International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 16, Issue 11 (November 2020), PP. 42-48

Aseismic wave velocity selection algorithm based on density peak clustering and similarity function

[1]Yang Zeli,[1]Jiang Murong,[2]Wang Yunsheng,[1]Huang Yaqun,[1]Lu Shaofei
[1]*School of Information Science Engineering, Yunnan University*
[2]*Yunnan aerospace engineering geophysical detection Co. Ltd*
Corresponding Author: Huang Yaqun.

ABSTRACT
In the three-dimensional imaging tunnel geological advanced forecasting instrument, the manual velocity scanning analysis method is often used when selecting the seismic wave velocity. The selected velocity is not unique, the efficiency is low, and the accuracy is low. The paper uses the density peak clustering algorithm to group the velocity spectrum. Then, the least square method is used to linearly fit the amplitude values of the grouped signals with different velocities, and finally, a similar function is used to optimize the amplitude values of the fitted velocities, so that the computer can automatically select the velocity in the seismic wave velocity spectrum. The verification of the model data and the measured data of the Caoba tunnel shows that the algorithm in this paper has achieved good results in terms of prediction accuracy, algorithm stability, and calculation velocity.
Keywords: *seismic wave velocity analysis; velocity spectrum; least squares algorithm; density peak clustering algorithm; similarity function;*

Date of Submission: 24-11-2020 Date of Acceptance: 07-12-2020

I. INTRODUCTION

Seismic wave velocity prediction is one of the most basic problems in Geological Advance Forecast. With the development of geological exploration technology, the requirements for the accuracy of seismic wave velocity prediction are getting higher and higher. In order to accurately and quickly select the seismic wave velocity, researchers have been continuously improving and proposing new methods.

A lot of research has been carried out on seismic wave velocity analysis at home and abroad, for example, Li Fei[1](2017)studied the application of the simulated annealing method in seismic wave travel time inversion of the three-dimensional velocity model, Mao Yuantong[2](2019)studied the method of seismic wave travel time calculation and seismic positioning in three-dimensional anisotropic media, Wang Pu[3](2015) studied the calculation of shear wave velocity based on an improved simulated annealing algorithm, Ren Yiqing[4] (1996) studied the inversion of shear wave velocity using the simulated annealing method, Kou Lan[5](2019) studied the Three-dimensional seismic wave velocity structure analysis method based on gravity anomaly inversion, and Liu Xinxin[6](2010) studied the application of simulated annealing algorithm to achieve seismic pre-stack inversion, Zhou Jie[7](2018) used the finite difference method to simulate the seismic wavefield in the tunnel and Radon transform to separate the wave field,Amin Gholami[8](2020) studied the intelligent prediction model of seismic wave shear wave velocity optimization based on the cuckoo search algorithm,Xu Wenjun[9](2017) proposed a matching pursuit algorithm based on the idea of realizing automatic calculation of superposition velocity,Mohamed Sedek[10](2017)proposed a normal time difference method and the linear interpolation algorithm for predicting the superposition velocity.

The three-dimensional imaging tunnel geological advanced forecaster (AGA-T3) is an instrument that predicts the conditions of the surrounding rock and stratum in front of and around the tunnel during tunnel excavation. It uses manual velocity analysis to select the velocity, there are often problems such as not unique velocity, low efficiency, and low accuracy. For this reason, the paper uses density peak clustering algorithm, least square method, and similar function to analyze the velocity to find the optimal velocity and overcome the shortcomings of manual methods.

II. VELOCITY SPECTRUM GENERATION

The velocity spectrum is the relationship curve of the energy of the seismic wave with respect to the velocity of the wave. There are mainly apparent velocity spectrum and superimposed velocity spectrum. The former seismic wave vibration varies with apparent velocity and the amplitude of the superimposed signal of the latter changes with the superimposed velocity [11](2014). In this paper, we use the latter to generate the velocity spectrum. In the velocity analysis, the velocity can be accurately selected by generating the velocity spectrum, which provides reliable velocity for dynamic correction and horizontal superposition. This velocity can also be used to calculate the layer velocity of underground rock formations, which provides a basis for identifying lithology, determining rock formation changes, and realizing automatic velocity selection [12] (1981).

A sampling of n seismic wave signal, the sending point of the $i$seismic wave signal is$S_i(S_x, S_y, S_z)$, the receiving point is $R_i(R_x, R_y, R_z)$,The scanning point is$C_j(C_x, C_y, C_z)$ (the scanning point is on the central axis), where the x-axis points to the direction of the tunnel face, the y-axis points to the height of the tunnel, and the z-axis points to the width of the tunnel.

The signal value $Y_i(Y_{i0}, Y_{i1}, \cdots, Y_{ip})$ and the signal value time series $X_i(X_{i0}, X_{i1}, \cdots, X_{ip})$ of the sending point S_i and the receiving point R_i are sampled. and p is the number of samples gathering. Sampling frequency f=96μs(0.096ms), velocity $V(V_0, V_1, \cdots, V_m)$, velocity change ΔV=30m/s, depth $H(H_0, H_1, \cdots, H_m)$ and, depth change ΔH= 0.2m.

When any one of $(H_k, V_k)(k = 0,1,2, \cdots, m)$ is input, the velocity spectrum generation algorithm is used to output the corresponding amplitude value $Y_{i,k}(i = 1,2, \cdots, n)$, the specific steps are as follows:

Step 1: Divide the exploration area into grids and scan along the grid points one by one, as shown in Figure 1.

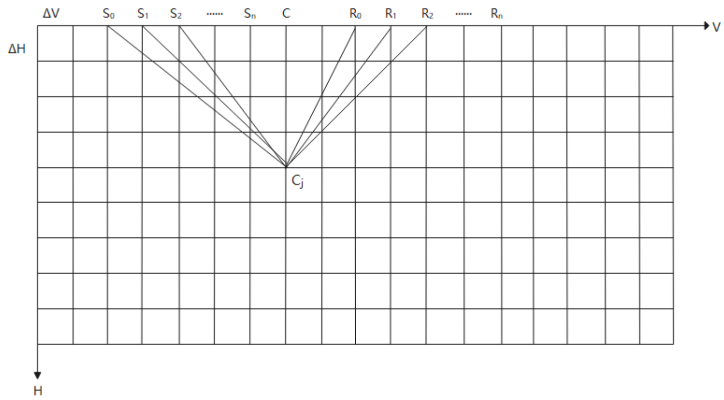

Figure 1: Earthquake wave path map

Step 2:For any scanning point C_j on the central axis, thenthe propagation distance from the channel $i$to C_j is defined in Equation (1) :

$$\begin{aligned} s_{i,j} &= S_iC_j + R_iC_j \\ &= \sqrt{(S_x - C_x)^2 + (S_y - C_y)^2 + (S_z - C_z)^2} + \sqrt{(R_x - C_x)^2 + (R_y - C_y)^2 + (R_z - C_z)^2} \end{aligned} \tag{1}$$

Step 3: Calculate the travel time of the reflected wave at any input velocity V_m which is described by Equation (2)

$$t_{j,m} = \frac{s_{i,j}}{V_m} \tag{2}$$

Step 4: Calculate the actual travel time $\Delta T_{j,m}$, and round down. The formula as follows:

$$\Delta T_{j,m} = \frac{t_{j,m}}{f} \times 1000 \tag{3}$$

Step 5:Extract the amplitude value $Y_{i,\Delta T_{j,m}}$ at the corresponding $\Delta T_{j,m}$ on the signal channel Y_i, obtain the amplitude value of n channels corresponding to any(H_k, V_k)which is defined in Equation (4) :

$$A_\varepsilon = (Y_{0,\Delta T_{j,m}}, Y_{1,\Delta T_{j,m}}, \cdots, Y_{n,\Delta T_{j,m}}) \tag{4}$$

Then add theamplitude values $Y_{i,\Delta T_{j,m}}$ of all the channel on the same scan point C_j to get the total amplitude value$A$as follow:

$$A = \sum_{i=0}^{n} (Y_{i,\Delta T_{j,m}}) \tag{5}$$

Connect the amplitude values $Y_{i,\Delta T_{j,m}}$ on the same scan point C_j with curves to obtain a dynamic calibration curve, as shown in Figure 2. The dotted line represents the dynamic calibration curve.

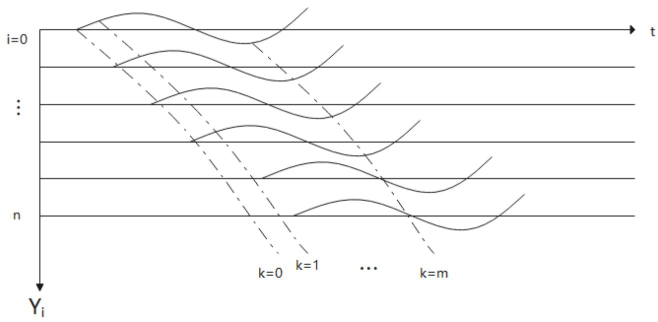

Figure 2: Dynamic calibration curve

Step 6: Take the average of the total amplitude value as

$$\bar{A} = \frac{A}{n} \tag{6}$$

Step 7: The range of $\bar{A} \in [-1,1]$ is divided into four intervals, and each interregional is given a different color to generate velocity spectra of different energy bands, according to the rule shown as follow:

$$\begin{cases} red, (-1 \le \bar{A} < -0.5) \cup (0.15 < \bar{A} \le 1) \\ bule, (0.1 < \bar{A} \le 0.15) \\ yellow, (0.06 < \bar{A} \le 0.1) \\ gray, (-0.5 \le \bar{A} \le 0.06) \end{cases}$$

The red area represents the best energy zone. If the selected velocity falls in the red area, it means that the velocity is the best velocity. At this time, the corresponding dynamic effect positive curve passes through more crests or troughs[12].

III. THEALGORITHM OF THIS PAPER

The Density Peak Clustering Algorithm (DPCA)[13] was proposed by Alex Rodriguez in 2014. The main idea is to find high-density regions separated by low-density regions. It can automatically find the cluster center to achieve efficient clustering of arbitrary shape data. The least-square method finds the best function match of the data by minimizing the square sum of the error. The similarity function is to compare the similarity of two data sets. There are many similar functions, such as Jaccard correlation coefficient, cosine similarity, and Pearson correlation coefficient. This article uses jaccard similarity coefficient, which represents the number of elements in the intersection of two sets A and B. The proportion of numbers in the union of A and B. The larger the jaccard value, the higher the similarity.

The algorithm steps in this paper are shown as follows:

Step 1: Initialization each parameter;

Step 1.1:Input the cutoff distance parameter d_c;

Step 1.2: Input the average velocity, select (H, V) that satisfies the range of $(-1 \le \bar{A} \le -0.5)$ and $(0.15 \le \bar{A} \le 1)$, and is denoted as $(Y_{i,\Delta T_{j,m}}, H, V, A_\varepsilon)$;

Step 1.3: Input class cluster classnum;

Step 2: Cluster (H, V) by density peak clustering algorithm

Step2.1:Calculate the Euclidean distance d_{ru} between any two data points $P_r(H_r, V_r)$ and $P_u(H_u, V_u)$as follow:

$$d_{ru} = \sqrt{(H_r - H_u)^2 + (V_r - V_u)^2} \tag{7}$$

Step2.2:From the cutoff distance, calculate the density ρ_r of any data point P_r, namely

$$\rho_r = \sum_{r \ne u} \varphi(d_{ru} - d_c) \tag{8}$$

Among them, the logical judgment function $\varphi(t) = \begin{cases} 1, & t \le 0 \\ 0, & t > 0 \end{cases}$.

Step2.3:For the data point P_r, among all the data points with greater density than P_r, the smallest distance selected to P_r is recorded as δ_r, namely

$$\delta_r = \min_{u \in \{u|\rho_u > \rho_r\}} d_{ru} \tag{9}$$

Step2.4:Draw a decision diagram with ρ_r as the horizontal axis and δ_r as the vertical axis;

Step2.5:Using the decision graph, mark the point with relatively high ρ_r and δ_r as the cluster center; set ρ_r

relatively low but δ_r relatively high Points marked as noise points;

Step2.6:Assign unmarked points

Assign each label point to its nearest neighbor and the cluster of data points with greater density.

Step 3:Fit each signal value with the least square method

Step 3.1:Linear fit the signal value $Y_{i,\Delta T_{j,m}}$ corresponding to the grouping (H,V) by the least square method

$$y_q = K_q X + b_q \tag{10}$$

Step 3.2:Calculate the shortest distance d_q from the signal value in the n channels to the fitted straight line

$$d_q = \frac{\left|K_q X_{i,j} + b_q - Y_{i,j}\right|}{\sqrt{K_q^2 + 1^2}} \tag{11}$$

Step 3.3:Record the amplitude value corresponding to the shortest distance

$$B = (Y_{0,d_q}, Y_{1,d_q}, \cdots, Y_{n,d_q}) \tag{12}$$

Step 4:Use similarity functions to optimize the selected velocity

Step4.1:Calculate the jaccard similarity coefficient of A_ε and B from the above formulas (4) and (12)

$$J(A_\varepsilon, B) = \frac{|A_\varepsilon \cap B|}{|A_\varepsilon \cup B|} \tag{13}$$

Step 4.2:Take the(H,V)that maximizes the value of $J(A_\varepsilon, B)$,that is the final depth position and predicted velocity.

IV. THEEXAMPLE VERIFICATION

This paper uses the model data and measured data in the tunnel geological advanced forecasting instrument to verify the effectiveness of this algorithm. The model data is a signal collected indoors. Its noise and interference are relatively small and it can intuitively reflect the dynamic calibration curve. The real measured data is the signal collected in the tunnel and there are many measured data. This article takes the Caoba tunnel as an example.

4.1 Model data

Using the velocity spectrum algorithm, the velocity spectrum, and signal diagram of the model signal are obtained, as shown in Figure 3. Using the density peak algorithm, the velocity cluster map of the model data is obtained, as shown in Figure 4. the result of the velocity selection of model data using the algorithm proposed in this paper, as shown in Figure 5.

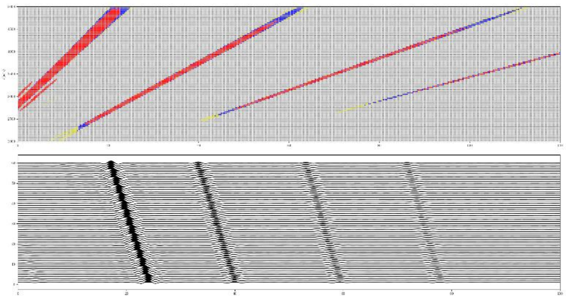

Figure 3: Velocity spectrum and signal diagram of model data

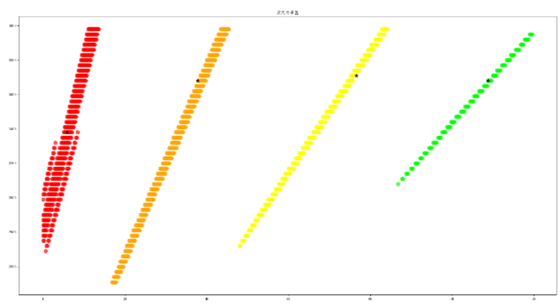

Figure 4:Cluster diagram of model data

In Figure 4, (H, V) is divided into 4 clusters, black dots represent cluster centers, and different colors represent different clusters.

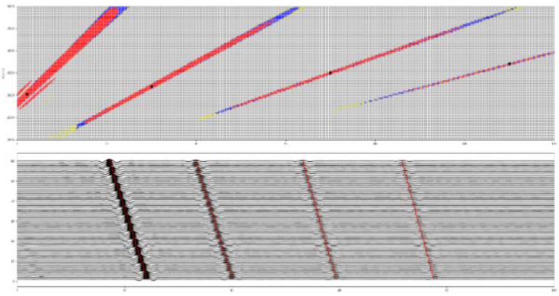

Figure 5:Velocity analysis result diagram of model data

In Figure 5, the upper graph is the velocity spectrum, and the lower graph is the signal. The red area in the velocity spectrum represents the optimal energy band, the black points represent the selected velocity and are all on the red energy band, and the red curve in the signal diagram is the dynamic correction curve.

In a tunnel with an actual length of 120 meters, a width of 12 meters, and a height of 7 meters, a 4-layer layered model interface is established based on the predicted length of the tunnel. The predicted range is 20 meters wide and 30 meters high. The error between the results of the algorithm in this paper and the model data is shown in Table 1.

Table 1:Comparison of theoretical model values and calculated values in this paper

Number of layers	Theoretical depth (m)	Depth of this article (m)	Depth error (%)	Theoretical velocity (m/s)	This article velocity (m/s)	Velocity error (%)
1	37	37.2	0.54	3000	3020	0.66
2	65	65.6	0.92	3200	3230	0.96
3	105	105.8	0.76	3500	3530	0.93
4	145	145.4	0.27	3700	3710	0.26

The results obtained by the algorithm in this paper have relatively small relative errors with the theoretical velocity and theoretical depth, and the velocity value is located within the optimal energy band, and the corresponding dynamic correction curve passes through the peak of each channel, and the optimal velocity value is better selected. The time efficiency of the selected velocity is 10 milliseconds.

4.2 Measured data of Caoba Tunnel

Using the algorithm in this paper, the velocity energy and signal graph of the real measured signal and the velocity clustering graph of the measured signal, and the velocity analysis result graph are obtained, as shown in Figure 6, Figure 7, and Figure 8.

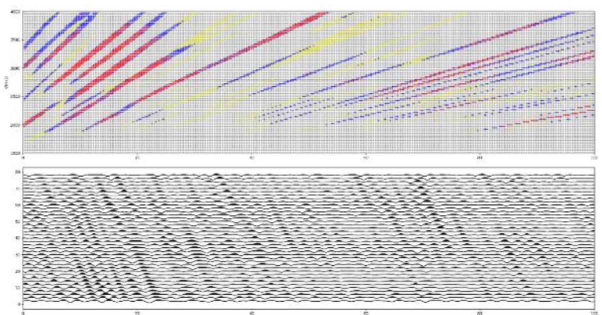

Figure 6: Velocity spectrum and signal diagram of measured data

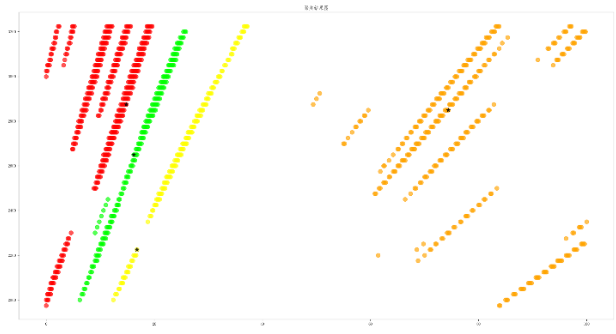

Figure 7: Cluster diagram of measured data

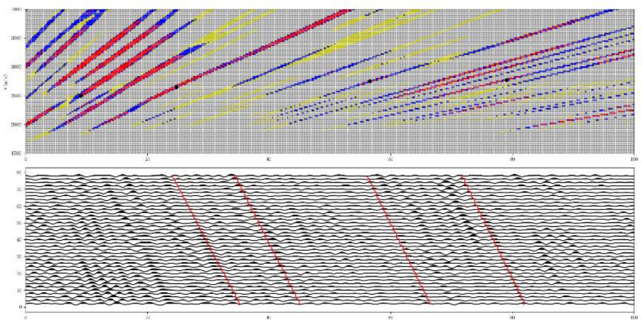

Figure 8: Velocity analysis result graph of measured data

The black point in the velocity spectrum in Fig. 8 represents the selected velocity, and the red curve in the signal diagram is the dynamic correction curve. Fig. 7 is similar to Fig. 4. Because the signal acquisition of the Caoba tunnel contains a lot of noise and interference, the original signal is first subjected to one-dimensional filtering and two-dimensional filtering. The predicted length of the tunnel is 100 meters, the predicted width is 20 meters, and the height is 30 meters. The error between the results of the algorithm in this paper and the measured data is shown in Table 2.

Table 2:Comparison of measured values and calculated values in this paper

Number of layers	Theoretical depth (m)	Depth of this article(m)	Depth error(%)	Theoretical velocity (m/s)	This article velocity(m/s)	Velocity error(%)
1	44.1	45	2.04	2525	2500	0.99
2	59.9	60.8	3.17	2600	2650	1.92
3	95	94.6	0.76	2675	2700	0.18
4	114	115	0.87	2750	2775	0.9

The velocity values automatically selected by the algorithm in this paper are located in the red energy band. and the corresponding dynamic calibration curve has more crests or troughs across each channel, the optimal velocity value can be selected better, and the time efficiency of selecting the velocity is 20 milliseconds.

V. CONCLUSION

This paper uses the density peak clustering algorithm, least squares method, and similar function to select the velocity of the sampled seismic wave signal. By bidirectionally approximating the objective function, the system is optimized as a whole, and the locally optimal solution is found to avoid being in the same energy band. Select multiple velocity values. The verification of theoretical model data and measured seismic wave signals shows that the algorithm in this paper effectively solves the problem of artificially selecting the velocity of seismic wave signals in the three-dimensional imaging tunnel geological advanced forecasting instrument.

The results obtained are close to the theoretical values, and the accuracy and time are efficiencies and other aspects have been improved, which has strong practical value for automatic selection velocity.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (61862062).Research Project on Education and Teaching Reform of Yunnan University(2020Y13).

REFERENCES

[1]. Li Fei,ZhangXuemei,ChenHongfeng,HeShaolin.Application of simulated annealing method in seismic traveltime inversion of 3D velocity model[J].Earthquake Research in China,2017,33(03):355-364

[2]. Mao Yuantong. Study on the method of seismic wave travel time calculation and seismic positioning in three-dimensional anisotropic media[D].Institute of Geophysics,China Earthquake Administration,2019

[3]. Wang Pu,WuGuochen,LiWei.Determination of shear wave velocity based on improved simulated annealing algorithm[J].Fault Block Oil and Gas Field,2015,22(03):330-333

[4]. Ren Yiqing,XuZhongda,MaZaitian.Inversion of shear wave velocity using simulated annealing method[J].Petroleum Geophysical Prospecting,1996(05):677-684

[5]. Kou Lan,ZhangJin.Research on 3D seismic wave velocity structure analysis method based on gravity anomaly inversion[J].Chinese Journal of Earthquake Engineering, 2019, 41(05):1259-1265

[6]. Liu Xinxin, Fang Guangjian. Using simulated annealing algorithm to achieve seismic pre-stack inversion[J]. Lithologic Reservoirs, 2010, 22(01): 87-92

[7]. Zhou Jie.Research on finite difference numerical simulation of tunnel earthquake advance prediction and Radon transform wavefield separation[D].Chongqing University, 2018

[8]. Amin Gholami,Seyed Mohsen Seyedali,Hamid Reza Ansari. Estimation of shear wave velocity from post-stack seismic data through committee machine with cuckoo search optimized intelligence models[J]. Journal of Petroleum Science and Engineering,2020,189

[9]. Xu Wen- Jun,Yin Jun- Feng,WANG Hua- Zhong.AUTOMATIC ESTIMATION OF STACKING VELOCITY BASED ON SPARSE INVERSION.2017,60(6):640-650

[10]. Sedek, M., & Gross, L. Normal move-out correction in anisotropic and laterally heterogeneous media using simultaneous velocity variation with offset. Journal of Natural Gas Science and Engineering, 45, 399–414. doi:10.1016/j.jngse.2017.05.014

[11]. Wang Junheng, Li Shuling, Yao Changli. Introduction to Solid Geophysics: Geological Publishing House, 2014

[12]. MouYongguang.Digital processing method of seismic exploration data [M].Beijing: Petroleum Industry Press 1981

[13]. Rodriguez A, Laio A. Clustering by fast search and find of density peaks [J]. Science, 2014, 344(6191): 1492-1496.

Huang Yaqun., et. al. "Aseismic wave velocity selection algorithm based on density peak clustering and similarity function." *International Journal of Engineering Research And Development*, vol. 16(11), 2020, pp 42-48.