Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
e-ISSN: 2395-3454

Vol. 18 | No. 5 (Special Issue) | pp. 1719-1725 | 2019

Original Research Paper | Open Access

Research on Coal Mine Gas Concentration Prediction Based on Cloud Computing Technology Under the Background of Internet

Lei Wang, Wei Li and Yan Li
Hebei Institute of Communications, Shijiazhuang Hebei, 051430, China

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 31-08-2019
Accepted: 27-10-2019

Key Words:
Coal mine
Gas concentration
Elma neural network
Genetic algorithm
Cloud computing

ABSTRACT

With the continuous expansion of the scale of gas concentration data, in order to meet the requirements of mass data processing, this paper used the strong advantages of cloud computing in the processing of large data sets to build the framework of coal mine gas concentration under the cloud platform, proposed a genetic optimization Elma neural network model based on cloud computing, and carried out experiments based on the massive data of a coal mine in Tangshan. It has been proved that its mean square error is basically stable within 0.05, reaching the acceptable error range. This algorithm is both efficient and feasible in short-term prediction of coal mine gas concentration.

INTRODUCTION

Coal has long been China's main energy source and will continue to occupy an important position in the long term. In China, most of the coal production is underground mining, the mine environment is complex and changeable, and the threat of natural disasters such as water, gas, fire, geothermal heat and dust has caused frequent gas accidents. Gas disasters not only seriously endanger the lives of employees, but also have a very bad impact on coal production (Yuan et al. 2018, Qiang et al. 2018). As a key indicator for evaluating coal mine safety production, real-time and accurate prediction of gas concentration can provide a strong guarantee for coal mine production.

With the rise of the Internet of Things and the rise of the mobile internet, coal mining companies have already experienced data explosions. Traditional algorithms are extremely inferior in the processing of massive data and will continue to increase with the amount of computation, and eventually collapse due to insufficient resources. Cloud computing, represented by the Map reduce distributed platform, can simultaneously use multiple processors for parallel computing and distributed processing, with huge storage space, ultra-large processing speed and many other advantages. In the cloud computing mode, the data acquisition layer continuously collects various gas-related monitoring data from the downhole, and the amount of data in the intermediate process of the prediction system can be up to 1PB per day. In this paper, cloud computing is applied to the prediction of gas concentration. The Elman neural network is used to establish the gas prediction model. Since it is difficult to obtain the global optimal solution by using the gradient descent method for network training, it is proposed to optimize the nerve with genetic algorithm (GA). The weights and thresholds of the network, and based on cloud computing big data processing techniques, speed up network convergence and generate thresholds. Based on the massive data of a mining area in Tangshan City, the experimental results show that the algorithm is efficient and feasible in the short-term prediction of massive gas concentration data, and the prediction accuracy is guaranteed.

EARLIER STUDIES

Over the years, the theory and methods of gas concentration prediction have emerged. Zhao et al. (2016) have fully studied the influence of gas concentration on coal safety accidents and the principle of gas sensors, and used self-correction technology to improve the accuracy of gas sensor measurement data. After fully studying the working environment of coal mine equipment. A study established a wireless sensor network management system for coal mine on-site using a wireless sensor network with simpler structure and higher transmission efficiency, which improved the reliability of coal mine equipment (Lin et al. 2015, Temel et al. 2018). A study built a real-time flow data processing framework spark streaming to construct a real-time prediction system for gas concentration based on flow regression (Che et al. 2017).

MATERIALS AND METHODS

Cloud-based Gas Concentration Prediction Analysis Framework

In this paper, the master/slave architecture is used to predict the gas concentration, and the parallel prediction algorithm is used to predict the massive data. The specific analysis framework is shown in Fig. 1.

The cloud computing master server accepts real-time data of gas concentration collected by each sensor in the data source, gas concentration manually collected underground, and historical gas concentration, and pre-processes all data, and combines valuable data to form a data model, data management. The subsystem allocates gas concentration data to the corresponding cloud from the server's data storage module according to the data model, and sets up a copy from the server in other clouds to record the storage location and the copy location. According to actual needs, one or more algorithms suitable for the data are selected in the prediction algorithm model library, and the tasks are decomposed and distributed to the slave servers. After the slave servers receive the tasks, the storage and execution modules cooperate with each other to achieve high efficiency. Accurately complete the gas concentration prediction analysis task (He et al. 2016).

Cloud-based Elman Neural Netwoek Model

Elman neural network structure and algorithm: Elman artificial neural network is a dynamic recurrent neural network. It contains special dynamic feedback links and recursive functions. It can not only solve the modelling problem of static networks, but also realize the mapping of nonlinear dynamic systems (Lanzerstorfer 2018). The network consists of an input layer, a hidden layer, a connection layer, and an output layer. The connection layer can accept the feedback signal from the hidden layer, and delays and stores, and self-links to the hidden layer, which increases the dynamic processing information capability of the network. Its state space expression is

$$\begin{cases} y(k) = g(w^3 x(k) \\ x(k) = f(w^1 x(k) + w^2 (u(k-1) \\ x_c(k) = x(k-1) \end{cases} \tag{1}$$

Where, u is the input layer vector; y is the output layer vector; x is the hidden layer vector; x_c connect the layer vector; w^1 a weight vector for the connection layer to the hidden layer;w^2 a weight vector for the input layer to the hidden layer; w^3 the weight vector from the hidden layer to the output layer. G0 is the excitation function of the output layer, generally a linear function; f0 is a hidden layer excitation function, often taken as the sigmoid function (Yong et al. 2018, Tröstl et al. 2016).

Elman uses the gradient descent algorithm to train the error function E between the predicted output value and the expected output value to derive the derivative value of each layer, and correct the weight of each layer of the feedforward connection part along the negative gradient direction. Correct the input layer and hidden layer, the weight between the hidden layer and the output layer ω. The formula is as follows:

$$\Delta\omega = -\eta \frac{\partial E}{\partial \omega} \qquad \ldots(2)$$

In the formula, η is the learning speed.

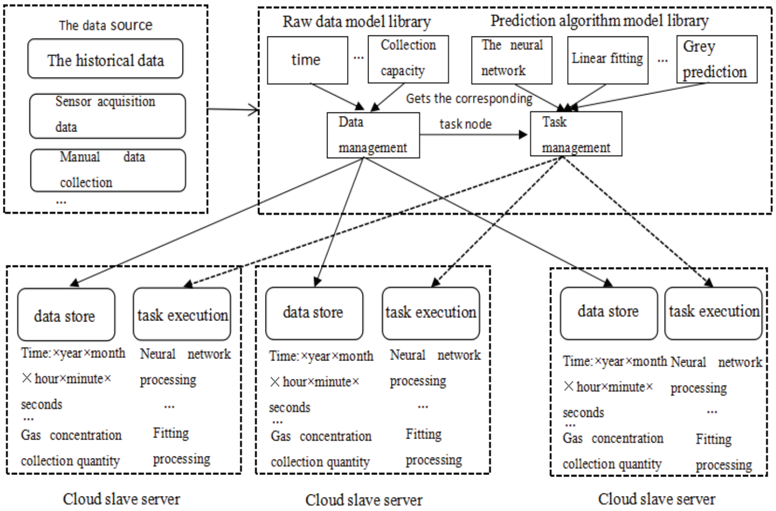

Fig. 1: Based on cloud computing gas concentration prediction analysis framework.

***Genetic algorithm optimized Elman neural network*:** Genetic algorithm is a parallel randomized search optimal solution which simulates the evolution process of biology in nature. It can be summarized as two operations: genetic operation (cross and mutation) and evolutionary operation (selection).

The genetic optimization Elman neural network first uses the genetic algorithm to find the optimal solution of the connection weight and threshold, and feeds it back into the network, and the optimal solution is obtained through sample training. Specific steps are as follows:

1. The weights and thresholds of the Elman neural network are randomly generated as the initial population individuals, and each individual is a binary coded symbol string form in which the connection weights of the layers are connected in a certain order with the threshold. Set the population size to N, the probability of hybridization p_c and variation probability p_m as the algorithm parameters.
2. Perform network learning, and the learning error e is taken as the fitness f after taking the reciprocal.

   $$F = 1/E = 1/\left[\frac{1}{2}\sum_{P=1}^{N}\ \sum_{k=1}^{M}\left(y_k^p - o_k^p\right)^2\right] \qquad \ldots(3)$$

   The fitness function F is the basis for evaluating the pros and cons of individuals in the population. Where y and o are the actual and expected values of the neural network, respectively; p is the individual label and k is the output node label.
3. Individuals with higher fitness carry good genes, have good adaptability in inheritance, and have a greater chance of producing excellent offspring. Individuals with large fitness values are selected as genetic parents with a certain probability.

The probability that the i-th individual is selected as a genetic parent:

$$P_i = F_i / \sum_{i=1}^{N} F_i \qquad \ldots(4)$$

4. The individual's reproductive process, the selected superior individuals are single-point crossover operation, and the intermediate individuals are obtained, and the basic position mutation operation is performed, that is, several genetic position points are randomly selected and inverted by the probability p_m to generate a new generation genetic group.
5. When the set number of iterations or iteration precision is satisfied, the optimization process is ended, and the optimized solution obtained by the search is decoded and used as the weight and threshold of the Elman neural network. After the training iteration, the weights and thresholds are updated again, and the prediction results of the network are output.

Cloud-based GA-Enn algorithm implementation: When genetic algorithm is used to optimize the neural network, if the individual data of the population is selected, the process of calculating the individual fitness value will consume a lot of time. This article uses Hadoop, the most popular open source cloud computing platform, to parallelize the genetic algorithm, which greatly saves the operation time. It is mainly composed of HDFS (Distributed File System) and Map Reduce. HDFS adopts master/slave architecture. HDFS cluster consists of a Namenode (management node) and several Datanodes (data nodes). Map Reduce is a parallel programming mode, including the Map phase and the Reduce phase. In the Map phase, each node server takes the split data block split as input and executes the map function on all records in each block to generate intermediate results.

The <key, value> key-value pair, in the Reduce phase, takes the key-value pairs generated by the Map as input, executes the Reduce function, processes it to generate a new key-value pair, and finally writes the result to HDFS. The realization of parallelization of genetic algorithms is divided into three stages, namely Map.

The phase, the Combine phase and the Reduce phase are as follows:

1. In the Map stage, the training set of the individual information of the population is divided into data blocks and allocated to each data node. The Map function converts it into a vector <key, value>. All Map tasks perform neural network learning in parallel to adjust the connection right of the network. Value and threshold, when the output error reaches the set value, the <individual, error value> key-value pair is output.

```
Pseudo code
Input: key, value
Output: individual, error value
Map(key, value)
{
Do{
        Elman neural network iterative training
        }While (output error does not reach the set value
of 0.01)
Obtain individual and error values;
       Emit (individual, error value)
}
```

2. In the Combine phase, the data is aggregated, and key-value pairs with the same individual are merged and sent to the Reduce function.

3. In the Reduce phase, accumulate all Map side $\Delta\omega_{ij}$ average value as global update $sum\Delta\omega_{ij}$, update all individuals.

$$\Delta\omega_{ij} = \Delta\omega_j + sum\Delta\omega_j \qquad \ldots(5)$$

Convert <individual, error value> to <individual, fitness value>, and perform genetic algorithm selection, crossover, and mutation operations.

```
Pseudo code
Input: key, value
Output: key, updated individual
Reduce(key, value)    ‘
Read weight
        While (individual not fully processed) {calculate the next individual
Weight update amount}
Adding the weight update amount to the average value and correcting the weight
For all weight individuals
        {
Converting the error to a fitness value;
        }
select;
Single point crossing;
Basic position variation
Emit (key, updated individual)
}
```

EXPERIMENTAL RESULTS AND ANALYSIS

Experimental environment and error evaluation criteria: The experiment chose to build a Hadoop cloud platform from 6 PCs. The four PCs are dual-core 2.4 GHz, 2 GB memory, and 2 dual-core 2.1 GHz, 1 GB memory. The Hadoop version is 1.2.1, using a Gigabit NIC, and the computer is connected through multiple switches.

The gas monitoring data of a fully mechanized mining face in a mining area of Tangshan City was selected for application analysis. The training data is historical data from 2016 to 2017. The sampling interval of each sampling device is 10 min. To predict the gas concentration data in January 2018, the sample data file size is about 2 GB, including 20,976,075 samples. Among them, there is a large amount of noise data.

Purely from the size of the data file is not enough to accurately indicate the concept of big data, this article has formed the concept of big data because of the huge number of samples in the sample library.

For the part of the data intercepted during this time period, see Table 1 and Table 2.

In order to ensure the accuracy of the experimental conclusions, the experiment was repeated 60 times and the average value was taken as the final prediction result.

Historical monitoring data of coal seam depth, coal seam thickness, temperature and gas concentration, which have a greater impact on gas prediction, were selected as inputs to the Elman neural network.

Table 1: Training data.

year	month	day	Time	Minute	Depth / m	Thickness / m	Temperature / °C	Gas concentration /%
2016	1	1	0	0	615	6.3	22.6	0.2590
2016	1	1	0	10	615	6.3	22.6	0.2040
2016	1	1	0	20	615	6.3	22.6	0.1850
...	...	...	...	...	...	...	...	...
2017	3	5	12	30	615	6.3	23.2	0.2590
...	...	...	...	...	...	...	...	...

Table 2: Gas concentration data.

year	month	day	Time	Minute	Depth / m	Thickness / m	Temperature / °C	Gas concentration /%
2018	1	1	0	0	615	6.3	21.8	0.2289
2018	1	1	0	10	615	6.3	21.8	0.1892
2018	1	1	0	20	615	6.3	21.8	0.2068
...	...	...	...	...	...	...	...	...
2018	1	15	1	0	615	6.3	21.3	0.2980
...	...	...	...	...	...	...	...	...

The initialization parameters of the network are

The number of input layer nodes is s1 = 4.

The number of hidden layer and connection layer nodes is s2 = 7.

The number of output layer nodes is s3 = 1.

The hidden layer uses the sigmoid function and the output layer uses a linear function. The initial weight of the network is a random number between [0, 0.1], the initial threshold is set between [0, 0.2], and the learning rate is set to 0.1. The genetic crossover rate was 0.4, the mutation rate was 0.06, and the evolutionary algebra was 50.

Due to the complicated mining environment in the underground, the data collected by the gas monitoring system contains a lot of rough and untimely "dirty data", so it should be cleaned and standardized before data analysis. Use valuation method instead of missing data value, reject noise

Sound item; the minimum-maximum normalization process is used, and after the processing is completed, the data falls within the interval of [0, 1].

To measure the effect of the forecast, the evaluation criteria used in this paper are as follows:

Let F(x) and f(x) be the actual and predicted values of gas concentration, respectively, then the relative error E1 is

$$e_1 = \frac{F(x) - f(x)}{F(x)} \quad \ldots(6)$$

Since the relative error calculation will come up positive and negative, the merits of the mean square error e_2 evaluation algorithm are introduced:

$$e_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{F(x) - f(x)}{F(x)}\right)^2 \quad \ldots(7)$$

Experiment and analysis of results: Experiment 1: In order to verify that the parallel algorithm selected in this paper has higher efficiency and prediction accuracy than the genetic optimization Elman neural network algorithm in processing massive data, two computers with the same configuration are selected, one of which is installed with data mining software. Weka can implement local genetic optimization Elman algorithm prediction, another node from the cloud platform, through testing 5 sets of different size data sets, to see the change in processing efficiency.

In order to avoid the error caused by accidental factors, repeat the test 10 times and take the average value as the final result, as shown in Table 3.

The experimental results show that the efficiency of the traditional algorithm is significantly better than that of the parallel algorithm when the amount of data is small. This is because, when dealing with small-scale data, the parallel algorithm is mainly spent on the start-up allocation and data transmission from the node task, but as the amount of data increases, the efficiency of the traditional algorithm drops sharply, which is due to the increase in data. When

Table 3: Comparison of processing efficiency.

Data size / mb	Genetic Optimization Elman Algorithm/s	Parallel algorithm/s
100	12	28
200	32	46
2048	136	86
5120	236	122
15360	446	224

Table 4: Comparison of prediction results.

Numbering	Predictive value	Actual value	Relative error
1	0.2385	0.2311	-0.0375
2	0.1865	0.1956	0.0358
3	0.2120	0.2068	0.0381
...	...	...	...
144	0.2956	0.3020	0.0311
...	...	...	...

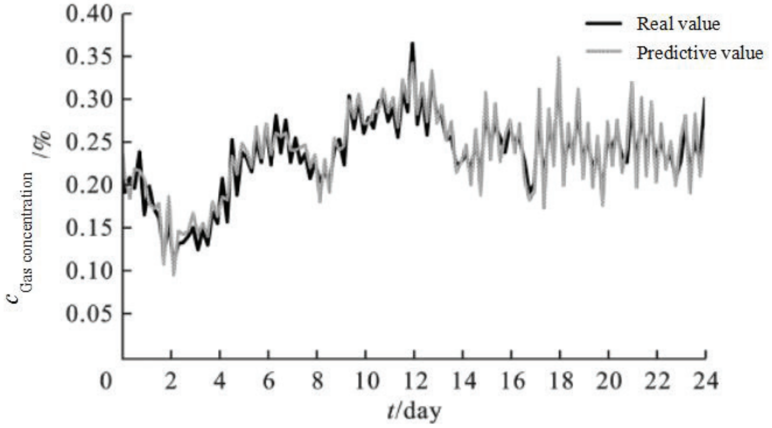

Fig. 2: Comparison of prediction results.

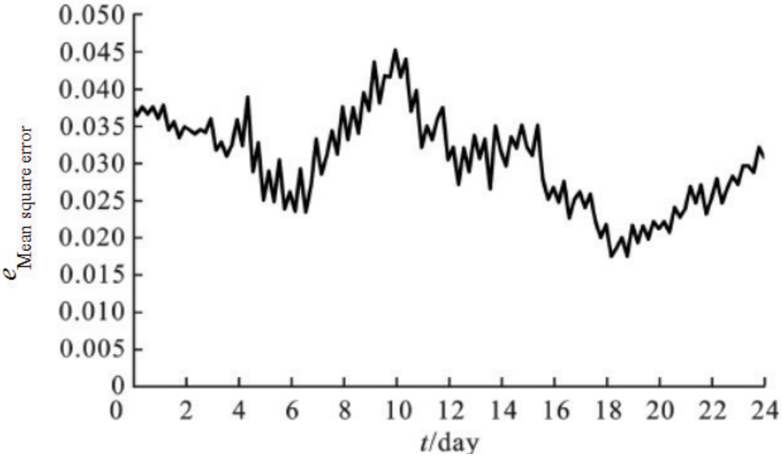

Fig. 3: Mean square error curve.

the calculation of weights consumes a lot of time, the traditional algorithm will eventually collapse due to insufficient resources, but the parallel algorithm based on distributed computing shows obvious advantages, indicating that the parallel algorithm is compared with the traditional algorithm in massive data processing. Aspects are more efficient and have obvious advantages.

Experiment 2: It combines the parallel Elman algorithm and training data proposed in this paper to complete the task of short-term prediction of gas concentration. The comparison between predicted and true values is shown in Table 4.

A comparison of the predicted and true values of a certain day in January 2014 is shown in Fig. 2.

It can be seen from the figure that the two curves are roughly the same, and the curve trends are generally similar. The mean square error curve of the day is shown in Fig. 3.

The mean square error is 0.037. It can be concluded from the figure that the mean square error is basically stable within 0.05 and an acceptable error range is reached. Through this experiment, it is proved that the cloud-based GA-Elman algorithm is feasible.

CONCLUSION

The experimental results show that the efficiency of the traditional algorithm is significantly better than that of the parallel algorithm when the amount of data is small. However, with the increase of the amount of data, the efficiency of the traditional algorithm drops sharply. This is because when the data increases, the calculation of the weight will consume a lot of time. The traditional algorithm will eventually collapse due to insufficient resources, but the parallel algorithm based on distributed computing. However, it shows obvious advantages, which shows that the parallel algorithm is more efficient and obviously superior to the traditional algorithm in massive data processing. Experiments show that the cloud-based genetic optimization Elman prediction algorithm has greatly improved processing speed and prediction accuracy.

REFERENCES

Che, H. C., Zhang, X. Y., Zhang, L., Wang, Y. Q., Zhang, Y. M., and Shen, X. J. 2017. Prediction of size-resolved number concentration of cloud condensation nuclei and long-term measurements of their activation characteristics. Scientific Reports, 7(1): 334-347.

He, J., Duan, C., Lei, M., and Zhu, X. 2016. The secondary release of mercury in coal fly ash-based flue-gas mercury removal technology. Environmental Technology, 37(1): 28-38.

Lanzerstorfer, C. 2018. Fly ash from coal combustion: dependence of the concentration of various elements on the particle size. Fuel, 228: 263-271.

Lin, M., Peng, W., and Gui, H. 2016. Heavy metals in deep groundwater within coal mining area, northern Anhui province, china: concentration, relationship, and source apportionment. Arabian Journal of Geosciences, 9(4): 1-10.

Qiang, Z., Li, C., Qiang, W., and Zhang, B. 2018. Effect of dry water on methane separation and recovery from coal mine gas based on hydrate. Rsc. Advances, 8(48): 27171-27180.

Temel, N. K., Sertakan, K., and Gürkan, R. 2018. Preconcentration and determination of trace nickel and cobalt in milk-based samples by ultrasound-assisted cloud point extraction coupled with flame atomic absorption spectrometry. Biological Trace Element Research, 186(2): 597-607.

Tröstl, J., Chuang, W. K., Gordon, H., Heinritzi, M., Yan, C., and Molteni, U. 2016. The role of low-volatility organic compounds in initial particle growth in the atmosphere. Nature, 533(7604): 527.

Yong, J., Wang, L., Zhang, L., An, W., Wei, W., and Wei, Z. 2018. Direct operation of solid oxide fuel cells on low-concentration oxygen-bearing coal-bed methane with high stability. Energy and Fuels, 32(4): 7b02968.

Yuan, D., Zheng, Y., Li, Q., Lin, B., Zhang, G., and Liu, J. 2018. Effects of pore structure of prepared coal-based activated carbons on CH 4 enrichment from low concentration gas by IAST method. Powder Technology, 333: 377-384.

Zhao, J., Zhao, Y., and Liang, W. 2016. Hydrate-based gas separation for methane recovery from coal mine gas using tetrahydrofuran. Energy Technology, 4(7): 864-869.