Computers, Materials & Continua CMC, vol.61, no.3, pp.1217-1231, 2019

Research on Data Fusion of Adaptive Weighted Multi-Source Sensor

Donghui Li[1], Cong Shen[2, *], Xiaopeng Dai[1], Xinghui Zhu[1], Jian Luo[1], Xueting Li[1], Haiwen Chen[3] and Zhiyao Liang[4]

Abstract: Data fusion can effectively process multi-sensor information to obtain more accurate and reliable results than a single sensor. The data of water quality in the environment comes from different sensors, thus the data must be fused. In our research, self-adaptive weighted data fusion method is used to respectively integrate the data from the PH value, temperature, oxygen dissolved and NH_3 concentration of water quality environment. Based on the fusion, the Grubbs method is used to detect the abnormal data so as to provide data support for estimation, prediction and early warning of the water quality.

Keywords: Adaptive weighting, multi-source sensor, data fusion, loss of data processing, grubbs elimination.

1 Introduction

Multi-sensor information fusion is an emerging interdisciplinary discipline that has been formed since the 1970s. It has been widely used in high-tech fields such as military, national defense, and aerospace [Wu (2016); Guo (2017)], and has become a hot area attracting much attention. The primary goal of a multi-sensor system is to unite, correlate, and assemble the data from sensors with related database information to obtain more comprehensive, higher, and more reliable data information. However, many problems should be encountered in areas of data correlation, sensor uncertainty, and data management. While the fundamental problem is the inherent uncertainty in the sensor measurement system. This uncertainty arises not only from inaccuracies and noise from measurements, but also from inconsistencies of different sensors. In the past, strategies were used to model this uncertainty and fuse different types of data to get consistent decisions. In the early 1980s, Bar-shalom [Bar-shalom (1981)] studied the correlation between two sensor subsystems and gave the formula for calculating the mutual covariance matrix. In the 1990s, Carlson proposed the well-known federal Kalman filter algorithm [Carlson (1990)], which uses the upper bound of variance to eliminate correlation and uniform information distribution principles. The algorithm avoids the calculation of cross-covariance matrix, but it is conservative in some respects.

---

[1] College of Information Science and Technology, Hunan Agricultural University, Changsha, 410128, China.

[2] College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, China.

[3] College of Computer Science and Engineering, National University of Defense Technology, Changsha, 410073, China.

[4] Faculty of Information Technology, Macau University of Science and Technology, Macau.

[*] Corresponding Author: Cong Shen. Email: m18374889745_1@126.com.

In 1994, Kim [Kim (2002)] proposed a maximum likelihood fusion estimation algorithm that requires random variables to be normally distributed so as to construct a likelihood function. Later, considering the significance of linear minimum variance, Sun et al. [Sun and Deng (2004)] proposed three fusion algorithms, that are matrix weighting, diagonal weighting and scalar weighting. In recent years, Olfati-Saber [Olfati-Saber (2008)] proposed a distributed algorithm based on consistency strategy-Kalman Consensus filter. And it has attracted extensive attention due to the simple distributed architecture as well as the scalability and robustness of the algorithm. Zhangpin and other people, after analyzing how to deal with the uncertainty and inconsistency of sensor nodes' data in sensor network data fusion, have proposed an optimized Bayesian estimation of multi- sensor data fusion method-a Bayesian fusion algorithm based on Kalman filter. According to the ways that filter is applied to sensor data, fusion data, or to both, three different techniques were proposed: forward filtering, backward filtering, and forward-backward filtering.

In the late 1970s, concepts and nouns concerning the information synthetic began to appear in some publicly-published documents, and in the later long period, the term "data fusion" was commonly used [Zuo (2005)]. Until the 1990s, the term "information fusion", after consideration of the diversity of sensor information, was then widely adopted [Ling, Chen, Gu et al. (2000)]. The concept of information fusion has been described in many ways. The literature Bar-shalom [Bar-shalom (1981)] gives many concepts of fusion based on specific behaviors and related fields of application. The literature Liu et al. [Liu, Zhao, Liu et al. (2005)] proposes a general description of information fusion. Based on which above, the literature Mao et al. [Mao, Xiao and Yang (2015)] gives the following description: The form of information fusion is a framework, which organizes, correlates, and synthesizes multi-source information through specific mathematical methods and technical tools in order to obtain high-quality and useful information. And a precise definition of "high quality" depends on the application object. Nagesware [Rao (2000)] of Tunisia proposed a method of physical system fusion based on physics rules, and achieved satisfactory results in the detection of methane hydroxide; Serge [Moigne (2000)] from France proposed the way of fusing wind speed and direction, which is better for solving the wind field problem; Reboul et al. [Reboul and Bruge (2000)] from the United States developed a fusion system (Landsat), which can be used to monitor the changes in the vegetation and so on.

The research of information fusion theory involves many basic theories. From the perspective of the algorithm, it can be roughly divided into two categories: probability statistics method and artificial intelligence method. Among them, probability statistics method is represented by Bayes [Guan (2013)] and its deformation methods. In artificial intelligence methods, Bayes estimation [Guan (2013)], evidence reasoning [Zhang (2016)] fuzzy theory [Guo (2013)] as well as neural network [Jia (2017)] in total account for 85 percent of the entire information fusion algorithm. And machine learning methods such as vector-supporting machines [Lu (2016); Ding, Zhang, Zhang et al. (2018); Xie (2014)], genetic algorithms [Li (2017)], rough sets [Guo (2017)] and some other artificial intelligence methods[Liu, Dong, Liu et al. (2018); Zhang, Cai, Liu et al. (2018); Sun, Cai, Li et al. (2018); Cui, Zhang, Cai et al. (2018)] have also been applied in information fusion. In our research, we also draw some methods [Cai, Wang, Zheng et al. (2013); Liu, Cai, Xu et al. (2015); Huang, Liu, Zhang et al. (2018)] in other articles. In this paper, on the basis of the research method of Li et al. [Li, Zhang, Xu et al. (2018); Liu, Tang, Li et al. (2019)], adaptive weighted data fusion method is adopted to eliminate the fused data by using Grubbs, so as to obtain data that can be actually used.

2 Adaptive weighted data fusion

2.1 General weighted data fusion

At present, the data fusion theory has been widely used in the fields of state estimation. Among the related algorithms, the weighted fusion algorithm is mature, and many research results have proved that the algorithm is optimal, unbiased, and has the smallest mean square error. The key to apply the weighted fusion algorithm lies in the determination of the weight coefficient, which in turn is inversely proportional to the variance measurement of sensors. Assume that there are N sensors of different accuracy, of which the variances are σ_1^2, σ_2^2, … , σ_n^2, and the measured values of the sensors are X_1, X_2, …, X_n, thus the result of the weighted fusion is:

$$X = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\cdots+\frac{1}{\sigma_n^2}}X_1 + \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\cdots+\frac{1}{\sigma_n^2}}X_2 + \cdots + \frac{\frac{1}{\sigma_n^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\cdots+\frac{1}{\sigma_n^2}}X_n \tag{1}$$

From the above equation, it can be seen that the measurement data with large variance gives a smaller weight, while the data with small variance gives a larger weight, so this data fusion method can obtain a more reliable measurement result than the method of the arithmetic average.

2.2 Adaptive weighted data fusion ideas

Assume that the mean square errors of the sensed data of the N sensors are σ_1^2, σ_2^2, ... , σ_n^2, respectively, the measured values of the sensor nodes are X_1, X_2, ... , X_n, correspondently, and the weight factors of each sensor is W_1, W_2, ... , W_3, respectively. Since the data are independent of each other and belongs to unbiased estimation of X, the truth factor and weight factor of X after fusion would respectively satisfy the following relationships:

$$\bar{X} = \sum_{i=1}^{n} W_i X_i\,,\ \sum_{i=1}^{n} W_i = 1 \tag{2}$$

The total mean square error is:

$$\sigma^2 = E[(X-\bar{X})^2] = E[\sum_{i=1}^{n} W_i^2 (X-X_i)^2 + 2\sum_{p=1,q=1,p\neq q}^{n} W_p W_q (X-X_p)(X-X_q)] \tag{3}$$

X_1, X_2, ... , X_n are independent from each other and are unbiased estimates of X. Thus conclusion can be drawn:

$$E[(X-X_p)(X-X_q)] = 0 \tag{4}$$

$p \neq q$, p=1, 2, ... , n; q=1, 2, ... , n. And σ^2 can be written as:

$$\sigma^2 = E[\sum_{i=1}^{n} W_i^2 (X-X_i)^2] = \sum_{i=1}^{n} W_i^2 \sigma_i^2 \tag{5}$$

According to the above formulas, the mean square error σ^2 is a multivariate quadratic function, so σ^2 must have a minimum value. According to the extreme value theory of the multivariate function, the minimum weight factor is:

$$W_i^* = \frac{1}{\sigma_i^2 \sum_{p=1}^{n} \frac{1}{\sigma_p^2}} \tag{6}$$

So, the correspondent minimum mean square error is:

$$\sigma_{\min}^2 = \frac{1}{\sum_{p=1}^{n} \frac{1}{\sigma_p^2}} \tag{7}$$

After adaptive weighted data processing, each sensor data from test points are fused into one data and sent to the monitoring and management center for applying water quality prediction and warning service.

The weighted data fusion algorithm does not require prior knowledge of the data from sensor measurement, and the data fusion results can be obtained merely by the data values through local decision after the fusion of a single sensor.

2.3 Adaptive weighted data fusion algorithm

Assume that two different sensors are used to measure a constant quantity. The observed values are: $z_1 = x + v_1$, $z_2 = x + v_2$. Among them, $v_i(i=1,2)$ is the random error during the observation. Assuming $v_i \sim N(0,\sigma_i^2)$, the two sensor observations are assumed to be independent of each other. The assumed estimate of x and the observed value of $z_i(i=1,2)$ are linearly related, and $\hat{x}$ is the unbiased estimation value of x, thus:

$$\hat{x} = \omega_1 z_1 + \omega_2 z_2 \tag{8}$$

$\Omega = (\omega_1, \omega_2)$ is the weight of the measured value for each sensor. Assume that the estimation error is $\tilde{x} = x - \hat{x}$. With the mean squared error of which the cost function is $\tilde{x}$, it can be inferred that:

$$J = E(\tilde{x}^2) = E\{[x - \omega_1(x+z_1) - \omega_2(x+z_2)]^2\} \tag{9}$$

Because $\hat{x}$ is the unbiased estimate of x, it can be inferred that:

$$E(\tilde{x}) = E[x - \omega_1(x+z_1) - \omega_2(x+z_2)] = 0 \tag{10}$$

Since $E(v_1) = E(v_2) = 0$ and $E(x) = E(\hat{x})$, it can be inferred that: $\omega_2 = 1 - \omega_1$

Then the cost function can be written as:

$$J = E(\tilde{x}^2) = E[\omega_1^2 v_1^2 + (1-\omega_1)^2 v_2^2 + 2\omega_1(1-\omega_1)v_1 v_2] \tag{11}$$

With $E(v_1^2) = \sigma_1^2$ and $E(v_2^2) = \sigma_2^2$, v_1, v_2 are independent of each other, then:

$$E(v_1 v_2) = 0 \tag{12}$$

thus:

$$J = E(\tilde{x}^2) = \omega_1^2\sigma_1^2 + (1-\omega_1)^2\sigma_2^2 \tag{13}$$

In order to ensure the smallest J, Ω is derivatived and thus:

$$\frac{\partial J}{\partial \Omega} = 0 \tag{14}$$

The optimal weight is:

$$\omega_1^* = \frac{\sigma_2^2}{\sigma_2^2 + \sigma_1^2} \qquad \omega_2^* = \frac{\sigma_1^2}{\sigma_2^2 + \sigma_1^2} \tag{15}$$

The optimal estimate is:

$$\hat{x} = \frac{\sigma_2^2 z_1}{\sigma_2^2 + \sigma_1^2} + \frac{\sigma_1^2 z_2}{\sigma_2^2 + \sigma_1^2} \tag{16}$$

The above formula shows that when the values of the two sensors are given suitable, the optimal estimation can be obtained by integrating the observed values through the observer. To extend this conclusion to multiple sensors, the variances of measured values from the multiple sensor groups should be setted as $\sigma_i (i = 1, 2...n)$ respectively, and they are independent from each other. The estimated value of the true value is $\hat{x}$, and is the unbiased estimate of x . The weighting factors of the sensors are $\omega_i (i = 1, 2...n)$ respectively. According to the extremum value-seeking theory of multivariate function, the weighting factor corresponding to the minimum mean-square error can be obtained as:

$$\omega_p^* = \frac{1}{|\sigma_p^2 \sum_{i=1}^{n} \frac{1}{\sigma_i^2}|} \tag{17}$$

3 Examples of adaptive weighted data fusion algorithm

This study mainly focused on the detection of PH value, dissolved oxygen degree, temperature, and ammonia nitrogen concentrations in aquaculture environment. Five sets of sensors were placed at the No. 2 pond of Quyuan aquaculture base in Yueyang City, Hunan Province. Taking into account the instability of the quality of the sensor equipment itself, each set of sensors was divided into experimental Group 1 and experimental Group 2. Both experimental Group 1 and experimental Group 2 contain 4 sensors, a PH value sensor, a dissolved oxygen sensor, a temperature sensor, and an ammonia nitrogen concentration sensor.

In order to verify the fusion effect of the adaptive weighted data fusion, experiments were conducted in fish ponds. Fish pond data of October 2017 were recorded every 10 minutes. 5 points of time, that are 8:05, 8:15, 8:25, 8:35, and 8:45 on October 1, 2017, were chosen to fuse. The extracted data are shown in Tabs. 1-5. Sensor Groups 1-4 denote sensors in the four corners of the fish pond. Sensor group 0 denotes the sensor group in the middle of the fish pond. Sensor 0-4 (A) denotes the Group 1 sensor in the experiment. Sensor 0-4 (B) indicates the Group 2 experimental sensors.

Table 1: Data in 8:05 at 1th October 2017

Sensor Group		Temperature (℃)	Dissolved oxygen (mg/L)	PH	NH_3 (mg/L)
Sensor Group 1	Sensor 1A	18.8	8.73	8.45	0.116
	Sensor 1B	18.1	9.2	8.58	0.119
Sensor Group 2	Sensor 2A	18.3	9.18	8.54	0.111
	Sensor 2B	18.3	9.16	8.49	0.104
Sensor Group 3	Sensor 3A	18.9	8.94	8.46	0.089
	Sensor 3B	19.1	8.85	8.3	0.065
Sensor Group 4	Sensor 4A	19.2	8.86	8.27	0.073
	Sensor 4B	18.6	8.44	8.29	0.072
Sensor Group 0	Sensor 0A	18.7	9.2	8.3	0.080
	Sensor 0B	18.8	9.08	8.37	0.080

Table 2: Data in 8:15 at 1th October 2017

Sensor Group		Temperature (℃)	Dissolved oxygen (mg/L)	PH	NH_3(mg/L)
Sensor Group 1	Sensor 1A	18.7	8.83	8.36	0.111
	Sensor 1B	18.9	9.1	8.52	0.109
Sensor Group 2	Sensor 2A	18.8	9.11	8.63	0.100
	Sensor 2B	18.8	9.17	8.26	0.096
Sensor Group 3	Sensor 3A	18.7	8.9	8.6	0.080
	Sensor 3B	18.6	8.88	8.29	0.078
Sensor Group 4	Sensor 4A	19.1	8.81	8.36	0.045
	Sensor 4B	19.1	8.46	8.15	0.064
Sensor Group 0	Sensor 0A	18.7	9.22	8.44	0.109
	Sensor 0B	18.9	9.03	8.62	0.084

Table 3: Data in 8:25 at 1th October 2017

Sensor Group		Temperature (℃)	Dissolved oxygen (mg/L)	PH	NH_3(mg/L)
Sensor Group 1	Sensor 1A	18.9	8.84	8.58	0.106
	Sensor 1B	19.1	9.21	8.62	0.087
Sensor Group 2	Sensor 2A	18.5	9.09	8.55	0.083
	Sensor 2B	18.6	8.97	8.72	0.088
Sensor Group 3	Sensor 3A	18.4	8.89	8.62	0.072
	Sensor 3B	19	8.62	8.33	0.091
Sensor Group 4	Sensor 4A	19.1	8.85	8.51	0.072
	Sensor 4B	18.8	8.59	8.78	0.069
Sensor Group 0	Sensor 0A	18.4	9.03	8.36	0.077
	Sensor 0B	18.6	9.11	8.29	0.085

Table 4: Data in 8:35 at 1th October 2017

Sensor Group		Temperature (℃)	Dissolved oxygen (mg/L)	PH	NH_3(mg/L)
Sensor	Sensor 1A	19.2	8.66	8.62	0.105
Group 1	Sensor 1B	18.3	9.15	8.55	0.080
Sensor	Sensor 2A	18.3	9.19	8.64	0.081
Group 2	Sensor 2B	19	9.1	8.32	0.074
Sensor	Sensor 3A	18.4	8.9	8.53	0.069
Group 3	Sensor 3B	18.9	8.69	8.29	0.068
Sensor	Sensor 4A	19.1	8.88	8.51	0.081
Group 4	Sensor 4B	18.6	8.52	8.29	0.079
Sensor	Sensor 0A	18.9	9.18	8.61	0.085
Group 0	Sensor 0B	18.8	9.06	8.28	0.088

Table 5: Data in 8:45 at 1th October 2017

Sensor Group		Temperature (℃)	Dissolved oxygen (mg/L)	PH	NH_3(mg/L)
Sensor	Sensor 1A	18.8	9.81	8.07	0.052
Group 1	Sensor 1B	18.6	10.16	8.48	0.081
Sensor	Sensor 2A	19.2	9.95	8.59	0.110
Group 2	Sensor 2B	19.3	9.69	8.57	0.128
Sensor	Sensor 3A	18.9	9.56	8.49	0.103
Group 3	Sensor 3B	18.8	9.18	8.74	0.165
Sensor	Sensor 4A	19.5	9.47	8.28	0.070
Group 4	Sensor 4B	19.8	9.02	8.3	0.075
Sensor	Sensor 0A	19.4	9.63	8.38	0.101
Group 0	Sensor 0B	19.5	10.16	8.42	0.088

3.1 Weighted data fusion of temperature data

Taking the temperature data out from Tab. 1 separately for weighted fusion analysis, the data are shown in the following Tab. 6.

Table 6: Data in 8:05 at 1th October 2017

Sensor 0-4(A)					Sensor 0-4(B)				
S_1A	S_2A	S_3A	S_4A	S_0A	S_1B	S_2B	S_3B	S_4B	S_0B
18.8	18.3	18.9	19.2	18.7	18.1	18.3	19.1	18.6	18.8

The arithmetic average value of the measurement data from Group A is:

$$T_{(1)} = \frac{1}{5}\sum_{i=1}^{5} T_{1i} = 18.78$$

The corresponding standard error is:

$$\sigma_{(1)} = \sqrt{\frac{1}{5-1}\sum_{i=1}^{5}(T_{1i} - T_{(1)})^2} = 0.327$$

The arithmetic average value of the measurement data from Group B is

$T_{(2)}=\frac{1}{5}\sum_{i=1}^{5}T_{2i}=18.58$

The corresponding standard error is:

$\sigma_{(2)}=\sqrt{\frac{1}{5-1}\sum_{i=1}^{5}(T_{2i}-T_{(2)})^2}=0.396$

Since $T_{(1)}$ and $T_{(2)}$ are the measurement data from the same batch, and there has been no statistical measurement of temperature, that is, the variance of the previous measurement $\sigma^1=\infty$,thus it can be concluded that $(\sigma^1)^{-1}=0$.

According to the batch estimation theory, the variance of the temperature fusion value can be obtained as:

$$\sigma^+=[(\sigma^1)^{-1}+H^TD^{-1}H]^{-1}=\left\{[1,1]\begin{pmatrix}\frac{1}{\sigma_{(1)}^2} & 0\\ 0 & \frac{1}{\sigma_{(2)}^2}\end{pmatrix}\begin{bmatrix}1\\1\end{bmatrix}\right\}=\frac{\sigma_{(1)}^2\sigma_{(2)}^2}{\sigma_{(1)}^2+\sigma_{(2)}^2} \tag{18}$$

H is the coefficient matrix of the measurement equation and H=[1,1]; D is the covariance of the measurement noise and $D=\begin{bmatrix}\sigma_{(1)}^2 & 0\\ 0 & \sigma_{(2)}^2\end{bmatrix}$.

The fusion value of the measurement data obtained by combining the above formula is:

$$T^+=\frac{\sigma_{(2)}^2T_{(1)}+\sigma_{(1)}^2T_{(2)}}{\sigma_{(1)}^2+\sigma_{(2)}^2} \tag{19}$$

Substituting the data from Tab. 6, result can be obtained as: The fusion value after weighted information fusion at 8:05 is: $T_1^+=18.6989$, and variance $\sigma_1^+=0.0636$.

Similarly, the temperature data in Tabs. 2-5 are fused through the same weighted information fusion method, and the following temperature and temperature data fusion results are obtained:

Table 7: Fusion data of temperature data in different time points at 1th October 2017

Time	8:05	8:15	8:25	8:35	8:45
Fusion value of Temperature Data (°C)	18.6989	18.828	18.766	18.739	19.23
Variance of Temperature Data	0.0636	0.0157	0.0346	0.0527	0.0676

3.2 Weighted data fusion of dissolved oxygen data

The weighted data fusion analysis of the dissolved oxygen data in Tables from 1 to 5 is performed. The weighted information fusion method is the same as the temperature data fusion calculation method in 3.1. Thus, the following results can be obtained:

Table 8: Fusion data of dissolved oxygen in different time points at 1th October 2017

Time	8:05	8:15	8:25	8:35	8:45
Fusion value of Dissolved Oxygen (mg/L)	8.97	8.96	8.93	8.94	9.68
Variance of Temperature Data	0.029	0.023	0.013	0.031	0.033

3.3 Weighted data fusion of PH data

The weighted fusion analysis is performed on the PH data from Tabs. 1 to 5, and the weighted information fusion method is the same as the temperature data fusion calculation method in 3.1. The following results can be obtained:

Table 9: Fusion data of PH value in different time points at 1th October 2017

Time	8:05	8:15	8:25	8:35	8:45
Fusion Value of PH Data	8.405	8.444	8.528	8.534	8.445
Variance of PH Data	0.007	0.012	0.008	0.003	0.016

3.4 Weighted data fusion of NH_3 concentration data

The weighted fusion analysis of the NH_3 concentration data from Tabs. 1 to 5 is performed. The weighted information fusion method is the same as the temperature data fusion calculation method in 3.1. Thus, the following results can be obtained:

Table 10: Fusion data of NH_3 value in different time points at 1th October 2017

Time	8:05	8:15	8:25	8:35	8:45
Fusion Value of NH_3 Data(mg/L)	0.091	0.087	0.083	0.079	0.093
Variance of NH_3 Data	0.00021	0.00021	0.00005	0.00004	0.00044

4 Treatment of loss data in aquaculture environment

Loss error, also known as fault error, is an error that is clearly inconsistent with the facts. It is usually caused by issues of operator error, damage to the internal components of the system, loose wiring, detachment, and sudden external shocks. The presence of negligence error is a serious distortion of the measurement result and must be removed. This paper mainly uses Grubbs judgment criteria to process data.

4.1 The Grubbs guidelines

Assume that the measurement data sequence of a sensor from multiple independent detections toward a certain object can be defined as $X_1, X_2, \cdots, X_n$,and the measurement data X_i(i=1, 2, ... , n)obey the normal distribution. Assume that this measurement column is setted according to the ascending order, that is:

$$X_1 \le X_2 \le \cdots \le X_n \tag{20}$$

and the measured value obeys the normal distribution, then

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i\ ,\quad v_i = X_i - X\ ,\quad \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} v_i^2} \tag{21}$$

According to the principle of sequential statistics, the exact distribution of the Grubbs statistic is found:

$$g_i = \frac{X_i - X}{\sigma}\text{(i=1 or n)} \tag{22}$$

Therefore, after giving the significant level of a (usually taking a=0.05 or a=0.01), we can use the table lookup method (Grubbs Threshold Table https://wenku.baidu.com/view/0f 3c083a172ded630a1cb6c8.html) to find out the critical value of the Grubbs statistic

$g_0(n,a)$。 $p[g_i \geq g_0(n,a)] = a$ as a small probability event ,and it should not appear when $X_i(i=1,2,\cdots,n)$ obeys the state distribution.

Measure g_i , the corresponding Grubbs statistic of the top value $X_i(i=1 \text{or} n)$, if it satisfies $g_i \geq g_0(n,a)$,it is considered that there is a significant difference in the distribution of the statistic g_i, and the corresponding X_i should contain a delinquent error (negligence error, or gross error, which means an error that is obviously inconsistent with the facts). X_i is a suspicious value and should be eliminated. If $g_i < g_0(n,a)$, it is considered that the corresponding X_i has no error negligence value X_i, and it cannot be rejected as a suspicious value.

4.2 Treatment of errors in aquaculture environment

In the aquaculture environment, the data collected by sensors will inevitably result in sparse errors due to various reasons. In this paper, we use Grubbs criterion to deal with the negligence errors in the environment. This section performed the Grubbs processing of negligence data on the combined data in Section 3, and the following table shows the fused data of temperature, PH, dissolved oxygen, and NH_3 concentration.

Table 11: Fusion data in different time points at 1th October 2017

Time	8:05	8:15	8:25	8:35	8:45
Fusion Value of Temperature Data(°C)	18.6989	18.828	18.766	18.739	19.23
Fusion Value of dissolved oxygen (mg/L)	8.97	8.96	8.93	8.94	9.68
PH value	8.405	8.444	8.528	8.534	8.445
NH_3 Concentration (mg/L)	0.091	0.087	0.083	0.079	0.093

4.2.1 Processing of temperature data gross errors

According to the known conditions, the Grubbs elimination is practiced to the temperature data of the first group, and the average value of five numbers in the temperature data is:

$$T_1 = \frac{1}{5}\sum_{i=1}^{5} x_i = 18.85$$

The residual error $V_i=X_i-T_1$ is calculated due to the average value, and the residual error of the first data group is shown in Tab. 12:

Table 12: Residual errors of temperature data

Time	8:05	8:15	8:25	8:35	8:45
V_i	-0.153	-0.024	-0.086	-0.113	0.378

The approximate error can be calculated by using the residual error of the temperature data:

$$\sigma_1 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} V_i^2} = 0.216$$

Using the table look-up method (Grubbs Threshold Table https://wenku.baidu.com/view/0f3c083a172ded630a1cb6c8.html) to find out the critical value of Grubbs $g_0(n,a)$, and it is known that $P[g \geq g_0(n,a)]=a$ (significant level a generally takes 0.05 or 0.01) is a small Probability event that should not occur when X_i obeys a normal distribution. Check the value table and it is known that $g_0(5,0.05)=1.764$, $g_0(5,0.05) \times \sigma_1 = 0.829$. At this time $|V_i|_{max}=0.378<g_0(5, 0.05) \times 0.216=0.829$, so the temperature data need not be eliminated and the subsequent calculation can be performed directly.

4.2.2 Treatment of dissolved oxygen data gross errors

The Grubbs elimination on the dissolved oxygen data can be performed by using the same calculations as in Section *Processing of Temperature Data Gross Errors,* and in the dissolved oxygen data:

$$T_2 = \frac{1}{5}\sum_{i=1}^{5} x_i = 9.096$$

Table 13: Residual error of dissolved oxygen data

Time	8:05	8:15	8:25	8:35	8:45
V_i	-0.126	-0.136	-0.166	-0.156	0.584

The approximate error can be calculated by using the residual error of the dissolved oxygen data:

$$\sigma_2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} V_i^2} = 0.327$$

Using the table look-up method (Grubbs Threshold Table https://wenku.baidu.com/view/0f3c083a172ded630a1cb6c8.html) to find out the critical value of Grubbs $g_0(n, a)$, and it is known that $P[g \geq g_0(n, a)]=a$ (significant level a generally takes 0.05 or 0.01) is a small probability event that should not occur when X_i obeys a normal distribution. Check the value table and it is known that g_0 (5, 0.05)=1.764, $g_0(5,0.05) \times \sigma_2 = 0.576$. At this time $|V_i|_{max}=0.584>g0(5,0.05) \times 0.327=0.576$, thus for dissolved oxygen data at 8:45 ,it should be eliminated. The data after elimination is shown in the following tab:

Table 14: Dissolved oxygen data after Grubbs elimination

Time	8:05	8:15	8:25	8:35
Oxygen Dissolved(mg/L)	8.97	8.96	8.93	8.94

The eliminated data should be re-calculated using the Grubbs criteria, and $T_2^{'}=8.95$, $\sigma_2^{'}=0.018$, and the values of $g_0(4,0.05)=1.496$, $g_0(4,0.05) \times \sigma_2^{'}=0.027$ were obtained by checking numerical table. At this point $|V_i|_{max}=0.02<g_0(4,0.05) \times 0.018=0.027$, thus, it is not necessary to eliminate, and the next calculation can be performed directly.

4.2.3 Handling gross errors of PH data

To perform Grubbs elimination on the PH value, the same calculation method as in section *Processing of Temperature Data Gross Errors* can be used in the PH value data:

Table 15: Residual error of PH data

Time	8:05	8:15	8:25	8:35	8:45
V_i	-0.066	-0.027	0.056	-0.063	0.026

The approximate error can be calculated by using the residual error of the dissolved oxygen data:

$$\sigma_3 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} V_i^2} = 0.057$$

Using the table look-up method (Grubbs Threshold Table https://wenku.baidu.com/view/0f3c083a172ded630a1cb6c8.html) to find out the critical value of Grubbs $g_0(n, a)$, it is known that $P[g \geq g_0(n, a)]=a$ (significant level a generally takes 0.05 or 0.01) is a small Probability event that should not occur when X_i obeys a normal distribution. Check the value table and it is known that g_0 (5, 0.05) = 1.764, $g_0(5,0.05)\times\sigma_2=0.1$. At this time $|V_i|_{max}=0.066<g_0(5, 0.05)\times0.057=0.1$, so the PH data need not be deleted.

4.2.4 Handling gross errors of NH_3 concentration data

To practice Grubbs elimination on the data of NH_3 concentration, the same calculation method as in section *Processing of Temperature Data Gross Errors* can be performed. In the NH_3 concentration data:

Table 16: Residual error of NH_3 concentration data

Time	8:05	8:15	8:25	8:35	8:45
V_i	0.0044	0.0004	-0.0036	-0.0076	0.0064

The approximate error can be calculated by using the residual error of the dissolved oxygen data:

$$\sigma_4 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} V_i^2} = 0.0057$$

Using the table look-up method (Grubbs Threshold Table https://wenku.baidu.com/view/0f3c083a172ded630a1cb6c8.html) to find out the critical value of Globes $g_0(n, a)$, it is known that $P[g \geq g_0(n, a)]=a$ (significant level a generally takes 0.05 or 0.01) is a small Probability event that should not occur when X_i obeys a normal distribution. Check the value table and it is known that g_0 (5, 0.05)=1.764, $g_0(5,0.05)\times\sigma_4=0.01$. In this case, $|V_i|_{max}=0.0076<g_0$ (5, 0.05)×0.0057=0.01, thus, it is not necessary to eliminate the NH_3 concentration data.

5 Summary

This study has placed importance on the PH value, temperature, oxygen dissolved a value, and NH_3 concentration in the aquaculture environment. Self-adaptive weighted methods are used to perform respectively the first-level fusion of the factors affecting the aquaculture environment, obtaining fusion result of sensor data at a 10-minutes interval between 8:05 and 8:45. Limited by environment and network transmission, the method of

Grubbs is used to detect abnormal data after fusion. Abnormal data of dissolved oxygen would be removed so as to provide a reliable data support for subsequent water quality judgment, prediction and early warning.
In this paper, achievements were made in aspects of data fusion and error data processing in the aquaculture environment. But the following work can also be conducted at levels of feature level fusion and decision level fusion so as to better consider the rationality of the integration at a more complete level. During the process of fusion, the combination of intelligent algorithms such as Bayesian, expert system and cluster analysis can bring more reliable results to the fusion. Meanwhile in processing the abnormal data, it would see comprehensive consideration in aspects of the eliminating the error data (using Wright criterion and histogram method, etc.) and the completion of missing data (using Newton interpolation and Lagrange interpolation method, etc.). Thus, we can constantly improve the fusion accuracy, and promote the development and application of data fusion.

Acknowledgement: This study was supported by National Key Research and Development Project (Project No. 2017YFD0301506), National Social Science Foundation (Project No. 71774052), Hunan Education Department Scientific Research Project (Project No. 17K044; 17A092).

References

Bar-shalom, Y. (1981): On the track-to-track correlation problem. *IEEE Transactions on Automaction Control*, vol. 26, no. 2, pp. 571-572.

Carlson, N. (1990): Federated square root filter for decentralized parallel processes. *IEEE Trasactions on Aerospace and Electronic System*, vol. 26, no. 3, pp. 517-525.

Cai, Z.; Wang, Z.; Zheng, K.; Cao, J. (2013): A distributed TCAM coprocessor architecture for integrated longest prefix matching, policy filtering, and content filtering. *IEEE Transactions on Computers*, vol. 62, no. 3, pp. 417-427.

Cui, J.; Zhang, Y.; Cai, Z.; Liu, A.; Li, Y. (2018): Securing display path for security-sensitive applications on mobile devices. *Computer, Materials & Continua*, vol. 55, no. 1, pp. 17-35.

Ding, S.; Zhang, J.; Zhang, X.; An, R. (2017): Research progress of multi-classification twin support vector machines. *Journal of Software*, vol. 29, no. 1, pp. 89-108.

Guan, L. (2013): *Research on Monitoring Data Fusion in Indoor Environment Method Based on Bayesian Network (Ph.D. Thesis).* Jilin University.

Guo, N. (2013): *Research on Multi-Sensor Data Fusion Algorithm Based on Fuzzy Sets and Statistical Theory (Ph.D. Thesis).* Taiyuan University of Technology.

Guo, Q. (2017): *Uncertainty Information System Based on Rough Set Theory and Its Decision Research (Ph.D. Thesis)*. Hefei University of Technology.

Guo, Y. (2017): *Port Customer Credit Risk Assessment System Based on Semi-Supervised Learning and Information Fusion (Ph.D. Thesis).* Beijing Jiaotong University.

Huang, M.; Liu, Y.; Zhang, N.; Xiong, N.; Liu, A. et al. (2018): A services routing based caching scheme for cloud assisted CRNs. *IEEE Access*, vol. 6, no. 1, pp. 15787-15805.

Jia, C. (2017): *Research on the Adaptive Control Method of Multiple Models Based on Neural Networks (Ph.D. Thesis).* Beijing University of Science and Technology.

Kaur, J.; Kaur, K. (2017): A fuzzy approach for an IoT-based automated employee performance appraisal. *Computers, Materials & Continua*, vol. 53, no. 1, pp. 23-36.

Kim, K. (2002): Development of track to track fusion algorithms. *American Control Conference*, vol. 1, pp. 1037-1041.

Li, X. (2017): *Research on the Balance of Automobile Assembly Line Based on Improved Genetic Algorithm (Ph.D. Thesis).* Beijing University of Science and Technology.

Li, D.; Zhang, G.; Xu, Z.; Lan, Y.; Shi, Y. et al. (2018): Modelling the roles of cewebrity trust and platform trust in consumers' propensity of live-streaming: an extended TAM method. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 137-150.

Ling, L.; Li, Z.; Chen, C.; Gu, Y.; Li, C. (2000): Optimal allocation of time weights for multi-sensor data fusion. *Journal of Chinese Inertial Technology*, vol. 8, no. 2, pp. 36-39.

Liu, F.; Tang, G.; Li, Y.; Cai, Z.; Zhang, X. et al. (2019): A survey on edge computing systems and tools. *Proceedings of the IEEE*.

Liu, J.; Zhao, Z.; Liu, J.; Zeng, D. (2005): Realizing soft measurement of ventilator flow based on data fusion method. *Journal of North China Electric Power University (Natural Science)*, vol. 32, no. 3, pp. 61-65.

Liu, S.; Cai, Z.; Xu, H.; Xu, M. (2015): Towards security-aware virtual network embedding, *Computer Networks*, vol. 91. pp. 151-163.

Liu, X.; Dong, M.; Liu, Y.; Liu, A.; Xiong, N. (2018): Construction low complexity and low delay CDS for big data codes dissemination. *Complexity.*

Lu, Z. (2016): *Study on the Early Fault Diagnosis Method of Rotating Machinery Based on Variational Mode Decomposition and Optimization of Multi-core Support Vector Machines (Ph.D. Thesis)*. Chongqing University.

Mao, L.; Xiao, W.; Yang, H. (2015): Using information fusion technology to improve aquaculture water quality monitoring system. *Journal of Fisheries Science*, vol. 29, no. 3, pp. 55-58.

Moigne, J. (2000): Image registration and fusion for NASA remotely sensed imagery. *IEEE International Conference on Information Fusion*, pp. 24-31.

Olfati-Saber, R. (2008): Distributed Kalman filtering for sensor networks. *IEEE Decision and Control,* pp. 5492-5498.

Rao, N. (2000): Fusion method for physical systems based on physical laws. *IEEE International Conference on Information Fusion*, pp. 15-21.

Reboul, S.; Bruge, D. (2000): Optimal segmentation by random process fusion. *IEEE International Conference on Information Fusion*, pp. 19-23.

Sun, S.; Deng, Z. (2004): Multi-sensor optional information fusion Kalman filter with application. *Aerospace Science and Technology*, vol. 8, no. 1, pp. 57-62.

Sun, W.; Cai, Z.; Li, Y.; Liu, F.; Fang, S. et al. (2018): Security and privacy in the medical internet of things. *Security and Communication Networks.*

Wu, R. (2016): *Key Technology Research on Information Fusion in Military Information System (Ph.D. Thesis).* University of Electronic Science and Technology.

Xie, Y.; Xie, X. (2014): Feature selection algorithm based on feature subset differentiation and support vector machine. *Journal of Computer Science,* vol. 37, no. 8, pp. 1704-1718.

Zhang, H.; Cai, Z.; Liu, Q.; Xiao, Q.; Li, Y. et al. (2018): A survey on security-aware network measurement in SDN, *Security and Communication Networks*.

Zhang, Z. (2016): *Structural Reliability Analysis Method Based on Evidence Theory (Ph.D. Thesis)*. Hunan University.

Zuo, J. (2005): *Multi-sensor Information Fusion Filter Based on Kalman Filtering Method (Ph.D. Thesis).* Heilongjiang University.