Data Fusion of WSN Based on CRFSD Algorithm

Qidong YONG, Xiaoqiang CI, Dongyong XIA, Qi ZHANG, Shifu ZHANG

Dept. of POL Application and Management, Logistic Engineering University, 401311, Chong Qing, China
National Engineering Research Center for Disaster and Emergency Rescue Equipment, Logistic Engineering University, 401311, Chong Qing, China

Tel.: +86-13983385898, fax: +86-023-86730850, E-mail:yqd170@sina.com

Received: /Accepted: 9 January 2014 /Published: 31 January 2014

Abstract: WSN are a sort of self-organizing networks, and the data redundancy will be produced in the procedure of gather of graphics and video or the data. The data fusion is needed. On the basis of the previous method of data fusion, the data method based on information support (that is CRFSD) is provided, and through the simulation experiment, which is held in the time domain, spatial domain, temporal and spatial domain, the data fusion is realized. Copyright © 2014 IFSA Publishing, S. L.

Keywords: Data redundancy, Data fusion, Information support.

1. Introduction

Wireless sensor networks, which are set include with information collection, communication and computing capabilities, is composed with hundreds of tiny, inexpensive sensor nodes. When it does, information is collected by self-organizing wireless network multi hop relay to the user terminal, and a big number of the graphics and data information for real-time collection transmitted through this self-organizing network, the perception of environmental information will be transmission through multi channel interaction. At the same time, in the perception of a big number of nodes and information transmission, data redundancy will be produced inevitably; the best solution is data fusion [1].

2. The Necessarily of Data Fusion in WSN

As wireless sensor networks are self-organizing network, in practical applications, most of the sensor nodes are deployed in harsh environments, by the constraints of cost and energy consumption, then sensor accuracy is generally lower. In addition, images, audio and videos collected by its node are transmission through radio, the amount of data that the introduction of the sensor information using inter node communication is more susceptible to interference, error data are generated, data transmitted to an aggregation node by each node individually is obviously inappropriate. At the same time, a big number of redundant information in the network will waste a lot of bandwidths and valuable energy resources will reduce the efficiency of the
information collected and affect the timeliness of the information collection, then data fusion will be essential to be applied in the WSN. That is to say that WSN will have a lot of redundant data in the procedure of data transmission [2]. The main purpose of the data fusion is to remove redundant data in the network, minimize the transmission of redundant information, and then accurate information will be sent in place in a timely manner. If these redundant data are not dealt with, the energy consumption of the sensor node will be so fast that the lifetime of the WSN. From above, the application of data fusion technology will play a great role in three aspects. One is to save energy, and the second aspect is to obtain more accurate information, and the last aspect is to improve the efficiency of data collection [3].

3. Description of the Algorithm

There are many sorts of data fusion algorithms currently. As known to all, if the priori knowledge of measurement information is known as the premise, such as noise intensity, priori probability, the associated probability, distributed monitoring algorithm, space integration algorithm, attribution fusion algorithm, situation assessment algorithm and threat estimation algorithm, and classical algorithm such as Bayesian statistical inference, evidence theory, adaptive algorithm, neural network are often used. In WSN, sensor nodes are arranged in a model of self-organizing network, the priori information can’t be determined, and therefore, the algorithm which is used without the dependent on the prior knowledge should be studied, then the algorithm will be more practical and accuracy [4].

3.1. Model of the Algorithm

In the WSN, when the same parameter is measured by multiple sensor nodes, the deviation result between sensor nodes is represented as mutual support degree of the nodes, just as the degree of mutual support between nodes. Here, support is used to indicate the consistency of the information measured among the sensor nodes. The CRFSD (Cluster Reliability Fusion Support Design) is provided to overcome the effectiveness of the energy consumption. Before the model is established, confidence distance is assumed to be $\alpha_{ij}$, which means that the degree of support of the $j$-th node to node I.

On the basis of the support matrix, to define the degree of support of the measured values of each sensor node, the exponential decay function is used. When spatial domain is fixed, the reliability of the sensor nodes in the entire measurement time are judged on the consistency sequence analyzing at the different time points in a time domain; when the time domain is fixed, the reliability of the sensor nodes are judged on the consistency by the measured value of one time point in a spatial domain. With the comprehensive utilization of space and time domain information of the sensor nodes, consistent reliability fusion based on the degree of support of the temporal and spatial domain will be achieved after the measured values weighted. Fusion algorithms is held in each cluster, and the sensor node is responsible for transmission to the perception of data to the cluster head node, then fusion algorithm will be completed by the cluster head node, and the fusion results will be sent to the sink node. Once the algorithm is feasible, conducive to network users for real target state or process information should be got, and accurate decisions and estimates will be decided [5].

3.2. Design of the Algorithm

Assume that there is an Array of sensor nodes composed with n sensor nodes, direct measurement is used, stationary or gradient parameter $X$ is monitored, then the formula can be got as follows.

$$z_i(k) = X + z_i(k) i = 1,2,...,n , \quad (1)$$

In the formula (1), $z_i(k)$ is defined as the measured value of the $i$-th sensor node at the time point of $k$; $X$ is the monitored target; $v_i(k)$ represents the measurement noise at the time point of $k$, at the same time $E[v_i]$ and $D[v_i]$ are unknown.

If the values of $z_i(k)$ and $z_j(k)$ are quite different, it indicates that the measured values of the two sensor nodes have low mutual support, and even deviate from each other. On the contrary, if he measured values of the two sensor nodes are close. It indicates that the measured values of the two sensor nodes have high mutual support. To quantify the measured value of the degree of support of each sensor node in the same time, the index attenuation function is introduced, and the support matrix is constructed.

At the time point of $k$, the support of the measured value of sensor node $I$ and sensor node $j$ can be described as follows.

$$\alpha_{ij} = \exp\left(-X\left(z_i(k) - z_j(k)\right)^{2}\right), \quad (2)$$

The exponential decay function can be used to quantify the degree of support for sensor nodes measured values and to avoid that the degree of support whether 1 or 0 in traditional methods absolutely. In addition, $X$ is used as an adjustable parameter to adjust the size of the measurement scale, and then the program is easy to be realized. Squinting modifier accordingly, at time point of $k$, the support matrix among the sensor nodes can be described as follows.
For a row element of the support matrix of $SD(k)$, if the value of $\sum_{j=1}^{n} \alpha_{ij}(k)$ is big, it indicates that the measured value of the i-th sensor node at the time point of k is consistent with the majority of the sensor nodes; otherwise, the measured value of the i-th sensor node at the time point of k is departing from the majority of the sensor nodes.

At the time point of k, the measured value of sensor nodes consistency of i-th sensor node with the other sensor nodes can be judged by the formula as follows.

$$r_i(k) = \frac{1}{n} \sum_{j=1}^{n} \alpha_{ij}(k) / \sum_{j=1}^{n} r_j(k),$$

(4)

Obviously, $1 \gg r_i(k) > 0$. In the formula (2), the closeness of the two sensor nodes measuring value is reflected at a measurement time. In the formula (4), the closeness of the measured value of the i-th sensor node measured with all the sensor nodes (including the i-th sensor node) is reflected at a measurement time. From the formula (4), it is learned that the data fusion of the measured values of the wireless sensor nodes can be done between the spatial domain and temporal and spatial domain [6].

The spatial domain is considered first. In the environment, only the $r_i(k)$ is considered, which is referred to that the consistency measure of the sensor nodes at a time of k, without the reliability of the measured value during the scale of measurement. $x_i(k)$ is referred to that the estimate of the target after data fusion for then sensor nodes in the cluster of $z$ at the time of k, then the formula of the consistency measure for availability can be described as follows.

$$x_i(k) = \frac{1}{n} \sum_{j=1}^{n} r_i(k) z_j(k) / \sum_{j=1}^{n} r_j(k) ,$$

(5)

From above, it is learned that the target fusion estimation algorithm based on the consistency measure is in low complexity, but the implication of the reliability of the sensor nodes consistency measured was not dug in different time.

The other domain is temporal and spatial domain. In the environment, reliability of the wireless sensor nodes in different time is considered, and information of the sensor nodes is leaded.

At a moment of measurement of the sensor nodes, $r_i(k)$ as the sensor node consistency measure of the measurement value may be high, but it does not indicate the high reliability of sensor nodes in the entire measurement range, the reliability of the sensor nodes is manifested through consistency measure of all measurement points in all time. For example, $r_i(k)$ may be big at a certain moment, but also may become small at a moment. It indicates that the measured consistency of sensor nodes is unstable and that is to say that the reliability of the sensor nodes within the entire measuring time is poor. In the process of data fusion of measurement values, the sensor node which has the big consistency of the measured value and a stable consistency variation can be used.

The implied reliable information of the consistency measure in the different moments could be studied by the sample mean and variance of the statistical theory:

1) At the time point of k, the measured value consistency mean of the i-th sensor node can be described as follows.

$$\bar{r}_i(k) = \frac{1}{K} \sum_{i=1}^{K} r_i(k) z_i(t),$$

(6)

In order to shorten the cluster head of the sensor nodes time of computing and saving storage space, recursive form can be used, then formula (7) can be derived from formula (6), as follows.

$$\bar{r}_i(k) = \frac{k-1}{k} r_i(k-1) + \frac{1}{k} r_i(k),$$

(7)

The measured value consistency variance can be described as follows.

$$\sigma^2_i(k) = \frac{1}{k} \sum_{i=1}^{k} \left[ \bar{r}_i(k) - \bar{r}_i(t) \right]^2 ,$$

(8)

The formula (9) can be derived from formula (8) as follows.

$$\sigma^2_i(k) = \frac{k-1}{k} \sigma^2_i(k-1) + \frac{1}{k-1} \left[ \bar{r}_i(k) - \bar{r}_i(t) \right]^2,$$

(9)

In the actual process of data fusion, the information of those sensor nodes which have big measured value consistency mean and small measured value consistency variance should be taken full advantage. In other words, the information of the sensor nodes which have high consistency and reliability should be fully used [7].

2) At the time point of k, $q_i(k)$ is defined as the weighting coefficient of the i-th sensor node measured value. In the procedure of data fusion, the information of the sensor nodes which have high consistency and reliability should be fully use, then $q_i(k)$ is in the direct bound up with $r_i(k)$, and in the opposite bound up with $\sigma^2_i(k)$. To ensure
\( q(k) \) nonnegative, the linear function is chosen as the weighting coefficient. Then formula (10) can be got as follows.

\[
q_i(k) = \left[1 - \varphi \sigma^2_i(k)\right] r_i(k),
\]

(10)

In the formula (10), \( \varphi \) is an adjustable parameter, the extent could be influenced by the adjust of the parameter for \( \sigma^2_i(k) \).

From above, for a time ray of \( k(1), k(2), k(3), \ldots, k(t) \), \( x_z(k) \) can be described as follows.

\[
x_z(k) = \sum_{i=1}^{n} q_i(k) z_i(k) / \sum_{i=1}^{n} q_i(k),
\]

(1)

In the formula (11), \( x_z(k) \) refers that the estimate values of the i-th sensor node in the sensor node cluster of \( z \) after data fusion in the temporal and spatial domain.

Obviously, formula (1) to formula (5) belong to the method of the first kind while formulas (1) to formula (11) are the formulas of the data fusion.

4. Simulation

To simulate the data fusion, sixteen sensor nodes and one cluster head node are chosen as the research objects. In the experiment, the ID numbers of the three sensor nodes are defined as S1, S2 and S3.

The target of the sensor nodes is the temperature, which has the true value of 20 degrees Celsius; however, the noise of the sensor node distribution is unknown. The received measured values of the cluster head sensor nodes are composed of 14 X 16 group numbers, listed as Table 1 [8].

For the numbers of Table 1, methods of the spatial domain fusion, temporal and spatial domain, and time domain fusion are used to be simulated. In the simulation, exponential decay coefficient \( X \) is defined as 0.8 while \( \varphi \) is defined as 0.01. The result of the data fusion after three kinds of methods used is listed in Fig. 1.

5. Conclusions

Two results are easy to be seen from Fig. 1 as follows.

1) Judging from the vertical axis, which is represented of the process of change of absolute error, the data fusion of the entire simulation process based on the temporal and spatial domain (shown by the blue line), the absolute error is generally lower than the result of time-average domain fusion method (shown by purple lines) and spatial domain fusion method (shown by brown lines), and the overall

<table>
<thead>
<tr>
<th>Sensor nodes</th>
<th>Measure times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>S1</td>
<td>19.4</td>
</tr>
<tr>
<td>S2</td>
<td>19.2</td>
</tr>
<tr>
<td>S3</td>
<td>19.6</td>
</tr>
<tr>
<td>S4</td>
<td>19.3</td>
</tr>
<tr>
<td>S5</td>
<td>19.7</td>
</tr>
<tr>
<td>S6</td>
<td>19.8</td>
</tr>
<tr>
<td>S7</td>
<td>19.2</td>
</tr>
<tr>
<td>S8</td>
<td>18.9</td>
</tr>
<tr>
<td>S9</td>
<td>20.1</td>
</tr>
<tr>
<td>S10</td>
<td>19.2</td>
</tr>
<tr>
<td>S11</td>
<td>19.6</td>
</tr>
<tr>
<td>S12</td>
<td>18.9</td>
</tr>
<tr>
<td>S13</td>
<td>19.8</td>
</tr>
<tr>
<td>S14</td>
<td>19.5</td>
</tr>
<tr>
<td>S15</td>
<td>19.2</td>
</tr>
<tr>
<td>S16</td>
<td>20.3</td>
</tr>
</tbody>
</table>

Fig. 1. Result of the data fusion.

In the Fig. 1, the horizontal axis represents the number of observations, while the vertical axis represents the absolute error.
performance of the absolute error value change is relatively flat, which is in the convergence trend; At the same time, from the results of the spatial domain data fusion, it can be seen that in the fourth measurement time, the absolute error has its minimum, but it has dramatic changes on the whole.

It is mainly because that only the consistency of the measurement data is considered in the process of data fusion, without considering the reliability of measured values on the entire timeline. It is can be learned that only the consistency of the measurement data is considered without the overall data reliable is not accurate.

2) Among all sensor nodes at the same time, big or small and the degree of support of the measured values reflect the degree of consistency of the target for the different sensor nodes, while the consistency of the different time points of the sensor node reflects the measurement value within the entire measurement interval reliability of the sensor nodes. After the integration of the mean and variance of the sequence, weighting coefficient built, and then the measured value of each time is merged together. In this way, consistency and reliability of the entire fusion data can be a combination of organic, the data fusion simulation is completed, and the validity of the algorithm is checked.

6. Summary

Without knowing priori conditions, the application of CRFSD algorithm in the WSN can reduce most redundant data as well as complete data fusion.; During the data transmission, both the time & energy will be saved, meanwhile, the accuracy will be improved.

Acknowledgements

The thesis is supported by National Philosophy & Social Science Funds 12FJS004 and key technology of emergency equipment & industrialization funds CSTC2012GG-SFGC00002.

References

[3]. Xuehua Zhang, Multiple resource data collection, Kunming Technology University, 2011, pp. 51-59.