Research on Modern Higher Education 2, 01013 (2017)
DOI: 10.24104/rmhe/2017.02.01013

Research on cost early warning of textile industry based on AC algorithm

Xiaolu Meng*
Fuzhou College of Foreign Studies and Trade, Fuzhou, Fujian, China

ABSTRACT: In recent years, the textile industry faces huge pressure domestically and internationally. With a small profit margin in the textile industry, it becomes very important to improve the economic efficiency and the cost management. Therefore, establishing an effective cost early warning system for the textile industry is necessary. The proposal of AC algorithm predicts the cost alert completely based on the historical data without estimating or assuming the future trend of the input variables in advance. Based on its advantages, the length of pattern (k) is 2, 3, 4 or 5, and then we respectively build the cost early warning model. According to the results of comparison, the model with k =3 has the strongest predictive ability. The predictive error of the total cost is 1.8% smaller than that of others. Then, this paper predicts the cost alert of the next five years by the model with k =3. The results show that the cost alert has certain volatility. Accordingly, we proposed some suggestions to cope with the crisis. This paper also provides references for the industry, government and relevant investors.

Keywords: cost early warning; AC algorithm; textile industry; self-organizing data mining method

1 INTRODUCTION

As a traditional pillar industry of the national economy in our country, the textile industry has made great contributions to the social development. However, with the rapid development of modern industry, the market is increasingly competitive. The ministry of industry and information technology of China formulated and issued the “Twelfth Five-Year Development Plan of Textile Industry”, pointing out that the textile industry faced huge cost pressure in the international and domestic. For the textile industry with small profit margins, cost management becomes very important. Cost management is an important component of enterprise management. Building the early warning system of cost management, that conforms to the enterprise production operation mode, is not only beneficial to the enterprise to control the cost risk and reduce the cost, also improves enterprise management level and market competitiveness[1-2].

As the saying goes, “Preparedness ensures success, unpreparedness spells failure”. Building an effective cost early warning system is very necessary. Cost early warning is based on cost forecast to get the information of costs alerts. By analyzing the alerts, the corresponding solutions are proposed. With the wide application of data mining technology, most scholars applied the technology to the research cost.

For example, some scholars used the association rule algorithm or the fuzzy rough set algorithm to deal with the problems of variables related with the cost [3-4]. Zhou Chao et al used the grey model (GM) [5] and made some improvements [6], such as combining GM with the nonlinear regression model, to predict the operation cost of Oil and gas. It helped the enterprises control the cost. Also, some scholars estimated the production cost based on support vector machine (SVM) [7-8], predicted the logistics cost based on BP neural network [9], and controlled the quality cost based on regression analysis, correlation analysis and fuzzy neural network [10]. Zhenyu Cai and Xingmin Ma analyzed the factors affecting coal production costs based on Wavelet neural network, and used SVM method to establish the model to predict the production cost of coal [11].

The research and application of these methods have been promoted the improvements of the researches on cost. But, generally, these studies are from the microscopic angle to research, such as studying the enterprise operating costs and the cost of a project. Re-

*Corresponding author: meng_xiaolu@126.com

searching costs early warning from the macroscopic angle is less. So the research scope is narrow and the applicability is not strong. Early only James B. Edwards from the macro point of view pointed out that the combination of data mining and cost management could achieve the advanced intelligent development of cost management and improve the cost competitiveness of industries. But unfortunately, he hadn't given a specific model of data mining combining with cost management. So, studying the cost alert from the macro point of view is of great significance. Because of wide range of the research, which involves in many influencing factors and unstable variations, the methods used before are not entirely suitable for this research. So we use self-organizing data mining method to solve this problem. This self-organizing data mining can overcome the problems of neural network, such as hardly explaining the results of practical significance, over-fitting, and also have the advantages of using a priori knowledge of system and inductive learning algorithm process. Meanwhile, the method is applicable to the complex system that has many interconnected influencing factors, especially the system exists noise. So, it is widely useful [12–15].

The self-organizing data mining methods includes parameter algorithm and non-parametric algorithm [16]. Among them, AC algorithm is one of non-parametric algorithms. Compared with the parameter algorithm, AC algorithm does not need to estimate in advance or assume the future trend of input variables, and it predicts completely according to the historical data. Therefore, AC algorithm is most suitable for this research.

2 THE BASIC PRINCIPLE OF AC ALGORITHM

Analog Complexing Algorithm referred to as AC algorithm, was proposed by Lorence [17]. When applying AC algorithm to the prediction, it's necessary to meet the assumptions:

First, the system is a multidimensional process.

Second, the observations are valid during the process.

Third, the multidimensional process is fully representative, namely the data set consists of the basic variables of the system.

Fourth, over a period of time, the behavior of the process will be repeated.

The steps of AC algorithm are as follows:

Step 1: the generation of candidate patterns.

Assuming: $x_i = \{x_{1t} \dots x_{mt}\}$, ($t$ =1, 2...N). Each row of the matrix represents a sample data; each column of the matrix represents an index. The pattern, $P_k(i)$ (i=1, 2…N-K+1) is a matrix that starts from i row to $i+k-1$ row, which totals k rows. And the k is named as the length of a pattern.

$$P_k(i) = \begin{bmatrix} x_{1i} & \cdots & x_{li} & \cdots & x_{mi} \\ \vdots & & \vdots & & \\ x_{1,i+j} & \cdots & x_{l,i+j} & \cdots & x_{m,i+j} \\ \vdots & & \vdots & & \vdots \\ x_{1,i+k-1} & \cdots & x_{l,i+k-1} & \cdots & x_{m,i+k-1} \end{bmatrix}_{k\times m} \quad (1)$$

Because AC algorithm uses the combination of continuation of similar pattern as the development of state of the reference pattern, so the prediction interval should be the same as the continuation of reference pattern. And we choose a known pattern that is the nearest neighbor point of prediction as a reference pattern, namely $P^R = P_k(N-k+1)$.

Step 2: converting the candidate patterns.

According to the principle of AC algorithm, it is possible one or several patterns are similar to the reference pattern. Their length of the patterns is k. However, due to the dynamic system, it may generate different standard variance, average between similar patterns and reference pattern during the different periods. If it happens, in order to make them be compared, it is necessary to convert the candidate patterns to the same as the reference pattern, and measures the similarity between the patterns. Generally, the linear method is taken as the way of converting between patterns. Patterns are converted as following:

$$T_i\left[P_k(i)\right] = \begin{bmatrix} x^*_{1i} & \cdots & x^*_{li} & \cdots & x^*_{mi} \\ \vdots & & \vdots & & \\ x^*_{1,i+j} & \cdots & x^*_{l,i+j} & \cdots & x^*_{m,i+j} \\ \vdots & & \vdots & & \vdots \\ x_{1,i+k-1} & \cdots & x^*_{l,i+k-1} & \cdots & x^*_{m,i+k-1} \end{bmatrix} \quad (2)$$

In the equation (2), $x^*_{l,i+j} = a^i_{0l} + a^i_{1l} x_{l,i+j}$, ($j$=0,1…$k$-1, i=1,2…N-k+1, l=1,2…m).

a^i_{0l} is the gap between the similar patterns and P^R. a^i_{1l} is an uncertain parameter. With the data of reference pattern as a target, it can use the least squares to estimate each candidate pattern.

$$T_i\left[P_k(i)\right] = \begin{bmatrix} x_{1i} & x_{2i} & \cdots & x_{mi} \\ x_{1,i+1} & \ddots & & x_{m,i+1} \\ \vdots & & \ddots & \vdots \\ x_{1,i+k-1} & x_{2,i+k-1} & \cdots & x_{m,i+k-1} \end{bmatrix} \times \begin{bmatrix} a^i_{11} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & a^i_{1m} \end{bmatrix} + \begin{bmatrix} a^i_{01} & a^i_{02} & \cdots & a^i_{0m} \\ a^i_{01} & \ddots & & a^i_{0m} \\ \vdots & & \ddots & \vdots \\ a^i_{01} & a^i_{02} & \cdots & a^i_{0m} \end{bmatrix} = P^R \quad (3)$$

According to equation (3), it will estimate a_{0l}^{i} and a_{1l}^{i}, and calculate the sum of error squares to judge the similarity between the patterns.

Step 3: the selection of similar patterns.

This step mainly identifies the similarity between the patterns. To measure the similarity between P^R and the converted $P_k(i)$, it needs to calculate the distance between the i-th candidate pattern and the reference pattern, and judge the degree of similarity between the patterns by Euclidean. The step goes as following:

$$d = \frac{1}{k+1}\sum_{j=0}^{k-1}\sqrt{\sum_{r=1}^{m}(x_{r,i+j} - x_{r,N-k+j+1})^2} \qquad (4)$$

Then, the degree of similarity between the patterns is $s_i = \frac{1}{d_i}$.

Step 4: combining the continuation of similar patterns to predict.

Assuming: selecting F similar patterns, they are $A_1, A_2, ..., A_F$. Reference pattern is P^R, and the corresponding predicted variables are $Y_1, Y_2, ..., Y_F$ and Y^R. Where, $Y_1, Y_2, ..., Y_F$ and Y^R are k-dimensional column vector. Then, the prediction problem turns into the problem of building model. The model has k samples and F variables that are one dependent variable (Y^R) and six independent variables ($Y_1, Y_2, ..., Y_F$). However, whether k is greater or smaller than F, it's hard to get good results by conventional regression method. Thus, it needs to use the GMDH method to solve this problem. Because the GMDH method is good for modeling with limited data samples, the linear GMDH model is written as following:

$$Y^R = \sum_{j \in J} g_j Y_j \qquad (5)$$

Where, J is a subset of the set {1, 2…F}, {g_i} is the weight coefficient of the patterns. If $g_i = 0$, it indicates that Y_j does not appear in the model. When $k \times m$ is small, the weight coefficient of the patterns will be the same, and totals 1.

$$y_{N+1} = \sum_{j=1}^{F} g_j y_{j+k+i-l} \text{ , } (i=1,2\ldots i) \qquad (6)$$

Where, i is the prediction range. y_{N+1} and $y_{j+k+i-l}$ are the continuation of predicted variables of reference pattern and similar pattern. Then it will predict the results by combining the F continuation of similar patterns.

3 APPLICATION OF AC ALGORITHM IN THE RESEARCH ON COST EARLY WARNING

3.1 *Determining the research object and relevant variables*

Because of researching on the cost early warning of textile industry, so this paper selects the total cost (y) as the research object. And the selection of relevant variables should follow the principles, such as comparability, representativeness, comprehensiveness, operability and so on. Because of the influence of the internal and external environment on the textile industry, the affect factors of cost mainly include the national policy, the macroeconomic regulation and control, changes in market demand, economic development, the management level of textile industry, technology, environmental pollution, the production of industrial products, etc. Thus, selecting 18 indexes as the relevant variables are as shown in the Table 1:

Table 1. 18 variables

Variable symbol	Variable	Unit
x_1	Gross output value of textile industry	100 million yuan
x_2	Exchange rate of RMB	RMB yuan
x_3	Registered unemployment rate in urban areas	%
x_4	Household consumption expenditure	yuan
x_5	Cost of principal business	100 million yuan
x_6	Accumulated depreciation	100 million yuan
x_7	Taxes and other charges on principal business	100 million yuan
x_8	Total exports of textiles and textile articles	USD 100 million
x_9	Total imports of textiles and textile articles	USD 100 million
x_{10}	Total liabilities of textile industry	100 million yuan
x_{11}	Revenue from principal business	100 million yuan
x_{12}	Total wage bill of textile industry	100 million yuan
x_{13}	Total investment in fixed assets	100 million yuan
x_{14}	Total volume of industrial waste water discharge	10 000 tons
x_{15}	Total energy consumption	10 000 tons of SCE
x_{16}	Total volume of industrial waste gas emission	100 million cu.m
x_{17}	Purchasing price indices for fuels and power	Preceding year=100
x_{18}	Purchasing price indices for textile materials	Preceding year=100

3.2 *Selecting and processing sample data*

3.2.1 *Selecting sample data*

In order to study the impact of variables on the research object, this paper chooses the sample interval from 1980 to 2012. The annual data of each variable gets from *China Statistical Yearbook*, *China Industrial Economic Statistical Yearbook*, *China Industrial Economy Statistical Yearbook*, *Almanac of China's Textile Industry* and *China Urban Life and Price Yearbook*.

3.2.2 *Processing sample data*

In the sample data, there are exception values that should be corrected. And the missing values will be filled by smooth index method.

3.3 *Diving the warning limit and warning degree*

Warning degree, namely the size of alert, is used to measure the severity of alert. Generally, warning degree can be divided into warning zone and no warning zone. The no warning zone means the total cost occurred is in the reasonable range, which is normal. The warning zone means the total cost occurred is high or low, that is abnormal. According to the magnitude of deviation, the warning zone will be divided into four zones, including higher, high, low, lower. High cost affects industry enterprise's profit and restricts their development. Low cost can help the industry to save resources and increase profits, but it doesn't mean that the lower, the better. If the industry blindly reduces investment to save costs, although it can bring short-term benefit, in the long view it's not good for the development of the industry. Thus appropriately increasing the cost can help the industry get higher profits, and promote its development.

Therefore, the warning degree will be divided into five grades, which are lower, low, normal, high and higher. Warning limit is the critical value between the grades. Based on the data set of total cost index from 1980 to 2012, the annual growth rate of cost is calculated. Then on the basis of systematic method, that is following the principles of the minority, the majority, the half and the mean to divide the warning limit. The result is in the Table 2.

3.4 *Filtering indicator variables*

The purpose of filtering indicator variables is to sort out the most predictive indicator variables, and delete variables with weak predictive ability. Meanwhile, it can eliminate collinearity among the variables. Thus the paper uses stepwise regression method to filter the variables. Taking the confidence level with 0.05, then with the variables whose p-value is less than 0.05 are significant variables, we reserve. The others are deleted. Table 3 only selects the variables whose p-value is less than 0.05.

Table 3. Variables with p-value<0.05

Variables	x_1	x_5	x_6	x_{12}	x_{16}	x_{17}
p-value	0.0008	0	0	0.0031	0.0001	0.0486

According to the result of Table 3, x_1, x_5, x_6, x_{12}, x_{16} and x_{17} are reserved as following research.

3.5 *Application of AC algorithm*

3.5.1 *Establishing model*

To reduce the impact caused by the dimensionless inconsistence and data oscillation, it needs the variable for indexation processing. The formula is as follows:

$$x_i(t) = \frac{x_i^0(t)}{x_i^0(t-1)} \times 100 \tag{7}$$

Where, $x_i^0(t)$ is the value of x_i in the t year. $x_i(1)=100$, $i = 1,2\ldots7$, $t = 1,2\ldots32$.

The length of pattern (k) is an important parameter in the AC algorithm. The different models will be built by different k. Taking into account the studied variables are affected by macroeconomic factors, and Chinese economic growth exists cyclical fluctuations for 4~5 years. So, the value of k is selected from 2 to 5. When k =2, 3, 4, 5, the data of variables from 2010 to 2011, from 2009 to 2011, from 2008 to 2011 and from 2007 to 2011 are respectively chosen as the data of variables of corresponding reference pattern. The other N-k patterns are chosen as candidate patterns. The size of sample (N) is 32, and the data of variables is set as the actual value for the comparison of that in 2012.

Based on the theory of AC algorithm, the prediction of the cost alert of textile industry is conducted by MATLAB. The range of prediction is 1. Table 4 is the results of the prediction.

The data of Table 4 are anti-normalized. Then comparing the predictive value of variables ($\hat{y}_i$) with the actual value (y_i), under the different k the corre-

Table 2. The division of warning limit of cost growth rate

Principle	Lower (%)	Low (%)	Normal (%)	High (%)	Higher (%)
Principle of the minority	(-∞, 8.99]	(8.99, 11.54]	(11.54, 16.59]	(16.59, 20.28]	(20.28, +∞)
Principle of the majority	(-∞, 4.97]	(4.97, 7.12]	(7.12, 23.27]	(23.27, 24.24]	(24.24, +∞)
Principle of the half	(-∞, 7.12]	(7.12, 8.99]	(8.99,20.28]	(20.28,23.37]	(23.37, +∞)
Principle of the mean	(-∞, 6.34]	(6.34, 10.51]	(10.51,18.84]	(18.84,23]	(23, +∞)
Weighted average	(-∞, 6.86]	(6.86, 9.54]	(9.54,19.75]	(19.75,22.72]	(22.72, +∞)

Table 4. Predictive value of variables in 2012 (normalized)

Length of pattern	Predictive value after indexation processing (%)						
	y	x_1	x_5	x_6	x_{12}	x_{16}	x_{17}
k =2	115.13	114.92	119.21	184.82	116.01	159.65	106.49
k =3	111.64	105.40	107.65	113.88	106.63	100.29	87.96
k =4	119.85	119.77	119.93	118.42	117.74	106.24	123.47
k =5	118.37	118.29	118.44	117.00	116.35	105.38	121.81

sponding relative errors are calculated by the formula $\frac{|\hat{y}_i - y_i|}{y_i} \times 100\%$, the results are shown as Table 5:

Table 5. The relative error in different k

Variables	The relative error（%）			
	k = 2	k = 3	k = 4	k = 5
y	4.98	1.80	9.11	7.93
x_1	1.36	9.53	2.80	1.53
x_5	19.76	8.15	20.48	18.99
x_6	72.59	6.35	10.58	9.25
x_{12}	0.12	8.20	1.36	0.16
x_{16}	58.68	0.32	5.60	4.74
x_{17}	16.97	3.38	35.62	33.81

It can be seen from Table 5, when k =2, the predictive errors of x_5, x_6, x_{16}, x_{17} are bigger than others, the maximum error of 72.59%, so the early warning model of cost with k =2 has not good prediction effect. When k =4 or 5, the predictive errors of x_5 and x_7 are bigger, so the corresponding early warning models are not suitable for predicting. When k =3, among the predictive value of variables, the maximum relative error is 9.53%, the minimum is 0.32%. The errors are within the range we can accept. And this paper studies the cost early warning, so the accuracy of predicting total cost is critical. The relative error is 1.80% by the early warning model of cost with k =3 predicting, it's smaller than the prediction errors of other models. In conclusion, the early warning model of cost with k =3 has a good prediction effect. Therefore we use the early warning model of cost with k =3 to predict.

3.5.2 *Predicting the alert of cost*

Observed values with 33 years, from 1980 to 2012, are used as the data samples to predict the alert of cost. The k is 3. The range of prediction is 5, namely predicting the alert of cost in five years from 2013 to 2017.The results are as shown in Table 6.

From Table 6, it's known that the warning degree of cost is in the "normal" zone in 2013, 2015 and 2016, which means the amplitude of the three-year cost growth is in a relatively reasonable range. The warning degree of cost is in the "low" zone in 2014 and 2017, which indicates the amplitude of the two-year cost growth is relatively low. But the alert is not serious. For this situation, we propose appropriate measures to resolve the situation by observing abnormal factors, the development of the industry and the relevant national policy. Overall, the total cost of textile industry will increase in the next five years.

Table 6. The strength of cost alert in 2013-2017

Year	Total cost (y)			
	Predictive value normalized	Predictive value anti-normalized	Growth rate of cost (%)	Warning degree
2013	110.22	36900.61	10.22	normal
2014	106.96	39470.31	6.96	low
2015	117.25	46278.59	17.25	normal
2016	111.60	51648.19	11.60	normal
2017	107.30	55419.74	7.30	low

4 ANALYSIS OF COST ALERT AND SUGGESTIONS

4.1 *Analysis of cost alert*

From the analysis above, in the next five years the total cost of the textile industry is in a state of sustained growing, and its growth rate has certain volatility. It may be caused by some reasons, such as the increased labour costs, inflation, changes in exchange rates, equipment updating, technical innovation and so on. These reasons will bring great cost pressure for industry. Meanwhile, the textile industry is a heavy pollution industry, with the promulgation of "Discharge Standards of Water Pollutants for Dyeing and Finishing of Textile Industry" and the implementation of the policy about energy saving and emission reduction, the investment cost, governance cost and irregularities cost are increasing.

4.2 *Suggestions*

For controlling the cost and promoting economic growth of the industry, the textile industry needs to speed up technological upgrading, improve the capability of independent innovation, enhance energy conservation and emissions reduction, improve resource recycling, optimize the industrial chain, transform of the mode of economic development and deal with the rising labor costs. While the government needs to strengthen macro-control and stable RMB exchange rate to promote the development of industry.

5 CONCLUSION

Based on the limitations of data collection, for the common data mining methods, it's hard to directly predict the future development trend of cost by historical data and the uncertain future development trend of input variables. But AC algorithm can do it. AC algorithm has great advantages to solve these problems. So the establishment of cost early-warning model based on AC algorithm is effective and reliable, and its predicted results can be used as decision-making for the industry, government and relevant investors.

However, this paper still has some deficiencies: in the paper, the similarity of pattern is measured by Euclidean, which is more sensitive to noise. So it needs to further optimize the measurement method. Meanwhile, the length of pattern has great impact on the establishment of model, but its selection hasn't been reached an agreement until now. So it also needs to be further discussed.

ACKNOWLEDGEMENT

The author is sincerely grateful to those who make suggestions for the paper.

REFERENCES

[1] Xiaoqing Zhang. 2013. Discussion on the establishment of cost management and early warning control system. *Shanxi Youth*, (3): 47-48.

[2] Yong Gang Nie & Pei Jing. 2013. Discussion on the establishment of enterprise cost management early warning system. *Advanced Materials Research*, (622):1860-1863.

[3] Hara Kostakis, George Boskou & George Palisidis. 2011. Modeling activity-based costing in restaurants. *Journal of Modeling in Management*, (6): 243-257.

[4] Shuyu Zi, Ruxiang Wei & Mingchi Lin. The method of arguments reduction method in cost forecasting system. *Statistics & Decision*, 366(18): 63-66.

[5] Zhou Chao & Wang Xiuzhi. Application of combination forecasting model in oil and gas operation cost prediction. *Journal of Southwest Petroleum University*, (4): 1-4.

[6] Yu Zude & Song Zhaoxia. 2003. Application of grey model in forecast of oil and gas operation cost. *Techno-economics in Petrochemicals*, (6): 3-9.

[7] Js Chou, My Cheng & Yw Wu, et al. 2011. Predicting high-tech equipment fabrication cost with a novel evolutionary SVM inference model. *Expert Systems with Applications*, 38(7): 8571-8579.

[8] Feng Xue, Shang Gao. Cost model on batch manufacturing of weapon system using SVM. Ship Science and Technology. 2013, (1): 126-130.

[9] Yi Chen & Meifang Wan. The application of RBF neural networks in logistical system. *Computer Simulation*, (4): 159-162.

[10] Guijiang Duan, Yi Yan & Yang Wang. 2013. Analysis and control of quality cost based on data mining. *Computer Integrated Manufacturing Systems*, 19(7): 1693-1703.

[11] Zhenyu Cai & Xingmin Ma. 2010. Research of coal manufacturing cost based on SVM-wavelet neural network. *E-product E-service and E-entertainment*, (2): 1-4.

[12] He Changzheng, Zhang Bi, Yu Hai. 2002. The Comparison between the self-organizing data mining and artificial neural network. *System Engineering-Theory & Practice*, 22(11): 11-14.

[13] Vakhnenko A G & Ivakhnenko G A. 1995. The review of problems solvable by algorithms of the group method of data handling (GMDH). *Pattern Recognition and Image Analysis*, 5(4): 527–535.

[14] A. G. Ivakhnenko, G. A. Ivakhnenko & N. M. Andrienko. 1998. Inductive computer advisor for current forecasting of Ukraine micro economy. *Systems Analysis Modeling Simulation*, 31(2): 143-151.

[15] Yu. P. Zaychenk & I. O. Zayets. 2003. The fuzzy group method of data handling and its application to the problems of the macroeconomic indexes forecasting. *Systems Analysis Modeling Simulation*, 43(10): 1321-1329.

[16] J.A.Müller & F.Lemke. 2000. *Self-organizing Data Mining*. Berlin: Libri Books, pp: 21-25.

[17] Lorence, E. N. 1969. Atmospheric predictability is revealed by naturally occurring analogues. *J. Atmos. Sc*i, (4): 636-646.