Journal of Science and Engineering Management
Vol. 1, No. 2 (2020), pp.33-44
http//dx.doi.org/10.33832/jsem.2020.1.2.05

THE APPLICATION OF THE BP NEURAL NETWORK IN TIME SERIES PREDICTION

Joje Sicad

Iloilo Science and Technology University, Philippines

***Abstract*— A time sequence is a collection of data collected at different points in time that reflect the change of a thing, an image, etc. over time. Through the study of this series of data, we can find the rules hidden behind the data, model and analyze the reality and solve the problems in the domain. Time series prediction is a prediction that has not come through the observed data, enabling the decision-taker to have the ability to be forward-looking and to make better decisions, and therefore has an important meaning. This paper first introduces the time series prediction several prediction models that are currently mainly used, and analyzes the advantages and disadvantages of the current method and model for non-line system prediction. Then the advantages of artificial neural networks, especially BP neural networks in non-line system forecasting were introduced. Summarizes some defects in the Back Propagation (BP) Neural-by-Network prediction model, mainly the initial weight selection does not lead to the Neural-by-Neural network easily falling into local minimal values. This paper presents a BP Neural-by-Neural network model that uses genetic algorithms to improve the predictive performance of the Neural by the web network. With the development of information technology such as interconnection, mobile communication, and internet access, the number of data increases rapidly. The BP Neural-by-Network model has problems such as slow network convergence when there are many hidden layer points and a large number of data samples. To solve the above problem based on Hadoop, an open source distributed cloud platform. This article presents a parallel prediction model based on MapReduce. The different stages of the prediction model are distributed and parallel, which improves the calculation efficiency.**

***Keywords*— Time Sequence; Predictions; BP Neural Networks**

1. INTRODUCTION

By analysing the characteristics of wave overmen and shock of the time series data, time sequence prediction techniques grasp the periodic law of change of the time series. [1][2] It is possible to achieve the prediction of values or trends over a certain period of time. An example of a moving image is when the ancient Egyptians recorded the rise and fall of the Nile Every day, and the number recorded was what we said here. Is there any meaning in recording the rise and fall of this river? At that time, the people did not take a look at the time sequence for a long time. As a result, they found that the rise and fall of the Nile River was very regular law. According to the knowledge of the Nile River they indiscriminate rules to help the ancient Egyptian people on farming and housing planning, send to make agriculture develop rapidly. From the creation of Egypt splendid prehistoric civilization, which can be seen in the time sequence in ancient Egypt when the generation has been used. In the present generation society, the meteorological data is a series of data that change over time. And the gas and elephant workers have used the modeling analysis of

Received: September 25, 2020
Reviewed: November 26, 2020
Accepted: December 3, 2020

ISSN: 2652-8045 JSEM

the gas data to predict the change of the weather to provide great help to social production and people's lives. Stock prices change over time in a moment, which can experience the mutual transformation of peaks and troughs. It can lead the market to study the stock market by describing the variation of the stock market in the form of time series. You can see that the time sequence in nature, economy and other areas have played a great role in helping.

Prediction [3] is a prediction of a person's progress based on historical data and the study of a line that has not come for a period of time. A sure and reasonable prediction is the basis for people to make reasonable decisions. In scientific research and engineering practice, the prediction of the fast and accurate time sequence has always been a subject of close attention by many learners. The time sequence [4][5] contains a very large number of messages and contains a number of systems of operations, so the analysis of these time series data, from found in the data in the evolution of the rules, from and to achieve in the system before the line of travel forecast, in the actual application has important significance and value. For example, through the stock price time sequence to the stock price into the forecast, can enable the investor to effectively avoid risk; Sequence into the line of prediction "advance hair gas like the ground quality disaster warning, to minimize the loss; Patient number of time sequence into the line forecast, can let the relevant department door in advance to prepare, reasonable allocation of public health resources. Based on an analysis of the materials relevant to time series prediction, some researchers and experts have developed some basic methods of modeling and forecasting. When establishing a time series prediction model, if the time series number is related to a near-full line, a statistical-based modeling and forecasting method can be used, such as self-return. The model sliding average models, self-sliding moving average models, and false time series of unresolved and historical values have a linear relationship. If the time series [6] data is non-wired, use a non-line model to make predictions. Because the line model is easy to build, understand and interpret, people usually use the line model to solve problems. The traditional statistical prediction method works better on the number of time series on the line, but the time series obtained in the actual environment should be used Data are usually complex, non-lineal, and uneven sequences, accurate prediction of these time series data is very difficult. The prediction of non-line system has been put forward a lot of methods, self-return model, slide transformation self-return model, *etc*. These models have a certain degree of predictive ability for non-linear sequences, but the generalization ability is poor. The rise of the artificial neural network provides a new way to solve the problem of non-line, but the artificial neural through the network global search ability is not strong, easy to fall into the local very small value. According to more sample number, the numerical fluctuation period of some time series data is long and the characteristics are loose, and the data information can not be obtained fully during the training of neural network, which lead the model to grasp the trend of timing movement accurately.

On the one hand, in order to improve the accuracy and efficiency of neural network prediction model, it is necessary to grasp the periodic law of timing motion and improve the accuracy of model prediction by highlighting the key information in the time series data, so as to grasp the periodic law of timing motion. On the other hand, to avoid the excessive learning of the network model by simplifying the redundant information of the time series data, and thus improve the training rate of the neural network, and finally solve the problem of the inefficiency of network model construction.

2. RELATED RESEARCH

In the process of studying time series prediction, many scholars have found that single prediction models, whether based on traditional statistics or intelligent models based on neural networks, have certain limitations, because different models have advantages and disadvantages in predicting performance. So some combined models are proposed, which can be more comprehensive and stable from different angles. Some scholars use several

models for weighting and find that the quality of prediction is improved by comparing to a single model. In the optimization method of BP neural network, the bio-intelligent optimization algorithm which has emerged in recent years has been recognized by many scholars. The group intelligent optimization algorithm is an heuristic optimization method based on the biological habits and behavior of nature in recent years. The algorithms that optimize BP network intelligently are: genetic algorithm, particle group algorithm, fish algorithm, bee group algorithm, cuckoo algorithm, genetic algorithm, *etc*.

BP neural networks have the ability to map complex problems nonlinearly and are very effective in dealing with complex internal mechanisms. However, neural networks also have limitations, such as the model needs to set more parameters, and there is no effective method to filter parameters; Time series data usually has a small amount of noise data, Symbolic Aggregate approximation (SAX) [7] notation can ignore the effect of subtle noise on feature extraction, but the more obvious noise data in time series data can still lead to errors in SAX notation. Stelios *et al.*, have achieved the lowest difference, and the SAX notation with limited symbolic error is proposed by analyzing the distribution characteristics of the values of the sequence to be processed, optimizing the number of split points, i.e. the number of symbols. So that the error of the SAX representing sequence is aimed at the defects of the BP neural network, and the researchers of the time series have proposed a variety of data mining techniques based on SAX notation by combining the time series descending algorithm with different fields of technology. Lin et al. proposed an improved symbolic notation experiencing SAX, that enables SAX notation to support timing mining techniques in the original data space. Shieh and Camerra *et al.*, [8] proposed time series index structures iSAX and iSAX2.0for ultra-large amounts of time series data, which effectively solved there trilevel difficulties of time series data of more than one million years. For ECG data on mobile devices. Tayebi *et al.*, proposed RA-SAX, a time series analysis technology based on the mobile device side, which enable real-time analysis and processing of time series data. Payam et al. introduced SAX notation into the mobile sensing data processing framework, improving the performance of its pattern analysis and identification. Luo *et al.*, [9] built SAX's dynamic index structure through self-built networks, enabling them to capture and recognize human motion data. Buranasing *et al.*, [10] proposed a neural network model that predicts storm intensity by combining SAX notation with artificial neural networks. These techniques solve the practical application needs by applying SAX notation to different fields, and in the process of studying time series prediction, many scholars find that single prediction models, whether based on traditional statistics or intelligent models based on neural networks, have certain limitations, because different models have advantages and disadvantages in predicting performance, so some combined models are proposed, which can be made from different angles, more comprehensive and stable prediction.

3. EXPERIMENTAL METHOD

Genetic Algorithm (GA)'s search strategy does not rely on gradient information, so it is widely used, especially for complex nonlinear problems that traditional methods cannot solve. And GA is a coding combination of parameters, not a set of parameters, so it's very versatile. However, GA also has some disadvantages, such as the characteristics of easy precociousness. The GA improvement method mainly maintains the individual diversity in the algorithm, such as adjusting the operating parameters, expanding the population size, etc., but does not improve the learning ability and robustness of the algorithm, which is the advantage of artificial neural network. Therefore, combining neural network and GA can make full use of the advantages of both, so that the new algorithm has both the learning ability of neural network and the powerful global random search ability of GA.

3.1. GENETIC ALGORITHMS EXCELLENT BP NEURAL BY NETWORK MODEL RESEARCH

First of all, we need to study the coding problem, the classic genetic algorithm generally uses binary coding, but this coding method in dealing with more complex problems, more arguments, which will cause the solution space to become large, not conducive to fast search. In the large scale of the species system, we can use real coding, a real number directly mapped to a gene bit of chromosomes, which can effectively shorten the length of chromosomes, while avoiding repeated coding and decoding operations, simplifying the operation of genetic algorithms. The encoded code string consists of connection weight thresholds. If the neural network input layer has p nodes, the implied layer has q nodes, the output layer has c nodes, and the encoded mapping relationship can be expressed as formula (1):

$$X = \{w_{11}, w_{12}, \cdots, w_{pq}, v_{11}, v_{12}, \cdots, v_{1p}, \tau_1, \tau_2, \cdots, \tau_q, s_1, s_2, \cdots, s_c\} \tag{1}$$

Connect to form a string of codes to form an individual. N such individuals form a group, *i.e.*, the size of the population is N. The initialization of the population is also studied, and the initial resolution space requires the design of N initial individuals, each of which has a specific range of changes. If the scope of the solution space design is too large, the search efficiency will be reduced, and the scope is too small, which may cause the situation of no solution. The size of the population has a great influence on the results of the genetic algorithm's global optimization and the efficiency of the search, and a more appropriate population size should be set in combination with the actual situation.

The second step is to determine the adaptability function, which GA determines which individuals remain and which individuals do not produce offspring, depending on the size of the individual's adaptability. So the adaptability function directly affects the quality of GA's excellence. Because GA is used to optimize the weight threshold of the neural network, which in turn affects the error of the neural network, and the sum of the squares of the errors is the adaptation function. Then the sum of the squares of the errors is the highest, and the adaptability function is represented as shown in formula (2):

$$\text{fitness} = \frac{1}{E} \tag{2}$$

Where E is the output error of the neural network as shown in formula (3):

$$E = \frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y}_k)^2 \tag{3}$$

Where the expected and actual results of the k node of the output layer are k, N is the number of training set samples. After calculating the adaptability value according to the adaptability function, the evolution of the population is carried out. In the process of population evolution, all individuals in the population are arranged according to the size of adaptability and the sequence to determine the probability of individual selection. The process is as follows:

1. Descending order of individuals according to the size of adaptability;

2. Based on practical problems, plan a probability distribution table that assigns probability values to individuals based on the results of the above adaptation arrangement, as shown in formula (4) and formula (5):

$$q = r(1 - p_{max})^{N(k-1)} \tag{4}$$

$$r = \frac{p_{max}}{1-(1-p_{max})}N(k) \tag{5}$$

The optimization of BP neural networks using GA is mainly focused on three aspects: first optimizing the topology of neural networks, second optimizing learning rules, and optimizing the value of neural network connectivity.

The topology and the number of nodes of BP neural network have a great influence on the network. And a good network structure does not have redundant nodes and redundant

connection rights. In fact, the determination of network structure is a very difficult thing. At present, the number of network nodes to determine mainly by increasing or decreasing the method of testing, this method can solve some problems, and BP network topology determination to date. There is no effective principle. Traditional BP neural networks are pre-determined for learning rules in sample set training, but this may not be very effective. Optimize the learning rules of neural networks and improve their ability to predict performance. The difficulty with using GA to optimize learning rules is how to encode them. Not long after these studies began, there were no successful examples. The training process of BP neural network is a kind of complex function optimization, which finds the optimal solution of approximate connection right value through many constant modifications. The neural network weight matrix contains all the information of the nervous system, and the method of value seeking excellence is to adopt some rules, adjust gradually, and finally get the approximate optimal weight. However, the BP network is based on the gradient drop rule, the initial weight has a great impact on the network, and the results of different weights can sometimes vary greatly. Using genetic algorithms to optimize the weight of neural networks can avoid these problems to a great extent. This chapter describes how to optimize the implementation of connection weights.

The probability that the chromosome q is the kth in the series is selected; r will be allocated p_{max}; the sequence number $N(k)$ for the individual to adapt to descending order, the probability of the selection of the best chromosome, p_{max} L is the population size.

3. Chromosomal selection is then made using roulette selection: the cumulative selection probability of each chromosome $q_i$is calculated to produce sequence random values from large to small in the $x_j, if\ q_{i-1} < x_j < q_i$interval, the ith chromosome is selected as the next generation of evolutionary chromosomes. Then start with two of the most important calculations in GA: crossover and variation.

(1). Cross operator: When chromosomes are encoded in real numbers, they are directly calculated.

(2). Variant actuation: a certain mutation of the genetic value of the chromosome, the result of which is a new gene value.

Finally, you have to set the termination conditions for evolution. In practice, in order to obtain reasonable values in practice, the terminated evolution alga is set to a relatively large value, and the change curve between adaptability and evolutionary alga is observed. This value is used as the maximum value of the evolutionary algae when there is little change in the function graph as the adaptive curve flattens as the precariat terminates agleam. The BP Neural Network (GA-BP) model process optimized with GA is as follows:

Step 1: Estimate the initial topology of the BP neural network based on specific issues;

Step 2: Pre-train the BP neural network a certain number of times to obtain the initial weight and threshold range;

Step 3: Determine chromosomal coding and encode the initial weight threshold using real-number encoding. At the same time, the size of the population, evolutionary alga while determining the parameters;

Step 4: Determine its adaptability function. Calculate the error between the output value of the BP neural network and the expected value.

The sum of squares is calculated as the adaptation value of individual i by the inverted sum of error squares between the output value and the expected value. $f(i)$

Step 5: Determine whether the individual's adaptability meets the optimization criteria and, if so, turn to step 8;

Step 6: Genetic Operation:

(1). Choose. Calculates and sorts the adaptability values of the individual bodies in the population.

(2). Cross. The cross-probability is crossed by arithmetic cross-method.

(3). Variation. Non-uniform variation is used to variation according to the probability of variation.

Step 7: Create a new group. Repeat steps 3 to 6 to optimize the initially determined weights and thresholds until the optimization criteria are met;

Step 8: Optimal decoding of GA optimization as the optimal authority and threshold for BP neural networks;

Step 9: The training is over.

3.2. DATA PRE-PROCESSING

Most of the original data series have a great degree of randomness, in order to improve the prediction accuracy, weaken the degree of discreteness of the original sequence can not directly predict the original data sequence. The original need to be pre-processed. There are generally two ways to preprocess the data of the gray prediction model, sequence generation and sequence reduction generation. The add-up generation is to generate a new data series for the data of the sequence time node in turn, and the subtracted generation is to generate a new data sequence for the two data subtracted of the adjacent time node of the sequence. The subtracted generation can be called the inverse operation of the add-up generation.

Set the original non-negative data column as shown in formula (6):

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)\} \tag{6}$$

In order for the original data column to fluctuate significantly and take on an incremental form, the original data column needs to be added up once, and the data column after the original data column is added up once, as shown in formula (7):

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(N)\} \tag{7}$$

The added sequence elements can be summed up as formula (8):

$$x^{(1)}(i) = \{x^{(0)}(i) | i = 1,2, \cdots, N\} \tag{8}$$

Among them $x^{(0)}(1) = x^{(1)}(1)$

3.3. THE STUDY OF FEATURE SELECTION KEY TECHNOLOGIES

Feature selection is the data pre-processing process. From the original data set collected, which select a part of the characteristic properties to optimize, so that the system can use the least number of indicator properties to complete the system's tasks. Thereby not only reducing the dimension of data set processing, but also the implementation of the algorithm are more effective and fast. At present, the data feature dimension is relatively high, which contains a lot of feature information. Some may not be related to personal credit or not related and some characteristic attributes are related to each other is relatively high, which leads to a large amount of information redundancy and noise throughout the data set. If not selectively all the feature indicators are used for model construction, the first to make the model's computational speed slower, run longer, and may even affect the classification of the model and reduce the accuracy of the results. Therefore, it is necessary to select a small number of more useful features from it, so as to build a model to achieve better results.

(1). Variance filtering:

The first thing is to look at whether there is exactly the same data, if most of the observations of a variable change little or even the same value, that is, the variance of the feature is relatively small. Then we think that such feature variables can not be used to

significantly sub-target variables. You can consider removing such feature variables. So first we should reject those variables with exactly the same value, that is, set the variance filter threshold to 0, and if you find that there are no identical features in the data, the number of feature variables remains the same. At this point, if you choose to increase the threshold of variance filtering to remove more features, after the experiment found that the threshold specific value is not good to determine, and no matter how many thresholds are selected. For the prediction effect will have a large discount, classification accuracy will decrease by 0.1 to 1 percentage point. A little bit is not worth the loss, so at this time choose to abandon variance filtering and consider other feature selection methods.

(2). Relevance test:

After variance filtering, it is more suitable to test the correlation of these characteristic properties and to check whether they have correlation with each other, and to eliminate some variables with high correlation coefficients, so as to eliminate the adverse effects of collinearity on subsequent models. Here I choose to remove the correlation coefficient greater than 0. One of the feature variables is 8. After correlation screening, some variables with high correlation were removed.

(3). Random forest algorithms filter features:

If you train directly with all the remaining features for a random forest model, you need to randomly select several features to form a subset of features. And the probability of selecting features are really helpful to modeling is relatively low, which may result in the resulting decision tree and random forest fitting effect may not be good enough; Therefore, it is necessary to continue the filtering of feature variables. There are many methods of feature selection, such as embedding, packaging, and so on. Based on efficiency considerations, this paper finally decides to use random forest algorithms to filter features. Because the random forest algorithm is convenient, takes less time and is efficient, and has a special interface to measure the importance of feature variables for model fitting effect. It provides quantitative support for filtering features.

After feature screening, we can use as few features as possible to get the best possible results. In general, the specific distribution of sample data also affects some experimental results more or less. So in order to minimize this effect and try to make the results of the model more stable, it is generally necessary to take a multi-cross-validation approach to determine the optimal combination of features. The main process of cross-validation is: first, all the training sample data randomly divided into equal n parts, and then select the n-1 data as the training set, the remaining one automatically as a test set for a number of classification experiments. This n-test will be n optimal feature subset selection, there is n characteristic importance sorting, and finally will get n groups of different subsets of features, synching the n-times results and selecting from which to appear more times the characteristic composition of the characteristic subset.

4. RESEARCH RESULT

This experiment is mainly to test the parallelization of the cluster of MapReduce-based GA-BP neural networks. So only the convergence time of the test algorithm is required. The selection data for this chapter is from Knowledge Discovery and Data Mining (KDD) CUP 99. This is a complete predictive data set, used to predict the type of cyberattack, a data set for data mining feature analysis. Although it is old, but still a classic in this field. Each sample has 38 property values, a total dataset size of 708M, and a sample of 4898431 rows. Starting with all the data sets, the number of test nodes is the convergence time of the 1 to 8 node algorithm, 7 tests per set of experiments, which taking average time, time consuming and accelerating for different nodes such as Table I.

Table I. The Time-Consuming and Acceleration Ratio of the Different Nodes

The number of nodes	The number of seconds to run	Acceleration ratio
1	7896	/
2	5522	1.43
3	3494	2.26
4	2437	3.24
5	1984	3.98
6	1732	4.56
7	1709	4.62
8	1666	4.74

To test the execution efficiency of parallel algorithms at smaller data volumes: another experiment was designed, starting with 10% of the data set, increasing the amount of data by 10% at a time, using four fixed nodes, and testing the algorithm's convergence time as Table II.

Table II. The Convergence Time of the Test Algorithm

The scale of the dataset	The number of seconds
10%	561
20%	986
30%	1064
40%	1238
50%	1470
60%	1760
70%	1876
80%	2244
90%	2302
100%	2437

As can be seen from the chart, under the fixed number of nodes, four nodes should take 2437 seconds at 100% of the dataset. And 10% of the data theory should take 247 seconds at2467 x0.1 and 561 seconds at the actual time. Because the Hadoop platform requires a certain amount of startup time and task allocation time, communication between the nodes takes more time when the amount of data is small. The percentage of time actually spent performing the job is reduced, and the actual utilization of the system is low. With the increase of data volume, the proportion of time to execute the job increases, and the acceleration ratio gap between the four nodes in the amount of big data decreases gradually, which indicating that the utilization of cloud platform has been improved.

Using the improved algorithm in this paper, the values of TAI defense from November 1, 1991 to December 28, 1991 are predicted, and the results are as follows:

Table III. True and Forecast Values of TAI Defense from 1 November 1991 to 28 December 1991

Date	True value	Forecast value	Date	True value	Forecast value
1991-l1-1	4332.58	4408.42	1991-12-2	4355.02	4405.84
1991-11-2	4299.74	4277.27	1991-12-3	4343.24	4282.61
1991-11-4	4278.84	4275.73	1991-12-4	4378.13	4281.84
1991-11-5	4423.43	4274.06	1991-12-5	4340.09	4404.83
1991-11-6	4430.59	4392.35	1991-12-6	4399.88	4281.48
1991-11-7	4424.85	4406.97	1991-12-7	4501.9	4386.61
1991-11-8	4439.49	4397.14	1991-12-9	4473.82	4510.68
1991-11-9	4495.13	4413.03	1991-12-10	4455.76	4504.58
1991-11-11	4491.68	4538.13	1991-12-11	4401.79	4474.86
1991-11-13	4469.73	4536.61	1991-12-12	4439.97	4404.06

1991-11-14	4491.52	4411.67	1991-12-13	4417.06	4404.74
1991-11-15	4511.36	4527.98	1991-12-14	4399.61	4385.94
1991-11-16	4442.46	4478.26	1991-12-16	4391.53	4404.09
1991-11-18	4420.69	4407.89	1991-12-17	4400.3	4404.50
1991-11-19	4480.94	4409.71	1991-12-18	4457.62	4405.02
1991-11-20	4478.71	44W.90	1991-12-19	4466.57	4474.39
1991-11-21	4496.89	4526.05	1991-12-20	4448.04	4511.06
1991-11-22	4543.84	4523.93	1991-12-23	4459.75	4407.94
1991-11-23	4505.8	4522.99	1991-12-24	4449.58	4470.42
1991-11-25	4534.62	4521.31	1991-12-26	4529.58	4418.79
1991-11-26	4529.62	4518.81	1991-12-27	4540.55	4551.39
1991-11-27	4471	4519.54	1991-12-28	4600.67	4552.22
1991-11-28	4453.6	4518.88			
1991-11-29	4391.61	4404.23			
1991-11-30	4378.5	4405.42			

From the data in the table, it can be concluded that the trend of the forecast data and the movement of the real data are shown in Figure 1 and Figure 2:

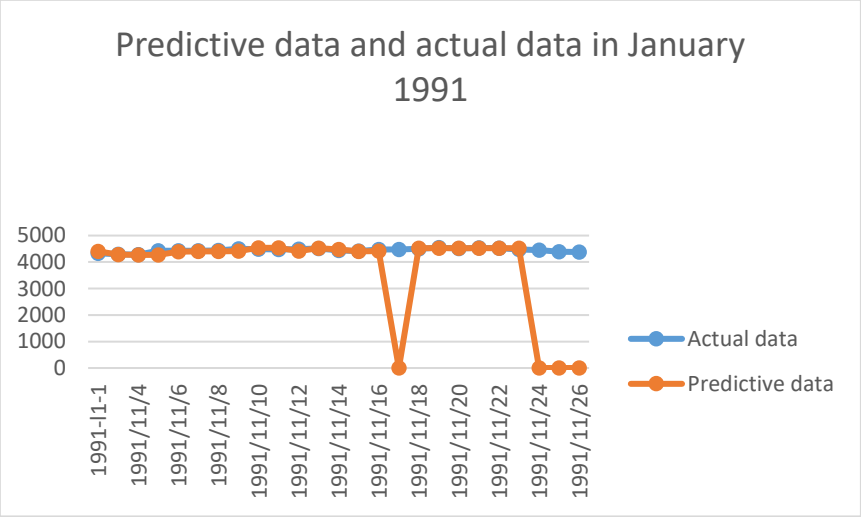

Fig. 1 Predictive Data and Actual Data in January 1991

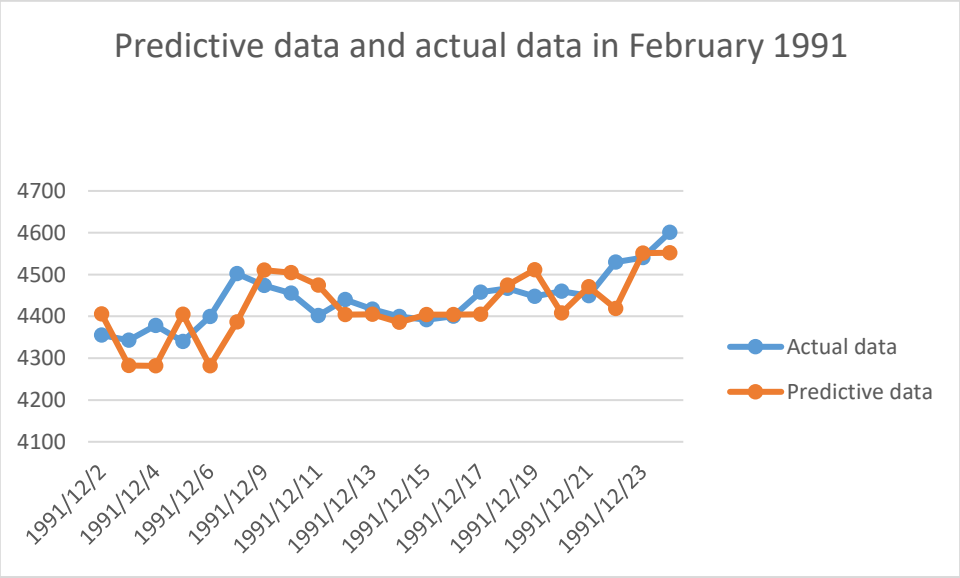

Fig. 2 Predictive Data and Actual Data in February 1991

5. CONCLUSION

It is important for people to make reasonable adjustments in advance by predicting the future changes of the study object through the observed time series data. In the past, most prediction methods were based on general statistical methods. But because time series prediction is a complex nonlinear relationship, traditional prediction methods are often powerless. Moreover, traditional forecasting methods [11]**Error! Reference source not found.**[13] generally take sampling when processing large-scale data sets, so that processing will lose information leading to increased prediction errors and take too long to use all the data. With the development of cloud computing, mobile Internet, internet of things and rapid growth of data volume, how to make rapid and accurate prediction has become an urgent problem to be solved.

In recent years, the rapid development of artificial neural networks [14] has provided people with a new method. BP neural networks with its powerful nonlinear capabilities, has become the most widely used artificial neural network model. Time series is a complex nonlinear system that is predicted using neural network models. Among all kinds of neural networks, BP neural networks have become the most widely used neural network models by virtue of their own advantages. There have been many successful cases for time series applications at home and abroad. But BP networks have some inherent defects, so how to improve the accuracy of BP neural networks has become an important research content.

For BP neural networks **Error! Reference source not found.** prone to local optimal solution and other natural deficiencies, network topology lacks effective theoretical basis, and need to optimize BP neural networks. GA is based on the natural genetic variation and other biological mechanisms of the group intelligent algorithm, using population evolution to find the global optimal solution. In this paper, a BP neural network algorithm optimized with GA is proposed to optimize the initial value of BP neural network, which can effectively improve the prediction quality of BP neural network. Experiments have also proved that the BP neural network model optimized by genetic algorithm has less error in influenza prediction than the traditional BP neural network model and Auto-Regressive Integrated Moving Average (ARIMA) model. For BP neural networks optimized by genetic algorithms, the model takes too long to train or even complete when the amount of data is large. Hadoop's open source cloud platform provides a Hadoop Distributed File System (HDFS) and MapReduce programming model that uses the MapReduce programming model to break down a large task into small tasks and solve problems that previously could not be solved by single nodes. In this paper, a data set parallelization method of BP neural networks based on MapReduce's GA optimization is proposed, mainly in the parallelization of GA and the parallelization of BP networks. GA parallelization uses a multi-group approach to parallelization by allocating different initial populations at different nodes, which can increase the diversity of populations, individual diversity, improve the chances of obtaining the best individual. The parallelized Map phase of the BP network calculates the amount of change for each connection weight in the neural network, which is temporarily local, and finally. The reduce side calculates the average of the change in ownership value, using this evaluation change to update the weight, which can speed up the convergence time of the BP network. The results show that the convergence speed is obviously accelerated in the cluster environment, which has obvious speed advantage over the traditional single-node training.

REFERENCES

[1] Yang, Y. and Zhang, L., “Research And Implementation of Iterative MapReduce Based On BP Algorithm [C]”, International Conference on Materials Engineering, Manufacturing Technology and Control, (2016).
[2] Attoh–Okine, N. O., “Machine learning: A basic overview, big data differ”, (2017), pp. 59-111, Doi: 10.1002/9781119229070.

[3] Su, C. H. and Cheng, C. H., "A hybrid fuzzy time series model based on ANFIS and integrated non linear feature selection method for forecasting stock [J]", Neural Computing, (2016), pp. 264-273.
[4] Panapakidis, I. P. and Dagoumas, A. S., "Day-ahead electricity price forecasting via the application of artificial neural network based models [J]", Applied Energy, (2016), pp. 172:132-151.
[5] Yang, Y., and Zhang, L., "Research and implementation of iterative MapReduce based on BP algorithm [C]", International Conference on Materials Engineering, Manufacturing Technology and Control, (2016).
[6] Cano, G., Garcia-Rodriguez, J. and Garcia-Garcia, A., "Automatic selection of molecular descriptors using random forest: Application to drug discovery [J]," Expert Systems with Applications, (2017), pp. 151-159.
[7] Lkhagva, B., Suzuki, Y., and Kawagoe, K., "Extended SAX: Extension of symbolic aggregate approximation for financial time series data representation[J]", (2006).
[8] Camerra, A., Palpanas, T. and Shieh, J., "ISAX 2.0: Indexing and mining one billion time series [C]", IEEE 13thInternational Conference on Data Mining, IEEE, (2013), pp. 58-67.
[9] Xiao, Q., Luo, Y. and Gao, S., "Human motion retrieval with symbolic aggregate approximation[C]", Control and Decision Conference, (2012), pp. 3632-3636.
[10] Buranasing, A. and Prayote, A., "Storm intensity estimation using symbolic aggregate approximation and artificial neural network[C]", International Computer Science and Engineering Conference, (2014).
[11] Barbara, D. C. and Liisa, H., "HAPT2D: high accuracy of prediction of T2D with a model combining basic and advanced data depending on availability[J]", European journal of endocrinology, vol. 178, no. 4, (2018), pp. 331-341.
[12] Wang, J. S. and Ning, X. C., "Time series prediction of them bank cash flow based on BP neural network Algorithmic," ICIC Express Letters, vol. 10, no. 7, (2016), pp. 1627-1633.
[13] Brocke, S. A., Alexandra, D. and Alexander, M. D., "Prediction of membrane permeation of drug molecules by combining an implicit membrane model with machine learning[J]", Journal of chemical information and modeling, vol. 59, no. 3, (2019), pp. 1147-1162.
[14] Davy, G., Kevin F, and Slade M., "Combining machine learning models of in vitro and in vivo bioassays improves rat carcinogenicity prediction[J]," Regulatory toxicology and pharmacology: RTP, vol. 94, no. 6, (2018), pp. 8-15.
[15] J.M., Lezcano - Mr Valverde F. Salazar, L. Le6n, E., Toledano J.A. Jover, B. Femandez - utierrez, E. Soudah, I. Gonzdlez - Alvaro, L. Abasolo, and L. Rodriguez - Rodriguez, "Development and validation of them a multivariate predictive model for rheumatoid arthritis mortality using a machine learning," (2017) DOI: 10.1038 / s41598-017-10558-W.

