



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Engineering and Information Sciences -
Papers: Part A

Faculty of Engineering and Information Sciences

2011

A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises

Xiaowei Yang

University of Technology Sydney

Guangquan Zhang

University of Technology Sydney, Guangquan.Zhang@uts.edu.au

Jie Lu

University of Technology Sydney, Jie.Lu@uts.edu.au

Jun Ma

University of Wollongong, jma@uow.edu.au

Publication Details

Yang, X., Zhang, G., Lu, J. & Ma, J. (2011). A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises. *IEEE Transactions on Fuzzy Systems*, 19 (1), 105-115.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:
research-pubs@uow.edu.au

A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises

Abstract

The support vector machine (SVM) has provided higher performance than traditional learning machines and has been widely applied in real-world classification problems and nonlinear function estimation problems. Unfortunately, the training process of the SVM is sensitive to the outliers or noises in the training set. In this paper, a common misunderstanding of Gaussian-function-based kernel fuzzy clustering is corrected, and a kernel fuzzy c-means clustering-based fuzzy SVM algorithm (KFCM-FSVM) is developed to deal with the classification problems with outliers or noises. In the KFCM-FSVM algorithm, we first use the FCM clustering to cluster each of two classes from the training set in the high-dimensional feature space. The farthest pair of clusters, where one cluster comes from the positive class and the other from the negative class, is then searched and forms one new training set with membership degrees. Finally, we adopt FSVM to induce the final classification results on this new training set. The computational complexity of the KFCM-FSVM algorithm is analyzed. A set of experiments is conducted on six benchmarking datasets and four artificial datasets for testing the generalization performance of the KFCM-FSVM algorithm. The results indicate that the KFCM-FSVM algorithm is robust for classification problems with outliers or noises.

Keywords

fuzzy, noises, kernel, support, vector, machine, algorithm, classification, problems, clustering-based, c-means, outliers

Disciplines

Engineering | Science and Technology Studies

Publication Details

Yang, X., Zhang, G., Lu, J. & Ma, J. (2011). A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises. *IEEE Transactions on Fuzzy Systems*, 19 (1), 105-115.

A Kernel Fuzzy c -Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises

Xiaowei Yang, Guangquan Zhang, Jie Lu, *Member, IEEE*, and Jun Ma

Abstract—The support vector machine (SVM) has provided higher performance than traditional learning machines and has been widely applied in real-world classification problems and nonlinear function estimation problems. Unfortunately, the training process of the SVM is sensitive to the outliers or noises in the training set. In this paper, a common misunderstanding of Gaussian-function-based kernel fuzzy clustering is corrected, and a kernel fuzzy c -means clustering-based fuzzy SVM algorithm (KFCM-FSVM) is developed to deal with the classification problems with outliers or noises. In the KFCM-FSVM algorithm, we first use the FCM clustering to cluster each of two classes from the training set in the high-dimensional feature space. The farthest pair of clusters, where one cluster comes from the positive class and the other from the negative class, is then searched and forms one new training set with membership degrees. Finally, we adopt FSVM to induce the final classification results on this new training set. The computational complexity of the KFCM-FSVM algorithm is analyzed. A set of experiments is conducted on six benchmarking datasets and four artificial datasets for testing the generalization performance of the KFCM-FSVM algorithm. The results indicate that the KFCM-FSVM algorithm is robust for classification problems with outliers or noises.

Index Terms—Classification, fuzzy c -means (FCM), fuzzy support vector machine (FSVM), high-dimensional feature space, kernel clustering, outliers or noises.

I. INTRODUCTION

THE support vector machine (SVM) is an important methodology for classification problems [10], [37], [44] and nonlinear function estimation problems [19] in the fields of pattern recognition and machine learning. The researchers have presented some important SVMs, such as Lagrangian SVM

(LSVM) [39], ν -SVM [48], and least-squares SVM (LS-SVM) [50], and obtained some bounds of classification errors [54]. In many real-world applications, such as financial time-series forecasting [5], image classification [6], [8], bioinformatics [9], face recognition [11], [29], and face authentication [25], [43], the SVM and its hybrid algorithms have been shown to yield higher performance. The SVM first maps the input points into a high-dimensional feature space and then constructs a separating hyperplane that maximizes the margin between two classes in this space. With no knowledge of the mapping, the SVM finds the optimal hyperplane by using the dot product functions, called kernels, in the high-dimensional feature space [53]. The solution of the optimal hyperplane can be written as a combination of a few input points that are called support vectors.

One of the main drawbacks in the application of the standard SVM is that the training process of the SVM is sensitive to the outliers or noises in the training dataset due to overfitting [64]. In many real-world classification problems, neither training point exactly belongs to any of the two classes when the outliers or noises exist. For example, one training point may belong 95% to one class and be 5% meaningless, and it may also belong 30% to one class and be 70% meaningless. This uncertainty results in some training points being more important than others for making decision. Considering that fuzzy approaches have been widely applied to uncertain problems, Huang and Liu [22] and Lin and Wang [35] have proposed fuzzy SVMs (FSVM) independently based on the unilateral weights and standard SVM model for classification problems with outliers or noises. As a fuzzy extension of the proximal SVM, the fuzzy linear proximal SVM has also been suggested [23]. A key issue for the FSVM is how to set the fuzzy memberships of training data points, and much work has been done on this. By introducing the confident factor and the trashy factor into the training dataset, Lin and Wang claimed that the approach for automatic setting of fuzzy memberships had been proposed [36]. In this method, many parameters have to be optimized, which makes it very difficult to set the confident factor and the trashy factor automatically. Based on the distance between a sample and its class center in the high-dimensional feature space, a new fuzzy membership function has also been designed [24], which is a kernel extension of the formulation in [35]. In the aforementioned algorithms, the fuzzy membership s_i can be regarded as the degree of the corresponding training point belonging to one class in the classification problem, and the value $1 - s_i$ can be regarded as the degree of meaningfulness. Considering that the same training sample may belong to multiple classes,

Manuscript received November 16, 2009; revised April 1, 2010 and July 15, 2010; accepted September 23, 2010. Date of current version February 7, 2011. This work was supported by the Australian Research Council under Discovery Grant DP0880739 and the Joint Development Project of Industry, Education, and Academy of Guangdong Province under Grant 2007B090400031.

X. Yang is with the Department of Mathematics, School of Sciences, South China University of Technology, Guangzhou 510641, China, and also with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney NSW 2007, Australia (e-mail: xwyang@scut.edu.cn).

G. Zhang, J. Lu, and J. Ma are with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: zhangg@it.uts.edu.au; jliu@it.uts.edu.au; junm@it.uts.edu.au).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TFUZZ.2010.2087382

a bilateral weights-based FSVM model has been presented [55]. At present, this model faces two main difficulties: how to set fuzzy memberships and how to decrease computational complexity. For the regression problems with outliers or noises, the weighted LS-SVM has been developed [49], where the sample weights are set by solving an unweighted LS-SVM model. Based on the fuzzy c -means (FCM) clustering in the original input space and the fuzzy IF-THEN rules, the ε -insensitive fuzzy model for multi-input single-output regression problems [32], [33], ε -insensitive fuzzy c -regression model for multi-input multi-output regression problems [31], and the ε -margin nonlinear classification model [30] have been proposed. In these three models, the strategies of iteratively setting sample weights and ensemble learning have been adopted, where the strategy of iteratively setting sample weights is similar to the two-stage strategy of solving the linear system in [49]. In the case of input variables being fuzzy numbers, Hong and Hwang [20] have proposed a support vector fuzzy regression machine. At present, how to set the reasonable sample memberships in the field of pattern recognition and machine learning is still an open and important problem.

The goals of this paper are to propose a strategy of setting the reasonable sample memberships and to design a robust FSVM algorithm for classification problems with outliers or noises. In the fields of machine learning and pattern recognition, clustering in the high-dimensional feature space is usually called kernel clustering [14]. In recent years, researchers have undertaken much work in kernel clustering. Wu *et al.* [58] have presented a kernel FCM (KFCM) clustering algorithm, Du *et al.* [12] have designed an entropic-regularization-based KFCM clustering algorithm, Mizutani and Miyamoto [40] have suggested an entropic-regularization-based kernel possibilistic FCM clustering algorithm, Yang *et al.* [60] have proposed a Mercer kernel-based deterministic annealing algorithm, Rhee *et al.* [46] have developed a kernel version of [42], and Kim *et al.* [27] have evaluated the performance of kernel fuzzy clustering algorithms. The results have shown that kernel fuzzy clustering is more robust than fuzzy clustering in the original input space for pattern-recognition problems with outliers or noises. Considering that fuzzy clustering can identify the distributions of the training points in different clusters and SVM can provide a good generalization performance, motivated by the design idea in [30], we develop in this paper a KFCM-clustering-based FSVM algorithm (KFCM-FSVM) to deal with the classification problems with outliers or noises. In the KFCM-FSVM algorithm, three steps will be conducted. First, the FCM clustering algorithm is used to cluster each of two classes from the training set in the high-dimensional feature space to obtain a fuzzy partition of the high-dimensional feature space. Second, the farthest pair of clusters, which belong to different classes and form one binary classification problem with membership degrees, is searched. Finally, the FSVM algorithm is adopted to induce the final classification results based on the training set composed of data points with membership degrees obtained in step 2. The experiments are conducted on six benchmarking datasets and four artificial datasets to test the generalization performance of the KFCM-FSVM algorithm.

The noises in classification problems are usually divided into two categories [56]: attribute noise and class noise. The attribute noises are represented by errors that are introduced to attribute values. Examples of these external errors include the erroneous attribute values, the missing attribute values, and the incomplete attributes values. The class noises are usually subdivided into two categories: 1) contradictory examples: The same examples appear more than once and are labeled with different classes; and 2) misclassifications: The examples are labeled with wrong classes. This type of error is common in situations, where different classes have similar symptoms. The outliers are the data points, which are far from the other data points, and they are not usually error data. In this paper, we mainly deal with classification problems with outliers or misclassification noise.

The rest of the paper is organized as follows. The FSVM model for classification problems is briefly reviewed in Section II. The KFCM clustering algorithm and its validity are briefly reviewed in Section III. In Section IV, we discuss a common misunderstanding of Gaussian-function-based kernel fuzzy clustering. In Section V, we give the KFCM-FSVM algorithm in detail. The computational complexity analysis of the KFCM-FSVM algorithm is discussed in Section VI. The experimental results and analysis are presented in Section VII. Finally, we give the related conclusions in Section VIII.

II. FUZZY SUPPORT VECTOR MACHINE MODEL FOR CLASSIFICATION PROBLEMS

Let us consider a training set of l pairs of data points $\{\mathbf{x}_i, y_i, s_i\}_{i=1}^l$ for a binary classification problem, where $\mathbf{x}_i \in R^n$ are the input data, $y_i \in \{-1, +1\}$ are the corresponding binary class labels, and $s_i \in (0, 1]$ is the fuzzy membership degree of \mathbf{x}_i belonging to y_i . The FSVM model for the binary classification problems is a quadratic programming (QP) problem based on the inequality constraints and can be described in the following [35]:

$$\min_{\mathbf{w}, b, \xi} J(\mathbf{w}, b, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \gamma \sum_{i=1}^l s_i \xi_i \quad (1)$$

s.t.

$$y_i [\mathbf{w}^T \varphi(\mathbf{x}_i) + b] \geq 1 - \xi_i, \quad i = 1, \dots, l \quad (2)$$

$$\xi_i \geq 0, \quad i = 1, \dots, l \quad (3)$$

where \mathbf{w} is a normal vector of the hyperplane, b is a bias, $\varphi(\mathbf{x}_i)$ is a nonlinear function that maps \mathbf{x}_i to a high-dimensional feature space, ξ_i is the error in misclassification, and γ is a regularization constant that controls the tradeoff between the classification margin and cost of misclassification.

The corresponding dual expression of the optimization problem (1)–(3) is as follows:

$$\max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

s.t.

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (5)$$

$$0 \leq \alpha_i \leq \gamma s_i, \quad i = 1, \dots, l \quad (6)$$

where α_i is a Lagrange multiplier, and $K(\mathbf{x}_i, \mathbf{x}_j)$ is a kernel function defined as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)\varphi(\mathbf{x}_j). \quad (7)$$

In the field of SVM, polynomial, Gaussian, and sigmoid functions are three commonly used kernel functions and are represented by the following equations, respectively,

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + b)^d \quad (8)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2}{2\sigma^2}} \quad (9)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\alpha(\mathbf{x}_i \cdot \mathbf{x}_j) + \beta). \quad (10)$$

In (8), d is a degree of the polynomial function, and b is an offset constant. In (9), σ^2 is a variance parameter, and in (10), α and β are constant parameters of the sigmoid kernel function.

The optimization problem (4)–(6) can be solved by sequential minimal optimization (SMO) [26]. Once the FSVM model has been solved, the class label of a testing example \mathbf{x} can be predicted as follows:

$$y(\mathbf{x}) = \text{sgn} \left[\sum_{j=1}^l \alpha_j y_j K(\mathbf{x}, \mathbf{x}_j) + b \right]. \quad (11)$$

III. KERNEL FUZZY *c*-MEANS CLUSTERING ALGORITHM AND ITS VALIDITY

Clustering is one of the most useful tasks in the data-mining process for discovering clusters and identifying interesting distributions and patterns in the underlying data. The clustering problem concerns partitioning a given dataset into clusters, such that the data points in a cluster are more similar to each other than the data points in different clusters. The partitions may define a hard boundary between subpartitions; this is called hard clustering [38]. In contrast, the boundaries between subpartitions generated by a fuzzy clustering algorithm are vague [3]. This means that each data point of a fuzzy partition belongs to different clusters with different membership degrees. FCM clustering is one of the most important fuzzy clustering algorithms in the field of pattern recognition and data mining and has been widely studied [13], [17], [21], [51]. In order to study the distribution of the data points in a high-dimensional feature space, a KFCM has been developed [58], and the detailed computational steps are as follows.

- 1) Choose the cluster number C and the termination parameter ε_1 .
- 2) Choose the kernel function K and its parameters.
- 3) Initialize the cluster centroids $\mathbf{v}_j, j = 1, 2, \dots, C$.

- 4) Compute the membership degrees u_{ij} of data point \mathbf{x}_i in the j th cluster $i = 1, 2, \dots, l$ and $j = 1, 2, \dots, C$

$$u_{ij} = \frac{(1/d^2(\mathbf{x}_i, \mathbf{v}_j))^{1/m-1}}{\sum_{p=1}^C (1/d^2(\mathbf{x}_i, \mathbf{v}_p))^{1/m-1}} \quad (12)$$

where

$$d^2(\mathbf{x}_i, \mathbf{v}_p) = K(\mathbf{x}_i, \mathbf{x}_i) - 2K(\mathbf{x}_i, \mathbf{v}_p) + K(\mathbf{v}_p, \mathbf{v}_p). \quad (13)$$

- 5) Compute the new kernel matrices $K(\mathbf{x}_i, \mathbf{v}_p^{\text{new}})$ and $K(\mathbf{v}_p^{\text{new}}, \mathbf{v}_p^{\text{new}})$

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{v}_p^{\text{new}}) &= \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{v}_p^{\text{new}}) \\ &= \frac{\sum_{k=1}^l (u_{kp})^m K(\mathbf{x}_k, \mathbf{x}_i)}{\sum_{k=1}^l (u_{kp})^m} \end{aligned} \quad (14)$$

$$\begin{aligned} K(\mathbf{v}_p^{\text{new}}, \mathbf{v}_p^{\text{new}}) &= \varphi(\mathbf{v}_p^{\text{new}}) \cdot \varphi(\mathbf{v}_p^{\text{new}}) \\ &= \frac{\sum_{k=1}^l \sum_{n=1}^l (u_{kp})^m (u_{np})^m K(\mathbf{x}_k, \mathbf{x}_n)}{(\sum_{k=1}^l (u_{kp})^m)^2} \end{aligned} \quad (15)$$

where

$$\varphi(\mathbf{v}_p^{\text{new}}) = \frac{\sum_{k=1}^l (u_{kp})^m \varphi(\mathbf{x}_k)}{\sum_{k=1}^l (u_{kp})^m}. \quad (16)$$

- 6) Update the membership degrees u_{ij} to u_{ij}^{new} , according to (12).
- 7) If $\max_{i,j} |u_{ij} - u_{ij}^{\text{new}}| < \varepsilon_1$, stop; otherwise, go to 5).

Cluster validity is one of the most important issues in cluster analysis [18]. A reliable validation index for a fuzzy clustering must consider both the compactness and the separation of the fuzzy partition. The optimal partitions require a maximal compactness for each cluster such that the clusters are located far from each other. Some validity indices suitable for fuzzy clustering, such as V_{PC} [2], generalized Dunn's index [4], V_{OS} [28], V_{CWB} [45], V_{XB} [59], and V_{SC} [62], have been developed. For FCM clustering, Pal and Bezdek have evaluated the performance of five validity indices and found that V_{XB} provided the best performance [41]. In this study, we use V_{XB} in [59]. In the high-dimensional feature space, it is denoted as follows when $m = 2$:

$$S = \frac{\sum_{j=1}^C \sum_{i=1}^l u_{ij}^2 d^2(\mathbf{x}_i, \mathbf{v}_j)}{l \min_{i,j} \{d^2(\mathbf{v}_i, \mathbf{v}_j)\}}. \quad (17)$$

Our goal is to find the fuzzy c -partition with the smallest S .

IV. COMMON MISUNDERSTANDING OF GAUSSIAN-FUNCTION-BASED KERNEL FUZZY CLUSTERING

The Gaussian-function-based kernel fuzzy clustering has been widely used in theoretical research and real-world applications. Chen and Zhang [7], Liao *et al.* [34], and Zhang and Chen [63] have applied it to magnetic resonance imaging brain image segmentation, Yang and Tsai [61] have proposed

a new version based on spatial bias correction, and Rhee *et al.* [46], Tushir and Srivastava [52], Wu and Zhou [57], and Zhou *et al.* [65] have applied it to possibilistic c -means clustering. Unfortunately, the researchers mentioned earlier believe that the preimages of the cluster centroids in the high-dimensional feature space exist in the original input space. In fact, this is a misunderstanding of Gaussian-function-based kernel fuzzy clustering. In the following Theorem 1, we will prove that the preimages of the cluster centroids in the high-dimensional feature space induced by the Gaussian kernel function do not necessarily exist in the original input space.

Theorem 1: In the high-dimensional feature space induced by the Gaussian kernel function, the preimages of the cluster centroids induced by the FCM algorithm do not necessarily exist in the original input space.

Proof: When the Gaussian kernel function is adopted, the data points in the original input space are mapped into the surface of the unit hypersphere in the high-dimensional feature space. Assuming that the cluster centroid $\varphi(\mathbf{v}_j^{\text{new}})$ in the high-dimensional feature space induced by Gaussian kernel function has one preimage $\mathbf{v}_j^{\text{new}}$ in the original input space, it is obvious that the cluster centroid $\varphi(\mathbf{v}_j^{\text{new}})$ locates on the surface of the unit hypersphere in the high-dimensional feature space. On the other hand, from (16), it is known that $\varphi(\mathbf{v}_j^{\text{new}})$ is a convex combination of the image vectors $\varphi(\mathbf{x}_k)$, $k = 1, 2, \dots, l$, which are images of some points in the original input space, in the high-dimensional feature space. However, because these image vectors $\varphi(\mathbf{x}_k)$, $k = 1, 2, \dots, l$ locate on the surface of the unit hypersphere, in general, $\varphi(\mathbf{v}_j^{\text{new}})$ should locate the interior of the unit hypersphere in the high-dimensional feature space, except that there is unique point in this cluster. This is a conflict, which indicates that the preimages of the cluster centroids in the high-dimensional feature space induced by the Gaussian kernel function do not necessarily exist in the original input space.

V. FUZZY SUPPORT VECTOR MACHINE ALGORITHM BASED ON KERNEL FUZZY C -MEANS CLUSTERING

Based on the KFCM clustering in the high-dimensional feature space and FSVM, in this section, we propose a KFCM-FSVM algorithm to deal with the binary classification problems with outliers or noises.

In the KFCM-FSVM algorithm, three steps will be conducted. First, the FCM clustering algorithm is used to cluster each of two classes from the training set in the high-dimensional feature space to obtain a fuzzy partition of the feature space and discover the distribution structures of the data points in the training set. In this step, the outliers are usually assigned to the lower membership degrees. Second, the farthest pair of clusters, which belong to different classes and form one binary classification problem with membership degrees, is searched. In real-world classification problems, the noises usually locate in the interaction area between two classes. When the farthest pair of clusters is adopted, the noises will be given lower membership degrees. From the two steps mentioned earlier, we can obtain a reasonable sample membership degree, which will degrade the effects the outliers or noises have on the decision function. Finally, the

FSVM is adopted to induce the final classification results. The detailed computational steps are as follows.

- Step 1:* Choose the kernel function K , and set the corresponding hyperparameters.
- Step 2:* Cluster each of two classes (ω_+ denotes the positive class and ω_- denotes the negative class) from the training set in the high-dimensional feature space by the FCM clustering algorithm and the validity index to obtain a fuzzy partition of the corresponding class in this feature space. Let the number of clusters in the class ω_+ be C_+ and the number of clusters in the class ω_- be C_- .
- Step 3:* Search for the farthest pair of clusters, in which one cluster belongs to the positive class ω_+ , and the other belongs to the negative class ω_- .
- Step 4:* The two farthest clusters obtained earlier compose a new training set with membership degrees, and the FSVM is conducted on this set to obtain a nonlinear classifier with the parameters α_i and b . It should be noted that the difference between the new training set and the original one is that the membership degrees are assigned to every single data point in the new training set.
- Step 5:* The class label of a testing example \mathbf{x} is predicted by (11).

If we replace KFCM with FCM in the KFCM-FSVM framework, we call this algorithm the FCM-FSVM algorithm.

Compared with the work in [30], the KFCM-FSVM has five changes. First, we partition the high-dimensional feature space, while [30] partitions the original input space. Second, we adopt the cluster validity index to determine the optimal cluster parameters C_+ and C_- , while [30] uses the equal cluster number. Third, we select the farthest pair of clusters to degrade the effects the outliers or noises have on the decision function, while [30] selects C pairs of clusters according to the nearest centroid distance. Fourth, we solve one QP optimization problem, while [30] solves C QP optimization problems. Finally, the single classifier is linear in [30], whereas the nonlinear is presented by the ensemble learning strategy, while there is only one nonlinear classifier in the KFCM-FSVM algorithm.

The flowchart of the KFCM-FSVM algorithm is depicted in Fig. 1.

VI. ANALYSIS OF THE COMPUTATIONAL COMPLEXITY OF THE KERNEL FUZZY C -MEANS CLUSTERING-BASED FUZZY SUPPORT VECTOR MACHINE ALGORITHM

The computational cost of the proposed algorithm is composed of the cost for KFCM clustering, the cost for searching the farthest pair of clusters, and the cost for solving the FSVM model. From (12)–(15), we know that the computational cost of the KFCM clustering is $O(t_+ \cdot C_+ \cdot l_+^2 + t_- \cdot C_- \cdot l_-^2)$ and that the computational cost of searching the farthest pair of clusters is $O(C_+ \cdot C_- \cdot l_+ \cdot l_-)$, where t_+ and t_- are the iterative times of the positive and the negative classes, respectively, C_+ and C_- are the corresponding cluster numbers, and l_+ and l_- are the corresponding sample numbers. Because the FSVM model is

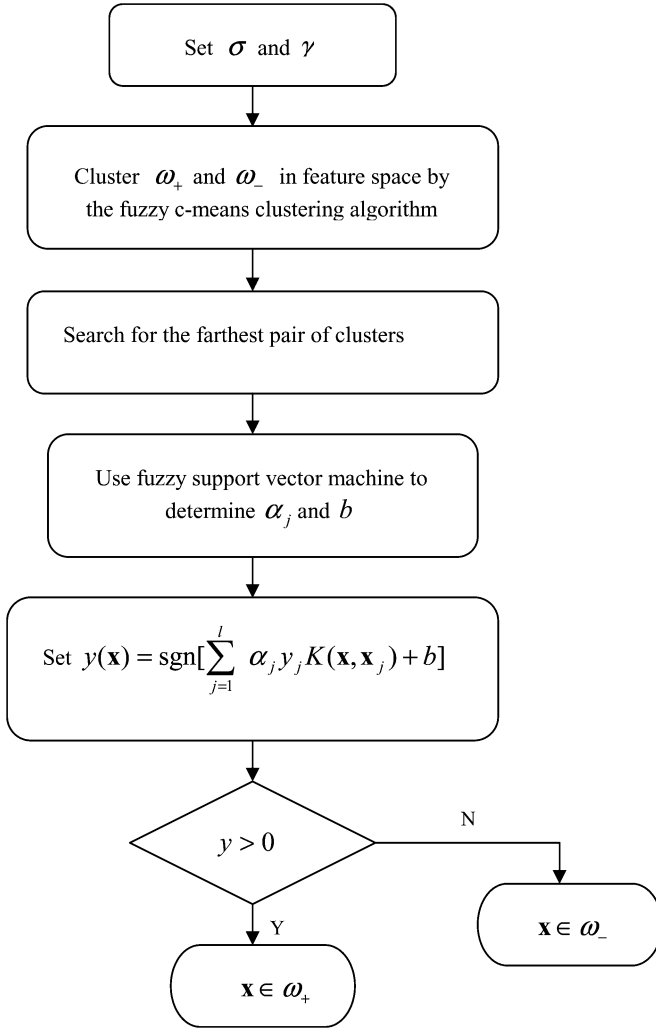


Fig. 1. Flowchart of the KFCM-FSVM algorithm.

solved by SMO, the computational cost of the SMO algorithm is $O(l_+ + l_-)^{2.2}$ [44]. Therefore, the computational cost of the proposed algorithm is $O(t_+ \cdot C_+ \cdot l_+^2 + t_- \cdot C_- \cdot l_-^2 + C_+ \cdot C_- \cdot l_+ \cdot l_- + (l_+ + l_-)^{2.2})$.

VII. NUMERICAL EXPERIMENTS AND DISCUSSIONS

In this section, we will conduct experiments on six benchmarking datasets and four artificial datasets to test the generalization performance of the KFCM-FSVM algorithm, for which the Pima Indians Diabetes (PIMA) dataset, waveform dataset, Statlog (Landsat satellite) dataset (Sat dataset), and MONK dataset can be downloaded from <http://archive.ics.uci.edu/ml>, and the Banana dataset can be downloaded from <http://ida.first.gmd.de/~raetsch/data/benchmarks.htm>. In order to show that the membership degrees are reasonable in the KFCM-FSVM algorithm, the SVM, FSVM, and FCM-FSVM algorithms are also implemented. In our experiments, the Gaussian kernel function is adopted. In the KFCM-FSVM and the FCM-FSVM algorithms, $m = 2$ and $\varepsilon_1 = 0.01$, and a random partition matrix is used for initialization. In the FSVM algorithm, the fuzzy

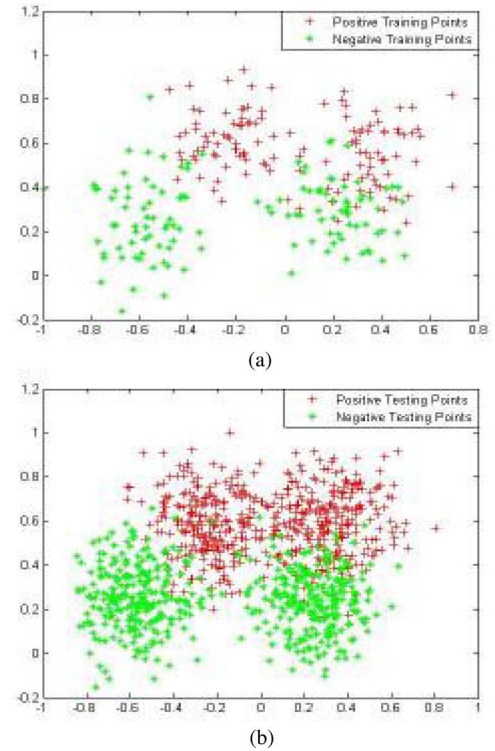


Fig. 2. Ripley dataset. (a) Training set. (b) Testing set.

membership degrees are set according to the strategy in [24], where $\delta = 0.01$. All computations are conducted on a server with a dual 2.0-GHz Intel Xeon processor and a maximum of 4.0-GB memory running CentOS Linux 5.3. All the programs are written in C++ and compiled using a GCC 4.2 compiler.

The datasets used in our experiments are described in detail as follows.

- 1) *Ripley dataset* [47]: This dataset consists of 1250 patterns with two features, and these patterns are assigned to two classes. Each class has a bimodal distribution obtained as a mixture of two normal distributions. The training set consists of 250 patterns (125 patterns belonging to each class), and the testing set consists of 1000 patterns (500 patterns belonging to each class), which are illustrated in Fig. 2.
- 2) *PIMA dataset*: This dataset consists of real-world data collected by the U.S. National Institute of Diabetes and Kidney Diseases. It contains 768 patterns with eight features, and these patterns are assigned to two classes. We randomly select 576 patterns (202 patterns belonging to the positive class and 374 patterns belonging to the negative class) for training and the remaining 192 patterns (66 patterns belonging to the positive class and 126 patterns belonging to the negative class) for testing.
- 3) *Banana dataset*: This dataset consists of patterns with two features and these patterns are assigned to two classes. The training set consists of 400 patterns (194 patterns belonging to the positive class and 206 patterns belonging to the negative class), and the testing set consists of 4900 patterns (2182 patterns belonging to the positive

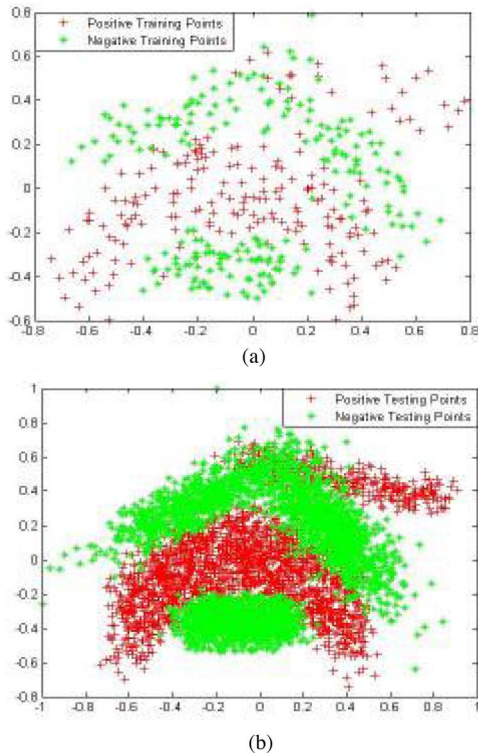


Fig. 3. Banana dataset. (a) Training set. (b) Testing set.

class and 2718 patterns belonging to the negative class), which are illustrated in Fig. 3.

- 4) *Waveform dataset*: This dataset consists of 5000 patterns with 21 features, and these patterns are assigned to two classes. The training set consists of 400 patterns (136 patterns belonging to the positive class and 264 patterns belonging to the negative class), and the testing set consists of 4600 patterns (1511 patterns belonging to the positive class and 3089 patterns belonging to the negative class).
- 5) *Statlog (Landsat satellite) dataset (Sat dataset)*: This dataset has 4435 training patterns and 2000 testing patterns with 36 attributes, whose values are real numbers, and six classes. These patterns are preprocessed to separate classes $\{1, 2, 5\}$ belonging to the positive class and classes $\{3, 4, 7\}$ belonging to the negative class.
- 6) *MONK dataset*: The MONK problem was the basis of the first international comparison of learning algorithms. There are three MONK problems. In this paper, the third MONK problem with a random noise added to the data was used. This dataset consists of 554 patterns with seven features, and these patterns are assigned to two classes. The training set consists of 122 patterns (60 patterns belonging to the positive class and 62 patterns belonging to the negative class), and the testing set consists of 432 patterns (228 patterns belonging to the positive class and 204 patterns belonging to the negative class).
- 7) *Artificial dataset 1*: This dataset consists of 560 patterns with two features, and these patterns are assigned to two classes. The training set consists of 279 patterns

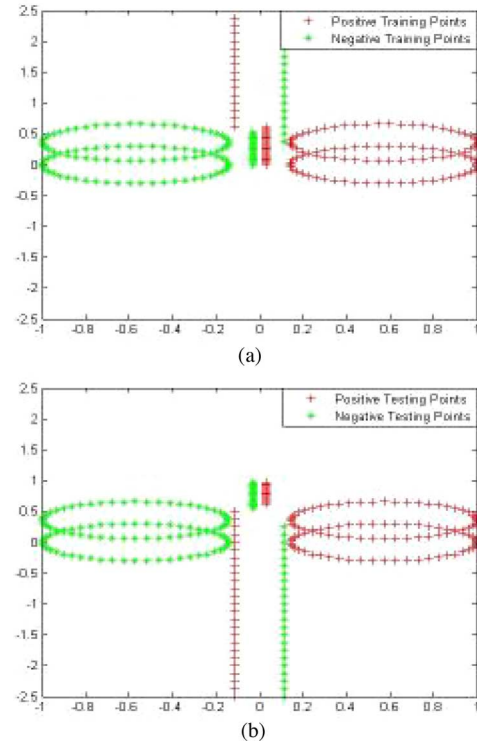


Fig. 4. Artificial dataset 1. (a) Training set. (b) Testing set.

(140 patterns belonging to the positive class and 139 patterns belonging to the negative class), and the testing set consists of 281 patterns (140 patterns belonging to the positive class and 141 patterns belonging to the negative class), which are illustrated in Fig. 4.

- 8) *Artificial dataset 2*: This dataset consists of 1260 patterns with two features, and these patterns are assigned to two classes. The training set consists of 630 patterns (293 patterns belonging to the positive class and 337 patterns belonging to the negative class), and the testing set consists of 630 patterns (297 patterns belonging to the positive class and 333 patterns belonging to the negative class), which are illustrated in Fig. 5.
- 9) *Artificial dataset 3*: This dataset is a “ring-shaped” dataset and consists of 680 patterns with two features, and these patterns are assigned to two classes. The training set consists of 344 patterns (173 patterns belonging to the positive class and 171 patterns belonging to the negative class), and the testing set consists of 336 patterns (163 patterns belonging to the positive class and 173 patterns belonging to the negative class), which are illustrated in Fig. 6.
- 10) *Artificial dataset 4*: This dataset is a “T-shaped” dataset. It consists of 440 patterns with two features, and these patterns are assigned to two classes. The training set consists of 221 patterns (110 patterns belonging to the positive class and 111 patterns belonging to the negative class), and the testing set consists of 219 patterns (109 patterns belonging to the positive class and 110 patterns belonging to the negative class), which are illustrated in Fig. 7.

In our experiments, two goals will be pursued and attained. The first goal is to show that the membership degrees are

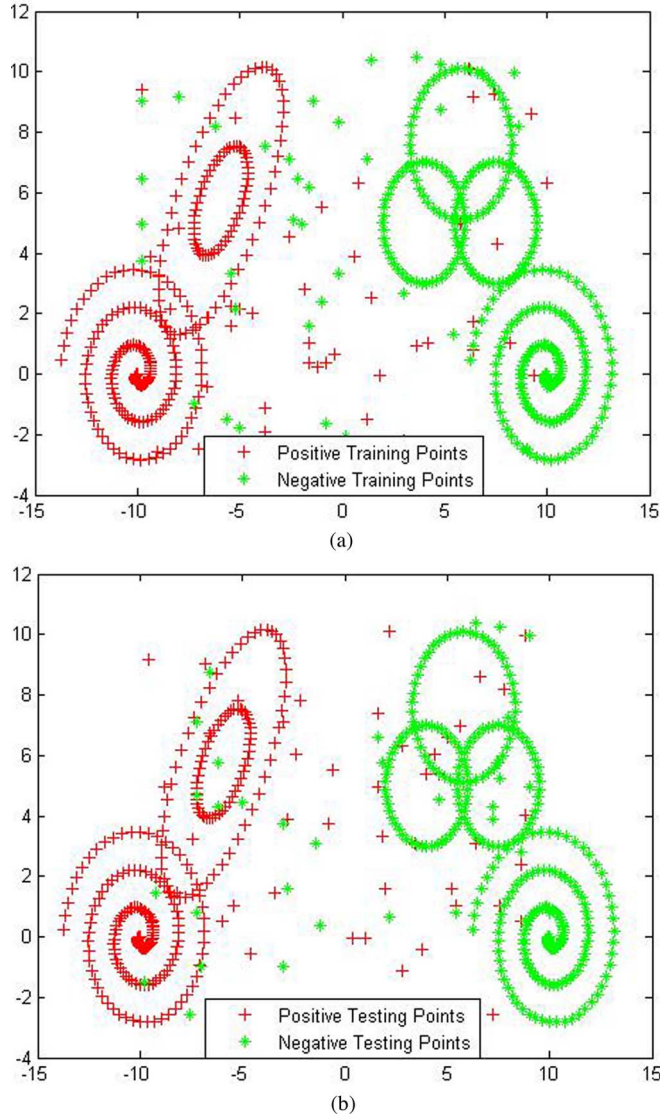


Fig. 5. Artificial dataset 2. (a) Training set. (b) Testing set.

reasonable in the KFCM-FSVM algorithm, and the second goal is to show that the KFCM-FSVM algorithm yields statistically reliable results. In order to reach the first goal, based on the grid method, where $\sigma = [2^{-4}, 2^{-3}, 2^{-2}, \dots, 2^5]$ and $\gamma = [2^0, 2^1, 2^2, \dots, 2^9]$, we first run four algorithms on the whole training set and obtain the decision function and then use the decision function to predict the data in the testing set. The testing accuracy, the optimal cluster number, and the corresponding optimal hyperparameters from four algorithms are reported in Table I. In consideration of the fact that the FCM-FSVM and KFCM-FSVM algorithms are sensitive to the initial cluster centroids, they run ten times independently. The testing accuracy corresponding to the FCM-FSVM and KFCM-FSVM algorithms in Table I is the best testing accuracy in ten trials. To show that the KFCM-FSVM algorithm is more robust than the FCM-FSVM algorithm for classification problems with outliers or noises, ten results conducted by two algorithms on the waveform dataset, artificial dataset 3, and artificial dataset 4

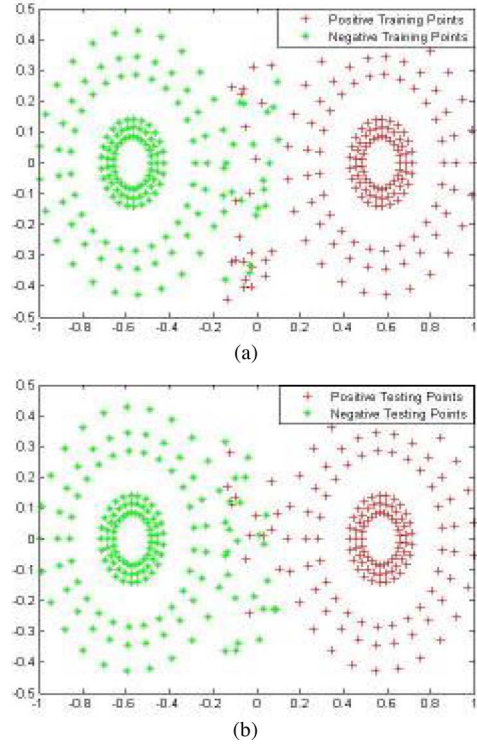


Fig. 6. Artificial dataset 3. (a) Training set. (b) Testing set.

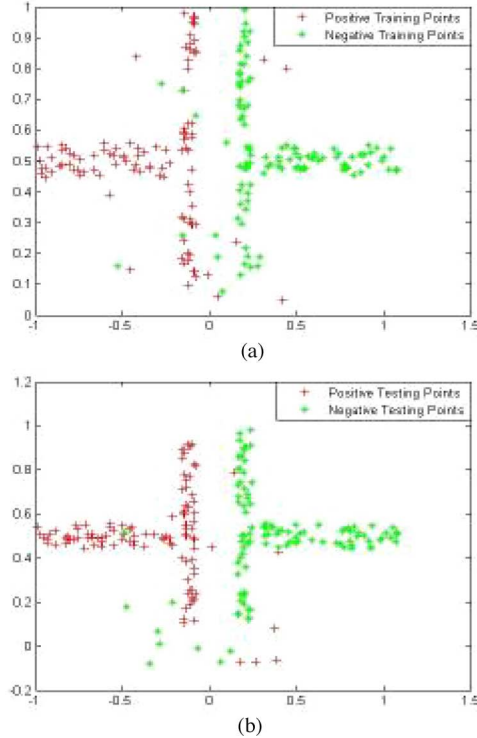


Fig. 7. Artificial dataset 4. (a) Training set. (b) Testing set.

are listed in Tables II–IV, respectively. In order to reach the second goal, we use the five-fold cross-validation strategy to reconstruct all the experiments for the SVM, FSVM, FCM-FSVM, and KFCM-FSVM algorithms. The detailed experimental steps are as follows. First, we randomly divide the given training set W into five subsets of approximately equal size, i.e.,

TABLE I
COMPARISON OF THE BEST RESULTS CONDUCTED BY THE SVM, FSVM,
FCM-FSVM, AND KFCM-FSVM ALGORITHMS ON THE TEN DATASETS

Dataset	Algorithm	Optimal cluster number	σ	γ	Testing accuracy
Ripley	SVM		0.5	1	0.907
	FSVM		0.125	16	0.913
	FCM-FSVM	(2,2)	0.5	256	0.883
	KFCM-FSVM	(2,2)	0.25	2	0.915
PIMA	SVM		1	1	0.776042
	FSVM		0.25	32	0.750000
	FCM-FSVM	(2,2)	4	512	0.786458
	KFCM-FSVM	(2,2)	2	64	0.786458
Waveform	SVM		1	2	0.902826
	FSVM		2	128	0.899348
	FCM-FSVM	(2,2)	1	4	0.904783
	KFCM-FSVM	(2,2)	1	2	0.906087
Banana	SVM		0.5	256	0.899184
	FSVM		0.25	512	0.894286
	FCM-FSVM	(4,3)	0.25	16	0.886939
	KFCM-FSVM	(4,3)	0.25	32	0.899388
MONK	SVM		4	32	0.976852
	FSVM		2	8	0.965278
	FCM-FSVM	(2,2)	2	4	0.974537
	KFCM-FSVM	(2,2)	2	8	0.976852
Sat	SVM		32	4	0.9845
	FSVM		32	64	0.9845
	FCM-FSVM	(2,3)	32	4	0.9860
	KFCM-FSVM	(2,3)	32	2	0.9860
Artificial Dataset 1	SVM		2	2	0.829181
	FSVM		0.0625	256	0.839858
	FCM-FSVM	(3,3)	1	16	0.854093
	KFCM-FSVM	(3,3)	0.5	1	0.861210
Artificial Dataset 2	SVM		1	512	0.936508
	FSVM		4	256	0.938095
	FCM-FSVM	(2,3)	4	2	0.941270
	KFCM-FSVM	(2,2)	16	64	0.939683
Artificial Dataset 3	SVM		4	16	0.955357
	FSVM		8	1	0.958333
	FCM-FSVM	(2,2)	4	2	0.967262
	KFCM-FSVM	(2,2)	4	1	0.961310
Artificial Dataset 4	SVM		0.0625	1	0.949772
	FSVM		0.0625	2	0.949772
	FCM-FSVM	(3,2)	0.5	512	0.954338
	KFCM-FSVM	(3,2)	0.0625	1	0.968037

W_1, W_2, \dots, W_5 , where $W_i \cap W_j = \emptyset$. Second, for each pair of hyperparameters (σ, γ) , where $\sigma = [2^{-4}, 2^{-3}, 2^{-2}, \dots, 2^5]$ and $\gamma = [2^0, 2^1, 2^2, \dots, 2^9]$, we run the given algorithm on $W \setminus W_k$, test the trained classifier on W_k to get the testing accuracy T_k , where k changes from one to five, and compute the average testing accuracy $T_{\text{average}} = 1/5 \sum_{k=1}^5 T_k$. Third, we find the maximum average testing accuracy and the corresponding hyperparameters. The results obtained by the five-fold cross-validation strategy are unbiased and listed in Table V.

Table I indicates that the KFCM-FSVM algorithm gives eight best results on ten datasets. For the artificial datasets 2 and 3, the results of the KFCM-FSVM algorithm are also the second best. The reason why the KFCM-FSVM algorithm can present better testing accuracy is that on the one hand, in the KFCM-FSVM algorithm, the data points in the original input space are first mapped into a hyperspherical surface in the high-dimensional feature space, and the FCM clustering is then conducted on these spherical data points, which completely takes advantage of the FCM clustering to degrade the effects the outliers have on the decision function; on the other hand, the strategy of selecting the farthest pair of clusters degrades the effects the noises have on the decision function. The FCM-FSVM algorithm gives seven in ten results that are superior to the results of the SVM and FSVM algorithms. Compared with the SVM algorithm, the FSVM algorithm based on the strategy of setting the fuzzy mem-

TABLE II
COMPARISON OF THE RESULTS OBTAINED BY THE FCM-FSVM AND
KFCM-FSVM ALGORITHMS ON WAVEFORM DATASET

Algorithm	Trial no.	Optimal cluster number	σ	γ	Testing accuracy
FCM-FSVM	1	(2,2)	1	8	0.898478
	2	(2,2)	1	4	0.901087
	3	(2,2)	1	4	0.903261
	4	(2,2)	1	2	0.903043
	5	(3,2)	1	2	0.904783
	6	(2,2)	1	4	0.903696
	7	(4,2)	1	8	0.900652
	8	(3,2)	1	8	0.901304
	9	(2,2)	1	8	0.898261
	10	(2,2)	1	8	0.901087
Average testing accuracy					0.901565
KFCM-FSVM	1	(2,2)	1	4	0.902826
	2	(3,2)	1	4	0.902826
	3	(2,2)	1	4	0.902826
	4	(2,2)	1	4	0.902609
	5	(2,2)	1	2	0.906087
	6	(2,2)	1	4	0.903043
	7	(2,2)	1	4	0.902826
	8	(3,2)	1	4	0.902826
	9	(2,2)	1	4	0.903043
	10	(2,2)	1	4	0.902826
Average testing accuracy					0.903174

TABLE III
COMPARISON OF THE RESULTS OBTAINED BY THE FCM-FSVM AND
KFCM-FSVM ALGORITHMS ON ARTIFICIAL DATASET 3

Algorithm	Trial no.	Optimal cluster number	σ	γ	Testing accuracy
FCM-FSVM	1	(2,2)	4	2	0.967262
	2	(2,2)	8	2	0.96131
	3	(2,2)	2	8	0.958333
	4	(2,2)	4	1	0.96131
	5	(2,2)	2	8	0.958333
	6	(2,2)	8	2	0.958333
	7	(2,2)	4	2	0.967262
	8	(2,2)	8	2	0.96131
	9	(2,2)	4	1	0.964286
	10	(2,2)	2	8	0.958333
Average testing accuracy					0.961607
KFCM-FSVM	1	(2,2)	4	1	0.96131
	2	(2,2)	4	1	0.96131
	3	(2,2)	4	1	0.96131
	4	(2,2)	4	1	0.96131
	5	(2,2)	8	4	0.96131
	6	(2,2)	4	1	0.96131
	7	(2,2)	8	2	0.96131
	8	(2,2)	8	4	0.96131
	9	(2,2)	4	1	0.96131
	10	(2,2)	16	16	0.96131
Average testing accuracy					0.96131

TABLE IV
COMPARISON OF THE RESULTS OBTAINED BY THE FCM-FSVM AND
KFCM-FSVM ALGORITHMS ON ARTIFICIAL DATASET 4

Algorithm	Trial no.	Optimal cluster number	σ	γ	Testing accuracy
FCM-FSVM	1	(3,2)	0.0625	256	0.917808
	2	(3,2)	0.5	512	0.954338
	3	(3,2)	0.0625	256	0.917808
	4	(3,2)	4	128	0.936073
	5	(3,2)	0.0625	256	0.917808
	6	(3,2)	0.25	32	0.949772
	7	(3,2)	0.5	256	0.954338
	8	(3,2)	0.25	32	0.949772
	9	(3,2)	0.5	256	0.954338
	10	(3,2)	0.0625	256	0.917808
Average testing accuracy					0.936986
KFCM-FSVM	1	(3,2)	0.0625	1	0.968037
	2	(3,2)	0.0625	1	0.949772
	3	(3,2)	0.0625	1	0.968037
	4	(3,2)	0.0625	1	0.949772
	5	(3,2)	0.0625	1	0.968037
	6	(3,2)	0.25	64	0.954338
	7	(3,2)	0.5	512	0.954338
	8	(3,2)	0.0625	1	0.949772
	9	(3,2)	0.0625	1	0.949772
	10	(3,2)	0.0625	1	0.949772
Average testing accuracy					0.956164

TABLE V
COMPARISON OF THE RESULTS CONDUCTED BY THE SVM, FSVM,
FCM-FSVM, AND KFCM-FSVM ALGORITHMS ON THE TEN TRAINING SETS
BASED ON FIVE-FOLD CROSS-VALIDATION STRATEGY

Dataset	Algorithm	σ	γ	Testing Accuracy
Ripley	SVM	0.5	32	0.880
	FSVM	0.5	256	0.892
	FCM-FSVM	0.0625	512	0.796
	KFCM-FSVM	0.125	1	0.892
PIMA	SVM	2	256	0.773913
	FSVM	0.5	8	0.761739
	FCM-FSVM	4	128	0.775652
	KFCM-FSVM	2	64	0.784348
Waveform	SVM	1	8	0.920000
	FSVM	0.5	512	0.920000
	FCM-FSVM	1	2	0.925000
	KFCM-FSVM	1	16	0.925000
Banana	SVM	0.25	4	0.905000
	FSVM	0.25	64	0.902500
	FCM-FSVM	0.0625	16	0.880000
	KFCM-FSVM	0.25	64	0.930000
MONK	SVM	4	256	0.925000
	FSVM	2	16	0.900000
	FCM-FSVM	2	64	0.908333
	KFCM-FSVM	2	8	0.925000
Sat	SVM	32	32	0.980383
	FSVM	32	512	0.980383
	FCM-FSVM	32	512	0.980609
	KFCM-FSVM	32	64	0.980609
Artificial Dataset 1	SVM	0.25	512	0.989091
	FSVM	0.5	256	0.974545
	FCM-FSVM	0.125	512	0.985455
	KFCM-FSVM	0.25	512	0.989091
Artificial Dataset 2	SVM	2	4	0.944444
	FSVM	1	4	0.942857
	FCM-FSVM	2	256	0.947619
	KFCM-FSVM	4	128	0.947619
Artificial Dataset 3	SVM	1	128	0.961765
	FSVM	0.5	512	0.961765
	FCM-FSVM	1	512	0.961765
	KFCM-FSVM	0.25	2	0.964706
Artificial Dataset 4	SVM	4	512	0.968182
	FSVM	2	512	0.968182
	FCM-FSVM	16	256	0.959091
	KFCM-FSVM	0.0625	1	0.972727

bership degrees in [24] can improve the testing accuracy for the datasets, where the outliers or noises are far from the data centroid. For example, the results of the FSVM algorithm are better than those of the SVM algorithm for the Ripley dataset and the artificial datasets 1 and 3. The reason is that the strategy in [24] is based on the hypersphere. When the outliers or noises are far from the data centroid, they will be assigned to lower membership degrees and will have lesser effects on the decision function.

In Tables II–IV, it can be seen that both the FCM-FSVM and KFCM-FSVM algorithms are sensitive to the initial cluster centroids, which is the main shortcoming of the presented algorithm. The reason is that both of them are based on FCM clustering. Generally speaking, the KFCM-FSVM algorithm is more robust than the FCM-FSVM algorithm for classification problems with outliers or noises. For example, six in ten results of the FCM-FSVM algorithm are less than that of the SVM algorithm, while one in ten results of the KFCM-FSVM algorithm is less than that of the SVM algorithm for the waveform dataset. For the artificial dataset 4, five in ten results of the FCM-FSVM

algorithm are less than the one of the SVM algorithm, while no result of the KFCM-FSVM algorithm is less than that of the SVM algorithm.

The unbiased results in Table V indicate that the testing accuracy of the KFCM-FSVM algorithm is the best on all of the training sets, which shows that the proposed KFCM-FSVM algorithm provides statistically reliable results.

VIII. CONCLUSION

This paper proposes a KFCM-clustering-based FSVM algorithm for classification problems with outliers or noises. The contributions of this paper are as follows. First, a common misunderstanding of Gaussian-function-based kernel fuzzy clustering in the high-dimensional feature space is corrected. Second, the KFCM-FSVM algorithm is proposed for classification problems with outliers or noises. The experiments have been conducted on six benchmarking datasets and four artificial datasets to test the generalization performance of the KFCM-FSVM algorithm. The results show that the KFCM-FSVM algorithm based on the FCM clustering in the high-dimensional feature space presents the most reasonable membership degrees and is more robust than the FCM-FSVM algorithm based on the FCM clustering in the original input space, SVM, and FSVM for classification problems with outliers or noises. Third, the computational complexity of the KFCM-FSVM algorithm is presented. It should be noted that although the KFCM-FSVM algorithm is tested on the binary classification problems, it can be applied to multiclassification problems without difficulty. However, it does not fit large-scale classification problems with outliers or noises.

In future work, we will investigate the techniques of data sampling and data compressing so that the proposed algorithm can be extended to large-scale classification problems with outliers or noises. Another interesting topic would be to solve the problem in which FCM is sensitive to the initial cluster centroids. A possible extension of the KFCM-FSVM algorithm would be to handle the nonlinear function estimation problems with outliers or noises. Further study on this topic will also include the accuracy measure of fuzzy classification [1], [15], [16] and many applications of the KFCM-FSVM algorithm in real-world problems.

REFERENCES

- [1] A. Amo, J. Montero, G. Biging, and V. Cutello, "Fuzzy classification systems," *Eur. J. Oper. Res.*, vol. 156, pp. 459–507, 2004.
- [2] J. C. Bezdek, "Cluster validity with fuzzy sets," *Cybern. Syst.*, vol. 3, no. 3, pp. 58–73, 1973.
- [3] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.
- [4] J. C. Bezdek and N. R. Pal, "Some new indices of cluster validity," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 28, no. 3, pp. 301–315, Jun. 1998.
- [5] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Trans. Neural Netw.*, vol. 14, no. 6, pp. 1506–1518, Nov. 2003.
- [6] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–1064, Sep. 1999.
- [7] S. C. Chen and D. Q. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.

- [8] Y. X. Chen and J. Z. Wang, "Image categorization by learning and reasoning with regions," *J. Mach. Learn. Res.*, vol. 5, pp. 913–939, 2004.
- [9] J. H. Chiang and T. L. M. Lee, "In silico prediction of human protein interactions using fuzzy-SVM mixture models and its application to cancer research," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 4, pp. 1087–1095, Aug. 2008.
- [10] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [11] O. Deniz, M. Castrillon, and M. Hernandez, "Face recognition using independent component analysis and support vector machines," *Pattern Recognit. Lett.*, vol. 24, no. 13, pp. 2153–2157, 2003.
- [12] W. W. Du, K. Inoue, and K. Urahama, "Robust kernel fuzzy clustering," in *Proc. 2nd Int. Conf. Fuzzy Syst. Knowl. Discov.*, Changsha, China, 2005, pp. 454–461.
- [13] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 773–781, Jul. 1989.
- [14] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 780–784, May 2002.
- [15] D. Gomez, G. Biging, and J. Montero, "Accuracy statistics for judging soft classification," *Int. J. Remote Sens.*, vol. 29, pp. 693–709, 2008.
- [16] D. Gomez, J. Montero, and G. Biging, "Improvements to remote sensing using fuzzy classification, graphs and accuracy statistics," *Pure Appl. Geophys.*, vol. 165, pp. 1555–1575, 2008.
- [17] L. Gröll and J. Jäkel, "A new convergence proof of fuzzy c -means," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 5, pp. 717–720, Oct. 2005.
- [18] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Inf. Syst.*, vol. 17, no. 2–3, pp. 107–145, 2001.
- [19] D. H. Hong and C. H. Hwang, "Interval regression analysis using quadratic loss support vector machine," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 2, pp. 229–237, Apr. 2005.
- [20] D. H. Hong and C. H. Hwang, "Support vector fuzzy regression machines," *Fuzzy Sets Syst.*, vol. 138, no. 2, pp. 271–281, 2003.
- [21] F. Höppner and F. Klawonn, "A contribution to convergence theory of fuzzy c -means and derivatives," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 5, pp. 682–694, Oct. 2003.
- [22] H. P. Huang and Y. H. Liu, "Fuzzy support vector machines for pattern recognition and data mining," *Int. J. Fuzzy Syst.*, vol. 4, no. 3, pp. 826–835, 2002.
- [23] Jayadeva, R. Khemchandani, and S. Chandra, "Fast and robust learning through fuzzy linear proximal support vector machines," *Neurocomputing*, vol. 61, pp. 401–411, 2004.
- [24] X. F. Jiang, Z. Yi, and J. C. Lv, "Fuzzy SVM with a new fuzzy membership function," *Neural Comput. Appl.*, vol. 15, no. 3–4, pp. 268–276, 2006.
- [25] K. Jonsson, J. Kittler, Y. P. Li, and J. Matas, "Support vector machines for face authentication," *Image Vision Comput.*, vol. 20, no. 5–6, pp. 369–375, 2002.
- [26] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Neural Comput.*, vol. 13, no. 3, pp. 637–649, 2001.
- [27] D. W. Kim, K. Lee, D. Lee, and K. H. Lee, "Evaluation of the performance of clustering algorithms in kernel-based feature space," *Pattern Recognit.*, vol. 38, pp. 607–611, 2004.
- [28] D. W. Kim, K. H. Lee, and D. Lee, "On cluster validity index for estimation of the optimal number of fuzzy clusters," *Pattern Recognit.*, vol. 37, pp. 2009–2025, 2004.
- [29] K. I. Kim, J. H. Kim, and K. Jung, "Face recognition using support vector machines with local correlation kernels," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 16, no. 1, pp. 97–111, 2002.
- [30] J. K. Leski, "An ϵ -margin nonlinear classifier based on fuzzy if-then rules," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 68–76, Feb. 2004.
- [31] J. K. Leski, " ϵ -insensitive fuzzy c -regression models: Introduction to ϵ -insensitive fuzzy modeling," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 4–15, Feb. 2004.
- [32] J. K. Leski, "Neuro-fuzzy system with learning tolerant to imprecision," *Fuzzy Sets Syst.*, vol. 138, no. 2, pp. 427–439, 2003.
- [33] J. K. Leski, "TSK-fuzzy modeling based on ϵ -insensitive learning," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 2, pp. 181–193, Apr. 2005.
- [34] L. Liao, T. Lin, and B. Li, "MRI brain image segmentation and bias field correction based on fast spatially constrained kernel clustering approach," *Pattern Recognit. Lett.*, vol. 29, pp. 1580–1588, 2008.
- [35] C. F. Lin and S. D. Wang, "Fuzzy support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 464–471, Mar. 2002.
- [36] C. F. Lin and S. D. Wang, "Training algorithms for fuzzy support vector machines with noisy data," *Pattern Recognit. Lett.*, vol. 25, no. 14, pp. 1647–1656, 2004.
- [37] C. T. Lin, C. M. Yeh, S. F. Liang, J. F. Chung, and N. Kumar, "Support-vector-based fuzzy neural network for pattern classification," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 1, pp. 31–41, Feb. 2006.
- [38] J. B. McQueen, "Some methods for classification and analysis of multivariate observation," in *Proc. 5th Berkeley Symp. Math. Statist. Probability*, Berkeley, CA, 1976, vol. 1, pp. 281–297.
- [39] O. L. Mangasarian and D. R. Musicant, "Lagrangian support vector machines," *J. Mach. Learning Res.*, vol. 1, pp. 161–177, 2001.
- [40] K. Mizutani and S. Miyamoto, "Possibilistic approach to kernel-based fuzzy c -means clustering with entropy regularization," in *Proc. 10th Int. Conf. Rough Sets, Fuzzy Sets, Data Mining Granular Comput.*, Regina, SK, Canada, 2005, pp. 144–155.
- [41] N. R. Pal and J. C. Bezdek, "On cluster validity for fuzzy c -means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, Aug. 1995.
- [42] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c -means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, Aug. 2005.
- [43] S. N. Pang, D. Kim, and S. Y. Bang, "Membership authentication in the dynamic group by face classification using SVM ensemble," *Pattern Recognit. Lett.*, vol. 24, no. 1–3, pp. 215–225, 2003.
- [44] J. C. Platt, "Sequential minimal optimization—A fast algorithm for training support vector machines," in *Advances in Kernel Methods-Support Vector Learning*. Cambridge, MA: MIT Press, 1998, pp. 185–208.
- [45] M. R. Rezaee, B. P. F. Lelieveldt, and J. H. C. Reiber, "A new cluster validity index for the fuzzy c -mean," *Pattern Recognit. Lett.*, vol. 19, no. 3–4, pp. 237–246, 1998.
- [46] F. C. H. Rhee, K. S. Choi, and B. I. Choi, "Kernel approach to possibilistic c -means clustering," *Int. J. Intell. Syst.*, vol. 24, pp. 272–292, 2009.
- [47] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 1996.
- [48] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, pp. 1207–1245, 2000.
- [49] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: Robustness and sparse approximation," *Neurocomputing*, vol. 48, no. 1–4, pp. 85–105, 2002.
- [50] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.
- [51] W. C. Tjhi and L. H. Chen, "Dual fuzzy-possibilistic coclustering for categorization of documents," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 3, pp. 532–543, Jun. 2009.
- [52] M. Tushir and S. Srivastava, "A new kernel based hybrid c -means clustering model," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2007, pp. 1–5.
- [53] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [54] V. N. Vapnik and O. Chapelle, "Bounds on error expectation for support vector machines," *Neural Comput.*, vol. 12, no. 9, pp. 2013–2036, 2000.
- [55] Y. Q. Wang, S. Y. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 6, pp. 820–831, Dec. 2005.
- [56] X. D. Wu, *Knowledge Acquisition from Databases*. Norwood, NJ: Ablex, 1995.
- [57] X. H. Wu and J. J. Zhou, "Possibilistic fuzzy c -means clustering model using kernel methods," in *Proc. 2005 Int. Conf. Comput. Intell. Modell., Control Autom. Int. Conf. Intell. Agents, Web Technol. Internet Commerce*, Vienna, Austria, 2005, vol. 2, pp. 465–470.
- [58] Z. D. Wu, W. X. Xie, and J. P. Yu, "Fuzzy c -means clustering algorithm based on kernel method," in *Proc. 5th Int. Conf. Comput. Intell. Multimedia Appl.*, IEEE Computer Society, Washington, DC, 2003, pp. 49–54.
- [59] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
- [60] X. L. Yang, Q. Song, and M. J. Er, "Robust data clustering in Mercer kernel-induced feature space," in *Proc. 3rd Int. Symp. Neural Netw.*, Chengdu, China, 2006, pp. 1231–1237.
- [61] M. S. Yang and H. S. Tsai, "A Gaussian kernel-based fuzzy c -means algorithm with a spatial bias correction," *Pattern Recognit. Lett.*, vol. 29, pp. 1713–1725, 2008.
- [62] N. Zahid, M. Limouri, and A. Essaid, "A new cluster-validity for fuzzy clustering," *Pattern Recognit.*, vol. 32, pp. 1089–1097, 1999.
- [63] D. Q. Zhang and S. C. Chen, "A novel kernelized fuzzy c -means algorithm with application in medical image segmentation," *Artif. Intell. Med.*, vol. 32, pp. 37–50, 2004.
- [64] X. G. Zhang, "Using class-center vectors to build support vector machines," in *Proc. IEEE Signal Process. Soc. Workshop*. New York: IEEE Press, 1999, pp. 3–11.

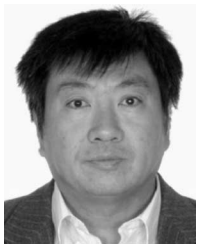
- [65] Y. Zhou, Y. Li, and S. X. Xia, "Robust fuzzy-possibilistic *c*-means algorithm," in *Proc. 2nd Int. Symp. Intell. Inf. Technol. Appl.*, Shanghai, China, 2008, pp. 669–673.



Xiaowei Yang received the B.S. degree in theoretical and applied mechanics, the M.Sc. degree in computational mechanics, and the Ph.D. degree in solid mechanics from Jilin University, Changchun, China, in 1991, 1996, and 2000, respectively.

From July 1991 to July 1993, he was with General Factory of Vibration Equipment of Xinxiang, Xinxiang, China. From July 1996 to May 2001, he was with the Department of Mathematics, Jilin University. Since June 2001, he has been with the Department of Mathematics, South China University of

Technology, Guangzhou, China, where he is currently a Professor. From January 2003 to May 2003, he was a Research Fellow with the Faculty of Mechanical Engineering, National University of Singapore, Singapore. From February 2006 to August 2006, he was a Visiting Scholar with the Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia, where he was also a Visiting Professor with the School of Software from July 2009 to July 2010. He has authored or coauthored more than 80 journal and refereed international conference articles. His research topics include the areas of structural reanalysis, interval analysis, soft computing, and support vector machines. His research interests include designs and analyses of algorithms for large-scale pattern recognitions, imbalanced learning, semisupervised learning, and evolutionary computation.



Guangquan Zhang received the Ph.D. degree in applied mathematics from Curtin University of Technology, Bentley, WA, Australia.

He is currently an Associate Professor with the Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia. From 1979 to 1997, he was a Lecturer, an Associate Professor, and a Professor with the Department of Mathematics, Hebei University, Baoding, China. He has authored or coauthored four monographs, four reference books, and more than

250 papers, including more than 120 refereed journal articles. His research interests include the area of multiobjective, bilevel and group decision making, decision support system tools, fuzzy measure, fuzzy optimization, and uncertain information processing.

Prof. Zhang was the recipient of four Australian Research Council discovery grants and many other research grants. He is also a Guest Editor of special issues for three international journals.



Jie Lu (M'05) received the Ph.D. degree from Curtin University of Technology, Bentley, WA, Australia, in 2000.

She is currently a Professor and a Director of the Decision Systems and e-Service Intelligence Research Laboratory, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia. She has authored or coauthored five research books and more than 250 papers in refereed journals and conference proceedings. Her research interests include the area of

intelligent information systems, decision-making modeling, decision support system tools, uncertain information processing, e-government, and e-service intelligence and personalization.

Prof. Lu was the recipient of four Australian Research Council discovery grants and many other research grants. She was an Editor-in-Chief for *Knowledge-Based Systems* and a Guest Editor of special issues for six international journals. She has also delivered four keynote speeches at international conferences.



Jun Ma received the Bachelor's and Master's degrees in applied mathematics from Southwest Jiaotong University, Chengdu, China, in 1999 and 1996, respectively.

He is currently a Research Associate with the Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia. He has authored or coauthored more than 40 publications in international journals and conferences. His research interests include automated and approximate reasoning with linguistic information

and their application in decision making.