J Electr Eng Technol.2018; 13(6): 2511-2520
http://doi.org/10.5370/JEET.2018.13.6.2511

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Multi-Radial Basis Function SVM Classifier: Design and Analysis

Zheng Wang *, Cheng Yang *, Sung-Kwun Oh† and Zunwei Fu**

Abstract – In this study, Multi-Radial Basis Function Support Vector Machine (Multi-RBF SVM) classifier is introduced based on a composite kernel function. In the proposed multi- RBF support vector machine classifier, the input space is divided into several local subsets considered for extremely nonlinear classification tasks. Each local subset is expressed as nonlinear classification subspace and mapped into feature space by using kernel function. The composite kernel function employs the dual RBF structure. By capturing the nonlinear distribution knowledge of local subsets, the training data is mapped into higher feature space, then Multi-SVM classifier is realized by using the composite kernel function through optimization procedure similar to conventional SVM classifier. The original training data set is partitioned by using some unsupervised learning methods such as clustering methods. In this study, three types of clustering method are considered such as Affinity propagation (AP), Hard C-Mean (HCM) and Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA). Experimental results on benchmark machine learning datasets show that the proposed method improves the classification performance efficiently.

Keywords: Multi-RBF SVM, Composite kernel, Dual RBF structure, Clustering method, Particle Swam optimization.

1. Introduction

Support Vector Machines (SVMs) is a set of related supervised learning models with associated learning algorithms which analyze data and recognize patterns for classification and regression, which is mainly grounded on the statistical learning theory [1]. By constructing a hyper-plane, SVMs is to maximize the margin of separation and also provide a good generalization performance on the real-world classification problems, in spite of the fact that it does not get any knowledge in the problem domain [3].

For SVM classification, there are some tasks with characteristics of high noise and a large number of input features, with a relatively small number of training data-sets. In these tasks, the distribution of dataset is filled with a variety of nonlinear characteristics. Nonlinear kernel SVM classifiers are also confronted with the potential over-fitting issue, especially for those linearly inseparable datasets in the high dimensional feature space. As a result, this would lead to severe over-fitting in the conventional nonlinear SVM classifiers.

To address this problem, "divide-and-conquer" algorithm is an important strategy and widely used as an available tool for solving conceptually difficult problems [4]. A "divide-and-conquer" algorithm works by breaking a difficult problem into two or more sub-problems. The solutions to the sub-problems are then combined to give a solution to the original problem [5]. A category of multiple linear SVM method based on subspace pattern recognition has been proposed to solve such over-fitting problem. In some previous works, [6] presented localized support vector machine which built up multiple linear SVM classifiers from training data and each model was designed to classify a particular test dataset. Z. Fu and A. Robles-Kelly proposed a mixture of linear SVMs by packaging linear SVMs into a probabilistic formulation and embedding them in the mixture of expert models [7]. [9] proposed a quasi-linear SVM by first composing a quasi-linear kernel function and then realizing multi-local linear classifiers with interpolation by implementing the SRM principle in the same way as a standard SVM.

In this study, a multi-RBF SVM classifier based on a composite multi-RBF kernel function is introduced. The proposed multi- RBF SVM methods is realized with the aid of the division of the entire input space into several subspaces, and each subspace is estimated by a RBF kernel function. Unlike as the conventional multi SVM classifier, the proposed multi-RBF SVM classifier incorporates a RBF mapping, which is nested as the combination of several input subspaces. The input space will be mapped into much higher feature spaces and become separable according to the "divide-and-conquer" strategy.

Unlike the conventional Multi-SVM methods designed by constructing multiple local SVM classifiers separately,

† Corresponding Author: School of Electrical and Electronic Engineering, The University of Suwon, Korea, and Key Laboratory of Complex Systems and Intelligent Computing in Universities of Shandong, Linyi University, Linyi, 276005, China (ohsk@suwon.ac.kr)
* School of Electrical and Electronic Engineering, The University of Suwon, Korea ({wangzheng, lyyc58}@suwon.ac.kr)
** Dept. of Computer Science, The University of Suwon, Korea, and Key Laboratory of Complex Systems and Intelligent Computing in Universities of Shandong, Linyi University, Linyi, 276005, China, and Department of Mathematics, Linyi University, Linyi, Shandong, 276005, China (zwfu@suwon.ac.kr)
Received: April 30, 2018; Accepted: July 19, 2018

the prior knowledge of local subsets is used to build a composite RBF kernel. Then the Multi-RBF SVM classifier is realized by using the composite kernel exactly in the same way as a single SVM classifier.

This paper includes the two main aspects as following.

Firstly, a novel Multi-RBF SVM classifier is introduced to map the nonlinear classification task into much higher feature space. The nested kernel structure has dual mapping process to a much higher-dimensional feature space for some specific nonlinear classification tasks. The multi-RBF SVM can make these tasks linearly separable for solving over-fitting problem.

Secondly, a composite kernel function with nested RBF structure is adopted. According to the distribution information of the training data, several local subsets from training data are obtained by using some clustering techniques for performance enhancement. Then, the composite kernel function with dual RBF mapping performs in the same way as a standard SVM.

The rest parts of the paper are organized as follows. Section 2 describes the conventional SVM Classifier. Section 3 formulates the Multi-RBF Classifier based on a Composite Kernel function. Section 4 discusses the preprocessing of the proposed multi-RBF SVM. Section 5 addresses the implementation of multi-RBF kernel for the SVM classifier in nested structure. Section 6 implements the experiment and results; and finally the conclusions are given in Section 7.

2. Conventional SVM Classifier

SVM is a typical two class classifier, that is, it only answers the problem that belongs to the positive class or the negative class.

Suppose there is a training set $\{x_i, y_i\}_{i=1}^{N}$: input $x_i \in R^n$; class labels $y_i \in \{-1,1\}$, with a hyper-plane $\langle w,x\rangle + b = 0$. Where x_i is the value of the ith input, y_i is the corresponding class label of the input x_i, which should be -1 or 1. The data set can be divided into two different sets A and B which have labels 1 and -1 respectively.

The set of vectors is said to be optimally separated by the hyper plane if it is separated without error and the distance between the closest vectors to the hyper plane is maximal.

Without loss of generality it is appropriate to consider a canonical hyper-plane, where the parameters, (w, b), are constrained by,

$$\min_i |(w,x^i)+b| = 1 \tag{1}$$

2.1 Linear SVM

To linear cases, a separating hyper plane must satisfy the following constraints,

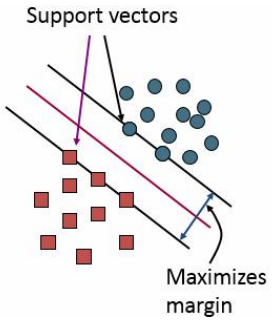


Fig. 1 Linear SVM

$$y^i[\langle w,x^i\rangle + b] \geq 1, i = 1,...,l. \tag{2}$$

where the Eq. (2) could be demonstrated in Fig. 1

As shown in Fig. 1, SVM maps the training points into a high-dimensional feature space and finds a separating hyper-plane that maximizes the margin between two classes.

2.2 Non-linear SVM

Usually the classification problems are non-linear and non-separable in the real world, so in the primal space, the low dimension input data sets is transformed into a high dimensional feature space using a mapping function $\varphi(\cdot)$, as Fig. 2 shows.

For non-separable cases in the feature space, the boundary function has a nonnegative variable as a slack variable ξ_i to make the margin to accept the violations, and then get a separating plane function,

$$y^i[\langle w,x^i\rangle + b] \geq 1-\xi_i, i = 1,...,l. \tag{3}$$

The optimal hyper plane problem becomes to find the solution of the following optimization problem,

$$\begin{aligned} &\min_{\omega,b,\xi} J(\omega,\xi) = \frac{1}{2}\omega^T\omega + c\sum_{i=1}^{N}\xi_i \\ &s.t \quad \begin{cases} y_i[w^T\psi(x_i)+b] \geq 1-\xi_i, i = 1,...,N \\ \xi_i \geq 0, k = 1,...,N \end{cases} \end{aligned} \tag{4}$$

where parameter c is used to control the degree of regularization, which is the only changeable parameter in SVM.

Introducing the vector of Lagrange multipliers $\alpha = (\alpha_1,...,\alpha_N)$, the Eq (4) above can be considered as a QP problem in dual space,

$$\begin{aligned} &\max_{\alpha} Q(\alpha) = -\frac{1}{2}\sum_{i,j=1}^{N} y_i y_j K(x_i,x_j)\alpha_i\alpha_j + \sum_{j=1}^{N}\alpha_j \\ &s.t. \quad \begin{cases} \sum_{i=1}^{N}\alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq c, \forall i \end{cases} \end{aligned} \tag{5}$$

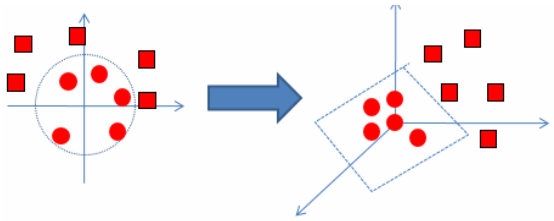


Fig. 2. Nonlinear SVM (via. mapping two dimensional into three dimensional)

where $K(x_i, x_j) = \phi(x_i)^T \phi(x_i)$ is the kernel function.

In the end, the decision making function can be gained.

$$y = sign[\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b] \tag{6}$$

Fig. 2 shows that the kernel method helps SVMs to deal with the linearly non-separable problems by mapping data into higher dimension feature space.

3. The Design and Analysis of Multi-RBF SVM Classifier based on Composite Kernel

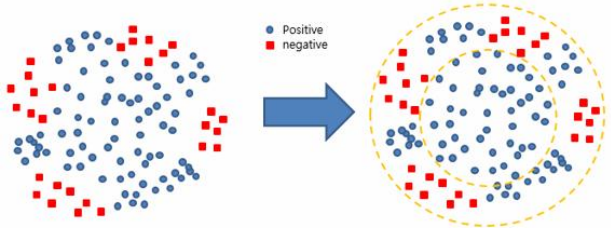
In this section, the expression of Multi-RBF kernel function is formulated.

If there are many discrete nonlinear characters in the dataset, or it is especially complex, conventional SVM can not obtain the described classification performance. As shown in Fig. 3, the yellow dotted line is considered as classification boundary in two dimensions (the classification hyper-plane in three dimensions). The dataset is not classified accurately according to the performance of both classification boundaries.

A proper solution is to divide the original training dataset into several subsets, and each subset is estimated by a nonlinear mapping. The combined classification result is more accurate for this kind of dataset.

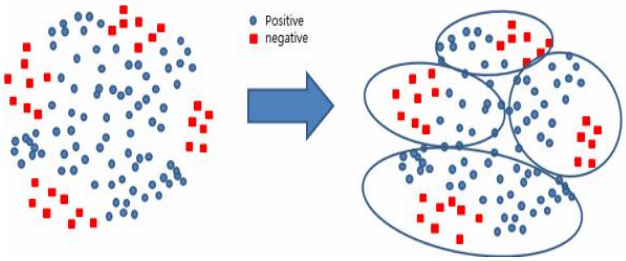
Then, the formulation is given as follows.

A nonlinear classification task can be treated as an aggregation of M nonlinear functions of curve fitting $A_j x^a e^{\frac{x^2}{\sigma}} + B_j, j = 1, ..., M$, as shown in Fig. 3.



(a) (b)

Fig. 3. Conventional SVM classification: (a) Artificial dataset filled with nonlinear characteristic; (b) Classification boundary constructed by conventional SVM with nonlinear kernel function



(a) (b)

Fig. 4. Multi-RBF classification; (a) Artificial dataset filled with nonlinear characteristic: (b) Classification boundary constructed by multi-RBF SVM with multiple classification boundary constructed by composite kernel function

Based on the piecewise linear approximation method, the piecewise linear model $f_P(x)$ can be written compactly as Eq. (7).

$$f_P(x) = \sum_{j=1}^{M} (A_j x^a e^{\frac{x^2}{\sigma}} + B_j) R_j(x) + b \tag{7}$$

where $R_j(x)$'s are the basis function, $A_j x^a e^{\frac{x^2}{\sigma}}$ is the curve fitting method of the nonlinear function.

The basis function is an element of a particular basis for a function space. Typical examples are wavelet basis and RBF basis. The final results of the nonlinear model has been obtained via an interpolation using the basis function $R(x)$. It also implies that the severe nonlinear model $f_P(x)$ can approximate an arbitrary nonlinear separating hyper-plane on a compact interval by merely increasing the value of M.

Assume parameter vectors $\Phi(x)$ and Θ as follows:

$$\begin{aligned} \Phi(x) &= [R_1(x), x^a e^{\frac{x^2}{\sigma}} R_1(x), ..., R_M(x), x^a e^{\frac{x^2}{\sigma}} R_M(x)]^T \\ \Theta &= [B_1, A_1, ..., B_M, A_M]^T \end{aligned} \tag{8}$$

Furthermore, the above parameter vectors $\Phi(x)$ and Θ is combined into the Eq. (7), the piecewise linear model can be transformed:

$$f_P(x) = \Theta^T \Phi(x) + b \tag{9}$$

The Structural Risk Minimization principle is introduced in the Eq. (9), the classification problem can be described as the QP optimization problem as following:

$$\begin{aligned} &\min_{\Theta, b, \xi} J_P = \frac{1}{2}\Theta^T\Theta + c\sum_{k=1}^{N} \xi_k \\ &s.t \quad \begin{cases} y_k[\Theta^T \Phi(x_k) + b] \geq 1 - \xi_k, k = 1, ..., N \\ \xi_k \geq 0, k = 1, ..., N \end{cases} \end{aligned} \tag{10}$$

The Lagrange function has been constructed, via introducing new variables (α_k , v_k) called Lagrange

multipliers:

$$L(\Theta, b, \xi; \alpha, v) = J_P(\Theta, \xi) - \sum_{k=1}^{N}(\alpha_k y_k[\Theta^T \Phi(x_k) + b] - 1 + \xi_k) - \sum_{k=1}^{N} v_k \xi_k \quad (11)$$

with Lagrange multipliers: $\alpha_k \geq 0$, $v_k \geq 0$, $k = 1,..., N$.

The solution is given by the saddle point of the Lagrange function:

$$\max_{\alpha, v} \min_{\Theta, b, \xi} L(\Theta, b, \xi; \alpha, v) \quad (12)$$

The dual problem becomes as follows:

$$\max_{\alpha} J_D(\alpha) = -\frac{1}{2}\sum_{k,l=1}^{N} y_k y_l K(x_k, x_l)\alpha_k \alpha_l + \sum_{k=1}^{N} \alpha_k$$
$$s.t. \quad \begin{cases} \sum_{k=1}^{N} \alpha_k y_k = 0 \\ 0 \leq \alpha_k \leq c, k = 1,..., N \end{cases} \quad (13)$$

In the quadratic form, the kernel trick is applied

$$K(x_k, x_l) = \Phi(x_k)^T \Phi(x_l) = (1 + x_l{}^a x_k{}^a e^{\frac{x_k{}^2 + x_l{}^2}{\sigma}}) \sum_{j=1}^{M} R_j(x_k) R_j(x_l) \quad (14)$$

Where a is constant according the nonlinear distribution property of training data. Hence, the non-linear model $f_P(x)$ is reduced to a standard SVM based on a composite kernel (Eq. 14) and each $K(x_k, x_l)$ can express a input subspace of training data. The nonlinear SVM classifier takes the following form:

$$y = sign[\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b] \quad (15)$$

with α_k positive real constants which are the solution to a QP problem.

In order to represent an inner product, the kernel is required to satisfy Mercer's condition [1]. In Eq. 14, factors $R_j(x_k)R_j(x)$ are outputs of basic functions. It follows that the kernel $K_c(x, x_i)$ satisfies Mercer's condition because it consists of a sum of products of Mercer kernels, cf. [1].

Standard SVM classifier with nonlinear mapping is often written as an implicit expression, especially for the Gaussian kernel, while the multi-RBF SVM is often described the nonlinear mapping with explicit formulation.

4. Preprocessing of Multi-RBF SVM Classifier

In multiple classifier systems, the original training data set is partitioned by using some unsupervised learning methods such as clustering methods and so on. In this study, three clustering methods are considered as preprocess of the proposed multi-RBF SVM classifier. Here a brief summary of these three methods is given.

4.1 Hard C-mean clustering method

Hard C-Mean Clustering Method (HCM) clustering method is a recently introduced clustering method which takes as input a set of measures of similarity between pairs of data points and outputs a set of clusters of the points with their balanced exemplars. In order to study the distribution of the data points in a high-dimensional feature space, Hard C-Means(HCM) is introduced into the proposed Multi-RBF SVM. The detailed computational steps are as follows.

Fix the number of clusters c(2≤c≤n) and initialize the partition matrix $U^{(0)} \in M_C$

Calculate the center vectors $\boldsymbol{v}_i$ of each cluster:

$$\boldsymbol{v}_i = \{v_{i1}, v_{i2}, \cdots, v_{ij}, \cdots, v_{im}\}$$
$$v_{ij} = \frac{\sum_{k=1}^{n} u_{ik} * x_{kj}}{\sum_{k=1}^{n} u_{ik}} \quad (16)$$

where, $[u_{ik}] = U, i = 1,2 \ldots . c \; ; k = 1,2 \ldots . m$

Update the partition matrix U ; these modifications are based on the standard Euclidean distance function between the data points and the prototypes,

$$d(x_k - v_i) = x_k - v_i = \left[\sum_{j=1}^{m}\left(x_{kj} - v_{ij}\right)^2\right]^{\frac{1}{2}} \quad (17)$$

$$u_{ik}^{(r+1)} = \begin{cases} 1 \;\; if \;\; d_{ik}^{(r)} = \min\left\{d_{jk}^{(r)}\right\} \;\; for \;\; all \;\; j \in c \\ 0 \qquad\qquad\qquad otherwise \end{cases} \quad (18)$$

Check the termination criterion. If

$$\| U^{(r+1)} - U^{(r)} \| <= \delta \quad \text{(tolerance level)} \quad (19)$$

the process is end. When otherwise set r=r+1 and return to calculate the new center vector. In step 4 the notation || || is any matrix norm such as the Euclidean norm.

4.2 Affinity propagation clustering

Affinity Propagation (AP) [7] is a recently introduced clustering method which takes as input a set of measures of similarity between pairs of data points and outputs a set of clusters of the points with their corresponding exemplars. The algorithm takes a matrix of similarity measures between each pair of points $s(i, k)$ as input. Instead of

requiring that the number of clusters be predetermined, the AP takes as input a real number $s(k, k)$ for each data point k. These values, which are called preferences, are a measure of how likely each point is to be chosen as exemplar.

The algorithm works by exchanging messages between the points until a stop condition is satisfied. There are two types of messages to be exchanged between data points. The responsibility $r(i,k)$, sent from data point i to candidate exemplar point k, reflects the accumulated evidence for how well-suited point k is to serve as the exemplar for point i, taking into account other potential exemplars for point i. The availability $a(i, k)$, sent from candidate exemplar point k to point i, reflects the accumulated evidence for how appropriate it would be for point i to choose point k as its exemplar, taking into account the support from other points that point k should be an exemplar.

The availabilities are initialized to zero: $a(i, k) = 0$. Then, the parameters are computed and updated using the rules as follows:

$$\begin{aligned} r(i,k) &\leftarrow s(i,k) - \max_{k's.t.k'\neq k}\{a(i,k') + s(i,k')\} \\ a(i,k) &\leftarrow \min\{0, r(k,k) + \sum_{i's.t.i'\neq\{i,k\}} \max\{0, r(i',k)\}\} \\ a(k,k) &\leftarrow \sum_{i's.t.i'\neq\{i,k\}} \max\{0, r(i',k)\} \end{aligned} \tag{20}$$

The AP method has been praised because of its ability to efficiently and quickly handle clustering problems.

4.3 ISODATA clustering method

In the HCM clustering method, the number of clusters c remains the same throughout the iteration, although it may turn out later that more or fewer clusters would fit the data better. This drawback can be overcome in the ISODATA Algorithm (Iterative Self-Organizing Data Analysis Technique Algorithm) [12], which allows the number of clusters to be adjusted automatically during the iteration by merging similar clusters and splitting clusters with large standard deviations. The algorithm is highly heuristic and based on the following pre-specified parameters

The detailed computational steps are as follows.

Choose randomly $K = K_0$ initial mean vector $\{m_1, m_2, \cdots m_k\}$ from the data set.

Assign each data point X to the cluster with closest mean $x \in \omega_i$ if $d(x, m_i) = \min\{d(x, m_1), d(x, m_2)\cdots, d(x, m_K)\}$.

Discard clusters containing too few members, i.e., if $n_j < n_{min}$, the discard ω_j and reassign its members to other clusters $K \leftarrow K - 1$.

For each cluster $\omega_j\ (j = 1, 2, \ldots, K)$ date the mean vector $m_j = \frac{1}{n_j}\sum_{x\in\omega_j} x$ and the covariance matrix:

$$\sum_j = \frac{1}{n_j}\sum_{x\in\omega_j}(x - m_j)(x - m_j)^T \tag{21}$$

The diagonal elements are the variances $\sigma_1^2, \sigma_2^2, \cdots \sigma_N^2$ along the N dimensions.

If $K \le K_0 / 2$ (too few clusters), go to splitting.

Else if $K > 2K_0$ (too many clusters), go to merging.

Else terminate if maximum number of iterations is reached.

5. Structural Design of Multi-RBF SVM Classifier Based on PSO

5.1 Structure of multi-RBF SVM classifier

Multi-RBF SVM is kind of multiple SVM which employ composite kernel function to estimate the dual mapping operation of between input and output. Firstly, the clustering

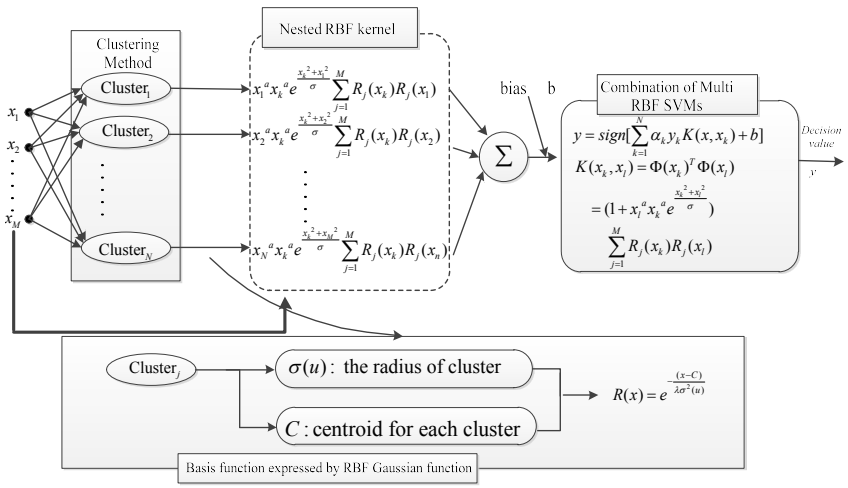


Fig. 5. Overall architecture of multi-RBF SVM classifier

method is implemented to partition the original dataset into several subsets.

According the description of proposed Multi-RBF SVM in previous, basis functions is used to interpolate the non-linear hyper-planes, and each basis function corresponds to a local linear hyper-plane. Therefore, the basis functions should remark the distribution information of dataset as accurate as possible. A RBF Gaussian function is chosen as basic function to capture distribution of partitioned subsets.

$$R(x) = e^{-\frac{(x-\mu)}{\lambda\sigma^2(u)}} \quad (23)$$

Where μ is the cluster center, σ is the radius for clusters. λ is the scale parameter.

There is also an important parameter in the multi-RBF SVM is the number of RBF Gaussian functions M. the number is obvious equal to the number of partitioning clusters. Theoretically, the composite kernel function is slide from linear kernel to nonlinear kernel. When M=1, composite kernel is acted as linear one. When $M = \infty$, composite kernel can be smoothly fit for nonlinear separating hyper-plane. If a proper number M is chosen along the nonlinear distribution of separating boundary, classification accuracy of the multi-RBF SVM classifier would be better.

5.2 Optimization procedure of multi-RBF SVM classifier

The design procedure of multi-RBF SVMs comprises the following steps.

[Step 1] Preprocessing data set using regularization

For most datasets, that have many dimensions in various ranges. To obtain data in similar scale, each data point would compare with the largest number in its dimension, and get the mapped value from -1 to 1.

[Step 2] Form training and testing data set

To make full use of data set and result the classification accuracy be no-biased, the dataset is divided into five parts. Each part is utilized for testing data, and the whole dataset will evaluate the quality of network. In this way, the data set can be take full advantage that there is not enough data. Furthermore, two performance indexes is introduced viz. the standard Root Mean Square Error(RMSE), and the Mean Squared Error (MSE)

[Step 3] Construct for multi-RBF SVM by modified parameters.

1. Choose the data point along the separating boundary in the interaction area between two classes
2. Select clusters $\omega_1, \omega_2, ..., \omega_M$ from training set in the training set in high-dimensional input space by clustering algorithm.
3. The several clusters are obtained to capture nonlinear distribution information of training dataset. A composite kernel is constructed to realize nested RBF SVM classifier by incorporating RBF function in each cluster. It should be noted that the difference between the new kernel function and the RBF is that the composite kernel implements the duple mapping to obtain separating hyperplane in much higher feature space.
4. The class label of a testing example x is predicted by (15) and (23). Finally, the binary classification of local SVMs will obtain the final decision.

[Step 4] Decide the PSO parameters used in the optimization.

Table 1. Parameter setting of PSO

PSO parameters		Values
Number of generations		20
Swarm size		20
Number of particles		3
Inertia weight		0.6
Acceleration constants		C1=2.0, C2=2.0
Objective function		Classification rate of training data
Boundary of search space	Learning rate	10e-3~10
	Momentum coefficient	10e-3~10

Here we decide upon the essential design parameters of PSO. Different from traditional PSO, M(the number of RBF Gaussian functions) , c(cost), λ (scale parameter), g (gamma) is chosen as optimization parameters. M is chosen as an integer and its range from 2 to 8. Other parameters would take random value as initialization.

For ISODATA algorithm, there are three extra parameters need be optimized. They are minimum number of one cluster, maximum variance of one cluster and minimum distance in one cluster. The range of these three parameters depends on the datasets.

[Step 5] Check the Termination criterion

As far as the accuracy is concerned, the generation of process is stopped a comparison of less than objective.

[Step 6] Output of the finally results.

An overall design flowchart for the proposed multi-RBF SVM is shown in Fig. 6

6. Experimental Studies

First, graphical experimental results are presented based on lib-SVM on the hayes data set in ordered to understand the behaviors of the proposed algorithm. Second, 15 benchmark machine learning data sets are used to experiments.

6.1 Graphical representation and its entity analysis

In the first experiment, proposed multi-RBF SVM and conventional SVM on the hayes dataset are compared.

Fig. 6 illustrates the classification results of the proposed classifier and conventional SVM on Benchmark dataset (hayes). In hayes dataset, we pick two class labels from

the original data and reduced into two dimensions. It can be easily seen that positive data and negative data are mixed together. It is different to distinguish one class of data from another by their properties. Fig. 6(b) shows the classification boundary produced by conventional (RBF) SVM.

The classification boundary distinguishes different label of classes around the bound area. In Fig. 6(c), the original dataset is divided into several subsets, and each subset is classified by a RBF kernel function. The classification results of these subsets combined with dual RBF structure are mapped into much higher feature space to make data separable. It is can be seen that the proposed multi-RBF SVM is more accurate for this dataset which filled nonlinear characters.

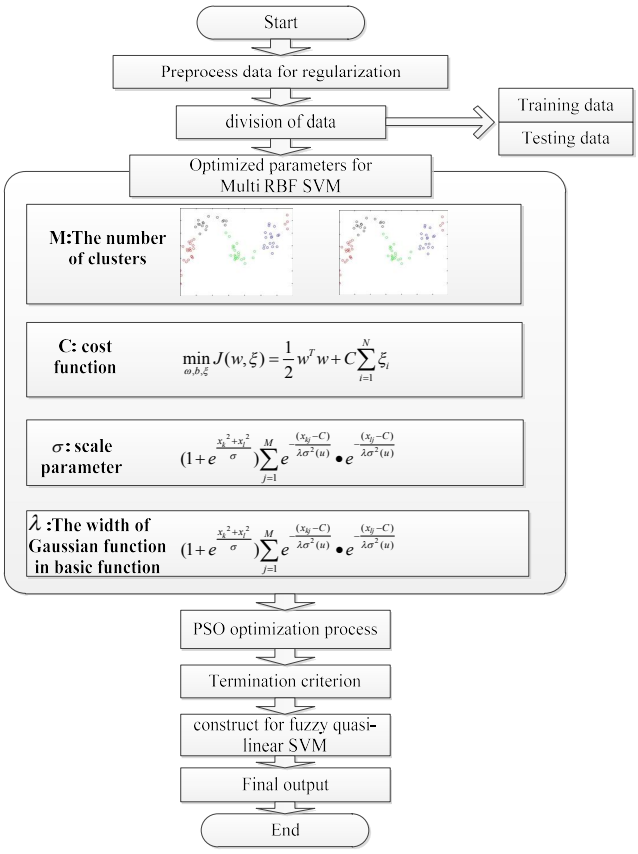


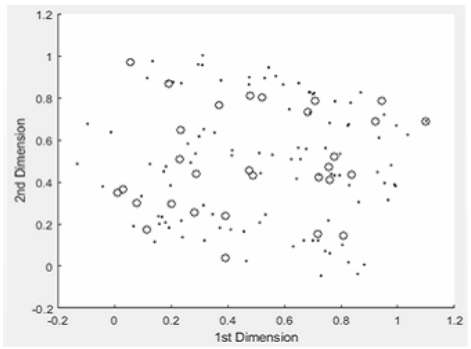
Fig. 6. Overall flowchart of design procedure of multi-RBF SVM classifier

Table 2. Summary of 10 benchmark data sets

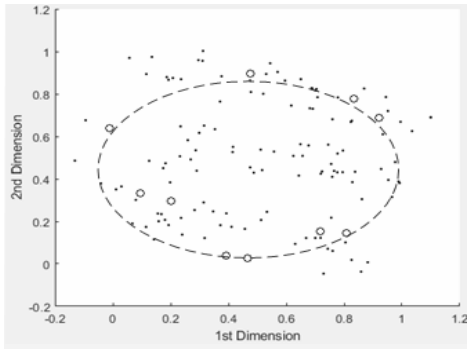
Data	No. of data	No. of Train	No. of Test	Label 1	Label 2	Label 3	Dimension
iris	150	120	30	50	50	50	4
Heart scale	270	216	54	150	120	/	12
vehicle	846	170	676	199	217	218/ 212	18
Segment	2301	1848	462	1150	1156	/	19
blood	749	599	150	178	571	/	4
mpg	392	313	79	245	68	79	7
oil	81	17	64	41	5	15/20	5
ionosphere	351	280	71	223	128	/	34
hayes	132	106	26	51	51	30	5
vowel	990	792	188	495	495	/	10

6.2 Benchmark data set

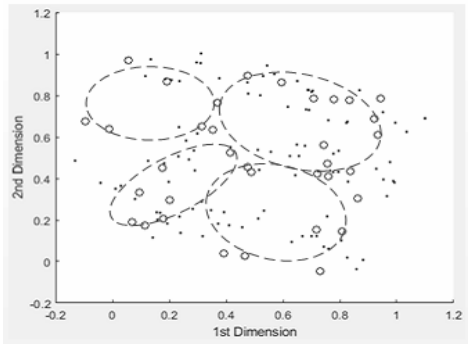
In order to evaluate the performance of the method further, we compare different algorithms on 10 well-known benchmark problems. The data sets are all preprocessed for



(a) Hayes data



(b) Conventional RBF SVM classification



(c) Multi RBF SVM classification

Fig. 7. Illustration of classification process of SVM and the proposed Multi-RBF SVM on hayes data set: (a) Data set filled with nonlinear characteristic: positive and negative data points are mixed together; (b) Circular classification boundary: data point around separating boundary can be correctly identified; (c) Multi Circular classification boundary: more data points around separating boundary can be correctly identified

binary classification simulations. These data sets include 10 benchmark real-world data sets from the University of California at Irvine (UCI) Machine Learning Repository.

A summary of the data sets is presented in Table 2.

All of the results are the average of 100 runs on the data sets. In addition, the data sets are all preprocessed to make the instances to have zero mean and unit standard deviation.

To make full use of data set and result the classification accuracy be no-biased, in all experiments, 5folds-Cross-Validation is used that dataset is randomly divided the into the training (80%) and testing (20%) part. Finally, to come up with a quantitative evaluation of the resulting proposed classifier. Two performance indexes are considered viz. the mean accuracy(MA) and the standard deviation (SD).

$$(MA \pm \mathrm{SD}) = \begin{cases} MA = \dfrac{\sum_{i=1}^{m} accuracy_i}{m} \\ SD = \sqrt{\dfrac{1}{m}\sum_{i=1}^{m}(accuracy_i - MA)^2}, \end{cases} \quad (24)$$

In this experiment, the total 10 benchmark datasets are employed in order to evaluate the performance of the conventional SVM classifier and other popular classifier are shown in Table 3. These dataset was downloaded from UCI machine learning laboratory website [36]. RandomTree, IBK (k=3) and NaiveBayes classifiers are used for comparison. The number of clusters needs to be optimized by PSO.

As shown in Table 3, Boldface entries in cells indicates better classification accuracy when compare with other SVM classifier.

Compare with conventional SVM, the proposed Multi-RBF SVM obtains the better performance when compared with some conventional SVM classifiers. The multi-RBF SVM with aid of HCM, AP and ISODATA clustering has 7, 6 and 5 best classification performances respectively. The classification performances show HCM clustering is better than ISODATA clustering method in benchmark dataset. The simple clustering method is best partition on these kinds of dataset.

ISODATA which is considered as supplement of HCM clustering do not show better performance. The reason for this is that ISODATA has too many parameters to optimized. It is hard to obtain proper parameters for ISODATA.

It is mentioned that hayes dataset has complex distribution and relatively small number input variables. AP and ISODATA can not obtain the proper partition on hayes.

When compare with other SVM classifier (RandomTree, IBK and bayes), the performance is improved on 9 results in 10 datasets, such as Ionosphere, vowel, Vehicle. These datasets have strong property of nonlinear distribution, whereas the classification accuracy of testing dataset is enhanced.

For oil dataset, the proposed Multi-RBF SVM is worse than other models. Oil dataset has relatively simple distribution with 80 inputs and 5 dimensions. The proposed Multi-RBF SVM can not match appropriately with too complex structure. Here the summary on these three clustering methods involved in this study is described as follows: K-means is a simple clustering method, which can obtain the better performance in small dataset (The number of input variables is less than 500). ISODATA and AP clustering can achieve better performance in dataset which have more than 500 input variables. Especially, AP clustering leads to preferred performance in vowel dataset because of the possibility of obtaining relatively more proper clusters in voice text related dataset (viz. vowel).

7. Concluding Remarks

In this study, a multi-RBF SVM classification classifier which is based on dual structure and unsupervised learning methods is introduced. To construct the classification classifier, firstly the original training set is divided into the balanced subsets by applying clustering techniques. Then, the local subsets have nonlinear characteristic and are used to estimate a composite kernel, and the composite kernel is utilized to realize the multiple local RBF SVM classifier by implementing the structural risk minimization in the same way as a single standard SVM. An important strength of the proposed classification approach is given as multi-RBF

Table 3. Comparison of the proposed Multi-RBF SVM , conventional SVM and other classifiers(MA±SD).

data	Conventional SVM(weka)[35]			Alternative classifiers(weka) [35]			Multi-RBF SVM (proposed)		
	linear	polynomial	RBF	RandomTree	IBK	bayes	HCM	AP	ISODATA
iris	96.67 ±0.41	71.00±0.43	95.33 ±0.16	94.13±0.18	94.80±0.16	95.33±0.15	97.33±0.03	96.00±0.04	96.67±0.03
heart scale	82.96 ± 0.41	64.26±0.60	55.74±0.67	70.56±0.29	78.15 ± 0.25	84.63± 0.36	84.81± 0.61	55.56±1.78	54.81±1.81
vehicle	74.29 ±0.36	31.27±0.59	58.98±0.45	70.00±0.39	69.88±0.39	44.68±0.46	79.79±0.75	79.79±0.75	79.79±0.75
segment	92.64± 0.14	76.14±0.26	87.77± 0.19	95.06±0.12	95.84 ±0.10	80.12 ±0.23	95.84± 0.15	96.84±0.12	96.84±0.13
blood	76.14 ±0.49	76.20±0.49	76.20±0.49	71.58±0.52	71.42±0.52	75.19±0.43	76.23±0.23	76.36±0.23	76.36±0.23
mpg	71.06 ± 0.44	62.50±0.50	67.73± 0.46	79.09±0.37	72.75±0.42	67.61±0.40	79.85±0.44	77.27±0.50	78.55±0.48
oil	92.57 ±0.17	67.28±0.40	80.26±0.31	95.35±0.10	98.79±0.05	98.04±0.05	96.67±0.03	96.67±0.03	96.67±0.03
ionosphere	88.05 ±0.34	66.67±0.58	86.19±0.37	86.90±0.28	87.01±0.36	82.33±0.39	93.44±0.07	93.44±0.07	93.44±0.07
hayes	52.66 ±0.55	70.05±0.44	75.03±0.40	72.61±0.42	65.79±0.47	80.93±0.35	84.62±0.30	49.29±0.81	48.46±0.91
vowel	68.54±0.24	87.07±0.15	77.07±0.20	74.49±0.22	94.75±0.09	54.90±0.23	95.25± 0.53	96.06±0.60	95.75±0.56

kernel function for the preprocessing aggregation of local clusters distribution. The learning method offers the advantages of the "divide-and-conquer" framework, i.e., extremely classification models may be employed that can be trained in parallel on complex (and usually hard to discriminate) training sets. The proposed multi-classifier methodology developed in this work could be applied for more effective as well as efficient system implementation through the clustering and classification techniques.

Acknowledgements

This work was supported by the GRRC program of Gyeonggi province under Grant number GRRC Suwon 2018-B2 Center for U-city Security & Surveillance Technology and also partially by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03032333).

References

[1] N. Cristianini, J. Shawe-Taylor, "An Introduction to Support Vector Machines," *Cambridge*, UK: Cambridge Univ. Press, 2000.

[2] Brodley CE, Friedl MA (1999) "Identifying mislabeled training data," *J Artif Intell Res*, 11:131-167.

[3] E. Osuna, R. Freund, F. Girosi. "Support Vector Machines: Training and Applications," *In A.I. Memo 1602*, MIT A.I.Lab., 1997.

[4] Cormen T T, Leiserson C E, Rivest R L. Introduction to algorithms =[M]. MIT Press, 2002.

[5] Li B, Wang Q, Hu J., "Multi-SVM classifier system with piecewise interpolation[J]." *Ieej Transactions on Electrical & Electronic Engineering*, 2013, vol. 8, no. 2, pp. 132-138.

[6] H. Cheng, P. Tan, and R. Jin, "Efficient algorithm for localized support vector machine," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 4, pp. 537-549, 2010.

[7] Z. Fu and A. Robles-Kelly, "On mixtures of linear SVMs for nonlinear classification," *Lecture Notes in Computer Science*, vol. 5342, pp. 489-499, 2010.

[8] D. S. Frossyniotis, and A. Stafylopatis, "A Multi-SVM Classification System," *Lecture Notes in Computer Science*, vol.2096/2001, pp. 198-207, 2001.

[9] B. Chen, F. Sun, and J. Hu, "Local linear multi-SVM method for gene function classification," *in Proc of the World Congress on Nature and Biologically Inspired Computing* (NaBIC'10) (Kitakyushu), Dec. 2010, pp. 183-188.

[10] C. Lin and S. Wang, "Fuzzy support vector machine," *IEEE Trans. Neural Netw*., vol. 13, no. 2, pp. 464-471, Mar. 2002.

[11] Y. Chen and J. Wang, "Support vector learning for fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst*., vol. 11, no. 6, pp. 716-728, Dec. 2003.

[12] J. Chiang and P. Hao, "Support vector learning mechanism for fuzzy rulebased modeling: A new approach," *IEEE Trans. Fuzzy Syst*., vol. 12, no. 1, pp. 1-12, Feb. 2004.

[13] Y. Wang, S. Wang, and 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.

[14] Q. Fan, Z. Wang, D. Li, D. Gao, and H. Zha, "Entropy-based fuzzy support vector machine for imbalanced data sets," *Knowl.-Based Syst*., vol. 115, pp. 87-99, 2017.

[15] X. Yang, L. Han, L. Yan, and L. He, "A bilateral-truncated-loss based robust support vector machine for classification problems," *Soft Comput.*, vol. 19, no. 10, pp. 2871-2882, 2015.

[16] S. Chen and X. Wu, "A new fuzzy twin support vector machine for pattern classification," *Int. J. Mach. Learn. Cybern*., vol. 3, pp. 1-12, 2017.

[17] S. Abe, T. Inoue, "Fuzzy Support Vector Machines for Multiclass Problems," *Proc. of ESANN'2002, Belgium*, pp. 113-118, 2002.

[18] Y. Zhung, S. W. Wu, Y. L. Wang, W. X. Wu, and Y. L. Chen, "Source separation of household waste: A case study in China," *Waste Management*, vol. 28, pp. 2022-2030, 2008.

[19] S.-B. Roh, S.-K. Oh, "Identification of plastic wastes by using fuzzy radial basis function neural networks classifier with conditional fuzzy C-means clustering," *Journal of Electrical Engineering and Technology*, vol. 11, no. 6, pp. 1872-1879, 2016.

[20] J. Kuiligowski, G. Quintas, S. Garrigues, and M. de la Guardia, "New background correction approach based on polynomial regressions for on-line liquid chromatography-Fouirer transform infrared spectroscopy," *Journal of Chromatography A*, vol. 1216, pp. 3122-3130, 2009.

[21] J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, and J. Tan, "Asymmetric least squares for multiple spectra baseline correction," *Analytica Chimica Acta*., vol. 683, pp. 63-68, 2010.

[22] C. Liu, S. X. Yang, L. Dong, "A comparative study for least angle regression on NIR spectra analysis to determine internal qualities of novel oranges," *Expert Systems with Applications*, vol. 42, pp. 8497-8503, 2015.

[23] Yang X, Zhang G, Lu J, et al. "A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises[J]," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 105-115, 2011.

[24]. V. Vapnik, The Nature of Statistical Learning Theory. Springer, Verlag Berlin, 1999.

[25] O. Chapelle and V. Vapnik, "Model selection for

Support Vector Machines," *Adv. Neural Inf. Proc. Syst*. 12, Cambridge, MA, MIT Press,2000.
[26] B. Li, Q. Wang and J. Hu, "A fast SVM training method for very large datasets," *IJCNN*, *2009 International Joint Conference on Neural Networks*, pp. 1784-1789, 2009.
[27] E. Osuna, R. Freund and F. Girosi., "Support Vector Machines: Training and Applications," *A.I. Memo 1602*, MIT A.I.Lab., 1997.
[28] Christian Platzer, Florian Rosenberg, and Schahram Dustdar, "Web Service Clustering using Multidimensional Angles as Proximity Measures Vienna University of Technology," *ACM Transactions on Internet Technology*, vol. 9, no. 3, Article 11, Publication date: July 2009.
[29] J. Suykens: Least Squares Support Vector Machines. Tutorial IJCNN, 2003.
[30] A. J. Smola and B. Sch¨okopf, "On a kernelbased method for pattern recognition, regression, approximation and operator inversion," *Algorithmica, 1998. emTechnical Report 1064*, GMD FIRST, April 1997.
[31] S. R. Gunn., "Support Vector Machines for Classification and Regression," *Technical Report, Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science*, 10 May, 1998.
[32] Chen B, Sun F, Hu J., "Local linear multi-SVM method for gene function classification[C]," *Nature and Biologically Inspired Computing. IEEE*, 2011: 183-188.
[33] Liu Y, Parhi K K., "Computing RBF Kernel for SVM Classification Using Stochastic Logic[C]," *IEEE International Workshop on Signal Processing Systems. IEEE*, pp. 327-332, 2016.
[34] Zhou B, Yang C, Guo H, et al. "A quasi-linear SVM combined with assembled SMOTE for imbalanced data classification[C]," *International Joint Conference on Neural Networks. IEEE*, pp. 1-7, 2013.
[35] [Online] Available weka platform: https://www.cs.waikato.ac.nz/ml/weka/.
[36] [Online] Available UCI dataset: http://archive.ics.uci.edu/ml/datasets.html.

Zheng Wang received his B.Sc. degree in Communication Engineering from Suwon University, Suwon, Korea, in 2012, and M. Sc. degree from Suwon University, Suwon, Korea, in 2016. He currently works as a PhD student in the Department of Electrical Engineering, Suwon University, Suwon, South Korea. His research interests include fuzzy set and fuzzy system, evolution computing and neural networks.

Cheng Yang received his B.Sc. degree in computer Science and technology from Liaocheng University in 2009, Shandong, and M. Sc. degree from Waseda University in 2013, Japan, and M. Sc. from Shanghai University in 2013,Shanghai. He currently works as a PhD student in the Department of Electrical Engineering, University of Suwon, Suwon, South Korea. His research interests include fuzzy system, Support Vector Machine and Neural Networks.

Sung-Kwun Oh received his B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Yonsei University, Seoul, Korea, in 1981, 1983, and 1993, respectively. During 1983-1989, he was a Senior Researcher of R&D Lab. of Lucky-Goldstar Industrial Systems Co., Ltd. From 1996 to 1997, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada. He is currently a Professor with the Department of Electrical Engineering, University of Suwon, Suwon, South Korea. His research interests include fuzzy system, fuzzy-neural networks, automation systems, advanced computational intelligence, and intelligent control. He currently serves as an Associate Editor of the KIEE Transactions on Systems and Control, International Journal of Fuzzy Logic and Intelligent Systems of the KFIS, and Information Sciences.

Zunwei Fu received his M. Sc., and Ph. D. degrees in mathematics from the University of Mining & Technology (Beijing) and Beijing Normal University, Beijing, China, in 2005 and 2008, respectively. He is currently a professor with the Department of Computer Science, University of Suwon, Suwon, South Korea and the Department of Mathematics, Linyi University, Shandong, China. His research interests include fuzzy set and system, Fourier transform and its applications. He served as an editor of Journal of Linyi University.