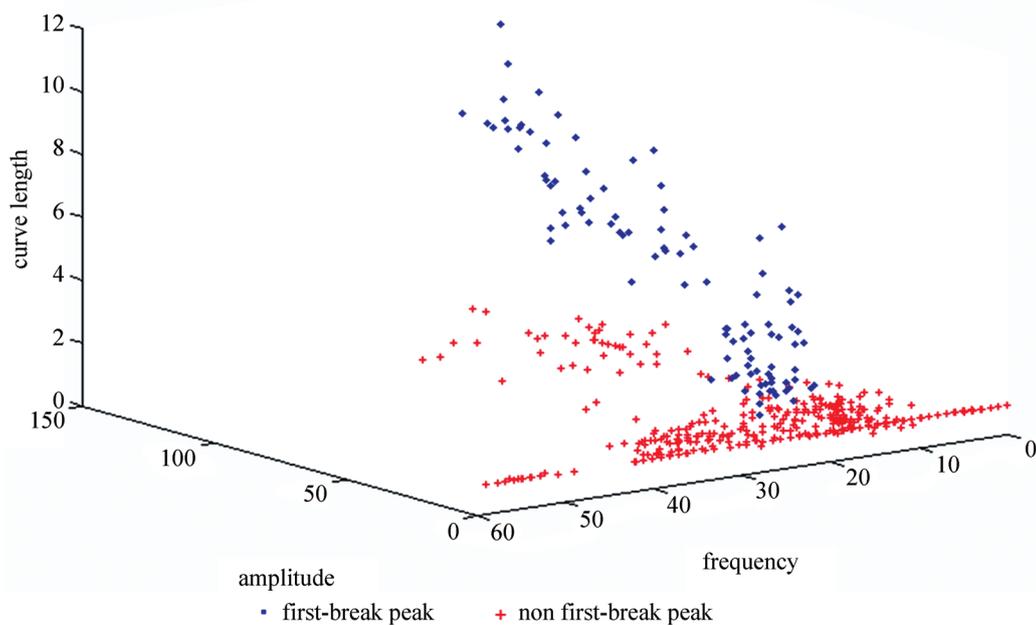


Figure 7
The Intersection Figure of Amplitude and Curve Length Ratio

The combination of instantaneous intensity ratio, amplitude, frequency, curve length ratio and adjacent seismic channel correlation can satisfy the need of neural network first break picking in different topography. The combination has strong stability and still keeps strong recognition ability for first break, even in low SNR seismic data.

3. NEURAL NETWORK'S APPLICATION IN FBK

3.1 The Process of Input and Output Data

Before neural network training, pick first arrival by hand as training samples. In order to get users to pick conveniently by hand and guarantee picking accuracy using neural network, we only investigate first break peak. Therefore, first break samples picked by hand must be first arrival peak time. Due to absorption of stratum

and wavefront diffusion effect, the amplitude difference of near offset and far offset in seismic data is obvious. Seismic data are normalized before extracting attributes in order to eliminate the difference. This contributes to enhancing the accuracy of first break picking. When extracting attributes, extract attributes of first arrival according to first arrival peak time picked by hand, and then obtain first arrival's 4 lockstep no first break peak as training samples of neural network.

Sigmoidal function is used as neural network activation function in this paper. According to its characteristic of intermediate high-gain and low gain at both ends, when data are learning in the region of far from 0, because the convergence speed of the network is very slow in saturation segment, the paralysis phenomenon will even appear. Thus we need to normalize the input samples. Normalized formula is as follows:

$$x' = \frac{x}{x_{\max} + x_{\min}} \quad (10)$$

in this formula, x' is normalized characteristic samples, x_{\max} is the maximum of characteristic samples, x_{\min} is the minimum of characteristic samples. After training samples are normalized, input data vary from 0 to 1. Thus this enhances learning speed of neural network, but the relation between data doesn't reduce.

When we use successfully trained neural network to pick first break, given its work efficiency, it is impossible to investigate each sample. So in order to enhance picking efficiency, the searching scope of first arrival is confined in this paper. The method is that first break peak time picked by hand and its offset are fitted by least-square method to obtain approximate formula of between first arrival peak and its offset, which is used in subsequent

picking process. According to offset information we can get first arrival's estimate value, which is centered to create a searching time window. Neural network searches for first break peak in the time window created above. Likewise we may set a minimum refraction wave's speed value and a maximum one, and calculate the searching scope of first arrival using the two speed and offset.

What neural network picks is first arrival's peak time. For vibrator, first break time is just peak time. But for dynamite source, first break time is jumping time of first arrival^[17]. For dynamite source, the time difference of first arrival's jumping position and peak position is related to first arrival's cycle, generally 3/4 cycle. Thus in order to get first arrival's jumping time for dynamite source, we need to do a time shift for the position of picked first break. We can directly use first arrival's frequency obtained from extracting attributes. Suppose that first arrival's frequency is f , and the shift is calculated by the formula below:

$$\Delta\tau = \frac{1}{4}T = \frac{1}{4f} \quad (11)$$

3.2 The Process of Actual Seismic Data

Now the northwest of China is the emphasis of oil exploration. Northwestern China is mostly covered with mountainous region, desert, gobi, gravel and swamp, and has complex topography. In order to test the algorithm's picking effect, we picked 3 seismic data from different regions in the northwest of China. Figure 8 is first break picking effect of seismic data of a work area in the middle of Junggar basin. As we can see from the figure below, severe noise exists before first arrival, but first arrival is clear and picking effect is very perfect.

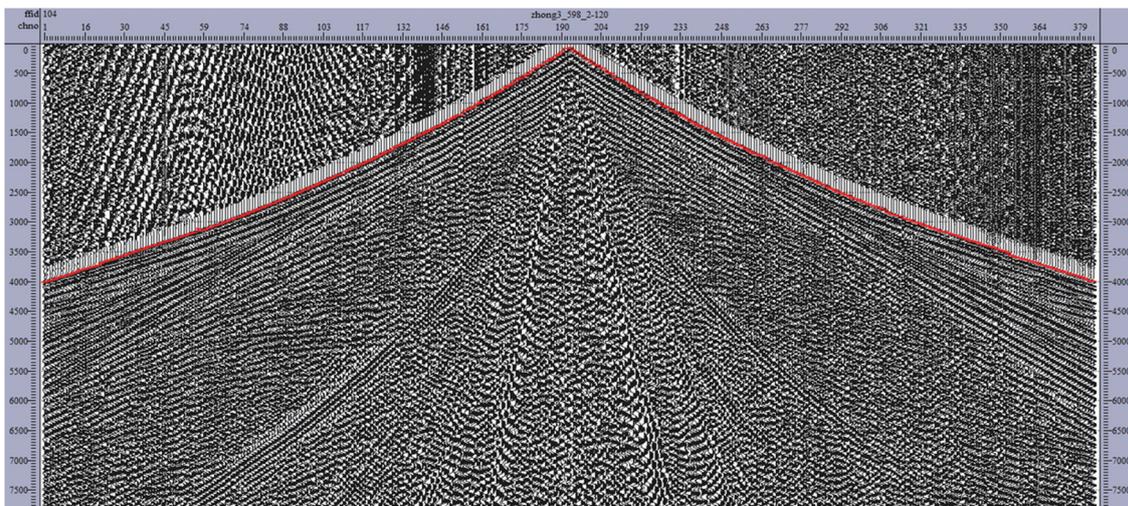


Figure 8 Seismic Data of a Work Area in Junggar Basin First Break Picking Figure

Figure 9 is first break picking effect figure of seismic data of a survey line of a work area in Ningxia Yinchuan basin. The survey line runs through the boundary of

desert and mountain front gravel strip. In Figure 9, the left is seismic data acquired in the desert, and the right is seismic data acquired in the mountain front bent. From the

figure, we can see that refraction wave is very obvious, and its energy is stronger, even exceeding direct wave's one. However, using our method discussed above has still

achieved good picking effect. Figure 10 is part enlarged figure of Figure 9, and is located on the boundary of direct wave and refraction wave. The picking effect is also ideal.

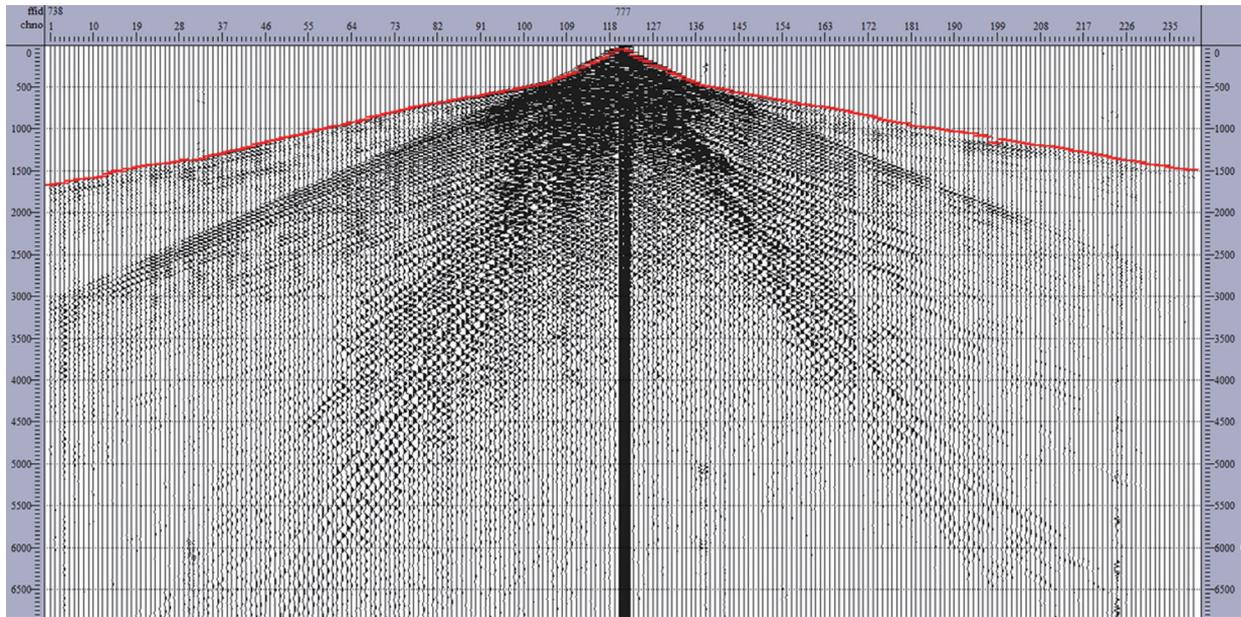


Figure 9
Seismic Data of a Work Area in Ningxia Yinchuan Basin First Break Picking Figure

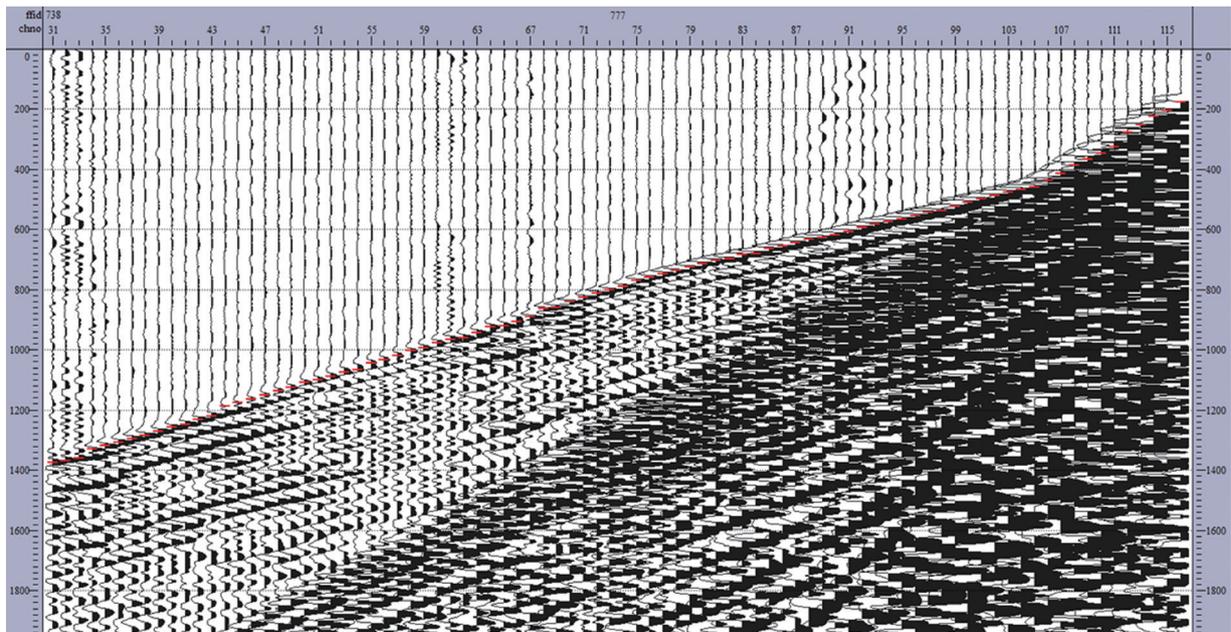


Figure 10
Enlarged Figure

Figure 11 is original seismic data figure of Yanan loess tableland areas. The region's surface undulates severely, and first arrival is undulant before making elevation correction. The result of FBP is not ideal when using other

traditional methods, but using our algorithm discussed above has achieved over 95% accuracy. Figure 12 is part enlarged figure of Figure 11.

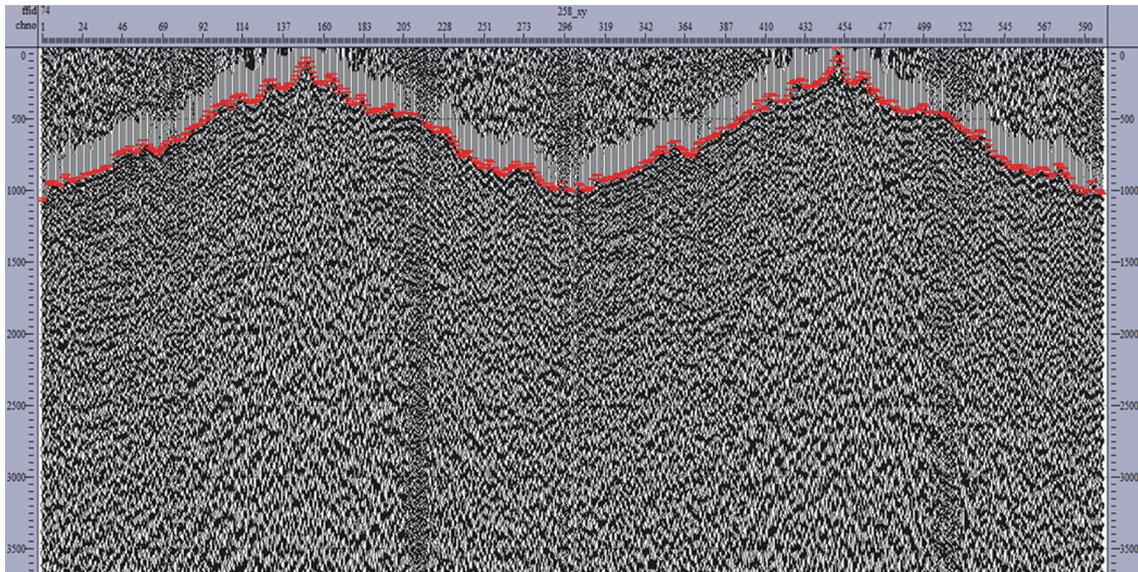


Figure 11
Seismic Data of Yanan Loess Tableland Areas First Break Picking Figure

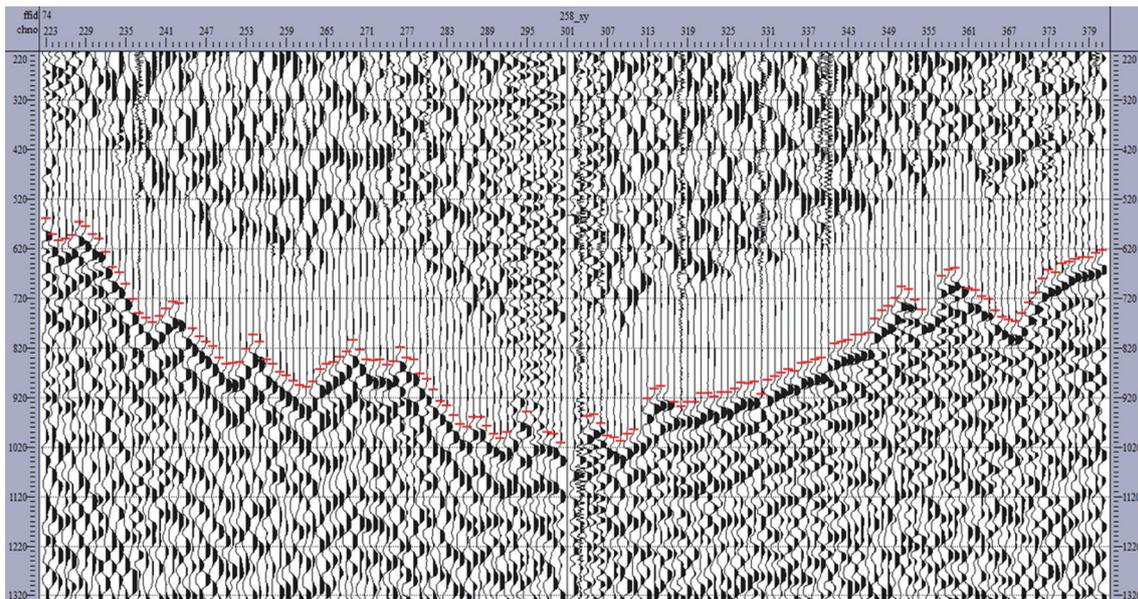


Figure 12
Enlarged Figure

CONCLUSION

The CC algorithm introduced in this paper has several advantages over BP algorithm in the field of first break picking. It has fast convergence speed and can determine own network topology. The more important is that it can expand network size to learn new training samples. The improvement of CC algorithm is also discussed in this paper. Simulation experiments show that the improved one has much faster convergence speed and stronger generalization ability over original one. The selection of attributes is one of the important aspects affecting the accuracy of first break picking. Five attributes are analysed in this paper. Intersection figures show that the combination of these five attributes has good distinction

and stability for first arrival. Our method has achieved good effect in testing actual seismic data.

REFERENCES

- [1] Zuo, G. P., Wang, Y. C., & Sui, R. L. (2004). An Improved Method for First Arrival Picking Using Energy Ratio. *Geophysical Prospecting For Petroleum*, 43(4), 345-347.
- [2] Zhang, W., Wang, Y. C., & Li, H. C. (2009). Seismic First Arrival Pickup Using Instantaneous Intensity Ratio. *Progress in Geophysics*, 24(1), 201-204.
- [3] Niu, P. C., & Zhang, J. Z. (2007). First Arrival Picking Based on Image Processing Methods. *Computer and Modernization*, 140(4), 27-30.

- [4] Cao, M. S., Ren, Q. W., & Wan, L. M. (2004). First Arrival Pickup Using Length Fractal Dimension Algorithm. *Oil Geophysical Prospecting*, 39(5), 509-514.
- [5] Yang, Z. J., & Shi, Z. K. Fast Approach for Cascade-Correlation Learning. *Mathematics in Practice and Theory*.
- [6] Yang, H. Z., Wang, W. N., & Ding, F. (2006). Two Structure Optimization Algorithm for Neural Networks. *Information and Control*, 35(6), 700-705.
- [7] Hart, D. I. (1996). Improving the Reliability of First-Break Picking with Neural Networks. *Annual Meeting, Society of Exploration Geophysicists*.
- [8] Fahlman, S. E., & Lebiere, C. (1990). The Cascade-Correlation Learning Architecture. *Advances in Neural Information Processing Systems*, 524-532.
- [9] Baluja, S., & Fahlman, S. E. (1994). Reducing Network Depth in the Cascade-Correlation Learning Architecture. Retrieved from <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA289352>
- [10] Wu, Q., & Nakayama, K. (1997). Avoiding Weight-IIIgrowth: Cascade Correlation Algorithm with Local Regularization. *Neural Networks*, 1954-1959.
- [11] Fahlman, S. E. (1989). Fast Learning Variations on Back-Propagation: Anempirical Study. In *Proceedings of the 1988 Connectionist Models Summer School*, 38-51.
- [12] McCormack, M. D., Zaucha, D. E., & Dushek, D. W. (1993). First-Break Refraction Event Picking and Seismic Data Trace Editing Using Neural Networks. *Geophysics*, 58(1), 67-78.
- [13] Kusuma, T., & Brown, M. M. (1992). Cascade-Correlation Learning Architecture for First-Break Picking and Automated Trace Editing. *Advance Geophysical*, 11, 11360-1139.
- [14] Gu, H. M., Zhou, H. Q., Zhang, X. Q. (1992). Automatic Pick of First Arrival Time. *Geophysical and Geochemical Exploration*, 16(2), 120-129.
- [15] Pei, Z. L., & Yu, Q. F. (1999). A Wavelet Transform and BP Neural Network-Based Algorithm for Detecting First Arrivals on Seismic Waves. *Investigation Science and Technology*, (4), 61-64.
- [16] Muret, M. E. *et al.* (1992). First Arrival Pickup: A Neural Network. Wei Cao trans. *Geophysical Prospecting*, (40), 587-604.
- [17] Liu, Z. C. (2007). First Break Intelligent Picking Technique. *Geophysical Prospecting For Petroleum*, 46(5), 521-530.
- [18] Wang, J. F., Luo, S. X. (2006). The Improved BP Neural Network and Its Application in Seismic First Break Picking. *Computing Techniques for Geophysical and Geochemical Exploration*, 28(1), 14-17.
- [19] Yin, Y. X., Han, W. G., Li, Z. C. (2006). *New Progress in Seismic Techniques* (1st ed., pp. 58-75). Dongying: Publishing House of China University of Petroleum.