International Journal of Database Theory and Application
Vol.9, No.11 (2016), pp.127-136
http://dx.doi.org/10.14257/ijdta.2016.9.11.12

Controllable Curve Fitting Based Swing Door Trending Algorithm and its Application in Process Data Compression

Song Renjie[1], Zhang Qinghe[2*], Liu Haiyang[3], Yang Shuo[4], Wang Zhaohui[5] and Bao Zhen[6]

[1,2]Department of Information Engineering, Northeast Dianli University, Jilin 132012, China
[3,4]State Grid Jilin Electric Power Co., Ltd. Jilin power company, Jilin 132012, China
[5,6]Beijing SP Zhishen Control Technology Co., Ltd, Beijing 102200, China
[1]1939811347@qq.coml, [2]1103348819@qq.com

Abstract

Swing door trending (SDT) algorithm is a lossy compression algorithm that be applied on the real-time database and proposed by OSI software company of American; SDT is widely used to compress process data generated by process industry. Using straight line for a data section to linear fitting in traditional SDT algorithm. However, the data generated in the process of industrial production are slightly fluctuating with time. So, the use of linear fitting will lead to a large decompression error. In order to overcome the large decompression error generated by the traditional SDT, we proposed Controllable Curve Fitting Based Swing Door Trending (CCFSDT). The CCFSDT algorithm uses curve line for a data section to fitting, the data restored are closer to the true value. And in order to reduce the cost of curve fitting, it can be appropriate to reduce the total number of points of curve fitting. We filter noise point before fitting to void the impact on the reduction data and achieve better compression effect. The experimental results on simulated data and actual plant data show that: under the same conditions, the CCFSDT can well reduce the errors of decompression and achieve satisfactory performance.

Keywords*: Data compression; SDT algorithm; Controlled curve fitting; Filter noise point*

1. Introduction

The real-time data come from process industrial provides a valuable reference value to industrial development, with the improvement of industrial automation, the data of real-time acquisition are more and more, if directly store, not only waste a lot of disk storage space, but also the data retrieval and transmission to the information processing system and network caused a great burden [1].The compression and storage of industrial process data are a very effective way to solve the above problems.

The SDT lossy compression algorithm is widely used in industrial process data compression, and many improved methods have been produced based on SDT algorithm. For example, Liang Zhang *et al* propose the improved method of adaptive compression offset ΔE [2]; Bing-liang Lu *et al* propose the improved method of proportion compression [3]; Song-tao Yu *et al* propose the improved method of tolerance dynamic adjustment [4]; Rui Xing *et al* propose the improved method of storing processed data without storing the original data [5]; Yi-lin Qu *et al* propose the improved method of an automatic parameter control [6]; But, in most cases, the data generated in the process industry are slightly fluctuating with time, the use of linear fitting will lead to a large

*Corresponding Author

ISSN: 2005-4270 IJDTA
Copyright ⓒ 2016 SERSC

decompression error. In this paper, we put forward a new compression strategy for the way of compression error is larger and considering the noise point and time of fitting by analyzing the SDT compression algorithm. It can achieve better compression effect by using controllable curve fitting and removing noise point. The experimental results show that the CCFSDT compression algorithm can better restore the data and consumption of the time difference is small under the same conditions.

2. Swing Door Trending Compression Algorithm

In the real-time database system, the amount of need data are enormous and the similarity degree of data is very high. Storing and transmission of uncompressed data will cause great waste of resources [7]. Swing door trending compression algorithm base on linear fitting, which has many advantages, such as achieve fast, high efficiency, error controllable, higher compression ratio and so on. Now it has become a dedicated algorithm [8]. It can save a lot of storage space by using the swing door trending compression algorithm for process data.

SDT algorithm uses the line between the starting point and the last point replace all data points between the two points when compression, and performs linear interpolation for the line segment when decompression. Compression data are the time and the value of the starting point and the last point, and the final point in the previous section of compression will be used as the starting point for the next section of compression [9].

The basic principle of SDT algorithm is shown in Figure 1, setting △E as the tolerance parameter of SDT compression algorithm. For some need compressor data, starting from t0 time. With the t0 time data point as the center, distance the center point △E and two points of the upper and lower as fulcrums to create two fictitious gates in vertical direction. when only there is one data point, the gates don't open. But, with the arrival of new data point, the gates will rotate open and don't close again until the end of the compression phase. When the next data point comes, the gates will open. Getting the slopes of the connection line between the upper and the lower support points with the present data point, and they are marked as upper slope and lower slope. Then every time when come a new data point, getting a line and its slope by connecting upper fulcrum and the new data point. Then compare with marked upper slope, if large, the upper slope will be marked again with the slope of the point; if small, upper slope marker unchange. It's the same principle for under slope, but replace the under slope mark with the new one when the new obtain slope is small than value of under slope. Every after update upper and lower slope, compared to the values of upper and lower slope marker, if the value of upper slope is larger or equal than the lower slope, the compression phase end. Then storage the data point of prior as the last point of the phase, and began compression for the next stage with the new storage data point. For example in Figure 1, there are two swing door trending compression stages. The first is use straight line segment of from t0 to t3 time instead of the data points of t0 to t3 time, and the second is use straight line segment of from t3 to t7 time instead of the data points of t3 to t7 time.

 Copyright ⓒ 2016 SERSC

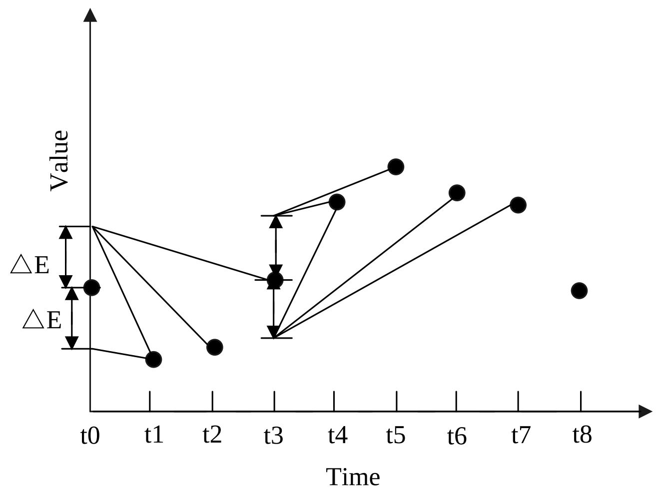

Figure 1. The Chart of Swing Door Trending Algorithm Compression Principle

Swing door trending algorithm can be more vivid to understand that according to the data to create one by one parallelograms of height fixed and rotate extension to hitch the data. When cannot "trap" all the data in a compression stage, archive storage near preceding a data point [10]. Every moment, the two parallel edges of the parallelogram are vertical in each stage, and its length are always twice △E [11]. Each time when rotate extension another two parallel edges, need observe the parallelogram is able to trap all data points in one compression stage. If one or more data points are not trapped, then the near preceding one data point will be archivalled storage, and the data will be compressed processing in parallelogram [12].

3. CCFSDT Compression Algorithm

3.1. Principle of CCFSDT

The tradition SDT algorithm uses linear fitting to compress a section of data, which leads to big compression errors. In order to solve this problem, this paper uses controllable curve fitting. The actual process data sometimes have noises interference, or due to the fault and other reasons produce too large or too small wrong values than true values. If these wrong data are also added to the controllable curve fitting, the curve may deviate from the original true change trend. It will lead to big compression error and data distortion increases. So before the data be determined whether to participate in the fitting, first calculate the difference value between the current data point and the last data point, and compare with noise limit value N that before setting. If the difference value is smaller than N, the data point participate in the fitting, otherwise not involved in fitting and end this compression. It can output the original value for the noise points when decompression, so the operator can observe the true data change.

By observing the process industry data, the change tendency of the data in a period of time is similar to the parabola. So we use parabola as the curve fitting basis and use least squares when fitting in this paper. Through experiments we found that store the intermediate data can achieve better compression effect than store the original data when curve fitting. Therefore, the time of data points and three parameters of parabola should be stored for compression data. When curve fitting, we can also control the number of fitting points to reduce the cost of fitting. We do a judgment before fitting: if the total

Copyright ⓒ 2016 SERSC

number of points is bigger than the limit value M that before setting in this fitting, then remove half of points through interval get points and the remaining points involved in curve fitting.

3.2. Description of CCFSDT

In the CCFSDT algorithm, set △E as tolerance, t is the time of compression and decompression, e is the average compression error.

Average compression error e is defined as

$$e=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-y_i^*\right)^2} \tag{1}$$

In formula (1): n is the total number of compression file data points, y_i is the original data value, y_i* is the data value of decompression reduction.

CCFSDT compression algorithm process is as follows:

Step 1 if it's the first time compression data, need to initialize relevant parameters.

Step 2 read a new data point in the waiting compression file.

Step 3 compare the difference value between the new data point D_j and the last data point D_{j-1} with the limit value N, if biger than N, store the D_j point, return to step 2; otherwise continue step 4.

Step 4 comparison and replacement of the upper and lower slopes.

Step 5 determine whether a compression interval is over, if not end, then return to step 2; otherwise continue to the next step.

Step 6 comparing the number of fitting points L with limit value M, if big than M, remove half of fitting points through interval get points, then next step; otherwise, direct next step.

Step 7 using the fitting points to do curve fitting.

Step 8 determine whether have need to compress data, if there are, then return to step 2; otherwise, end the compression.

The CCFSDT is shown in Figure 2.

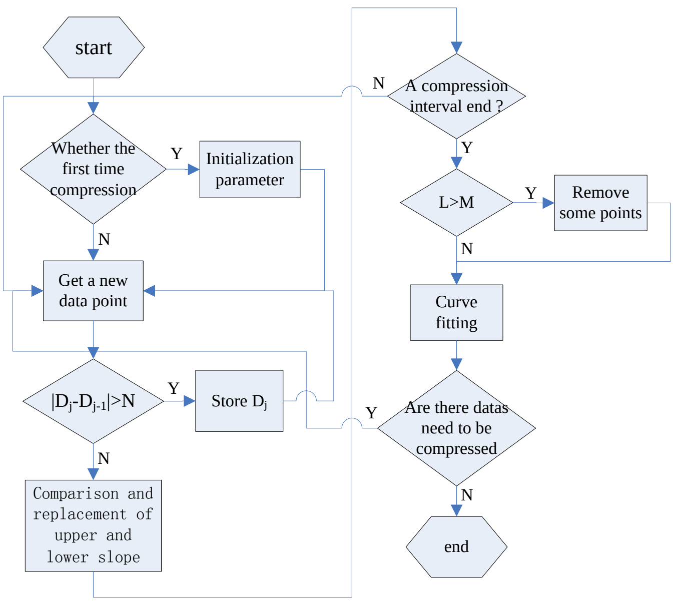

Figure 2. The Compression Flow Chart of CCFSDT Algorithm

Copyright ⓒ 2016 SERSC

4. Experiment

In order to fully verify effectiveness and feasibility of the CCFSDT algorithm, we use the simulation data and the actual power plant data to test.

4.1. Validation of Simulation Data

Select sinusoidal signals of different period and amplitude as the simulation signal. Here, we chose a sinusoidal signal that extract more than 700 data points and exist two noise points. Setting tolerance △E is 0.1, comparing the differences of compression and decompression data between the general swing door trending and the CCFSDT algorithm. The simulation results are shown in Figure 3 and Figure 4. Figure 3 shows the differences original data and decompression data of the general SDT algorithm. Figure 4 shows the differences original data and decompression data of the CCFSDT algorithm.

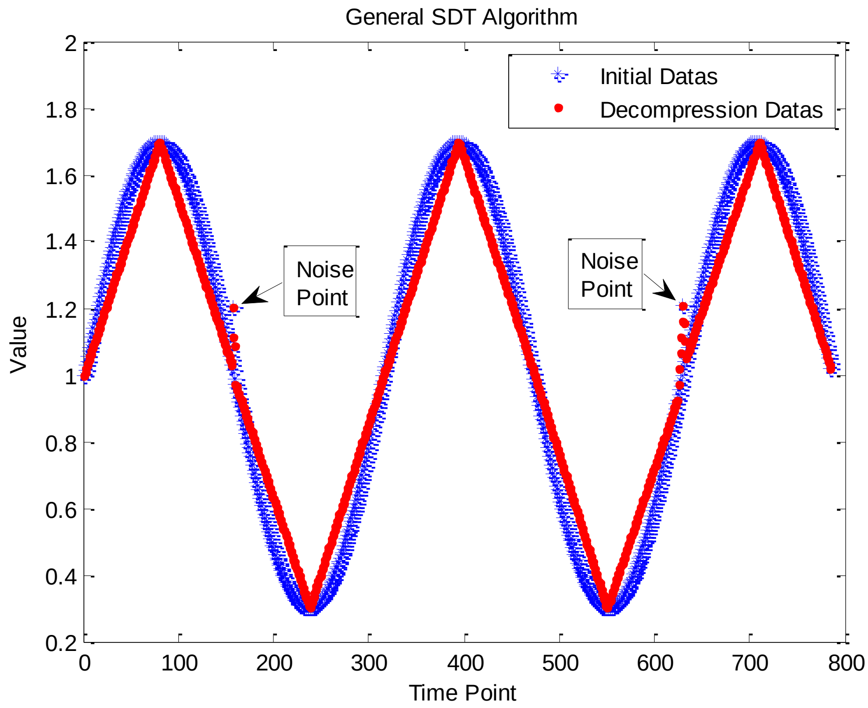

Figure 3. The Comparison Chart of General SDT Algorithm Decompressed Data

From Figure 3, it can be seen that the decompression data deviate from the actual value in some parts especially near noise points. The data distortion is large and affect the actual analysis of data.

Copyright ⓒ 2016 SERSC

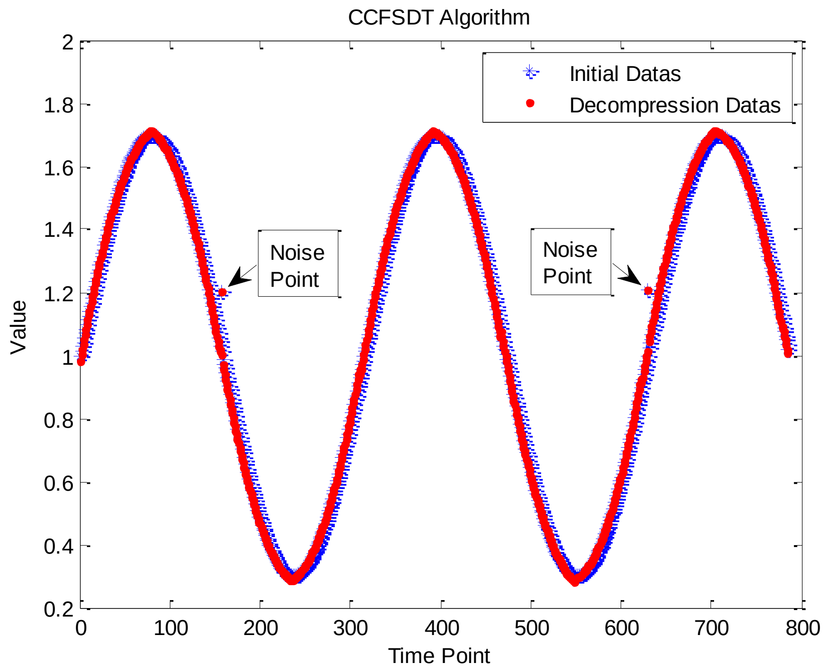

Figure 4. The Comparison Chart of CCFSDT Algorithm Decompressed Data

From Figure 4, we can see that the CCFSDT algorithm makes the data of decompression not only more approach the true value, but also has good processing ability for noise points. The decompression data near noise points don't very deviation from the original value, easy to analyze and deal with data.

Table 1 shows the interrelated parameters about compression and decompression of simulation data, from Table 1 we can see that the CCFSDT algorithm and tradition SDT algorithm don't big difference in the time of compression and decompression, but CCFSDT algorithm has great reduction in the average compression error.

Table 1. The Experimental Data Table of Simulation Data Decompression

Simulation data	Tradition SDT	CCFSDT
△E	0.1	0.1
Average compression error	0.103	0.013
Time consuming	6.2ms	9.3ms

4.2. Actual Data Validation

The actual process industry data come from the temperature of inverter in a photovoltaic power system, and choose 2000 data points at about noon of one day to analyze the compression and decompression effect of data.

Figure 5 and Figure 6 are actual data decompression graphical comparison. By comparing Figure 5 and Figure 6, we can see that the data of decompression more

 Copyright ⓒ 2016 SERSC

closer to real values when use the CCFSDT algorithm, especially in data fluctuation parts have greatly improved.

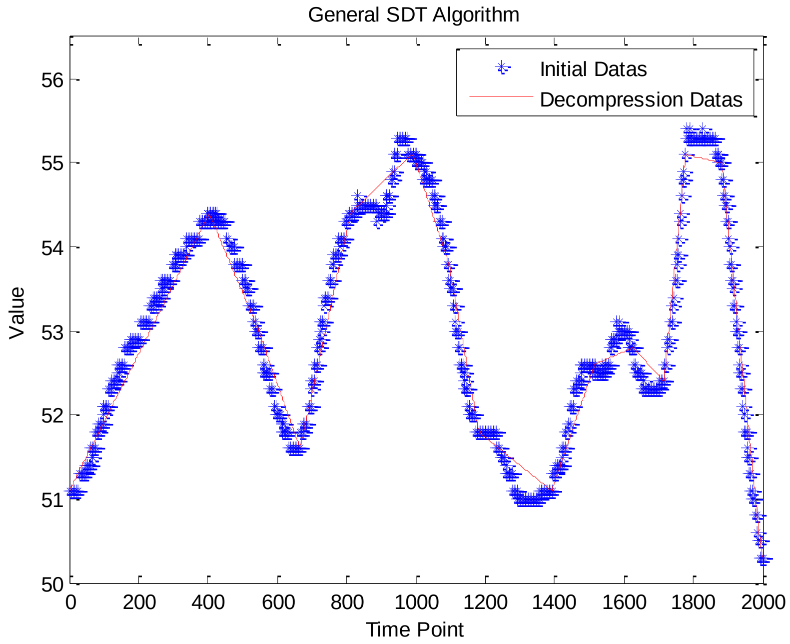

Figure 5. The Chart of General SDT Algorithm for the Actual Value Unzip Effect

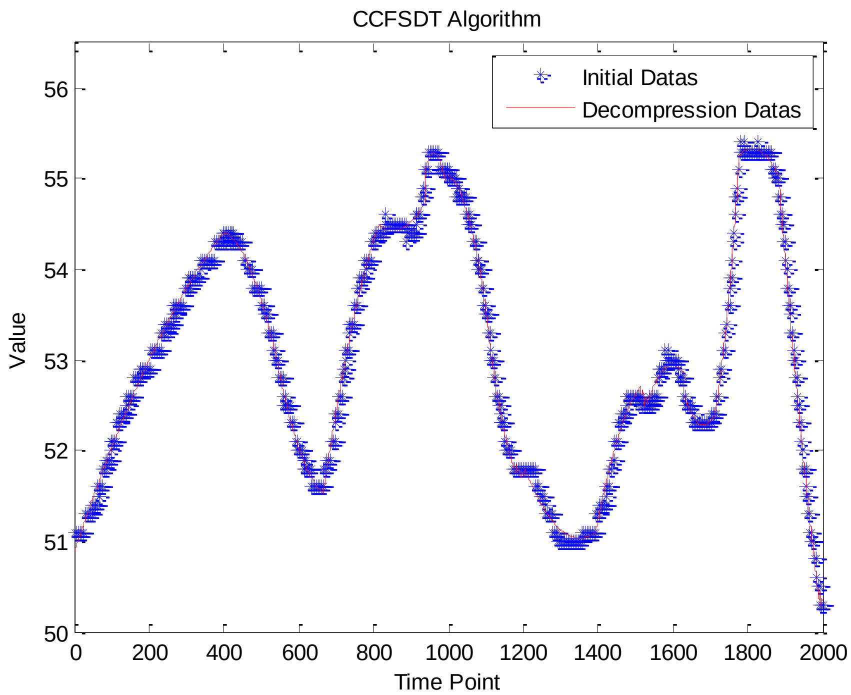

Figure 6. The Chart of CCFSDT Algorithm for the Actual Value Unzip Effect

Table 2 shows the interrelated parameters about compression and decompression of actual data. From Table 2, we can see that the CCFSDT algorithm has more lower

Copyright ⓒ 2016 SERSC

average compression error than the traditional algorithm, and time consuming isn't big difference in actual power plant data compression and decompression test under the same conditions.

Table 2. The Experimental Data Table of Actual Data Decompression

Simulation data	Tradition SDT	CCFSDT
△E	0.3	0.3
Average compression error	0.218	0.08
Time consuming	12.6ms	15.7ms

5. Concluding

In this paper, we proposed a controllable curve fitting(CCFSDT) compression algorithm based on swing door trending algorithm aim to solve the problem of decompression error larger in traditional swing door trending (SDT) algorithm. By curve fitting reduces the distortion of data, and we can also control the number of fitting points to reduce the cost of curve fitting. And considering the influence of noises and time consuming. Through the simulation and actual data test, it can be seen that the CCFSDT algorithm reduces the average compression error and better restore the authenticity of the data under the same conditions. And there is good processing power for nise points, that will more convenient to analyze and process data.

References

[1] N. Hainan, "A New Process Data Compression Algorithm Based on SDT Algorithm", Computer Technology and Development, vol. 20, no. 1, **(2010)**, pp. 25-28.
[2] Z. Wang, C. Xinchu and L. Dingxing, "Research and Application of Improved Process Data Compression Algorithm-SDT", Industrial Control Computer, vol. 22, no. 8, **(2009)**, pp. 1-4.
[3] L. Bingliang, Z. Jian and Z. Lei, "An Improved Data Compression Algorithm Based on Real-time Database", Computer Applications and Software, vol. 26, no. 11, **(2009)**, pp. 109-112.
[4] Y. Songtao, W. Xiaokun and Z. Liqiang, "Swing Door Trending (SDT) Improved Algorithm Based on Tolerance Dynamic Adjustment", Journal of Beijing University of Chemical Technology (Natural Science Version), vol. 40, no. 3, **(2013)**, pp. 109-113.
[5] X. Rui, Q. Qi and Z. Tao, "Improved SDT Algorithm", Computer Engineering and Design, vol. 34, no. 2, **(2013)**, pp. 515-518.
[6] Q. Yilin and W. Wenhai, "Automatic Parameter Control SDT Algorithm for Process Data Compression", Computer Engineering, vol. 36, no. 22, **(2010)**, pp. 40-42.
[7] R. S. H. Math, A. C. Tamhane and S. H. Tung, "Process Trending With Piecewise Linear Smoothing", Computer & Chemical Engineering, vol. 19, no. 2, **(1995)**, pp. 129-137.
[8] F. Xiaodong, C. Changling and L. Changling, "Improved SDT Process Data Compression Algorithm", High Technology Letters, vol. 9, no. 2, **(2003)**, pp. 91-96.
[9] J. R. Haritsa and K. Ramamritham, "Real-time Database Systems in the New Millennium", Real-Time Systems, vol. 19, no. 3, **(2000)**, pp. 205-208.
[10] W. Jian and C. Yong, "Overview of Real-time Database Management System Design for Power System SCADA System", Proceedings of the IEEE, vol. 31, no. 2, **(2005)**, pp. 62-66.
[11] Y. Dezhi, W. Jiechen and L. Guonian, "Study of Realization Method and Improvement of Douglas-peucker Algorithm of Vector Data Compression", Bulletin of Surveying and Mapping, vol. 11, no. 7, **(2002)**, pp. 18-22.

Copyright ⓒ 2016 SERSC

[12] S. Sivalingam, "Effect of Data Compression on Controller Performance Monitoring", Norwegian Univ of Sci. & Technol., Trondheim, Norway, **(2011)**, pp. 594-599.
[13] Q. Zhaoyang, C. Shuai, Y. Fan, Y. Jia and X. Shaoqing, "Intelligent Substation Data Classification Method Based on Two-level Analysis", Journal of Northeast Dianli University, vol. 34, no. 2, **(2014)**, pp. 61-65.
[14] Y. Mao, S. Yong, S. Zhaojian, Y. Yongliang and H. Jinfu, "Design and Development of Large-scale Data Management System of Wind Farm", Journal of Northeast Dianli University, vol. 34, no. 2, **(2014)**, pp. 27-31.

Authors

Song Renjie, (1963), female, Jilin, China, professor. Research Direction: Computer Applications in Power System. Email: 1939811347@qq.com.

Zhang Qinghe, (1989), male, Jiangsu, China, postgraduate. Research Direction: Intelligent Information Processing and Power Applications. Email: 1103348819@qq.com.

 Copyright ⓒ 2016 SERSC