Agricultural Data Rectification Algorithm Based on Internet of Things

Na Li, Qingxue Li and Huarui Wu

ABSTRACT

Agricultural information data in Internet of Things is the basis of abnormal environment warning system, however, these data in the measurement process will inevitably contain random error and gross error. For the problem of exceptional data existing in wireless sensor networks, a data rectification model based on the detection of gross error is proposed, which can automatically eliminate and compensate the gross error in measured data, in order to achieve the purpose of accurate data rectification. First, utilize the statistical tests method to detect and identify gross error in measured data, and the method can detect more than three gross errors, which show superiority in the number of gross error than other gross error detection methods. Then, median filter is exploited to estimate the value of gross error identified. At last, three order exponential smoothing is used to data reconciliation. Experimental results show that this method can achieve effective performance.

KEYWORDS

Wireless sensor networks; Gross error detection; Data rectification; Statistical tests

INTRODUCTION

In practical applications, the environmental monitoring data, such as soil temperature and humidity, air temperature and humidity, will inevitably contain random error and gross error in the process of measurement. Compared with the

Na Li, Qingxue Li(corresponding author), Huarui Wu*, National Engineering Research Center for Information Technology in Agriculture, Beijing Nongke Mansion, No 11 Shuguang Huayuan Middle Road, Banjing, Haidian, District, Beijing, P.R. China
random error, the probability of gross error occurred significantly lower but the error amplitude is greater that it can’t be ignored. If using data rectification[1][2] to the measured data directly, gross error will be diffused into the normal data, therefore, combing the gross error detection and data coordination technology for data rectification can improve the accuracy of data rectification when removing gross error.

At present, the most common and effective gross error detection method is based on statistical hypothesis testing. Gross error detection methods based on statistical hypothesis testing include Global Test[3], node Test[4], Measurement Test [5] and methods based on Akaike information criterion[6]. For multiple gross errors detection methods are mainly MT-NT joint methods[7][8], sequence removal methods[9], series compensation methods [10] and synchronous compensation methods[11] and so on.

Existing gross error detection methods[12] usually detected only three or fewer gross errors. In this paper, we propose a new two-stage gross error detection method, in the first stage, hypothesis testing method is adopted to detect the gross error in measured data, and then, median filter method is used to eliminate gross error, in the second stage, data reconciliation based on three order exponential smoothing is exploited to improve the accuracy of data rectification.

The remainder of this paper is organized as follows. In section gross error detection algorithms, we describe the proposed data rectification algorithm. In section experiments, we show experimental results to demonstrate the effectiveness of our proposed algorithm. Finally, we conclude this paper.

**GROSS ERROR DETECTION ALGORITHMS**

Set \( y \) as measured data vector, and \( x \) is the corresponding true value. Suppose that there is no gross error in measured data \( y \), and then, the model of measuring process can be described as

\[
y = x + \varepsilon
\]

where \( \varepsilon \) is the random error, and \( \varepsilon \) is assumed that \( \varepsilon \sim N(0, Q) \), which follows a multivariate normal distribution with zero mean and a \( \sigma_i^2 \) variance, then the corresponding steady state data reconciliation problem can be written as follows

\[
\min \left( y - x \right)^T Q^{-1} \left( y - x \right)
\]

s.t. \( Ax = 0 \)  

(2)

where \( A \) is the coefficient matrix of constraint equations. According to the theory of least square method, measured data \( y \) can be estimated by

\[
\hat{x} = y - QA^T \left( AQA^T \right)^{-1} A y
\]

(3)

206
In the meanwhile, we can get the residual error $\hat{\varepsilon}$ between $y$ and $\hat{x}$

$$\hat{\varepsilon} = y - \hat{x} = QA^T (AQA^T)^{-1} Ay \quad (4)$$

Hypothesis testing method is adopted to detect gross error, first of all, establish the following null hypothesis and alternative hypothesis:

Null hypothesis $H_0$ : there is no gross error in $y$, namely $E(\hat{\varepsilon}) = 0$,

Alternative hypothesis $H_1$ : gross errors exist in $y$, namely $E(\hat{\varepsilon}) \neq 0$.

On this Occasion, record the probability of type I class mistake as $P \left( H_0 \text{ is true but rejected} \right)$, certainly, it’s impossible to rule out the possibility of this kind of mistake, therefore, we should try to reduce the possibility of this kind of mistake as small as possible, namely

$$P \left( H_0 \text{ is true but rejected} \right) \leq \alpha \quad (5)$$

Assuming that measured data $y$ follows the normal distribution, and only contains random error $\varepsilon \sim N \left(0, \sigma^2 \right)$. Build the statistical magnitude $|Z| = \left| \frac{\varepsilon}{\sigma} \right|$, due to the biggest probability for this kind of mistake allowed is $\alpha$, so

$$P \left( H_0 \text{ is true but rejected} \right) = P \left( \left| \frac{\varepsilon}{\sigma} \right| > Z_{\alpha/2} \right) = \alpha \quad (6)$$

with $|Z| = \left| \frac{\varepsilon}{\sigma} \right| \sim N \left(0, 1 \right)$, see figure 1 for the standard normal distribution.
When $H_0$ is true, If measured data satisfy

$$|Z| = \left| \frac{\varepsilon}{\sigma} \right| > Z\alpha/2$$

then, reject $H_0$, else accept $H_0$. Equation (7) is the gross error detection model. Data rectification method is summarized in **Algorithm 1**, see figure 2 for details.

**Algorithm 1**

**Input:** True data $x_i$, $\sigma_j = x_j * 2.5\%$; synthetic test data $y_i$, $\alpha = 0.05$.

1. Compute $|Z| = \left| \frac{\varepsilon}{\sigma} \right|$ by equation (4).
2. If $|Z_i| > Z\alpha/2 = 1.96$, rectification value $\hat{x}_i = \text{median} \left( y_i, \hat{5} \right)$.
3. Else $\hat{y}_i = y_i$.
4. Fitting $\hat{y}_i$ with three order exponential smoothing[13].

**Output:** Estimated data $\hat{x}$

---

**EXPERIMENTS**

To verify the performance of our proposed data rectification method, Monte Carlo method is used to generate the measured data, which is also used by many other scholars in simulation test of data reconciliation.
Table I. Experimental data.

<table>
<thead>
<tr>
<th>No.</th>
<th>True data (℃)</th>
<th>Random error (℃)</th>
<th>Test data (℃)</th>
<th>Gross error (℃)</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28.6</td>
<td>-0.395</td>
<td>28.21</td>
<td>-0.552</td>
<td>0.552</td>
</tr>
<tr>
<td>2</td>
<td>28.5</td>
<td>0.711</td>
<td>29.21</td>
<td>-</td>
<td>0.999</td>
</tr>
<tr>
<td>3</td>
<td>28.5</td>
<td>0.003</td>
<td>28.50</td>
<td>-</td>
<td>0.004</td>
</tr>
<tr>
<td>4</td>
<td>28.4</td>
<td>0.343</td>
<td>28.74</td>
<td>-</td>
<td>0.483</td>
</tr>
<tr>
<td>5</td>
<td>28.2</td>
<td>-0.427</td>
<td>27.77</td>
<td>-</td>
<td>0.606</td>
</tr>
<tr>
<td>6</td>
<td>28</td>
<td>-0.538</td>
<td>27.46</td>
<td>-</td>
<td>0.768</td>
</tr>
<tr>
<td>7</td>
<td>27.9</td>
<td>-0.045</td>
<td>27.85</td>
<td>-</td>
<td>0.065</td>
</tr>
<tr>
<td>8</td>
<td>27.8</td>
<td>-0.126</td>
<td>27.67</td>
<td>-</td>
<td>0.182</td>
</tr>
<tr>
<td>9</td>
<td>27.8</td>
<td>0.597</td>
<td>28.40</td>
<td>-</td>
<td>0.860</td>
</tr>
<tr>
<td>10</td>
<td>27.7</td>
<td>0.303</td>
<td>28.00</td>
<td>-</td>
<td>0.438</td>
</tr>
<tr>
<td>11</td>
<td>27.8</td>
<td>0.270</td>
<td>32.07</td>
<td>4</td>
<td>6.144</td>
</tr>
<tr>
<td>12</td>
<td>28</td>
<td>-0.722</td>
<td>27.28</td>
<td>-</td>
<td>1.032</td>
</tr>
<tr>
<td>61</td>
<td>29.8</td>
<td>-0.038</td>
<td>29.76</td>
<td>-</td>
<td>0.051</td>
</tr>
<tr>
<td>62</td>
<td>29.6</td>
<td>-0.237</td>
<td>24.36</td>
<td>-5</td>
<td>7.077</td>
</tr>
<tr>
<td>63</td>
<td>29.1</td>
<td>1.092</td>
<td>30.19</td>
<td>-</td>
<td>1.501</td>
</tr>
<tr>
<td>64</td>
<td>28.9</td>
<td>0.405</td>
<td>29.30</td>
<td>-</td>
<td>0.560</td>
</tr>
<tr>
<td>65</td>
<td>28.6</td>
<td>0.358</td>
<td>28.96</td>
<td>-</td>
<td>0.501</td>
</tr>
<tr>
<td>66</td>
<td>29.3</td>
<td>0.390</td>
<td>35.69</td>
<td>6</td>
<td>8.723</td>
</tr>
<tr>
<td>67</td>
<td>29.2</td>
<td>0.192</td>
<td>29.39</td>
<td>-</td>
<td>0.263</td>
</tr>
<tr>
<td>68</td>
<td>29.1</td>
<td>0.150</td>
<td>29.25</td>
<td>-</td>
<td>0.206</td>
</tr>
<tr>
<td>69</td>
<td>29</td>
<td>0.348</td>
<td>25.35</td>
<td>-4</td>
<td>5.037</td>
</tr>
<tr>
<td>70</td>
<td>28.9</td>
<td>-0.056</td>
<td>28.84</td>
<td>-</td>
<td>0.078</td>
</tr>
<tr>
<td>71</td>
<td>28.7</td>
<td>-0.019</td>
<td>28.68</td>
<td>-</td>
<td>0.027</td>
</tr>
<tr>
<td>72</td>
<td>28.6</td>
<td>0.044</td>
<td>28.64</td>
<td>-</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Figure 3. Experimental and estimated data.

Experimental Data

Experimental data are from the Beijing Xiaotangshan modern agricultural demonstration base of the real air temperature data, choose air temperature data a day of the monitoring station from 2:00 on July 5, 2016 to 2 o’clock on July 6, 2016 as test data, the experiment selected the air temperature data follows normal distribution. Monitoring frequency for 3 times per hour, then, we get a total of 72 data records. Random noise data vector is added to true data vector, we get the measured data, as shown in table I. Data rectification results see Figure 3.
Results

Select average number of type I error (AVTI) and Overall Power (OP) to measure the effectiveness of gross error detection, the definition of AVTI and OP are as follows.

\[
AVTI = \frac{\text{Numbers of gross errors incorrectly identified}}{\text{Numbers of simulation trials}} \quad (8)
\]

\[
OP = \frac{\text{Numbers of gross errors correctly identified}}{\text{Numbers of gross errors simulated}} \quad (9)
\]

In experiments, these two indicators are obtained after 100 simulation experiments by Matlab software, see Table II. From Table II, it’s easy to find that the indicator OP is close to 1, which indicates that gross error detection method based on hypothesis testing can achieve a higher cognition rate. Additionally, the results of AVTI are rather small, indicating that this method makes lower probability of Type I error.

<table>
<thead>
<tr>
<th>No.</th>
<th>Value</th>
<th>AVTI</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>6</td>
<td>0.13</td>
<td>0.88</td>
</tr>
<tr>
<td>66,62</td>
<td>6, -5</td>
<td>0.10</td>
<td>0.93</td>
</tr>
<tr>
<td>66,62,11</td>
<td>6, -5,4</td>
<td>0.09</td>
<td>0.95</td>
</tr>
<tr>
<td>66,62,11,69</td>
<td>6, -5,4, -4</td>
<td>0.14</td>
<td>0.91</td>
</tr>
</tbody>
</table>

CONCLUSION

Through the research on the normal distribution model, a new data rectification method is proposed by jointly statistical hypothesis testing method and smoothing method. Experimental results show that this method can achieve effective performance.

ACKNOWLEDGEMENT

This work was supported by Youth Found of Beijing Academy of Agriculture and Forestry Sciences(Research on Management and Analysis of Internet of Things of Agricultural Facility Data) and Beijing Municipal Natural Science Foundation (Grant number 4151001).

REFERENCES


