International Journal of Multimedia and Ubiquitous Engineering
Vol.12, No.4 (2017), pp.25-32
http://dx.doi.org/10.14257/ijmue.2017.12.4.03

An Adaptive Compression Strategy Based on A* Algorithm

Yun Wu[1], Guannan Xu[2,*], Jieming Yang [3]and Lishan Bao[4]

[1,2,3] *School of Information Engineering, Northeast Dianli University,Jilin,132012,China*
[4]*State Grid Jiangsu electric power company information and Communications Branch, Nanjing, 210000, China*
838558160@qq.com,931530510@qq.com,670172713<yjmlzy@sina.com>,13801 3800@qq.com
** Corresponding Author: Guannan Xu*

Abstract

In this paper, a new adaptive compression strategy based on the A algorithm called SA* algorithm is proposed which combined the data compression with the heuristic method. A new evaluation function construction method is devised for the SA* algorithm, which is suitable for the frequency optimization of data compression. The simulation experiments are carried out to compare the adaptive compression algorithm based on heuristic search with the traditional equal-interval sample method applying Matlab platform, the experimental results show that the adaptive compression strategy based on A* algorithm can reduce the storage space of the data with lower distortion, and reach the balance of the quantity and quality of the data. This algorithm will be used to deal with the problem of large amount of historical load data in SCADA system, and it has the phenomenal ability to capture the tendency of changing load data, reducing the burden of the network and information processing system. The data cleaning and load forecasting time can be reduced as well.*

Keywords*: power load data, A* algorithm, adaptive sampling, data compression*

1. Introduction

With the rapid development of the current socio-economic and scientific technology, the demand for electricity and grid load is increasing from all sectors of society. The load data of high quality and the load forecasting result of high accuracy are very important for the planning, operation and control of power system, and have the far-reaching significance to improve the security, stability and economy of the power grid [1]. At present, with the development of the automation of electric power operation which has been applied to monitor and control the operation equipment of power system. The quantity of data collected by SCADA system in time becomes larger [2-3], and then the collected data are stored into historical database, which causes heavy burden to the network [4] and information processing system as well as increases the processing time to data cleaning, load forecasting [5-9]. Because of the more and more intelligent of the power grid, to improve the accuracy of load forecasting, demand for the accuracy of the data increases. Meanwhile, the increase of the sampling frequency has a heavier burden in information system, and reduction of processing capacity on information system becomes significant, a large amount of data storage is a big challenge to historical database as well.

Efficient data compression algorithm can reduce the amount of data and improve the utilization ratio of the data, the data compression using reasonable storage strategy can further reduce data storage space and the burden on the system. SCADA system of historical data compression generally use lossy compression algorithm. It can be divided

ISSN: 1975-0080 IJMUE

into 3 categories: signal transformation method, vector quantization method, and piecewise linear method. These all traditional data compression method is on the basis of equal interval sampling to data compression, which cannot avoid the disadvantages of static sampling from original database, so the traditional data compression method has been unable to meet the demand of the system[10]. Therefore, this paper proposes an efficient data compression technology aiming at reducing the amount of data and guaranteeing the data information at the same time. Based on the analysis of the historical load data in real time in the power system, so as to constantly adjust the sampling frequency. As a result, the sampling frequency can capture the data under test in real time, accurately. The adaptive data compression can reduce the storage as well as achieve the purpose to improve the cleaning efficiency of the late data.

In this paper, the organizational structures as follows, Section 2 introduces correlation algorithm; Section 3 introduces load data compression procedure; Section 4 presents the experimental results and analysis; And there is a conclusion finally.

2. The Principle of Adaptive Compression Algorithm Based on A* Algorithm

2.1. A * Algorithm

In the field of intelligent control, heuristic search is a kind of important method. It is the simulation of the human brain about cognitive model, which weighs the part of possibility but not all, and gives the last judgment[11].

A* algorithm is an important kind in the heuristic algorithm, it is widely applied to the optimal path to solve and some strategies in the design[12]. Algorithm uses heuristic function to find the terminal point to the starting point with minimum cost, the general form for evaluation function of the node *n* is[13-16].

$$f(n) = g(n) + h(n) \tag{1}$$

In the formula (1), $f(n)$ represents that valuation function from the initial points to the target point through the current extension point, it is the sum of two parts. 1） $g(n)$ represents the actual cost from the initial search point to the current extension point (usually expressed in depth, values are not fixed). 2） $h(n)$ represents the estimated value from the current extension point to the target node (the most important part of the search, values are fixed).

2.2. SA * Algorithm.

In the process of searching the optimal solution of the set δ , the common search strategy is improved, which makes it more suitable for the search of optimal frequency in this paper.

In the search of the optimal solution, considering the valuation function which is commonly used in the search target node *n* this paper designs a valuation function which is suitable for the frequency optimization.

$$f(n) = y(n) - y(n') \tag{2}$$

In the formula （2）, $f(n)$ represents that valuation function from the initial points to the target point through the current extension point, it is the sum of two parts. 1） $y(n)$ represents the actual cost of the current extension point ； 2） $y(n')$ represents the estimated value from the current extension point to the target node.

2.3. Sampling Strategy

Usually, the change of the normal load data is a smooth curve, and environment or other causes lead a significant influence on the change of the date, so the data information should be retained in order to ensure information from data and to achieve the aim of on-demand compression. When the change of data is smooth, the sampling points of historical data should be reduced; when the change of data is dramatic, the sampling points of historical data should be increased [17-20]. The sampling frequency of compression algorithm with traditional high quality is static, designed to greatly reduce the data storage space and improving the efficiency of data storage. However, it cannot eliminate the insufficiency of interval sampling of the raw data, so it cannot meet the needs of the compression of load data. The disadvantages of the sampling interval can be divided into two kinds: when the sampling interval is too large, though it reduces the burden of data storage space and the management of information system, it can make the distortion degree of data become bigger, leading to the loss of information, and unable to maintain the characteristics of the data itself. when the sampling interval is too small, though it can ensure the characteristics of the data itself, it increases the burden on the processing of information system, which is not conducive to later analysis of data[21-22]. In order to eliminate the influence of these factors, this paper presents a method of adaptive compression according to the curve features of load data. Therefore, the allowable fluctuation range for the values of load data should be given in a period of time. Assuming that the minimum value of fluctuation of the load data is δ_{min} while the maximum value is δ_{max}, the specific sampling strategy is as follows,

(1) When the fluctuation of the load data is smaller than δ_{min}, the data change is not obvious, and can decrease the sampling frequency, sampling and storage of little valuable data, to reduce the burden of management of information system at this time;

(2) When the fluctuation of the load data is larger than δ_{max}, the data change is obvious, and can increase the sampling frequency to improve the amount of information of the load data and to provide full information for later analysis of data at this time;

(3) When the fluctuation of the load data is between δ_{min} and δ_{max}, the sampling frequency keeps unchanged at this time.

In the above strategy, δ_{min} and δ_{max} should be determined by the rate of statistical average data from the historical database.

3. Load Data Compression Procedure

3.1. Adjustment Strategy of Sampling Frequency

Based on the characteristics of the normal curve of load data [23], adaptive data compression method is to reduce the sampling frequency when data changes smoothly, in order to reduce the load of information processing system; when the change of data is dramatic, the sampling frequency should be increased to keep the amount of information of the data. This method firstly calculates variations in the data in a period of time, and then judges the changes of data according to the given range of fluctuation, and the last is to adjust the sampling frequency [24-25].

Definition

In order to realize the sampling strategy, which can be defined the average change of the load data in a period of time as,

$$\Delta y = \frac{\left|y_n - y_{n-1}\right| + \left|y_{n+1} - y_n\right| + \cdots + \left|y_{n+k-1} - y_{n+k-2}\right|}{k} \quad (3)$$

y_n : Load data value of the *n* point

Minimum value of the average fluctuations δ_{min} -In the sampling period, the data is allowed to fluctuate within the minimum average value relative to a moment before;

Maximum value of the average fluctuations δ_{max} -In the sampling period, the data is allowed to fluctuate within the maximum average value relative to a moment before;

k-in the sampling period, monitoring points should be considered in average movement;

n -the sequence of sampling points in the current time;

$S(v_s)$ -the data set after the adaptive compression;

δ - a collection of Δy

T_d : sampling change step

To sum up, the thought of combining with heuristic algorithm, in this paper, the adaptive sampling algorithm steps are as follows, Such as Figure1.atep

1) Firstly defines δ_{min}、δ_{max}、k、ΔT、 $S(v_s)$ and T_d;

2) Data sampling by ΔT intervals;

3) In the sampling process, every step must be to judge whether the (3) within the $[\delta_{min}, \delta_{max}]$. If it is, just to keep the original frequency sampling, and input the sampling data to the data set; otherwise, to take the step 4;

4) Using the (3), if the calculation result is smaller than δ_{min}, then search the optimal solution within the data set to achieve the aim of reducing sampling frequency; if the

calculation result is bigger than δ_{max}, then search the optimal solution within the data set to

achieve the aim of increasing sampling frequency; otherwise, just to keep sampling frequency.

The calculated Δy in the current step will be put into data sets δ, while sampling data will continue to be input to data sets $S(v_s)$.

3.2. The Realization of the A* Algorithm on the Frequency Selection

The solution of optimal frequency search process can be viewed as a state space search process. The process of starting point as a sampled data of the first cycle, the purpose is find the optimal frequency value is the target node is the optimal sampling frequency and keep the original characteristic curve and reduce the data storage space. Using the heuristic search algorithm of optimum frequency selection of sampling, should first consider how to construct $f(w)$. According to the compressed data distortion to represent the extension point to the target node to estimate the value of $f(w)$,

$$f(w) = y(w) - y(w') \tag{4}$$

In the formula (4), $y(w)$ represents the sampling frequency of w, the data distortion; $y(w')$ represents the best sampling frequency of w', the distortion of the data.

Calculating $f(w)$ according to the size of the values of $y(w)$ and $y(w')$, in combination with the main idea of the A* algorithm to find the optimal sampling frequency, which can be completed by the following steps.

1） Initialization;

2）Within the current sampling time, comparing the calculated data Δy with the date in the data set δ;

3）To evaluate and calculate the sampling frequency in the current node, the distortion degree of compressed data, namely the valuation function as shown in the （4）;

4）To sort all frequency according to the distortion degree in the data set δ;

5）Choosing frequency set according to the size of the distortion degree; In the proximal data set, to detect the valuation function of the remaining nodes and to select a frequency value with minimum distortion degree;

6）To end the search, and to remember the optimal frequency sampling value of the current node

7）Before the arrival of next sampling time, to find the optimal frequency sampling values of the time according to the above method to loop iteration repeatedly until meeting the termination conditions of algorithm, finally to output the best data sample point.

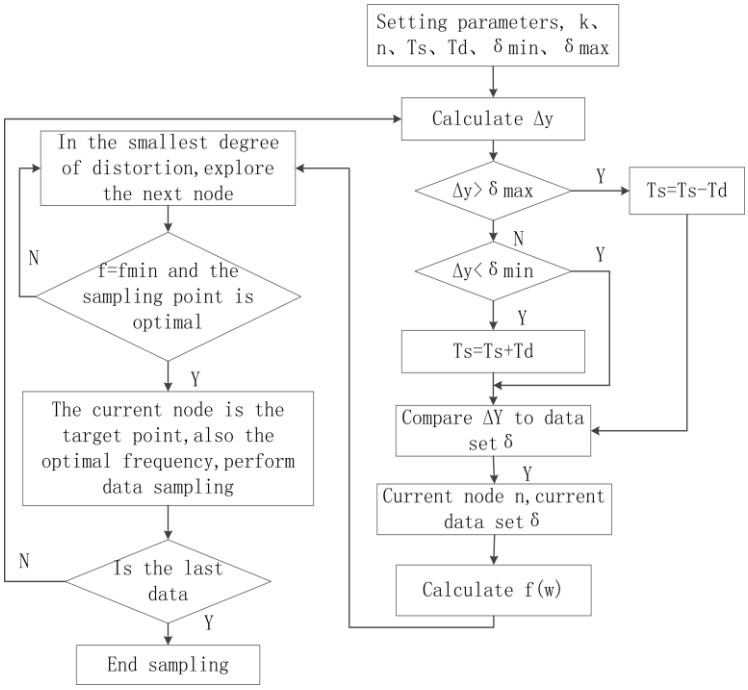

Figure 1. The Flow Chart of Load Data Compression Step

4. The Analysis of Experiments

This experiment takes load data from a province in 2015 as sample, using the traditional interval sampling method and the adaptive sampling algorithm for simulation experiments. The first is to make the division between the daily load curve and its daily power consumption, which can avoid the problems of large daily load data and maintain the original characteristics of the load curve at the same time. The sampling interval of sample data is 5 min, the points of daily load data is 288.

Figure 2and Figure 3 for the results of simulation diagram, it can be clearly seen that all the sample points use interval sampling are equally spaced through the comparisons between two figures, while the distribution of the sample points use adaptive sampling is uneven. In adaptive sampling, when data curve fluctuations are relatively flat, there are few data sampling points; while in the area where the data curve's volatility is larger, the points of data sampling are more, having the strong ability to capture data-changing trend. This is the effect of adaptive compression algorithm proposed in this paper. Through the comparison between interval sampling and adaptive sampling, in the acute phase, the data

change in interval sampling has been using the same sample frequency, distortion degree of the data is higher, which can lead to the loss of information of data and cannot show the features of the original data; while use the adaptive sampling in this phase, the distortion degree of the data is relatively lower. In the gentle phase of the change of data, the number sampling points in this phase is as dense as that in dramatic area in interval sampling, which increases the number of mining point of the load data and brings the burden for later processing of data, while use adaptive sampling in this phase can show the original features of the data with few sampling points. Table 1 is the data analysis, the distortion is divided into 4 levels, the 0 is the lightes

Table 1. Comparison of Simulation Results

Interval range	Equal interval sampling rate	Adaptive sampling rate	Equal interval sampling distortion degree	Adaptive sampling distortion degree
1-40	25%	12.5%	0	1
40-80	25%	50%	3	0
80-120	25%	27.5%	1	0
120-160	25%	72.5%	3	0
160-200	25%	15%	1	1

Comprehensive two indicators: distortion and sampling rate, the comparison shows that the adaptive compression algorithm is obviously superior to the traditional interval sampling method, such as the sampling results in guaranteeing to reduce the amount of data and the reflection for characteristics of the original data at the same time, which is in favor of the later processing work of the data.

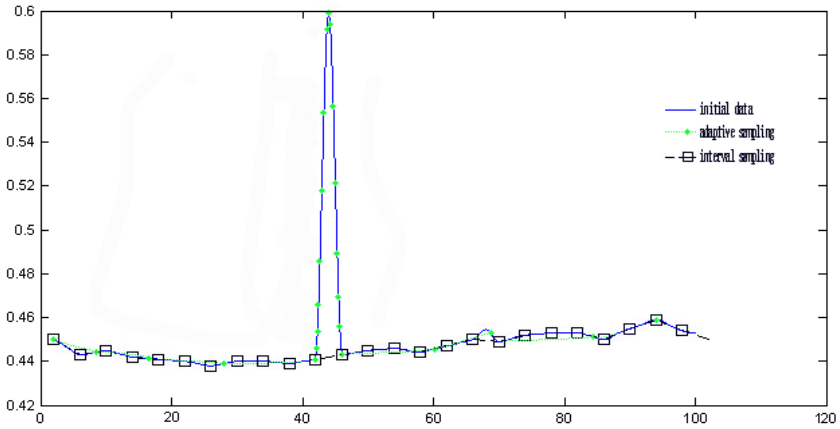

Figure 2. Experimental Result

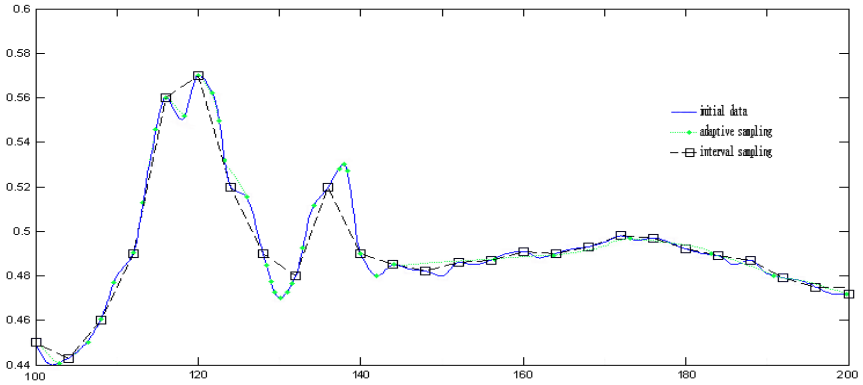

Figure 3. Experimental Result

5. Conclusions

Traditional interval sampling methods cannot achieve the aim of on-demand compression, so an adaptive compression algorithm is designed in this paper. This algorithm can reduce the amount of data without losing track of critical information about load data at the same time, and reduce the burden of the network and information processing system, the cost of the later work, such as data cleaning and load forecasting, additionally, it has provided the security for the safe and stable operation of interconnected power grid. Through the verification and analysis for the matlab simulation experiment, compared with the traditional sampling method, adaptive compression method has some advantages, for example, the ability to capture the change trends for load data is stronger, and is a kind of feasible and substituted method to the storage of load data compression.

Acknowledgments

This work was supported by Project development plan of science and technology of Jilin Province (No.20140204049GX).

References

[1] K. Wang and T. X. Wang .“Identification of Abnormal Power Load Data Based on Two Times Clustering” ,Electrical technique, vol. 11, **(2014)**, pp. 1-3+17.
[2] Y. Q. Song, G. L. Zhou and Y. L. Zhu. “Present Status and Challenges of Big Data Processing in Smart Grid” ,Power System Technology, vol.37, no. 4, **(2013)**, pp.927-935.
[3] D. W. Wang and Z. W. Sun. “Big Data Analysis and Parallel Load Forecasting of Electric Power User Side” ,Proceedings of the CSEE, vol.35, no.3, **(2015)**, pp.527-537.
[4] J. Y. Wang, Z. X. Ji, M. J. Shi. “Scenario Analysis and Application Research on Big Data in Smart Power Distribution and Consumption Systems”, Proceedings of the CSEE, vol.35, no.8, **(2015)**, pp.1829-1836.
[5] C. L. Tang, J. P. Lu, Y. Z. Xie, and L. Zhang. “Improved Data Stream On-Line Segmentation Based Ultra Short-Term Load Forecasting”, Power System Technology, vol.38, no.7, **(2014)**, pp.2014-2020.
[6] S. X. Zhang, B. Z. Zhao, F. Y. Wang and D. Zhang. “Short-term Power Load Forecasting Based on Big Data”, Proceedings of the CSEE, vol.35, no.1, **(2015)**, pp.7-42.
[7] J. Y Chen, W. Y .Li, Adriel .Lau, J. G. Cao and K. Wang. “Automated load curve data cleansing in power systems”, IEEE Trans on Smart Grids, vol.1, no.2, **(2010)**, pp.213-221.
[8] M. Yang, T. F. Lu and B. M. Ji. “A Review of Chaos Theory in Power System Load Prediction”, Journal of Northeast Dianli University, vol.35, no.3, **(2015)**, pp.18-21.
[9] B. Xiao, X. Xu, K. Song, Y. Bai and J. F. Li. “Abnormal Data Identification and Treatment in Spatial Electric Load Forecasting”, Journal of Northeast Dianli University, vol.33, no.1/2, **(2013)**, pp.45-50.
[10] F. Y. Ma, Q. P. Li and Z. B. Ma. “The Research of Historical Data Compression and Storage Strategy in Power Dispatch SCADA System”, Power System Technology, vol.38, no.2, **(2014)**, pp.109-1114.

[11] W. F. Liu and X. Li. "A Fast Multi-neighborhood Iterated Local Search Algorithm for Vehicle Routing Problem", Journal of Shenzhen University Science and Engineering, vol.32, no.2, **(2015)**, pp.196-204.
[12] Y. Zhang, G. Q. Tang and L. Chen. "Improved A* algorithm for time-dependent vehicle routing problem" ,Control Engineering of China, vol.19, no.5, **(2012)**, pp.750-752+756
[13] H. C. Wu, J. H. Liu, C. T. Tang, L. J. Xu and J. S. Liu. "Automatic Pipe Layout Design and Optimization Method Based on Improved A* Algorithm", Computer integrated manufacturing system, vol.7, no.7, **(2015)**, pp.1-17.
[14] Z. H. Xi, H. B. Li, H. P. Liu and F. C. Sun. "Multiple Object Tracking Using A* Algorithm Optimization", Journal of Tsinghua University (Science and Technology), vol.54, no.12, **(2014)**, pp.1549-1554.
[15] Q. Zhao, and A. Q. Zhao "A Multi-path Routing Algorithm Base on A* Algorithm", Journal of Electronics & Information Technology, vol.35, no.4, **(2013)**, pp.952-957.
[16] Y. Xin, H. W. Liang, M. B. Du, T. Mei, Z. L. Wang and R. H. Jiang. "An Improved A* Algorithm for Searching Infinite Neighbourhoods", Robot, vol.36, no.5, **(2014)**, pp.627-633.
[17] Z. H. Ji, Z. G. Kuang and H. Ni. "A Novel Two-Dimension Adaptive Data Collection Method for Network Management", presented at International Conference on Communications and Mobile Computing, **(2009)**.
[18] P. Lu, X. Cong, K. Song and D. D. Zhou. "Research Progress and Prospect of Computerized Adaptive Testing", Journal of Northeast Dianli University, vol.35, no.3,**(2015)**, pp.82-90.
[19] L. H. Wang. ". Product Model for Load Forecasting Based on Decomposition Technique in Time Series", Journal of Northeast Dianli University, vol.33, no.6,**(2013)**, pp.45-47.
[20] Z. Y. Sun, and C. Zhou. "Adaptive Clustering Algorithm in WSN Based on Energy and Distance", Journal of Northeast Dianli University, vol.36, no.1,(2016), pp.82-86.
[21] B. Zhang, and D. L. Zhang. "Parametric Compression Algorithm for Power System Steady Data", Proceedings of the CSEE, vol.31, no.1, **(2011)**, pp.72-79.
[22] C. Huang, S. X. Yang, Y. C. Liang, K. Liu, C. Wen and Z. H. Guo. "Practical Compression Method for Power System Fault Recording Data", Electric Power Automation Equipment, vol.34, no.6,**(2014)**, pp.62-167.
[23] G, J. Zhang, J. J. Qiu and J. H. Li. "Outlier Identification and Justification Based on Neural Network", Proceedings of the CSEE, vol.21, no.8, **(2011)**, pp.105-108+114.
[24] B. S. Gai, J. L. Wang and X. Liu. "An Adaptive Sampling Frequency Adjustment Method in Network Management System", Network new media technology, vol.1, no.1, **(2011)**.
[25] B. Jia, E. P. Eyisi. and X. Yuan. "Distributed Sampling Rate Adaptation for Networked Control Systems" , Proceedings of the First International Workshop on Cyber-Physical Networking Systems ,Shanghai,**(2011)**pp.768-774.

Authors

Yun Wu, She is an Associate Professor in the Northeast Dianli University. She received the MS degree from the Jilin University. Her research interests in intelligent information processing.

