Consensus Multisensor Data Fusion Algorithm Based on Dynamic Hierarchical Clustering Analysis

Liang-wang DIAO* and Xiao-xuan WANG

S&T on Information System Engineering Lab., The 28th Research institute of CETC,
Nanjing 210007, China

*Corresponding author

Keywords: Consensus multisensor data fusion, Hierarchical cluster, Connection matrix, Precision, Degree of confidence.

Abstract. To improve the precision of consensus multi-sensor data fusion method, the reliability of sensor data must be considered. The shortcomings of the existing consensus multi-sensor data fusion algorithms, such as asymmetry of distance matrix, and the subjectivity of the threshold determination of connection matrix were discussed, and a new confidence distance was defined. A new consensus multi-sensor data fusion algorithm based on dynamic hierarchical clustering was proposed. The results of simulation experiments show that this method is better than the existing methods.

Introduction

The idea of the multisensor consensus data fusion algorithm proposed in [1] is to measure the authenticity and credibility of the sensor data through the mutual support degree between the multisensor data. Then some improved algorithms had been proposed. The correlation function of fuzzy theory to calculate and sort the support degree of each sensor was proposed in [2-3]. Data fusion algorithm based on widened median and approach coefficients was discussed in [4-5]. Using the statistical hypothesis testing theory of multivariate normal distribution, a new confidence distance to measure the distance between different sensor data, thus improving the accuracy of data consistency was discussed in [6-7]. Although a lot of research works had been done on consensus data fusion, there were still some shortcomings that need to be improved[8]. Aiming at these problems, using the idea of hierarchical clustering analysis, the sensor nodes are dynamically divided into groups, not only can we find the consistency between the measured data of different sensors, but also find the consistency of the "composite sensor" data of different kinds of sensors, so that the sensor data can be more embodied, so it has high application value.

Analysis of Existing Consensus Data Fusion Methods

Suppose a sensor measures an object, its measurement model can be described by normal distribution \( X \sim \mathcal{N}(\mu, \sigma_i^2) \), that is:

\[
p_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2\sigma_i^2}(x_i - \mu)^2\right\}
\]  

(1)

In the formula, \( x_i \) is the measured value of the \( i \)th sensor, \( \mu \) is the true value of the measurement feature, \( \sigma_i \) is the standard deviation of the sensor's measurement error, \( i = 1, 2, \ldots, n \).

Basic Principles of Existing Conformance Fusion Methods

The basic principle of the existing consensus fusion method is shown in Figure 1. In Figure 1, the core of the consensus data fusion algorithm mainly had three steps: one is computing the distance between different sensor data, the second is measuring the consistency between the sensor data, and the third is finding the maximum sensor connection Group.
Analysis of Existing Consensus Data Fusion Methods

The existing consensus data fusion methods have two shortcomings: (1) the selection of the maximum sensor group only considers the largest number of sensors, and does not take into account the accuracy of sensors. (2) multisensor data are processed only once, without considering the effect of data accuracy caused by the fusion processing results of each sensor group.

Fusion Algorithm based on Dynamic Hierarchical Clustering

Based on the above analysis, a new metric distance is first defined to represent the preference of high-precision sensors, and fusion algorithm based on dynamic hierarchical clustering was proposed.

Statistical Confidence Distance

Suppose \( x_i \sim N(\mu, \sigma_i^2) \), \( x_j \sim N(\mu, \sigma_j^2) \), a new confidence distance is defined as follows:

\[
d_y = P_x \left( \left| Z \right| \leq \frac{|x_i - x_j|}{\sqrt{2} \min(\sigma_i, \sigma_j)} \right)
\]

Formula (2) is a probabilistic measure of the distance between two sensors measured data. According to formula (2), probability can be defined to determine the consistency between the measured data of sensors. The method of calculation is:

\[
r_y = \begin{cases} 
1, & d_y \leq 1 - \alpha \\
0, & d_y > 1 - \alpha 
\end{cases} \quad i, j = 1, 2, \ldots, n
\]

It indicates that, if \( r_{ij} = 1 \) the two sensor data are consistent and can be fused. Otherwise, \( r_{ij} = 0 \) the data of the two sensor is not consistent, and it is not suitable for the fusion processing.
The Basic Steps of the New Algorithm

The basic steps of the proposed algorithm are as follows:

Assuming that all sensors' measured data, \( x_1, x_2, \ldots, x_n \), the corresponding measurement accuracy of each sensor is, \( \sigma_i, i = 1, 2, \ldots, n \).

Step 1: each data is regarded as a class object, each class has only one object, and the distance \( d_{ij} \) between them is calculated by formula(2). The distance between class and class is the distance between the objects which they contain, thus the distance matrix \( D = (d_{ij})_{n \times n} \) is obtained.

Step 2: determine the two sensors that are satisfied \( \arg \min \{d_{ij}\} \).

1) if \( \arg \min \{d_{ij}\} \leq 1 - \alpha \), the data of sensors \( i \) and sensors \( j \) are processed according to the following formula:

\[
x_{(i,j)} = \frac{\sigma_j^{-2}}{\sigma_i^{-2} + \sigma_j^{-2}} x_i + \frac{\sigma_i^{-2}}{\sigma_i^{-2} + \sigma_j^{-2}} x_j
\]

(3)

\[
\sigma_{\text{com}}^{-2} = \frac{1}{\sigma_i^{-2} + \sigma_j^{-2}}
\]

(4)

The measurement data of the two categories of sensors and sensors are merged into a new class called "composite sensors" \((i, j)\). And return to step 1;

2) if \( \arg \min \{d_{ij}\} > 1 - \alpha \), then go to step 3;

Step 3: there are two situations at this time.

1) case 1: all classes are finally merged into one class, and the final data obtained is the result of multi-sensor consensus fusion.

2) case two: all classes are merged into \( m \) class \((n \geq m \geq 2)\). Then, the data with the minimum of the "compound sensor" precision index is used as the result of the multisensor conformance fusion.

The greatest advantage of the proposed algorithm is that the sensors of different precision grades are not put together for consensus judgment, but based on the idea of hierarchical clustering analysis. Two sensor data with consistency were fused first, and then the fusion results were used to be the data of "compound sensor". That is, not only to measure the consistency between the data of each sensor, but also to measure the data of "compound sensor". According to the consistency, this obviously makes full use of sensor data. On the other hand, after the hierarchical clustering algorithm is merged into a new class object, it is necessary to recalculate the distance between the merged classes, that is, the distance matrix, which undoubtedly increases the amount of computation. But fortunately, the number of sensors in multi-sensor fusion is not large, so this is not a big impact.

Calculating Example

The advantages of proposed algorithm are illustrated by a calculating example.

A detection system consisting of 10 sensors is used to measure a target characteristic parameter (its true value is 10), and the measured data are shown in Table 1.

Table 1. The precision and data of ten sensors.

| numb. | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|------|----|----|----|----|----|----|----|----|----|----|
| x_i  | 9.9545 | 9.8744 | 10.0827 | 10.3464 | 10.4006 | 9.9646 | 9.0638 | 10.1838 | 9.0565 | 9.0497 |
| \sigma_i^2 | 0.05 | 0.07 | 0.10 | 0.20 | 0.30 | 0.25 | 0.10 | 0.10 | 0.20 | 0.30 |

By using the confidence distance defined by (5), the matrix \( D = (d_{ij})_{10 \times 10} \) can be obtained.
It can be seen that the confidence distance between sensors 9 and 10 is the smallest and less than 0.95, and the measured data are fused according to formula (3) and (4): $x_{(9,10)} = 9.05378$, $\sigma_{x_{(9,10)}} = 0.34641$

For the limited space, the concrete calculation process is no longer shown. The calculation results are as follows:

(1) the confidence distance matrix can be calculated (at this time with 9 sensors), and the sensor 7 is minimum and less than 0.95 with the composite sensor (9, 10), so the fusion results can be obtained as follows: $x_{(7,9,10)} = 10.1338$, $\sigma_{x_{(7,9,10)}} = 0.2335$

(2) there are 8 sensors at this time, the confidence distance matrix is calculated, and the confidence distance between 1 and 6 of the sensors can be determined to be minimum and less than 0.95, so the fusion results can be obtained: $x_{(1,6)} = 9.9562$, $\sigma_{x_{(1,6)}} = 0.20412$

(3) there are 7 sensors at this time. By calculating the confidence distance, the distance between 4 and 5 of the sensors is determined to be minimum and less than 0.95. Therefore, the fusion results can be obtained as follows: $x_{(4,5)} = 10.36808$, $\sigma_{x_{(4,5)}} = 0.34641$

(4) there are 6 sensors at this time. By calculating the confidence distance, the distance between 3 and 8 of the sensors is determined to be minimum and less than 0.95, so the fusion results can be obtained: $x_{(3,8)} = 10.10825$, $\sigma_{x_{(3,8)}} = 0.2236$

(5) there are 5 sensors at this time. By calculating the confidence distance, the distance between 2 and the composite sensor (1, 6) is determined to be minimum and less than 0.95, so the fusion results can be obtained: $x_{(1,2,6)} = 9.92567$, $\sigma_{x_{(1,2,6)}} = 0.16161$

(6) there are 4 sensors at this time. By calculating the confidence distance, the distance between 3 and the composite sensor (1, 2, 6) and the composite sensor (3, 8) is minimized and less than 0.95. Therefore, the fusion results can be obtained as follows: $x_{(1,2,6;3,8)} = 9.98832$, $\sigma_{x_{(1,2,6;3,8)}} = 0.13098$

(7) there are 3 sensors in this process. By calculating the confidence distance, the distance between them can be judged to be greater than 0.95.

(8) by comparison, the final fusion result can be obtained as follows: $x_{(1,2,6;3,8)} = 9.98832$, $\sigma_{x_{(1,2,6;3,8)}} = 0.13098$

The above calculation process is shown in Figure 2.

Figure 2. The hierarchical clustering process of 10 sensors’ data.
If we use the existing data fusion method based on statistical distance and consistency[8,12,14] can be judged by threshold 0.95.

There are 6 sensors supporting the sensor 1, and 6 sensors supporting the sensor 2.
There are 6 sensors supporting the sensor 3, and 6 sensors supporting the sensor 4.
There are 6 sensors supporting the sensor 5, and 8 sensors supporting the sensor 6.
There are 2 sensors supporting the sensor 7, and 6 sensors supporting the sensor 8.
There are 2 sensors supporting the sensor 9, and 4 sensors supporting the sensor 10.
If 6 sensors are supported as the criteria, the initial sequence numbers of the original sensors are 1, 2, 3, 4, 5, 6, 8, and the results are as follows: \( x_{1,2,3,4,5,6,8} = 10.0358 \), \( \sigma_{1,2,3,4,5,6,8} = 0.1225 \).

Comparing the fusion results of the two algorithms, it is obvious that the estimation of the true value is not as good as the algorithm proposed in this paper.

**Summary**

The data fusion method proposed in this paper combines the data consistency of the sensor with the measurement accuracy of the sensor, thus avoiding the fusion process which is involved in the large error data, and improves the accuracy of the fusion results without affecting the fusion precision. Therefore, the data fusion results are more effective.

**Acknowledgement**

This research was financially supported by CETC Science and Technology Innovation Foundation.

**References**

[1] Lou R C, Lin M, Scherp P S. Dynamic multi-sensor data fusion system for intelligent robots[J]. IEEE, Journal of Robotics and Automation, 1988, 4(4): 386-396.

[2] Wang Tingjie Shi Huichang. Consensus Data Fusion Method Based on Fuzzy Theory [J], Journal of Transducer Technology , 1999, 18(6): 50-53) (in Chinese)

[3] Yang Hui, Yu Xun. Data fusion algorithm based on consensus processing and fuzzy theory[J]. Science & Technology Information, 2011(34): 39-41.(in Chinese)

[4] Ji Linna, Yang Fengbao. Data fusion algorithm based on widened median and approach coefficients[J]. Fire Control & Command Control, 2013, 38(1): 114-117.(in Chinese)

[5] Xue Fei; Yang Youliang; Dong Futao. CCD temperature measurement system design based on multi-sensor fusion algorithm[J], Instrument Technique and Sensor, 2014(8): : 88-91.(in Chinese)

[6] Diao Lianwang, Wang Changwu. Improved and generalized consensus data fusion method[J], Systems Engineering and Electronics, 2002, 24(9): 60-63. (in Chinese)

[7] Zeng Li; Jiang Yuan. Mathematical methods of multi-sensor data fusion [J], Journal of Yunnan University of Nationalities (Natural Sciences Edition), 2010, 19(5): 321-325.. (in Chinese)

[8] Xong Zhaohua, Diao Liangwang. Distributed consensus information fusion in multisensor detection cloud and its development[J]. Command Information System and Technology, 2018, 9(2): 8-18.