Cyber-Physical System Modeling Based on iGMDH Algorithm

Shengbin Ren, Fei Huang

School of Software, Central South University Changsha,China

Email:154711009@csu.edu.cn

Abstract—Aiming at the problem of error and error accumulation caused by data imprecision and uncertainty in CPS modeling , this paper proposes an iGMDH algorithm based on the GMDH and the idea of interval analysis. Firstly, the contractor is introduced to improve the SIVIA algorithm. The algorithm solves the problem of large amount of computation, long time consuming and deadlock in the SIVIA algorithm performs dichotomous search. The input and calculation of GMDH are transformed into interval number and interval operation, and the model parameters are estimated by using the improved SIVIA algorithm. Finally, the midpoint of the interval parameter is taken as the point estimation of the parameter to be estimated, and then the intermediate model is filtered by using the external criterion to establish the final system model. The experiment shows that the iGMDH algorithm can significantly improve the accuracy and noise immunity compared with the original algorithm, and effectively solve the problem of error and error accumulation in CPS modeling.

Keywords—Cyber-Physical System;Group Method of Data Handling; parameter estimation;interval analysis.

I. INTRODUCTION

Cyber-Physical system (CPS)[1,2] is a next-generation intelligent system which integrates computing system and physical system. It has realized the deep fusion and real-time interaction between the two systems through embedded system and network since it was proposed by the National Commission of the United States. The US President's Advisory Council on Science and Technology has listed CPS as one of the key research areas. A series of seminars on CPS have been held. Since its introduction in 2006, the development of CPS has been greatly promoted by many governments. Support and funding has become an important direction of academic and scientific research.Because discrete computational processes and continuous physical processes coexist in the CPS, the fusion characteristics the two can hardly be characterized by a single model. The temporal and spatial characteristics and dynamic non-determinism of CPS are also difficult to adapt to existing modeling languages. Therefore, how to establish a reliable CPS model becomes a bottleneck for the research and application of CPS.

On the study of CPS modeling, Lee and Edward A of Berkeley University proposed the concept of "information physics and physics of information systems" by summarizing the existing modeling methods of simulation control to solve the problem of physical entity and information entity modeling and their interaction[3]. The CPS system can be refined into computation entity and physical entity,by analyzing the feasibility of using UML language to build computation entity model and using simulink/RTW modeling tools to construct physical entity, the approach for integration of heterogeneous models based on UML framework are proposed[4]. Sun Z et al use AADL to extend the CPS modeling and gives a new method for compiling CPSADL. Taking the lunar vehicle self-propelled system as an example, the application of CPSADL is illustrated[5].Wang Li et al. put forward a modeling method of vehicular Cyber-Physical System with extended hybrid automata by analyzing the features of vehicular CPS software in depth Finally, taking vehicle speed control system as an example, the validity of the modeling method is illustrated[6].

Wang B and Baras J S have designed and implemented an integrated modeling and simulation tool chain of HybridSim ,which is used in the design and simulation of CPS.Taking the integrated water-cooled heating system model of intelligent building as an example, the convenience and effectiveness of HybridgeSim are demonstrated. The effects of packet loss rate and sampling rate on the system performance are also studied[7].The modeling method of CPS proposed above requires the modeler to master the running environment and the state of the system before establishing the model. And the events required for each state transition. However, with the increasing complexity of CPS , the environment is becoming more and more complex, these methods become difficult to adapt when building a system framework or analyzing a system.

Data-driven CPS modeling is another important method for modeling complex CPS. This method uses empirical data to describe the main attributes (self-similarity, non-stationary) of the CPS operation process, and establishes a model based on the general rules of the system characteristics found in the data.

Yang Fan et al proposed a method to build the CPS system model from data, extracted the characteristic value that can reflect system attributes from the obtained discrete data, and skillfully use GMDH algorithm to build the system model. The Deep integration of continuous physical events and discrete computing systems is realized[8].

As a heuristic self-organizing modeling method, the GMDH algorithm can construct the model according to the original information of input and output variables, and select the optimal model by using the outer criterion to realize the simulation of the internal structure of the research system.

GMDH algorithm is a heuristic self-organization construction. The model method can construct the model according to the original information of the input and output variables, and select the optimal model by using the external criterion to realize the simulation of the internal structure of the research system.Because of its self-organization and global optimization, it can effectively identify complex multivariable systems[9, 10, 11]. However, it has been proved in practice that the GMDH algorithm makes the data inaccurate and uncertain because of the limitation of computer word length and the existence of random noise in the process of model parameter estimation. The error caused by its calculation will be further generated on the basis of the original calculation when participating in the operation again.Especially in the process of establishing a layered model, the effect of error accumulation is enlarged and continuously expanded, and eventually, invalid results may be obtained.

Interval analysis was proposed by R.E. Moore et al in the 60s of the last century, in order to study the error caused by computer floating-point operation. Interval analysis can realize data storage and operation in the form of interval in the form of interval. In addition, some uncertainties can be expressed as interval, which can be included directly in interval algorithm, and its feasible set can be well represented. Interval analysis is considered as a powerful tool for system parameter estimation[12].

Based on the GMDH algorithm and the idea of interval analysis, this paper regards the parameter estimation problem of the model as a set of inversion problems, uses the improved SIVIA algorithm to estimate the model parameters, and then proposes a new modeling method of interval GMDH (iGMDH)algorithm.

The remainder of this paper is structured as follows: We describe our research models in Section 2. In Sections 3, we provide the experiment design, results and validation. Then at last in Section 4, we present conclusions.

II. The Establishment and Improvement of the Algorithm

As shown in Fig 1, the iGMDH algorithm proposed in this paper transforms the input of the GMDH algorithm into interval numbers ,and estimates the parameters of the model by using improved SIVIA algorithm.After the point estimation of the interval parameters is further calculated, the intermediate model is screened by the external criteria. Thus, the optimal CPS system model including interval parameters is established.

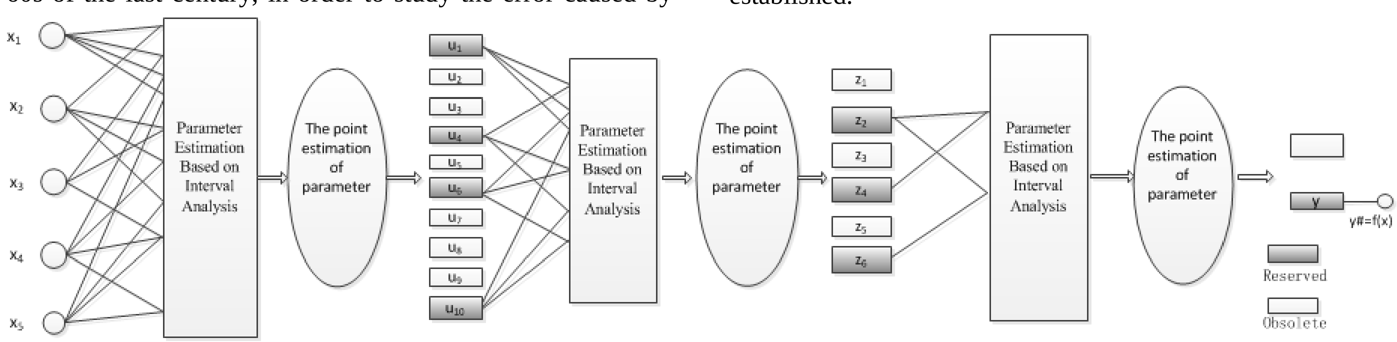

Fig.1 The modeling process of iGMDH algorithm

The input variables $x_1, x_2 \cdots x_5$ are five possible input variables of the system. After the input variables are combined, the estimated set of the parameters to be estimated is obtained by using the SIVIA algorithm, and the intermediate model $z_1, z_2 \cdots z_6$ of the first layer is generated. The center of the estimated set is calculated. As the point estimation of parameter estimation, the input variables of the second layer are screened out by the external criterion of GMDH.Repeat the above process, the model can produce intermediate model $z_1, z_2 \cdots z_6$ of the second layer. When the external criteria reaches optimal, the model is stopped and the optimal model $y^\#$ is obtained.

A. *SIVIA algorithm and its improvement*

(1)SIVIA algorithm

As a classical algorithm of interval analysis in parameter estimation of nonlinear systems, SIVIA algorithm can obtain all the global optimal solutions of parameter values within a given precision. According to the prior knowledge, the errors between the data and the output of the corresponding model are included in the known feasible domain, according to the prior knowledge, the error between the data and the output of the corresponding model is included in the given feasible domain. Ishii D et al proposed an interval method that verifies the properties described by a bounded signal temporal logic. This algorithm performs a forward simulation of a continuous-time dynamical system, detects a set of time intervals in which the atomic propositions hold, and validates the property by propagating the time intervals[18].

Liu B X proposed a bounded error estimation method based on interval analysis to identify the viscous damping coefficient and Coulomb friction coefficient of rod pumping system in directional wells. Compared with traditional estimation methods, this method is global, it bypasses the problem of initialization. Moreover, the method avoids amount error due to transformation of measurement output data[19].

Wang Li et al proposed a method to estimate the position of robot based on the real-time data of robot on-board sensors, environment sensors and other robots.This method treats the nonlinear boundary error estimation problem as an inversion set, to deal with outliers of specific types and model errors in the imprecise environment. The location map is obtained and the location problem of mobile robots is solved[20].

Let the system model containing unknown parameters p is $f(p), p \in P$, P is a priori Search set for the parameters .The actual observation data of the system is $y(x) \in R^k$, the theoretical output of the model is $y_m(p,x) \in R^k$, and the model output error $e(p,x) = y - y_m(p,x)$, then the $e(p,x)$ must belong to some known feasible domain $E = [\underline{e}, \overline{e}]$, $\overline{e}$ and $\underline{e}$ are the lower bound and upper bound of the known acceptable output error,respectively. if (x_i, y_i) and E are known ,then the prior feasible set for the model outputs can be expressed as $y_m(x_i) \in Y = [y - \overline{e_i}, y - \underline{e_i}]$. if and only the parameter estimation value $\hat{p}$ satisfies $y_m(x_i, \hat{p}) \subseteq Y$, then the parameter value is called a feasible value to meet the error requirement.Define the set S of all feasible values of p, as shown by formula (1).

$$S = \{p \in R^n \mid y(x_i) - y_m(p, x_i) \in [\underline{e}, \overline{e}]\} = \bigcap_{i=1}^{k} \{p \mid y_m(x_i, p) \in Y\} = \bigcap_{i=1}^{k} S_i \tag{1}$$

In formula (1), $y(x_i)$ represents the observed values of input x_i , $y_m(p, x_i)$ is the theoretical output of input x_i . It can be seen from the formula that with the increase of the sample size, the range of S will be gradually reduced, the length of the parameter set S will be smaller,when the sample size is large enough, the feasible value of the parameter to be estimated is closer to the real parameter of the system model.Because f is a multiparameter nonlinear function, the S expressed by the formula (1) can not be calculated accurately.Therefore, the inverse function f^{-1} of f is transformed into the set inversion problem of formula(2) to solve the problem.

$$S = f^{-1}(Y) \cap P \tag{2}$$

From the above,by using the SIVIA algorithm of interval analysis, we can always obtain two regular subpavings $\underline{S}, \overline{S}$ such that the formula (3) holds.

$$\underline{S} \subset S \subset \overline{S} \tag{3}$$

As the formula (3) shows, the unknown solution set S is contained in two known sets $\underline{S}$ and $\overline{S}$, that is, by finding the set $\underline{S}$ and $\overline{S}$, we can obtain the globally optimal solution set S.

(2) improved SIVIA algorithm CSIVIA

In the process of parameter estimation, the SIVIA algorithm uses bipartite to search priori search sets recursively, which makes the computation of the algorithm increase exponentially, thus consuming a lot of time and memory. In this paper, we first improve the algorithm and propose the CSIVIA algorithm. By introducing the contractor, the search domain of the estimated parameters can be compressed under the condition that the solution set remains invariant. Thus reducing the computational time complexity and space complexity (the number of boxes in the regular subpavings, mention High algorithm speed and efficiency is.

An contractor C is any of the contractors that can be used to contract the search domain of the algorithm,it can replace $[P]$ with a smaller prior search domain $[P]'$ under the condition that the solution set remains unchanged, that is, $S \in [P]' \in [P]$ .Thus to reduce its computational complexity and its computation time by reducing the number of boxes to be bisected. Figure 2 describes the improved CSIVIA algorithm by introducing the Forward-Backward contractor Ctc[21,22,23]. the 1 to 6 steps in the algorithm is to contract the prior search domain by using contractor and the fixed point fp to obtain the minimum value containing the true value of the parameter to be estimated.The 4 to the 10 step in the algorithm is to compress the prior search domain by using contractor and fixed point fp, and obtain the minimum $[p]'$ that contains he true value of the estimated parameters, so as to improve the computation speed of SIVIA algorithm. Steps 11 to 21 are test judgment according to the simulation output set $[y]_m([p]')$ .If $[y]_m([p]')$ is completely within Y, then $[p]'$ is called the definite solution, and stored in $\underline{S}, \overline{S}$. In this case, it needs to judge according to its width $w([p]')$, if $w([p]')$ is larger than the given tolerance parameter ε_0 ,then the binaries of $[p]'$ and the resulting subinterval are tested again. If the width $w([p]')$ is less than ε_0, it is used as the external approximate solution of the solution set $\overline{S}$. After the finite recursion, two sets $\underline{S}$ and $\overline{S}$ containing the solution set S can be obtained, and then the solution set S can be approximated.

Algorithm 1:improved SIVIA algorithm CSIVIA

Input: actual observational value y ; prior feasible set: Y ; prior search domain: $[P]$; tolerance parameter: ε_0 , fixed-point *fp*

Output: regular subpavings: $\underline{S}, \overline{S}$

```
1: S̲ ← S̄ ← φ                          //initialization
2: L ← {[P]}
3: while L ≠ φ do
4:   [P]' ← [−∞,+∞]^n                   //use forward-backward contractor Ctc
5:   while contraction([P],[P]') < fp do
6:     [P]' ← [P]
7:     Y ← Y ∩ [y]_m([P]')              //forward propagation
8:     [P]' ← [P] ∩ [y^{-1}]_m(Y)       //Backward propagation
9:   End while
10:  Return [P]'
11:  if [P]' = φ then                  //[P]' is a non-solution
12:    Continue;
13:  Else if ([y]_m([P]') ⊂ Y )then    //[P]' is a definite solution
14:    S̲, S̄ ← [P]'.
15:  Else if w([P]') < ε_0 then         //[P]' an external approximate solution.
16:    S̄ ← [P]'.
17:  Else                               //[P]' may contain a partial solution set, then it is dichotomized.
18 :     {L[P]', R[P]'} = Bisect([P]')  // bisects each box in [P]'
19:  End if
20: End while
21: Return S̲, S̄
```

Fig.2 Improved SIVIA algorithm CSIVIA

Algorithm 2:iGMDH

Input: sample data set: N ; actual observational value y ; prior error set E ; prior search field: $[P]$; tolerance parameter: ε_0 , fixed-point *fp*

Output:optimal complex model $y^{\#}$

1:the sample data set N be divided into training data set N_t and test data set N_c .

2:The general relation function of input variable and output variable is established by $K-G$ polynomial function.

3: The combination of two input variable are generated for C_m^2 intermediate model.

4:On the training data set N_t ,obtain the parameters estimated set S of each intermediate models by CSIVIA algorithm.

5: Find the point estimate q of interval estimation by $q = 1/2w(S)$

6:On the test data set N_c , the external standard value R_j is calculated according to the q and the experience value L is selected to judge.

7: if $R_j \le L$:

8: Select the corresponding intermediate model as input variable to the new layer and mark the smallest R_j as $R_{\min}$

9: While new layer $R_{\min}$ < top layer $R_{\min}$ do :

10: $\{3,4,5,6,7,8\}$

11: End while

12: End if

13:when the external criteria becomes optimal,the iterative process is stop and optimal complex model $y^{\#}$ is obtained.

Fig.3 iGMDH Algorithm

Table 1 Properties of experimental data sets

symboling	Normalized-losses	Wheel-base	length	width	height
-3,-2,-1,0,1,2,3	From 65 to 256	From 86.6 to 120.9	From 141.1 to 208.1	From 60.3 to 72.3	From 47.8 to 72.3
Curb-weight	Engine-size	bore	stroke	Compression-ratio	horsepower
From 1488 to 4066	From 61 to 326	From 2.54 to 3.94	From 2.07 to 4.17	From 7 to 23	From 48 to 288
Peak-rpm	City-mpg	Highway-mpg	price		
From 4150 to 6600	From 13 to 49	From 16 to 54	From 5118 to 45400		

B. *iGMDH algorithm*

Under the assumption of UBB(unknow-but-bounded error), the iGMDH algorithm only needs to know the range of parameters to be estimated and does not need other artificial assumptions when estimating parameters of the system model, which to some extent avoids the influence of human subjective factors on the result of parameter estimation. Compared with GMDH algorithm, the input and model calculation of iGMDH algorithm can be transformed into interval number and interval operation,the problem of system model parameters estimation is viewed as a set inversion, and the CSIVIA algorithm is used to obtain the approximate but reliable estimation set of the parameters, thus solving the problem of error and error accumulation caused by the floating point operation. After further calculation, the point estimation of the parameters to be estimated can also be obtained. Moreover,

The estimation of parameters can be carried out even with only a small amount of input data, And as the amount of data increases, the more information is provided, the higher the accuracy of the estimation is.

If the system has m inputs $x_1, x_2, \cdots, x_m$, y is the output of the system, A given set of input values must have a y value corresponding to it. Suppose there are N group x values and corresponding output y, then the process of establishing the system model by the iGMDH algorithm is shown in Figure 3.

Figure 3 is the pseudo code for the iGMDH algorithm, the first step of the algorithm is to process the data set. By extracting the characteristic attribute m that reflects the essence of the system, the sample set N of the input variable and output variable of the system is determined, and it is divided into training set N_t and test set N_c, where $N_t \geq N_c$, and $N_t \cup N_c = N, N_t \cap N_c = \phi$. The $K-G$ polynomial $y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_i x_j$ is constructed by 2 ~ 3 steps, and the general function relation of input variable and output variable is established. According to the relation function, combine the input variable in pairs generating for C_m^2 intermediate model. By using 4~6 steps to estimate the parameters of each intermediate model, the estimated values of the parameters to be estimated are obtained. After the solution set of each set of data is obtained, the estimated set S of parameters to be estimated is obtained by means of intersection. The center of S is calculated as the estimated value q of parameters to be estimated. On the test set N_c, 7~14 calculate the outer criterion of the model, the root mean square value R_j (such as formula 5), The intermediate model $\{u_i\}$ of the first layer is screened, and the better intermediate model u_j ($j \leq i$) is selected as the input of the next layer to generate the new layer of $R_{\min}$. Repeat the above steps to produce the intermediate model of the second layer and the third layer in turn . If the $R_{\min}$ of the new layer is no longer reduced, the modeling stops and the final model of the system is obtained (optimal complexity model).

$$R_j = \left[\frac{\sum_{i=nt+1}^{n} (y_i - z_{ij})^2}{\sum_{i=nt+1}^{n} y_i^2} \right]^{\frac{1}{2}}, j = 1,2,\cdots k \tag{5}$$

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. *Experimental Data*

The experimental data set is selected from the literature[12]. The data set was measured by UCI (University of California, Irvine) under the actual operating environment. Table 1 shows some relevant attributes of the data set.

The data set described in Table 1 contains the auto-reduction values of the car under various characteristic conditions, the risk level, and the normalized loss values of the specified risk levels compared with other cars. The second type of data is used to reflect the safety performance of the car. If the car test shows that the risk level increases (decrease), it means that its corresponding safety performance is reduced (increased). Therefore, +3 represents the lowest safety performance of the car, and -3 presents the highest safety factor.

In fact,the original data set contains 26 attributes, but because 10 of these attributes have a low impact on the analysis object (safety factor), they are negligible. In the experimental process, the system's continuous model is constructed by using the risk level as an output variable and the other 15 attributes as input variables.Because the metric forms of the discrete data matrix are different, it is necessary to standardize the value of each sample data on the attributes first. Therefore, all the experiments performed in this paper are carried out after the data is normalized.

B. Parameter estimation performance test

Taking the Car “normalized-losses” as the output variable, “wheel-base” and “horsepower” as the input variables, the system model is established as an example. When the tolerance parameter ε_0 is 0.01 and the fixed point fp is 0.001, the parameters of the model are estimated using the CSIVIA algorithm and the SIVIA algorithm, respectively.

As shown in Table 2, the number of box produced by the two algorithms in the pavings of the binary block and the computation time are compared. It can be seen that the CSIVIA algorithm proposed in this paper is superior to the SIVIA algorithm in time and the number of boxes produced.

It can be seen that the CSIVIA algorithm proposed in this paper is superior to the SIVIA algorithm in both computation time and the number of boxes generated.

Table 2 Comparison of two algorithms

Algorithm	Computation time	Proccessed boxes
SIVIA	136S	212381
CSIVIA	105S	168175

C. CPS modeling experiment

The purpose of this experiment is to verify the feasibility of the iGMDH algorithm. Based on the experiment 4.1, "City-mpg", "highway-mpg" and "Curb-weight" are added as input variables to model system.Figure 4 shows the parameter estimation results of two pairs of variables selected in each iteration process of iGMDH algorithm.

Figure 4 is the parameter estimation of the iGMDH algorithm that selects two pairs of variables in each iteration. Since $\overline{S}$ contains an external approximate solution, we can't get accurate S. For the sake of convenience and insurance, the experiment selects set $\underline{S}$ as the approximate estimate of solution set S.

Since the priori search domain of the algorithm is compressed by a contractor, seven numerical constraints are allowed to be established at one time, so we obtain the final solution set $S(x_i, x_j)$ by taking the intersection after all $\underline{S}_n(x_i, x_j), (0 \le i \le M)$ is obtained.Take the midpoint of $S(x_i, x_j)$ as the point estimate of the parameters to be estimated, and select the optimal intermediate models as the input of the next layer according to the external criterion R_j. Here, u_1, u_2 represent the two input variables (x_1, x_2) and (x_2, x_{13}) in the second layer,the variables x_1, x_2 and x_{13} respectively represent the "Normalized-losses," "Wheel-base" and "Peak-rpm".

The system continuous model shown in Fig. 4 can be expressed by formula (6), where $u_{1,} u_2$ are Formula (7) and Formula (8), respectively.

$$y = [-0.03022, -0.02799] + [0.69604, 0.709892]u_1 + [0.260018, 0.27515]u_2 + [-1.61314, -1.49687]u_1^2 - [1.279510, 1.292515]u_2^2 + [0.31763, 0.335104]u_1 u_2 \tag{6}$$

$$u_1 = [-0.06513, -0.06457] + [0.38912, 0.390266]x_1 + [-0.37203, -0.36676]x_2 + [-0.31157, -0.31098]x_1^2 + [0.13986, 0.147363]x_2^2 + [-0.88703, -0.79971]x_1 x_2 \tag{7}$$

$$u_2 = [-0.50133, -0.47986] + [-0.68241, -0.647312]x_2 + [0.40137, 0.422950]x_{13} + [0.646594, 0.669873]x_2^2 + [0.679783, 0.690016]x_{13}^2 + [0.03976, 0.045105]x_2 x_{13} \tag{8}$$

D. Modeling Error Experiment

To corroborate the effectiveness of the iGMDH Algorithm，two groups of comparative experiments have been done.One is the experiment without random noise, the other is the experiment with random noise.used to illustrate the comparison between minimum root mean square (RMS) and model error of iGMDH algorithm and GMDH algorithm when modeling data is uncertain and data uncertainty.when minimum root mean square and model error are smaller,the predicting accuracy is higher.

	$\underline{S}_1(x_1, x_2)$	$\underline{S}_2(x_1, x_2)$	$\underline{S}_3(x_1, x_2)$	$S(x_1, x_2)$
A	[-0.06709,-0.06369]	[-0.065851,-0.0640]	[-0.06457,-0.06513]	[-0.06513,-0.06457]
B	[0.38601,0.399658]	[0.38578,0.390266]	[0.38912,0.40059]	[0.38912,0.390266]
C	[-0.37214,-0.36653]	[-0.37302,-0.36676]	[-0.37203,-0.36672]	[-0.37203,-0.36676]
D	[-0.31157,-0.31026]	[-0.31168,-0.30965]	[-0.31253,-0.31098]	[-0.31157,-0.31098]
E	[0.12548,0.147893]	[0.12721,0.147363]	[0.13986,0.148494]	[0.13986,0.147363]
F	[-0.89897,-0.79971]	[-0.88703,-0.79224]	[-0.89033,-0.79220]	[-0.88703,-0.79971]

	$S_1(x_2, x_{13})$	$S_2(x_2, x_{13})$	$S_3(x_2, x_{13})$	$S(x_2, x_{13})$
A	[-0.50223,-0.46733]	[-0.51004,-0.46593]	[-0.50133,-0.47986]	[-0.50133,-0.47986]
B	[-0.69376,-0.64731]	[-0.69203,-0.64892]	[-0.68241,-0.64297]	[-0.68241,-0.647312]
C	[0.399453,0.42295]	[0.40032,0.435716]	[0.40137,0.429731]	[0.40137,0.422950]
D	[0.646594,0.67313]	[0.64319,0.669873]	[0.642989,0.673921]	[0.646594,0.669873]
E	[0.679783,0.69213]	[0.679325,0.69002]	[0.6795219,0.69256]	[0.679783,0.690016]
F	[0.03976,0.045105]	[0.03969,0.045873]	[0.039709,0.046063]	[0.03976,0.045105]

	$\underline{S}_1(u_1, u_2)$	$\underline{S}_2(u_1, u_2)$	$\underline{S}_3(u_1, u_2)$	$S(u_1, u_2)$
A	[-0.03022,-0.02754]	[-0.03054,-0.02747]	[-0.02973,-0.02799]	[-0.03022,-0.02799]
B	[0.69604,0.713121]	[0.69592,0.709892]	[0.694814,0.79795]	[0.69604,0.709892]
C	[0.259935,0.28025]	[0.260018,0.27515]	[0.259893,0.279715]	[0.260018,0.27515]
D	[-1.61579,-1.49322]	[-1.61314,-1.49687]	[-1.61989,-1.49721]	[-1.61314,-1.49687]
E	[1.278745,1.29372]	[1.279365,1.29298]	[1.279510,1.292514]	[1.279510,1.292515]
F	[0.31763,0.335104]	[0.317692,0.33547]	[0.317619,0.336063]	[0.31763,0.335104]

Fig.4 Estimation results of Model parameters

(1)Case without random noise

In this paper, the average relative error is chosen as the model error to measure the deviation between the predicted value of model and the true value. Fig.5 and Fig. 6 respectively given comparison the minimum mean square root value and model error of iGMDH algorithm with GMDH algorithm on the same data set.From Fig.5 and Fig.6, it is very clear that the proposed iGMDH algorithm not only outperforms the GMDH algorithm in terms of root mean square but also in terms of model error.

Because of the convergence of the algorithm itself, it can stop the modeling and obtain the optimal complexity model when the external criterion value reaches optimal. However, even if the error generated by each layer in the established GMDH polynomial neural network is small, it has a certain effect of error accumulation., and as the number of layers increases, the accumulation of errors increases. From Fig. 6, it is very clear that the error of iGMDH model and its accumulation are obviously smaller than those of GMDH model.so the model established by the proposed algorithm is more accurate and reliable than that built by GMDH algorithm.

Table 3 Comparison of parameters estimation results between two algorithms with or without noise in experimental data

	Without random noise		With random noise	
	iGMDH	GMDH	iGMDH	GMDH
A	[-0.030218,-0.027986]	-0.0288	[-0.035607,-0.025419]	-0.0145
B	[0.696037,0.7098921]	0.7052	[0.679346,0.72314715]	0.8164
C	[0.260018,0.275146]	0.2738	[0.2589316,0.289710]	-0.0221
D	[-1.613138,-1.496873]	-1.5375	[-1.654271,-1.4867359]	-0.6956
E	[1.279510,1.2925138]	1.2859	[1.27698465,1.3109412]	0.9670
F	[0.317692,0.335104]	0.3248	[0.3068923,0.3449271]	0.3078

(2) cases with random noise

In this sub-section,we added normal random noise with a standard deviation of 0.1 and a mean of 0 to the experimental data set.In order to ensure that all the errors of parameter estimation are only derived from the sample data and the parameter estimation method itself, is convenient to study the effect of random noise on the model parameter estimation results.The experiment takes the estimated value without random noise as the "true value" and the estimated value with random noise as the "predicted value."

IV. CONCLUSION

In this paper ,a novel method of CPS modeling through data-driven is introduced.the proposed method.The method transforms the input and calculation of GMDH algorithm into interval number and interval calculation, and uses the improved SIVIA algorithm to estimate the parameters of the model.After further calculating the point estimate of interval parameters,the intermediate models generated are filtered according to the external criteria,and the optimal complexity model with interval parameters is established. In order to verify the effectiveness of the algorithm, the comparison experiments of multiple groups between the algorithm and the GMDH algorithm are carried out in this paper.the results shows that the proposed iGMDH algorithm outperformed the GMDH algorithm.

REFERENCES

[1] Rajkumar R, Lee I, Sha L, et al. Cyber-Physical Systems: The Next Computing Revolution[J]. 2010, 14(6):731-736.

[2] Li Ren-Fa, Xie Yong, Li Rui, et al. Survey of Cyber-Physical Systems[J]. Journal of Computer Research and Development, 2012, 49(6):1149-1161.

[3] Lee E A. Cyber Physical Systems: Design Challenges[C]// IEEE Symposium on Object Oriented Real-Time Distributed Computing. IEEE Computer Society, 2008:363-369.

[4] Liu Sha, Wang Yu-ying, Zhou Xing-She,et al. Research and Design for the Modeling of Simulation of CPS[J]. Computer Science, 2012, 39(7):32-35.

[5] Sun Z, Zhou X. Extending and Recompiling AADL for CPS Modeling[C]//Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE, 2013: 1225-1230.

[6] Ye-jing L, Ming-cai C, Guang-quan Z, et al. A Model for Vehicular Cyber-Physical System Based on Extended Hybrid Automaton[C]//Computer Science & Education (ICCSE), 2013 8th International Conference on. IEEE, 2013: 1305-1308.

[7] Wang B, Baras J S. Hybridsim: A modeling and co-simulation toolchain for cyber-physical systems[C]//Distributed Simulation and Real Time Applications (DS-RT), 2013 IEEE/ACM 17th International Symposium on. IEEE, 2013: 33-40.

[8] Yang Fan,Liu Yan,Li Ren-Fa,et al. A Modeling Method Research Based on Data in Cyber-Physical System[J].Chinese Journal of Computers,2016,39(5), 961-972.

[9] Ivakhnenko A G. Heuristic self-organization in problems of engineering cybernetics[J]. Automatica, 1970, 6(2): 207-219.

[10] Dargahi-Zarandi A, Hemmati-Sarapardeh A. Modeling gas/vapor viscosity of hydrocarbon fluids using a hybrid GMDH-type neural network system[J]. Journal of Molecular Liquids, 2017, 236:162-171.

[11] Kondo T, Ueno J. Deep Feedback GMDH-Type Neural Network Using Principal Component-Regression Analysis and Its Application to Medical Image Recognition of Abdominal Multi-Organs[J]. Working Papers, 2015, 2(2):94.

[12] Moore R E, Bierbaum F. Methods and applications of interval analysis[M]. Philadelphia: Siam, 1979.

[13] Mitchell R, Chen I R. Effect of Intrusion Detection and Response on Reliability of Cyber Physical Systems[J]. IEEE Transactions on Reliability, 2013, 62(1):199-210.

[14] Yalei Y, Xingshe Z. Cyber-Physical Systems Modeling Based on Extended Hybrid Automata[C]//Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on . IEEE, 2013: 1871-1874.

[15] Li T, Tan F, Wang Q, et al. From offline toward real-time: A hybrid systems model checking and CPS co-design approach for medical device plug-and-play (MDPnP)[C]//Cyber-Physical Systems (ICCPS), 2012 IEEE/ACM Third International Conference on. IEEE, 2012: 13-22.

[16] Tan F, Wang Y, Wang Q, et al. Guaranteeing Proper-Temporal-Embedding safety rules in wireless CPS: A hybrid formal modeling approach[C]// Ieee/ifip International Conference on Dependable Systems and Networks. IEEE, 2013:1-12.

[17] Song Xiang-Jun, Zhang Guang-Quan. Modeling and Analysis of CPS Unmanned Vehicle Systems Based on Extended Hybrid Petri Net[J]. Computer Science, 2017, 44(7):21-24.

[18] Ishii D, Yonezaki N, Goldsztejn A. Monitoring Temporal Properties using Interval Analysis[J]. Ieice Transactions on Fundamentals of Electronics Communications & Computer Sciences, 2016, E99.A(2):442-453.

[19] Liu B X. Dynamic Parameters Estimation Based on Set Inversion and Interval Analysis Algorithm[J]. Advanced Materials Research, 2012, 591-593:1954-1957.

[20] [20] Wang Ying, Zhang Bo. Robot localization method using inversion set estimation in sensor networks[J]. Application Research of Compution, 2017, 34(4):1055-1059.

[21] Chabert G, Jaulin L. Contractor programming[J]. Artificial Intelligence, 2009, 173(11):1079-1100.

[22] Jaulin L, Walter E. Guaranteed nonlinear parameter estimation via interval computations[J]. Mathematics & Computers in Simulation, 1993, 35(2):123-137.

[23] Jaulin L, Kieffer M, Braems I, et al. Guaranteed non-linear estimation using constraint propagation on sets[J]. International Journal of Control, 2001, 74(18):1772-1782.

[24] Neto E D A L, Carvalho F D A T D. Nonlinear regression applied to interval-valued data[J]. Pattern Analysis & Applications, 2017, 20(3):809-824.