[1]Ping Huang

Energy-Saving Data Acquisition Model of Wireless Sensor Network Based on Nonlinear Algorithm

Abstract: - Wireless sensor network significant data acquisition has a high cost, long completion time, and low accuracy. This paper adopts a sensor network data acquisition method based on a nonlinear algorithm to solve the above problems. In this paper, a distributed data acquisition method based on nonlinear regression is established by combining the time series relationship of data. First, this paper uses nonlinear regression analysis technology to establish the sensing data model and retain the characteristics of the sensing data. This makes the node pass only the parametric data of the regression model. This paper uses this method to replace the transmission of actual monitoring sensory data information. Experiments show that this method can effectively reduce data redundancy and network traffic under any conditions. The method has been verified in practical WSN applications.

Keywords: Wireless sensor network; Correlation; Time series; Nonlinear algorithm; Network traffic; Collected data

I. INTRODUCTION

With the rapid development of Internet of Things technology, WSN has become the most basic technology. It can monitor, perceive and collect data from various environments and targets in real-time. With the continuous development of wireless communication technology, sensor technology, and embedded computer technology, wireless sensor network has been used more and more in war monitoring, environmental monitoring, disaster rescue, and so on. Wireless sensor data monitoring generates much data during remote monitoring in narrow areas such as transmission lines, coastal areas, or highways. The cost of data acquisition and long-distance transmission is high. In a limited space, the bandwidth and energy consumption of the network will increase due to the limitations of each node's computing power, storage space, and geographical location [1]. How to effectively collect and transmit data in a narrow area to save energy resources has become an essential topic in the current WSN research.

While the WSN node performs continuous sampling, the collected data is chronologically organized into a time series. Some scholars have conducted time series research on this model. The results show that the data obtained from different physical processes in a specific period have vital timeliness. Regression analysis is a method of analyzing the interdependence of various factors. Many experts and scholars are currently working on collecting large-scale, multi-resolution network data. The multi-resolution data acquisition method based on linear regression establishes the model of perceptual data. This ensures the characteristics of the perceptual data [2]. At the same time, this method makes the node only transmit the parameters of the regression model and replace the perception data in the accurate monitoring. Although the multi-resolution characteristics of the network can be ensured and real-time acquisition can be realized, there are still many imperfect problems. This paper adopts a data acquisition method based on a nonlinear algorithm of the sensor network to solve the above problems. This paper establishes a distributed data acquisition method based on linear regression by combining the time series relationship between data. The system uses linear regression analysis techniques to build a model of the sensed data and retain the characteristics of the sensed data. This makes the node pass only the parametric data of the regression model.

II. DISTRIBUTED LINEAR REGRESSION MODEL FOR SENSOR NETWORKS

The method can select t sensing values within a short distance in a specific time interval through parameters such as storage space and processing capacity of different sensor nodes. Suppose it is $(r_1, f_1),(r_2, f_2),\cdots,(r_t, f_t)$. r_i

[1] Shenzhen Power Supply Bureau Co., Ltd., Shenzhen, Guangdong, 518000, China

E-mail: huangping201512@163.com

and $f_i (i \in [1,t])$ represent the sampling moment and the measurable result that was measured. t perceptual data constructor $F(r)$ has very little approximation error $\eta_i = (F(r_i) - f_i$. The format of $F(r)$ depends on the specific problem. Here $F(r)$ can be expressed in equation (1):

$$F(r) = \sum_{j}^{n} \zeta_j \beta_j(r) \tag{1}$$

Here and the number of items n and the specific basic function β_j depend on the solution of the problem. Usually, the basic equation chosen in this paper can be $\beta_j(r) = r^{j-1}$, then equation (1) can be expressed by an $n-1$ degree polynomial of r.

$$F(r) = \zeta_1 + \zeta_2 r + \zeta_3 r^2 + \cdots + \zeta_n r^{n-1} \tag{2}$$

Although $n = t$ can accurately evaluate the corresponding f_i value, the operation of the higher-order function value F is likely to affect the data. It has a particular impact on its accuracy for the unpredictable value of r corresponding to f. It is better to use n much lower than t. That is $n \ll t$. An estimated function F corresponding to the measurement result f is obtained by choosing the value of the factor ζ_i. In the wireless sensor network, it is assumed that 50 sets of temperature measurement data closest to a certain point are selected to establish a cubic polynomial model. $F(r) = \zeta_1 + \zeta_2 r + \zeta_3 r^2 + \zeta_4 r^3$ estimates the measured value $f_i (i = 1, 2, \cdots, 50)$. The node does not need to transmit 50 accurate data and builds a mathematical model of the function. It only needs to transmit 4 parameters in the network. $\zeta_1, \zeta_2, \zeta_3$ and ζ_4 are used as compressed representations of the data. The amount of information in the network is reduced [3]. If the parameters obtained by the linear regression method are used, the expression mode of the polynomial must be converted into the expression form of a matrix. This method does not need to solve higher-order polynomials but only needs to maintain the corresponding matrix. Assume that the required n dimensional vector is $\zeta = (\zeta_1, \zeta_2, \cdots, \zeta_n)^T$, and the actual measurement value of t dimension is $f = (f_1, f_2, \cdots, f_t)^T$. The basis function matrix of the corresponding sampling point r_i at the time is expressed as follows:

$$T = \begin{bmatrix} \beta_1(r_1) & \beta_2(r_1) & \cdots & \beta_n(r_1) \\ \beta_1(r_2) & \beta_2(r_2) & \cdots & \beta_n(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1(r_m) & \beta_2(r_m) & \cdots & \beta_n(r_m) \end{bmatrix}$$

If $t_{ij} = \beta_j(r_i)$ is a matrix, then the t dimensional vector $F = (F(r_1), F(r_2), \cdots, F(r_t))^T$ of the prediction function of Equation (1) at the sampling time of r_i is expressed by Equation (3), that is

$$F = \begin{bmatrix} F(r_1) \\ F(r_2) \\ \vdots \\ F(r_t) \end{bmatrix} = T\zeta = \begin{bmatrix} \beta_1(r_1) & \beta_2(r_1) & \cdots & \beta_n(r_1) \\ \beta_1(r_2) & \beta_2(r_2) & \cdots & \beta_n(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1(r_m) & \beta_2(r_m) & \cdots & \beta_n(r_m) \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \vdots \\ \zeta_n \end{bmatrix} \tag{3}$$

Then this paper can use equation (4) to express the approximate error vector η , that is,

$$\eta = T\zeta - f \tag{4}$$

The smallest approximate error vector η is selected as the optimal object to obtain the optimal solution.

$$\min(\|\eta\| = (\sum_{i=1}^{t} \eta_i^2)^{1/2}) \tag{5}$$

This paper combines equation (4) with the optimal objective equation (5) to obtain

$$\min(\|\eta\|^2 = \|T\zeta - f\|^2 = \sum_{i=1}^{t} (\sum_{j=1}^{n} t_{ij}\zeta_j - f_i)^2) \tag{6}$$

Each $\zeta_s (s = 1, 2, \cdots, n)$ can differ from $\|\eta\|^2$ to 0 to get the minimum η :

$$\frac{d\|\eta\|^2}{d\zeta_s} = \sum_{i=1}^{t} 2(\sum_{j=1}^{n} t_{ij}\zeta_j - f_i)t_{is} = 0, s = [1, n] \tag{7}$$

The following matrix formula equivalent to formula (7) can be derived from formula (4)

$$(T\zeta - f)^T T = 0 \tag{8}$$

$$T^T (T\zeta - f) = 0 \tag{9}$$

$$T^T T\zeta = T^T f \tag{10}$$

Since its basic equation system is $\beta_j(r) = r^{j-1}$, its basic equation T is a column fully ordered matrix. It can be concluded that $T^T T$ is positive definite for any column, so $T^T T$ makes sense [4]. The solution to the coefficient vector ζ from equation (10) is:

$$\zeta = (T^T T)^{-1} T^T f \tag{11}$$

Make

$$C = T^T T = \begin{bmatrix} \langle\beta_1 \cdot \beta_1\rangle & \langle\beta_1 \cdot \beta_2\rangle & \cdots & \langle\beta_1 \cdot \beta_n\rangle \\ \langle\beta_2 \cdot \beta_1\rangle & \langle\beta_2 \cdot \beta_2\rangle & \cdots & \langle\beta_2 \cdot \beta_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\beta_n \cdot \beta_1\rangle & \langle\beta_n \cdot \beta_2\rangle & \cdots & \langle\beta_n \cdot \beta_n\rangle \end{bmatrix} \tag{12}$$

$$v = T^T f = \begin{bmatrix} \langle\beta_1 \cdot f\rangle \\ \langle\beta_2 \cdot f\rangle \\ \vdots \\ \langle\beta_n \cdot f\rangle \end{bmatrix} \tag{13}$$

Then according to equations (12) and (13), equation (11) can be written as $\zeta = C^{-1}v$, that is

$$C\zeta = v \tag{14}$$

C is a basic number product matrix. v is a base vector, which is a base projection. Knowing the measured values and the basic equation, we can use a classical equation (14) to obtain the optimal regression factor.

III. THE PARAMETER UPDATE METHOD OF THE REGRESSION MODEL

The data collected by nodes in the real-time monitoring of wireless sensor networks will gradually increase over time [5]. Due to the limitation of its energy, storage, and processing capacity, it can only store sampled data for some time. If the linear regression model obtains the required data, the following methods can be used for correction. Assume that the target product moment C of the basis functions and the projection vector v have been calculated in the r_1 to r_{m-1} sampling period. Then for the new measurement at time r_t we have

$$C(r_t)=\begin{bmatrix} \langle \beta_1(r_t)\cdot\beta_1(r_t)\rangle & \cdots & \langle \beta_1(r_t)\cdot\beta_n(r_t)\rangle \\ \langle \beta_2(r_t)\cdot\beta_1(r_t)\rangle & \cdots & \langle \beta_2(r_t)\cdot\beta_n(r_t)\rangle \\ \vdots & \vdots & \vdots \\ \langle \beta_n(r_t)\cdot\beta_1(r_t)\rangle & \cdots & \langle \beta_n(r_t)\cdot\beta_n(r_t)\rangle \end{bmatrix},$$

$$v(r_t)=\begin{bmatrix} \langle \beta_1(r_t)\cdot f(r_t)\rangle \\ \langle \beta_2(r_t)\cdot f(r_t)\rangle \\ \vdots \\ \langle \beta_n(r_t)\cdot f(r_t)\rangle \end{bmatrix}$$

Then the product matrix C of the number of basis functions and the projection vector v in the new sampling period is:

$$C \leftarrow C + C(r_t); v \leftarrow v + v(r_t) \tag{15}$$

In the scale of matrix C and vector v, the sliding window structure is used in this paper [6]. The computing power, storage capacity, and the size of the sliding window set by the actual demand increase with the increase of the scale of matrix C and vector v. After the data at time r_1 exceeds the sliding window setting, the revised C and v can be calculated according to formula (16).

$$C \leftarrow C + C(r_1); v \leftarrow v - v(r_1) \tag{16}$$

This method can not only obtain the regression factor from the linear equation system of $C\zeta = v$, but also can use the incremental method to revise the equation by vector and matrix.

IV. PERFORMANCE ANALYSIS AND SIMULATION TEST

4.1 Performance Evaluation

The time sequence n is to determine the storage difficulty $O(n)$ of the method in case of adjustment [7]. According to the operation process of this method, the amount of operation is also $O(n)$. The algorithm has $O(1)$ level computational complexity and $O(1)$ level storage capacity without occupying memory. So the overall complexity of the method is $O(n)$. For the actual data acquisition, data transfer, and average relatively low-performance index of WSN, this paper proposes a model based on RROEAQ to evaluate its performance:

$$ROEAQ = ERR \times QTY \tag{17}$$

ERR indicates the mean absolute error between the known data and the expected data. QTY refers to the sum of data collection and data transfer [8]. This method is also less error-prone as the value of $ROEAQ$ gets lower. The method's overall performance is related to the amount of model conditioning and more $ROEAQ$. Because the characteristic of this method is to collect and process the data of period and time series in real-time. It can be used in ecological environments, soil moisture, meteorological, and other fields. Therefore, its scalability is excellent.

4.2 Simulation test

Four performance parameters, including the number of model adjustments, data acquisition, data transmission, and mean absolute error, were compared and analyzed through Matlab simulation experiments [9]. In this paper, a miniature wireless sensor network monitoring device is established to verify the prediction accuracy of the method for temperature sampling data with time series correlation.

4.2.1 Simulation experiment

The characteristic of this method is that it is based on WSN acquisition and data processing and does not consider the network topology and other issues. In this paper, Matlab simulation software can well reflect the characteristics of this method [10]. The method proposed in this paper meets the needs of simulation experiments. The data for the simulation experiments came from the Berkeley-Intel team monitoring 10,000 temperature data of 1-1 through the existing sensor network. The parameter settings of the simulation test are shown in Table 1 below.

Table 1 Simulation experiment parameter settings

Simulation parameters	Value
Time series length n	6
Acquisition period T	1min
Upload cycle C	10
Total data	10000

Figure 1 compares the adjustment numbers of the two algorithms at different error thresholds ε. It can be seen from Figure 1 that the number of model adjustments decreases as the error threshold ε increases. Since the model adjustment using the piecewise linear regression method is smaller than the data encoding method, the computational complexity is slight.

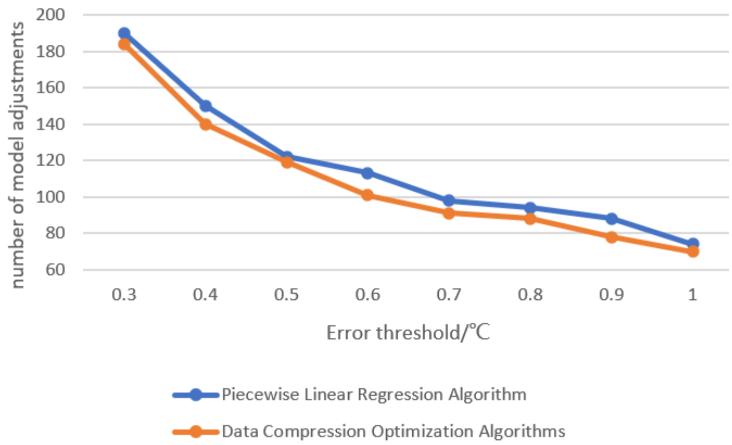

Figure 1 The effect of ε on the number of model adjustments

Figure 2 shows the ROEAQ comparison of the two algorithms under different error thresholds ε. Because ROEAQ represents the sum of errors and data, it only does numerical operations and has no other units.

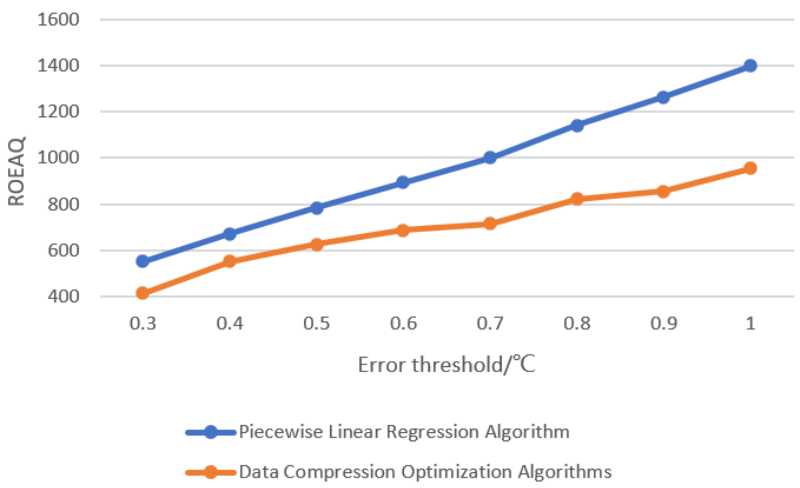

Figure 2 The effect of ε on ROEAQ

Simulation experiments show that the data collected and transmitted data decreases when the error threshold increases. The average absolute error increases. In sampling nodes, the piecewise linear regression algorithm adaptively reduces the data collection speed by adjusting the sampling points. Its processing speed is much better than the traditional compression algorithm [11]. Compared with the traditional data encoding method, the ROEAQ using the piecewise linear regression method is lower, and the overall effect is better. The user can recover the collected data within the allowable error range. At the same time, users can weigh this against the amount of data collected and transmitted.

4.2.2 Test inspection

The experiment of this part is to establish a virtual test platform to test the correctness of the method. This paper uses the CC1110 chip and SHT11 sensor to form a WSN network. The sensor of SHT11 is used for temperature measurement, and CC1110 is used for data processing and wireless transmission. Two intra-cluster nodes, a cluster head node, and a single node distributed indoors form a miniature wireless sensor network. The system is used for real-time monitoring of the indoor environment.

The temperature information obtained from node one from 9:00 to 18:00 is the initial data. There are two types of daily temperature variation trends: large temperature fluctuations and the same temperature variation trends. The rate of change is significant, and the trend of temperature change is the same. This paper embeds the method in two nodes and collects and processes the temperature data in two cycles. The corresponding time and temperature change characteristics are shown in Table 2. In this paper, the collection period of each sample is T = 10 minutes, and there are 25 sampling points. The data sequence is 6, and the data upload period is C=10.

Table 2 Time and temperature change characteristics of each period

Period	Start and end time	Change characteristics
1	9:00-13:00	large temperature fluctuations
2	14:00-18:00	The temperature drops faster

Figures 3 to 4 compare the monitoring results of the two periods with the predicted results of the regression model. Table 3 compares the amount of data collected, the amount of transmitted data, the number of model adjustments, and the absolute mean error between the two periods.

From Figure 3 and Table 3, it can be seen that the data flow of data collection and transmission in Phase 1 is substantial, caused by the violent fluctuation of air temperature. It can be seen from Figure 4 and Table 3 that the second stage is because the temperature change rate is significant and the collection period is short, so a large amount of data is collected. Because the sensory data conforms to a regression model, the amount of data transmitted is small.

Table 3 Comparison table of performance indicators for each period

Period	Collection amount	Transmission volume	Adjustment times	Mean absolute error/℃
1	23	19	3	0.0581
2	19	13	2	0.0879

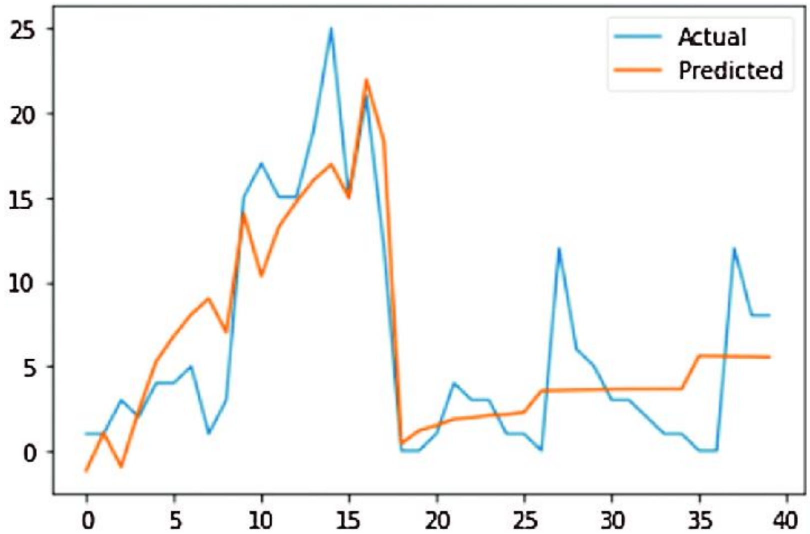

Figure 3 Comparison of large temperature fluctuations in period 1

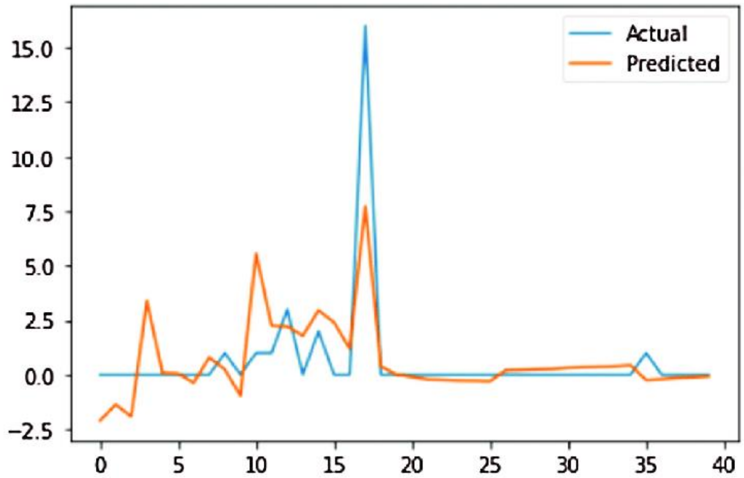

Figure 4 Comparison of rapid temperature reduction in period 2

V. CONCLUSION

In this paper, the data acquisition method of WSN is given by using the time dependence and nonlinear regression method. The purpose is to overcome the shortcomings of current WSN data collection and compression methods. This paper compares the performance of the piecewise linear regression method with the experimental results. It proves that this method can reduce the data's sampling amount and transfer rate. The deviation between the forecast results and the monitoring results is minimal. This illustrates the effectiveness and efficiency of the algorithm. The algorithm proposed in this paper is suitable for applying the period and timing relationship.

REFERENCES

[1] Pu, Z., Zhu, M., Li, W., Cui, Z., Guo, X., & Wang, Y. Monitoring public transit ridership flow by passively sensing Wi-Fi and Bluetooth mobile devices. IEEE Internet of Things Journal.,2020; 8(1): 474-486

[2] Zou,H. & He, D. Technology sharing game from ecological perspective. Applied Mathematics and Nonlinear Sciences.,2021; 6(1): 81-92

[3] Jun, H., Chuan-Xi, L., Chun-Long, X., Wei-Xiong, L., Jun-Jian, Y. & Lozano, A. Decision-making system and verification of pavement diseases treatment scheme for highway reconstruction and extension. Applied Mathematics and Nonlinear Sciences.,2021;6(1): 151-162

[4] Sassi, A., Bachir, A., & Bechkit, W. Evaluating Regression Models for Temporal Prediction of Wi-Fi Device Mobility. Wireless Personal Communications.,2021; 116(3): 2169-2186

[5] Zhao, B., Hu, K., Gong, Y., & Shen, Y. A tensor-based joint AoA and ToF estimation method for Wi-Fi systems. IEEE Wireless Communications Letters.,2021; 10(11): 2543-2546

[6] Sridher, T., Sarma, A. D., & Naveen Kumar, P. Performance Evaluation of Onboard Wi-Fi Module Antennas in Terms of Orientation and Position for IoT Applications. International Journal of Engineering.,2022; 35(10): 1918-1928

[7] Sakhare, R., & Vanajakshi, L. Reliable corridor level travel time estimation using probe vehicle data. Transportation Letters.,2020; 12(8): 570-579

[8] Zhao, B., Hu, K., Gong, Y., & Shen, Y. A tensor-based joint AoA and ToF estimation method for Wi-Fi systems. IEEE Wireless Communications Letters.,2021; 10(11): 2543-2546

[9] MASRİ, İ., & ERDAL, E. Developing A New Scheduling Algorithm for Wi-Fi 6 Technology Based on Machine Learning. International Journal of Engineering Research and Development.,2020; 14(1): 322-337

[10] Seo, J., Cho, C., & Won, Y. Enhancing the reliability of Wi-Fi network using evil twin AP detection method based on machine learning. Journal of Information Processing Systems.,2020; 16(3): 541-556

[11] Khan, M. A., Hamila, R., Gastli, A., Kiranyaz, S., & Al-Emadi, N. A. ML-Based Handover Prediction and AP Selection in Cognitive Wi-Fi Networks. Journal of Network and Systems Management.,2022; 30(4): 1-21