

Research on the E-commerce Sales Forecast Based on the Improved Parallel Chaos Algorithm in the Cloud Computing

Guanghe Han

*School of Economics and Management, Heilongjiang Bayi Agricultural University
Daqing, Heilongjiang 163319, China
dqhanguanghe@126.com*

Abstract

In recent years, Chinese e-commerce industry has developed rapidly. And it makes a series of remarkable achievements. To forecast the sales of the electronic commerce can evaluate the developed status of the e-commerce better. And it also brings a great help for the management of the electronic commerce. To combine the electronic commerce with the cloud computing provides the power to the further development of the electronic commerce. In this paper, we use the related cloud computing knowledge and propose the improved parallel chaos algorithm. And we use the algorithm to forecast the electronic commerce sales. The experimental results show that the algorithm can get the predictive values quickly and accurately.

Keywords: *Electronic commerce; Cloud computing; Parallel chaos algorithm; MapReduce*

1. Introduction

Since the emergence of the electronic commerce, the industry has been developed rapidly. The rapid development of the electronic commerce provided the strong driving for Chinese economic boom. Therefore, many scholars began to study the electronic commerce industry [1-2]. While the electronic commerce industry developed rapidly, how to manage the electronic commerce better has attracted the attention of many scholars [3-4]. To forecast the electronic commerce sales can better evaluate the development status of the electronic commerce industry. And it can also manage the electronic commerce industry better. The large usage of the cloud computing made it possible to combine it with the electronic commerce. And the cloud computing made the development of the electronic commerce reach a new high [5].

At present, there were many scholars studying the electronic commerce. Zhang Qiang studied on the collaborative forecasting on supply chain in electronic business [6]. The author combined the quantitative method with the qualitative method to study the influence of the supply chain member demand on the system cost. Under the electronic commerce environment, the paper studied the significance of the supply chain management for the joint forecast demand. He proposed that when the online shopping competed with the traditional shopping, the online shopping should make more focused on exploring the new needs. Zhu Bangzhu and Zhang Qiuju researched on the E-business customer churn prediction based on three-step integration [7]. The author constructed the eclectic commerce customer of the three stage prediction model based on the SMC and the LSSVM. Firstly, the model adopted the SMC model to compute the customer activity. It made the 0.5 as the threshold determine the status of the churn of the customers, and it can identify the sentenced customers and the misjudge customers. Then the model sent the training sample to LSSVM to train and learn, and it can identify the customer churn

status. Finally, it input the misjudge customers into the nearest neighbor classifier to prejudice. Zhu Bangzhu studied the SMC-RS-LSSVM model based on the customer churn forecast model of the eclectic commerce [8]. Firstly, the author used the SMC model to compute the customer activity. It made the 0.5 as the threshold determine the status of the churn of the customers. And it can identify the sentenced customers and the misjudge customers. Then, the author adopted the rough set theory to get the customer churn forecast index system. Then he sent the training set into the least square support vector machine. At last, he can determine the customer churn status for the training sample. Later, Zhu Bangzhu and other people introduced the self organizing data mining method electronic commerce customer churn prediction. They proposed customer churn prediction model of the electronic commerce based on the objective system analysis (OSA) and the group method of data handling (GMDH) [9]. In order to improve the prediction accuracy, the paper also made the up sampling method to balance the data. And it made the quantity of the lost and the non-lost customers roughly equal. Cheng Jun researched on the trust prediction method for C2C E-commerce [10]. The author thought that the information forecast had an important significance for the risk of C2C transactions. However, the existing trust models were computed based on the evaluation of transaction, which was a kind of post trust prediction. So, it was more meaningful for the future transactions. The authors proposed a trust predicting method of the future transactions for a merchant to help the consumers to predict the possible degree of the trust in future transactions before trading. Xia Han researched on the E-commerce credit risk prediction [11]. The author combined the accounting indicators of the enterprises with the electronic commerce operation ability to construct the enterprise e-commerce credit risk prediction index. Then, the author used the principal component analysis and the support vector regression machine to predict the degree of the credit risk of the electronic commerce. In addition, there were many scholars studying the e-commerce [12-15].

Lu Lu and other people presented Morpho, which was a modified version of the Hadoop MapReduce framework, which decoupled storage and computation into physical clusters and virtual clusters respectively [16]. When MapReduce computation was performing, the map tasks could get data directly from physical machines without any extra data transfers. Morpho also achieved high performance by two complementary strategies for data placement and VM placement, which could provide better map and reduce input locality. Xuyun Zhang etc. proposed a hybrid approach that combined TDS and BUG together for efficient sub-tree anonymization over big data [17]. Further, they designed MapReduce algorithms for the two components (TDS and BUG) to gain high scalability. Experiment evaluation demonstrated that the hybrid approach significantly improved the scalability and efficiency of sub-tree anonymization scheme over existing approaches. Christoforos Kachris, Georgios Ch. Sirakoulis and Dimitrios Soudris presented a novel scratchpad memory architecture that was used accelerate MapReduce applications by indexing and processing the key/value pairs [18]. Shin-Jer Yang and Yi-Ru Chen conducted a thorough research on one of the critical technologies in cloud computing, the MapReduce programming model [19]. Due to the different computing power and system resources between the nodes, such uniform distribution of tasks would lower the performance between nodes, and hence this paper made improvements on the original speculative execution method of Hadoop (called Hadoop Speculative) and LATE Scheduler by proposing a new scheduling scheme known as Adaptive Task Allocation Scheduler (ATAS).

In this paper, we combine the cloud computing with the electronic commerce. Then we put forward the improved parallel chaos algorithm and forecast the sales volume of the electronic commerce. The structure of this paper is as follows. The first part is the introduction. The second part is the basic knowledge. In this part, we introduce the cloud computing, the C-C algorithm and the weighted first order local prediction method. The

third part is the improved parallel chaos prediction method. The fourth part is the conclusion of the fifth part of the experiment.

2. The Basic Knowledge

2.1 The Cloud Computing

The cloud computing is a new network computing model. It can provide the computing, storage, network and software resource sharing and service by the low cost and high efficiency. The cloud computing is the results that the Utility Computing, Virtualization, Haas, SaaS and PaaS apply. The cloud computing is a revolutionary move. It means that computing power can also be used as a commodity in circulation. Of course, this commodity is transmitted by the Internet. Simply, the cloud computing uses the software of the large data centers on the internet and the ability to process the data to move the complex computing to computer from the user terminal to the cloud.

Hadoop is the cloud computing platform which can be more easily developed and processed in parallel with the large scale data. Its main features are the strong expand capacity, the low cost, the high efficiency and the good reliability etc. The Hadoop platform is composed of two parts. They are Hadoop distributed file system (HDFS) and Map Reduce computing model. MapReduce is the high efficient distributed programming model, and it is also a way to process and generate large scale data sets. For all stages of the MapReduce calculation model, the workflow has input, map, shuffle, reduce, output five stages.

2.2 The Traditional C-C method

For the embedding dimension and the time delay, we adopt the C-C method. The C-C method considers the dependence between the embedding dimension and the time delay. For each time series, the C-C method calculates is as follows.

$$\bar{S}(\tau) = \frac{1}{16} \sum_{m=2}^7 \sum_{j=1}^4 S(m, r_j, \tau) \quad (1)$$

$$\Delta \bar{S}_1(\tau) = \frac{1}{4} \sum_{m=2}^7 \Delta S(m, \tau) \quad (2)$$

$$S_{cor}(\tau) = \Delta \bar{S}_1(\tau) + |\bar{S}(\tau)| \quad (3)$$

Where, $S(m, r_j, \tau) = C(m, r_j, \tau) - C^m(1, r_j, \tau)$ reflects the system's self-correlation characteristic.

r is size of neighborhood radius. $C(m, r, \tau) = \frac{1}{M^2} \sum_{1 \leq j \leq k \leq M} \theta(r - \|X_j - X_k\|)$ is the correlation integral of the system. It describes the degree of correlation of all variables in the system dynamic process. And it reflects that the sophistication degree $\theta(\cdot)$ of the attracted geometric structure is Heaviside function.

$$\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (4)$$

For the estimation of the related dimension, we adopt the G-P algorithm. When $r \rightarrow 0, N \rightarrow \infty$, the associated integral function is approximate to $\log_2 C(m, r, \tau) \rightarrow v \log_2 r$.

When the m becomes bigger, the v reaches the limit value which is the correlation dimension of the system. Then, we select the first zero point of $\bar{S}(t)$ or the first minimum point that the first through $\bar{S}(t)$ as the most appropriate time delay τ . The minimum value point of $S_{cor}(r)$ is the optimal time delay window.

2.3 Weighted First Order Local Prediction Method

We assume that the single variable load time series is $\{x(1), x(2), \dots, x(n)\}$. Therefore, the phase points in the reconstructed phase space are as follows.

$$\begin{cases} Y_{1+(d-1)\tau} = (x(1+(d-1)\tau), x(1+(d-2)\tau), \dots, x(1)) \\ \qquad \qquad \qquad = x_{1+(d-1)\tau}^1, x_{1+(d-1)\tau}^2, \dots, x_{1+(d-1)\tau}^d, \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \vdots \\ Y_i = (x(i), x(i-\tau), \dots, x(i-(d-1)\tau)) = (x_i^1, x_i^2, \dots, x_i^d), \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \vdots \\ Y_n = (x(i), x(n-\tau), \dots, x(n-(d-1)\tau)) = (x_n^1, x_n^2, \dots, x_n^d), \end{cases} \quad (5)$$

In the above formula, Y_i is the phase point in the reconstructed phase space. $i = 1+(d-1)\tau, 2+(d-1)\tau, \dots, n$. τ is the delay time. d is the embedding dimension.

If the embedding dimension satisfies $d \geq 2D+1$ and D is the attraction dimension, the reconstruction system and the original system maintain a diffeomorphism in R^d space. For the deterministic mapping $f^{(d)} : R^d \rightarrow R^d$, there is,

$$Y_{i+1} = f^{(d)}(Y_i) \quad (6)$$

And its equivalent form is,

$$x_{i+1}^1 = \tilde{f}^{(d)}(x_i^1, x_i^2, \dots, x_i^d) \quad (7)$$

In the formula 7, the expressed space of $\tilde{f}^{(d)}$ becomes the reconstructed phase space.

Only $\tilde{f}^{(d)}$ is determined, can it achieve the forecast of the x_{i+1}^1 or $x(i+1)$.

In general, the chaotic attractor has the characteristics of the general stability, the attraction and the internal fractal. The predicted points and the nearest neighbor points follow the similar motion law. According to the determined nearest point (usually using Euclidean distance method), we establish k function relation between the nearest same evolutionary point and the subsequent time series evolution. For example, for the map, we adopt the first order linear approximation,

$$X_{i+1}^i = a + \sum_{i=1}^d b_i x_i^i = a + b_1 x_i^1 + b_2 x_i^2 + \dots + b_d x_i^d \quad (8)$$

In the formula 8, $x_i^1, x_i^2, \dots, x_i^d$ is the each component for the prediction center point Y_i . x_{i+1}^1 or $x(i+1)$ is the prediction measurement. a, b_1, \dots, b_d are the parameters which are to be determined. In general, we adopt the Euclidean distance method to determine the k nearest neighbor point.

In the determined neighborhood, because of the different spatial distances between the neighboring points and the predicted points, the influence of the different points is also different. Therefore, there is,

$$W_j = \frac{\exp(-(d_j - d_{\min}))}{\sum_{j=1}^k \exp(-(d_j - d_{\min}))} \quad (9)$$

Where, $d_{\min} = \min\{d_j, j=1, 2, \dots, k\}$. d_j is the Euclidean distance from the nearest neighbor Y_j to the prediction center point Y_i . k is the number of the neighbor points. w_j is the normalized weight.

Therefore, according to the weighted least square method, the fitting parameters a, b_1, \dots, b_d have,

$$\sum_{j=1}^k W_j (x_{i+1}^1 - a - \sum_{j=1}^d b_j x_i^j)^2 = \min \quad (10)$$

3. Improved Parallel Chaos Prediction Method

In this paper, we propose the improved parallel chaos algorithm to predict the customer churn of the electronic commerce.

Using the traditional C-C method, we can calculate the embedding dimension m and the delay time τ . However, the traditional C-C method has three shortages when calculating the embedding dimension m and the delay time τ .

1. For the time series that the sampling period is T , $t = kT$ is possible to the first zero point for $S(t)$ or the global minimum point for $Scor(t)$. Therefore, it may get the conflicting conclusions when calculating the embedding dimension.

2. For the statistic $S(m, N, r, \tau)$, we adopt Block averaging strategy. When $t = kT$, $S(t)$ is zero. With t becomes bigger, $S(t)$ shows the rising frequency. When the optimal delay time is large, it will influence the selection of the first local minimum point for $S(t)$.

3. Under the ideal conditions, the global minimum point of $Scor(t)$ is the optimal embedding window τ_w . In fact, the $Scor(t)$ exists some local minimum points and the global minimum point. They are quite close in the numerical value. It influences t that optimal embedding window of the global minimum point τ_w corresponding. t may not the global minimum point. Finally, it leads to the error estimation of the optimal embedded window τ_w .

Based on the shortages of the traditional C-C method, we put forwards the following improved C-C method. Then we reconstruct the phase space of the chaotic time series. Finally, we determine the embedding dimension m and delay time τ .

(1) The improved C-C method re-determines the optimal delay time τ_d is,

$$S_1(m, N, r, \tau) = C(m, N, r, \tau) - C'(1, N, r, \tau) \quad (11)$$

Where, $m = 2, 3, \dots$. We choose the largest and the smallest r .

Then, we have the following definition:

$$\Delta S_1(m, N, r, \tau) = \max_{j \in \{1, \dots, N\}} \{ |S_1(m, N, r, \tau_j) - S_1(m, N, r, \tau_{j-1})| \} \quad (12)$$

$$\overline{\Delta S_1}(\tau) = \frac{1}{4} \sum_{m=2}^s \Delta S_1(m, \tau) \quad (13)$$

$$S_{cor}(\tau) = |\overline{S_1}(\tau) - \overline{S_1}(\tau)| \quad (14)$$

Next, we make the first local minimum of $\overline{\Delta S_1}(\tau)$ as the optimal delay time τ . Then, compared with the original method, the first local minimum point is selected. The selection of the new method for τ is more accurate. Especially, when τ_d is bigger, it is more obvious.

(2) For the time series that the sampling period is T , when m, r are fixed and $N \rightarrow \infty$, $t = KT$ is not only the local maximum point of the formula but also the zero point of the formula. Compared with the original C-C method, the period position of $S_{cor}(\tau) = |\overline{S_1}(\tau) - \overline{S_1}(\tau)|$ exists the obvious local peak value. It is more reliable for the selection of the optimal embedding window τ_w . Therefore, the improved C-C method find the period point of $S_{cor}(\tau) = |\overline{S_1}(\tau) - \overline{S_1}(\tau)|$ as the optimal embedding window τ_w , and the embedding dimension is $m = \tau_w / \tau + 1$.

According to the analysis of weighted first order local prediction method, we can know that the Euclidean distance d_j and the weight w_j only reflect the distance between the nearest neighbor and the prediction center. In fact, we hope not that the neighboring points and center distance are close. We hope that the angle between the evolution

direction of the predicted points and the neighboring points is small in order to ensure that they change in the same direction. The correlation degree is evaluated by the similar degree among the curves. The greater the correlation degree is, the better the degree of fitting is. Therefore, the correlation degree between the nearest neighbor points and the prediction center is,

$$\begin{cases} r_{ij}^{(l)} = (\Delta \min + \rho \Delta \max) / (\Delta_{ij}^{(l)} + \rho \Delta \max) \\ r_{ij} = \frac{1}{d} \sum_{j=1}^d r_{ij}^{(l)}, j = 1, 2, \dots, k \end{cases} \quad (15)$$

Where, $\Delta_{ij}^{(l)} = |x_i^l - x_j^l|$, $\Delta \min = \min_{1 \leq l \leq d} \min_{1 \leq j \leq k} |x_i^l - x_j^l|$, $\Delta \max = \max_{1 \leq l \leq d} \max_{1 \leq j \leq k} |x_i^l - x_j^l|$. x_i^l is the first component of the nearest neighbor Y_j . d is the embedding dimension. k is the number of the neighbor points. r_{ij} is the correlation degree between the neighborhood point Y_j and the center point Y_i . Where, $0 \leq r_{ij} \leq 1$. ρ is the distinguish coefficient.

According to the formula (15), we know that if the correlation degree of the two points is bigger, the similar degree among the curves is bigger. Then the prediction accuracy is higher. Similarly, for the formula (12), the Euclidean distance between two points is near, the influence on the prediction point is greater. Therefore, considering the neighbor effects of the nearest neighbor points for the prediction center point, and the correlation between the nearest neighbor points and the prediction center points, the normalized weight of the j neighborhood point Y_j can be expressed as follows.

$$r_{ij} = \frac{r_{ij} w_j}{\sum_{j=1}^k}, j = 1, 2, \dots, k \quad (16)$$

Therefore, according to the weighted least square method, the fitting parameters a, b_1, \dots, b_d have,

$$\begin{aligned} \sum_{j=1}^k r_j (x_{j+1}^1 - a - \sum_{i=1}^d b_i x_j^i) &= 0 \\ \sum_{j=1}^k r_j (x_{j+1}^1 - a - \sum_{i=1}^d b_i x_j^i) x_j^1 &= 0 \\ &\vdots \\ \sum_{j=1}^k r_j (x_{j+1}^1 - a - \sum_{i=1}^d b_i x_j^i) x_j^d &= 0 \end{aligned} \quad (17)$$

Where, k is the number of the nearest neighbor of the prediction center. x_{ij} is the i component of the nearest neighbor Y_j . x_{j+1}^1 is the first component of the one step evolution point Y_{j+1} for the nearest neighbor Y_j . Then we make the above formula into the matrix form.

$$\begin{bmatrix} \sum_{j=1}^k r_j x_{j+1}^1 \\ \sum_{j=1}^k r_j x_{j+1}^1 x_j^1 \\ \vdots \\ \sum_{j=1}^k r_j x_{j+1}^1 x_j^d \end{bmatrix} = \begin{bmatrix} 1 & \sum_{j=1}^k r_j x_j^1 & \cdots & \sum_{j=1}^k r_j x_j^d \\ \sum_{j=1}^k r_j x_j^1 & \sum_{j=1}^k r_j (x_j^1)^2 & \cdots & \sum_{j=1}^k r_j x_{j+1}^1 x_j^d \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^k r_j x_j^d & \sum_{j=1}^k r_j x_{j+1}^1 x_j^d & \cdots & \sum_{j=1}^k r_j (x_j^1)^2 \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_d \end{bmatrix} \quad (18)$$

After simplification,

$$C = AB \quad (19)$$

Where,

$$C = \begin{bmatrix} \sum_{j=1}^k r_j x_{j+1}^1 \\ \sum_{j=1}^k r_j x_{j+1}^1 x_j^1 \\ \vdots \\ \sum_{j=1}^k r_j x_{j+1}^1 x_j^d \end{bmatrix} \quad (20)$$

$$A = \begin{bmatrix} 1 & \sum_{j=1}^k r_j x_j^1 & \cdots & \sum_{j=1}^k r_j x_j^d \\ \sum_{j=1}^k r_j x_j^1 & \sum_{j=1}^k r_j (x_j^1)^2 & \cdots & \sum_{j=1}^k r_j x_{j+1}^1 x_j^d \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^k r_j x_j^d & \sum_{j=1}^k r_j x_{j+1}^1 x_j^d & \cdots & \sum_{j=1}^k r_j (x_j^1)^2 \end{bmatrix} \quad (21)$$

$$b = \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_d \end{bmatrix} \quad (22)$$

The prediction parameter vector is $\hat{B} = (A^T A)^{-1} A^T C$. We can get the forecast value.

$$\hat{x}_{i+1}^1 = \hat{a} + \sum_{j=1}^d \hat{b}_j x_i^j = \hat{a} + \hat{b}_1 x_i^1 + \hat{b}_2 x_i^2 + \cdots + \hat{b}_d x_i^d \quad (23)$$

We set that the optimization problem is to find the minimum value of the objective function.

$$\begin{aligned} \min \quad & f(x_q) \\ & q = 1, 2, \dots, m \quad x_q \in [g_q, h_q] \end{aligned} \quad (24)$$

Where, f is the objective function. x_q is the optimal variables. m are the number of the optimal variables. h_q and g_q are the upper and lower bounds of x_q . Firstly, we initialize the algorithm. We assume that the iteration number is $k=0$. The chaotic sequence number of the optimal variable x_q is $l_q (q=1, 2, \dots, m)$. We follow the logistics map as the chaotic quote generator, it produces $M_1 = \sum_{q=1}^m l_q$ chaotic variables of different trajectories $\{y_{q,n}^p\} (q=1, 2, \dots, m, p=1, 2, \dots, l_i)$. That is,

$$\begin{aligned} y_{n+1} &= 4y_n(1 - y_n) \\ n &= 0, 1, 2, \dots, \quad y_0 \in [0, 1] \end{aligned} \quad (25)$$

Then, according to the following formula, we carrier $y_{q,n}^p$ into the chaotic variable $x_{q,n}^p$ and enlarge the ergodic interval into the optimal variable interval.

$$x_{q,n}^p = g_q + (h_q - g_q) y_{q,n}^p \quad (26)$$

We search iteratively by using the chaotic variables which are obtained from the first carrier. For the M_1 chaotic variable, it produces $M_2 = \prod_{q=1}^m l_q$ parallel feasible solution.

Then, we calculate the objective function value. Then, we take the minimum value as the calculated value $f(k)$ in this step. The Map Reduce process of the algorithm is as follows.

$$\begin{aligned} & \text{if } f(k) < f^*, \text{ then } f^* = f(k), x_q^* = x_q(k) \\ & \text{else if } f(k) \geq f^* \text{ then abandon } x_q(k) \end{aligned} \quad (27)$$

$k = 1$

If the termination criterion is satisfied, the search ends. We output the optimal solution x_q^{**} . If not, we return the formula (25) until the objective function no longer decreases. In the prediction problem, the smaller of the gap between the predicted and the actual value is, the better the prediction results are. The optimization objective function of the prediction problem is as follows.

$$\min E = \sum_{o=1}^N \frac{1}{2} (d_o - y_o)^2 \quad (28)$$

Where, d_o is the actual value and y_o is the predicted value.

Therefore, the above formula is recorded as the objective function of the optimization problem.

The optimal variable x_q is corresponding to the l_q chaotic series $\{y_{q,n}^p\} (p=1,2,\dots,l_q)$. In addition, the parallel search requirements $\{y_{q,n}^p\}$ are different. Because the chaotic variable is sensitive to the initial value, we give the different initial values $y_{q,0}^p$ for the logistics mapping. Then it produce the different chaotic series $y_{q,n}^p$.

$$\begin{cases} y_{q,0}^1 = a \\ y_{q,0}^2 = a + \Delta \\ y_{q,0}^k = a + (k-1)\Delta \\ y_{q,0}^{l_q} = a + (l_q-1)\Delta \end{cases}$$

(29)

Where, $a \in (0,1)$. Δ is a very small positive number.

Initially, the chaotic sequences are very close. According to several iterations, the chaotic sequence will be separated gradually. If we know the difference of the initial of the chaotic sequences is Δ , after iteration, the different value becomes δ . We determine that the two chaotic sequences are completely separated. Then, according to the Lyapunov index λ of the chaotic mapping, we get the iteration steps T .

$$T = \left\langle \frac{1}{\lambda} \ln\left(\frac{\delta}{\Delta}\right) \right\rangle \quad (30)$$

Where, the $\langle \rangle$ expresses that we select the integer.

4. Experiment and Simulation

In order to verify the effectiveness and feasibility of the improved parallel chaos prediction method, we carry out the numerical simulation experiments to verify the effectiveness and feasibility of the method. In this experiment, we use for IBM system X3650 servers, and make each core as a Hadoop node. There are eight nodes. One is the NameNode. Others are the DataNodes of the data sets. We select 1000 electronic commerce stores from 2005 to 2013. There are 108 months. We make these sales data as a training set. We make 24 months of the sales data from 2014 to 2015 as the test set. Firstly, we analyze the computation time of the improved parallel chaos prediction algorithm.

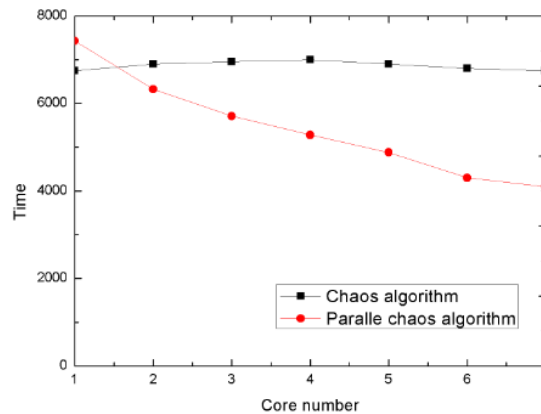


Figure 1. Execution time Contrast Diagram

From the fig.1, we can see that the computation time of the sample is different in the processing under the same data dimension. When we only use a machine to perform, due to the time consuming of communication, the parallel algorithm is time-consuming a little more. However, when the machine group is to expand, we can see that the execution time decreases rapidly. When the group uses are 6 to 7 units, the time is basically stable.

After that, we calculate the accuracy of the algorithm and compare it with other algorithms. The results are shown as follows.

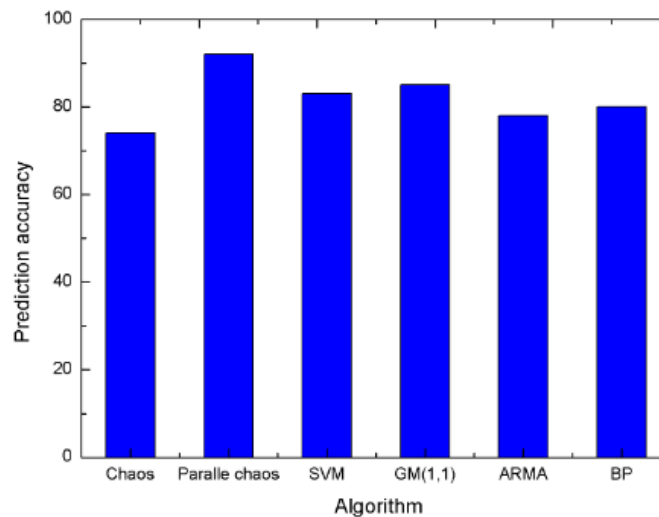


Figure 2. Prediction Accuracy Comparison Chart

From the fig.2, we can see that, compared with other algorithms, the prediction accuracy of the parallel chaos algorithm is the highest and it reaches to 92%. On the contrary, the other prediction results are not very ideal. It repeats to show that the feasibility and effectiveness of the parallel algorithm.

5. Conclusion

The rapid development of the e-commerce provides a strong driving force for Chinese economic take-off. The cloud computing has the characteristics of the low cost, the large scale, and it has a very good combination with the electronic commerce. It can effectively improve the speed and efficiency of computing by using the cloud computing to study the relevant knowledge of the e-commerce. In this paper, we use the parallel chaos algorithm to study the sales forecast of the electronic commerce and obtain a good result. The main

contents of this paper are as follows. Firstly, in this paper, we introduce the research background of the related knowledge. Secondly, we introduce the related basic algorithm. Thirdly, we propose that the improved parallel chaos algorithm. Fourthly, we apply the improved chaos parallel algorithm to predict the sales of the electronic commerce. Finally, the experimental results show that the algorithm is feasible and effective.

Acknowledgment

The Scientific Research Foundation of Heilongjiang Provincial Education Department of China (NO: 125HQ003)

References

- [1] Dimitris Potoglou, Juan-Francisco Palacios, Claudio Feijóo. An integrated latent variable and choice model to explore the role of privacy concern on stated behavioural intentions in e-commerce [J]. *Journal of Choice Modelling*, 2015, 17: 10-27.
- [2] Neil Towers, Kiki Xu. The influence of guanxi on physical distribution service quality availability in e-commerce sourcing fashion garments from China [J]. *Journal of Retailing and Consumer Services*, 2016, 28: 126-136.
- [3] Giovanni Acampora, Daniyal Alghazzawi, Hani Hagrass, Autilia Vitiello. An interval type-2 fuzzy logic based framework for reputation management in Peer-to-Peer e-commerce [J]. *Information Sciences*, 2016, 333: 88-107.
- [4] Muhammad Aljukhadar, Sylvain Senecal. The user multifaceted expertise: Divergent effects of the website versus e-commerce expertise [J]. *International Journal of Information Management* 2016, 36(3): 322-332.
- [5] Georg Lackermair. Hybrid cloud architectures for the online commerce [J]. *Procedia Computer Science*, 2011, 3: 550-555.
- [6] Zhang Qiang. The study of collaborative forecasting on supply chain in electronic business [D]. Xi'an Electronic and Science University, Management science and Engineering, 2005.
- [7] Zhu Bangzhu, Zhang Qiuju. E-business customer churn prediction based on three-step integration of SMC and least square support vector machine [J]. *Chinese Soft Science*, 2010, 6: 186-192.
- [8] Zhu Bangzhu. E-business customer churn prediction based on integration of SMC rough set and least square support vector machine [J]. *System engineering theory and Practice*, 2010, 11: 1960-1967.
- [9] Zhu Bangzhu, Zhang Qiuju, Zhou Haofei, Wei Yiming. E-business customer churn prediction based on integration of object system and group method of data handing network [J]. *Chinese Journal of Management Science*, 2011, 11: 64-70.
- [10] Cheng Jun. Research of trust prediction method for C2C E-commerce [D]. Nanjing Normal University, School of computer science and technology, 2014.
- [11] Xia Han. E-commerce credit risk prediction base on PCA-SVR [J]. *Journal of Modern Information*, 2015, 35: 76-79.
- [12] Liu Xuewei. Research for eBusiness customer churns based on assembled model of Pareto/NB and Naïve Bayes [D]. Sichuan University, Management Science and Engineering, 2006.
- [13] Shen Peilan. Research of the model of customer prediction base on E-Business [D]. Jiangsu University of Science and Technology, Management Science and Engineering, 2011.
- [14] Zi Zhidao. Forecasting and analysis of cross border electronic commerce based on GM(1,1) [J]. *Mathematics in practice and theory*, 2015, 1: 97-105.
- [15] Zhang Yanrong, Yu Jing, Zhao Zhijie. Research on prediction of E-commerce consumer behavior based on rough set theory [J]. *Commercial research*, 2014, 452: 101-179.
- [16] Lu Lu, Xuanhua Shi, Hai Jin, Qiuyue Wang, Daxing Yuan, Song Wu. Morpho: A decoupled MapReduce framework for elastic cloud computing [J]. *Future Generation Computer Systems*, 2014, 36: 80-90.
- [17] Xuyun Zhang, Chang Liu, Surya Nepal, Chi Yang, Wanchun Dou, Jinjun Chen. A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on cloud [J]. *Journal of Computer and System Sciences* 2014, 80(5): 1008-1020
- [18] Christoforos Kachris, Georgios Ch. Sirakoulis, Dimitrios Soudris. A MapReduce scratchpad memory for multi-core cloud computing applications [J]. *Microprocessors and Microsystems*, 2015, 39(8): 599-608.
- [19] Shin-Jer Yang, Yi-Ru Chen. Design adaptive task allocation scheduler to improve MapReduce performance in heterogeneous clouds [J]. *Journal of Network and Computer Applications*, 2015, 57: 61-70.