Computers and Geosciences 110 (2018) 81–89

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Intelligent inversion method for pre-stack seismic big data based on MapReduce

Xuesong Yan [a], Zhixin Zhu [a], Qinghua Wu [b,*]

[a] School of Computer Science, China University of Geosciences, Hubei, Wuhan, 430074, China
[b] Faculty of Computer Science and Engineering, WuHan Institute of Technology, Hubei, Wuhan, 430205, China

ARTICLE INFO

2016 MSC:
00-01
99-00

Keywords:
Intelligent optimisation algorithm
Pre-stack seismic data
Elastic parameter inversion
MapReduce

ABSTRACT

Seismic exploration is a method of oil exploration that uses seismic information; that is, according to the inversion of seismic information, the useful information of the reservoir parameters can be obtained to carry out exploration effectively. Pre-stack data are characterised by a large amount of data, abundant information, and so on, and according to its inversion, the abundant information of the reservoir parameters can be obtained. Owing to the large amount of pre-stack seismic data, existing single-machine environments have not been able to meet the computational needs of the huge amount of data; thus, the development of a method with a high efficiency and the speed to solve the inversion problem of pre-stack seismic data is urgently needed. The optimisation of the elastic parameters by using a genetic algorithm easily falls into a local optimum, which results in a non-obvious inversion effect, especially for the optimisation effect of the density. Therefore, an intelligent optimisation algorithm is proposed in this paper and used for the elastic parameter inversion of pre-stack seismic data. This algorithm improves the population initialisation strategy by using the Gardner formula and the genetic operation of the algorithm, and the improved algorithm obtains better inversion results when carrying out a model test with logging data. All of the elastic parameters obtained by inversion and the logging curve of theoretical model are fitted well, which effectively improves the inversion precision of the density. This algorithm was implemented with a MapReduce model to solve the seismic big data inversion problem. The experimental results show that the parallel model can effectively reduce the running time of the algorithm.

1. Introduction

Seismic exploration is a type of method for oil exploration based on seismic information. Because earthquake information can reflect the variations in the trends of the reservoir parameters, the reservoir parameters can be predicted by establishing a correspondence between the seismic attributes and the reservoir parameters obtained from logging. With the rapid development of seismic technology, the prediction of the reservoir characteristic parameters by using seismic attributes will become an inevitable trend in future oil and gas exploration and development. Moreover, people hope to extract more reliable seismic attributes from seismic data in order to further reduce the uncertainty in a seismic attribute analysis. The pre-stack seismic data have some characteristics such as a large amount of data, abundant information, and so on; thus, the abundant information of the reservoir parameters can be obtained through the inversion of pre-stack seismic data. Furthermore, the pre-stack inversion method has some obvious advantages such as stable results, a high resolution, and a strong controllability. Therefore, in recent years, inversion based on pre-stack seismic data has always been a popular topic in the field of seismic exploration.

The pre-stack seismic data contain a large amount of useful information for predicting the conditions of underground oil and gas (Shaopeng, 2009), including the three elastic parameters, namely, the P-wave velocity V_p, S-wave velocity V_s, and density ρ. The three elastic parameters can reflect the saturation conditions of the underground oil and gas from the side. Moreover, the relationship between the P-wave velocity V_p and the gas saturation is non-linear, and the relationship between the density ρ and the gas saturation is linear. Further, the S-wave velocity can reflect some of the characteristics of rock. Therefore, it is necessary to collect information related to the variations in the three elastic parameters when determining the oil and gas saturation. In the inversion of the elastic parameters of the pre-stack seismic data, it is necessary to determine the elastic parameters. Then, the goal of matching the corresponding parameters of the actual terrain will be achieved by

* Corresponding author.
E-mail address: wuqinghua@sina.com (Q. Wu).

https://doi.org/10.1016/j.cageo.2017.10.002
Received 28 April 2017; Received in revised form 15 August 2017; Accepted 6 October 2017
Available online 10 October 2017
0098-3004/

continuously adjusting the three parameters. It can be seen that the parameter inversion of the pre-stack seismic data is an optimisation process, and an intelligent optimisation algorithm can be used to solve this problem. It is necessary to construct a suitable objective function for the parameter inversion of seismic data; then, this objective function is optimised to obtain the optimal solution. However, this objective function is often non-linear. At the beginning of development, experts and scholars solved this problem by using a linear or quasi-linear method for the objective function. Although the linear inversion method is widely used and forms a system, it has a strong dependence on the initial model and other defects. If the initial model is incorrect, it will lead to unreliable inversion results. When using intelligent optimisation algorithms for the above problem, existing single-machine environments have been unable to meet the calculation demands required for the huge amount of data because of the large amount of pre-stack seismic data. Thus, the development of a method with a high speed and efficiency is urgently needed to solve the problem of parameter inversion of pre-stack seismic big data.

In the middle 1980s, experts and scholars in the field of geophysics began to be concerned with non-linear global intelligent optimisation inversion technology. Rothman was the first to use the simulated annealing method to optimise the automatic residual static correction problem (Rothman, 1986). Berg was the first to theoretically analyse the possibility of using a genetic algorithm for multi-parameter inversion (Berg, 1990). Mallick used a genetic algorithm to solve the Amplitude Variation with Offset (AVO) inversion problem (Mallick, 1995). Yang Wencai carried out a systematic study of the application of a genetic algorithm for solving the problem of geophysics (Wencai, 1995). Misra et al. applied the fast simulated annealing algorithm to the pre-stack AVO inversion problem based on boundary protection, which can obtain a better retrieval accuracy (Misra and Sacchi, 2008). Lu Pengfei et al. improved the simulated annealing algorithm and carried out inversion with some important parameters, which also effectively improved the retrieval accuracy (Lu et al., 2008). According to the defects of the particle swarm optimisation algorithm, the algorithm was improved by Zhu Tong. The convergence speed of the improved algorithm was faster, and it can be used to effectively solve the inversion problem of the elastic parameters (Tong and Xiaofan, 2011). In the past 30 years, this series of non-linear global intelligent optimisation inversion techniques has been widely used for various inversion problems, and many significant research results have been obtained. In the inversion process, an intelligent algorithm has many shortcomings; thus, experts and scholars have attempted to combine intelligent algorithms with other algorithms to realise hybrid optimisation inversion in order to improve the retrieval accuracy. Porsani et al. were the first to mix the genetic algorithm and linear inversion methods together and applied it to the seismic waveform inversion problem (Porsani et al., 1993). Then, Priezzhev et al. combined a neural network with a genetic algorithm to research the non-linear inversion of multiple seismic data and optimised the learning process of the neural network by using a genetic algorithm (Priezzhev et al., 2008). Pantelis Soupios et al. proposed and applied a mixed genetic algorithm to different geophysical problems and combined a genetic algorithm with a local optimisation method to improve the shortcomings of various algorithms. The experimental results showed that this method can effectively utilise the advantages of various algorithms to improve the accuracy of the results and to ensure the calculation efficiency (Soupios et al., 2011). Bai Junyu et al. combined a genetic algorithm with the conjugate gradient method to solve the problem of seismic wave impedance inversion, which leveraged the local optimisation capability of the conjugate gradient method and the global optimisation capability of the genetic algorithm. The inversion results obtained by their theoretical model and the real data were both very good (Junyu et al., 2014).

Although an intelligent algorithm is one of the main methods for solving problems in the field of geophysical inversion, these algorithms face some difficulties in the non-linear inversion of geophysics. First, although these intelligent algorithms have been widely used in various fields and have been demonstrated to be effective, they have their own shortcomings. For example, although the simulated annealing algorithm has a robust initial value, universal easy implementation, and so on, the annealing temperature setting has considerable influence on the algorithm. If the temperature is not set properly, it may cause algorithm failure. A genetic algorithm is good at global searching, but it is easy trapped into local optimal and pre-maturity. The ant colony algorithm adopts a positive feedback mechanism and has a strong ability to find a solution, but it is complicated. The advantage of the particle swarm algorithm is the high convergence speed, but the overall performance of the algorithm is affected by a lower accuracy, easy divergence, and other shortcomings.

Second, the computational efficiency of these intelligent algorithms is generally low. When using them to carry out the study of non-linear inversion problems, the important parameters are solved according to the optimisation of an objective function. In the process of searching for a solution, there are always some problems such as a low computational efficiency. For example, for the ant colony algorithm, the number of pheromones will affect the efficiency of the algorithm. Further, in the early stages of the algorithm, if a pheromone is deficient, the inversion speed will be reduced. For a genetic algorithm in the late stages, the poor local search efficiency will cause a reduction in the search efficiency and an increase in the time consumed. The search efficiency of the algorithm is closely related to the efficiency of solving the inversion problem; that is, the search efficiency is low, and the efficiency of solving the problem must be also low. Seismic inversion technology has continuously been developed, which makes the research trend of inversion transform from post-stack to pre-stack. The amount of the data used in pre-stack inversion is huge, leading to rapid growth in the number of calculations during the inversion process, which greatly increases the amount of data processing by the intelligent algorithm in the non-linear inversion of geophysics.

The ideal way to solve the non-linear inversion problem is the stochastic optimisation method (Yin Cheng, 2001), but because this method is based on the probability statistics and often requires a long computation time, it is difficult to achieve. However, this method has no dependence on the initial model; thus, it can be used to solve the non-linear inversion problem. With the development of computer technology, especially the emergence of parallel machines, it provides the basic conditions for the use of non-linear stochastic optimisation methods that require enormous amounts of computing resources. In recent years, the emergence of cloud computing (Misra and Sacchi, 2008; Dean and Ghemawat, 2008) technology has received more attention, which provides the possibility for large-scale computing. MapReduce can be combined with a number of commonly used intelligent algorithms, especially with an algorithm that is similar to a genetic algorithm with its own parallelism. According to the previous studies, it is known that a genetic algorithm usually can find a satisfactory solution for problems within a reasonable period of time. However, with the increase in the complexity and scale of the problem to be solved, the computational time will be greatly increased, and the operation efficiency will be also reduced. With the extensive application of genetic algorithms, methods for increasing the efficiency of the genetic algorithm are urgently required. Because of the inherent parallelism of a genetic algorithm, parallel processing of a genetic algorithm can be determined by the inherent characteristics of the algorithm and the problem's demands for the algorithm. Parallel processing can improve the solution speed of the genetic algorithm and expand the population size to maintain the diversity of the population and prevent premature convergence, thus improving the quality of the solution of the problem.

The basic steps for solving the pre-stack seismic data parameter inversion problem are as follows: first, calculate and obtain the reflection coefficient using an approximate equation; then, convolute the reflection coefficient with a seismic wavelet to obtain the seismic data; and finally, compare the seismic data with the real seismic data. If both data sets are more fitting and the three parameters also fit the three parameters from log data, it indicates that the inversion precision is high. In this paper, we convert the pre-stack seismic parameter inversion problem into an optimisation problem and then use an intelligent optimisation algorithm to solve it.

2. Intelligent parameter inversion problem

2.1. Basic procedures of parameter inversion

The establishment of a convolution model of inversion is one of the main procedures for the inversion of the pre-stack elastic parameters, and the main steps for establishing the convolution model are as follows.

The first step is to calculate the reflection coefficient R_{pp}. R_{pp} is calculated by using the approximate equation by Aki & Rechard (Aid and Richards), as shown in formula (1).

$$R_{pp}(\theta)=\frac{1}{2}\left(1+\tan^2\theta\right)\frac{\Delta V_p}{\overline{V_p}}-\left(4\gamma^2\sin^2\theta\right)\frac{\Delta V_s}{\overline{V_s}}+\frac{1}{2}\left(1-4\gamma^2\sin^2\theta\right)\frac{\Delta\rho}{\overline{\rho}} \tag{1}$$

ΔV_p, ΔV_s, and $\Delta\rho$ respectively represent the differences in V_p and V_s between the upper and lower two layers and the difference in ρ. $\overline{V_p}$, $\overline{V_s}$, and $\overline{\rho}$ respectively represent the means of V_p and V_s in the upper and lower layers and the mean of ρ. θ is the angle, and $\gamma=\frac{V_s}{V_p}$, which is calculated using real data. R_{pp} can be calculated according to this formula and is used as a component of the convolution operation of seismic records.

The second step is to obtain a seismic wavelet. A seismic wavelet is another component of the convolution model of seismic records, and the data of seismic records are obtained according to the convolution calculation of the wavelet and reflection coefficient, which is suitable for the establishment of a forward model and the creation of a synthetic seismogram. A Ricker wavelet was used in this study, which is a type of zero-phase seismic wavelet, and it is expressed in formula (2) (Kallweit and Wood, 1982).

$$f(t)=\left(1-2\pi^2V_mt^2\right)e^{-\pi^2V_mt^2} \tag{2}$$

V_m is the dominant frequency, and t is time, which can be set manually.

The third step is to carry out the convolution operation with the reflection coefficient and Ricker wavelet, as shown in formula (3) (Liu and Liu, 2003).

$$s(\theta)=R_{pp}\times f(t)+n(t) \tag{3}$$

R_{pp} represents the reflection coefficient function, $f(t)$ represents the seismic wavelet, and $n(t)$ represents the noise. Noise factors were not considered in this study. The calculated $s(\theta)$ is used to construct the objective function.

2.2. Objective function

Because an individual would be evaluated by the algorithm according to the fitness function converted from the objective function, the quality of the constructed objective function is the main factor that affects the inversion effect of the pre-stack elastic parameters. In this study, the Aki & Rechard approximate equation was first used to calculate the value of R_{pp}, i.e. the reflection coefficient of the reflected P-wave. Then, R_{pp} and a wavelet were utilised during convolution to obtain the synthetic seismogram data. The number of sampling points was n, and each sampling point required m different angles to calculate $n\times m$ seismic record data points. Finally, the m sets of seismic record data points obtained by the optimisation of each sampling point were subtracted from the actual seismic record data and then squared. This result was divided by m after the cumulative sum; then, the data obtained from the n sampling points were divided by n after accumulation. After the final results were squared, that is what is required. According to the above formula, the inversion objective function can be expressed as formula (4).

$$f(x)=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(s(\theta_{i,j})-s'(\theta_{i,j})\right)^2}{n\times m}} \tag{4}$$

$s(\theta_{i,j})$ is the seismic record of forward, and $s'(\theta_{i,j})$ is the seismic record of inversion.

2.3. Inversion framework of the intelligent algorithm

When using an intelligent optimisation algorithm to solve the pre-stack seismic data parameter inversion problem, its framework can be represented as shown in Fig. 1.

First, the ranges of the three elastic parameters are determined according to the existing empirical information, and individual initialisation is carried out. Then, the reflection coefficient is calculated according to the Aki & Rechard approximate equation. The difference between the

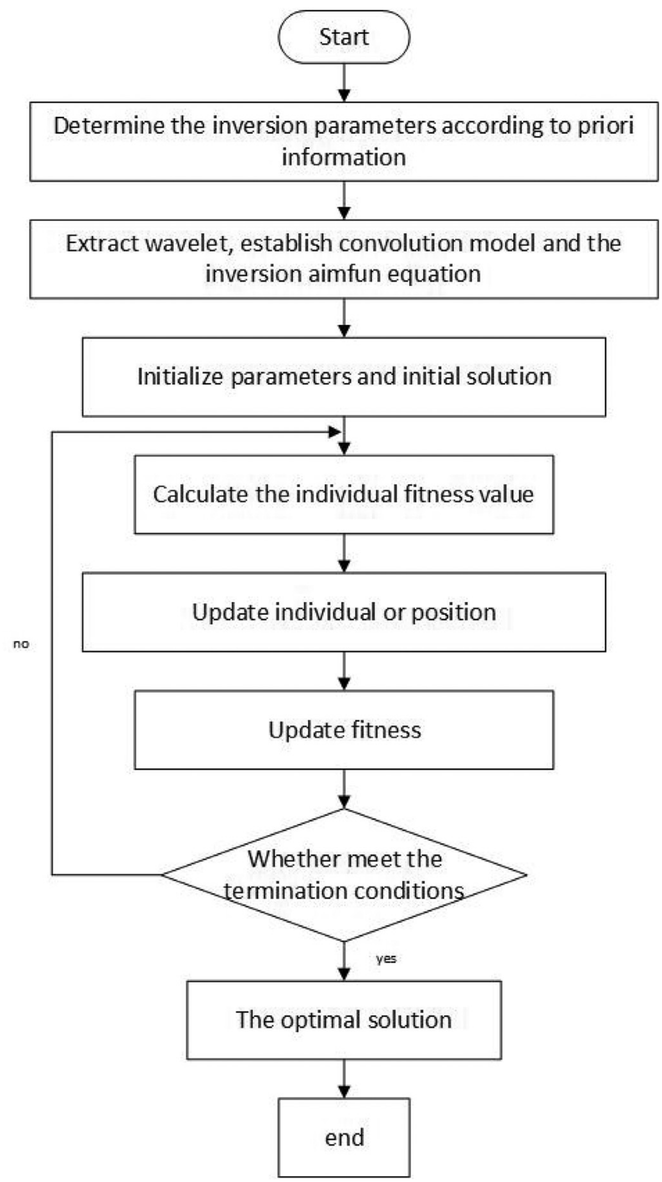

Fig. 1. Flow chart of intelligent parameter inversion.

seismic recording data and the results obtained from the convolution of the above reflection coefficient and wavelet is calculated. Then, the individual screening is carried out according the above difference. The fitness value is updated according to a series of optimisation operations, and the optimal result is output when the terminal condition of the algorithm is reached. In here, the terminal condition include two conditions: one is the algorithm has attained the optimal solution and other is the algorithm has reached the maximum iterations.

3. Intelligent parameter inversion algorithm based on MapReduce

3.1. Individual representation and initialisation

In this paper, an individual in the algorithm was composed of inversion parameters. For the actual logging curve model, each sampling point served as a layer, i.e. it was one-dimensional. Because the inversion of the three parameters was studied, the individual length was three times that of the sampling point. Assuming that there are n sampling points, the number of parameters in the model to be solved is $3n$; thus, the corresponding individual encoding is

$$G_i = \left(V_{p1}, V_{s1}, \rho_1, \cdots, V_{pj}, V_{sj}, \rho_j, \cdots, V_{pn}, V_{sn}, \rho_n\right)$$

In this study, traditional real-number coding was used to design an individual (chromosome) in the population, which was initialised according to the method of random initialisation within a certain range. Each chromosome was composed of a set of real numbers. The population size was assumed to be N, among which V_{pj}, V_{sj}, and ρ_j represented the values of the three parameters corresponding to the $j-th$ sampling point of the individual G_i. In addition, the range of variation was set according to the actual logging data. The structures of each chromosome and the population space are shown in Fig. 2.

It was found through an analysis of the logging data that although logging can provide the density ρ, P-wave velocity V_p, and S-wave velocity V_s of rock, inversion is generally carried out by establishing a simple relationship among the three parameters in practical applications. For different rock properties, Gardner used a statistical method to establish the relationship between the rock density and the P-wave velocity, as shown in formula (5) (Gardner et al., 1974).

$$\log(\rho) = a \bullet \log\left(V_p\right) + b \tag{5}$$

With the help of this relation, the corresponding logarithmic forms of $\log(\rho)$ and $\log(V_p)$ are generated according to the data of ρ and V_p in the logging curve model. Then, the parameters a and b are obtained from these data according to the least-squares method, as shown in formula (6) (Lawson and Hanson, 1995).

$$\begin{gathered} a = \frac{n \times sum\left(\log\left(V_p\right) \times \log(\rho)\right) - sum\left(\log\left(V_p\right)\right) \times sum(\log(\rho))}{n \times sum\left(\log\left(V_p\right)^2\right) - sum\left(\log\left(V_p\right)\right)^2} \\ b = \frac{sum(\log(\rho))}{n} - a \times \frac{sum\left(\log\left(V_p\right)\right)}{n} \end{gathered} \tag{6}$$

Hard constraints on the P-wave velocity and density are established according to the above formula. The P-wave velocity V_p and S-wave velocity V_s at each sampling point are still constrained by the bounding range, whereas the density ρ is determined according to V_p determined by the bounding function and its corresponding constraint relation. This formula can be used to better reflect the relationship between the two, and the obtained density is closer to the real data. The constraint for the bounding range is shown in formula (7).

$$\begin{aligned} &0.8 \bullet V_{pwell} \le V_p \le 1.2 \bullet V_{pwell} \\ &0.8 \bullet V_{swell} \le V_s \le 1.2 \bullet V_{swell} \\ &0.9 \bullet \rho_{well} \le \rho \le 1.1 \bullet \rho_{well} \end{aligned} \tag{7}$$

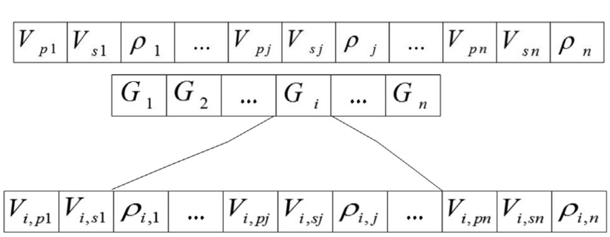

Fig. 2. Structure of a chromosome and the population space.

3.2. Crossover operator

Because real-number encoding was used in this study, the search capability will be weak if the single-point crossover method is simply used. Therefore, a crossover operator that is more suitable for real-number encoding was adopted in this study with the arithmetic crossover strategy. This strategy adopted the mode of a linear combination of two individuals to generate two new individuals, where λ is a number taken from a random distribution in the [0, 1] interval. The expressions are shown in formula(8).

$$\begin{aligned} &child1 = \lambda \times parent2 + (1-\lambda) \times parent1 \\ &child2 = \lambda \times parent1 + (1-\lambda) \times parent2 \end{aligned} \tag{8}$$

Suppose that λ is 0.3 and that a gene in a parent individual is crossed over. Then, the values of the corresponding genes in the progeny individuals are shown in Fig. 3.

3.3. Fitness function

The final aim of the algorithm is to obtain a set of ideal solutions for the elastic parameters, which will result in the minimum differences among the elastic parameter optimised by the intelligent algorithm, the seismic record obtained from forward inversion, and the actual seismic record. In this study, the fitness function was transformed from the objective function, and the objective function was for the minimum value to match the selection operator. Therefore, the fitness function was designed as the objective function, as shown in formula(9).

$$fit(f(x)) = f(x) \tag{9}$$

3.4. MapReduce realisation of the algorithm

In this study, MapReduce was used to realise pre-stack elastic parameter inversion based on a genetic algorithm. Multi-copy replication with the input data was carried out, where the number of iterations of the serial genetic algorithm is m, and the number of individuals is k. Thus, the fitness value is calculated mk times. If the number of copies is n, it is equivalent to n sub-populations that carry out the genetic algorithm at the same time, where the number of individuals in each sub-population is k/n, the number of iterations is m, and the fitness value is still calculated

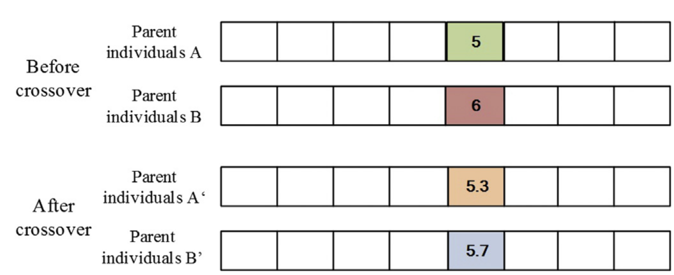

Fig. 3. Arithmetic crossover.

mk times. In this way, the fitness value of an individual can be calculated in parallel, and the total running time can be reduced effectively.

First, the input data are duplicated to obtain multiple copies, and the number of Mappers are determined according to the number of copies. Each Mapper is created with a sub-population, and each sub-population executes a complete genetic algorithm. The fitness values of all of the individuals in a sub-population are first calculated with the Map function to select the optimal individual according to the fitness value. The optimal individual is preserved as offspring according to the selection strategy. Then, crossover, mutation, and the ergodic sub-group are carried out. The algorithm is iterated until the termination condition is reached, and the optimal individual in this sub-population is output for transfer to the Reducer. The Map function is used to implement multiple parallel tasks, and each Mapper performs the same task.

A Mapper reads the number of individuals in a population and the number of iterations from the main function. Then, the actual logging data and the data of the seismic record are read. The genetic algorithm designed above is executed through the Map $<String, Object>$ mapping to obtain a set of optimal values of V_p, V_s, and ρ and the corresponding fitness values.

The pseudo-code of the Map function is as follows:

Map part. Mapper (Key, Value)
Input: Data of logging curve; data of seismic record
Output: Fitness value of optimal individual; optimal individual

1. For i: = 0 to iteration-1 do
2. For j: = 0 to popsize-1 do
3. Evaluation of fitness:fit(j) = evaluate (population [j]);
4. Selection:population [j] = select (population [j]);
5. Crossover:population [j] = crossover (population [j]);
6. Mutation:population [j] = mutate (population [j]);
7. End for
8. End for
9. Return optimal fitness value
10. Return optimal individual

The Reducer is responsible for collecting the output results of multiple Mapper tasks, and further operation is then carried out. The input type of the Reducer must match the output type of the Mapper. The number of Mappers is set as duplication, and the number of Reducers is set as 1, which will summarise all of the results obtained from the Map process, i.e. a summary of the individual with the optimal fitness value obtained by the Mapper. According to a comparison of the fitness value, the individual with the optimal fitness value is output, namely, the final optimal solution.

The pseudo-code of the Reduce function is as follows:

Reduce part. Reducer (Key, Value)
Input: Fitness value of the optimal individual; optimal individual
Output: Fitness value of the global optimal individual; global optimal individual

1. *minFitness* = *MAXVALUE*
2. For i: = 0 to duplication do
3. If individual [i]. fitness < minFitness
4. MinFitness = individual [i]. fitness;
5. FinalIndividual = individual [i];
6. End for
7. Return optimal fitness value
8. Return optimal individual

MapReduce parallel technology is used to improve the solution efficiency. The first thing to consider is how to distribute or cut the data to be processed. Then, a suitable solution is found according to the problem and the characteristics of the MapReduce framework, to. The input data of the problem have a certain relevance; therefore, cutting is not considered. Instead, the data are copied and distributed. The MapReduce architecture is composed of a Mapper and Reducer, and the Mapper is mainly responsible for task processing, whereas the Reducer is responsible for summarising the results. Therefore, the specific steps are designed as follows:

Step 1: Duplicate the input data (including the data of the logging curve and the data of the seismic record) to obtain multiple copies. The input Key is the offset address of the input file, and the value is the text data read in;
Step 2: Distribute every copy to each Mapper;
Step 3: Each Mapper receives one copy of the input data. Establish the sub-population on this basis and finish a complete operation of the genetic algorithm to optimise the pre-stack seismic data parameter;
Step 4: Each Mapper outputs the individual with the optimal fitness value in its sub-population, namely, the best set of parameter values of V_p, V_s, and ρ. Because there is only one Reducer in the design, the value of the key output by the Mapper is set as the integer 0 so that all of the data output by the Mapper are transmitted to the same Reducer. The value refers to the individual with the optimal fitness value and the fitness value of this individual;
Step 5: The Reducer collects the optimal individual of each Mapper. Compare the fitness values of all of the optimal individuals to select an individual with best fitness value and output this individual and its fitness value. The designed parallel framework is shown in Fig. 4.

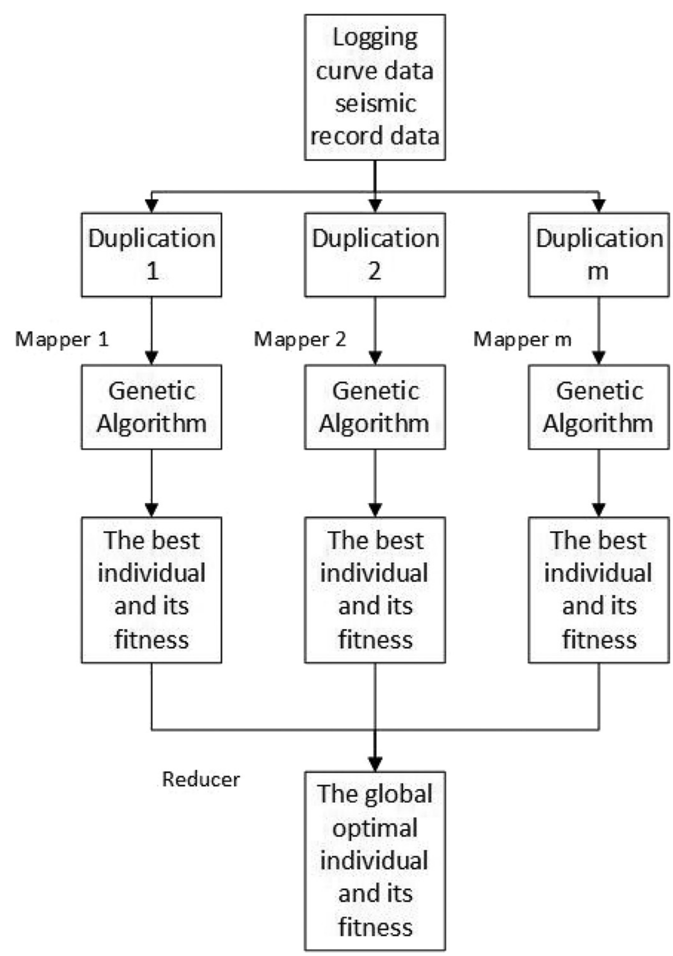

Fig. 4. Framework of the algorithm based on MapReduce.

4. Experimental results and analysis

4.1. Experimental environment

The operating system used the experiment to test the proposed algorithm is Ubuntu 12.10. In order to facilitate a comparison of the results obtained in a single-machine experimental environment with those obtained in a parallel experimental environment, the same configuration is used for both environments. The JDK version is 1.7.0.45. The cluster configuration is summarised in Table 1.

4.2. Experimental data

The experimental data utilised in this study contain log data and seismic record data, and two different sizes of data sets were used for the model trial. The log data in data set 1 comprise the data of 241 sampling points, including the P-wave velocity V_p, shear-wave velocity V_s, and density ρ. Each sampling point corresponds to eight different angles: 0°, 6°, 11°, 17°, 23°, 29°, 34°, and 40°. Each of the data sets uses these eight angles, and the Aki & Rechard formula is used to create a forward model of the logistic model of the log. The logistic model of the log is used to calculate the reflection coefficient, and the reflection coefficient is convoluted with the wavelet. The relationship between the upper and lower groups of sampling points needs to be used to form seismic records; therefore, the seismic record data contain 240 × 8 data points. Data set 2 contains the data of 24,001 sampling points, corresponding to 24,000 × 8 seismic record data point. Data set 1 is used to compare the optimised elastic parameter results of the basic genetic algorithm (GA), basic particle swarm optimisation algorithm (PSO), and the proposed intelligent algorithm. Data set 2 is used to compare the results of the serial algorithm and the algorithm based on MapReduce. The resulting logging data are shown in Fig. 5 and Fig. 6.

On the basis of the above logging data, a synthetic seismic record is established. The wavelet is zero-phase Ticker wavelet with a frequency of 50 Hz, a wavelet length of 50 ms, and a sampling interval of 1 ms. The resulting seismic record data are shown in Fig. 7.

4.3. Experimental results and analysis

To more clearly observe the performance of the algorithms for the experimental sets, the termination condition of the algorithm is 5000 iterations, the population size is 20 individuals, the crossover probability is set to 0.7, the mutation probability is set to 0.05, and an experiment is repeated 20 times. When the number of individuals is 20 and the number of iterations is 5,000, the values of a and b are 0.12980109381164917 and −0.1476920424395448, respectively, according to formula (6). The inverted log generated by the proposed intelligent algorithm is compared with the model log, as shown in Fig. 8, and the inverted synthetic seismic records are shown in Fig. 9. The inverted elastic parameters of the basic GA are shown in Fig. 10, and the inverted elastic parameters of the basic PSO are shown in Fig. 11.

The logistic model of the log uses the algorithm presented in this paper for the inversion of the elastic parameters. By comparing the three inversion results of the elastic parameters, the inversion results of the P-wave velocity V_p and shear-wave velocity V_s already fit the theoretical model log. The density ρ has a large improvement, but the effect is less than that of the P-wave velocity and shear-wave velocity. There is a certain error. In general, the inversion effect of the three parameters have greatly improved. We can see that our proposed intelligent algorithm has a better inversion effect compared with the basic GA and basic PSO.

Table 1
Cluster configuration parameters.

Compute nodes	5
Processor	8.0 GHz
Memory	8 GB
Operating system	Ubuntu 12.10
Hadoop	Hadoop 2.4.0
JDK	1.7.0.45

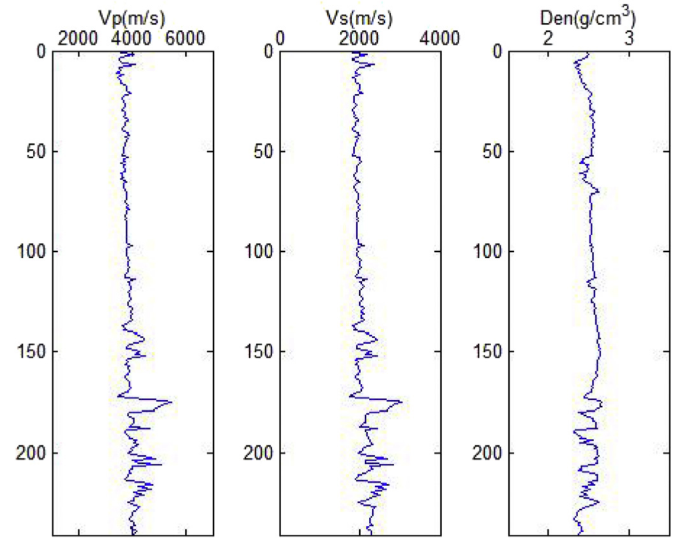

Fig. 5. Logging data chart of data set 1.

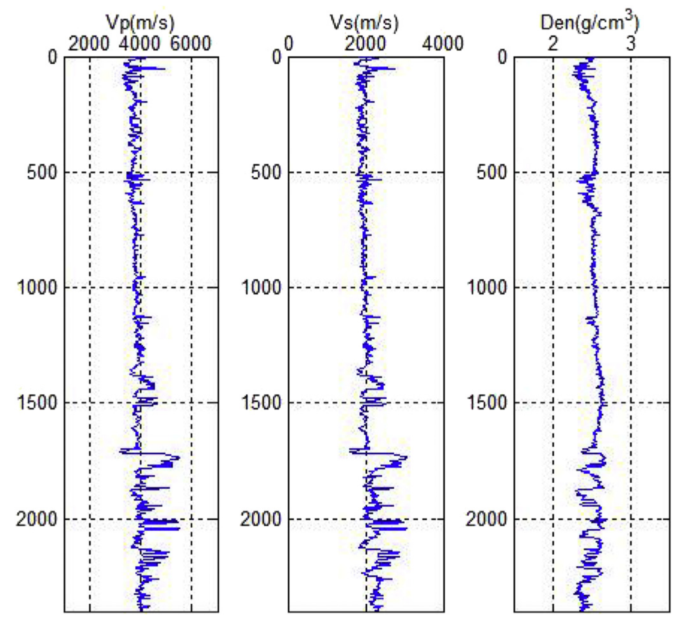

Fig. 6. Logging data chart of data set 2.

One-twentieth of the sampling point data is taken from the data set to compare the three parameters. These results are listed in Table 2. It can be seen from the table that the effect of optimisation by the proposed intelligent algorithm for the three elastic parameters is better than those of the basic GA and basic PSO, and the results are close to the actual logging data. Through the inversion model trial, it is found that the intelligent optimisation algorithm proposed in this paper has a better inversion effect than the genetic algorithm. The difference between the three algorithms is that the proposed hybrid intelligent optimisation algorithm adopts new population initialisation, selection, crossover, and mutation strategies. We set the crossover probability to 0.7 and the mutation probability to 0.05.

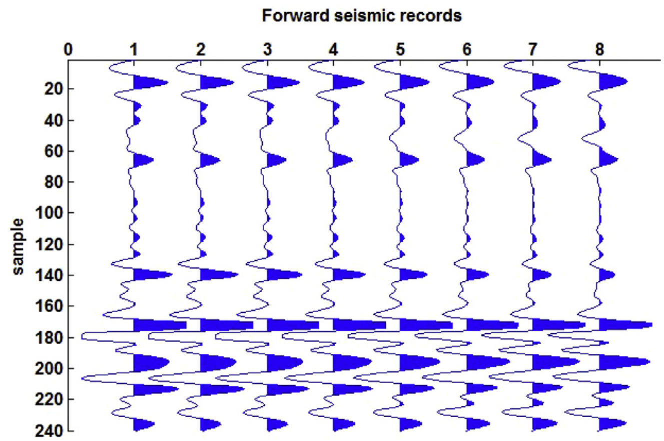

Fig. 7. Theoretical model for the synthetic seismic records.

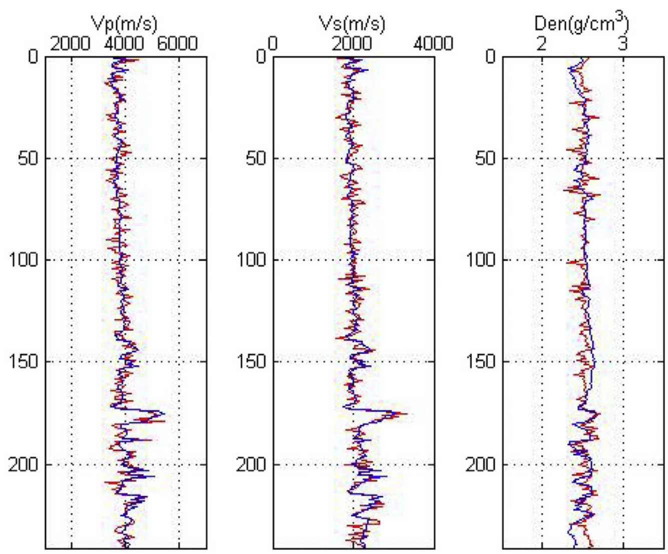

Fig. 8. Comparison of the Inverted Log and Model Log Using the Proposed Intelligent algorithm (5000 Iterations).

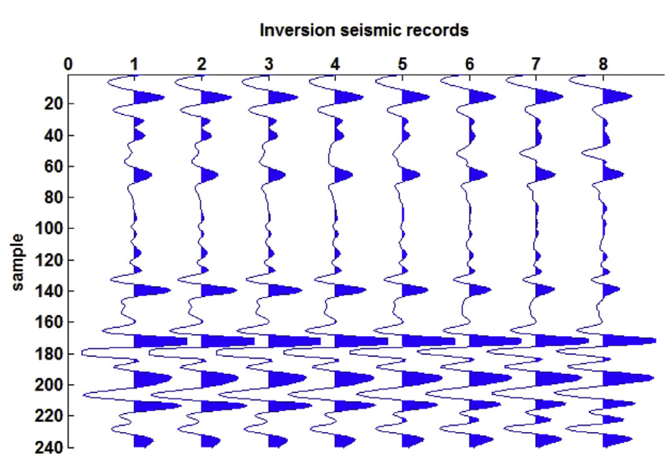

Fig. 9. Inverted synthetic seismic records using the proposed intelligent algorithm.

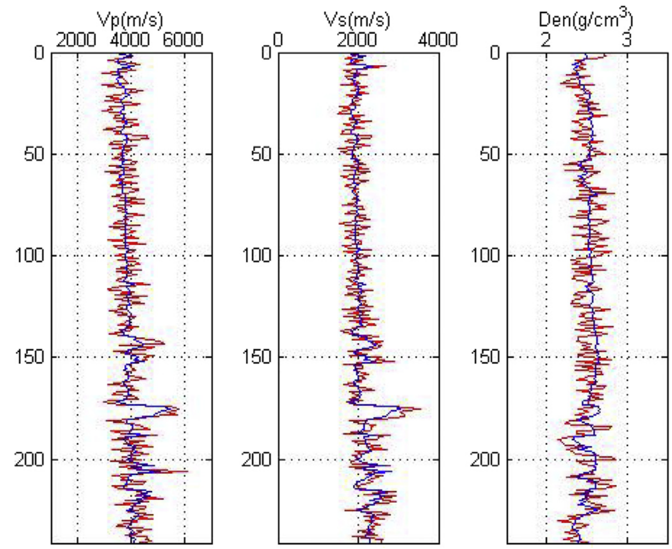

Fig. 10. Comparison of inverted log and model log using the basic GA (5000 iterations).

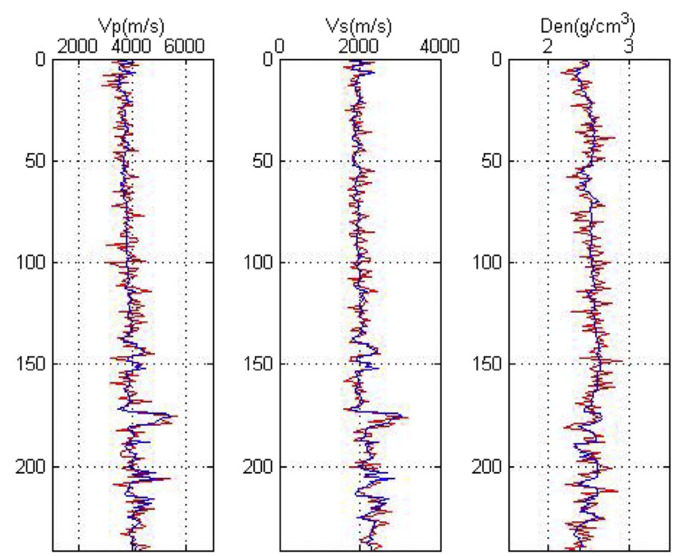

Fig. 11. Comparison of the inverted log and model log using the basic PSO (5000 iterations).

A comparison of the three optimisation algorithms regarding the convergence for inversion of the elastic parameters of the pre-stack seismic data is shown in Fig. 12. It can be seen from the analysis shown in Fig. 12 that the convergence rates of the proposed intelligent algorithm and basic PSO in the early stages are faster, the objective function value rapidly declines, and the genetic algorithm converges slowly. When the algorithms enter the middle and late stages, the genetic algorithm is prematurely convergent, and the function value remains at about 0.02. The objective function value of the basic PSO remains at 0.008. Although the intelligent algorithm proposed in this paper is still carrying out a global search, the overall effect is better than that the genetic algorithm, the objective function value remains at about 0.005, and the error is reduced by an order of magnitude. It can be seen that in the early stages, the rate of convergence is increased owing to the use of a population initialisation strategy that is more suitable for the problem, which minimises the difference between ρ obtained from the intelligent algorithm and ρ obtained from the log data, and the adjustment of the

Table 2
Comparison of the inversion results and model log results using three algorithms.

	V_p				V_s				ρ			
	GA	PSO	Our	well	GA	PSO	Our	well	GA	PSO	Our	well
1	2894.63	3719.41	3529.83	3534.34	2049.81	1824.23	1703.22	1818.18	2.1868	2.2498	2.4876	2.4876
2	4094.86	3805.20	3473.10	3512.97	1580.10	1940.79	2024.67	1963.53	2.5650	2.3571	2.4652	2.3532
3	3866.21	3389.56	3548.09	3556.64	2129.88	1971.75	1787.21	1898.77	2.6257	2.3721	2.5099	2.4055
4	3323.15	3959.01	3749.53	3745.83	2061.21	1934.23	1948.49	1984.91	2.6095	2.3003	2.6002	2.4647
5	4158.59	3732.96	3744.54	3691.67	2145.51	1643.64	1960.77	1867.41	2.5336	2.4542	2.5179	2.5266
6	3200.91	3481.93	3688.33	3720.09	2034.82	2255.41	1884.47	1913.77	2.4124	2.5005	2.5297	2.5125
7	3765.68	3674.14	3662.99	3718.24	1510.21	2015.19	1866.86	1887.69	2.6237	2.5021	2.5088	2.5237
8	4387.15	4009.68	3722.84	3741.65	2093.42	1822.86	2109.54	1900.32	2.3203	2.4131	2.5187	2.5571
9	3261.08	3602.76	3654.89	3694.35	1570.61	1924.03	1917.73	1863.21	2.4899	2.5092	2.5292	2.5734
10	3744.94	3596.55	3678.60	3671.08	2036.97	1602.56	2149.43	1984.94	2.6550	2.3375	2.5081	2.4276
11	3659.02	3791.08	3768.45	3756.23	2305.18	2348.04	2226.55	2046.00	2.2229	2.4454	2.4951	2.4313
12	3515.94	3993.62	3646.81	3651.18	1545.08	1847.81	1944.24	1896.27	2.5401	2.3445	2.5610	2.4609

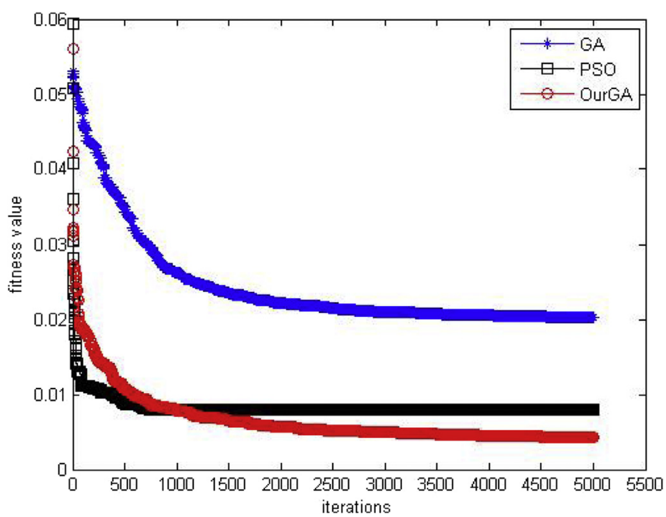

Fig. 12. Convergence curves of the three algorithms.

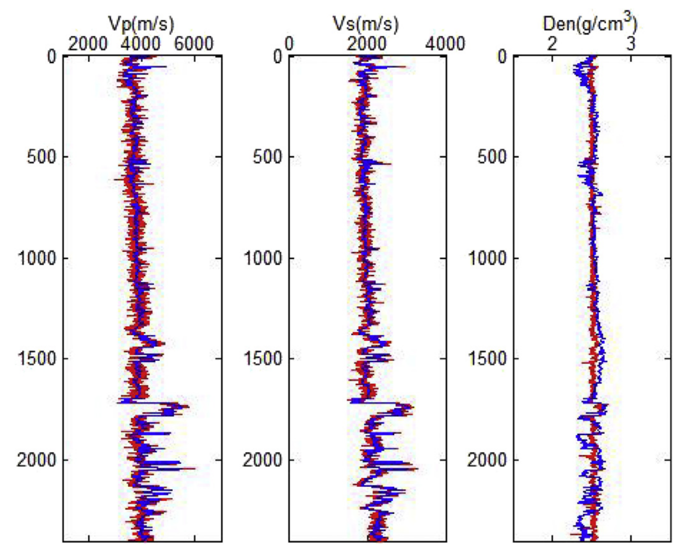

Fig. 13. Comparison of the inverted log and model log in a single-machine environment (500 iterations).

selection strategy. Therefore, relative to the genetic algorithm, the intelligent algorithm proposed in this paper improves the population initialisation, selection, crossover, and mutation strategies, which makes the algorithm more suitable for solving the problem and effectively improves the inversion precision.

We experimented with the algorithm based on MapReduce and used data set 2 as the experimental data. The parallel strategy replicates multiple copies of the experimental data. In order to ensure the same number of fitness calculations when the experimental parameters are set, the single-machine experimental parameters are set to 400 individuals iterated 50, 100, 250, and 500 times. The parameters of the parallel experiment are as follows: 20 individuals per population, 20 copies, and 50, 100, 250, and 500 iterations. The number of data nodes is set to 2, 3, 4, and 5. The experiments in two different environments are run 20 times, and the experimental results were statistically analysed. The mean value and the performance of the algorithm were evaluated. Taking 500 iterations as an example, the log's P-wave velocity, shear-wave velocity, and density results obtained through the intelligent inversion of the elastic parameters of the pre-stack seismic data in the single-machine environment are shown in Fig. 13, and the results of the log's P-wave velocity V_p, shear-wave velocity V_s, and density ρ obtained using the intelligent parameter inversion algorithm based on MapReduce are shown in Fig. 14. As can be seen from the figure, the inverted V_p V_s, and ρ fit well with the theoretical model log, but there are some errors.

In this experiment, the experimental results are shown in Fig. 15 and Fig. 15 for different numbers of data nodes and iterations. It can be seen

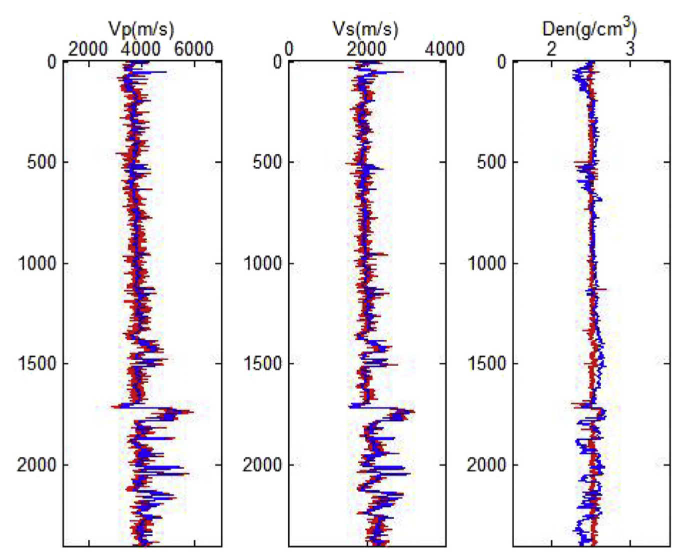

Fig. 14. Comparison of the inverted log and model log based on MapReduce (500 iterations).

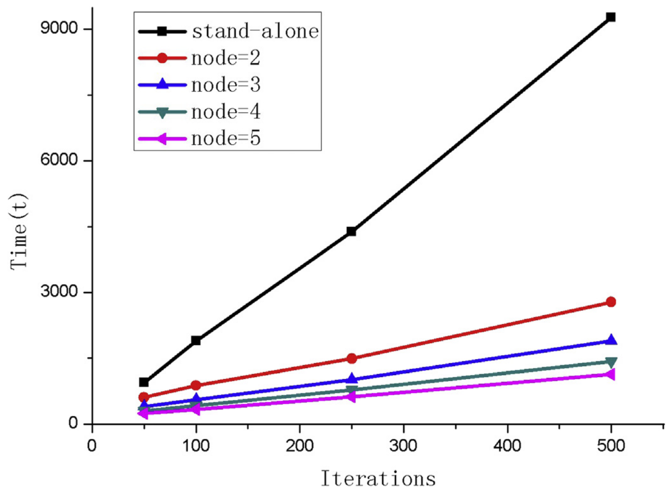

Fig. 15. Comparison of the running time.

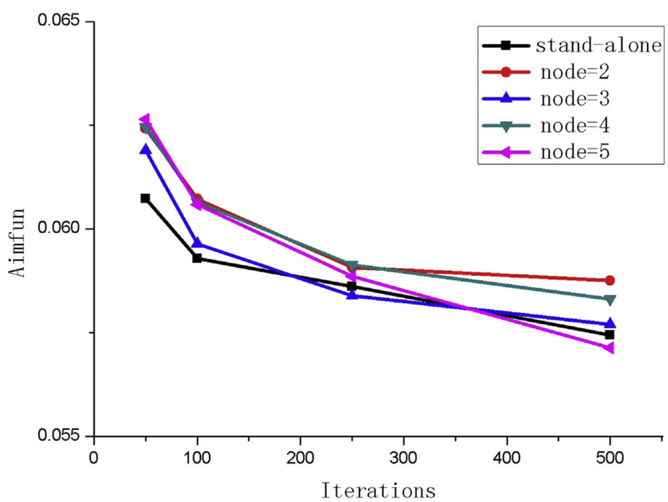

Fig. 16. Comparison of the objective function values.

from Fig. 15 that the running time of the algorithm increases as the number of iterations increases in both the single-machine and MapReduce environments. In the distributed environment, the running time of the algorithm is significantly lower than that of the single-machine environment, and the computation time of the algorithm decreases as the number of computing nodes increases. By applying multiple copies, we can calculate the fitness of an individual population in parallel, guaranteeing both the same fitness value and calculation time while effectively reducing the running time of algorithm, thereby improving the efficiency of the algorithm. Thus, the use of parallel technology can effectively reduce the running time of the algorithm. At the same time, we can see from Fig. 16 that the objective function value results obtained in the single-machine and MapReduce environments do not significantly differ, even better than that in the single-machine environment. It can be seen that the parallel strategy can ensure both the total number of individuals and the fitness value calculation times. Further it also has experimental results similar to or even better than those of the single-machine environment while reducing the execution time of the algorithm. In conclusion, this method can effectively solve the inversion of pre-stack seismic data.

5. Conclusion

In this paper, an intelligent algorithm that is more suitable for solving the parametric inversion of pre-stack seismic data is proposed. The efficiency of the algorithm is improved by improving the population initialisation, selection, crossover, and mutation strategies. Through a trial calculation of the inversion of the logistic model of the log, the effects of parametric inversion by the proposed intelligent algorithm are evaluated. From the experimental results, the proposed intelligent algorithm is more efficient in inversion than the genetic algorithm, and the degree of fitting of the parameters is also improved. A parallel strategy based on MapReduce is used to improve the parametric inversion of pre-stack seismic big data and proposed to solve the non-linear inversion problem according to the problem characteristics. The use of multiple copies of the parallel strategy can guarantee the same calculation times for a fitness value as those of a single-machine environment while reducing the running time of the algorithm and improving the efficiency of the algorithm.

Acknowledgment

This paper is supported by the Natural Science Foundation of China (No. 41404076 and 61673354) and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan).

References

Aid, K., Richards, P., Quantitative Seismology: Theory and Methods, San Francisco.
Berg, E., 1990. Simple Convergent Genetic Algorithm for Inversion of Multiparameter Data, pp. 1126–1128. https://doi.org/10.1190/1.1889928.
Dean, J., Ghemawat, S., 2008. Mapreduce: simplified data processing on large clusters. Commun. ACM 51 (1), 107–113.
Gardner, G., Gardner, L., Gregory, A., 1974. Formation velocity and density the diagnostic basics for stratigraphic traps. Geophysics 39 (6), 770–780.
Junyu, B., Zilong, X., Yunfei, X., Tianshou, X., 2014. Nonlinear hybrid optimization algorithm for seismic impedance inversion. In: Beijing 2014 International Geophysical Conference & Exposition, Beijing, China, 21-24 April 2014. Society of Exploration Geophysicists and Chinese Petroleum Society, pp. 541–544.
Kallweit, R., Wood, L., 1982. The limits of resolution of zero-phase wavelets. Geophysics 47 (7), 1035–1046.
Lawson, C.L., Hanson, R.J., 1995. Solving Least Squares Problems. SIAM.
Liu, X., Liu, H., 2003. Survey on seismic blind deconvolution. Prog. Geophys. 18 (2), 203–209.
Lu, P.-F., Yang, C.-C., Guo, A.-H., Wang, Z.-l., Liu, P., 2008. Modified simulated annealing algorithm and its application in pre-stack inversion of reservoir parameters. Prog. Geophys. 1, 014.
Mallick, S., 1995. Model-based inversion of amplitude-variations-with-offset data using a genetic algorithm. Geophysics 60 (4), 939–954. https://doi.org/10.1190/1.1443860.
Misra, S., Sacchi, M.D., 2008. Global optimization with model-space preconditioning: application to avo inversion. Geophysics 73 (5), R71–R82. https://doi.org/10.1190/1.2958008.
Porsani, M.J., Stoffa, P.L., Sen, M.K., Chunduru, R., Wood, W.T., 1993. A Combined Genetic and Linear Inversion Algorithm for Seismic Waveform Inversion, pp. 692–695.
Priezzhev, I., Shmaryan, L., Bejarano, G., 2008. Nonlinear multitrace seismic inversion using neural network and genetic algorithm. In: 3rd EAGE St. Petersburg International Conference and Exhibition on Geosciences-geosciences: from New Ideas to New Discoveries.
Rothman, D.H., 1986. Automatic estimation of large residual statics corrections. Geophysics 51 (2), 332–346. https://doi.org/10.1190/1.1442092.
Shaopeng, L., 2009. Research on Avo Seismic Parameter Inversion Method and its Application. Master's thesis.
Soupios, P., Akca, I., Mpogiatzis, P., Basokur, A.T., Papazachos, C., 2011. Applications of hybrid genetic algorithms in seismic tomography. J. Appl. Geophys. 75 (3), 479–489.
Tong, Zhu, Xiaofan, Li, 2011. Inversion of seismic scalar wave equation based on improved particle swarm optimization algorithm. Chin. J. Geophys. 54 (11), 2951–2959.
Wencai, Y., 1995. Genetic algorithm for geophysical inversion. Geophys. Prospect. Petrol. 34 (1), 116–122.
Yin Cheng, X.G., 2001. Seismic inversion and nonlinear stochastic optimization. Comput. Tech. Geophys. Geochem. Explor. 23 (1), 6–10.