ISSN : 0974 - 7435 Volume 10 Issue 13

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(13), 2014 [7073-7080]

Study on the pricing mechanism of coal enterprises based on time series model

Zhang Zhaoyin
School of Computer Science and Technology in Heilongjiang University, 150010, (CHINA)
E-mail : zhaoyingaga@126.com

ABSTRACT

In this paper, a structure-adaptive piece-wise linear segments representation of time series is proposed. The algorithm can automatically produce the K piece-wise segments of time series, which can approximate the original time series. The time series prediction method based on support vector machine is proposed to predict coal price. The experiment results show that each error index of our proposed algorithm is less than that of any other model. The predictive accuracy is improved greatly, which can provide reference for pricing mechanism of coal enterprises.

KEYWORDS

Coal; Support vector machine; Accuracy.

INTRODUCTION

Coal is main fuel for enterprises of iron and steel, the price of coal is important factor for enterprises to decide how to purchase fuel[1,2]. Basing on this background, predicting the price of coal is very necessary[3-5]. Recently the study on data mining of time series mainly concentrates on both the similarity search in a time series database and the pattern mining from a time series[6-9]. In the time series similarity search, due to the high dimensionality of the data, an efficient technique for the similarity computation is very anxious. In the pattern mining the trend prediction is a new domain. It extracts the static attributes from a time series that are the most important predicting attributes. Those attributes can be created a static database. Then, a high generalized classification technique can be used to mine rules from the time series. The definition of similarity between object and measure research is of great significance in the fields of statistical theory, machine learning and data mining and so on. In the time series data mining research, similarity search is an important research content, this is mainly driven by a practical need.

In the similarity research, an efficient representation is a key of descreasing the burden of the similarity computation. In this paper, we present a structure-adaptive piece-wise linear segments representation of time series. The algorithm can automatically produce the K piece-wise segments of time series, which can approximate the original time series.

The paper is organized as follows. In the next section, A structure adaptive segmented linearization description method is given, including structure adaptive time series segmented linearization description and time series similarity measurement. In Section 3, the time series prediction method based on support vector machine[10-16] is proposed. In Section 4, in order to test the performance of proposed algorithm, coal price is predicted using four algorithms. Finally, we conclude our paper in section 5.

A STRUCTURE ADAPTIVE SEGMENTED LINEARIZATION DESCRIPTION METHOD

Structure adaptive time series segmented linearization description

Piecewise linearization is that the data of time series can be divided into linear sequence, which is made up of a series of approximate straight line to replace the original raw data. The key problem of piecewise linearization is how to select the appropriate number of segments K of straight line, using the K line segment to approximate the original time series. If K value is too small, it will lose useful information. On the contrary, if the K value is too large, it will produce too many redundant information. This problem is solved as follows.

If time series S has n number of points and each value of S is function of time, then S can be expressed as

$$S_t = y(t) \quad t = 1, 2, \cdots, n \tag{1}$$

$(t_i, y(t_i))$ represents landmark, $\frac{dy(t)}{dt}|_{t=t_i} = 0$. Time sequence points between two landmarks can be estimated by straight line.

$$\alpha_i = \frac{y(t_{i+1}) - y(t_i)}{t_{i+1} - t_i} \tag{2}$$

The approximation error set of this area is $\{\beta_j, j = 1, 2, \cdots, M\}$ and M represents the number of $y(t)$ between t_i and t_{i+1}.

$$\beta_j = y(t_{i+j}) - (y(t_i) + \alpha_i \cdot (t_{i+j} - t_i)) \tag{3}$$

The approximation error of the whole straight line segment is

$$B_i = \frac{\sqrt{\sum_{j=1}^{M} \beta_j^2}}{M} \tag{4}$$

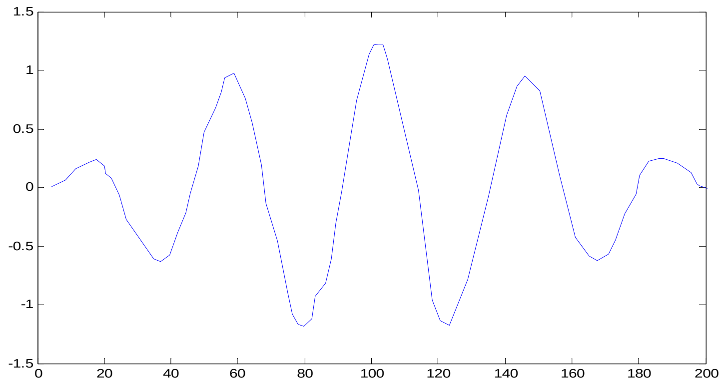

Figure 1 : The original time series model

In order to test the effectiveness of artificial time series, the original time series model is shown in Figure 1, model with additive Gaussian noise is shown in Figure 2 and piecewise linearization result of artificial time series is shown in Figure 3.

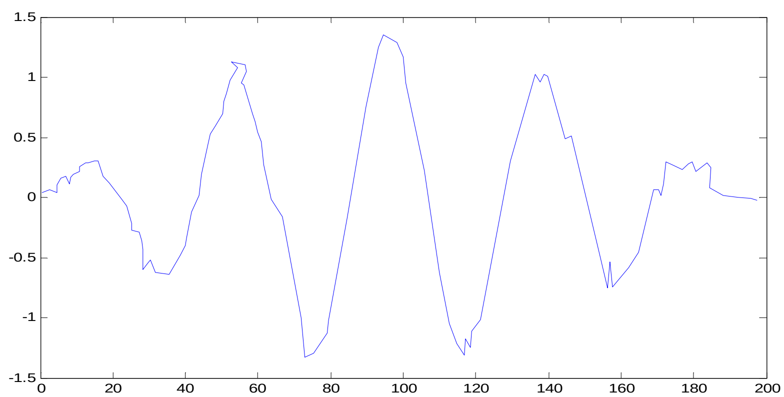

Figure 2 : Model with additive Gaussian noise

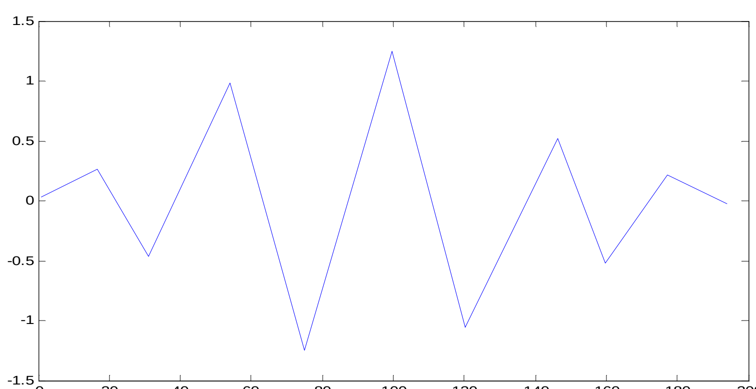

Figure 3 : Piecewise linearization result of artificial time series

Time series similarity measurement

After piecewise linearization, the original time series will be composed of a series of approximate straight segment end to end. A original time series not divided is labeled by capital English letters, such as S. After the piecewise linearization, time series is represented by boldface capital English letters S, the length of which is K and it contains four variables, STL, STR, SYL, SYR. The i-th straight line is represented by the left end (STL_i, SYL_i) and the right end (STR_i, SYR_i). A time series can be represented by the function $y(t)$, all kinds of deformation can be roughly defined as the following.

(1) $NO(y(t)) = y(t) + e(t)$, $e(t)$ represents random noise.

(2) Shifting operation is $SH(y(t)) = y(t) + C_s$, C_s represents a constant.

(3) Amplitude scaling is $AS(y(t)) = c_A y(t)$, c_A represents a positive constant.

(4) Time scaling is $TS(y(t)) = y(c_t t)$, c_t represents a positive constant.

(5) Linear drift is $LD(y(t)) = y(t) + L(t)$, $L(t)$ represents linear equation.

Below based on piecewise linearization we will define a simple formula of similarity measurement, the formula is not sensitive to deformation. Time series after segmentation is S is normalized, the value of which is between -1 and +1. The value of some end is Y, after normalization it is

$$\overline{Y} = \frac{Y - \frac{Y_{\min} + Y_{\max}}{2}}{\frac{Y_{\max} - Y_{\min}}{2}} \tag{5}$$

$Y_{\min} = \min(SYL, SYR)$, $Y_{\max} = \max(SYL, SYR)$. Similarity distance formula between time series A and B is defined as follows.

$$D(A,B) = \sum_{i=1}^{K} \left| \frac{(AYR_i - AYL_i)}{(ATR_i - ATL_i)} - \frac{(BYR_i - BYL_i)}{(BTR_i - BTL_i)} \right| \tag{6}$$

The formula meets the following inequality defined in accord with the distance.

$$D(A,B) = D(B,A) \tag{7}$$

$$D(A,B) \geq 0 \tag{8}$$

$$D(A,A) = 0 \tag{9}$$

If $D(A,B) = 0$, it means $A = B$ or A is similar to B.

THE TIME SERIES PREDICTION METHOD BASED ON SUPPORT VECTOR MACHINE

According to Takens principle, the time series is $\{x_t \mid t = 1, 2, \ldots, M\}$, which is can be expressed as $X_t = (x_t, x_{t-\tau}, \ldots, x_{t-(d-1)\tau})$. d represents dimension and τ represents delay time. The following mapping F can be get.

$$F : R^d \to R^d$$

It meets $X_{t+p} = F(X_t)$. X_t is a vector of d dimension, which represents the current state. p represents forward prediction step length, and F represents predictive model.

$$x_{t+p} = f(x_t, x_{t-\tau}, \ldots, x_{t-(d-1)\tau})$$

f represents a mapping from a state of d dimension to a real number.

$$f : R^d \to R$$

Use the known training samples, we can predict the support vector machine. The one step predictive model of support vector machine at time t is

$$\hat{x}_{t+1} = \sum_{l=1}^{L} \overline{\lambda_l} K(X_l, X_{lt}) + \overline{\lambda_0} \quad \textbf{(10)}$$

$X_{lt} = (x_t, x_{t-\tau}, \ldots, x_{t-(d-1)\tau})$ and L represents the number of points.

$X_{l(t+1)} = (\hat{x}_{t+1}, x_{t+1-\tau}, \ldots, x_{t+1-(d-1)\tau})$. The prediction of $t+2$ point is

$$\hat{x}_{t+2} = \sum_{l=1}^{L} \overline{\lambda_l} K(X_l, X_{l(t+1)}) + \overline{\lambda_0} \quad \textbf{(11)}$$

The predictive model of the p step is

$$\hat{x}_{t+p} = \sum_{l=1}^{L} \overline{\lambda_l} K(X_l, X_{l(t+p-1)}) + \overline{\lambda_0} \quad \textbf{(12)}$$

p represents the predictive step and $X_{l(t+p-1)} = (\hat{x}_{t+p-1}, x_{t+1}, \ldots, x_{t+p-1-(d-1)\tau})$. After $K(u,v)$ is chosen, training samples can be used to train support vector machine and $\overline{\lambda}$ is obtained.

EXPERIMENT AND ANALYSIS

Average price of coal from year 1990 to year 2004 is shown in TABLE 1. The proposed predictive method is used to predict the coal price from year 1994 to year 2004. Observed value and predictive value is shown in TABLE 2. The line chart of observed value and predictive value is shown in Figure 4. Predictive result of linear regression model, ARMA model and three exponential smoothing is shown in TABLE 3. Evaluation index of predictive effect of four algorithms are shown in TABLE 4.

$$SSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$$MSE = \frac{1}{N} \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| y_i - \hat{y}_i \right|$$

TABLE 1 : Average price of coal of each year

Year	GDP(trillion)	GDP increasing rate(%)	Average price of coal(ton/yuan)
1990	1.85479	3.8	69.61
1991	2.16178	9.2	74.75
1992	2.66381	14.2	68.34
1993	3.46344	13.5	130.61
1994	4.67594	12.6	116.76
1995	5.84781	10.5	128.88
1996	6.78846	9.6	165.00
1997	7.44626	8.8	166.60
1998	7.83452	7.8	160.20
1999	8.20675	7.1	143.98
2000	8.94681	8	139.69
2001	9.73148	7.5	150.59
2002	10.51723	8	162.00
2003	11.72519	9.3	174.98
2004	15.9878	9.5	207.06

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|(y_i - \hat{y}_i)/y_i\right|$$

$$MSPE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i / y_i)^2}$$

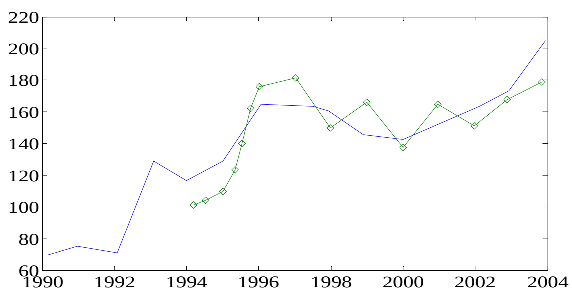

Figure 4 : The line chart of observed value and predictive value

TABLE 2 : Observed value and predictive value

Year	Observed value	Predictive value
1994	116.76	101.30
1995	128.88	107.65
1996	165	174.43
1997	166.6	181.69
1998	160.2	149.44
1999	143.98	167.74
2000	139.69	136.61
2001	150.59	166.10
2002	162	150.25
2003	174.98	167.09
2004	207.06	178.71

TABLE 3 : Predictive value of three models

Year	Observed value	Predictive value		
		Linear regression model	ARMA model	Three exponential smoothing
1994	116.76	120.73	101.30	137.24
1995	128.88	135.04	107.65	120.43
1996	165	146.23	174.43	144.25
1997	166.6	145.02	181.69	148.92
1998	160.2	138.74	149.44	150.53
1999	143.98	123.08	167.74	121.25
2000	139.69	148.45	136.61	157.36
2001	150.59	146.19	166.10	164.09
2002	162	179.45	150.25	168.68
2003	174.98	193.44	167.09	180.09
2004	207.06	225.81	178.71	198.32

TABLE 4 : Evaluation index of predictive effect

Predictive model	SSE	MSE	MAE	MAPE	MSPE
Linear regression model	2861.992	4.86	14.61	0.0909	0.0299
ARMA	2940.756	4.93	14.76	0.0961	0.0321
Three exponential smoothing	2485.741	4.53	13.76	0.0927	0.0314
Support vector machine	279.952	1.52	3.39	0.0225	0.0098

It can be seen that each error index of our proposed algorithm is less than that of any other model. The predictive accuracy is improved greatly, which can provide reference for pricing mechanism of coal enterprises.

CONCLUSION

In this paper, we present a structure-adaptive piece-wise linear segments representation of time series. The algorithm can automatically produce the K piece-wise segments of time series, which can approximate the original time series. Then the time series prediction method based on support vector machine is proposed. In order to test the performance of proposed algorithm, coal price is predicted using four algorithms. The results shown that the proposed algorithm has high predictive accuracy.

ACKNOWLEDGEMENTS

This work has been supported by Heilongjiang Social Science Planning Fund Project (Project No. 13E063), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 10YJC630348), Heilongjiang Social Science Planning Fund Project (Project No. 12E039).

REFERENCES

[1] Z.Liu,Y.Ding; "On the establishment and improvement of coal supply emergency response mechanism,"China Coal, **34(9)**, 8-20 **(2008)**.

[2] X.Zhu, D.Zhao; "The problems and countermeasures of coal emergency management in China,"China Energy, **33(1)**, 38-41 **(2011)**.

[3] T.Lv, Y.Gu, M.Q.Fang; "Site selection and scale measure-ment of regional coal reserves: taking jiangsu province as an example,"Energy Procedia, **5**, 2072-2077 **(2011)**.

[4] Y.B.Sun, P.Guo, Y.J.Wang; "Ascertaining reasonable reserve deal about coal resources," Journal of the China Coal Society, **30(5)**, 661-664 **(2005)**.

[5] J.L.Jiao, H.Z.Ge, Y.M.Wei; "Impact analysis of China's coal-electricity price linkage mechanism: results from a game model," Journal of Policy Modeling, **32(4)**, 574-588 **(2010)**.

[6] D.A.Galbi; "Regulating prices for shifting between service providers,"Information Economics and Policy, **13(4)**, 393-410 **(2001)**.

[7] Y.Sun, H.Zhang; "A study of China's coal pricing mech-anism under the financial crisis,"Reformation & Strategy, **26(6)**, 34-35 **(2010)**.

[8] H.J.Peng, M.H.Zhou, M.Z.Liu, Y.Zhang, Y.B. Huang; "A dynamic optimization model of an integrated coal supply chain system and its application," Mining Science and Technology (China), **19(6)**, 842-846 **(2009)**.

[9] S.S.Keerthi, O.Chapelle, D.DeCoste; "Building support vector machines with reduced classifier complexity,"J. Mach. Learn. Res., **7**, 1493-1515 **(Jul. 2006)**.

[10] G.H.Bakir, J.Weston, L.Bottou; "Breaking SVM complexity with cross-training," inProc. 17th Neural Inf. Process. Syst. Conf., **17**, 81-88 **(2005)**.

[11] K.De Brabanter, J.De Brabanter, J.A.K.Suykens, B.De Moor; "Optimized fixed-size least squares support vector machines for large data sets," Comput, Stat. Data Anal., **54(6)**, 1484-1504 **(Jun. 2010)**.

[12] J.Y.Wang; "Application of support vector machines in bioinformatics," M.S. thesis, Dept. Comput. Sci. Inf. Eng., National Taiwan University,Taipei, Taiwan, **(2002)**.

[13] B. Schölkopf, S.Mika, C.J.C.Burges, P.Knirsch, K.Müller, G.Rätsch, A.J.Smola; "Input space versus feature space in kernel-based methods," IEEE Trans. Neural Netw., **10(5)**, 1000-1017 **(Sep. 1999)**.

[14] J.A.K.Suykens, T.Van Gestel, J.De Brabanter, B.De Moor, J.Vandewalle; Least Squares Support Vector Machines. Singapore: World Scientific, **(2002)**.

[15] G.H.Bakir, J.Weston, L.Bottou; "Breaking SVM coplexity with cross-training," inProc. 17th Neural Inf. Process. Syst. Conf., **17**, 81-88 **(2005)**.

[16] M.Fauvel, J.Chanussot, J.A.Benediktsson, J.R.Sveinsson; "Spec-tral and spatial classification of hyperspectral data using SVMs and morphological profiles," IEEE Trans. Geosci. Remote Sens., **46(11)**, 3804-3814 **(Nov. 2008)**.