An Improved Fusion Method of Fuzzy Logic Based on k-mean Clustering in WSN

1, 2 Feng WANG, 2 Kang WANG, 2 Fuzhen JIANG

1 College of Information Engineering, Taiyuan University of Technology, Shanxi Taiyuan 030024, People’s Republic of China
2 Faculty of Resource & Safety Engineering, China University of Mining & Technology Beijing 100083, People’s Republic of China
1 Tel.: 13033488721
1 E-mail: wangfeng@tyut.edu.cn

Received: 11 July 2013 /Accepted: 25 September 2013 /Published: 31 October 2013

Abstract: Data fusion in wireless sensor networks is one of the key Technologies. In this paper, for the problem of large deviation of data fusion based on weighted fuzzy logic algorithm in wireless sensor networks, a new method is proposed. First to eliminate the flawed data through analysis of initial data using the idea of k-mean clustering, and to revise the weighting factors of weighted fuzzy logic algorithm with the rest authentic data, and then integrate all data by means of weighted method to get the final fusion result. Experimental results show that this method can achieve higher integration accuracy compared with the other same fusion methods.

Copyright © 2013 IFSA.

Keywords: Data fusion, Fuzzy logic algorithm, Wireless sensor networks, K-mean clustering, Weighting factors.

1. Introduction

Due to their vast potential applications, wireless sensor networks (WSN) have attracted significant attention in recent years 0. WSN consists of a large number of sensor nodes which are spatially distributed densely deployed and automatically organized to detect environment and communicate the physical information of the network covered area which is often harsh and inhospitable. As sensor nodes are usually deployed densely in large numbers, and the data collected from adjacent regions is bound to have a certain degree of redundancy, so how to deal with the redundant data to minimum energy consumption and reliably deliver the sensory information to the sink in an effective way is indeed a major challenge in such networks. Data fusion in WSN systems makes possible to solve this challenging issue, it can improve the performance of a network by aggregating information from a set of different source data, and eliminating redundancy and power consumption, ensuring fault-tolerance between nodes, and managing effectively the available communication bandwidth between network components 0. However, the data from different nodes must have a certain degree of deviation under the influence of the difference in quality of each sensor or other unpredictable factors, and that excessive deviation of data can be considered as the invalid data and should be eliminated. So it is necessary to get rid of these data for ensuring information accuracy and eliminating power consumption before making data fusion.
Many domestic and overseas algorithms with data fusion have been proposed for achieving significant energy savings by allowing nodes aggregate the data streams. These algorithms mainly include the Bayes estimation, Fisher information, the BP neural network, D-S evidence theory and etc. 0 0 0 0, and adaptive weighted fusion algorithm is adopted in Ref. 0, although this method without any priori knowledge, using the initial measurements to fusion directly, but it is complicated to calculate the variance of each node and weighting factors, which would lead to more power consumption. Literature 0 then makes use of the Least Squares with data fusion which not taking the invalid data into account could result in non-ideal fusion data despite the fact that it is simple to implement. In the Ref. 0, then a novel and improved algorithm of IRTC (Interference-Reducing Topology Control) for 3D WSN is proposed to prolong the network lifetime, but it still needs to be modified by using in the real environment. The fusion algorithm based on Fuzzy Logic provides an appropriate tool in processing the imprecise data, as it is uncomplicated in calculating weighting factors and easy to implement so it’s widely used in some specific applications, but in these practical applications, the deviation of measurements would lead to a large deviation of fusion result. In this paper we propose an improved fusion algorithm through deep study on fuzzy logic theory 0-0 and weighted fusion algorithm based on fuzzy logic 0, firstly, eliminating the error data with the method of k-mean clustering, and then revising the weighting factors of weighted fuzzy logic algorithm with the rest effective data, and aggregating the rest data to get the final fusion data. This method is simple in computation, requiring no priori knowledge and can achieve higher integration accuracy.

2. The Weighted Algorithm Based on Fuzzy Logic

The most weighted fusion algorithms are always based on clustering routing, so is the weighted fuzzy logic algorithm.

Firstly, suppose there are $n$ source nodes to measure a parameter, the sensor measurements are respectively $x_i$ at time $k$, where, $i = 1, 2, ..., n$, so $x_i(k)$ is the data transmitted by node $i$ to cluster head at time $k$ and the same meaning with $x_j(k)$, then we define the correlation between this two measurements as following:

$$a_{ij}(k) = \frac{\min \{x_i(k), x_j(k)\}}{\max \{x_i(k), x_j(k)\}}$$

(1)

According to equation 1, obtaining the $a_{ij}(k)$ range of (0,1]. If $a_{ij}(k)$ is less than a certain value, by related regulation, then we can think that these two measurements are not completely relevant 0, the computational formula of $a_{ij}(k)$ as following:

$$a_{ij}(k) = \begin{cases} a_{ij}(k) & a_{ij}(k) \geq P \\ 0 & a_{ij}(k) < P \end{cases}$$

(2)

The $a_{ij}(k)$ in the formula (2) can be defined as the membership function of fuzzy logic which can be used to ascertain the correlation between data by a simple way. But the above formula only defines the correlation between $i$ and $j$ and not reflect the correlation of $i$ or $j$ with the other nodes, and the validity of node $i$ should be reflected by $a_{1i}(k)a_{2i}(k)....a_{ni}(k)$, so we define the correlation matrix $A$ as following:

$$A = \begin{bmatrix} 1 & a_{12}(k) & \cdots & a_{1n}(k) \\ a_{21}(k) & 1 & \cdots & a_{2n}(k) \\ \vdots & \vdots & \ddots & \vdots \\ a_{ni}(k) & a_{n2}(k) & \cdots & 1 \end{bmatrix}$$

According to the matrix $A$, the data validity of node $i$ is obtained by $A_{sum}(i) = a_{1i}(k)+a_{2i}(k)+\cdots+a_{ni}(k)$ at time $k$. The higher the value of $A_{sum}(i)$, the greater the validity of data. In addition, we can see that the leading diagonal value of $A$ are all 1, it indicates the node itself is completely relevant.

Secondly, the key of weighted fusion algorithm based on fuzzy logic is to obtain each node weighting factors which can be indicated as $w_i(k)$ in there, it means the $x_i(k)$’s proportion of all data in this weighted fusion process. All weighting factors in the same cluster should satisfy with the following formula:

$$\sum_{i=1}^{n} w_i(k) = 1,$$

(3)

In the above-mentioned, the validity of $x_i(k)$ should be determined by $a_{ij}(k)a_{1j}(k)....a_{nj}(k)$, thus, all the information that attached to the correlation $a_{1i}(k)a_{2j}(k)....a_{ni}(k)$ of $x_i(k)$ should be considered when we determine the weighting factor of $x_i(k)$, then the following formula can be obtained:

$$w_i(k) = c_1a_{1i}(k)+c_2a_{2i}(k)+\cdots+c_na_{ni}(k)$$

(4)

Consider to the correlation between $i$ and $j$ is equal to the correlation between $j$ and $i$, that is $a_{ij}(k) = a_{ji}(k)$, which means that the matrix $A$ is symmetric matrix so that the matrix $A$ must have
one Eigen value \( \lambda(\lambda > 0) \) and one eigenvector \( R \) relate to this \( \lambda \), the \( R \)'s elements can be recorded as \( R_i(k) \) 0, then we can get:

\[
AR = \lambda R(\lambda > 0), \tag{5}
\]

The formula (5) can be turned into the following according to related matrix properties:

\[
W = AC, \tag{6}
\]

where \( W = [w_1(k), w_2(k), \ldots, w_n(k)]^T \), and \( C = [c_1, c_2, \ldots, c_n]^T \). Above formula can be turned into formula (7) by relevant linear transformation.

\[
W = \lambda P, \tag{7}
\]

In equation (7), we record elements of \( p \) as \( p_1(k) + p_2(k) + \cdots + p_n(k) \) and make process to get the following formula:

\[
W_i(k) = p_i(k)/(p_1(k) + p_2(k) + \cdots + p_n(k)) \tag{8}
\]

From the formula (8) it can be seen that key of getting the weighting factor of each node is to obtain the matrix \( P \), however, the matrix \( P \) is turned by matrix \( A \) through many times, so the most significant point of all process of data fusion is how to get matrix \( A \). The correlation can be represented by membership function, it is most suitable for membership function with Gaussian distributions representing the correlation through a lot of experiments and data analysis.

\[
\phi(x) = e^{-\frac{(x-a)^2}{2b^2}}, \quad a > 0, b > 0, \tag{9}
\]

where \( a \) and \( b \) are the mean and variance of measurements in Gaussian distributions respectively, \( x_i(k) \) is the transmission value by \( i \) at time \( k \), and the membership of \( x_i(k) \) can be obtained by the following formula:

\[
p_i(k) = e^{-\frac{(x_i(k)-a)^2}{2b^2}}, \tag{10}
\]

The formula (10) substituted into the formula (8), and then achieving data fusion result according to the following formula:

\[
Y = w_1(k)x_1(k) + w_2(k)x_2(k) + \cdots + w_n(k)x_n(k) \tag{11}
\]

The data fusion model is shown in Fig. 1.

Using the method of error analysis based on k-mean clustering proposed in this paper, eliminating the invalid data beyond the required precision by analyzing the initial data, constructing correlation matrixes through weighted fuzzy logic algorithm, and calculating weighted factors, then multiplying with the corresponding data to get the final fusion data.

3. Error Analysis Method Based on k-mean Clustering

Assume that there are \( n \) sensors to measure a parameter at one time, and getting \( n \) measurements, the sensor measurements are respectively \( x_i(k) \), we can adopt the k-mean clustering method to find out the flawed data. In the control areas, it is used to divide the \( n \) measurements into several classes, and to determine whether each cluster is the efficient data set according to compare the clustering center distance, then to eliminate these invalid data to achieve the purpose of error data separation. The procedures of k-mean clustering algorithm are as follows:

1) Sort the \( n \) data according to the order of magnitude for expediting the separation of failure data, and choosing the minimum \( X_{\text{min}} \), median \( X_{\text{mid}} \), maximum \( X_{\text{max}} \) as the initial centers which are recorded respectively \( Z_1(1), Z_2(1), Z_3(1) \);

2) All of the \( n \) data are partitioned respectively to the corresponding clusters if that is closest to center points, the constructed datasets of \( k \) centers are respectively recorded as \( S_j(k) \), \( j = 1,2,3 \)

\[
D_j(k) = \min_i (\| X_i - Z_j(k) \|), \quad j = 1,2,3 \tag{12}
\]
where \( k \) is the number of iterations, \( D_i(k) \) represents the distance that the \( i \)th data from its nearest cluster center, and \( Z_j(k) \) is the cluster center.

3) By equation (13), recalculate the value of clustering centers after each round of iteration, equation (13) as following:

\[
Z_j(k+1) = \frac{1}{N_j} \sum_{x \in \text{set}_j(k)} X_j, \quad j=1,2,3, \quad (13)
\]

4) If \( Z_j(k+1) \neq Z_j(k), \quad j=1,2,3 \), then return to step 2) to repeat iteration until \( Z_j(k+1) = Z_j(k), \quad j=1,2,3 \).

5) Calculate the distance from \( Z_2(n) \) to \( Z_1(n) \) and \( Z_3(n) \) which are respectively \( d_1 \), \( d_2 \), the formula as following:

\[
d_1 = Z_2(n) - Z_1(n), \quad d_2 = Z_3(n) - Z_2(n)
\]

and now we need to drawing one conclusion that which \( k \) cluster should be eliminated according to the relationship of \( |d_1| \) or \( |d_2| \) with \( |\delta| \) which determined by error-tolerance \( \delta \) which being set up in advance. If \( |d_1| \) or \( |d_2| \) is greater than \( \delta \), then the data belong to \( Z_1(n) \) or \( Z_3(n) \) can be thought as failure data and need to be removed, otherwise, keep them. The value of \( \delta \) is set based on data attribute and the requested precise.

4. Experimental Results

In this section, in order to detect the validity of the improved method, we perform an experiment on the simulation datasets with Matlab. Assumed that there are 10 measurements from 10 sensors need to be fused at one time, that is \( n=10 \), and the estimated true value is set to 1. Specific measurements and the variance of each sensor are shown in the following Table 1.

<table>
<thead>
<tr>
<th>Sensor number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance ( \delta_i )</td>
<td>0.07</td>
<td>0.1</td>
<td>0.05</td>
<td>0.2</td>
<td>0.3</td>
<td>0.25</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Measurements ( x_i )</td>
<td>0.99</td>
<td>0.96</td>
<td>1.05</td>
<td>0.90</td>
<td>0.55</td>
<td>0.635</td>
<td>1.02</td>
<td>1.01</td>
<td>1.03</td>
<td>1.48</td>
</tr>
</tbody>
</table>

First of all, divide the initial data into several clusters by formula (12), and 3 clusters are achieved after the first iteration, denoted as \{\( x_2, x_8 \)\}, \{\( x_1, x_3, x_4, x_7, x_8, x_9 \)\}, \{\( x_{10} \)\}, respectively. Then calculate \( Z_j(2)=0.5925, \quad Z_2(2)=0.9943, \quad Z_3(2)=1.48 \), where \( Z_j(2) \neq Z_j(1), \quad j=1,2,3 \), so return to step 2) and repeat. After the second iteration we can get \( Z_j(3)=Z_j(2), \quad j=1,2,3 \), and the algorithm converges at this point. So the final clusters are \{\( x_2, x_8 \)\}, \{\( x_1, x_3, x_4, x_7, x_8, x_9 \)\}, \{\( x_{10} \)\} and \( Z_3(3)=0.5925, \quad Z_2(3)=0.9943, \quad Z_3(3)=1.48 \) this time, according to formula (14), \( d_1=0.4018, \quad d_1=0.4857 \).

Due to the magnitude of measurements is very small and accuracy required precisely, so the error-tolerance \( \delta \) is set to \( \pm 0.1 \) which is \( \pm 10 \% \) of true value, so it can be seen that \( |d_1| \) and \( |d_2| \) are all larger than \( |\delta| \), so the data belong to \( Z_1(3) \) and \( Z_3(3) \) can be considered to be the failure data and need to be eliminated, only take \( Z_2(3) \) as the final fusion dataset, where \( Z_2(3) \) is \{\( x_1, x_2, x_3, x_4, x_7, x_8, x_9 \)\}.

To fusion the selected dataset with formula (10) and formula (11) in section 2, we can get the fusion value of 1.0046.

The following Fig. 2 shows the simulation diagram of weighting factors which obtained respectively by the improved fusion algorithm and unimproved fusion algorithm.

From the Fig. 2 we can see that the weighted factors of two methods are definitely different, for example, in the original algorithm, the weight of sensor 5 is relatively big, as we know that its measurement deviation is large which would impact the precise of fusion, but in the proposed algorithm,
its weight turns to 0, this is due to the measurement of sensor 5 is eliminated as invalid data which being ensured by k-means. Then take sensor 2 as the other example, its weighting factor is relatively small because of bad impact of the other error data in unimproved algorithm, but in the improved method the weight becomes larger, the reason is that the data measured by sensor 2 would never be mislead by the particularly large deviation of those data and considered to be high credibility by the proposed method, just as the fact. The other weighted values are all reassigned like that for improving the precise of fusion.

The following Fig. 3 shows the simulation diagram of fusion value of four algorithms which include traditional weighted fuzzy logic algorithm, adaptive weighted fusion algorithm, Least Square method and the improved algorithm.

![Fig. 3. The fusion data of four algorithms.](image)

From Fig. 3 we can visually see that the fusion data obtained by the traditional weighted fuzzy logic algorithm is farthest from the true value 1, and the results of both adaptive weighted fusion algorithm and Least Square method are almost same, and being better than the weighted fuzzy logic, but there’s still little error, and that the fusion data of improved algorithm deviating from the true value is the minimum, that is the result is the most accurate, which indicates the validity of the improved method.

The Table 2 then specifically lists the fusion values and the error with different algorithms.

<table>
<thead>
<tr>
<th>Fusion method</th>
<th>Result</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive weighted fusion algorithm in ref. 7</td>
<td>1.0113</td>
<td>0.0113</td>
</tr>
<tr>
<td>Least Square method in ref. 8</td>
<td>0.9900</td>
<td>0.01</td>
</tr>
<tr>
<td>Traditional weighted fuzzy logic algorithm</td>
<td>0.8879</td>
<td>0.1121</td>
</tr>
<tr>
<td>Improved algorithm based on k-mean clustering</td>
<td>1.0046</td>
<td>0.0046</td>
</tr>
</tbody>
</table>

In the Table 2, we have got that the fusion results of adaptive weighted fusion algorithm and Least Square algorithm are really almost same and their error are both around 0.01, and the error is greatest by using weighted fuzzy logic alone, and that the performance of improved algorithm is best for its least error 0.0046 which further explains the significant improvement of proposed method.

### 4. Conclusion

In this paper, we propose an improved weighted method based on the k-means. The proposed method consists of pretreatment and fusion stages, it can be simply implemented in WSN data fusion. Moreover the proposed method is robust to noises. Experiments demonstrate the proposed clustering method outperform the traditional fuzzy logic and weighted methods. However the proposed method has not been tested in realistic environments. And our future work will focus on it.

### Acknowledgements

The authors are thankful for the financial support from the research Grant “Natural Science Foundation of Shanxi Province”, Grant no. 2012011013-4, offered by the Taiyuan University of Technology.

### References


[12]. Hui-Lei Yang, Research of data fusion technology based on LEACH in wireless sensor network, Nanjing University of Science and Technology, 2012.


_______________________________________________________________

2013 Copyright ©, International Frequency Sensor Association (IFSA). All rights reserved.
(http://www.sensorsportal.com)