A Modified Version of Sugeno-Yasukawa Modeler

Amir H. Hadad[1], Tom Gedeon[1], Saeed Shabazi[2], and Saeed Bahrami[3]

[1] Department of Computer Science and Information Technology,
Australian National University, Canberra, Australia
{Amir.Hadad,Tom.Gedeon}@anu.edu.au
[2] NICTA Victoria Laboratory,
Department of Computer Science and Software Engineering,
University of Melbourne
shahbazi@csse.unimelb.edu.au
[3] Department of Computer Science and Information Technology,
Abhar Azad University, Tehran, Iran
saeed.bahrami@aau.edu.ir

Abstract. One of the most significant steps in fuzzy modeling of a complex system is Structure Identification. Efficient structure identification requires good approximation of the effective input data. Misclassification of effective input data can significantly degrade the efficiency of the inference of the fuzzy model. In this paper we present a modification to the Sugeno-Yasukawa modeler [1] to improve structure identification by increasing the accuracy of effective input data detection. We improved Sugeno-Yasukawa Modeler by modifying the algorithm in two ways. Firstly, we used a new Trapezoid Approximation method based on [2] to improve estimation of membership functions. Secondly we change the modeling process of modeling. There exist some intermediate models in the Sugeno-Yasukawa modeling process, a combination of which will result in the final fuzzy model of the system. In the original modeling process, parameter identification is only done for the final fuzzy model. By doing the parameter identification for the intermediate fuzzy models, we have improved the accuracy of these intermediate models. The RC (Regularly Criterion) error has been reduced for intermediate fuzzy models and the MSE decreased without using the new Trapezoid Approximation method. By using the new trapezoid method, the RC value for the intermediate models and MSE for the final model improved even more. This accuracy increase, result in a better detection of effective input data among input data records of a system.

Keywords: Fuzzy Logic, Fuzzy Modeling, Trapezoid Estimation, Structure Identification, Parameter Identification.

1 Introduction

A common real world problem is to find the effective input parameters for each new environment on the basis of required outputs and create a relationship between these input/output parameters to be able to react appropriately in different situations. After

H. Sarbazi-Azad et al. (Eds.): CSICC 2008, CCIS 6, pp. 852–856, 2008.

creating these relationships, the rules and the way input/output parameters related to each other should be adjusted. The whole process is called modeling or identifying a model [3], and can be divided into: structure identification, parameter identification. Structure identification is of two types. The first is to find the effective parameters is called structure identification type I (some would call that feature selection [9])

The fuzzy model was firstly introduced by Zadeh ([1], [4]) and is a good choice to model a nearly unknown environment, as humans do. In this paper we have introduced a modified version of the Sugeno-Yasukawa modeling process.

In section two of this paper the modified version of Sugeno-Yasukawa modeling process will be explained. In the third section, we present the modified version of Trapezoid Approximation. In the simulation results section we will compare different methods from different perspectives and the result of the modifications in the original modeling process and the benefits and drawbacks of using this modified version of modeling process will be presented.

2 Modified Sugeno-Yasukawa Fuzzy Modeling Method

Our main modification is in "structure identification type I" phase. In the original process there is only one parameter identification phase for the final detected parameters. But in our method in each of the phases in detecting effective parameters we run the parameter adjustment phase for the membership function of the input and output parameter(s). Afterwards the RC[1] criterion is calculated on the basis the formula which is presented in [6].

By applying this algorithm on the intermediate models, their membership function parameters would be adjusted. This results in lower RC values in both effective and non-effective input data intermediate models.

RC values have decreased in comparison with the RC values in [1] by using the modified modeling method for the following sample function:

$$y = (1 + x_1^{-2} + x_2^{-1.5})^2, \quad 1 \le x_1, x_2 \le 5. \tag{1}$$

3 Creating Fuzzy Membership Functions for Sparse Data

In our method we modified the algorithm which was presented in [2]. The data of each cluster for creating the fuzzy membership functions of the original algorithm is selected based on the following steps:

1. $MaxU = Max\{U_1(x), U_2(x), ..., U_n(x)\}$
2. $\forall x \in M \; if \; U_i(x) = MaxU \; then \; x \in cluster_i$

Where:

n: Number of clusters,
U_i: Membership function of cluster i,
$cluster_i$: it represents the *i'th* cluster dataset,
M: Dataset.

[1] Regulatory Criterion.

In cases where the dataset is sparse, the number of data which belongs to each of the clusters in this method can be restricted and it is not possible to create the fuzzy membership functions by using the trapezoid algorithm which is presented in [2]. We created a new algorithm for these cases. We changed the data categorization part by using the following selection method:

$$\forall x \in M \textit{ if } U_i(x) > \textit{boundary then } x \in \textit{cluster}_i$$

Where:

U_i: Membership function of cluster i,

$cluster_i$: it represents the i'th cluster dataset,

boundary: it is a constant number less than or equal to 0.1

M: Dataset.

In this new method there is no need the calculate *MaxU*.

After the cluster data selection, we use the Trapezoidal Approximation based on [2].

(a) Old method trapezoidal estimation result, q = 3

(b) Our method trapezoidal estimation result, q = 3

Fig. 1. (a) Old method trapezoidal estimation result (b) Our method trapezoidal estimation result (q=3 in both (a) and (b)).

It is clear that without changing each cluster data selection method it is not possible to apply the trapezoidal algorithm on data.

4 Simulation Results

For the simulation, part we generate sample datasets for formula (1) 5 times. For each of the sample dataset generations, we generate 3 input variables one hundred times randomly and calculate the output based on input variables 1 and 2. We applied three structure identification algorithms on these datasets:

1. Original modeling method with original Trapezoid Approximation method: SY_{SY}
2. Our modeling method with Tikk's Trapezoid Approximation method: HA_{TI},
3. Our modeling method with our modified version of Tikk's Trapezoid Approximation method: HA_{HA}

To make it clearer how we improved the structure identification phase, we have divided the intermediate fuzzy models into two groups:

Group I: Models which only include effective input parameters
Group II: Models which include both effective and non-effective input parameters

We did a comparison between the average improvements in RC values of these groups. The average RC improvement value for the HA_{TI} method, group I is equal to 51.74%, while it is equal to -6.33% for group II. On the other hand, average RC improvement value for HA_{HA} method, group I is equal to 47.93%, while it is equal to

3.87% for group II (Table 1). We can conclude that the RC improvement for non-effective parameters in the modeling process is either negative or a small positive value. As a result, the detection of the effective parameters will be improved and the likelihood of detecting non-effective parameters as effective will be decreased. We calculated the RC improvement based on the following formula:

$$RC\ improvement(\%) = \frac{RC_{old} - RC_{new}}{RC_{old}} * 100 \tag{2}$$

We also calculated the final models, MSE shares based on the following formula:

$$MSE\ Share(Method_i) = \frac{MSE\,(Method_i)}{MSE(Method_i) + MSE(Method_j) + MSE(Method_k)} * 100 \tag{3}$$

Since we ran the simulation five times, the results presented in Table 2, is the averages of MSE shares for each of the methods.

We are doing extra calculation to decrease MSE of the final model and the drawback will be processing time increase for HA_{HA} and HA_{TI} method. Time taken for each of the methods for 5 runs is presented in Table 3, as we can see; the time for the SY_{SY} method is the lowest.

Table 1. Comparison of group I intermediate fuzzy models and group II intermediate fuzzy models RC improvement

	HA_{HA}	HA_{TI}
Group I Improvement	47.93%	51.74%
Group II Improvement	3.87%	-6.33%

Table 2. Average of MSE shares of final models for 5 runs of the algorithm for random data

	HA_{HA}	HA_{TI}	SY_{SY}
MSE Share	16.3%	33.8%	49.9%

Table 3. Average of time shares of final models for 5 runs of the algorithm for random data

	HA_{HA}	HA_{TI}	SY_{SY}
MSE Share	51.2%	42.4%	6.4%

5 Conclusion

We modified the Trapezoid Approximation presented in [2] and created broader membership functions. This slightly increased the RC values of the intermediate models in the HA_{HA} method in comparison to the HA_{TI} method. However, after applying parameter identification on the final model the MSE was lower in the HA_{HA}

method. Since membership functions are broader the chance of finding a more optimum membership function will be increased.

We have also boosted the process of structure identification strongly by doing the parameter identification process for intermediate fuzzy models. In this way, we have decreased the RC error of intermediate models for intermediate models. What makes our modified process more accurate in detecting effective input parameters from non-effective input parameters is that the RC values reduction in models of group I is more than that of group II.

The HA_{HA} method has the following advantages in comparison with the SY_{SY} method. Firstly, it is very useful when number of probably effective input parameters is large. Secondly, it is useful for sparse data for trapezoidal approximation.

Acknowledgments. Amir Hadad is funded by a grant form the Australian Research Council (ARC), Saeed Shahbazi is funded by National ICT Australia (NICTA).

References

1. Fukuyama, Y., Sugeno, M.: A new method of choosing the number of clusters for fuzzy c-means method. In: Proc. 5th Fuzzy system Symposium, pp. 247–250 (1989)
2. Tikk, D., Biró, G., Gedeon, T.D., Kóczy, L.T., Yang, J.D.: Improvements and Critique on Sugeno's and Yasukawa's Qualitative Modeling. IEEE Transaction on fuzzy systems 10(5) (October 2002)
3. Sugeno, M., Yasukawa, T.: A Fuzzy-logic-based Approach to Qualitative Modeling. Fuzzy Systems, IEEE Transactions on Fuzzy Systems 1(1), 7–31 (1993)
4. Zadeh, L.A.: Towards a theory of fuzzy systems. In: Kalman, R.E., DeClaris, R.N. (eds.) Aspects of Network and System Theory, pp. 469–490. Holt, Rinehart and Winston, New York (1971)
5. Zadeh, L.A.: Fuzzy sets and information granularity. In: Gupta, M., Ragade, R., Yager, R. (eds.) Advances in Fuzzy Set Theory and Application, pp. 3–18. North Holland, Amsterdam (1979)
6. Ihara, J.: Group method of data handling towards a modeling of complex systems-IV. Systems and Control (in Japanese) 24, 158–168 (1980)
7. Fukuyama, Y., Sugeno, M.: A new method of choosing the number of clusters for fuzzy c-means method. In: Proc. 5th Fuzzy System Symposium (in Japanese), pp. 247–250 (1989)
8. Chua, L.O., Yang, L.: Cellular neural networks: theory. IEEE Transactions on Circuits and Systems 35(10), 1257–1272 (1988)
9. Langley, P.: Selection of relevant features in machine learning. In: Proceedings of the AAAI Fall Symposium on Relevance. AAAI Press, Menlo Park (1994)