A New Data Fusion Method–Generalized Fusion Method Postprint

Authors: Wei Yanfei, Geng Jianping, Shi Huli, Li Jingjing, Li Lin, Xu Xichong

Date: 2017-10-11T00:00:00+00:00

Abstract

The purpose of data fusion is to merge or synthesize multiple datasets into a more complete and higher-performance dataset. Based on the generalized extension interpolation extrapolation method, this study proposes a mathematical model for the fusion optimization of two or more sets of equation solutions and data sequences, referred to as the generalized fusion method. This approach can fully exploit the potential of data and measurement resources, employing innovative integration and mutual calibration algorithms to address the challenges faced in data fusion. The main steps for establishing and solving the generalized fusion mathematical model are introduced; the allocation principles of weights in the model are analyzed; and finally, it is applied to the field of satellite navigation data processing. Experimental results demonstrate that this method exhibits good practicality, excellent data stability, and high extrapolation stability, showing promise as a new universal solution method for data fusion problems.

Full Text

A New Data Fusion Method: The Generalized Fusion Method

Journal: Astronomical Research & Technology
Volume: 13, No. 3, July 2016
ISSN: 53-1189 (Print), 1672-7673 (Online)

Authors:
Geng Jianping[1], Shi Huli[2], Lin[2], Xu Xikun[1], Li Jingjing[4], Li

Affiliations:
[1]. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China
[2]. Beijing Riyue Jiutian Technology Co., Ltd., Beijing 100012, China

[3]. National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
[4]. School of Electronic Information, Zhongyuan University of Technology, Zhengzhou, Henan 450007, China

Abstract

The purpose of data fusion is to merge or synthesize multiple datasets into a complete dataset with improved performance. Based on the generalized extension interpolation and extrapolation method, this paper proposes a mathematical model that establishes and optimally fuses two or more sets of equation solutions and data sequences, referred to as the generalized fusion method. This approach can fully exploit the potential of data and measurement resources, employing innovative integration and mutual calibration algorithms to solve challenging problems in data fusion. The paper introduces the main steps for establishing and solving the generalized fusion mathematical model, analyzes the weight distribution principles within the model, and finally applies the method to satellite navigation data processing. Experimental results demonstrate that the method offers good practicality, stable data performance, and high extrapolation stability, showing promise as a new universal solution for data fusion problems.

Keywords: Data fusion; Generalized fusion; Generalized extrapolation; Model solving; Weight distribution; Satellite navigation data processing

2. Generalized Extension Interpolation and Extrapolation Method

The generalized fusion method is built upon the foundation of generalized extension interpolation and extrapolation. The fundamental design concept of this extrapolation method involves implementing piecewise least squares approximation, employing interpolation processing at segment endpoints and fitting approximation within segments. During intrasegment fitting, data from neighboring extension domains are utilized for auxiliary fitting, thereby enhancing the approximation accuracy of the polynomial within the definition domain. When segments are connected, the transition at connection endpoints is smooth, resulting in relatively high approximation accuracy across the entire domain. The extrapolation model of the generalized extension least squares approximation method integrates both interpolation and fitting functions, achieving excellent approximation precision.

Assuming the optimal estimate $\hat{x}_n$ at epoch t_n and the data values before t_n are known, the problem of solving for the next moment' s value x_n represents a typical data extrapolation challenge. The following generalized extension interpolation and extrapolation model can be established:

$$\min_{\mathbf{a}} \sum_{i=n-m}^{n-1} [a_0 + a_1 t_i + a_2 t_i^2 - x_i]^2$$
$$\text{s.t. } a_0 + a_1 t_n + a_2 t_n^2 = \hat{x}_n$$

where $\mathbf{a} = [a_0, a_1, a_2]^T$ represents the coefficients to be determined. By optimizing the objective function of approximation residuals, the unknown coefficients of the generalized extension approximation polynomial can be obtained. Subsequently, the estimated value $\hat{x}_n$ at epoch t_n can be extrapolated and solved using:

$$\hat{x}_n = a_0 + a_1 t_n + a_2 t_n^2$$

[Figure 1: see original paper] shows the schematic diagram of generalized extension extrapolation.

---

3. Mathematical Model of Generalized Fusion Method and Its Solution

3.1 Establishment of the Generalized Fusion Mathematical Model

The generalized fusion solution model draws on the generalized extension interpolation and extrapolation formula. During approximation, data from a segment of epochs are selected, where prior optimal estimates (which can be replaced by measurement values initially) are used for fitting approximation processing. The optimal weighted combination of the predicted value of the generalized state variable at the current epoch to be solved and the measurement value serves as the interpolation constraint locking point. The polynomial coefficients **a** are used as optimization variables, and the feasible interval numbers of the variables serve as constraints, limiting the solution to a certain range and thereby improving solution accuracy.

The generalized fusion optimization solution model is constructed as:

$$\min_{\mathbf{a}} \sum_{i=n-m}^{n-1} [a_0 + a_1 t_i + a_2 t_i^2 - \hat{x}_i]^2 + k_1 [a_0 + a_1 t_n + a_2 t_i^2 - \tilde{x}_n]^2$$
$$+ k_2 [a_0 + a_1 t_n + a_2 t_n^2 - x_n]^2$$
$$\text{s.t. } \mathbf{a} \in \Omega$$

where i is the epoch index, x_i is the measured state value, $\hat{x}_i$ is the optimal estimate of the state value, $\tilde{x}_n$ is the predicted state value at epoch t_n obtained through low-order state variable recursion, and $\mathbf{a} = [a_0, a_1, a_2]^T$ represents the coefficients to be optimized. The objective function can be either linear or nonlinear. For nonlinear relationships, the generalized fusion method directly employs nonlinear solution algorithms. The model can incorporate nonlinear state equations, nonlinear constraint equations, and other nonlinear measurement equations, which expands its application scope, increases equation redundancy and continuity, and improves solution accuracy.

3.2 Solution Steps for the Generalized Fusion Method

Step 1: Establish the residual minimization optimization model with constraint equations. The method is not limited by the number of equations during optimization and can be solved directly using nonlinear algorithms such as the simplex method. This approach offers simplicity, intuitiveness, wide applicability, and fast solution speed.

Step 2: Solve for the predicted state value $\tilde{x}_n$. When solving for the predicted state value $\tilde{x}_n$ at epoch t_n, the optimal estimate $\hat{x}_n$ at epoch t_n has already been obtained. The predicted value can be recursively obtained through:

$$\tilde{x}_n = \hat{x}_{n-1} + \dot{x}_n \Delta t + \ddot{x}_n \Delta t^2$$

where $\dot{x}_n$ is the first derivative (velocity) at epoch t_n, $\ddot{x}_n$ is the second derivative (acceleration), and Δt is the time interval between adjacent epochs. Higher-order state variables obtained through extrapolation can improve solution accuracy and correlation. During initial fusion when the optimal initial state estimate is unknown, measurement values from the first few moments can be used.

Step 3: Solve for the optimal state estimate $\hat{x}_n$. After obtaining the state value and predicted state value $\tilde{x}_n$ at epoch t_n, along with the directly measured state value x_n at epoch t_n, these values are combined. Using direct optimization algorithms, the optimal estimate $\hat{x}_n$ at epoch t_n can be obtained from the generalized fusion optimization model.

Step 4: Iterate to obtain an optimized set of optimal estimates. By continuously iterating through these steps, an optimized set of optimal estimates $\hat{x}_i$ $(i = 1, 2, 3, ...)$ can be obtained.

3.3 Weight Analysis in the Generalized Fusion Model

In the generalized fusion mathematical model, the term $[a_0 + a_1 t_n + a_2 t_n^2 - \tilde{x}_n]$ represents the residual between the approximation function and the predicted state value, while $[a_0 + a_1 t_n + a_2 t_n^2 - x_n]$ represents the residual between the approximation function and the measured state value. Their weight coefficients

k_1 and k_2 significantly affect the contribution of their respective residuals to the objective function, making weight selection critically important.

The magnitude of these weights influences the proximity of the optimal estimate to the measured or predicted values. Before determining the weight coefficients, it is necessary to analyze the accuracy and error characteristics of the state predicted values, measured values, and fusion data itself. During data fusion, adjusting k_1 and k_2 can control the closeness of the generalized fusion optimal estimate to either the measured or predicted values.

To verify the method' s feasibility through simulation, two datasets were generated: one with large random errors but small bias (range: -3 to 3), and another with small random errors but large offset (range: -1 to 1). After data generation, different weight combinations were analyzed using the generalized fusion solution model. The results are shown in [Figure 2: see original paper]. The four subplots (a), (b), (c), and (d) demonstrate that as k_1 gradually decreases and k_2 gradually increases, the optimal estimate of the state quantity moves away from the measured value and approaches the predicted value. This effect becomes more pronounced with larger weight differences. If certain measurements in a data sequence experience sudden jumps causing data inaccuracy, adjusting the weight combination can reduce the contribution of these anomalous data points to the optimal estimate, yielding smoother results.

---

4. Application of Generalized Fusion Method in Satellite Navigation and Simulation/Measured Data Analysis

4.1 Application of Generalized Fusion in Satellite Navigation

Using satellite positioning and navigation as an example, this section demonstrates the practicality of the generalized fusion method. For simplicity, only one directional component of user position is analyzed in detail; the other two components can be corrected using the same method to obtain complete user coordinates.

Step 1: Obtain the optimal position estimate $\hat{x}_m$ at time t_m and the corresponding velocity $\hat{v}_m$. The receiver' s position measurements from the initial moment to epoch t_m are denoted as x_i $(i = 1, 2, \ldots, m)$. Position measurements are processed using piecewise least squares fitting. After processing, the value at time t_m obtained through the approximation polynomial serves as the optimal position estimate $\hat{x}_m$ at time t_m. In navigation positioning, velocity measurements have relatively high accuracy, so the optimal velocity values at the initial moment and subsequent moments can directly use the corresponding measured velocity values.

Step 2: Obtain the predicted position value $\tilde{x}_m$ at time t_m. Having obtained the optimal position estimate $\hat{x}_m$ and the corresponding optimal velocity

$\hat{v}_m$ at time t_m, the predicted position value $\tilde{x}_m$ at time t_m can be calculated using formula (5), where the first derivative $\dot{x}_n$ represents velocity at epoch t_n and the second derivative $\ddot{x}_n$ represents acceleration.

Step 3: Obtain the optimal estimate $\hat{x}_m$ at time t_m. Through the above steps, the measured position value at time t_m and the predicted position value $\tilde{x}_m$ at time t_m have been obtained. These values must now be combined to obtain the optimal estimate $\hat{x}_m$ at time t_m. A threshold is added between the measured and predicted values. When their absolute difference exceeds this threshold, the measured value is considered to have experienced a jump, and the weight combination in the formula is adjusted to steer the optimal estimate toward the predicted value. When the absolute difference does not exceed the threshold, the measured data is considered within the error range, and an appropriate weight combination is selected to obtain the optimal estimate $\hat{x}_m$.

Step 4: Iterate to obtain an optimized set of optimal estimates. By continuously iterating through these steps, an optimized set of optimal estimates $\hat{x}_m, \hat{x}_{m+1}, \hat{x}_{m+2}, ...$ can be obtained.

4.2 Simulation and Measured Data Analysis

4.2.1 Correction of Jump Data by Generalized Fusion The accuracy of satellite positioning systems depends on two main factors: the spatial distribution of observed satellites (commonly called Geometric Dilution of Precision, GDOP) and the accuracy of each observation measurement. Residual errors from tropospheric delay correction and multipath effects can cause measurement data to develop biases or sudden jumps. The generalized fusion method can effectively correct these erroneous data.

A dataset with random errors in the range of (-0.5, 0.5) was generated, with larger offset values added at points 10, 20, 30, and 40 to create jump points. This dataset was then processed using the generalized fusion method. The simulation results in [Figure 3: see original paper] show that the method provides real-time correction of jump data. The results not only faithfully follow the variation pattern of the original measured values in non-jump sections but also effectively eliminate the influence of jumps in anomalous sections.

4.2.2 Comparison Between Least Squares Fitting and Generalized Fusion Method When the number of satellites exceeds four and each measurement is independent and follows a Gaussian normal distribution, the least squares algorithm is the most classical approach for solving the receiver' s three-dimensional coordinates. This algorithm minimizes the sum of squared residuals among observed pseudoranges, thereby achieving optimal estimates. However, in practice, measurements are not independent, and errors from multipath effects constantly affect positioning accuracy. In some cases, the data may not even follow a normal distribution, making it difficult for the least squares algorithm to guarantee optimal estimates.

To verify the effectiveness of the generalized fusion method under such conditions, a simulation experiment was conducted. The theoretical value of the measurements was set, and noise was added to generate a dataset with random errors. Both the generalized fusion method and the least squares algorithm were applied to process this dataset. The correction effects are shown in [Figure 4: see original paper]. The standard deviation of the least squares fitted values was $\sigma_{LS} = 3.626396349003966$, while the standard deviation of the generalized fusion optimal estimates was $\sigma_{GF} = 0.758068193045756$. The generalized fusion method improved stability by approximately 4.8 times compared to the least squares method, with smaller bias as well, bringing the corrected error data closer to the theoretical values.

4.2.3 Analysis of Measured Data On May 26, 2015, a dynamic test was conducted at the National Astronomical Observatories, Chinese Academy of Sciences, using a receiver with an ATGM332D satellite positioning chip. A segment of raw positioning data with obvious jumps and corresponding raw velocity values in the Y-direction was extracted from the final obtained receiver three-dimensional coordinates. Both the generalized fusion method and the least squares fitting method were applied to process this data. The comparison results in [Figure 5: see original paper] show that the least squares fitting method did not effectively repair certain jump points, and although it followed the variation pattern of the original data at a few points, the overall data deviated from the measured data. In contrast, the generalized fusion method can real-time correct bad data in the original measurements and faithfully follow the trajectory of the raw data in non-jump sections.

Modern society demands increasingly higher precision for time-frequency references. While cesium atomic clocks and hydrogen atomic clocks have been established as high-precision time-frequency standards, even higher precision and stability can be achieved through combined processing and refined processing of atomic clock ensemble data. This paper applies the generalized fusion method to fuse the output data from hydrogen and cesium atomic clocks. [FIGURE:6(a)] shows the frequency data output from cesium and hydrogen atomic clocks. Comparison reveals that the cesium clock has relatively large random errors but excellent long-term stability, whereas the hydrogen clock exhibits small random error amplitude but significant frequency drift over time. The generalized fusion method can combine the advantages of both atomic clocks' frequency characteristics to obtain a time-frequency reference with small random errors and good long-term stability. This method was applied to fuse time-frequency data from cesium and hydrogen atomic clocks at the National Time Service Center, using the hydrogen clock' s frequency variation rate to correct the cesium clock' s frequency output. The results are shown in [FIGURE:6(b)].

Conclusion

This paper proposes a new data fusion method—the generalized fusion method—based on the generalized extension interpolation and extrapolation method, specifically addressing bad data, particularly jump data, in satellite navigation positioning. Through simulation and measured data analysis, compared with the least squares algorithm, the generalized fusion method demonstrates superior capability in repairing jump data generated during the positioning process, offering higher reliability in data processing. The generalized fusion method exhibits good practicality, stable data performance, and high extrapolation stability, and is expected to become a new universal solution for data fusion problems.

References

[1] Wang Jiongqi, Zhou Haiyin, Wu Yi. The theory of data fusion based on state optimal estimation. *Applied Mathematics*, 2006.

[2] Shi Zhong, Huo Xing. Application of multi-sensor information fusion technique in INS/GPS integrated navigation system. *Journal of Bohai University (Natural Science Edition)*, 2014, 403–407.

[3] Shang Lin, Liu Guohua, Zhang Rui, et al. An information fusion algorithm for integrated autonomous orbit determination of navigation satellites. *Acta Astronautica*, 2015.

[4] Ning Xiaolin, Fang Jiancheng. An autonomous celestial navigation method for LEO satellite based on unscented Kalman filter and information fusion. *Aerospace Science and Technology*, 2015, 222–228.

[5] Tzschichholz T, Boge T, Schilling K. Relative pose estimation of satellites using PMD/CCD-sensor data fusion. *Acta Astronautica*, 2015, 25–33.

[6] Wang Xinlong, Zhang Qing, Li Hengnian. An autonomous navigation scheme based on starlight, geomagnetism, and gyros with information fusion for small satellites. *Acta Astronautica*, 2015, 708–717.

[7] Shi Huli, Yan Yihua, Xu Guohua, et al. Generalized extension approximation method in engineering science. Beijing: Science Press, 2015.

[8] Geng Jianping, Yi Wei, Liu Cheng, et al. A new method for detection and correction of single-frequency cycle slips using data extrapolation based on the method of generalized extended interpolation. *Astronomical Research & Technology*, 2015, 174–182.

[9] Shi Huli, Huang Kang, Yi Wei, et al. A new approximation filtering method: China Patent, 201410208791.7, 2014-08-19.

[10] He Ruizhu, Liu Cheng, Huang Kang. A method of multiple-step weighted least square estimate for satellite positioning. *Astronomical Research & Technology*, 2015, 36-43.

[11] Zhu Limei, Zhang Xuchun. Analysis of the positioning error on elevation during the high dynamic. *Journal of Naval Aeronautical and Astronautical University*, 2015, 203-207.

Note: Figure translations are in progress. See original paper for figures.