Quasi Support Vector Data Description (QSVDD)

Yonos Allahyari[1] and Hadi Sadoghi-Yazdi[1,2]

[1]*Department of Computer Engineering,*
Ferdowsi University of Mashhad, Mashhad, Iran
[2]*Center of Excellence on Soft Computing and Intelligent Information Processing,*
Ferdowsi University of Mashhad

Ynsalh@gmail.com, h-sadoghi@um.ac.ir

Abstract

In this paper it is proposed a boundary based classifier that is inspired by SVDD and makes an important role for gravity center of training samples. In the proposed method all training samples intervene in determining the classifier boundary. Consequently, the relevant classifier isn't placed in the group of the support vector machines. Due to the employment of this idea, this method is called "Quasi Support Vector Data Description (QSVDD)". The ability of this method to eliminate the effect of noisy training samples on synthetic data is shown. Experiments on real data sets show that the proposed method describes more accurately lots of real data sets than SVDD.

Keywords: *Support Vector Data Description; one-class classification, SVM*

1. Introduction

The one-class classification problem is an interesting field in pattern recognition and machine learning researches. In this kind of classification, we assume the one class of samples classified as the target class and the rest of samples are classified as the outlier. One-class classification is particularly significant in Applications where only a single class of data objects is applicable and easy to obtain. Objects from the other classes could be too difficult or expensive to be made available. So we would only describe the target class to separate it from the outlier class. Since several models of one-class classifier design has been introduced. There are two approaches have been proposed to resolve this problem [1]. First way is estimate the probability density function of learning samples and uses this probability density function with a threshold on its density. One of the popular methods to approximate probability density function is Gaussian model, the mixture of Gaussian and parzen density [2, 3]. In the second approach an optimized boundary around the learning samples was searched. KNN and SVDD are examples of boundary methods [4, 5].

The SVDD is a kind of one-class classification method based on Support Vector Machine [6]. It tries to construct a boundary around the target samples by covering the target samples within a minimum hyper sphere. Inspired by the support vector machines (SVMs), the SVDD decision boundary described by a few target objects, that known as support vectors (SVs). A more flexible boundary can be obtained with the introduction of kernel functions, by which samples are mapped into a high-dimensional feature space. The most commonly used kernel function is Gaussian kernel [7]. After introducing this method by Tax [5] some of researchers try to improve the method for generating better results and some researchers try to reduce the runtime complexity and others try to generate new classifiers based on SVDD. Lee et. al., [8] presented an improving support vector data description using local density degree (D-SVDD); results showed that the D-SVDD had better performance than SVDD and a k-nearest-neighbor data description method. A method named weighted SVDD [9] in some articles has been referred to. This method presented a new approach to eliminate the effect of noisy. Zhang et al. [9] proposed a novel fuzzy classifier for multi-classification problems based on SVDD and improved PCM, and it reduced the effect of outliers and it yield lower error rate.

Xiao et. al., [10] proposed a new method for designing a one-class classifier based on SVDD for multi-distributed samples. Tingting Mu et. al., [11] propose two variations of the support vector data description with negative samples that learn a closed spherically shaped boundary around a set of samples in the target class by involving different forms of slack vectors. Wang et. al., [12] proposed a novel approach to generate artificial outliers for support vector data description with boundary value method. Implementing this method leads to decrease error rate. Park et al. [13] presented a new denoising method that uses the SVDD, the geodesic projection of the noisy point to the surface of the SVDD ball in the feature space, and a method for finding the preimage of the denoised feature vectors. Liu et. al., [14] proposed a method to reduce runtime complexity in classification process. Huang et. al., [15] presented an improved support vector data description method named two-class support vector data description (TC-SVDD). The proposed method can give each class of objects in the target data set a hyper sphere-shaped description simultaneously, if the target data set contains two classes of objects.

2. Support Vector Data Description

The objective of SVDD is to find a sphere or domain with minimum volume containing all or most of the data. Let $\{x_i | i = 1,2,\ldots \mathrm{n}\} \subset R^d$ be the given training data set. Let a and R denote the center and radius of the sphere, respectively. This goal is formulated as a constrained convex optimization problem:

$$\begin{aligned} &\text{Minimize } R^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{subject to} (x_i - a)^T (x_i - a) \le R^2 + \xi_i,\ \xi_i \ge 0,\ i = 1,\ldots,\mathrm{n}. \end{aligned} \tag{1}$$

where ξ_i is a slack variable that allows the possibility of outliers in the training data set. The parameter C controls the trade-off between the volume and the training errors. Constructing the Lagrangian function with Lagrange multipliers α_i, γ_i gives

$$L = R^2 + C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\alpha_i (x_i - a)(x_i - a)^T - R^2 - \xi_i) - \sum_{i=1}^{n}\gamma_i\xi_i. \tag{2}$$

Setting partial derivatives of R, a and ξ_i to zero gives the constraints

$$\frac{\partial L}{\partial R} = 2R(1 - \sum_{i=1}^{n}\alpha_i) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n}\alpha_i = 1. \tag{3}$$

$$\frac{\partial L}{\partial a} = 2\sum_{i=1}^{n}\alpha_i (x_i - a) = 0 \quad \Rightarrow \quad a = \sum_{i=1}^{n}\alpha_i x_i. \tag{4}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad \alpha_i + \gamma_i = C. \tag{5}$$

Substituting (3)–(5) into (2), gives dual problem of (1)

$$\text{Maximize} \sum_{i=1}^{n} \alpha_i x_i^T x_i - \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j x_i^T x_j .$$
$$\text{subject to} \begin{cases} 0 \le \alpha_i \le C, \; i = 1,\ldots,\text{n}. \\ \sum_{i=1}^{n} \alpha_i = 1. \end{cases} \tag{6}$$

Solving problem (6) gives a set α_i. A training object x_i and its corresponding α_i satisfy one of the three following conditions:

$$\begin{aligned} \|x_i - a\|^2 < R^2 &\Rightarrow \alpha_i = 0 \\ \|x_i - a\|^2 = R^2 &\Rightarrow 0 < \alpha_i < C \\ \|x_i - a\|^2 > R^2 &\Rightarrow \alpha_i = C \end{aligned} \tag{7}$$

The objects with the coefficients $\alpha_i > 0$ are called the support vectors. From the above relations we can see only the support vectors are needed in the description of the sphere. The center of the sphere could be calculated by (4). The radius R of the sphere can be obtained by calculating the distance from the center of the sphere to any support vector with $0 < \alpha_i < C$, which provides the sparse representation of the domain description.

To determine whether a test point z is within the sphere, the distance from z to the center of the sphere has to be calculated. A test object z is accepted when this distance is smaller than the radius, i.e.,

$$\|z - a\|^2 = z^T z - 2\sum_{i=1}^{n} \alpha_i (z^T x_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j (x_i^T x_j) \le R^2 \tag{8}$$

The method can be made more flexible, the inner product $x_i^T x_j$ can be replaced by a new inner product $K(x_i, x_j)$ satisfying Mercer's theorem. An ideal kernel function would map the target data onto a bounded, spherically shaped area in the feature space and outlier objects outside this area. The polynomial kernel and the Gaussian kernel.

According to (4) and (7) center of sphere is only affected by support vectors. Therefore decision boundary tends to be located near the rejected samples that identified as noisy or outlier samples. Even reducing the parameter C does not solve the problem and will increase training error. Figure 1 demonstrates disability of SVDD to overcome outliers.

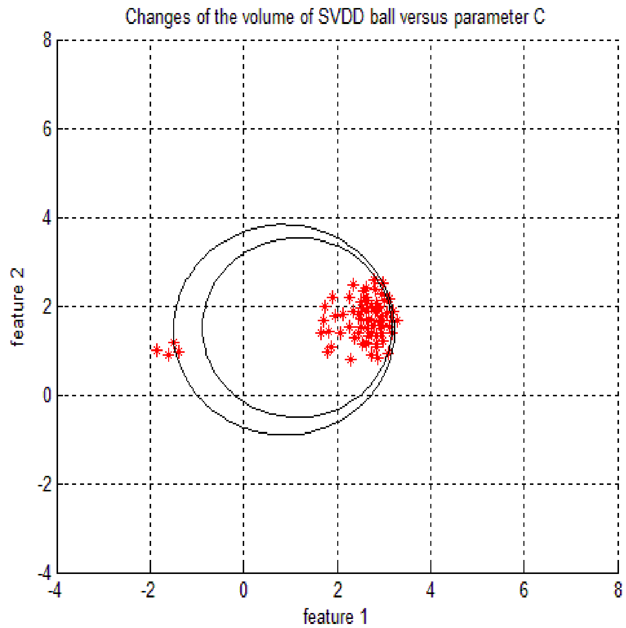

Figure 1. The Center of Sphere is Affected by Outliers that Results in Inappropriate Description for Target Class. Even Reducing Parameter *C* does not Solve this Problem and will Increase Training Error

To overcome this disadvantage, the "Quasi Support Vector Data Description (QSVDD)" method is presented in the next section which the gravity center of the samples has a decisive role.

3. Quasi Support Vector Data Description (QSVDD).

In this section, mathematical model of SVDD will be changed as purposeful. The goal is to give importance to gravity center of samples and incorporate this idea with SVDD method. Distance between center of sphere and samples gravity center is formulated as:

$$\|\bar{x}-a\|^2 = \left\|\sum_{i=1}^{n}\frac{x_i}{n}-a\right\|^2 = \frac{1}{n^2}\left\|\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}a\right\|^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-a)^T(x_j-a) \tag{9}$$

and proposed idea is formulated as a constrained optimization problem:

$$\begin{aligned}&\text{Minimize} R^2+\frac{B}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-a)^T(x_j-a)+C\sum_{i=1}^{n}\xi_i.\\&\text{subject to}\begin{cases}(x_i-a)^T(x_i-a)\le R^2+\xi_i.\\ \xi_i\ge 0,\ i=1,\ldots,\text{n}.\end{cases}\end{aligned} \tag{10}$$

In (10), the parameter, *B,* implies the importance degree of gravity center and the parameter, *C*, has the same role that it has in SVDD method and ξ_i, is a slack variable. Constructing the Lagrangian function with Lagrange multipliers α_i, γ_i gives

$$L = R^2 + \frac{B}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - a)^T(x_j - a) + C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\alpha_i((x_i - a)^T(x_i - a)^T - R^2 - \xi_i) - \sum_{i=1}^{n}\gamma_i\xi_i. \quad (11)$$

Setting partial derivatives of *R*, *a* and ξ_i to zero gives the constraints

$$\frac{\partial L}{\partial R} = 0 \quad \Rightarrow \quad 2R - 2R\sum_{i=1}^{n}\alpha_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{n}\alpha_i = 1. \quad (12)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad \alpha_i + \gamma_i = C \quad \Rightarrow \quad 0 \le \alpha_i \le C. \quad (13)$$

$$\frac{\partial L}{\partial a} = 0 \quad \Rightarrow \frac{B}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(-(x_i - a) - (x_j - a)) + \sum_{i=1}^{n}\alpha_i(-2(x_i - a)) = 0 \;\Rightarrow\; a = \frac{\sum_{i=1}^{n}(\alpha_i + \frac{B}{n})x_i}{1 + B}. \quad (14)$$

Substituting (12)–(14) into (11), gives dual problem of (10)

$$\text{Maximize } \frac{1}{1+B}\left(-2\frac{B}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i x_i^T x_j - \frac{B^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^T x_j - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j x_i^T x_j\right) + \sum_{i=1}^{n}\alpha_i x_i^T x_i + \frac{B}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^T x_j. \quad (15)$$
$$\text{subject to } \begin{cases} \sum_{i=1}^{n}\alpha_i = 1. \\ 0 \le \alpha_i \le C.\ i = 1,\dots,n. \end{cases}$$

This optimization problem is equivalent to a convex quadratic problem with global minimum, when $B>-1$ holds. Solving problem (15) gives a set α_i. A training object x_i and its corresponding α_i satisfy one of the three conditions that are mentioned in (7). The radius *R* of the sphere can be obtained by calculating the distance from the center of the sphere to any support vector with $0<\alpha_i<C$. To determine whether a test point *z* is within the sphere, the distance from z to the center of the sphere has to be calculated. A test object *z* is accepted when this distance is smaller than the radius, i.e.,

$$\|z - a\|^2 = x^T x - \frac{2}{B+1}\left(\sum_{i=1}^{n}\alpha_i x_i^T z + \frac{B}{n}\sum_{i=1}^{n}x_i^T z\right) + \frac{1}{(B+1)^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j x_i^T x_j + \frac{B^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^T x_j + 2\frac{B}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i x_i^T x_j\right) \le R^2 \quad (16)$$

Considering (14), resulted hyper-spherical discriminator gets affected by all training samples. Therefore, this method is called "Quasi Support Vector Data Description (QSVDD)". Now the capability of QSVDD is revealed in comparison with SVDD in order to overcome the problem of outliers. Figure 2 shows that the center of sphere is affected by all training samples. Therefore, the effect of outliers in determining the center is decreased.

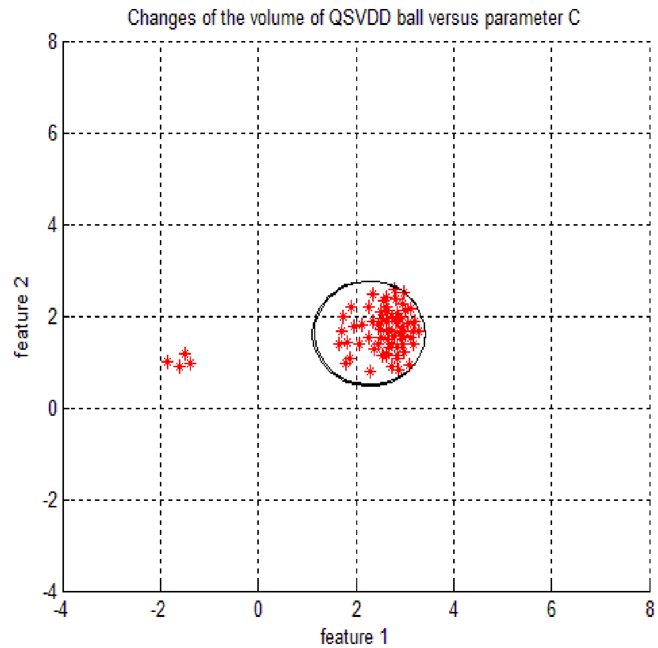

Figure 2. The Center of Sphere is affected by All Target Samples. Thus, the Effect of Outliers is Reduced

4. Mathematical Discussion for QSVDD Method

Optimization problem (10) would be equivalent to (2) in a special case, when $B=0$ holds. Therefore, QSVDD is a more general method than SVDD. Hence, it's expected to find better solutions for QSVDD than SVDD. Regarding (15), problem has a global and unique solution when the parameter B take any value in the interval $(-1,+\infty)$. Consequently, if it assumes that error function with respect to parameter B is a monotonically increasing or decreasing function at the vicinity of zero, then it's expected to find better solutions for QSVDD than SVDD at neighborhood of zero. So, to demonstrate superiority of QSVDD method, in experiments it is tried to search at the vicinity of zero for finding optimum value of B. Experimental results verify that this assumption is close to reality. Also relation (16) results that runtime complexity for QSVDD is O(n) while runtime complexity for SVDD and D-SVDD[8] is O(|SV|), such that SV is the set of support vector samples. This is a disadvantage of QSVDD compared with SVDD and D-SVDD.

5. Experimental Results and Comparative Analysis

Before conducting our experiments on real datasets, we perform our experiments on syntactic data in high dimensional space with RBF kernel. In Figures 3 and 4 flexible boundaries that are resulted by SVDD and QSVDD are shown respectively. It is obvious that the effect of outliers is reduced when we use QSVDD method.

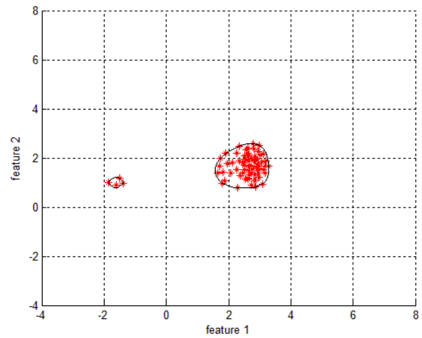

Figure 3. The Description is Affected by Outliers in svdd Method. In this Figure Boundary of Description Tends to be Located at Near of the Outliers

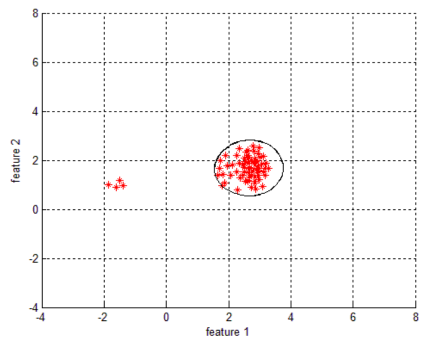

Figure 4. The Description isn't affected by Outliers in QSVDD Method therefore Drawback of SVDD is Neutralized

Now we compare results of applying SVDD and QSVDD on image (a) in Figure 5. In our experiment we select training samples from the rectangular piece of the flower in image (a) and training process is done for SVDD and QSVDD methods with equal parameters for *C* and sigma. After applying test process on image (a), SVDD method yields Image (b) and QSVDD method yields Image (c). We observe that the result of QSVDD method is better than SVDD in flower identification.

(a)-main image. (b)-resulted by svdd. (c)-resulted by Qsvvd.

Figure 5. This Figure Presents Quality of Flower Identification for SVDD and QSVDD Methods

To investigate the success of these attempts on real datasets, we conducted various tests in which the SVDD, QSVDD and D-SVDD[8] methods are applied to Iris, Haberman, Glass, wine and balance-scale Datasets from UCI Repository of machine learning Database[1]. We assume that our optimum parameters are those that simultaneously decrease error type-I and error type-II. Therefore, we introduce the error function as relation (17) that must be minimized in parameter tuning process:

$$error\ \ function = 1 - \sqrt{(1 - \frac{FN}{TP+FN}) \times (1 - \frac{FP}{TN+FP})}. \tag{17}$$

TP: number of target test samples that classified correctly.
FP: number of outlier test samples that classified incorrectly.
TN: number of outlier test samples that classified correctly.
FN: number of target test samples that classified incorrectly.

The model parameters were found by cross validation for proposed methods. The N-fold M times cross validation method is a scheme invoked to predict the error ratio of learning technology. The idea is to divide the original data into target data and outlier data. The target data is divided randomly into N parts, in each of which the class is presented in approximately the same proportions as in the full dataset. Each part is held out in turn and learning scheme trained on the remaining (N-l) parts; then its error ratio is calculated on the holdout set. Thus the learning procedure is executed a total of N times on different training sets. Finally, the N×M error estimates are averaged to yield an overall error estimate. In our experiments we Invoke 3 fold 15 times (total of 45 run) method to predict error ratio of learning process. The results of our experiments are shown in Tables 1-5. In these experiments, the samples are mapped into a high-dimensional feature space with RBF kernel for flexible description.

Table 1. Results of QSVDD, D-SVDD and SVDD on Iris Dataset

	Target class	SVDD	QSVDD	D-SVDD
Iris Dataset	Class1	96.26	97.09	**97.97**
	Class2	**92.78**	92.78	92.78
	Class3	90.92	**91.12**	90.92

Table 2. Results of QSVDD, D-SVDD and SVDD on Haberman Dataset

	Target class	SVDD	QSVDD	D-SVDD
Haberman Dataset	Class1	60.36	**64.14**	61.12
	Class2	51.17	**52.80**	51.17

[1] UCI repository of machine learning database, http://archive.ics.uci.edu/ml/datasets.html..

Table 3. Results of QSVDD, D-SVDD and SVDD on Glass Dataset

	Target class	SVDD	QSVDD	D-SVDD
Glass Dataset	Class1	69.54	**70.57**	69.54
	Class2	69.95	**70.88**	69.95
	Class3	71.85	**72.16**	71.85
	Class5	69.67	**71.78**	69.67

Table 4. Results of QSVDD, D-SVDD and SVDD on Wine Dataset

	Target class	SVDD	QSVDD	D-SVDD
Wine Dataset	Class1	91.82	97.74	**98.02**
	Class2	77.36	**81.24**	78.89
	Class3	92.91	**95.44**	95.17

Table 5. Results of QSVDD, D-SVDD and SVDD on Balance-scale Dataset

	Target class	SVDD	QSVDD	D-SVDD
Balance-scale Dataset	Class1	87.82	87.95	**88.12**
	Class2	87.67	**87.96**	87.67
	Class3	**87.88**	87.88	87.88

If we compare the results that obtained in experiments, we see that the answer of QSVDD and D-SVDD method in the worst case is equal to the answer of SVDD method, while each of the QSVDD and D-SVDD methods are not dominant over another.

6. Conclusion and Future Works

We have proposed QSVDD method that emphasis on direct the center of sphere towards gravity center of training samples. Our aim is reducing the effect of outlier samples. Furthermore, QSVDD is equivalent to SVDD, when B=0 holds. Thus, proposed method is a general extension of the SVDD. Moreover, the drawback of proposed method is that it isn't a support vector method and all training samples intervene in determining the classifier boundary and runtime complexity is O(n). One possible future research is to examine our method for pattern de-noising analogous pattern de-noising based on SVDD that proposed in [13]. Another possible future work is extending Two-class svdd [15] by QSVDD.

References

[1] D. M. J. Tax, "One-class classification: concept learning in the absence of counter-examples", Technische Universiteit Delft, Netherlands, vol. 65, **(2001)**.
[2] E. Parzen, "On estimation of a probability density function and mode", Annals of Mathematical Statistics, vol. 33, **(1962)**, pp. 1065-1076.
[3] C. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, Walton Street, Oxford OX2 6DP, **(1995)**.

[4] A. Ypma and R. Duin, "Support objects for domain approximation", Skovde (Sweden), Proceedings of Int. Conf. on Artificial Neural Networks (ICANN98), **(1998)**.
[5] D. M. J. Tax and R. P. W. Duin, "Support Vector Data Description", Kluwer Academic Publishers, Machine Learning, vol. 54, **(2004)**, pp. 45-66.
[6] D. M. J. Tax and R. P. W. Duin, "Support vector domain description", Pattern Recognition Letters, vol. 20, **(1999)**, pp. 1191-1199.
[7] S. M. Guo, L. C. Chen and J. S. H. Tsai, "A boundary method for outlier detection based on support vector domain description", Pattern Recognition, vol. 42, **(2009)**, pp. 77-83.
[8] K. Y. Lee, D. W. Kim, D. Lee, K. H. Lee, "Improving support vector data description using local density degree", Pattern Recognition, vol. 38, **(2005)**, pp. 1768 – 1771.
[9] Y. Zhang, Z. -X. Chi, K. -Q. Li, "Fuzzy multi-class classifier based on support vector data description and improved PCM", Expert Systems with Applications, vol. 36, **(2009)**, pp. 8714–8718.
[10] Y. Xiao, B. Liu, L. Cao, X. Wu, C. Zhang, Z. Hao, F. Yang and J. Cao, "Multi-sphere Support Vector Data Description for Outliers Detection on Multi-Distribution Data", IEEE International Conference on Data Mining Workshops, **(2009)**.
[11] T. Mu and A. K. Nandi, "Multiclass Classification Based on Extended Support Vector Data Description", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, vol. 39, no. 5, **(2009)** October.
[12] C. -K. Wang, Y. Ting, Y. -H. Liu and G. Hariyanto, "A Novel Approach to Generate Artificial Outliers for Support Vector Data Description", IEEE International Symposium on Industrial Electronics (ISlE 2009) Seoul Olympic Parktel, Seoul, Korea, **(2009)** July 5-8.
[13] J. Park, D. Kang, J. Kim, J. T. Kwok and I. W. Tsang, "Pattern De-Noising Based on support Vector Data Description", Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, **(2005)** July 31 - August 4.
[14] V. -H. Liu and V. -C. Liu, "Reducing the Run-Time Complexity of Support Vector Data Descriptions", Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, **(2009)** June 14-19.
[15] G. Huan, H. Chen, Z. Zhou, F. Yin and K. Guo, "Two-class support vector data description", Pattern Recognition, vol. 44 , **(2011)**, pp. 320–329.

Author

Hadi Sadoghi Yazdi is currently an Associate Professor of Computer Science and Engineering at Ferdowsi University of Mashhad (FUM). He received his B.S. degree in Electrical Engineering from FUM in 1994, and received his M.S. and Ph.D. degrees in Electrical Engineering from Tarbiat Modares University in 1996 and 2005, respectively. Dr. Sadoghi Yazdi has received several awards including Outstanding Faculty Award and Best System Design Award in 2007. His research interests are in the areas of Pattern Recognition, Machine Learning, Machine Vision, Signal Processing, Data Mining and Optimization.