FUZZY LOGIC-BASED INTELLIGENT CONTROL FOR SVM SPEAKER VERIFICATION WITH THE SUPPORT OF GMM PRIOR INFORMATION

Ing-Jr Ding, Chih-Ta Yen and Zih-Jheng Lin
Department of Electrical Engineering, National Formosa University, Taiwan
E-mail: eugen.ding@gmail.com; chihtayen@gmail.com

ICETI 2012-F1060_SCI
No. 13-CSME-34, E.I.C. Accession 3492

ABSTRACT

In this paper, a fuzzy logic-based intelligent control (FLIC) scheme for support vector machine (SVM) speaker verification, called FLICSVM, is developed. The proposed FLICSVM method enhances SVM training by considering the property of training utterances for establishing the SVM model and therefore could further ensure the robustness of the SVM classifier on speaker verification. In FLICSVM, when establishing the SVM model in the training procedure, the popular fuzzy control methodology is employed to tune certain specific SVM parameter according to the prior information of SVM training utterances that is derived from Gaussian mixture model (GMM) calculations. Experimental results demonstrated that proposed FLICSVM is apparently superior to conventional SVM in the recognition accuracy.

Keywords: SVM; FLICSVM; speaker verification.

CONTRÔLE INTELLIGENT BASÉ SUR LE SYSTÈME DE LOGIQUE FLOUE POUR CONTRÔLE DE VÉRIFICATION DE LOCUTEUR PAR SVM AVEC SUPPORT D'INFORMATIONS GMM PRÉALABLES

RÉSUMÉ

Dans cet article, nous développons un schéma de contrôle intelligent par système de logique floue (FLIC) pour machines à vecteurs de support pour le contrôle de vérification de locuteur, appelé FLICSVM. Le FLICSVM proposé rehausse la formation SVM en considérant les propriétés de la formation d'énoncés pour établir le modèle SVM, et par conséquent pourrait assurer la robustesse du SVM pour la vérification de locuteur. Dans le FLICSVM, pendant l'établissement du modèle SVM, dans la procédure de formation, la méthodologie de contrôle par logique floue est employer pour accorder certains paramètres spécifiques SVM en accord avec les informations des énoncés SVM dérivés des calculs de modèles de mélange gaussien (GMM). Les résultats des expériences démontrent que le FLICSVM est manifestement supérieur au SVM conventionnel pour l'exactitude de la reconnaissance.

Mots-clés : SVM ; FLICSVM ; vérification du locuteur.

1. INTRODUCTION

Speaker recognition has been a popular technique in recent years. Speaker recognition can be further divided into two main categories, speaker identification and speaker verification. Speaker identification is used to determine the identity of a particular person. The purpose of a speaker verification system is to verify the identity of people based on their uttered voices to decide whether the speaker is accepted or not. The paper focuses on the speaker verification application. According to the author's knowledge, the support vector machine (SVM) approach [1] provides the most favorable option in performing speaker verification [2–4]. SVM technique has been popular in lots of technical areas and also used to some specific applications, e-mail spam filtering [5], for example.

When performing speaker verification by SVM, the quality of the training data for establishing SVM models will largely affect the classification accuracy of the trained SVM classifier. When encountering improper training data on SVM designs, extremely little dissimilarity between valid speakers' and imposters' utterances characteristics, for example, a model-based SVM enhancement methodology may be adopted to improve the classification accuracy of the trained SVM by adjusting the certain parameter of the SVM model to adapt itself to the training data [2–4]. Boujelbene et al. [3] in 2010 proposed an improved SVM method by modifying the kernel function. In addition, Dong et al. [4] developed a fast SVM training algorithm to solve the problem that establishes an SVM on a very large-sized training data set.

Although previous studies on SVM improvements could effectively increase the performance of SVM, few works are done on tuning SVM model to match the training data characteristics to achieve the optimal recognition accuracy of SVM classification. To tackle this issue, this paper employs the popular fuzzy technique to carry out the work of SVM parameter adjustments. This study proposes fuzzy logic-based intelligent control (FLIC) scheme [6] for SVM tuning on speaker verification application. Fuzzy methods have been successfully used to lots of fields [7, 8], including speaker recognition in this work. The developed method, FLICSVM, uses the evaluated training data characteristics to regulate the free parameter C of SVM formulations so that the known SVM margin size decision problem in SVM separating hyperplane could be effectively solved.

2. SPEAKER VERIFICATION USING SVM

For speaker verification tasks, the widely used SVM classifier is appropriate for verifying whether the input speech utterance from a test speaker is acceptable for the system [1]. The SVM is commonly used as a data classifier. The SVM classifies new input data using a separating hyperplane. If the SVM model attempts to determine whether an input speech datum belongs to the valid speaker set, it first attempts to locate the SVM model of the valid speaker set in the SVM database. The separating hyperplane of the SVM model of the valid speaker set then classifies the input speech datum as either valid or invalid.

Suppose a set of labeled training points is (x_1, y_1), (x_2, y_2), …, (x_n, y_n). Each training point x_i belongs to either of two classes and is assigned a label, $y_i \in \{-1,\ 1\}$ for $i = 1,2,\ldots,n$. Based on these training data, the hyperplane is

$$w \cdot x + b = 0, \tag{1}$$

which is defined by the pair (w,b), such that the point x_i can be separated according to the function

$$f(x_i) = \mathrm{sign}(w \cdot x_i + b) = \begin{cases} 1, & \text{if } y_i = 1, \\ -1, & \text{if } y_i = -1. \end{cases} \tag{2}$$

The set S is linearly separable if a pair (w,b) exists such that the inequalities

$$\begin{cases} (w \cdot x_i + b) \geq 1, & \text{if } y_i = 1, \\ (w \cdot x_i + b) \leq -1, & \text{if } y_i = -1, \end{cases} \qquad i = 1,2,\ldots,n, \tag{3}$$

are valid for all elements of set S. Equation (3) can be rewritten as one set of inequalities as follows:

$$y_i(w \cdot x_i + b) - 1 \geq 0, \qquad \forall i. \tag{4}$$

If the set S is linearly separable, a unique optimal hyperplane exists, and for this hyperplane, the margin between the projections of the training points of two different classes is maximized. Conversely, when encountering the case that the set S is not linearly separable, Eq. (4) will then be modified to allow classification violations in the SVM formulation. In this case, some non-negative variable $\xi_i \geq 0$ will be introduced to generalize Eq. (4) for dealing with data that are not linearly separable:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \qquad i = 1, 2, \ldots, n. \tag{5}$$

When performing the SVM speaker verification procedure, Eq. (5) can be used to indicate the signal x_i as one of the two sound classes, the valid speaker class, and the imposter class. When the point x_i does not satisfy Eq. (4) where an error of misclassification occurs, the corresponding ξ_i in Eq. (5) for the point x_i must be the value of nonzero to satisfy the SVM classification inequality. The term $\sum_{i=1}^{n} \xi_i$ can be thus viewed as some measure of the number of training errors. The objective function to be minimized to find an optimal solution for Eq. (4), $\|w^2\|/2$, will be changed by assigning an extra cost for misclassification errors. Equation (6) shows such the modified solution to the optimal hyperplane problem of the non-linear separable condition as follows:

$$\begin{aligned}
&\text{minimize } \frac{1}{2} w \cdot w + C \sum_{i=1}^{n} \xi_i \\
&\text{subject to } y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, n, \\
&\xi_i \geq 0, \quad i = 1, 2, \ldots, n,
\end{aligned} \tag{6}$$

where C denotes a constant.

Note that the parameter C in Eq. (6) is a tunable parameter and could be chosen by the user. The parameter C is the only free parameter in SVM formulations. When a larger value of C is given, a higher penalty to misclassification errors is assigned. The value of the adjustable parameter C will directly decide the degree of balance between margin maximization and classification violation. The characteristics of the dissimilarity degree of SVM training data could be viewed as important clue to tune the value of the parameter C to an optimal one. The presented FLICSVM by using fuzzy logic-based intelligent control to tune the parameter C of conventional SVM will be introduced in the following section.

3. DEVELOPED FLICSVM WITH SUPPORTS OF GMM

In SVM-based speaker verification, the quality of the training data for establishing SVM models will largely affect the classification accuracy of the trained SVM classifier. When encountering substandard training data on SVM designs, the classification accuracy of the estimated SVM separation hyperplane will be extremely doubtful. To tackle this issue, this section presents the improved SVM method, FLICSVM, which needs the support of GMM additionally to analyze the training data of SVM. FLICSVM is performed under fuzzy regulations and involves three types of fuzzy control operations, EDoMV-driven fuzzy control, LLR-driven fuzzy control and EDoMV&LLR-driven fuzzy control.

3.1. Prior Information from GMM for Fuzzy Inference

GMM models are adopted in the evaluation of the distinguishability of two classes of data that trains the SVM classifier. Mathematically, a GMM is a weighted sum of M Gaussians [9, 10], denoted as

$$\lambda = \{w_i, \mu_i, \Sigma_i\}, \quad i = 1, 2, \ldots, M, \quad \sum_{i=1}^{M} w_i = 1, \tag{7}$$

where w_i is the weight, μ_i is the mean and Σ_i is the covariance. In this work, two parameter sets, λ_1 and λ_2 (two GMM models, that is), for representing the acoustic characteristics of valid speakers and imposters are determined, respectively.

The distinction between valid speakers' and imposters' data for training SVM could then be viewed as an uncertain measure and calculated by estimating the likelihood of the training data on parameter sets λ_1 and λ_2. Consider that the likelihood calculation is operated with a segment of training data composed of both valid speakers' and imposters' speech utterances, covering n acoustic feature vectors of D dimensions, $X = \{x_i \mid i = 1, 2, \ldots, n\}$, combined with two GMM acoustic models, λ_1 for normal speakers (valid speakers) and λ_2 for abnormal speakers (imposters). The estimated likelihood score that an observation frame x_i belongs to a GMM of class $Z \in \{\lambda_1, \lambda_2\}$ is then

$$L(x_i \mid Z) = \sum_{j=1}^{M} w_j \cdot \frac{1}{(2\pi)^{D/2} \cdot \left|\Sigma_j\right|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x_i - \mu_j)^T (\Sigma_j)^{-1} (x_i - \mu_j)\right\}. \tag{8}$$

Then the evaluation of the training data segment X containing n frames is done by comparing X with both the valid speaker model and the imposter model, i.e. λ_1 and λ_2. For the overall training frames X, the score is locally computed and then accumulated as log-likelihood ratio (LLR):

$$LLR(X) = \log \frac{L(X \mid \lambda_1)}{L(X \mid \lambda_2)} = \log(L(X \mid \lambda_1)) - \log(L(X \mid \lambda_2)) = \sum_{i=1}^{n} \log L(x_i \mid \lambda_1) - \log L(x_i \mid \lambda_2). \tag{9}$$

The calculated LLR score for the training data segment X indicate whether the difference between valid speakers' data and imposters' data is distinct. A large value of LLR implies that it is easy to separate valid speakers' data from imposters' data. Conversely, these two categories of training data are vague with a relatively small LLR value.

Although the abovementioned LLR index could be used to evaluate the distinguishability degree of two classes of training data of the SVM classifier, additional calculations in Eqs. (8) and (9) are required. In addition, both the valid GMM speaker model and the imposter GMM model, i.e. λ_1 and λ_2 in Eq. (9), need to be established in advance. Such the GMM training work usually also need to acquire a large amount of training data. A simple and direct way to assess the distinguishability degree of two classes of training data is to carry out the dissimilarity measure in two GMM distributions of training data. A natural measure between two Gaussian distributions, λ_r and λ_s for example, is to derive the Euclidean distance of mean vectors of two Gaussian distributions. The derivation formula of the Euclidean distance of mean vectors, μ_r and μ_s, which are corresponding to two GMM models, valid speaker model λ_r and imposter model λ_s of the overall training frames X, respectively, is as follows:

$$\mathrm{EDoMV}(X) = \sqrt{(\mu_{r_1} - \mu_{s_1})^2 + (\mu_{r_2} - \mu_{s_2})^2 + \cdots + (\mu_{r_n} - \mu_{s_n})^2}. \tag{10}$$

The value of the derived Euclidean distance of two mean vectors for the training data segment, X, imply the degree of dissimilarity between valid speakers' data and imposters' data. A large value of EDoMV implies that it is easy to separate valid speakers' data from imposters' data. Conversely, these two classes of training data are ambiguous with a relatively small EDoMV value. Figure 1 illustrates the index EDoMV between valid speaker and imposter Gaussian distributions.

3.2. Enhancing SVM by EDoMV-Driven Fuzzy Control

As already noted, the EDoMV index can be used as the key to the control of the free parameter C of SVM and, as a result, a rule set of seven IF-THEN rules is designed to fulfill the desired requirements as follows:

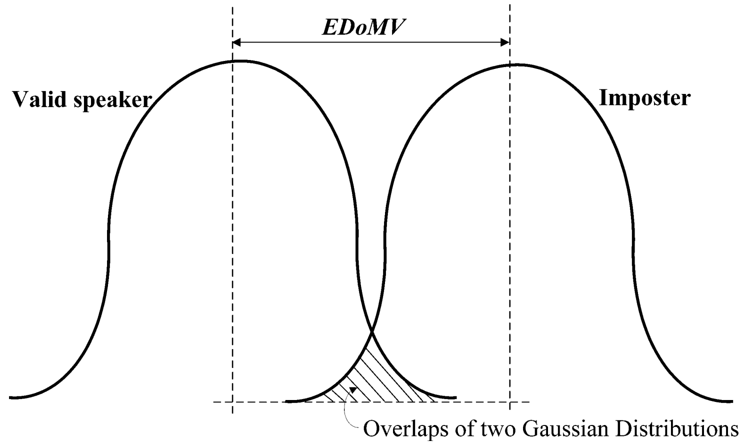

Fig. 1. The index EDoMV between valid speaker and imposter Gaussian distributions.

Rule 1: If EDoMV is large, then parameter C of SVM is small.
Rule 2: If EDoMV is rather large, then parameter C of SVM is rather small.
Rule 3: If EDoMV is slightly large, then parameter C of SVM is slightly small.
Rule 4: If EDoMV is medium, then parameter C of SVM is medium.
Rule 5: If EDoMV is slightly small, then parameter C of SVM is slightly large.
Rule 6: If EDoMV is rather small, then parameter C of SVM is rather large.
Rule 7: If EDoMV is small, then parameter C of SVM is large.

For translating lingual statements into quantitative expressions for computation, fuzzy techniques quite naturally come into use. Here a specific type of fuzzy logic control mechanism by Takagi-Sugeno (T-S hereafter) [6] is employed. EDoMV-driven fuzzy control formulation is explained in the following.

Let $M_1(\text{EDoMV})$, $M_2(\text{EDoMV})$, $M_3(\text{EDoMV})$, $M_4(\text{EDoMV})$, $M_5(\text{EDoMV})$, $M_6(\text{EDoMV})$ and $M_7(\text{EDoMV})$ be membership functions associated respectively with large, rather large, slightly large, medium, slightly small, rather small and small values of EDoMV. Also let functions $f_1(EDoMV)$, $f_2(EDoMV)$, $f_3(EDoMV)$, $f_4(EDoMV)$, $f_5(EDoMV)$, $f_6(EDoMV)$, and $f_7(EDoMV)$ set small, rather small, slightly small, medium, slightly large, rather large and large values of C respectively in each of the seven cases. The previous set of rules can then be further clarified as:

Rule 1: If EDoMV is $M_1(\text{EDoMV})$, then $C = f_1(\text{EDoMV})$,
Rule 2: If EDoMV is $M_2(\text{EDoMV})$, then $C = f_2(\text{EDoMV})$,
Rule 3: If EDoMV is $M_3(\text{EDoMV})$, then $C = f_3(\text{EDoMV})$,
Rule 4: If EDoMV is $M_4(\text{EDoMV})$, then $C = f_4(\text{EDoMV})$,
Rule 5: If EDoMV is $M_5(\text{EDoMV})$, then $C = f_5(\text{EDoMV})$,
Rule 6: If EDoMV is $M_6(\text{EDoMV})$, then $C = f_6(\text{EDoMV})$,
Rule 7: If EDoMV is $M_7(\text{EDoMV})$, then $C = f_7(\text{EDoMV})$.

The final fuzzy system defuzzified output is as follows [6]:

$$C = \frac{\sum_{i=1}^{7} M_i(\text{EDoMV}) \cdot f_i(\text{EDoMV})}{\sum_{i=1}^{7} M_i(\text{EDoMV})} \qquad (\text{EDoMV} - \text{driven FLICSVM}). \tag{11}$$

3.3. Improvements of SVM by LLR-Driven Fuzzy Control

As mentioned, the index LLR is an alternative opinion for determining the value of the free parameter C if one discards additional calculations of log-likelihood scores between speech frames and GMM speaker modes. A rule base of five regulations is designed to achieve the purpose of appropriately adjusting the parameter C of SVM as follows:

Rule 1: If LLR is large, then parameter C of SVM is small.
Rule 2: If LLR is rather large, then parameter C of SVM is rather small.
Rule 3: If LLR is medium, then parameter C of SVM is medium.
Rule 4: If LLR is rather small, then parameter C of SVM is rather large.
Rule 5: If LLR is small, then parameter C of SVM is large.

Within the framework of the fuzzy process, the statements of linguistic terms with uncertainty to some degree can be formulated in quantized forms for subsequent computations. The formulation of the above implications is given as a set of five fuzzy IF-THEN fuzzy rules and the system output C. The developed fuzzy control formulation for SVM, called LLR–Driven FLICSVM, is explained in the following.

Let $A_1(\mathrm{LLR})$, $A_2(\mathrm{LLR})$, $A_3(\mathrm{LLR})$, $A_4(\mathrm{LLR})$, and $A_5(\mathrm{LLR})$ be membership functions associated respectively with large, rather large, medium, rather small and small values of LLR. Also let functions $p_1(\mathrm{LLR})$, $p_2(\mathrm{LLR})$, $p_3(\mathrm{LLR})$, $p_4(\mathrm{LLR})$, and $p_5(\mathrm{LLR})$ set small, rather small, medium, rather large and large values of C respectively in each of the five cases. The previous set of rules can then be further clarified as:

Rule 1: If LLR is $A_1(\mathrm{LLR})$, then $C = p_1(\mathrm{LLR})$,
Rule 2: If LLR is $A_2(\mathrm{LLR})$, then $C = p_2(\mathrm{LLR})$,
Rule 3: If LLR is $A_3(\mathrm{LLR})$, then $C = p_3(\mathrm{LLR})$,
Rule 4: If LLR is $A_4(\mathrm{LLR})$, then $C = p_4(\mathrm{LLR})$,
Rule 5: If LLR is $A_5(\mathrm{LLR})$, then $C = p_5(\mathrm{LLR})$.

The final fuzzy system defuzzified output C is as follows [6]:

$$C = \frac{\sum_{i=1}^{5} A_i(\mathrm{LLR}) \cdot p_i(\mathrm{LLR})}{\sum_{i=1}^{5} A_i(\mathrm{LLR})} \qquad \text{(LLR-driven FLICSVM).} \tag{12}$$

3.4. Fuzzy Reasoning by Both EDoMV and LLR for SVM

EDoMV-driven FLICSVM and LLR-driven FLICSVM methods could tune the parameter C of SVM according to the derived values of EDoMV and LLR respectively under fuzzy reulations. However, the tunning accuracy of the parameter C is a little somewhat dissatisfactory, which could be further improvemed using fuzzy reasoning by both of two indexes EDoMV and LLR.

A rule base with 25 fuzzy implications is given to govern the regulation of C under the circumstance of two antecedents (EDoMV and LLR), as follows:

Rule 1: If EDoMV is large and LLR is large, then C is set to be small,
Rule 2: If EDoMV is large and LLR is rather large, then C is set to be rather small,
$\vdots$
Rule 13: If EDoMV is medium and LLR is medium, then C is set to be medium,
$\vdots$
Rule 24: If EDoMV is small and LLR is rather small, then C is set to be rather large,
Rule 25: If EDoMV is small and LLR is small, then C is set to be large.

Let $U_1(\mathrm{EDoMV})$, $U_2(\mathrm{EDoMV})$, …, and $U_5(\mathrm{EDoMV})$ be the membership functions associated respectively with large, rather large, …, and small values of EDoMV, and $V_1(\text{LLR})$, $V_2(\text{LLR})$, …, and $V_5(\text{LLR})$ be the membership functions associated respectively with large, rather large, …, and small values of LLR.

In addition, let the functions $q_1(\text{EDoMV},\text{LLR})$, $q_2(\text{EDoMV},\text{LLR})$, …, and $q_{25}(\text{EDoMV},\text{LLR})$ set small, rather small, …, and large values of C respectively in each of the 25 cases. The previous set of rules can then be further clarified as:

Rule 1: If EDoMV is $U_1(\text{EDoMV})$, and LLR is $V_1(\text{LLR})$, then $C = q_1(\text{EDoMV},\text{LLR})$,
Rule 2: If EDoMV is $U_1(\text{EDoMV})$, and LLR is $V_2(\text{LLR})$, then $C = q_2(\text{EDoMV},\text{LLR})$,
$\vdots$
Rule 25: If EDoMV is $U_5(\text{EDoMV})$, and LLR is $V_5(\text{LLR})$, then $C = q_{25}(\text{EDoMV},\ \text{LLR})$.

And the final system output, as follows [6]:

$$C = \frac{\sum_{i=1}^{25} w^i \cdot q_i(\text{EDoMV},\text{LLR})}{\sum_{i=1}^{25} w^i} \qquad \text{(EDoMV\&LLR-driven FLICSVM)}, \tag{13}$$

where

$$w^1 = U_1(\text{EDoMV}) \cdot \mathrm{V}_1(\text{LLR}), \mathrm{w}^2 = \mathrm{U}_1(\text{EDoMV}) \cdot \mathrm{V}_2(\text{LLR}), \ldots, \mathrm{w}^{25} = \mathrm{U}_5(\text{EDoMV}) \cdot \mathrm{V}_5(\text{LLR}). \tag{14}$$

4. EXPERIMENTS AND RESULTS

All the uttered data were recorded in an office by a close-talking microphone. The speech signal was sampled at 44.1 kHz and recorded at the settings of the mono channel and the 8-bit resolution. The analysis frames were 20-ms wide with a 10-ms overlap. For each frame, a 10-dimensional feature vector was extracted. The feature vector for each frame was a 10-dimensional cepstral vector. In the training phase, the training data for establishing SVM were collected from 27 male speakers where 13 speakers were chosen as the valid speakers and the other 14 speakers were imposters. Each of the 27 speakers was asked to offer 20 utterances of his names in Mandarin as the training data, totally 540 training utterances. Two GMM models that represent the valid speakers and the imposters were also built for the index LLR determination. Furthermore, T-S fuzzy models of EDoMV-driven, LLR-driven and EDoMV&LLR-driven FLICSVM were built up with the support of the MATLAB FIS toolbox. In the recognition phase, each of the 27 speakers in the training phase was again requested to provide 20 utterances of his names in Mandarin as test data, which were divided into 27 test databases, DB-1 to DB-27, each of which contained 20 utterances from some specific speaker.

Table 1 shows the recognition performance of the conventional SVM with various values of the parameter C. Observed from Table 1, in the average instance involving 27 test databases, the SVM model with the parameter $C = 16$ has the highest recognition rate of 72.96% and is therefore chosen for recognition performance comparisons with the proposed fuzzy-based enhancing SVM method. Note that a special case with parameter C set to be zero will have the worst recognition accuracy, which just achieves 62.68%.

Table 2 shows the recognition accuracy comparisons of the conventional SVM, the proposed EDoMV-driven, LLR-driven and EDoMV&LLR-driven FLICSVM. As indicated in Table 2, it is observed that the proposed EDoMV&LLR-driven FLICSVM approach demonstrated the highest average recognition rate of 78.88%, followed by LLR-driven FLICSVM and then EDoMV-driven FLICSVM. In comparison with the developed FLICSVM methods, conventional SVM achieved the most dissatisfactory recognition performance. Experimental results demonstrate the effectiveness of SVM with a tunable parameter C mechanism. All of the developed EDoMV-driven, LLR-driven and EDoMV&LLR-driven FLICSVM methods are superior to the conventional SVM method in recognition accuracy due to well-designed fuzzy logic-based intelligent control schemes for parameter C adjustments. Among these three FLIC-regulated SVM approaches,

Table 1. Recognition performances of conventional SVM with various values of parameter C.

	Recognition rates (%)				
Testing	Parameter C				
data set	0	16	128	256	2048
DB-1	30	100	100	100	100
DB-2	0	80	80	80	100
DB-3	0	60	100	100	60
…	…	…	…	…	…
DB-13	100	95	100	100	100
DB-14	50	85	80	80	85
DB-15	100	100	95	100	100
…	…	…	…	…	…
DB-25	50	50	50	50	65
DB-26	80	65	10	10	65
DB-27	80	40	40	40	40
Avg.	62.68	**72.96**	65.37	65.93	70.11

Table 2. Recognition accuracy comparisons of the proposed EDoMV-driven FLICSVM, LLR-driven FLICSVM, EDoMV&LLR-driven FLICSVM and the conventional SVM.

	Recognition rates (%)			
	Methods for speaker verification			
Testing data set	EDoMV FLICSVM	LLR FLICSVM	EDoMV&LLR FLICSVM	Conventional SVM ($C = 16$)
DB-1	100	100	100	100
DB-2	80	80	80	80
DB-3	80	80	80	60
…	…	…	…	…
DB-13	100	100	100	95
DB-14	80	80	95	85
DB-15	100	100	100	100
…	…	…	…	…
DB-25	75	75	75	50
DB-26	100	100	90	65
DB-27	40	40	40	40
Avg.	73.51%	77.59%	**78.88%**	72.96

EDoMV&LLR-driven FLICSVM performs best since the FLIC in EDoMV&LLR-driven FLICSVM uses two indexes EDoMV and LLR to be inputs for accurately reasoning the value of parameter C.

In this work, MATLAB fuzzy toolbox is used for online simulations. Fig. 2 shows fuzzy reasoning process of the fuzzy control scheme of developed EDoMV&LLR-driven FLICSVM by MATLAB fuzzy toolbox. It can be observed from Fig. 2 that the values of two input indexes EDoMV and LLR are 0.15 and 1200 respectively, and the free parameter C of SVM using such the EDoMV and LLR settings as the input of the fuzzy model was inferred to be the value of 28.2. Note that the value of parameter C (the output) will be updated when the input values of EDoMV and LLR are changed (two vertical red lines in Fig. 2 moved). Online simulation testing experiments using MATLAB fuzzy toolbox show the efficiency and effectiveness of the proposed EDoMV-driven, LLR-driven and EDoMV&LLR-driven FLICSVM in an online practical speaker recognition application.

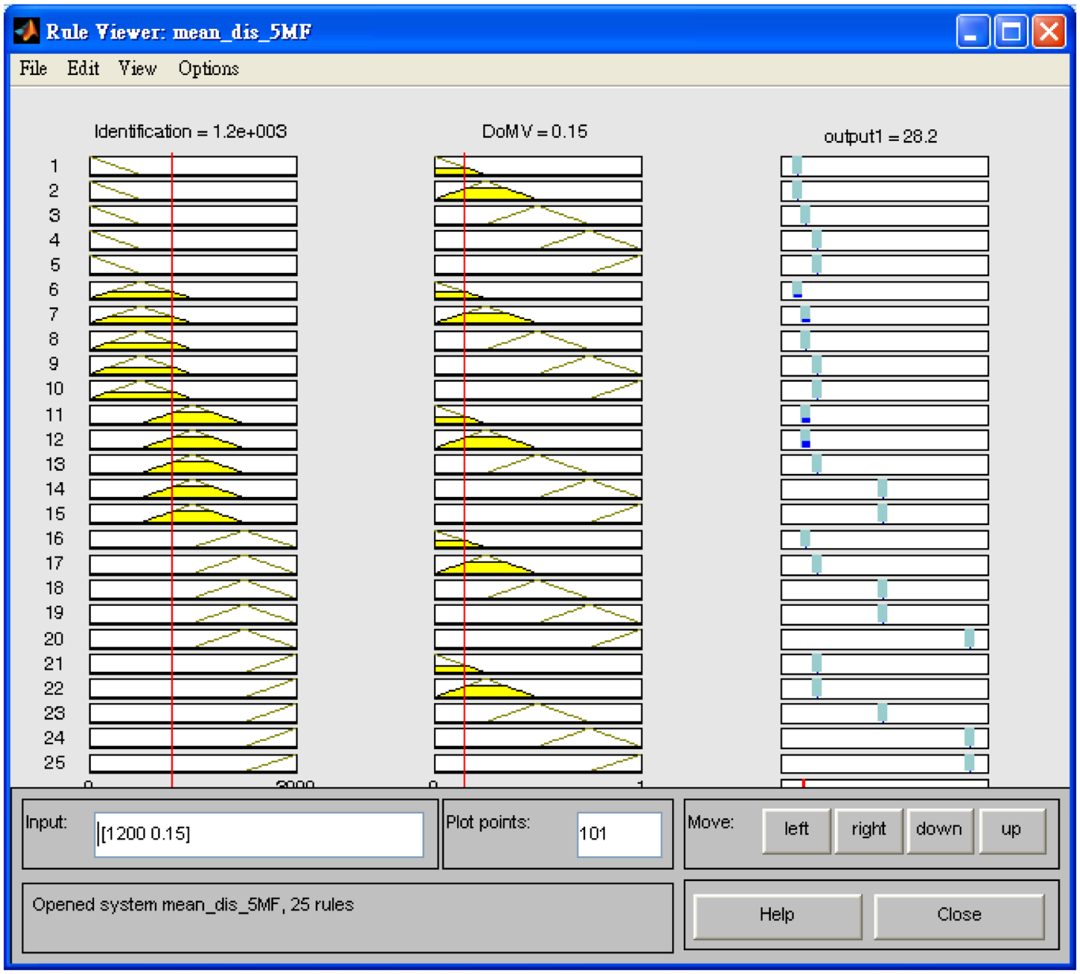

Fig. 2. Fuzzy reasoning of EDoMV&LLR-driven fuzzy control using the MATLAB fuzzy tool box.

5. CONCLUSIONS

This paper proposes an FLISSVM method to improve the performance of conventional SVM. The developed FLICSVM utilizes fuzzy logic-based intelligent control to tune the only free parameter *C* of SVM according to the dissimilarity degree of the acquired two categories of training data, valid speakers and imposters in speaker verification applications, in the establishment stage of SVM. The presented FLICSVM involves three types of fuzzy control schemes, EDoMV-driven fuzzy control, LLR-driven fuzzy control and EDoMV&LLR-driven fuzzy control. Experimental results in speaker recognition showed that all of three developed fuzzy-based SVM enhancement approaches outperform the conventional SVM, and EDoMV&LLR-driven FLICSVM has the best performance in recognition accuracy.

ACKNOWLEDGEMENT

This research is partially supported by the National Science Council (NSC) in Taiwan under grant NSC 101-2221-E-150-084.

REFERENCES

1. Burges, C.J.C., "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp. 121–167, 1998.
2. McLaren, M., Vogt, R., Baker, B. and Sridharan, S., "Data-driven background dataset selection for SVM-based speaker verification", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 6, pp. 1496–1506, 2010.
3. Boujelbene, S.Z., Mezghani, D.B.A. and Ellouze, N., "Improving SVM by modifying kernel functions for speaker identification task", *International Journal of Digital Content Technology and its Applications*, Vol. 4, No. 6, pp. 100–105, 2010.
4. Dong, J.X., Krzyzak, A. and Suen, C.Y., "Fast SVM training algorithm with decomposition on very large datasets", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, pp. 603–618, 2005.
5. Hsu, W.C. and Yu, T.Y., "Support vector machines parameter selection based on combined Taguchi method and Staelin method for e-mail spam filtering", *International Journal of Engineering and Technology Innovation*, Vol. 2, No. 2, pp. 113–125, 2012.
6. Zimmermann, H.-J., *Fuzzy Set Theory and Its Applications*, 3rd edn., Kluwer Academic, 1996.
7. Soliman, M.A., Kaldas, M.M.S., Barton, D.C. and Brooks, P.C., "Fuzzy-skyhook control for active suspension systems applied to a full vehicle model", *International Journal of Engineering and Technology Innovation*, Vol. 2, No. 2, pp. 85–96, 2012.
8. Ding, I.J., "Speech recognition using variable-length frame overlaps by intelligent fuzzy control", *Journal of Intelligent & Fuzzy Systems*, Vol. 25, No. 1, pp. 49–56, 2013.
9. Reynolds, D.A. and Rose, R.C., "Robust text-independent speaker identification using Gaussian mixture speaker models", *IEEE Transactions on Speech and Audio Processing*, Vol. 3, No. 1, pp. 72–83, 1995.
10. Kenny, P., Boulianne, G., Ouellet, P. and Dumouchel, P., "Speaker and session variability in GMM-based speaker verification", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 4, pp. 1448–1460, 2007.