Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Model Structure Identification of Hybrid Dynamical Systems based on Unsupervised Clustering and Variable Selection

Duc-An NGUYEN * Jude NWADIUTO * Hiroyuki OKUDA * Tatsuya SUZUKI *

** Department of Mechanical Systems Engineering, Nagoya University, Nagoya, 464-8603, JAPAN.(e-mail: nguyen.an.duc@h.mbox.nagoya-u.ac.jp, j_nwadiuto@nuem.nagoya-u.ac.jp, h_okuda@nuem.nagoya-u.ac.jp, t_suzuki@nuem.nagoya-u.ac.jp)*

Abstract: This paper proposes a systematic identification process for the hybrid dynamical system (HDS) estimating not only the coefficients but also the structure of the model. Beneficially speaking, the proposed system identification is used for the HDS system that the model structure, the number of modes, and the explanatory variables of the model are unknown. In the proposed method, a quantitative index to evaluate the number of modes is deployed and the optimal number of modes is determined from the measurement. The variable selection method is also introduced to determine the explanatory variables in each mode in a systematic manner. Two of piece-wise linear models which are prepared for the proposed system to identify and the validity of the proposed method is then demonstrated. Finally, the result of the proposed in comparison with the conventional system identification method for HDS is discussed.

Keywords: Hybrid Systems Identification, Variable Selection, Unsupervised Classification.

1. INTRODUCTION

The system identification is well known as an important issue on the controller design in modern system control theory. Especially, the need for the system identification algorithm for complex nonlinear systems is increasing. There are several approaches to identify the nonlinear system. The first approach is based on the first principle if the backgrounding physical phenomena of the system is known. The second approach which is getting popular is the use of machine learning technique. The neural network which considerably improved with the development of convolution network is one of the most promising techniques to imitate the input-output relation of the original system. The third approach is the approximation of the original complex behavior by using several more simple submodels. This kind of system is known as the continuous/discrete hybrid dynamical system (HDS) which consists of several simple continuous dynamical models combined with the discrete switching event.

In the past year, the identification of hybrid dynamical systems (HDS) has been broadly discussed in many different contexts. One of the applications of the system identification for the HDS is to represent the complexity of human behavior in Choi and Farrell (2000), Kim et al. (2005), Sekizawa et al. (2007), Akita et al. (2008), Okuda et al. (2009), Okuda et al. (2012).While modern machine learning-based control methods improve the performance and robustness, they add-up complexity and it is not possible to understand the underlying physical phenomena of the system. On the other hand, it is confirmed in these studies that the HDS identification technique helps the understanding of human behavior thanks to its simple structures and describability. Researchers have investigated the properties of various model classes to identify HDS model properly in Ferrari-Trecate et al. (2003), Bemporad et al. (2003), Roll et al. (2004). One of the well-known class of HDS for the system identification is the Piece-Wise Auto-Regressive system with eXogenous inputs (PWARX). PWARX models are a generalization of the classical ARX models, in the sense that the regressor space is partitioned into a finite number of polyhedral regions, wherein each region the input-output relation is defined through an ARX model. PWARX model represents a broad class of hybrid systems, they form a subclass of piecewise affine (PWA) models, Sontag (1981), which is under mild condition equivalent to other hybrid modeling formalisms, such as mixed logic dynamics (MLD) systems Bemporad (1999) and linear complementarity (LC) models, A. J v. d. Schaft and J.M. Schumacher (1996); W. Heemels et al., (2000), and Hofbaur and Williams (2004). However, the main limitation of these identification algorithms is the requisition of a prior knowledge of the model structure. The model structure must be known or decided before starting the system identification process, in other words. System identification technique is used sometimes if the first principle modeling is not available since the prior knowledge is not enough to describe the model. The system identification approach can be more attractive when the above drawbacks are overcame. In HDS, there are two features to determine the structure of the model. The first one is the number of modes or sub-models, which

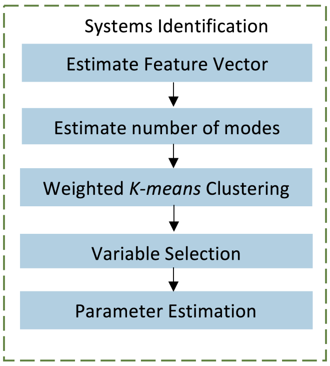

Fig. 1. Procedure of Systems Identification

is the most significant factor to identify the accurate HDS model. Another feature is the designing of the explanatory variables. However, it is known that the model can provide better performance if the model includes more number of explanatory variables for the learning data, despite its reduction of the advantages in simplifying the model complexity to understand the physical phenomena of the model and also in the model accuracy, for the 'non-learning data' from the over-fitting problem. In this research, a new framework for the system identification of hybrids dynamical systems focusing on the estimation of the model structure is presented by extending the previous HDS identification method based on the clustering Ferrari-Trecate et al. (2003). In this paper, one of the class of HDS, PieceWise AutoRegression eXogenous (PWARX) model is examined. Its identification process starts with estimating the feature vector from the raw data. All measurements are partitioned into several submodels (modes) based on the distribution of the computed feature vector. The main difficulty in the automatic partitioning process is solved by adopting the unsupervised clustering technique, which consists of the *'weighted K-means'* algorithm with Davies Bouldin Criterion (DBC). These techniques allow automatic classification of the submodels that share similar dynamics but are defined in different regions, which will be shown in **Example 2**. After the mode classification is done, the model structure in each mode, i.e. the selection of the explanatory variables, is estimated by comparing the model structure with Akaike Information Criterion (AIC) to overcome the overfitting problem and estimate the actual model structure. Once the model structure and the mode assignment for all measurements have been estimated, multivariate linear regression can be used to estimate the parameter of PWARX model in each mode. The proposed algorithm is tested for the mathematical example of a PWARX model which has different variable structures in each mode.

2. PROBLEM FORMULATION

For given a discrete-time nonlinear dynamic systems with input $\Phi_k \in \mathbb{R}^n$, output $y(k) \in \mathbb{R}$, A Piece-Wise Affine (PWA) model establishes a relationship between observations $\Phi(k)$ and the predicted output $y(k)$ in the form

$$y(k) = f(\Phi_k) + \varepsilon_k \tag{1}$$

in which $\varepsilon_k \in \mathbb{R}$ is Gaussian distributed error term with mean μ and variance σ^2, $\Phi_k \in \mathbb{R}^n$ is the regression vector, and $f : \mathbb{R}^n \to \mathbb{R}$ is a PWA function of the following form:

$$f(\Phi_k) = \begin{cases} \boldsymbol{\theta}^{1T}\boldsymbol{\Phi_1(k)} & \text{if } \Phi(k) \in \mathscr{R}_1 \\ \boldsymbol{\theta}^{2T}\boldsymbol{\Phi_2(k)} & \text{if } \Phi(k) \in \mathscr{R}_2 \\ \vdots & \vdots \\ \boldsymbol{\theta}^{sT}\boldsymbol{\Phi_j(k)} & \text{if } \Phi(k) \in \mathscr{R}_s \end{cases} \tag{2}$$

With $\varphi = \phi_1, \phi_2, \phi_3, ..., \phi_k$ denoted as the set of the entire input variables, and $\boldsymbol{\Phi_j(k)} \in \{\varphi\}$ represented the regressor vectors of mode j (j $\in$ {1,2,...s}), where $\{\varphi\}$ is the vector which is the concatenation of the elements of the set φ. $\{\mathscr{R}_i\}_{i=1}^s$ gives a complete partition of $\Phi(k)$ domain $\mathscr{R} \subseteq \mathbb{R}^n$. Each region $\mathscr{R}_i$ is described by:

$$\mathscr{R}_i = \left\{\boldsymbol{\varphi_k} \in \mathbb{R}^n : \boldsymbol{H}_i(\Phi_j) \preceq_{[i]} 0\right\}$$

where $\boldsymbol{H_i}$ is a matrix that defines the partition $\{\mathscr{R}_i\}_{i=1}^s$. The symbol $\preceq_{[i]}$ denotes a vector whose elements can be the symbols $\leq$ or $<$.
For easier understanding, we reinterpreted an example of $f(x_k)$ as follow:

$$f(x_k) = \begin{cases} \theta_1^1\phi_1 \quad + \theta_3^1\phi_3 + ... + \theta_k^1\phi_k & \text{if } \boldsymbol{H}_i(\Phi_1) \in \mathscr{R}_1 \\ \quad \theta_2^2\phi_1 + \theta_3^2\phi_3 + ... + \theta_k^2\phi_k & \text{if } \boldsymbol{H}_i(\Phi_2) \in \mathscr{R}_2 \\ \qquad \vdots & \\ \theta_1^s\phi_1 \quad + \theta_4^s\phi_4 + ... + \theta_k^s\phi_k & \text{if } \boldsymbol{H}_i(\Phi_k) \in \mathscr{R}_s \end{cases} \tag{3}$$

where the blank spaces indicate the explanatory variables that are not used in the origin model. The identification problem can be stated as follow:

Problem 1. Assume that the data set $\mathscr{S} = \{(\Phi(k), y(k)), k = 1, ..., N\}$ is generated from the PWARX model (3). Automatically estimate the partition $\mathscr{R}_i$, variables $\boldsymbol{\Phi_j(k)}$ and parameter vectors $\{\theta_i\}_{i=1}^s$ that characterizing the PWARX model (3).

3. IDENTIFICATION PROCEDURE

In order to solve Problem 1., we proposed a framework depicted in Fig. 1, composing of five main stages. Firstly, feature vectors are estimated from the generated data. Secondly, based on the estimated feature vectors, an automatic identification of the structure of the target PWARX is proceeded, including evaluating the optimum number of submodels s then partition data into s different segments (modes) using clustering technique. Then, for each submodels the influential variable is selected by the Variable Selection method. Finally, the parameters of each selected variable is being estimated by weighted- Least Square method. In furtherance of illustrating such framework, we will generate the following example:
Example 1.
Let the data be generated as PWARX linear systems with 3 variables, where:

$$f(x_k) = \begin{cases} [\;1 \quad 0 \quad 0.5]\Phi_1(k) & \text{if } \boldsymbol{H}_i(\Phi_1) \in \mathscr{R}_1 \\ [\;0 \quad 0.5 \quad 0.7]\Phi_2(k) & \text{if } \boldsymbol{H}_i(\Phi_2) \in \mathscr{R}_2 \\ [\;1 \quad 1 \; -1.1]\Phi_3(k) & \text{if } \boldsymbol{H}_i(\Phi_3) \in \mathscr{R}_3 \end{cases} \tag{4}$$

for which N = 500 of the input sample $x(k) \in \mathbb{R}^p$ sampled from a uniform distribution, with $\mu = 1$ and variance $\sigma^2 = 0.1$. The zero parameter terms corresponding to the variables that does not contribute any effect in the original model. The illustration data (y, ϕ_1, ϕ_2) is depicted in Fig. 2.

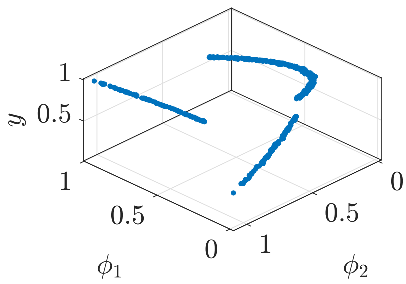

Fig. 2. Generated measurements of targeting model in $\phi_1 - \phi_2$ space.

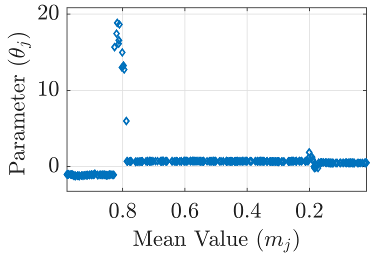

Fig. 3. Plot of estimated feature vectors Ex.1

3.1 *Estimate feature vector*

In the first stage, we initiate a small local data set (LDs) from the generated data. With the assumption that a PWA map is *locally linear*, a small subsets of points $x(k)$ that are close to each other are likely to dwell in the same region $\bar{\mathscr{R}}_i$. For each datapoint $(x(j), y(j) : j = 1, ..., N)$ we build a LDs $\mathscr{C}_j$ collecting $(x(j), y(j))$ and the $c-1$ distinct datapoint (x̃,ỹ) that satisfy:

$$||x(j) - \tilde{x}||^2 \leqslant ||x(j) - \hat{x}||^2 \quad \forall (\hat{x}, \hat{y}) \in \mathscr{S} \backslash \mathscr{C}_j, \qquad (5)$$

where $||.||$ is the euclidean norm. The vector of coefficient θ_j^{LDs} estimated from the data in $\mathscr{C}_j$ is computed by the following equation:

$$\theta_j^{LDs} = (\phi_j^T \phi_j)^{-1} \phi_j^T \, y_{\mathscr{C}_j}, \; \phi_j = \begin{bmatrix} x_1 & x_2 & \dots & x_c \\ 1 & 1 & \dots & 1 \end{bmatrix} \qquad (6)$$

where x_i are the vector of regressors belonging to $\mathscr{C}_j$ and y_{C_j} is the vector of the output sample in $\mathscr{C}_j$. We also calculate the mean value:

$$m_j = \frac{1}{c} \sum_{(x,y) \in \mathscr{C}_j} x, \qquad j = 1, ..., N \qquad (7)$$

Now, we consider the *feature vector*:

$$\zeta_j = [(\theta_j^{LDs})^T, m_j^T] \quad \forall j = 1, ..., N. \qquad (8)$$

Fig. 3 shows a part of feature values distribution for Example 1.

3.2 *Estimate number of modes*

The clustering technique is applied to make the segmentation of the feature vectors into s classes, and criterion is required to determine number of modes. Within this work, Davies-Bouldin Criterion (DBC) is used to inspect the optimum number of the cluster to output the optimized number of submodels *quantitatively* .

DBC is constructed on the idea that the inter-cluster separation as well as intra-cluster homogeneity and compactness, should be high, Davies and Bouldin (1979) In this step the density and the distance of the sample points are calculated, which contain the information of the density distribution leading to the optimum clusters. The appropriate number of cluster number of cluster centers is obtained through calculating the Davies-Bouldin value of every clustering result and choose the one which has the minimum Davies-Bouldin value.

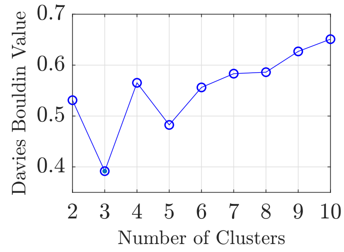

Fig. 4. DBC index for number of cluster Ex.1

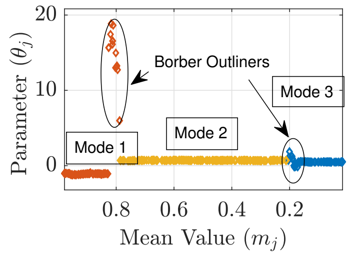

Fig. 5. Plot of clustered feature vector Ex.1

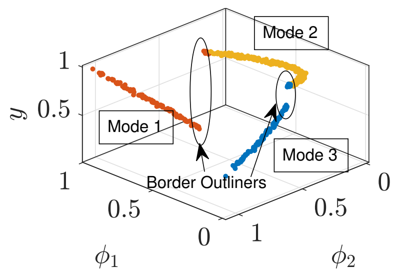

Fig. 6. Clusters of target PWARX system for Ex.1 in $\phi_1 - \phi_2$ space.

$$s = \underset{1,2,...,N}{\operatorname{argmin}} V_{DB} = \frac{1}{N} \sum_{i=1}^{N} P_i \qquad (9)$$

where s is the optimal number of clusters and P_i is defined as:

$$P_i = max_{(i \neq j)} P_{ij}, \qquad (10)$$

where P_{ij} is the similarity measure between clusters R_i and R_j, and is defined as:

$$P_{ij} = \frac{S_i + S_j}{D_{ij}}, \qquad (11)$$

$$S_i = \left(\frac{1}{|R_i|} \sum_{x \in R_i} ||x - \mu_i||^p \right)^{1/p}, p > 0 \qquad (12)$$

where $|R_i|$ is the number of vectors in cluster R_i and μ_i is the center of cluster R_i, and:

$$D_{ij} = \left(\sum_{i=l}^{N} ||v_{il} - v_{jl}||^t \right)^{1/t}, t > 1 \qquad (13)$$

where v_i and v_j are the centroids of the clusters R_i and R_j respectively. According to DBC Theory, the minimum

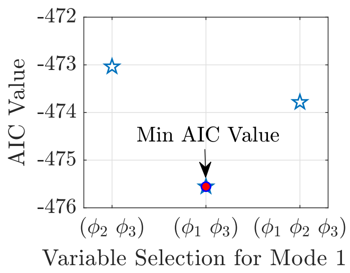

Fig. 7. AIC for mode 1 variables

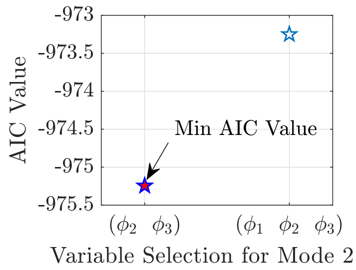

Fig. 8. AIC for mode 2 variables

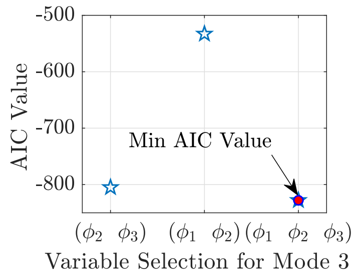

Fig. 9. AIC for mode 3 variables

value of DBC index refers to the optimum number of cluster. Fig. 4 shows the Davies Bouldin values for different number of cluster of the exampled PWARX system. For Example 1. $s_{optimum} = 3$ is obtained successfully.

3.3 Clustering feature vectors

The feature vectors are classified into s subsets after the decision of optimum s. Normally, any clustering algorithm can be used. However, within this paper, Ferrari-Trecate et al. (2003) *weighted* K-means method is used. In the *weighted* K-means method, the following cost function considered:

$$J(\{\mathscr{R}_i\}_{i=1}^{s}, \{\mu\}_{i=1}^{s}) = \sum_{i=1}^{s}\sum_{\zeta_j\in\mathscr{R}_i} w_j ||\zeta_j - \mu_i||^2 \tag{14}$$

where μ_i are the centers of the cluster $\mathscr{R}_i$, and with w_j is the weighting factor defined by:

$$w_j = \frac{1}{\sqrt{(2\pi)^{2n+1} det(C)}} \tag{15}$$

with C is the covariance matrix constructed from empirical covariance matrix and scatter matrix **?**. The goal of weighted *K-means* is to find the subset $\mathscr{R}_i$ and center μ_i that minimize J.

The result of clustering algorithm implemented into Example 1. is visualized in Fig. 5, 6.

3.4 Variable Selection

Variable selection is considered as a powerful method to eliminate unnecessary variables out of the final model. For example, during the motion of autonomous vehicle, it has to keep on sensoring surrounding objects to measure information such as: relative position, velocity, acceleration, steering angle, lane changing intention, and so on. The number of information which can be measured will be increasing in the case of the driving situation with many surrounding cars. It is very wasteful to measure all these variables as some of them do not contribute for the prediction of surrounding cars. Furthermore, this reduction of unnecessary information is beneficial when the obtained model is analyzed to understand the principle underlying model. In this research, the variable selection is adopted to realize the reduction of unnecessary information. In each region $\mathscr{R}_i$, $2^k - 1$ models corresponding to all the possible combinations of $\phi(k)$ are tested based on Akaike Information Criteria:

$$AIC = N * log\left(\frac{1}{N}\sum_{i=1}^{N}(y^i - f_j(\phi_j^i))^2\right) + 2(n_j+1) + N*(log(2\pi)+1) \tag{16}$$

where:
N : Number of data,
k : Number of candidates of input variables,
j : $j \in \{1, 2, ..., 2^k - 1\}$ (All possible combinations),
n_j: Number of input variables in combination j,
ϕ_j^i: Selected input vector of the $i-th$ data,
y^i : Output of the $i-th$ data.

According to Akaike's theory, the most appropriate model can be obtained as the model that giving the smallest AIC value.

For the Example 1., the result for variable selection are shown in Figs. 7, 8, 9, and the obtained optimal PWARX model structure is:

$$f(x_k) = \begin{cases} \theta_1^1\phi_1 \quad\quad\quad\quad + \theta_3^1\phi_3 & \text{if } \boldsymbol{H}_i(\Phi_1) \in \mathscr{R}_1 \text{ (mode 1)} \\ \quad\quad\quad \theta_2^2\phi_2 + \theta_3^2\phi_3 & \text{if } \boldsymbol{H}_i(\Phi_2) \in \mathscr{R}_2 \text{ mode 2)} \\ \theta_1^3\phi_1 + \theta_2^3\phi_2 + \theta_3^3\phi_3 & \text{if } \boldsymbol{H}_i(\Phi_3) \in \mathscr{R}_3 \text{ (mode 3)} \end{cases} \tag{17}$$

3.5 Estimating parameter in each submodels

Now, the parameter estimation of each submodels in each is achieved by using the weighted-Least Square Method (weighted-LS), which minimizes the cost function:

$$\sum_{\phi(j),y(j)\in\mathscr{R}_j} (y(j) - \theta_k\Phi_j(k))^T w_j (y(j) - \theta_k\Phi_j(k)) \tag{18}$$

It can be seen in the formula (15) that weighted factor w_j is proportional to inverse of the square root of the determinant of the covariance matrix C. This weight helps to eliminate the effect of the outliner data points near the border (showed in Fig. 5) in the parameter estimation step. For each mode $j = 1, 2, ...s$ the parameter is estimated using the following weighted-Least Square formula:

$$\theta^s = [(\Phi_j(k))^T W_j(\Phi_j(k))]^{-1}\Phi_j(k)^T W_j \; y(k). \tag{19}$$

where W_j is the diagonal matrix of the weighted factor w_j.

As the result of all processes we proposed, the following PWARX models are successfully identified for the Example 1:

$$\begin{aligned} \theta^{1T} &= [\ 1.0109 \quad 0 \quad 0.5001] \\ \theta^{2T} &= [\ \quad 0 \quad 0.4999 \quad 0.7045] \\ \theta^{3T} &= [\ 1.0202 \quad 1.0011 \quad \text{-}1.0997] \end{aligned} \tag{20}$$

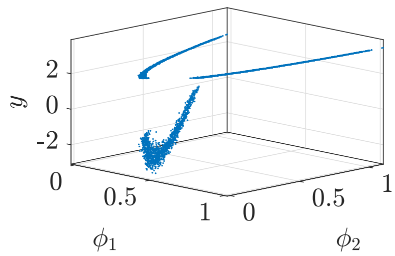

Fig. 10. Generated measurements of PWARX Ex.2

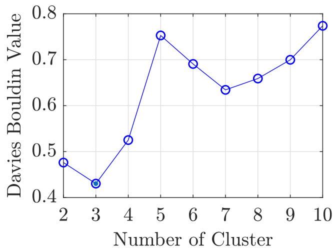

Fig. 11. DBC values for number of clusters Ex.2

4. EVALUATION OF SYSTEMS IDENTIFICATION

In this section, the proposed method is demonstrated with another example that is made more complicated. **Example 2.** The target model of identification is set as follows:

$$f(x_k) = \begin{cases} \phi_1 + 0.4\,\phi_2 - 0.01\phi_3 \qquad\qquad\qquad + \;\; \phi_6 & \text{if } \boldsymbol{H}_i(\Phi_1) \in \mathscr{R}_1 \\ \phi_1 + 0.4\,\phi_2 \qquad\qquad\qquad\qquad\qquad\;\; + \;\; \phi_6 & \text{if } \boldsymbol{H}_i(\Phi_2) \in \mathscr{R}_2 \\ \qquad\qquad\qquad - 1.2\,\phi_3 \; + 7\,\phi_4 \quad + 3\,\phi_6 & \text{if } \boldsymbol{H}_i(\Phi_3) \in \mathscr{R}_3 \end{cases}$$

To generate the testing measurement, N = 500 of the input sample $x(k) \in \mathbb{R}^n$ are sampled from an uniform distribution, with $\mu = 1$ and variance $\sigma^2 = 0.1$. The blank terms corresponding to the variables that does not contribute any effect in the original model. The result of proposed identification procedure implemented to Example 2 are visualized in Figs.(10),(11),(12),(13). And the variable selection method and parameter estimation applied to the target model for the three mode of target PWARX systems is depicted in the following result Eq.(21). The white boxes represent for explanatory variables eliminated by AIC Model Selection.

$$\begin{aligned} \theta^{1T} &= [1.0099\ 0.3991\ \text{-}0.0066 \quad \square \quad \square \quad 2.0127] \\ \theta^{2T} &= [0.9999\ 0.4000 \quad \square \quad \square \quad \square \quad 1.9999] \\ \theta^{3T} &= [\quad \square \quad \square \quad \text{-}1.1922\ \text{-}6.9775 \quad \square \quad 2.9780] \end{aligned} \tag{21}$$

The results show in Figs.(10),(11),(12),(13) confirmed that the proposed framework has successfully identified the correct number of modes for the targeting PWARX systems. When comparing with the parameter obtained by Ferrari-Trecate et al. (2003) technique, it is noticed that the estimation of parameters without AIC, is not able to produce the reliable model that contained the zero terms, resulting in a set of vague parameters:

Example 1. Parameter Estimation without AIC:

$$\begin{aligned} \theta^{1T} &= [\ 1.0185 \quad \mathbf{-0.0004} \quad 0.4951] \\ \theta^{2T} &= [\ \mathbf{0.0031} \quad 0.5001 \quad 0.6999] \\ \theta^{3T} &= [\ 1.0085 \quad 0.9984 \quad \text{-}1.1066] \end{aligned}$$

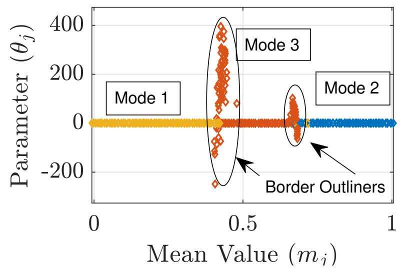

Fig. 12. Plot of clustered feature vectors for Ex.2

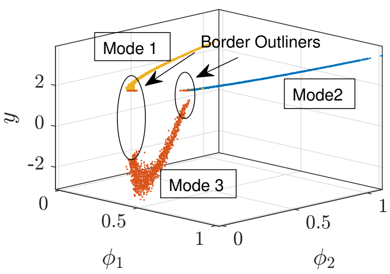

Fig. 13. Clusters of target PWARX systems Ex.2

Example 2. Parameter Estimation without AIC:

$$\begin{aligned} \theta^{1T} &= [\ 1.0186\ \ 0.4634\ \ \text{-}0.0187\ \ \mathbf{0.0133}\ \ \mathbf{0.0192}\ 2.0523] \\ \theta^{2T} &= [\ 1.0488\ \ 0.4208\ \ \mathbf{-0.0261}\ \ \mathbf{0.0058}\ \ \mathbf{0.0192}\ 2.0642] \\ \theta^{3T} &= [\mathbf{-0.0123}\ \mathbf{0.1155}\ \ \text{-}1.0127\ \ \text{-}7.3775\ \ \mathbf{-0.0044}\ 3.1280] \end{aligned}$$

In Example 1., it is noticed that conventional method could not estimate the actual model structure for the target PWARX systems. Even though the parameter $\phi^1_1 = \mathbf{-0.0004}$ and $\phi^2_1 = \mathbf{0.0031}$ seems meaningless since the coefficients are very small, there is no reason to ignore those explanatory variables. Of course the variable selection can be done by setting appropriate threshold to eliminate the variables. For example, the appropriate variables can be selected by setting the threshold, $|\theta^s_j| < 0.02$. However, the result of Example 2., without AIC shows that it is not possible to set the appropriate threshold obviously. The parameter $\phi^1_3 = -0.0187$ which is meaningful term and have to be included in identified model while the parameter $\phi^2_3 = \mathbf{-0.0261}$ that does not include any information of the systems and should be eliminated in the resulting model. In contrast, as the result shown in Eq.(20) and Eq.(21), the proposed system identification process could obtain the actual model structure successfully by applying the variable selection. Consequently, it is confirmed that the variable selection method must be included if the original model structure is not given. Generally speaking, the system identification technique is used when the original model structure can not be estimated from the first principle modeling. This means the model selection have to be applied normally in the process of HDS identification as proposed method.

Furthermore, in considering the system that containing number of variables in hybrid systems, a compact model representation with model selection is effective to avoid the over-fitting problem and to reduce the computational burden in the application of the model. For example, in the case of the driving behavior analysis introduced in the section 1, the vehicle has to keep on surveilling its surrounding objects or vehicles by analyzing their data, such as relative velocity, acceleration, lane changing intention, traffic signal, and so on. The number of variables multiplied with the number of vehicles gives a large number of data to be analyzed. Therefore, it will become more efficient if the insignificant variables to be unloaded from the recognition and signal processing.

Another big contribution of this paper is an automatic decision of the number of modes in HDS. Prior work by Ferrari-Trecate et al. (2003), Bemporad et al. (2004) have demonstrated the effectiveness by adopting the clustering method in identifying PWARX model. However, these studies can not be applied when the number of modes is not given. Although, this assumption is hold for the system identification or parameter estimation for the known artificial system, the systematic approach to determine the appropriate number of modes must be exploited for more complex system such as human behavior. Another approach to estimate the number of modes is using an unsupervised clustering, such as hierarchical clustering Okuda et al. (2012). This approach helps to know the relationship of each modes for understanding the physical meaning as model interpretation, however, the systematic estimation of optimal sub-models segmentation is not considered, intuitively speaking the accuracy of such approach cannot be verified in standard method, but rather experience of the author. The proposed method is expected to provide more reliable and robust mode segmentation without the prior knowledge on the decision of the numeber of modes. In this research, the automated sub-models identification by examining Dabvies-Bouldin Criterion has shown a good result for both examples as it has shown the precise number of modes for both simple and complex examples. We proposed the Davies Bouldin Criterion for the balanced trade-off between the computational time and precision. Specifically, for the same performance in sorting the correct number of clusters, within our knowledge, DBC usually gives the fastest result with satisfying precision, compared with the other method. Moreover, DBC functioning similarly to an adjusted density over distance inspecting index, it is expected that the increasing number of data helps DBC inspecting better density corresponding with the correct number of cluster. However, since it will be a trade-off with the computing time, we recommend the DBC index for the implementation of the system contain the number of data point less than 10,000.

5. CONCLUSION

In this paper, a systematic framework to identify hybrid system that contains unknown model structure, number of modes, and explanatory variable, is presented. In the proposed method, the quantitative index DBC had been integrated to automatically decide the correct number of models. Also the Variable Selection method is employed to determine the explanatory variables in each mode in systematic manner. Our results suggested that Variable Selection is a valuable addition for the system identification, due to its ability to explicitly eliminate trivial factors from the system. The future work would be the implementation of the proposed framework to analyze and model driver behavior from real-world data as introduced in Section 1.

REFERENCES

Akita, T., Suzuki, T., Hayakawa, S., and Inagaki, S. (2008). Analysis and synthesis of driving behavior based on mode segmentation. In *2008 International Conference on Control, Automation and Systems*, 2884–2889. IEEE.

Bemporad, A., Garulli, A., Paoletti, S., and Vicino, A. (2003). A greedy approach to identification of piecewise affine models. In *International Workshop on Hybrid Systems: Computation and Control*, 97–112. Springer.

Bemporad, A., Garulli, A., Paoletti, S., and Vicino, A. (2004). Data classification and parameter estimation for the identification of piecewise affine models. In *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, volume 1, 20–25. IEEE.

Choi, J.Y. and Farrell, J.A. (2000). Nonlinear adaptive control using networks of piecewise linear approximators. *IEEE Transactions on Neural Networks*, 11(2), 390–401.

Davies, D.L. and Bouldin, D.W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2), 224–227.

Ferrari-Trecate, G., Muselli, M., Liberati, D., and Morari, M. (2003). A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2), 205–217.

Hofbaur, M.W. and Williams, B.C. (2004). Hybrid estimation of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(5), 2178–2191.

Kim, J.H., Hayakawa, S., Suzuki, T., Hayashi, K., Okuma, S., Tsuchida, N., Shimizu, M., and Kido, S. (2005). Modeling of driver's collision avoidance maneuver based on controller switching model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(6), 1131–1143.

Okuda, H., Ikami, N., Suzuki, T., Tazaki, Y., and Takeda, K. (2012). Modeling and analysis of driving behavior based on a probability-weighted arx model. *IEEE Transactions on Intelligent Transportation Systems*, 14(1), 98–112.

Okuda, H., Suzuki, T., Nakano, A., Inagaki, S., and Hayakawa, S. (2009). Multi-hierarchical modeling of driving behavior using dynamics-based mode segmentation. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 92(11), 2763–2771.

Roll, J., Bemporad, A., and Ljung, L. (2004). Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1), 37–50.

Sekizawa, S., Inagaki, S., Suzuki, T., Hayakawa, S., Tsuchida, N., Tsuda, T., and Fujinami, H. (2007). Modeling and recognition of driving behavior based on stochastic switched arx model. *IEEE Transactions on Intelligent Transportation Systems*, 8(4), 593–606.

Sontag, E. (1981). Nonlinear regulation: The piecewise linear approach. *IEEE Transactions on Automatic Control*, 26(2), 346–358.