A Multi-source Experimental Data Fusion Evaluation Method Based on Bayesian Method and Evidence Theory

Huan Zhang, Wei Li, Ping Ma, Ming Yang
Control and Simulation Center
Harbin Institute of Technology, Harbin, Heilongjiang, China
e-mail: zhanghuan_1996@163.com, fleehit@163.com, pingma@hit.edu.cn, myang@hit.edu.cn

***Abstract*—The experimental data used for system performance evaluation have many sources and multiple granularity. Thus, it is necessary to fuse multi-source experimental data for evaluation. Aiming at multi-source experimental data fusion problem, a multi-source experimental data fusion evaluation method based on Bayesian and evidence theory is proposed. According to the size of the data sample size, the large sample data are fused by the classical frequency method, while the small sample data are fused by Bayesian method. Then the parameter information is updated by the Bayesian method and data fusion result is obtained. The experimental data of different scenarios are fused by evidence theory. The evidence combination method is used to fuse the data when the evidence bodies do not conflict, while the weighted average correction method is used for data fusion when there is conflict between the evidence bodies. The result of the multi-source experimental data fusion is obtained based on Bayesian method and evidence theory.**

Keywords- data fusion; performance evaluation; Bayesian method; evidence theory

I. INTRODUCTION

As the complexity of the system increases, the performance evaluation of the system becomes more and more important [1]. The performance evaluation of the system is based on experimental data. However, only a small amount of measured data can be obtained because the cost of system real experiments is more and more expensive due to the use of high technology [2]. Besides, experimental data can be obtained through semi-physical simulation and digital simulation experiments. Therefore, the experimental data of performance evaluation have the characteristics of multi-source, multi-capacity, multi-granularity and multi-type, etc., and it is necessary to fuse multi-source data and then conduct comprehensive evaluation. The concept of data fusion originated in the 1970s, first used in the military field [3], and later gradually applied to various non-military fields [4]. There is no uniform definition of data fusion, and researchers have given multiple definitions from different aspects. The typical definition is: data fusion is a multi-level and multi-faceted data processing process that automatically detects, correlates, estimates and combines data from multiple sources [5].

Data fusion is applied to various theoretical knowledge and cutting-edge technologies, but no uniform algorithm is suitable for all scenarios due to its wide range of applications [6][7]. Data fusion method can be mainly divided into three categories: signal processing and estimation theory method, statistical inference theory method, and information theory method. The commonly used data fusion methods mainly include weighted average, classical reasoning, Bayesian fusion [8] and fuzzy theory. Aiming at the multi-source experimental data of performance evaluation, Bayesian method, as a commonly used method [9], can process data in the form of static probabilities and has good validity; evidence theory, second only to Bayesian method, has no requirement for data sample size and has good data fusion effect; classical reasoning method is also suitable for processing large sample size data and also has good validity; other commonly used data fusion methods, such as fuzzy theory, are applicable to processing experimental data.

Aiming at multi-source experimental data, a data fusion method is proposed. In section 2, we analyze the characteristics of experimental data and the applicable conditions of Bayesian method and evidence theory. Then we propose a data fusion evaluation scheme based on the characteristics of experimental data. In section 3, data fusion method is introduced in detail. According to the sample data, multi-type Bayesian fusion can be divided in two cases and evidence theory fusion is also related to whether the evidence bodies are in conflict. In section 4, we present the conclusion of experimental data fusion method in system performance evaluation and the future work in next step.

II. MULTI-SOURCE EXPERIMENTAL DATA FUSION SCHEME

Experimental data used for system performance evaluation can be divided into measured data, semi-physical simulation data, digital simulation data, etc. The measured data and semi-physical simulation data are obtained through real experiments and semi-physical simulation respectively, while the digital simulation data are obtained by running the digital simulation system. Therefore, there are multiple types of data in the experimental data. Otherwise, the experimental data can also come from different experimental scenarios. The experimental scenario defines the scope and constrains of the problems studied in the system, variables, activities and interactive relations related to the experimental objects, etc., and it includes the setting and organization of all kinds of data in the system [10]. Thus, multi-source experimental data fusion problem can be divided into two aspects: one is the multi-type data fusion problem, and another is different experimental scenarios' data fusion problem.

Bayesian fusion method can process the data in the form of static probability and has no requirement on sample size, which is also the most efficient fusion method. Therefore, Bayesian method is considered for using to fuse multi-type data. For data fusion problem of different experimental scenarios, Bayesian fusion method is also considered for data fusion in order to ensure the validity of fusion results and the uniformity of methods. Because the experimental data are from different types of data under different scenarios, the posterior density $\pi\left(\theta,\lambda|X\right)$ is considered in Bayesian fusion, where parameter λ and parameter θ characterize the mean of the variables and different experimental scenarios respectively and X is observation sample. However, θ's posterior density $\pi\left(\theta|X\right)$ is focused on generally. The calculation of $\pi\left(\theta|X\right)$ needs the information about $\pi\left(\lambda\right)$ and $\pi\left(\theta\,|\,\lambda\right)$. $\pi\left(\lambda\right)$ is the parameter distribution of different experimental scenarios, which is easily to obtain, while $\pi\left(\theta\,|\,\lambda\right)$ is conditional probability which is hard to obtain directly in the actual engineering application. Therefore, Bayesian method is not suitable for data fusion of different experimental scenarios. Evidence theory also has great advantages in data fusion and can process data of different types or different sample sizes. Thus, evidence theory is used to fuse data from different scenarios.

Above all, Bayesian fusion method is suitable for multi-type data fusion while evidence theory is suitable for fusing sample data from different scenarios. Thereout, the multi-source experimental data fusion scheme based on Bayesian method and evidence theory is obtained (see Figure 1). That is, classify the multi-source experimental data firstly, fuse the multi-type data of the same experimental scenario using Bayesian data fusion method and then fuse the single data from different scenarios using evidence theory fusion method. Considering multi-type data mainly from simulation data, semi-physical simulation data and measured data, it is necessary to classify data before data fusion, perform data preprocessing for each type of data, and then fuse data. The simulation data have a large amount of data and low authenticity, and the classical frequency estimation method is used for data fusion. However, the semi-physical simulation data and the measured data have a small number of data, and Bayesian method is used. Then, regarding large sample frequency data fusion result and small sample Bayesian data fusion result as prior information and observation sample data, respectively, the Bayesian method is used to update the prior information parameters to obtain the multi-type data fusion results under the same scenario. After obtaining the multi-type data fusion results of each scenario, the evidence theory method is used to fuse data of different experimental scenarios. Firstly, model the sample data of each scenario and describe it as a form of evidence body. When the evidence bodies do not conflict, the conventional evidence body fusion method is used for processing; when the evidence bodies are in conflict, process using the weighting method and preprocess the conflict evidence before the evidence combination. Finally, multi-source experimental data fusion results are obtained.

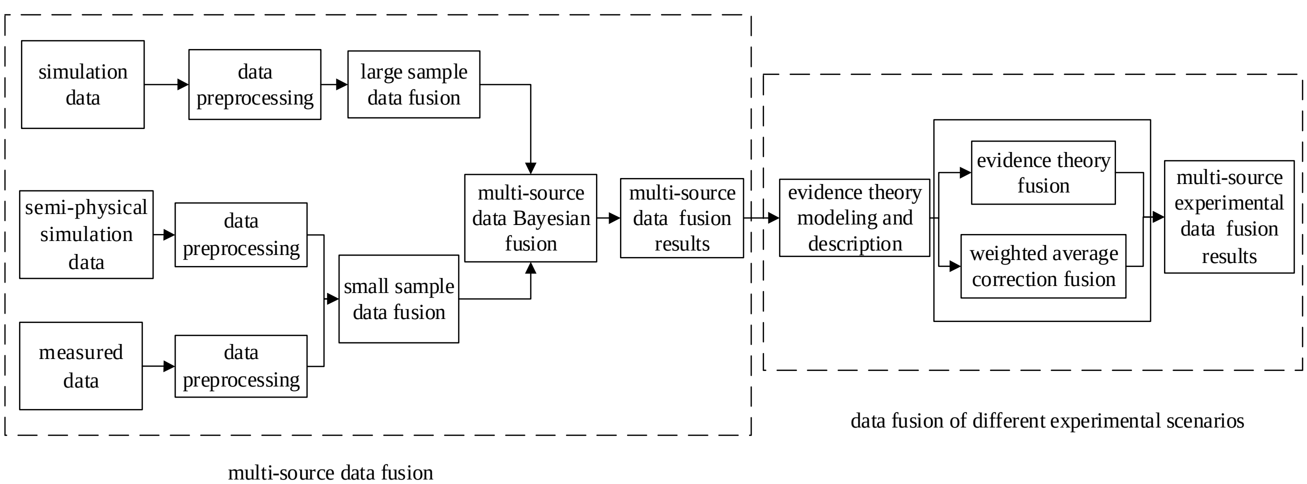

Figure 1. Multi-source experimental data fusion scheme

III. MULTI-SOURCE DATA FUSION METHOD BASED ON BAYESIAN AND EVIDENCE THEORY

According to the multi-source experimental data fusion scheme described above, multi-source experimental data fusion mainly uses multi-type data Bayesian fusion method and evidence theory fusion method of different experimental scenarios. The description of these two data fusion methods is given.

A. *Multi-type Bayesian Fusion*

The multi-type data Bayesian fusion method mainly uses the Bayesian method to update the prior information parameters, and the obtained result is the data fusion result.

The principle of the Bayesian update method is described as follows. Assuming that parameter θ is the mean of a variable, X is observation sample and $\pi(\theta)$ is the prior density of parameter θ. Then, the posterior density $\pi(\theta|X)$ of θ according to Bayesian formula is

$$\pi(\theta|X)=\frac{f(X|\theta)\pi(\theta)}{\int_{\Theta} f(X|\theta)\pi(\theta)d\theta} \tag{1}$$

where Θ is the parameter space and $f(X|\theta)$ is the likelihood function of X given by θ. From (1), we can see that the observation sample and the prior density of θ are used to calculate the posterior density. In system performance evaluation, the measured data and the semi-physical simulation data generally have small sample sizes, but the degree of authenticity is high, which can be used as the observation sample; the digital simulation data generally have large sample size, but the degree of authenticity is low, from which can obtain the prior density of θ. Thus, digital simulation data, semi-physical simulation data and measured data need to be fused before Bayesian fusion.

1) Classical frequency fusion of large sample data

Generally, digital simulation data are large sample data and classical frequency parameter estimation is widely used in data fusion under large sample conditions. Therefore, the large sample data are fused by the classical frequency method. Firstly, assume that there are n independent sample data sets in the simulation data. Their estimated value of the same distribution parameter θ is $\{\theta_1,\theta_2,...,\theta_n\}$. And the observation for each sample data set is $\theta_i=\theta+\xi_i, i=1,2,...,n$, where ξ_i is a random error and independent of each other and ξ_i also obeys normal distribution. The estimated value $\hat{\theta}$ of ξ_i is described as arithmetic mean of observations, i.e., $\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}\theta_i$ and the mean variance of $\hat{\theta}$ is $\hat{\sigma}^2=\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\theta_i-\bar{\theta}\right)^2$.

For ease of processing data, n observations are divided into k batches, where the j th batch is described as $\left\{\theta_{j1},\theta_{j2},...,\theta_{jn_j},n_j\geq 2, j=1,2,...,k\right\},\sum_{j=1}^{k}n_j=n$. The mean of the j th batch is $\bar{\theta}_j=\frac{1}{n_j}\sum_{i=1}^{n_j}x_{ji}, j=1,2,...,k$ and its variance is $\hat{\sigma}_j^2=\frac{1}{n_j(n_j-1)}\sum_{i=1}^{n_j}\left(\theta_{ji}-\bar{\theta}_j\right)^2, j=1,2,...,k$. When $\left\{\bar{\theta}_1,\bar{\theta}_2,...,\bar{\theta}_k\right\}$ is seen as k unequal precision observations or observations from k different sample sets of distribution parameter θ, each $\bar{\theta}_j$ can be expressed as $\bar{\theta}_j=\theta+\xi'_j, j=1,2,...,k$, where ξ'_j is a random error and independent of each other and $\xi'_j \sim N\left(0,\hat{\sigma}_j^2\right)$. The likelihood function of $\left\{\bar{\theta}_1,\bar{\theta}_2,...,\bar{\theta}_k\right\}$ is obtained and it is

$$L=\prod_{j=1}^{k} f\left(\bar{\theta}_j,\hat{\sigma}_j^2,\theta\right) \tag{2}$$

where $f\left(\bar{\theta}_j,\hat{\sigma}_j^2,\theta\right)=\frac{1}{\sqrt{2\pi}\hat{\sigma}_j}\exp\left[-\frac{\left(\bar{\theta}_j-\bar{\theta}\right)^2}{2\hat{\sigma}_j^2}\right]$.

Then, the estimated value of parameter θ is obtained by maximum likelihood estimation. The estimated value is $\hat{\theta}=\bar{\theta}=\left(\sum_{j=1}^{k}\frac{1}{\hat{\sigma}_j^2}\bar{\theta}_j\right)\left(\sum_{j=1}^{k}\frac{1}{\hat{\sigma}_j^2}\right)^{-1}$ and the mean variance is $\hat{\sigma}^2=\left(\sum_{j=1}^{k}\frac{1}{\hat{\sigma}_j^2}\right)^{-1}$. Thus, the prior information of parameter θ is obtained by fusing digital simulation data through classical frequency fusion method.

2) Bayesian fusion of small sample data

The measured data and the semi-physical simulation data may be small sample data, and Bayesian method is used to fuse small sample data. Assume that there are n sets of the same type of prior information before the small sample data, which is compatible with small samples. The prior density $\pi_i(\theta)$ of the distribution parameter θ can be obtained from each set of prior samples, whose weight is b_i, $(i=1,2,\dots,n)$. The prior distribution density of parameter θ is

$$\pi(\theta)=\sum w_i\pi_i(\theta) \tag{3}$$

where $w_i=b_i/\sum b_i$, $(i=1,2,\dots,n)$. Then, the posterior density of the distribution parameter θ is obtained by Bayesian formula, which is

$$\pi(\theta|X)=\frac{f(X|\theta)\pi(\theta)}{\int_{\Theta} f(X|\theta)\pi(\theta)d\theta} \tag{4}$$

where Θ is parameter space and X is the sample obtained in the field test. $f(X|\theta)$ is the distribution density of subsample X when given θ. Thus, the posterior density of θ is

$$\pi\left(\theta \middle| X\right)=\frac{1}{m\left(X \middle| \pi\right)}\sum_{i=1}^{n} w_i \pi_i\left(\theta\right) f\left(X \middle| \theta\right) \tag{5}$$

where $m\left(X \mid \pi\right)=\int_{\Theta} f\left(X \mid \theta\right)\pi\left(\theta\right)d\theta$, which is the edge density of X .

For easy to calculation, an expression of posterior density when $n=2$ is given. In this case, there are two prior information, i.e.,

$$\pi\left(\theta \middle| X\right)=\frac{1}{m\left(X \middle| \pi\right)}\sum_{i=1}^{2} w_i \pi_i\left(\theta\right) f\left(X \middle| \theta\right) \tag{6}$$

Where

$$\pi_i\left(\theta \middle| X\right)=\frac{\pi_i\left(\theta\right) f\left(X \middle| \theta\right)}{m\left(X \middle| \pi_i\right)}, i=1,2 \tag{7}$$

Then, we can get

$$\pi\left(\theta \middle| X\right)=\frac{\sum_{i=1}^{2} w_i m\left(X \middle| \pi_i\right)\pi_i\left(\theta \middle| X\right)}{m\left(X \middle| \pi\right)} \tag{8}$$

Assume that

$$\lambda_1=\frac{w_1 m\left(X \middle| \pi_1\right)}{m\left(X \middle| \pi\right)}, \lambda_2=\frac{w_2 m\left(X \middle| \pi_2\right)}{m\left(X \middle| \pi\right)} \tag{9}$$

Finally, we obtain that

$$\pi\left(\theta \middle| X\right)=\lambda_1\left(X\right)\pi_1\left(\theta \middle| X\right)+\lambda_2\left(X\right)\pi_2\left(\theta \middle| X\right) \tag{10}$$

Using similar mathematical derivation like above, when n kinds of prior information exist in general, there is

$$\pi\left(\theta \middle| X\right)=\sum_{i=1}^{N}\lambda_i\left(X\right)\pi_i\left(\theta \middle| X\right) \tag{11}$$

where $\lambda_i\left(X\right)=\frac{w_i m\left(X \middle| \pi_i\left(\theta\right)\right)}{m\left(X \middle| \pi\left(\theta\right)\right)}, \left(i=1,2,\ldots,n\right)$ .From the above we can see, the posterior distribution of θ is fused by multiple posterior distributions when there is a variety of prior information and the weighted average of these posterior distributions is fusion posterior distribution. Then, the Bayesian fusion result of small sample data can be obtained.

Finally, the large sample data fusion result and the small sample data fusion result are regarded as prior information and observation sample data, respectively. The Bayesian method is used to update the prior information parameters, and the obtained result is the multi-type data fusion result under the same experimental scenario. Furthermore, multi-type data fusion results under different scenarios can be obtained.

B. Evidence Theory Fusion

The experimental data used in performance evaluation mostly come from different experimental scenarios, which evidence theory fusion method is used to fuse. After obtaining multi-type data fusion results of different experimental scenarios, the sample data of each scenario are described as evidence body using evidence theory. The evidence combination method is adopted when the evidence bodies do not conflict, while the weighted average correction fusion method is adopted when the evidence bodies conflict. Then, the data fusion results of different experimental scenarios are obtained.

1) Evidence combination

The evidence theory fusion criterion is used to synthesize the nonconflicting evidence bodies. This combination method is a strict AND operation method. The basic probability distribution of common focal elements of multiple belief functions is proportional to the respective basic probability distribution. Therefore, the evidence method has a focusing effect. This effect will strengthen support for common goals and weaken the impact of divergent goals. The principle of evidence combination is: if $Bel_1, Bel_2, ..., Bel_n$ is the n belief functions on the same identification frame; $m_1, m_2, ..., m_n$ is the corresponding basic probability assignment functions; $A_i, A_j, A_k, ...$ is the corresponding focal elements. Assume that

$$K=\sum_{A_i \cap A_j \cap A_k \cap \ldots = \varnothing} m_1\left(A_i\right) m_2\left(A_j\right) m_3\left(A_k\right)\ldots < 1 \tag{12}$$

Belief function Bel_D synthesized by $Bel_1, Bel_2, ..., Bel_n$ is determined by the basic probability assignment function m_D given below:

$$m_D(A)=\begin{cases}\frac{1}{1-K}\sum_{\cap A_i = A}\prod_{1\le i\le N} m_i(A_i) & \mathrm{A}\neq\varnothing\\ 0 & \mathrm{A}=\varnothing\end{cases} \tag{13}$$

Belief function Bel_D given by m_D is called the direct sum of $Bel_1, Bel_2, ..., Bel_n$, i.e.,

$$Bel_D = Bel_1 \oplus Bel_2 \oplus \ldots \oplus Bel_n \tag{14}$$

According to (14), the core of the belief function Bel_D is equal to the intersection of the cores of $Bel_1, Bel_2, ..., Bel_n$. If the cores of $Bel_1, Bel_2, ..., Bel_n$ do not intersect, $Bel_1, Bel_2, ..., Bel_n$ could not be synthesized, i.e., the evidence they correspond to supports completely different propositions. When $Bel_1, Bel_2, ..., Bel_n$ represents multiple batches of completely different evidence, it is the completely conflict evidence and cannot be synthesized by the evidence combination method.

2) Weighted average correction fusion

When evidence bodies are in high conflict, the result of combination is contrary to the common sense. The solutions to this problem can be mainly summarized into two categories: one is to modify the data fusion rules, and the other is to modify the data model and pre-process the conflict evidence before evidence combination. Because modifying the data fusion rules can lead to the destruction of the commutative law and associative law of Dempster's combination rule, the second solution conforms to the theoretical framework of the evidence theory method [11]. Weighted average correction method can reduce evidence conflict and guarantee the focusing effect of fusion to some extent. Thus, weighted average correction fusion method is used to deal with the contradictory evidence bodies.

The principle of weighted average correction is as follows:

$$m(A) = \frac{1}{n}\sum_{1}^{n} w_i m_i(A) \tag{15}$$

where w_i is the weight of each evidence and $\sum_{i=1}^{n} w_i = 1$. Weighted average correction method can achieve the suppression of evidence conflict through using different weighting methods and comprehensively consider the information of multiple conflicting evidence bodies. Furthermore, this method reduces the evidence conflicts by weighting, which has been widely used in the fusion of conflicting evidence bodies.

IV. CONCLUSION

In the performance evaluation, the experimental data need to be fused since it comes from digital simulation experiments, semi-physical simulation experiments and real experiments, etc., and it also comes from different experimental scenarios in some cases. The multi-source experimental data fusion problem is decomposed into multi-type data fusion problem and data fusion problem under different experimental scenarios, where multi-type data fusion adopts Bayesian method and data fusion of different scenarios uses evidence theory method. In the case of multi-type data fusion, the results obtained by the fusion of large sample data with the classical frequency method and the results obtained by the fusion of small sample data with the Bayesian method are respectively taken as prior information and observation sample data, and the Bayesian method is used to obtain the multi-type data fusion result. In the case of data fusion of different scenarios, each sample is described as the form of evidence body, and different data processing methods are adopted according to the relationship between the evidence bodies. The final result of multi-source experimental data fusion/is obtained through the proposed method.

The proposed data fusion method is suitable for processing static data. In the next step, we will select the appropriate data using the proposed method and other existing methods to do experiments. Then, according to the experimental results, compare fusion results to verify the effectiveness of this method.

REFERENCES

[1] Z. Xiaoge and M. Sankaran. Aircraft re-routing optimization and performance assessment under uncertainty. Decision Support Systems, vol. 96, pp. 67-82, Apr. 2017, DOI:10.1016/j.dss.2017.02.005.

[2] C. Fei and W. Xin. Application of Data Fusion in Anti-jamming Performance Evaluation of Missile Weapon Equipment System. Aerospace Electronic Warfare, vol. 30, no. 3, pp. 1-4, Jun. 2017.

[3] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," in Proceedings of the IEEE, vol. 85, no. 1, pp. 6-23, Jan. 1997.

[4] G. Li, C. Tiejun, W. Yuhai, and L. Chenyan. Study For Multi-resources Geospatial Data Integration. Geomatics World, no. 1, pp. 62-66, Feb. 2007.

[5] C. Kewen, Z. Zuping, and L. Jun. Multisource Information Fusion: Key Issues, Research Progress and New Trends. Computer Science, vol. 40, no. 8, pp. 6-13, Aug. 2013.

[6] R. R. Yager. Set Measure Directed Multi-Source Information Fusion. in IEEE Transactions on Fuzzy Systems, vol. 19, no. 6, pp. 1031-1039, Dec. 2011.

[7] Q. Youjie and W. Qi. Review of Multi-source Data Fusion Algorithm. Aerospace Electronic Warfare, vol. 33, no. 6, pp. 37-41, Dec. 2017.

[8] H. Jiang and S. Youchao. Application Study of Multi-sourced Reliability Data Fusion Method Based on Bayes Theory. Aircraft Design, vol. 32, no. 6, pp. 52-55, Dec. 2012.

[9] L. Qingmin, W. Hongwei, and L. Jun. Small sample Bayesian analyses in assessment of weapon performance. in Journal of Systems Engineering and Electronics, vol. 18, no. 3, pp. 545-550, Sept. 2007.

[10] Z. Ying, L. Jiangtao, and Z. Meng. Operational Experiment Scenario Design Based on UML. Command Control & Simulation, vol. 37, no. 2, pp. 86-91, Apr. 2015.

[11] Q. Xiaochao. Research on Simulation Model Validation and Calibration Methods Under Uncertainty. Harbin Institute of Technology, 2016.

ISBN: 978-1-61208-741-2