Comput Syst Sci & Eng (2018) 6: 409–420
© 2018 CRL Publishing Ltd

International Journal of
Computer Systems
Science & Engineering

Improving support vector domain description by maximizing the distance between negative examples and the minimal sphere center's

Mohamed EL Boujnouni[1] **and Mohamed Jedra**[2]

[1] *Chouaib Doukkali University, Laboratory of Information Technologies, National School of Applied Sciences, El Jadida Morocco*
[2] *Mohammed V University, Faculty of Sciences Rabat, Laboratory of Conception and Systems (Microelectronic signals and Informatics) Avenue Ibn Battouta B.P 1014, Rabat - Morocco*
E-mail: med.elbouj@gmail.com,jedra@fsr.ac.ma

Support Vector Domain Description (SVDD) is an effective kernel-based method used for data description. It was motivated by the success of Support Vector Machine (SVM) and thus has inherited many of its attractive properties. It has been extensively used for novelty detection and has been applied successfully to a variety of classification problems. This classifier aims to find a sphere with minimal volume including the majority of examples that belong to the class of interest (positive) and excluding the most of examples that are either outliers or belong to other classes (negatives). In this paper we propose a new approach to improve the classification accuracy of SVDD. This objective will be achieved by exploiting the existence of negative examples in the training step, without increasing the computational time and memory resources required to solve the quadratic programming problem of that classifier. Simulation results on two challenging artificial problems, namely chessboard and two spirals, and four benchmark datasets have successfully validated the effectiveness of the proposed method.

Keywords: Support Vector Domain Description, Negative examples, Classification accuracy, Space and time complexities.

1. INTRODUCTION

Data domain description aims to characterize a given set of objects by modeling the boundary enclosing them [1], [2]. A good description covers all target data but includes no superfluous space or negative examples. Support Vector Domain Description is one of data domain description methods. It has been developed by Tax et al. [2], [3], [4] to solve one-class classification problems basing on Vapnik's Support Vector Machine learning theory [5], [6]. In contrast to SVM, which discriminates between two classes by fitting a hyperplane with maximum margin between the examples of the opposite classes, SVDD tries to find a minimal enclosing sphere around the target class. Thus, it provides the best representation of the class of interest and offers inferences that can be used to detect outliers. This classifier has many interesting and desirable properties. It's based on an elegant and rigorous mathematical foundation from optimization and statistical learning theory. It's derived from the structural risk minimization principle as opposed to empirical risk. It can model arbitrarily distributions without the need to make assumptions concerning data distribution. Training SVDD involves solving a constrained convex quadratic programming problem (QP) which can get unique global minimum [7], [8]. As a nonlinear kernel based method, this classifier can map hardly separable data of opposite classes into a high-dimensional feature space in which they become separable through a hyper-spherical surface. The transformation is performed implicitly by using kernel functions [9], [10].The nature of this classifier makes it suitable to be used in outliers detection problems to detect samples that are different from a learned dataset or it can be used as one-class classifier when the training data of one class is well sampled while the data of the other classes are not. This classifier uses few examples of the training dataset named Support Vectors to describe the boundaries between different classes. This process accelerates heavily the computational time required to classify new instances. Both Support Vector machine and Support Vector Domain Description have been applied successfully to a variety of research fields such as: Event and novelty detection [11], [37], Fault Diagnosis [12], [13], [14], credit ratings [15], [16], [17], image classification [18], [19], Computer and network security [20], [21], [22], etc.

Support Vector Domain Description can be used in two different ways: Firstly, when negative examples are not available, this case is called Normal Data Description or one-class SVDD. Secondly, when negative data are available, this case can be considered as an extension of the first one and is named two-classes SVDD. In the last case this classifier searches the minimal sphere that includes the majority of positive examples and excludes the most of negative ones. This approach improves the classification accuracy of Normal Data Description but increases drastically the space and time complexities required to solve the latter's QP, because negatives samples participate also in the training task. Thus, the major problem of training SVDD with negative examples is the requirement of large memory and enormous training time especially for large-scale applications.

In this paper we propose a new approach to extend Support Vector Domain Description from one-class to two-classes by using negative examples. We aim to reach this goal without increasing the space and time complexities of Normal Data Description. The rest of this paper is organized as follows: Section 2 presents both conventional versions of this classifier (One-class and Two-classes), Section 3 gives a detailed description of our approach and the last section contains several experimental results to demonstrate the validity of the proposed method.

2. SUPPORT VECTOR DOMAIN DESCRIPTION

2.1 One-class Support Vector Domain Description

Suppose we are given a dataset $S = \{x_1, x_2, \ldots, x_N\}$ where N is the number of samples, one-class SVDD attempts to find the smallest ball with a center a and a radius R, that contains most of the patterns in S [4], [23], [24], [25]. This is an optimality problem that can be formulated mathematically as follows:

Minimize:

$$R^2 \quad \text{Subject to } ||x_i - a||^2 \leq R^2 \; \forall i = 1, \ldots, N \tag{1}$$

Where $||\cdot||$ is the Euclidean norm. To allow the presence of outliers a positive constant C was introduced, the latter determines the tradeoff between the volume of the sphere to minimize and the rejection of target objects. The optimization problem (Eq. (1)) then becomes:

Minimize:

$$R^2 + C\sum_{i=1}^{N}\varepsilon_i \quad \text{Subject to } ||x_i - a||^2 \leq R^2 + \varepsilon_i \forall i = 1, \ldots, N \tag{2}$$

Where ε_i with $i = 1, \ldots, N$ are slack variables. The Lagrangian reformulation of this problem can be written as:

$$L(R, a, \alpha_i, \varepsilon_i) = R^2 - \sum_{i=1}^{N}(R^2 + \varepsilon_i - ||x_i - a||^2)\alpha_i - \sum_{i=1}^{N}\varepsilon_i\mu_i + C\sum_{i=1}^{N}\varepsilon_i \tag{3}$$

Where $\alpha_i \geq 0$ and $\mu_i \geq 0$ are Lagrange multipliers. Annulling the partial derivatives of L gives the following constraints:

$$\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i = 1 \tag{4}$$

$$\frac{\partial L}{\partial a} = 0 \Rightarrow a = \sum_{i=1}^{N}\alpha_i x_i \tag{5}$$

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \Rightarrow \alpha_i = C - \mu_i \tag{6}$$

Substituting the equations Eq. (4), Eq. (5) and Eq. (6) into Eq. (3) gives the following dual problem:

Maximize:

$$W = \sum_{i=1}^{N} x_i^2\alpha_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j x_i x_j$$

$$\text{Subject to } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{N}\alpha_i = 1 \; \forall i = 1, \ldots, N \tag{7}$$

The main goal is to determinate each α_i with $i = 1, \ldots, N$ by maximizing the equation above with respect to the constraints.

2.2 Two-class Support Vector Domain Description

When negative examples are available, they can be incorporated in the training to improve the description. In opposition to the positive examples which should be within the minimal sphere, the negative ones should be outside it. In the following, the target objects are enumerated by the indices, i, j and the negative examples by l, m. To allow some classification errors for both classes (Target and Non-target) slack real positive variables ε_i and ε_l are introduced [4]. The new optimization problem can be formulated as follows:

Minimize

$$R^2 + C1\sum_{i=1}^{N}\varepsilon_i + C2\sum_{l=1}^{M}\varepsilon_l$$

Subject to $||x_i - a||^2 \le R^2 + \varepsilon_i$ *and* $||x_l - a||^2 \ge R^2 - \varepsilon_l$ ε_i, $\varepsilon_l \ge 0 \; \forall\, i = 1, \ldots, N$ *and* $l = 1, \ldots, M$ (8)

Where $C1$ and $C2$ are constants real positives, $C1\sum_i \varepsilon_i$ and, $C2\sum_l \varepsilon_l$ are penalty terms. By using Lagrange multipliers, α_i, α_l, γ_i, γ_l the equation (8) can be re-written as follows:

$$L(R, a, \alpha_i, \varepsilon_i, \alpha_l, \varepsilon_l) = R^2 - \sum_{i=1}^{N}\alpha_i\left(R^2 + \varepsilon_i - ||x_i - a||^2\right) - \sum_{i=1}^{N}\gamma_i\varepsilon_i + C1\sum_{i=1}^{N}\varepsilon_i - \sum_{l=1}^{M}\alpha_l\left(||x_l - a||^2 - R^2 + \varepsilon_l\right) - \sum_{l=1}^{M}\gamma_l\varepsilon_l + C2\sum_{l=1}^{M}\varepsilon_l \quad (9)$$

With $\alpha_i \ge 0$, $\alpha_l \ge 0$, $\gamma_i \ge 0$, $\gamma_l \ge 0$ are Lagrange multipliers. Setting the partial derivatives of L to zero gives the following constraints:

$$\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i - \sum_{l=1}^{M}\alpha_l = 1 \quad (10)$$

$$\frac{\partial L}{\partial a} = 0 \Rightarrow a = \sum_{i=1}^{N}\alpha_i x_i - \sum_{l=1}^{M}\alpha_l x_l \quad (11)$$

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \text{ and } \frac{\partial L}{\partial \varepsilon_l} = 0 \Rightarrow \alpha_i = C1 - \gamma_i \quad \text{and } \alpha_l = C2 - \gamma_l \quad (12)$$

After substituting the equations Eq. (10), Eq. (11) and Eq. (12) into the Eq. (9). The dual problem can be written as:

Maximize

$$W = \sum_{i=1}^{N}\alpha_i x_i x_i - \sum_{l=1}^{M}\alpha_l x_l x_l - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j x_i x_j + 2\sum_{l=1}^{M}\sum_{j=1}^{N}\alpha_l\alpha_j x_l x_j - \sum_{l=1}^{M}\sum_{m=1}^{M}\alpha_l\alpha_m x_l x_m$$

Subject to $0 \le \alpha_i \le C1$ and

$0 \le \alpha_l \le C2 \quad \forall i = 1, .., N, \; \forall l = 1, .., M$

and $\sum_{i=1}^{N}\alpha_i - \sum_{l=1}^{M}\alpha_l = 1$ (13)

The main goal is to determinate the set of α_i with $i = 1, \ldots, N$ that corresponds to the positive examples and the α_l with $l = 1, \ldots, M$ related to the negative examples by maximizing the new dual problem.

In real-world applications, datasets are rarely distributed spherically [26]. In order to have a flexible data description, a kernel trick $k(x_i, x_j) = \phi(x_i)\phi(x_j)$ was introduced [4], [23], [27], [28]. The kernel function maps a dataset of hardly inseparables instances into a higher dimensional feature space where they become easily separable. In this new feature space the dot products are replaced by a suitable kernel function. In literature, many kernels have been proposed. Table 1 lists some commonly used ones.

Table 1 Some commonly used kernel functions.

Linear kernel	$k(x_i, x_j) = x_i \cdot x_j + c$
Gaussian kernel	$k(x_i, x_j) = e^{(-\|\|x_i - x_j\|\|^2/2\sigma^2)}$
Exponential kernel	$k(x_i, x_j) = e^{(-\|\|x_i - x_j\|\|/2\sigma^2)}$
Sigmoid kernel	$k(x_i, x_j) = \tanh(\alpha(x_i, x_j) + c)$
Polynomial kernel	$k(x_i, x_j) = (\alpha \cdot x_i \cdot x_j + c)^d$
Laplacian kernel	$k(x_i, x_j) = e^{(-\|\|x_i - x_j\|\|/\sigma)}$

As a conclusion, training a given dataset with Support Vector Domain Description implies solving a quadratic programming problem with constraints expressed by the equations (7) or (13) that correspond respectively to the versions : one-class and two classes. As said before, this process requires large amounts of computational time and memory, specifically when dealing with large datasets. Denote the number of training samples by N, the space complexity for solving the one-class QP is $O(N^2)$ and its time complexity is $O(N^3)$ [29], [30]. By comparing the equations (7) and (13) we remark that: When negative examples are incorporated in SVDD, the space and time complexities required to solve the QP increase from $O(N^2)$ and $O(N^3)$ to $O((N+M)^2)$ and $O((N+M)^3)$ respectively.

2.3 Multi-class Support Vector Domain Description

To solve a K-classes classification problem, multiple minimal hyperspheres (a_k, R_k) with $k = 1, \ldots, K$ will be constructed. In each training step the N samples of the k^{th} class are considered as target while the M remaining samples that belong to the other classes are considered as negatives. This strategy is called one-against-all decomposition. To predict the membership of an unknown sample x_z a similarity measure function $sim(x_z, a_k, R_k)$ has to be evaluated. This function can be expressed as follows:

$$\text{class of } x_z \equiv \arg\max_{k=1,\ldots,K} sim(x_z, a_k, R_k) \quad (14)$$

In literature, several similarity functions were proposed, the simplest one is given by the following equation:

$$sim(x_z, a_k) = -||x_z - a_k||^2 \quad (15)$$

Using the function above with the equation (14) implies affecting to x_z the class where the center a_k is the closest. Zhu et al. [33] proposed a similarity function that considers not only the distance between x_z and the center of the k^{th} class a_k but also the radius R_k:

$$sim(x_z, a_k, R_k) = R_k^2 - ||x_z - a_k||^2 \quad (16)$$

Other similarity measure supposes that the training examples from the class k are generated from a Gaussian distribution in a high-dimensional feature space with the mean a_k and the variance R_k^2. Then, according to the Bayesian decision rule, the similarity function can be expressed as follows:

$$sim\,(z, a_k, R_k) = \frac{1}{R_k}\exp\left(\frac{-||z-a_k||^2}{R_k^2}\right) \tag{17}$$

In the same context, Wu et al. [34] focused on the position of x_z with respect the minimal hyperspheres. Then they proposed similarity functions which involved three cases:

First case: When the test point x_z is outside all the minimal hyperspheres, they choosed the nearest one to x_z:

$$sim(x_z, S_k) = R_k - ||x_z - a_k|| \tag{18}$$

Second case: When the test point x_z is inside only one hypersphere, they choose the corresponding class.

Third case: When the test point x_z is located in the space between a set of hyperspheres. Then they obtained the class of x_z by comparing the distance between this point and the center of each of those hyperspheres. To eliminate the effect of different spherical radii, a relative distance is applied:

$$sim(x_z, a_k, R_k) = -\frac{||x_z - a_k||}{R_k} \tag{19}$$

In another work, Hao et al. [35] proposed a fuzzy membership function to determinate the class that an unknown test example x_z belongs to. Their proposed similarity function is given by the equation below:

$$sim(x_z, a_k, R_k) = \begin{cases} 0.5 \times \left[\frac{1-\frac{||x_z-a_k||}{R_k}}{1+\lambda_1\left(\frac{1}{R_k}\right)||x_z-a_k||}\right] + 0.5 \\ \quad \text{if } ||x_z - a_k|| \le R_k \\ 0.5 \times \left[\frac{1}{1+\lambda_2(||x_z-a_k||-R_k)}\right] \\ \quad \text{otherwise} \end{cases} \tag{20}$$

Where λ_1 and λ_2 are user predefined parameters that satisfy:

$$\lambda_2 = \frac{1}{R_k(1+\lambda_1)} \tag{21}$$

The similarity functions described above are evaluated based on the equations Eq. (22), Eq. (23), Eq. (24) and Eq. (25) given by the following expressions:

In the case of one-class Support Vector Domain Description:

$$||x_z - a_k||^2 = x_z \cdot x_z - 2\sum_{i=1}^{N}\alpha_{ki}x_i x_z + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ki}\alpha_{kj}x_i x_j \tag{22}$$

$$R_k^2 = x_s \cdot x_s - 2\sum_{i=1}^{N}\alpha_{ki}x_i x_s + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ki}\alpha_{kj}x_i x_j \tag{23}$$

With α_{ki} is the i^{th} Lagrangian multiplier corresponding to the k^{th} class and $x_s \in SV$. SV is the set of Support Vectors having $0 < \alpha_{ks} < C$.

In the case of two-classes Support Vector Domain Description:

$$\begin{aligned}||x_z - a_k||^2 = {} & x_z.x_z - 2\left(\sum_{i=1}^{N}\alpha_{ki}x_i x_z - \sum_{l=1}^{M}\alpha_{kl}x_l x_z\right) \\ & + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ki}\alpha_{kj}x_i x_j + \sum_{l=1}^{M}\sum_{m=1}^{M}\alpha_{kl}\alpha_{km}x_l x_m \\ & - 2\sum_{i=1}^{N}\sum_{l=1}^{M}\alpha_{ki}\alpha_{kl}x_i x_l \end{aligned} \tag{24}$$

$$\begin{aligned}R_k^2 = {} & x_s.x_s - 2\left(\sum_{i=1}^{N}\alpha_{ki}x_i x_s - \sum_{l=1}^{M}\alpha_{kl}x_l x_s\right) \\ & + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ki}\alpha_{kj}x_i x_j + \sum_{l=1}^{M}\sum_{m=1}^{M}\alpha_{kl}\alpha_{km}x_l x_m \\ & - 2\sum_{i=1}^{N}\sum_{l=1}^{M}\alpha_{ki}\alpha_{kl}x_i x_l \end{aligned} \tag{25}$$

For any $x_s \in SV$. SV is the set of support vectors having $0 < \alpha_{ks} < C1$ (with x_s is a target object) or $0 < \alpha_{ks} < C2$ (with x_s is negative object).

3. OUR APPROACH

As mentioned before, our objective is to improve Normal Data Description by using negative examples without increasing the complexity to solve the Quadratic Programming problem of this classifier. In this section, mathematical model of this latter will be changed as purposeful. Rather than excluding the negative points outside the minimal hypersphere that surrounds the target data, we propose to maximize the separability between the negative examples and the center of the minimal hypersphere. This idea will be incorporated into the mathematical formulation of SVDD. Our new criterion to maximize is given by the equation below:

$$\frac{1}{M}\sum_{l=1}^{M}||a - x_l||^2 \tag{26}$$

Where M is the number of negative data. The Lagrangian function of this problem can be expressed as follows:

$$\begin{aligned}L(R, a, \alpha_i, \varepsilon_i) = {} & R^2 - \frac{\rho}{M}\sum_{l=1}^{M}||a - x_l||^2 \\ & - \sum_{i=1}^{N}(R^2 + \varepsilon_i - ||x_i - a||^2)\alpha_i - \sum_{i=1}^{N}\varepsilon_i\mu_i \\ & + C\sum_{i=1}^{N}\varepsilon_i \end{aligned} \tag{27}$$

Where C is a constant positive and real, $C\sum_i \varepsilon_i$ is penalty term, and α_i are Lagrange multipliers. It can be seen that minimizing the cost function Eq. (27) will make the square of the radius R^2 as small as possible and the term $\frac{\rho}{M}\sum_l ||a - x_l||^2$ as large as possible. The parameter ρ is a real strictly positive number, another condition on ρ is given later by the equations (29) and (31), which implies that the value of ρ must be strictly lower than one. This parameter plays a compromise between the minimization of the radius of the hypersphere, and the maximization of the separability between the center of the hypersphere and negative examples. By annuling the partial derivatives of L with respect to R, a, ε_i we obtain following constraints:

$$\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i = 1 \tag{28}$$

$$\frac{\partial L}{\partial a} = 0 \Rightarrow a = \frac{1}{(1-\rho)}\left(\sum_{i=1}^{N} \alpha_i x_i - \frac{\rho}{M}\sum_{l=1}^{M} x_l\right) \tag{29}$$

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \Rightarrow \alpha_i = C - \mu_i \tag{30}$$

The dual problem is then:

Maximize:

$$W = \frac{-1}{(1-\rho)}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j x_i x_j + \sum_{i=1}^{N}\alpha_i \underbrace{\left(\frac{2\rho}{(1-\rho)M}\sum_{l} x_i x_l + x_i^2\right)}_{Term\ I} \underbrace{- \frac{\rho^2}{(1-\rho)M^2}\sum_{l=1}^{M}\sum_{m=1}^{M} x_l x_m - \frac{\rho}{M}\sum_{l=1}^{M} x_l^2}_{Term\ II}$$

$$\text{Subject to} \quad 0 \le \alpha_i \le C \quad and \sum_{i=1}^{N}\alpha_i = 1 \ \forall i = 1, \ldots, N \tag{31}$$

The optimization problem described by the Eq. (31) is equivalent to a convex quadratic problem with global minimum, when $\frac{-1}{(1-\rho)} < 0$, we conclude that ρ must belong to the interval $]0, 1[$.

The objective is to find the values of Lagrange multipliers that maximize the Eq. (31); in this case the second term of the equation is just a constant, it doesn't contains any Lagrange multipliers, as consequence it's not concerned by the maximization. By contrast, the first term contains N Lagrange multipliers $(\alpha_1, \ldots, \alpha_N)$ labeled by the indexes i and j (which refer only to the target examples). By comparing Eq. (7) and the first term of Eq. (31) we remark that: Solving both QPs requires the same space and time complexities: $O(N^2)$ and $O(N^3)$ respectively.

To classify a test point x_z, we apply the same strategy quoted in the previous section basing on the following equations:

$$||x_z - a_k||^2 = x_z.x_z - \frac{2}{(1-\rho)}\left(\sum_{i=1}^{N}\alpha_{ki} x_i x_z - \frac{\rho}{M}\sum_{l=1}^{M} x_l x_z\right) + \underbrace{\frac{1}{(1-\rho)^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ki}\alpha_{kj} x_i x_j - \frac{2\rho}{M}\sum_{i=1}^{N}\sum_{l=1}^{M}\alpha_{ki} x_i x_l + \frac{\rho^2}{M^2}\sum_{l=1}^{M}\sum_{m=1}^{M} x_l x_m\right)}_{constant}$$

$$R_k^2 = x_s.x_s - \frac{2}{(1-\rho)}\left(\sum_{i=1}^{N}\alpha_{ki} x_i x_s - \frac{\rho}{M}\sum_{l=1}^{M} x_l x_s\right) + \underbrace{\frac{1}{(1-\rho)^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ki}\alpha_{kj} x_i x_j + \frac{2\rho}{M}\sum_{i=1}^{N}\sum_{l=1}^{M}\alpha_{ki} x_i x_l + \frac{\rho^2}{M^2}\sum_{l=1}^{M}\sum_{m=1}^{M} x_l x_m\right)}_{constant} \tag{32}$$

With α_{ki} is the i^{th} Lagrangian multiplier corresponding to the k^{th} class and $x_s \in SV$ a Support Vector having $0 < \alpha_s < C$.

The equation (32) is used to classify a new unknown sample x_z, it contains three terms: the two first ones depend on x_z and must be evaluated for each new sample x_z, by contrary the last one is a constant that must be evaluated just one time and still available for each new x_z. This is an advantage because it minimizes the time to classify a new instance. The equation (33) is used to evaluate the radius of the hypersphere corresponding to the k^{th} class, based on an example x_s belonging to the set of Support Vectors. The example x_s appears just in the first two terms of the equation, while the last term doesn't depend on x_s and can be evaluated just one time and stills available whatever the value of x_s.

4. EXPERIMENTS AND RESULTS

4.1 Datasets and experimental setting

The performance of our new method is evaluated using three challenging classification problems: Two- spirals, Checkerboard, and four benchmark datasets namely: Fertility, Libras Movement, Blood Transfusion Service Center, and Tic-Tac-Toe Endgame available from the UCI repository of machine learning data [36].

The goal behind the first experiment is to show graphically the classification effectiveness of our method compared to normal data description and to highlight the effect of the parameter ρ. To achieve this objective we chose to describe a highly nonlinear separating surface. This test will be performed on two-spirals dataset represented in 2 dimensions. Spiral shape exists in several natural and physical domains such as: The motion of particles in cyclotrons, the classic double helix in DNA, the spiral feed in manufacturing. Spiral forms are particularly interesting because of their high levels of nonlinearity and resistance to shape transformation under translation, rotation or other scalar operations. Spirals structures are also attractive for their temporal properties and are found to be particularly hard to classify for pattern recognition purposes.

The second experiment has the same objectives as the first but with another dataset. This latter was chosen to be hard to describe as the first. Namely checkerboard [31], [32], it consists 81 black and white points in 2 dimensions taken from nine black and white squares.

Table 2 Description of the datasets used in the experiments.

Dataset name	Number of instances	Subsets		Number of classes	Number of attributes
		Training subset	Testing subset		
Fertility	100	80% of samples/ class	The remaining samples/ class	2	9
Libras Movement	360	80% of samples/ class	The remaining samples/ class	15	90
Blood Transfusion Service Center	748	80% of samples/ class	The remaining samples/ class	2	4
Tic-Tac-Toe Endgame	958	80% of samples/ class	The remaining samples/ class	2	9

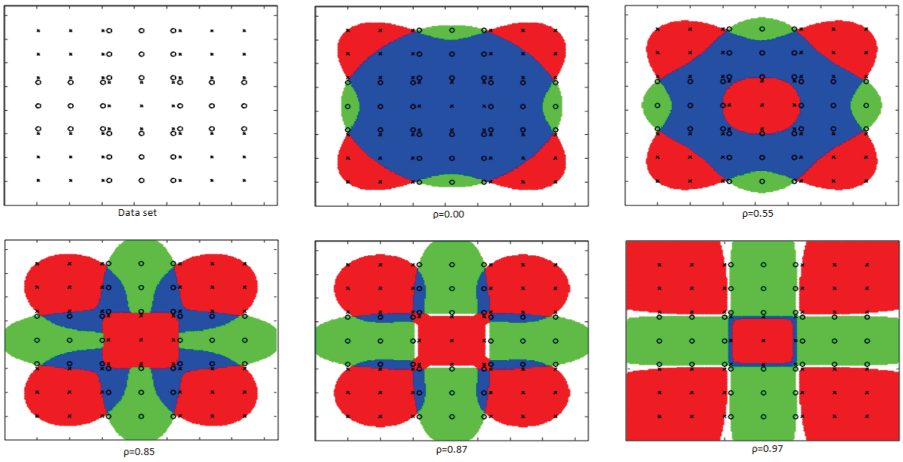


Figure 1 Classification of chessboard dataset using SVDD with negative examples and = 100, $\sigma = 1.25$.

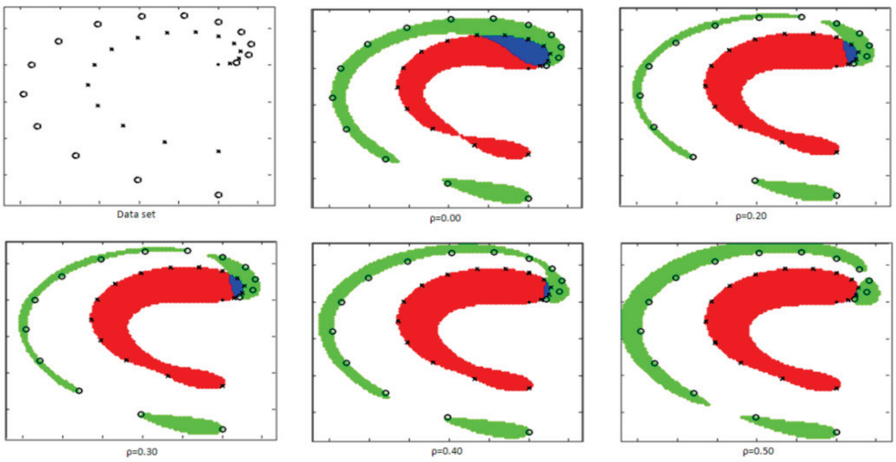


Figure 2 Classification of Two spirals dataset using SVDD with negative examples and = 100, $\sigma = 1.5$.

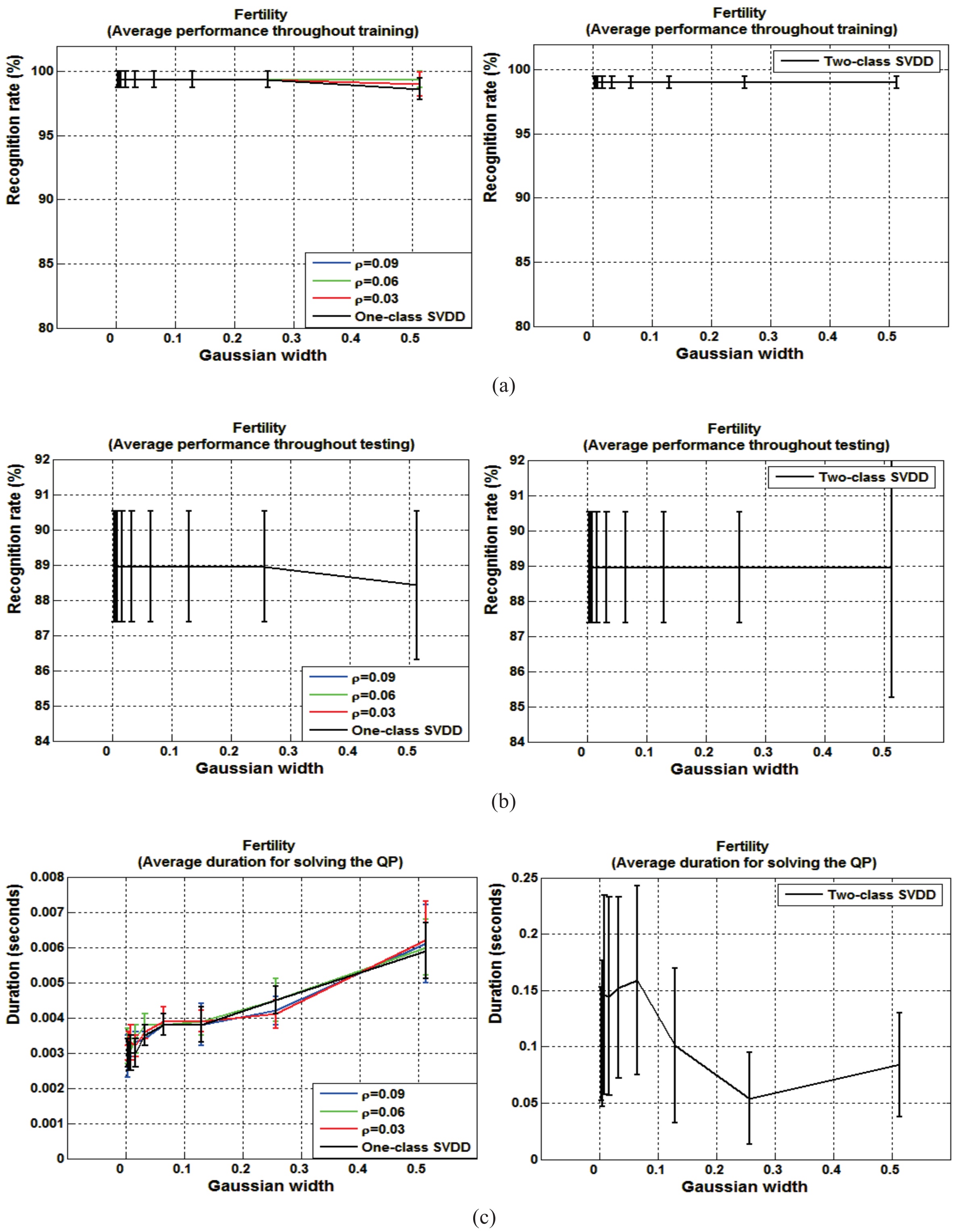


Figure 3 The average recognition rates and runtimes to minimize the QPs pb of One-class, Two-classes and the proposed SVDD using Fertility dataset.

The third experiment aims to compare three versions of Support Vector Domain Description which are : One-class, two classes and the proposed one. This comparison regards: Learning performance, generalization ability, and the computational time required to solve the QPs problems represented by the equations Eq. (7), Eq. (13) and Eq. (31). Four datasets described in Table 2 were used to perform those experiments: In each test we split randomly each dataset into two disjoints subsets that serve for training and testing with respect to the rates indicated in this table. To evaluate the average and the standard deviation the training and testing processes are repeated 20 times.

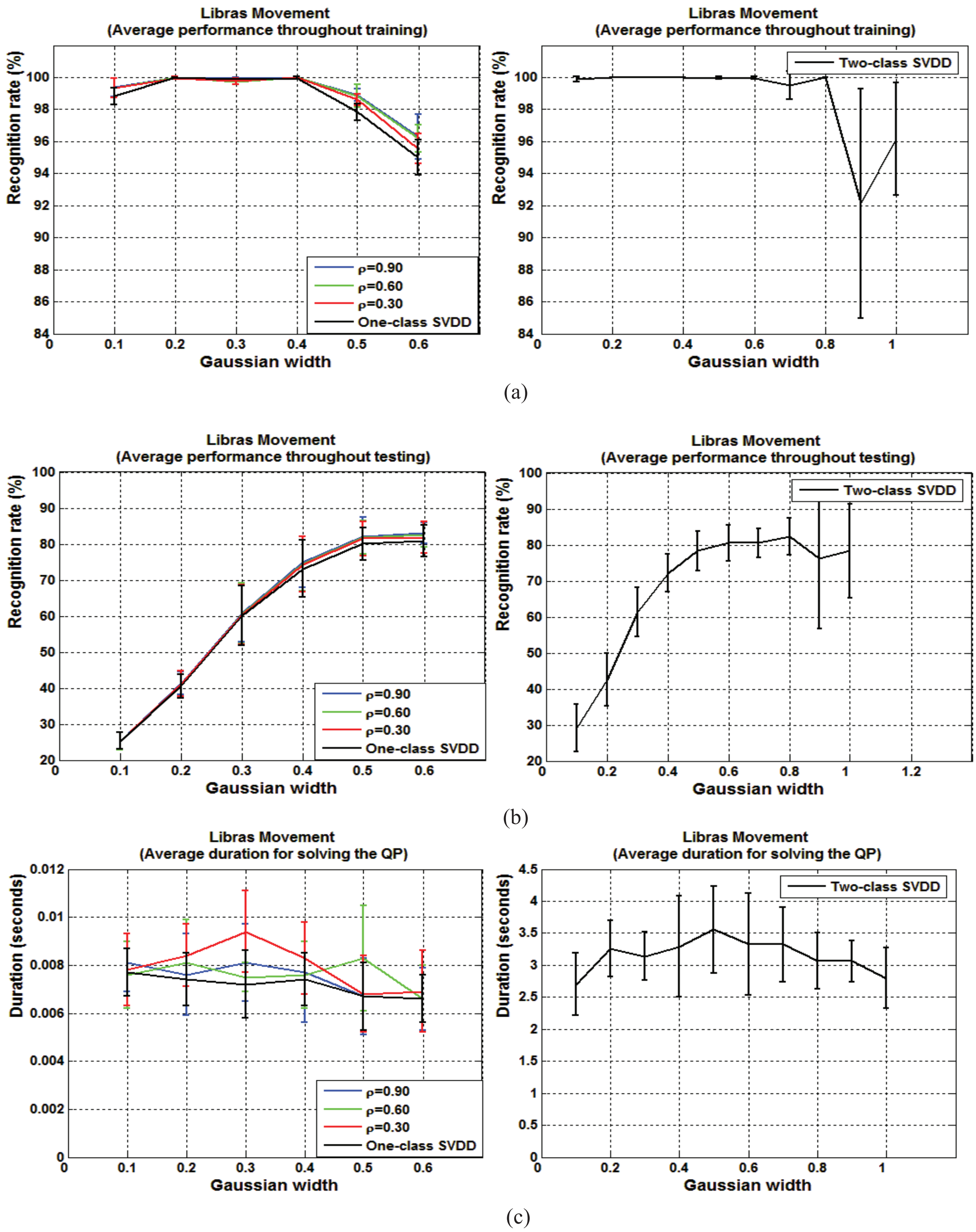


Figure 4 The average recognition rates and runtimes to minimize the QPs pb of One-class, Two-classes and the proposed SVDD using Libra Movement dataset.

4.2 Experimental Results

The two first experiments (Figure 1 and 2) aim to determinate the effect of the new parameter ρ on the classification accuracy of one-class SVDD ($\rho = 0$) compared to our new approach ($0 < \rho < 1$). To show the effectiveness of our approach, we plot the classification result in 2D with three colors. The red, green, and blue represent respectively the first class, the second one, and the overlaps areas (when the samples belong to both classes). The regularization parameter C is equal to 100, the Gaussian width σ is fixed at 1.25 (Chessboard) and 1.5 (Two-spirals) and the new parameter ρ is increased gradually in the sets $\{0, 0.55, 0.85, 0.87, 0.97\}$ and $\{0, 0.2, 0.3, 0.4, 0.5\}$ corresponding respectively to Chessboard and two spirals.

Figures 1 and 2 show that: When using the conventional One-class version the smallest hypersphere encloses not only the normal examples but also a large number of negative data (blue color). This could be due to the fact that this classifier does not

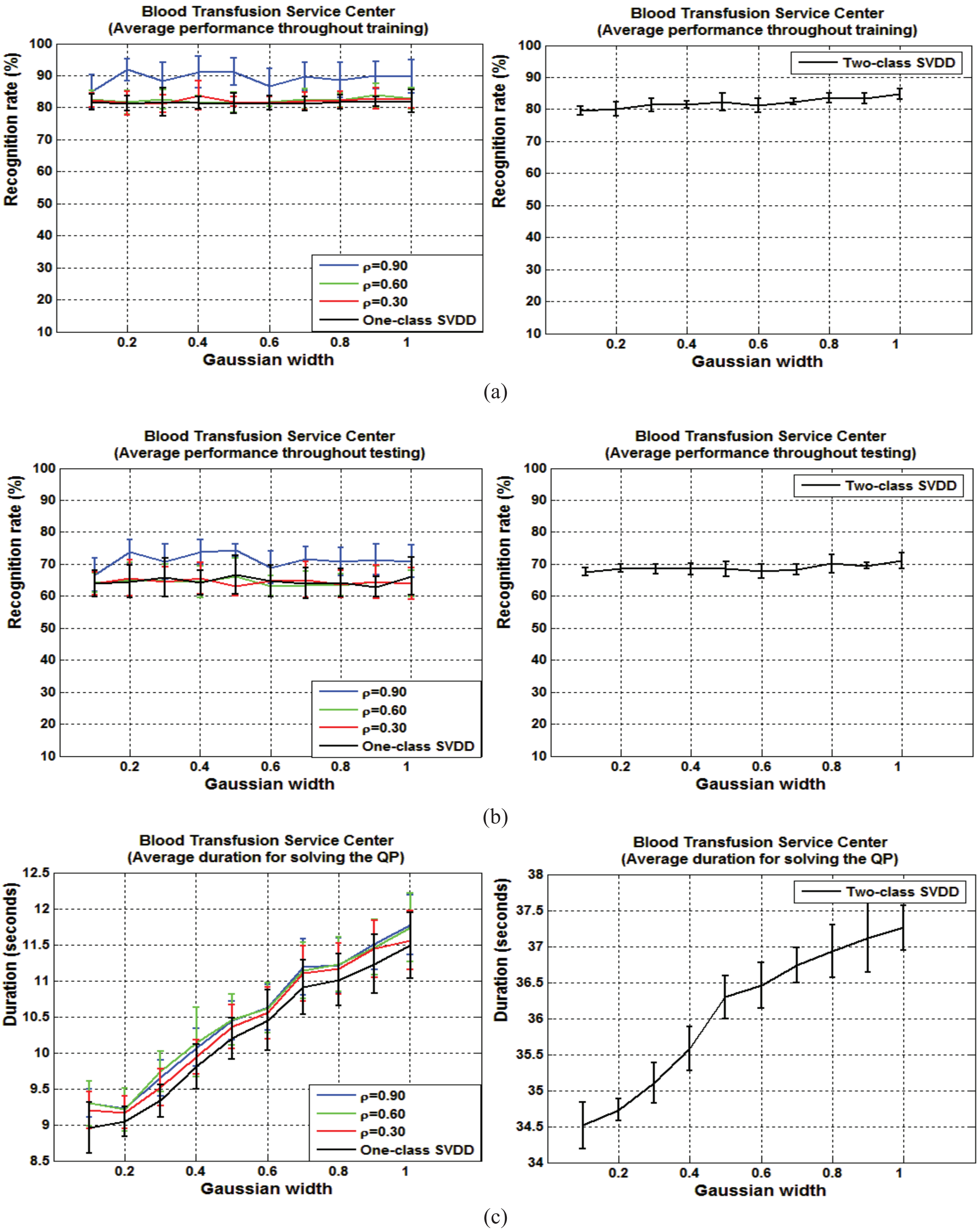
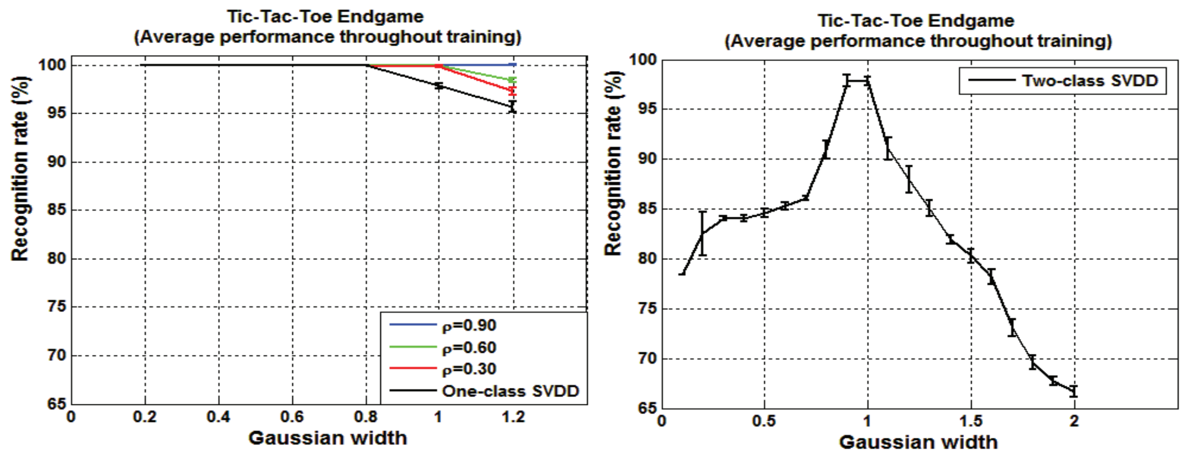


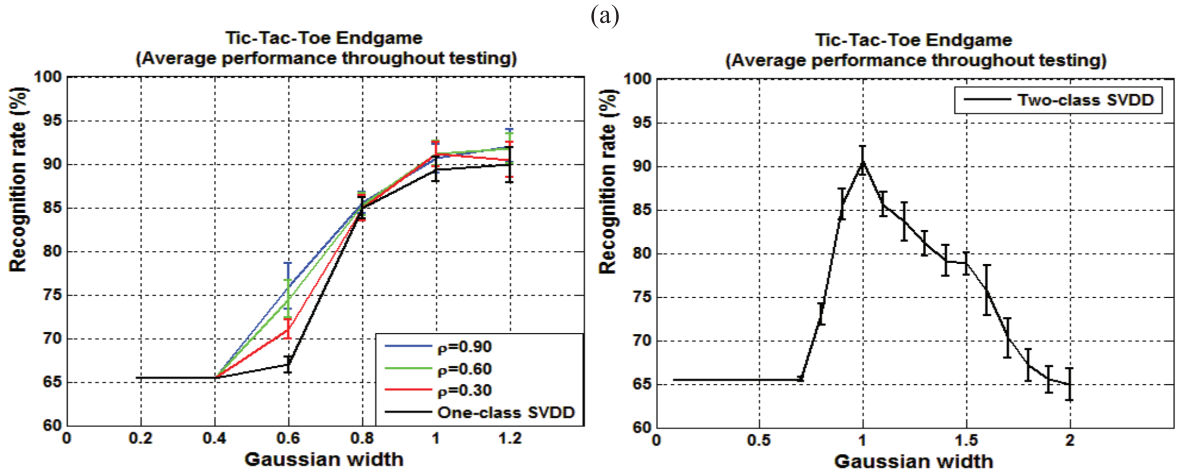
Figure 5 The average recognition rates and runtimes to minimize the QPs pb of One-class, Two-classes and the proposed SVDD using Blood Transfusion Service Center dataset .

take into consideration the negative data during training. When we switch to our approach by increasing the value of ρ, we remark that the description become more tightening, i.e blue color is disappeared progressively and is replaced by the right one. This shows that the proposed method, under the same conditions of the conventional one-class SVDD (i.e the same values of the parameters C and ρ, and the same calculation complexity) gives higher precision rates.

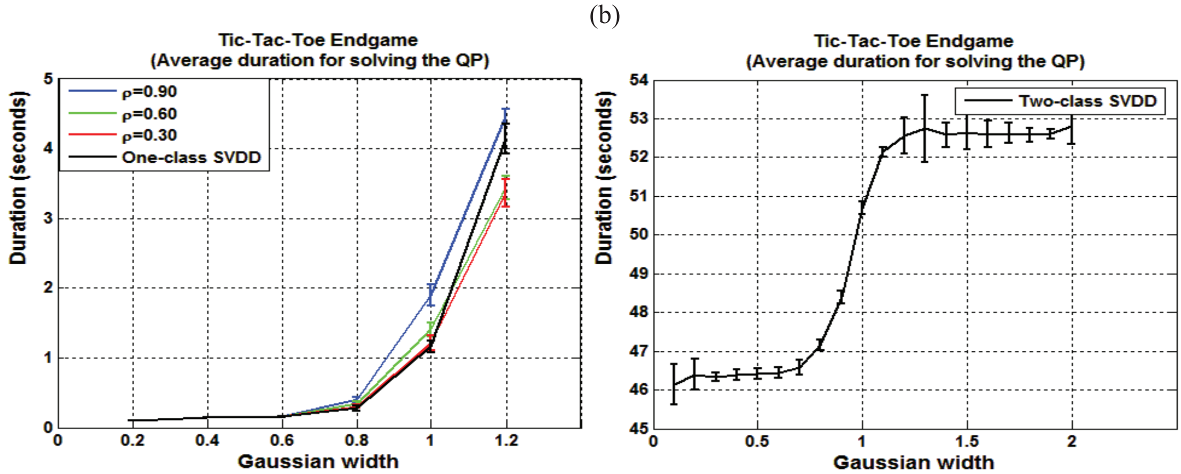
The third experiment is performed using a Gaussian kernel with the similarity function described in [34]. The parameter ρ is varied in the set $\{0.3, 0.6, 0.9\}$ the regularization parameters C, $C1$ and $C2$ are fixed at 100. The Figures 3, 4, 5 and 6 show the average recognition rate and the standard deviation calculated in the training and testing processes also they provide the average



(a)



(b)



(c)

Figure 6 The average recognition rates and runtimes to minimize the QPs pb of One-class, Two-classes and the proposed SVDD using Tic-Tac-Toe Endgame dataset.

computational time required to solve the QPs of each classifier in seconds. All experiments are conducted on a PC with 64 bit operating system, having 8 GB of RAM and a Core 5 processor. Table 3 analyses the figures and gives the maximum and the minimum values found in the experiments, the best values are highlighted. From the Figures 3, 4, 5, 6 and the Table 3 we observe that:

- In the training step: When using our approach, the recognition rate presents high accuracy compared to One-class and Two-classes SVDD. This means that each minimal hypersphere found by our method encloses successfully the target samples of the corresponding class. The performance of our method depends heavily on the choice of the value of the regularization parameter $\rho \in]0, 1[$ and it reaches its maximum when ρ equals to 0.90. These promising results must be reevaluated to check against over-fitting.

- In the testing step: When using our method, the novelty detection (the generalization ability) rate gives the best re-

Table 3 Analysis of the experimental results obtained from applying One-class, Two-classes and the proposed SVDD on the four datasets.

Dataset name			Fertility	Libras Movement	Blood Transfusion Service Center	Tic-Tac-Toe Endgame
One class SVDD	Training	[Min, Max]	[98.64, 99.38]	[95.00 100.00]	[81.27, 82.00]	[95.67, 100]
	Testing	[Min, Max]	[88.42, 88.95]	[25.33, 80.83]	[62.95, 66.61]	[65.44, 89.9]
	Duration	[Min, Max]	[0.00, 0.01]	[0.01, 0.01]	[8.96, 11.49]	[0.10, 4.14]
Two Classes SVDD	Training	[Min, Max]	[99.01, 99.01]	[92.13, 100.00]	[79.55, 84.72]	[66.66, 97.86]
	Testing	[Min, Max]	[88.95, 88.95]	[29.17, 82.33]	[67.62, 70.97]	[64.97, 90.58]
	Duration	[Min, Max]	[0.05, 0.16]	[2.70, 3.56]	[34.51, 37.26]	[46.13, 52.81]
The proposed approach	Training	[Min, Max]	[99.01, 99.38]	[95.57, 100]	[81.26, 91.84]	[97.29, 100.00]
	Testing	[Min, Max]	[88.42, 88.95]	[25.17, 83.17]	[63.12, 74.40]	[65.44, 92.04]
	Duration	[Min, Max]	[0.00, 0.01]	[0.01, 0.01]	[9.17, 11.77]	[0.10, 4.45]

sults in the majority of cases as compared to One-Class and Two-Classes SVDD. Also the recognition rate depends on the values taken by ρ and almost grows with increasing values of this latter. This improvement can be explained by the fact that when increasing the value of ρ, the distance between the center of the minimal hypersphere that encloses each target class and the negative samples increases which gives a tight description of the classes of interest.

- The duration to solve the Quadratic Programming problems: In the first two experiments, the duration to minimize the QPs problem of our method is equals to that of One-class SVDD and is lower compared to Two-classes version. In the second two experiments the duration to solve the QPs of our method grows little (by fractions of second) compared to the version One-class but still very lower than Two-classes.

As a result, the proposed approach outperforms both One-Class and Two-Classes SVDD in terms of learning performance, generalization ability and training time especially when choosing a good couple of the variables $\rho \in]0, 1[$ and σ.

5. CONCLUSION AND FUTURE WORK

In this paper, a novel approach to improve Support Vector Domain Description by using negative examples was proposed and successfully incorporated with the mathematical formulation of this classifier. The proposed method retains the same space and time complexity of One-class version and improves its classification accuracy. The performance of this new version was evaluated empirically on two challenging artificial problems and four benchmark datasets. The experimental results show that our method has achieved high classification accuracy with low training time compared to both One-class and Two-Classes SVDD.

In future, we plan to make a complete experimental study regarding the mutual dependencies between the regularization parameter ρ, the choice of the kernel function and its parameter (s), and the constant C of Support Vector Domain Description classifier. Also, we intend to found a function that can generate automatically the optimal value of the new parameter ρ. This function will take into consideration principally the distribution of the training dataset points in the feature space.

REFERENCES

1. G. Ritter and M. Gallegos, Outliers in statistical pattern recognition and an application to automatic chromosome classification. Pattern Recognition Letters, vol. 18, pp. 525-539, 1997
2. D. Tax and R. Duin, Data Domain Description Using Support Vectors. In Proceedings of the European Symposium on Artificial Neural Networks, pp. 251-256, Bruges, Belgium, 1999.
3. D. Tax, R. Duin, Support vector domain description. Pattern Recognition Letters. vol. 20, no. 11-13, pp. 1191-1199, 1999.
4. D. Tax, R. Duin, Support Vector Data Description. Machine Learning. vol. 54, pp. 45-66, 2004.
5. V. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, New York. 1995.
6. C. Cortes, V. Vapnik, Support-vector networks. Machine Learning, vol. 20, no. 3, pp. 273-297,1995.
7. J-C. Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods –Support Vector Learning. MIT Press, Cambridge, pp. 41-65, 1999.
8. P-J. Kim, H-J. Chang, D-S. Song, J-Y. Choi, Fast Support Vector Data Description Using K-Means. In Proceedings of the 4^{th} International Symposium on Neural Networks, ISNN 2007, pp. 506-514, Nanjing, China, 2007.
9. B. Schölkopf and A-J. Smola, Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge Mass: MIT Press, London, 2002.
10. K. Lee, D-W. Kim, D. Lee, K-H. Lee, Improving support vector data description using local density degree. Pattern Recognition, vol. 38, no. 10, pp. 1768-1771, 2005.
11. Li. Yin, H. Wang, W. Fan, Active learning based support vector data description method for robust novelty detection, Knowledge-Based Systems, vol. 153, pp. 40-52, 2018.
12. D. Lixiang, X. Mengyun, B. Tangbo, W. Jinjiang, A new support vector data description method for machinery fault diagnosis with unbalanced datasets, Expert Systems with Applications, vol. 64, pp 239-246, 2016.

13. S. Hu, D. Shi, X. Song, L. Fang, W. Yang, Q. Tong, Fault Diagnosis of Analog Circuits Based on Multi Classification SVDD Aliasing Region Identification. In: Qiao F., Patnaik S., Wang J. (eds) Recent Developments in Mechatronics and Intelligent Robotics. ICMIR 2017. Advances in Intelligent Systems and Computing, vol. 691. Springer, Cham, 2018.
14. K. Zhu, F. Mei, J. Zheng. Adaptive fault diagnosis of HVCBs based on P-SVDD and P-KFCM, Neurocomputing, vol. 240, pp. 127-136, 2017.
15. C. Gangolf, R. Dochow, G. Schmidt, T. Tamisier, SVDD: A proposal for automated credit rating prediction, International Conference on Control, Decision and Information Technologies (CoDIT),Metz, France, 2014.
16. Y. Cai, Y. Jiang, Credit scoring using incremental learning algorithm for SVDD, International Conference on Computer, Information and Telecommunication Systems (CITS),Kunming, China, 6-8 July 2016.
17. J. Shi and B. Xu, Credit Scoring by Fuzzy Support Vector Machines with a Novel Membership Function, Journal of Risk and Financial Management, vol. 9, no. 4, pp. 13, 2016.
18. V. Mygdalis, a. Iosifidis, a. Tefas, I. Pitas, Semi-supervised subclass support vector data description for image and video classification, Neurocomputing, vol. 278, pp. 51-61, 2018.
19. F. Sukru Uslu, H. Binol, M. Ilarslan, A. Bal, Improving SVDD classification performance on hyperspectral images via correlation based ensemble technique, Optics and Lasers in Engineering, vol. 89, pp. 169-177, 2017.
20. J.D. Bodapati, N. Veeranjaneyulu, Abnormal Network Traffic Detection Using Support Vector Data Description. In: Satapathy S., Bhateja V., Udgata S., Pattnaik P. (eds) Proceedings of the 5^{th} International Conference on Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing, vol 515. Springer, Singapore, 2017.
21. B-A. Tama, K-H. Rhee, An extensive empirical evaluation of classifier ensembles for intrusion detection task, International Journal of Computer Systems Science and Engineering, vol. 32, no 2, pp. 149–158, 2017.
22. M. El Boujnouni and M. Jedra, New Intrusion Detection System Based on Support Vector Domain Description with Information Gain Metric, International Journal of Network Security, vol. 20, no. 1, pp.25-34, Jan. 2018.
23. A. Ben-Hur, D. Horn, H-T. Siegelmann, V. Vapnik, Support vector clustering. Journal of Machine Learning Research, vol. 2, no. 12, pp. 125-137, 2001.
24. J. Lee and D. Lee, An improved cluster labeling method for support vector clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 27, pp. 461-464, 2005.
25. M. Wu and J. Ye. A Small Sphere and Large Margin Approach for Novelty Detection Using Training Data with Outliers. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 11, pp. 2088-2092, 2009.
26. J. Wang, P. Neskovic, L-N. Cooper, Pattern Classification via Single Spheres. Lecture Notes in Artificial Intelligence. vol. 3735, pp. 241-252, 2005.
27. K-R. Mäller, S. Mika, G. Rätsch, K. Tsuda and B. Schölkopf, An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks, vol. 12, no. 2, pp.181- 201, 2001.
28. A. Tavakkoli, M. Nicolescu, G. Bebis, M. Nicolescu, A support vector data description approach for background modeling in videos with quasi stationary backgrounds. International journal on artificial intelligence tools,vol. 17, no. 4, pp. 635-658, 2008.
29. C-S. Chu, I-W. Tsang, J-T. Kwok, Scaling up support vector data description by using core-sets. In Proceedings of the International Joint Conference on Neural Networks, pp. 425-430, Budapest, Hungary, 2004.
30. J. Liang, S. Liu, and D. Wu, Fast Training of SVDD by Extracting Boundary Targets, Iranian journal of electrical and computer engineering, vol. 8, no. 2, pp. 133-137, 2009.
31. T-K. Ho and E-M. Kleinberg, Building projectable classifers of arbitrary complexity. In Proceedings of the 13^{th} International Conference on Pattern Recognition, pp. 880-885, Vienna, Austria, 1996.
32. L. Kaufman, Solving the quadratic programming problem arising in support vector classification. In Advances in Kernel Methods Support Vector Learning, B. Schollkopf, C-J-C. Burges, and A-J. Smola, eds., MIT Press, pp. 147-167. 1999.
33. M. Zhu, Y. Wang, S. Chen and X. Liu, Sphere-structured support vector machines for multi-class pattern recognition. Lecture Notes in Computer Science, vol. 2639, pp. 589-593, 2003.
34. Q. Wu, X. Shen, Y. Li, G. Xu, W. Yan, G. Dong and Q. Yang, Classifying the Multiplicity of the EEG Source Models Using Sphere-Shaped Support Vector Machines IEEE Transactions on Magnetics, vol. 41, no. 5, pp. 1912-1915, 2005.
35. P-Y. Hao, J-H. Chiang and Y-H. Lin, A new maximal-margin spherical-structured multi-class support vector machine Applied Intelligence, vol. 30, no. 2, pp. 98-111. 2009.
36. "UCI repository of machine learning databases," http://archive.ics.uci.edu/ml/.
37. L. Yuan, E. Yao and G. Tan, Automated and precise event detection method for big data in biomedical imaging with support vector machine, International Journal of Computer Systems Science and Engineering, vol. 33, no. 2, 2018.