A Combination-of-Tools Method for Learning Interpretable Fuzzy Rule-based Classifiers from Support Vector Machines

Tamas Kenesei, Johannes A. Roubos, Janos Abonyi

Department of Process Engineering, University of Pannonia
Veszprem P.O.Box 158, H-8201 Hungary
email: abonyij@fmt.uni-pannon.hu
www.fmt.uni-pannon.hu/softcomp

Abstract. A new approach is proposed for the data-based identification of transparent fuzzy rule-based classifiers. It is observed that fuzzy rule-based classifiers work in a similar manner as kernel function-based support vector machines (SVMs) since both model the input space by nonlinearly maps into a feature space where the decision can be easily made. Accordingly, trained SVM can be used for the construction of fuzzy rule-based classifiers. However, the transformed SVM does not automatically result in an interpretable fuzzy model because the SVM results in a complex rule-base, where the number of rules is approximately 40-60% of the number of the training data. Hence, reduction of the SVM-initialized classifier is an essential task. For this purpose, a three-step reduction algorithm is developed based on the combination of previously published model reduction techniques. In the first step, the identification of the SVM is followed by the application of the Reduced Set method to decrease the number of kernel functions. The reduced SVM is then transformed into a fuzzy rule-based classifier. The interpretability of a fuzzy model highly depends on the distribution of the membership functions. Hence, the second reduction step is achieved by merging similar fuzzy sets based on a similarity measure. Finally, in the third step, an orthogonal least-squares method is used to reduce the number of rules and re-estimate the consequent parameters of the fuzzy rule-based classifier. The proposed approach is applied for the Wisconsin Breast Cancer, Iris and Wine classification problems to compare its performance to other methods.

Key words: Classification, Fuzzy classifier, Support Vector Machine, Model Reduction

1 Introduction

Fuzzy logic helps to improve the interpretability of knowledge-based classifiers through its semantics that provide insight in the classifier structure and decision making process. Application of SVM methods for fuzzy logic is not a completely new idea. SVMs based on sigmoidal kernels are functionally equivalent to feed-forward neural networks. When Gaussian kernel functions are applied, the SVM can be used to initialize Radial Basis Function (RBF) classifiers [8]. In [9] a support vector regression model was used to identify a dynamic process by a RBF network. A restricted class of fuzzy systems is functionally equivalent to RBF models [11]. These SVM-based approaches can also be applied to identify fuzzy systems as shown in [10]. However, only a simple example for the approximation of an univariate function was presented, while interpretability and applicability issues were not discussed.

An example for the incorporation of fuzzy logic into SVM can be found in [12], where fuzzy sigmoid functions has been used as a kernel in the SVM framework. An another interesting example can be found in [13], where the concept of fuzzy regression has been adopted. Another approach for the fusion of fuzzy and SVM based modeling techniques us proposed in [14], where fuzzy membership values are obtained from fuzzy clustering have been applied to form ..ed input variables of the SVM algorithm. The linguistic interpretability of the kernel functions of SVMs has been already recognized in [15]. It has been showed that the kernel matrix may be interpreted in terms of linguistic values based on the premises of if–then rules. A simple real–life application example for SVM based fuzzy modeling has been presented in [16], where the SVM based fuzzy model is used for modeling the arc welding process. Though simple SVM models can handle with a variety of classification tasks, fuzzy support vector machines have been investigated for a period of time to make classification become more effective Lin and Wang (2002) first proposed a prototype of fuzzy SVM (FSVM), where one applies a fuzzy membership function to each input data of SVM. This kind of fuzzy SVM has been applied in [17] for text categorization.

The main advantage of rule-based fuzzy rule-based classifiers over SVM is the transparency and the linguistic interpretability. Fuzzy logic, however, does not guarantee interpretability [1, 3]. Hence real effort must be made to keep the resulting rule-base transparent [2, 4, 5]. For this purpose, two main approaches are followed in the literature: (i) selection of a low number of input variables (features) in order to create a compact classifier [6], and (ii) construction of a large set of possible rules by using all inputs, and subsequently use this set to make a useful selection out of these rules [4, 7]. In both approaches, further model reduction can be realized by generalization and/or similarity-based set-reduction techniques [3, 6]. However, if the fuzzy model is obtained from a SVM, the resulted initial fuzzy model is usually not very interpretable. This is because the SVM results in a complex rule-base where the number of rules is approximately 40-60% of the number of the training data. Therefore, fuzzy models constructed from SVM are presumably more complex than necessary. This suggests that a simple transformation of a SVM into a fuzzy model should be followed by rule-base simplification steps. Hence, to obtain a parsimonious and interpretable fuzzy rule-based classifier a three-step model reduction algorithm is proposed:

1. *Application of the Reduced Set method*
   The identification of the SVM is followed by the application of the Reduced Set (RS) method to decrease the number of kernel functions. Originally, this method has been introduced by [18] to reduce the computational complexity of SVMs. The obtained SVM is subsequently transformed into a fuzzy rule-based classifier.
2. *Similarity-based fuzzy set merging*
   The Gaussian membership functions of the fuzzy rule-based classifier are derived from the Gaussian kernel functions of the SVM. The interpretability of a fuzzy model highly depends on the distribution of the membership functions. Hence, the next reduction step is achieved by merging fuzzy sets based on a similarity measure [3].
3. *Rule-base simplification by orthogonal transformations*
   Finally, an orthogonal least-squares method is used to reduce the number of rules and re-estimate the consequent parameters of the classifier. The application of orthogonal transforms for reducing the number of rules has received much attention in the recent literature [19, 20]. These methods evaluate the output contribution of the rules to obtain the order of importance. The less important rules are then removed according this ranking to further reduce the complexity and increase the transparency.

The remainder of this article is organized as follows. Section 2 explains the structure of the fuzzy rule-based classifier. In Section 3, the reduction of the classifier is discussed. First, the SVM learning method is reviewed. The other subsections deal with the similarity-based and orthogonal transform-based model simplification techniques. In Section 4, the proposed approach is experimentally evaluated for the two-class Wisconsin Breast Cancer classification problem and the three-class Iris and Wine classification problems. Finally, the conclusions are given in Section 5.

2 The Fuzzy Rule-based Classifier

2.1 The Structure of the Fuzzy Rule-based Classifier

Fuzzy rule-based models have been successfully applied to many pattern recognition and classification problems. The classical fuzzy rule-based classifier classifies an example according to the rule with the greatest degree of association. By using this reasoning method the information provided by the other rules is lost. In this paper a new fuzzy rule-based classifier structure is presented for the classification of N_c labeled classes.

One widely used approach to solve non-fuzzy N_c-class pattern recognition problems is to consider the general problem as a collection of binary classification problems. Accordingly, N_c classifiers can be constructed, i.e. one for each class.

The c–th classifier, $c = 1, \ldots, N_c$, separates class c from the $N_c - 1$ other classes. This one-against-all method results in a hierarchical classifier structure that allows for a sequential model construction and evaluation procedure. Based on this, we propose fuzzy rule-based classifier that consists of N_c fuzzy subsystems with a set of Takagi-Sugeno-type fuzzy rules [21] that describe the c–th class in the given data set as

$$R_i^c : \quad \textbf{If } x_1 \text{ is } A_{i1}^c \textbf{ and } \ldots x_n \text{ is } A_{in}^c \textbf{ then } y_i^c = \delta_i^c, \quad i = 1, \ldots, N_R^c \,, \tag{1}$$

where R_i^c is the i–th rule in the c–th fuzzy rule-based classifier and N_R^c denotes the number of rules. $A_{i1}^c, \ldots, A_{N_i}^c$ denote the antecedent fuzzy sets that define the operating region of the rule in the N_i-dimensional input space; $\mathbf{x} = [x_1, x_2, \ldots, x_{N_i}]^T$. The rule-consequent δ_i^c, $i = 1, \ldots, N_R^c$, is a crisp (non-fuzzy) number. The **and** connective is modeled by the product operator allowing for interaction between the propositions in the antecedent. Hence, the degree of activation of the i–th rule is calculated as $beta_i^c(\mathbf{x}) = \prod_{j=1}^{N_i} A_{ij}^c(x_j), \;\; i = 1, \ldots, N_R^c$.

The output of the classifier is determined by the following decision function

$$y^c = sgn\left(\sum_{i=1}^{N_R^c} \beta_i^c(\mathbf{x})\delta_i^c + b^c\right), \tag{2}$$

where b^c is a constant threshold. If $y^c = +1$ then the observation $\mathbf{x}$ is part of class c, otherwise, when $y^c = -1$, then it is not an item in class c.

For simplicity in the notation, the superscript c that denotes the index of the fuzzy subsystem is neglected in the sequel. It is not necessary there because the following sections only deal with the identification of the submodels that are independent from each other.

2.2 Formulation of the Fuzzy Rule-based Classifier as a Kernel Machine

Recent years have witnessed a surge of interest in learning methods based on Mercer kernels, i.e. functions $k(\mathbf{x}_i, \mathbf{x}_j)$ which for all data pairs $\{\mathbf{x}_1, \ldots, \mathbf{x}_{N_d}\} \subset \mathbb{R}^{N_i}$ give rise to positive matrices $K_{ij} := k(\mathbf{x}_i, \mathbf{x}_j)$, where N_d denotes the number of data points. Using k instead of a dot product in $\mathbb{R}^{N_i}$ corresponds to mapping the data into a possibly high-dimensional space F, by a usually nonlinear map $\phi : \mathbb{R}^{N_i} \to F$, and taking the dot product there $k(\mathbf{z}_i, \mathbf{x}) = (\phi(\mathbf{z}_i) \cdot \phi(\mathbf{x}))$ [18].

The main principle of kernel-based support vector classifiers is the identification of a linear decision boundary in this high-dimensional feature-space.A link to the fuzzy model structure will now be established. The fuzzy sets are represented in this paper by Gaussian membership functions

$$A_{ij}(x_j) = \exp\left(\frac{(x_j - z_{ij})^2}{2\sigma^2}\right) \; \beta_i(\mathbf{x}) = k(\mathbf{z}_i, \mathbf{x}) = \exp\left(\frac{\|\mathbf{x} - \mathbf{z}_i\|^2}{2\sigma^2}\right) . \tag{3}$$

The degree of fulfillment $\beta_i(\mathbf{x})$ can be written in a more compact form by using the Gaussian kernels

This kernel interpretation of fuzzy systems shows that fuzzy models are effective in solving nonlinear problems because they map the original input space into a nonlinear feature space by using membership functions similar to the SVM that utilizes kernel functions for this purpose.

3 Fuzzy Rule-based Classifier based SVM

In the previous section, a new type of fuzzy rule-based classifier structure has been presented that is formed as a set of binary classifiers. We continue with the identification of the separate fuzzy rule-based classifiers which means the determination of the number of rules, the parameters of the membership functions and the rule consequents. The known fuzzy identification methods do not estimate these parameters simultaneously, but apply heuristic tools like fuzzy clustering. However, this results in a suboptimal model structure, although it may also be subject to further optimization [5].

We aimed at a new identification method that quickly, at low computational costs, results in an optimal classifier that generalizes well. In this section we show how such fuzzy rule-based classifiers can be identified by means of SVMs. After a brief review on SVM learning, the model transformation and simplification techniques are discussed.

In the previous sections it has been shown how a SVM, that is structurally equivalent to a fuzzy model, can be identified. Unfortunately, this identification method cannot be used directly for the identification of interpretable fuzzy systems because the number of the support vectors is usually very large. Typical values are 40-60% of the number of training data which is in our approach equal to the number of rules in the fuzzy system. Therefore, there is a need for an interpretable approximation of the support vector expansion. For this purpose a step-wise algorithm will be introduced, where the first step is based on the recently published Reduced Set (RS) method developed for reducing the computational demand of the evaluation of SVMs [18].

3.1 Reduction of the Number of Fuzzy Sets

In the previous section, it has been shown how kernel-based classifiers with a given number of kernel functions N_R, can be obtained. Because the number of the rules in the transformed fuzzy system is identical to the number of kernels, it is extremely important to get a moderate number of kernels in order to obtain a compact fuzzy rule-based classifier.

From (3) it can be seen that the number of fuzzy sets in the identified model is $N_s = N_R N_i$. The interpretability of a fuzzy model highly depends on the distribution of these membership functions. With the simple use of (3), some of the membership functions may appear almost undistinguishable. Merging similar fuzzy sets reduces the number of linguistic terms used in the model and thereby increases the transparency of the model. This reduction is achieved by a rule-base simplification method [3, 4], based on a similarity measure $S(A_{ij}, A_{kj})$,

$i, k = 1, \ldots, n$ and $i \neq j$. If $S(A_{ij}, A_{kj}) = 1$, then the two membership functions A_{ij} and A_{kj} are equal. $S(A_{ij}, A_{kj})$ becomes 0 when the membership functions are non-overlapping. During the rule-base simplification procedure similar fuzzy sets are merged when their similarity exceeds a user-defined threshold $\theta \in [0, 1]$.

The set-similarity measure can be based on the set-theoretic operations of intersection and union $S(A_{ij}, A_{kj}) = \frac{|A_{ij} \cap A_{kj}|}{|A_{ij} \cup A_{kj}|}$, where $|.|$ denotes the cardinality of a set, and the $\cap$ and $\cup$ operators represent the intersection and union, respectively, or it can be based on the distance of the two fuzzy sets[3]. Here, the following expression was used to approximate the similarity between two Gaussian fuzzy sets [4]

$$S(A_{ij}, A_{kj}) = \frac{1}{1 + d(A_{ij}, A_{kj})} = \frac{1}{1 + \sqrt{(z_{ij} - z_{kj})^2 + (\sigma_{ij} - \sigma_{kj})^2}} \ . \tag{4}$$

3.2 Reduction of the Number of Rules by Orthogonal Transforms

By using the previously presented SVM identification and reduction techniques, the following fuzzy rule-based classifier has been identified

$$y = sgn\left(\sum_{i=1}^{N_R} \prod_{j=1}^{N_i} \exp\left(\frac{(x_j - z_{ij})^2}{2\sigma^2}\right) \delta_i + b\right) . \tag{5}$$

Because the application of the RS method and the fuzzy set merging procedure the obtained membership functions only approximate the original feature space identified by the SVM. Hence, the $\delta = [\delta_1, \ldots, \delta_{N_r}]^T$ consequent parameters of the rules have to be re-identified to minimize the difference between the decision function of the support vector machine and the fuzzy model (5)

$$MSE = \sum_{j=1}^{N_d} \left(\sum_{i=1}^{N_x} \gamma_i k(\mathbf{x}_j, \mathbf{x}_i) - \sum_{i=1}^{N_R} \delta_i \beta_i(\mathbf{x}_j) \right)^2 = \|\mathbf{y}_s - \mathbf{B}\delta\|^2 \ , \tag{6}$$

where the matrix $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_{N_R}] \in \mathbb{R}N_d \times N_R$ contains the firing strength of all N_R rules for all the inputs $\mathbf{x}_i$, where $\mathbf{b}_j = [\beta_j(\mathbf{x}_1), \ldots, \beta_j(\mathbf{x}_{N_d})]^T$. As the fuzzy rule-based classifier (5) is linear in the parameters δ, (6) can be solved by a least-squares method $\delta = \mathbf{B}^+ \mathbf{y}_s$, where $\mathbf{B}^+$ denotes the Moore-Penrose pseudo inverse of $\mathbf{B}$.

The application of orthogonal transforms for the above mentioned regression problem (6) for reducing the number of rules has received much attention in recent literature [19, 20]. These methods evaluate the output contribution of the rules to obtain an importance ordering. For modeling purposes, the Orthogonal Least Squares (OLS) is the most appropriate tool [19]. The OLS method transforms the columns of $\mathbf{B}$ into a set of orthogonal basis vectors in order to inspect the individual contribution of each rule. To do this, Gram-Schmidt orthogonalization of $\mathbf{B} = \mathbf{WA}$ is used, where $\mathbf{W}$ is an orhogonal matrix $\mathbf{W}^T\mathbf{W} = \mathbf{I}$ and $\mathbf{A}$ is an upper triangular matrix with unity diagonal elements. If $\mathbf{w}_i$ denotes the i–th column of $\mathbf{W}$ and g_i is the corresponding element of the OLS solution vector $\mathbf{g} = \mathbf{A}\delta$, the output variance $\mathbf{y}_s^T\mathbf{y}_s/N_d$ can be explained by the regressors $\sum_{i=1}^{N_r} g_i\mathbf{w}_i^T\mathbf{w}_i/N_d$. Thus, the error reduction ratio, ϱ, due to an individual rule i can be expressed as $\varrho^i = \frac{g_i^2\mathbf{w}_i^T\mathbf{w}_i}{\mathbf{y}_s^T\mathbf{y}_s}$. This ratio offers a simple mean for ordering the rules, and can be easily used to select a subset of rules in a forward-regression manner. Evaluating only the approximation capabilities of the rules, the OLS method often assigns high importance to a set of redundant or correlated rules. To avoid this, in [20] some extension for the OLS method were proposed.

4 Application Example

In order to examine the performance of the proposed identification method, firstly it was applied to the Wisconsin Breast Cancer data (WBCD)[1], which is a benchmark problem in the classification and pattern recognition literature. The data is divided into a training and an evaluation subset that have similar size and class distributions (We used 342 cases for training and 341 cases for testing the classifier).

First, the advanced version of C4.5 was applied to obtain an estimate for the useful features. This method gave 36 misclassification (5.25%) for this problem. The constructed decision-tree model had 25 nodes and used mainly three inputs; x_1, x_2 and x_6. Based on this pre-study, only the previous three inputs were applied to identify the SVM classifier with $N_x = 71$ support vectors. The application of this model resulted in 3 and 15 misclassifications on the training and testing data, respectively. This model has been reduced by the RS method, by which we tried to reduce the model by a factor of 10, $N_R = 8$. By this step, the classification performance slightly decreased on the training set to 12 misclassifications, but the validation data showed a slightly better result with 14 misclassifications. Next, the reduced kernel-classifier was transformed into a fuzzy system. Figure 1 shows the membership functions that were obtained. The obtained model with eight rules is still not really well interpretable; however, some of the membership functions appear very similar and can probably be merged easily without loss in accuracy. The performance of the classifier slightly increased after this merging step (Table 1). Subsequently, using the OLS method, the rules were ordered according to there importance. Then, we reduced the number of rules one-by-one according the OLS ranking, till a major drop in the performance was observed. To our surprise, only two rules and four membership functions were necessary to have a good classification performance on this problem: 14 and 16 misclassification on the learning and validation data, respectively (Table 1). The obtained rules are:

R_1: **If** x_1 is Small **and** x_2 is Small **and** x_6 is Small **then** Class is Benign;
R_2: **If** x_1 is High **then** Class is Malignant;

[1] The WBCD is available from the University of California, Irvine, URL: *http://www.ics.uci.edu/˜mlearn/*

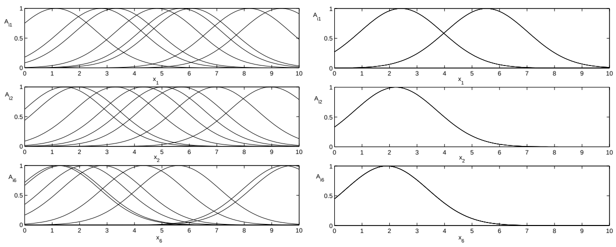

Fig. 1. Non-distinguishable membership functions obtained after the application of the RS method (left); Interpretable membership functions of the OLS reduced fuzzy model (right)

where x_1 is the clump thickness, x_2 the uniformity of cell size, and x_6 a measure for bare nuclei.

Table 1. Classification rates and model complexity for classifiers constructed using the proposed techniques.

Method	♯ Miss. Train (Accuracy)	♯ Miss. Test	♯ Rules	♯ Conditions
SVM	3 (99.1%)	15 (95.6%)	71	213
RS method	12 (96.5%)	14 (95.9%)	8	24
Membership merging	11 (96.8%)	13 (96.2%)	8	10
OLS	14 (95.9%)	16 (95.3%)	2	4

The previously presented illustrative example showed the details of how the application of the model reduction steps effect the model complexity and model performance. To give more, detailed analysis of the proposed additional benchmark datasets were also used. In the following the results of this analysis are only briefly presented. The IRIS and the Wine datasets are also taken from the UCI repository. It is interesting to note that these datasets do not define binary classification problems, hence, there was a need for the integration of three compact models identified based on the one-class from all strategy. Table 2 concludes these results. In these experiments the same parameterset was used, that points on that the proposed algorithm is robust, and its parameters can be tuned easily.

5 Conclusions

It has been shown in a mathematical way that SVMs and fuzzy rule-based classifiers work in a similar manner as both models maps the input space of the classifier into a feature space with the use of either nonlinear kernel or membership functions. The main difference between SVMs and fuzzy rule-based classifier

Table 2. Classification rates and model complexity for classifiers constructed using the proposed technique for the IRIS/WINE datasamples.

Method	♯ Miss. Train (Accuracy)	♯ Rules	♯ Conditions
SVM	4 (97%)/5 (97%)	264/117	1056/351
RS method	8 (95%)/7 (96%)	84/63	336/189
Membership merging	12 (92%)/21 (88%)	60/27	240/81
OLS	15 (90%)/12 (93%)	36/27	54/81

systems is that fuzzy systems have to fulfill two objectives simultaneously, i.e., they must provide a good classification performance and must also be linguistically interpretable, which is not an issue for SVMs. However, as the structure identification of fuzzy systems is a challenging task, the application of kernel-based methods for model initialization could be advantageous because of the high performance and the good generalization properties of these type of models. Accordingly, support vector-based initialization of fuzzy rule-based classifiers is proposed. First, the initial fuzzy model is derived by means of the SVM learning algorithm. Then the SVM is transformed into an initial fuzzy model that is subsequently reduced by means of the reduced set method, similarity-based fuzzy set merging, and orthogonal transform-based rule-reduction. Because these rule-base simplification steps do not utilize any nonlinear optimization tools, it is computationally cheap and easy to implement.

The application of the proposed approach is shown for the Wisconsin Breast Cancer classification problem. This classification study showed that a proper rule structure is obtained by the proposed model identification procedure. The obtained classifier is very compact but its accuracy is still comparable to the best results reported in the literature-based on nonlinear optimization tools. Besides, it might be clear that still real progress can be made in the development of novel methods for feature selection.

We consider this paper also as a case study for further developments in the direction of a combination–of–tools methodology for modelling and identification, aiming at models that perform well on multiple criteria, i.e, here different soft-computing tools, namely support vector machines and fuzzy techniques are combined to achieve a predefined trade-off between performance and transparency. In this sense, it is expected that the current work is not only useful for the identification of fuzzy classifiers but also provides insights in the understanding and analysis of SVM-based classifiers.

Acknowledgments. The authors would like to acknowledge the support of the Cooperative Research Centre (VIKKK, project III/1) and Hungarian Research Found (OTKA T049534). Janos Abonyi is grateful for the support of the Bolyai Research Fellowship of the Hungarian Academy of Sciences and the veges Fellowship.

References

1. J. Valente de Oliveira: Semantic constraints for membership function optimization, IEEE Trans. FS Vol. 19 (1999) 128–138
2. D. Nauck and R. Kruse: Obtaining interpretable fuzzy classification rules from medical data, Artificial Intelligence in Medicine Vol. 16 (1999) 149–169
3. M. Setnes and R. Babuška and U. Kaymak and H.R. van Nauta Lemke: Similarity measures in fuzzy rule base simplification, IEEE Trans. SMC-B, Vol. 28 (1998) 376–386
4. Y. Jin: Fuzzy Modeling of High-Dimensional Systems,IEEE Trans. FS Vol. 8 (2000) 212–221
5. M. Setnes and J.A. Roubos: GA-Fuzzy Modeling and Classification: Complexity and Performance, IEEE Trans. FS. Vol. 8 (2000) 509–522
6. J.A. Roubos and J. Abonyi and M. Setnes: Learning fuzzy classification rules from labeled data, International Journal of Information Sciences Vol. 150 (2003) 77–93
7. H. Ishibuchi and T. Nakashima and T. Murata: Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems, IEEE Trans. SMC–B Vol. 29 (1999) 601–618
8. B. Schlkopf and K. Sung and C. Burges and F. Girosi and P. Niyogi and T. Poggio and V. Vapnik:Comparing support vector machines with gaussian kernels to radial basis function classifiers, IEEE Trans Sign. Processing Vol. 45 (1997) 2758–2765
9. W.C. Chan and K.C. Cheung and C.J. Harris: Modelling of Nonlinear Dynamic Systems Using Support Vector Machines, Proc. of the IFAC Symposium on Artificial Intelligence, Budapest, Hungary (2000) 217–222
10. J. T. Jeng and T.T. Lee: Support Vector Machines for Fuzzy Neural Networks, Proc. of the IEEE SMC conference (1999) VI–115–120
11. J.-S.R. Jang and C.-T. Sun: Functional Equivalence Between Radial Basis Function Networks and Fuzzy Inference Systems, IEEE Trans. NN Vol. 4 (1993) 156–159
12. G. Camps–Valls, J.D. Martn–Guerrero, J.L. Rojo-lvarez, E. Soria–Olivas: Fuzzy sigmoid kernel for support vector classifiers, Neurocomputing 62 (2004) 501–506
13. Chih–Chia Yao, Pao-Ta Yu: Fuzzy regression based on asymmetric support vector machines, Applied Mathematics and Computation 182 175–193
14. Asli Celikyilmaz, I. Burham Trksen: Fuzzy Functions with Support Vector Machines, Accepted Manuscript to appear in Information Sciences
15. Jacek M. Leski: On support vector regression machines with linguistic interpretation of the kernel matrix, Fuzzy Sets and Systems 157 (2006) 1092–1113
16. Xixia Huang, Shanben Chen: SVM–based fuzzy modeling for the arc welding process, Material Science and Engineering A 427 (2006) 181–187
17. Tai–Yue Wang, Huei-min Chiang: Fuzzy support vector machine for multi–class text categorization, Information Processing and Management 43 (2007) 914–929
18. B. Schlkopf and S. Mika and C.J.C. Burges and P. Knirsch and K.-R. Mller and G. Rtsch and A. Smola: Input space vs. feature space in kernel-based methods, IEEE Trans. on Neural Networks, Vol. 10(5) (1999) 1000–1017
19. J. Yen and L. Wang.: Simplifying fuzzy rule-based models using orthogonal transformation methods, IEEE Trans. SMC-B Vol. 29 (1999) 13–24
20. M. Setnes and H. Hellendoorn: Orthogonal transforms for ordering and reduction of fuzzy rules, in FUZZ-IEEE 700-705,San Antonio, Texas, USA 2000
21. T. Takagi and M. Sugeno: Fuzzy identification of systems and its application to modeling and control, IEEE Trans. SMC Vol. 15 (1985) 116–132