International Journal of Database Theory and Application
Vol.8, No.6 (2015), pp.223-232
http://dx.doi.org/10.14257/ijdta.2015.8.6.20

Study on Validation Method of Simulation Model Combination

[1]Zheng Kun, [2] and Zhang Jianyan

[1,2] *School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang City, China*
E-mail: zheng_kun18@126.com

Abstract

Aimed at improving the efficiency of the development of complex simulation system and the problem of lower development costs, a validation development framework is proposed based on combination model. The system development process can be broken in the discovery of decomposition for the model, combination and combination of validation phase. This thesis gives the validation method based on model of tag transfer system behavior, judging by strong simulation and semantic similarity relations after the combination model of dynamic behavior and request model conforms to the degree. Simulation model combined the experimental results show that the method can automatically finish the combination of the simulation model validation, the validity of the model validation can meet the requirements of practical application.

Keywords: *simulation model, behavior verification, semantic similarity, data type validation*

1. Introduction

With the development of simulation technology, simulation technology has been widely used in the national defense, economic, material, and chemical industry and other fields. Because the simulation is an activity based on model, the credibility of the model has become the focus of user model. In the simulation process, there are usually several simulation models to choose [1]. At this time, you need to verify their credibility and choose the believers.

It is a common simulation model validation method through comparative analysis of the characteristics of the difference between the simulation output and the reference output. Naylor pointed out that according to intuition to surface simulation model validation, from the typical event analysis, assumptions and other internal characteristics of different angles the consistency between the simulation model and real object. Thus he put forward a lot of validation method based on the characteristics, such as analysis of variance and some nonparametric test, including chi square test, Kolmogorov Smirnov has inspection, *etc.*. Kheir used Theil coefficient of inequality of missile simulation system effectiveness which is verified. This paper analyzed the characteristics of grey correlation model, and applied to missile simulation system reliability evaluation work. Some researchers used error analysis method evaluating the credibility of the econometric model [2]. Some researchers analyze the connection between the credibility and the similarity is discussed, and given the simulation model credibility quantification based on similarity method. Someone is using the classical Fourier method to estimate frequency spectrum, through the spectrum analysis to the evaluation of missile system simulation model credibility, frequency spectrum analysis method and other window spectrum, the maximum entropy spectrum and cross spectrum, *etc.* [3-5]. In addition, some researcher gradually formed a relatively unified the theory and method of system based on the research of the predecessors' work and summary [6].

ISSN: 2005-4270 IJDTA

According to the above problem, this paper presents a simulation model validation method based on feature differences. First of all, according to different types of data, their respective characteristics difference measurement model is given. Then, extract the few independent representative principal component based on principal component analysis from multiple related data characteristic differences. The multiple outputs for simulation model are divided into m class. Finally, determine whether the reference output belongs to one kind of Fisher's discriminate analysis based on the m class. If it is true, the class of all elements that are to be believed, the corresponding simulation model and the reference output corresponding simulation model is regarded as the most "recent" believers.

2. Model Development Framework

The description of the differences of between simulation output and the reference output is the basis of the simulation model validation and choose a job. The classical simulation model validation method the simulation output and the difference between the reference output from a single or a few aspects. For example, suppose test only focus on static data difference between the mean or variance; *TIC* coefficient method, error analysis method only pay attention to the distance between the graded data difference; Grey correlation coefficient method and the similarity method only focus on differences between graded data appearance; Spectrum analysis focuses on the difference between someone who comported himself data item smooth spectral density [7]. This section according to the data characteristics of the simulation output and the reference output can be divided into static, graded and someone who comported himself three kinds of data, and then the characteristics of each type of data difference measurement model is given respectively, and fully describe the difference between the simulation output and the reference output, as shown in Figure1.

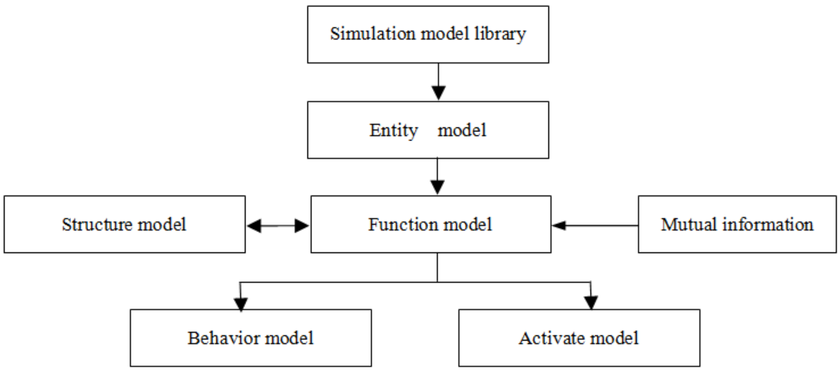

Figure 1. Simulation Model Development Framework

Simulation model library is a hierarchical database with index values, which contains basic type and reuses type two kinds of simulation model. Reuse can be made up by basic type or as a whole to join. After the simulation model library receives the request, first look for reuse layer, if there is the reuse model of the same structure, will return and the model over; If not found, it will need to request simple model is decomposed into a collection, according to the set in the simulation model library search, search process still layered implementation [8]. When the request of elements in the collection in the model library query to the results, the results in the collection to the model elements are combined, the combination of rules according to the simulation model of input and output data and receive relationship, and to request the model formulated on the basis of the final implementation.

2.1. The Key Variables and Indicators of Statistical Analysis Methods

Statistical method is the most commonly used techniques, the test data and simulation data are applicable. Statistical characteristics of the method are used to extract data. Statistical analysis methods include hypothesis test and confidence interval method, *etc.*

2.2. The Analysis of Continuous Dynamic Process Consistency

This method is used to analyze the degree of consistency of simulation and test data. The correlation coefficient method can be used to check the simulation and test data. The linear relationship between the correlation coefficient is defined as

$$s = \frac{n\left(\sum_{i=1}^{n} x_i \bullet x_i'\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i'\right)}{\sqrt{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2} \bullet \sqrt{n(\sum_{i=1}^{n} x_i'^2) - (\sum_{i=1}^{n} x_i')^2}} \tag{1}$$

BC inequality method can be used to check the fitting degree between the simulation and test data. *BC* is defined as

$$BC = \frac{\sqrt{\sum_{i=1}^{n} (x_i - x_i')^2}}{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i'} \tag{2}$$

Maximum error and the error integral error norm can also be used to analyze continuous dynamic process consistency.

2.3. Discrete Event Validation

This method is used to determine the effectiveness of discrete events. In generally, the discrete event validation can be realized by analyzing the conditions, the probability, the relationship between the influence and its behavior. Discrete event happened often with random uncertainty, such as the distribution of probability, arrival time, changes of state variables caused by events, *etc.* Classical statistical analysis method is often used for the validation of random uncertainty.

2.4. Mixed Behavior

This method is used to validate specific hybrid dynamic continuous dynamics, discrete events and their relationships. In the process of event tree sequence analysis, logical relationship and behavior can be used in the validation work.

Validation data type is used to ensure consistent portfolio model with the request data type. Model has a "black box" structure, interface is the only way to model the interaction with the outside world, and so model combination of data types match is authentication interface to set the data type of the match. Combined model could eventually be has the desired functionality of the target system only through the validation data type.

3. Dynamic Behavior Model Validation

3.1. Simulation Model Matching

Set M to present combined simulation model of interface, $M=\{m_1, m_2,... ,m_x\}$. X is the number of interfaces in *A*, $m_1, m_2,... , m_x$ present x the data type of the interface respectively in the M. Request for N said model interfaces, $N=\{n_1, n_2,... , n_y\}$, *y* is for the interface number *N*. $n_1, n_2,... , n_y$ *y* the data type of the interface respectively in *N*. Simulation model matching of data types can be divided into precise, surplus and missing type 3 kinds. For a given initial conditions and a set of input values, a simulation model of behavior should request and the behavior of the model is very close.

3.2. Method Validation Process

Behavior, the behavior of the simulation model is united said, and according to the simulation model of the average execution time and the number of missions. Behavior combinations for a set of meet the execution time of solution according to the constraints on the model execution time. Transferring behavior represented as a tag System (Labeled Transition System), calculates the time is worth to the execution of the simulation model of the behavior of the process sequence and is expressed as the *LTS*. Need to be validated, a simulation model validation process is divided into two steps; compare the combination simulation model with the request of the LTS. If there is strong equivalence relation, the combined simulation model is effective; if there is no strong equivalence relation, and then enters the step 2. This paper defines a weakening of the semantic similarity relation with parametric epsilon *V* epsilon. This method has high credibility because the validation method based on the simulation time and the semantics of these two important factors in the model. As shown in Figure2, the first three steps are independent on the combined simulation model and request model execution, step 4 for the combination of simulation model with the request to participate.

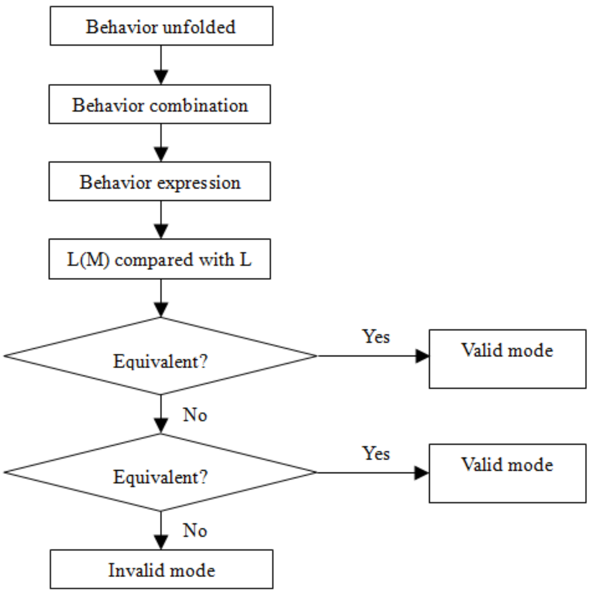

Figure 2. Combination Model Validation Process

4. The Simulation Model Representation

The difference between the simulation output and the reference output characterization is the basis of the simulation model validation and choose a job. Classic simulation model validation method describes the differences only from a single or a few aspects between simulation output and the reference output. For example: suppose test only focus on static data difference between the mean or variance; *TIC* coefficient method, error analysis method only pay attention to the distance between the graded data difference; Grey correlation coefficient method and the similarity method only focus on differences between graded data appearance; Spectrum analysis focuses on the difference between someone who comported himself data item smooth spectral density. In this section, the simulation output and the reference output can be divided into static, graded and someone who comported himself three kinds of data based on the data characteristics, and then characteristics of each type of data difference measurement model is given respectively, and fully describe the difference between the simulation output and the reference output.

In some special simulation application, we need to be focused on certain physical quantities, such as: control system response of the maximum overshoot and rise time, steady-state error and change tendency of the location data, the noise of the measurement data, *etc.* At this time, we need to put these parameters extracted from the data on the basis of the specific domain knowledge, in order to measure differences between simulation output and the reference output; And then, according to their data types, use below method of depicting their differences; Finally, with the original simulation data with the difference between the reference data structure characteristic difference vector data can be obtained to support the simulation model validation and choose a job.

The system output data type is different, because of the characteristics are also different. Therefore, in depicting the difference between the simulation output and the reference output before, need to categorize them. Assume that as a random variable X, MX for its corresponding sample space, $X(i)$, i=1, 2,…, N is the i time observe value of X. X(i), i=1, 2,..., N constitute a collection of {X(1), X(2),…, $X(N)$} is called a static data; If $X(i)$, i=1, 2,..., N for in observed data, in order to $X(i)$, i=1, 2,..., N of time sequence X(1), X(2),…, $X(N)$ called dynamic data. Further, assume that dynamic data X(1), X(2),…, $X(N)$ corresponding to the time sequence of T(1), T(2),…, $T(N)$, define the X(1), X(2),…, $X(N)$ along with the change of time frequency is as follows:

$$F = \frac{\sum_{i=1}^{N-1} \frac{\Delta X(i)}{\Delta T(i)}}{\overline{X}} \tag{3}$$

Among them, the F≥0 is data change frequency of X(1), X(2), ..., X(N); ΔX(i)=$X(i+1)-X(i)$; $\Delta T(i)=T(i+1)-T(i)$; $X=Ni=|X(i)|/N=0$. Thus, assumed $F_0 \geq 0$ to be the evaluation data changes how the critical value. If $F \geq F_0$, argues that X(1), X(2),…, $X(N)$ someone who comported himself as data; Otherwise, think X(1), X(2) ,…, $X(N)$ are the graded data.

4.1. Static Data Characteristic Difference Model

Assume the reference output X_r and simulation output X_s to be the static data, respectively by the corresponding system run N times and M times after the output of the collection.

$$X_r=\{x_r(1), x_r(2),\ldots, x_r(\mathrm{N})\} \tag{4}$$

$$X_s=\{x_s(1), x_s(2),\ldots, x_s(\mathrm{M})\} \tag{5}$$

Among them, X_r(i), i=1, 2,…, N is the output in the i times run of the reference system; $X_s(j)$, j=1, 2,…, M is the output in the j times run of the simulation system. For static data, the sample data of concentration and dispersion is the main trend of two characteristics, respectively is described with the mean and variance. As a result, get X_s relative concentration and dispersion characteristics of X_r differences are as follows:

$$m = \left| x_s - x_r \right| = \left| \frac{\sum_{i=1}^{e} x_s(i)}{E} - \frac{\sum_{i=1}^{f} x_r(i)}{F} \right| \tag{6}$$

$$n = \left| s_s - s_r \right| = \left| \sqrt{\frac{\sum_{i=1}^{E} \left[x_s(i) - x_s \right]^2}{E-1}} - \sqrt{\frac{\sum_{i=1}^{F} \left[x_r(i) - x_r \right]^2}{F-1}} \right| \tag{7}$$

Among them, m is present for the central tendency characteristic difference; n is present for the discrete trend features.

4.2. Graded Data Characteristic Difference Measurement Model

Assume that the reference output X_r and simulation output X_s as graded data, respectively to get series by the corresponding N and M times for system operation.

$$X_{ri} = \left[x_{ri}(1), x_{ri}(2), ..., x_{ri}(N) \right] \tag{8}$$

$$X_{sj} = \left[x_{sj}(1), x_{sj}(2), ..., x_{sj}(M) \right] \tag{9}$$

Among them, i=1, 2,..., N, j=1, 2,…, M.

To run multiple times at each sampling point average output processing, get the reference output and simulation output is as follows:

$$X_r = \left[x_r(1), x_r(2), ..., x_r(N) \right] \tag{10}$$

$$X_s = \left[x_s(1), x_s(2), ..., x_s(M) \right] \tag{11}$$

In the further clarification, X_r and X_s sampling time may be inconsistent, number of sampling points may be different too. At this point, we need to select one of the two as a benchmark, adopt the method of interpolation consistency processing for another time. In this paper, the processing of data assumes are after dealing with the time consistency of the data.

5. Combination Model's Inspection

5.1. Improvement of Periodogram Estimation

Spectrum estimation on random sequence segmentation is improved by reducing the variance of periodogram spectrum estimation. However, random sequence length limited the amount of separable segment, which can be increased by piecewise sequence there is overlap between segments. This method can increase the variance of each paragraph, so deal with need of the section and balanced between overlap rates. Another method is Welch method to estimate the periodogram improved, which improved periodogram estimation is based on sampling sequence segments and on each segment of the sequence using a rectangular window method. Due to the rectangular window edge value close to zero, the block sequence using a rectangular window can reduce the piecewise dependence on overlap rate. According to the need to choose the appropriate window function, choose the overlap rate 50%, can greatly improve the performance of spectrum estimation.

$$x(n) = x(n + iL - L) \tag{12}$$

Each sequence contains a sample, *L* calculation cycle Figurebefore the data window *w(n)* and the multiplication, the data get *X* modified periodogram.

$$M_L = \frac{1}{L}\left|\sum_{n=0}^{L-1} x(n)w(n)e^{\omega n}\right|^2, i = 1,2,...., X \tag{13}$$

W is present the average energy of window function data.

$$W = \frac{1}{L}\sum_{n=0}^{L} w^2(n) \tag{14}$$

Use of window spectrum estimation of power spectrum estimation, Welch method is used, the data segment in a cover section, the section have overlap, the segments after the data using the rectangular window to add data window handle. Using this method not only can reduce the frequency leakage, and can reduce the estimation error, and to a certain extent to ensure the resolution.

Spectrum analysis is mainly analysis of transformation of time series to the frequency domain. The simulation model is valid conclusions by comparing the power spectrum density of consistency. Next is to compare whether the agreement of simulation time series of power spectral density and reference the statistical features of time series of power spectral density, and the degree of consistency. Therefore, we need compatibility test on two groups of spectrum estimation.

5.2. The Test Method of Wavelet

Description of the signal time domain and frequency domain is often used to two forms. Also, on most of model output sequence from the aspects of time domain or frequency domain model validation, the model output sequence is verified. But in real life and the engineering practice, such as dynamic output process of weapon system, many output for non-stationary signal. In order to the non-stationary signal model validation work, must first to smooth processing of signals and information loss of original data. Wavelet transform can detect signal contained in the trend, mutation and other characteristics, has the ability of local analysis and refined.

The nature of wavelet transform and Fourier transform. The signal is with a set of wavelet basis function is a linear combination of the said come out, the difference is selected wavelet basis function is different, is produced by the basic wavelet, wavelet basis function basic wavelet, also known as the mother wavelet, there is a concentration of power in time domain is very has the nature of special function, it meet the conditions of mathematical integral is zero.

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{15}$$

Its spectrum satisfies the following conditions:

$$W_{\psi} = \int_{-\infty}^{\infty} \frac{\psi(s)^2}{S} ds > 0 \qquad (16)$$

The basic wavelet in the frequency domain also has good damping properties. Wavelet basis function is passed by the mother wavelet scale and displacement of W_{ψ} , composed of a set of standard orthogonal basis.

6. Validation Method Instance

The simulation model validation process is usually tedious and done by hand. It requires the presence of system experts, especially when it used in key scenes, such as military combat. Validating the model is crucial. This article takes a tank repair model as an example, and describes the validation method in detail.

Simulation model and the corresponding real system under the condition of the same input output can be seen as two random processes, respectively on two random processes, can get two sets of data curve, correlation analysis to get correlation coefficient. Correlation coefficient is close to 1, shows that the higher the degree of two random process related, then the higher the credibility of the simulation model or system, when the correlation coefficient is 1, the output of the simulation system and the output of the real system are exactly the same.

The following for a certain missile over a period of high output for example. Respectively under the condition of the same initial input of the simulation model and the output curve of three groups of the height of the real system. Respectively the simulation system output and the output of the real system is i=1, 2, 3. *A* representative simulation model of the height data represents the height of the real system data, which is shown in Figure3.

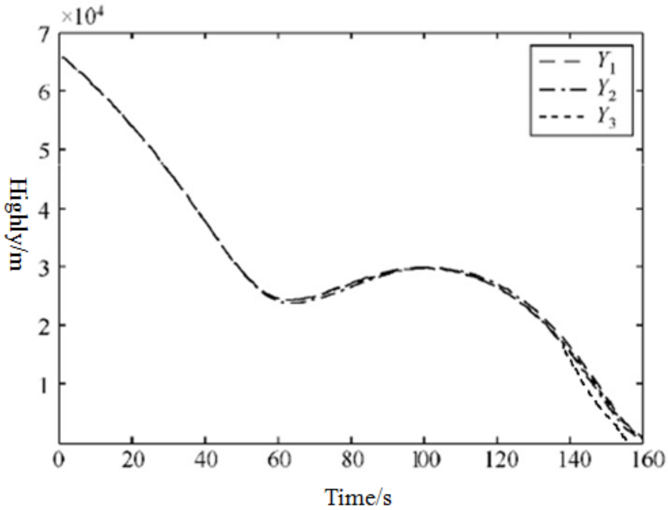

Figure 3. The Simulation Data Plot

Then, the simulation data and real data are fitting for the functional form. Through many experiments, selecting nine fitting polynomial fitting method, the error in the acceptable range. In the process of synthesize data fitting function, does not require the real data and simulation data meet the requirements of time sequence consistency, namely don't simulation data and real data meet the requirement of the same starting point and sampling interval and so on. After the method of calculation are all based on the function after fitting, so using the method of model validation does not require the data to meet the time sequence consistency. This is where the method compared with other model validation method is a different place.

7. Conclusion

In this paper, we research on the verification method of the simulation model combination. We proposed a verification method of simulation model combination. It verifies the effectiveness of the combination model from data type and behavior aspect. Comparing the relationship of the two LTS verifies the effectiveness of the combination model.

While it is acceptable to assume their existence, how they are obtained is still an open question. Reliability verification method is proposed in this paper, a simulation model based on function fitting for often appeared in the process of weapon development of real system data and simulation data do not match the sampling interval, the starting point is not consistent, and so on and so forth. This paper introduces the principle of this method in detail, and introduces the application of the method for model validation steps. This method avoids the traditional simulation model validation method must require real data and simulation data to meet the requirements of time sequence consistency, and is verified by a missile model instance analyses the credibility of high output data. Analysis shows that the method can get the real data and simulation data does not meet the time series under the condition of complete consistency model validation.

A valid model is defined as close enough with respect to the states, sequence and duration of component execution, to a request model. While it is acceptable to assume their existence, how they are obtained is still an open question.

References

[1] Kavussanos M. G. and Alizadeh A. H., “Econ. Mode”, vol. 5, **(2002)**, pp. 747-782.
[2] Sutherland L. M., Middleton P. F. and Anthony A. J., “Anna. of Surg.”, vol. 3, **(2006)**, pp. 291-300.
[3] W. Tianmiao, Z. Dapeng and L. Da, “Chin. J. of Med. Inst.”, vol. 4, **(2008)**, pp. 235-238.
[4] L. Fei, M. Ping, Y. Ming and W. Z. Cai, “J. of Harb. Ins. of Tech.”, vol. 1, no. 1-3 **(2007)**.
[5] C. Z. Dong, Z. Y. Jin, F. Xiang and Z. Bin, “Acta. Auto. Sin.”, vol. 10, **(2010)**, pp. 1361-1370.
[6] W. Jing, W. X. Yan, C. Y. Xing and T. J. Chuan, “Sys. Eng. and Elec.”, vol. 8, **(2010)**, pp. 1677-1679.
[7] M. F. Yan, Y. Ming and W. Z. Cai, “Simu. Mode. Prac. and Theo”, vol. 5, **(2010)**, pp. 500-515.
[8] X. A. Liu, H. Xiao and W. Suping, “J. of Tact. Miss. Tech”, vol. 6, **(2010)**, pp. 107-112.