International Journal of Fuzzy Logic and Intelligent Systems, vol. 7, no. 1, March 2007 pp. 1-6

A Modified Approach to Density-Induced Support Vector Data Description

Jooyoung Park and Daesung Kang

Department of Control and Instrumentation Engineering, Korea University
Jochiwon, Chungnam, 339-700, Korea

Abstract

The SVDD (support vector data description) is one of the most well-known one-class support vector learning methods, in which one tries the strategy of utilizing balls defined on the feature space in order to distinguish a set of normal data from all other possible abnormal objects. Recently, with the objective of generalizing the SVDD which treats all training data with equal importance, the so-called D-SVDD (density-induced support vector data description) was proposed incorporating the idea that the data in a higher density region are more significant than those in a lower density region. In this paper, we consider the problem of further improving the D-SVDD toward the use of a partial reference set for testing, and propose an LMI (linear matrix inequality)-based optimization approach to solve the improved version of the D-SVDD problems. Our approach utilizes a new class of density-induced distance measures based on the RSDE (reduced set density estimator) along with the LMI-based mathematical formulation in the form of the SDP (semi-definite programming) problems, which can be efficiently solved by interior point methods. The validity of the proposed approach is illustrated via numerical experiments using real data sets.

Key words : one-class problems, D-SVDD, SVDD

1. Introduction

With a great deal of recent successes in theoretical and empirical studies, the support vector learning method has grown up as a viable tool in the area of intelligent systems [1, 2]. Among the important application areas for the support vector learning, we have the one-class classification problems [3, 4, 5, 6, 7, 8, 9, 10]. In the problems of one-class classification, we are in general given the training data most of which are from the normal class, and after the training phase is finished, we are required to decide whether each testing vector belongs to normal class or abnormal class. One of the most well-known support vector learning methods for the one-class problems is the SVDD (support vector data description) [3, 4, 5]. In the SVDD, balls are used for expressing the region for the normal class. Since balls on the input domain can express only limited class of regions, the SVDD in general enhances its expressing power by utilizing balls on the feature space instead of the balls on the input domain. Recently, with the objective of overcoming a possible drawback of the SVDD which treats all data with equal importance, the so-called D-SVDD (density-induced support vector data description) was proposed incorporating the idea that the data in a higher density region are more significant than those in a lower density region in describing the normal class data [6]. In this paper, we consider the problem of further improving the D-SVDD, and utilize an LMI (linear matrix inequality)-based optimization approach to solve the proposed improved version of the D-SVDD problems. More specifically, the main issues addressed in this paper are as follows: First, observing that the D-SVDD utilizes the notion of the local density degree which is computed based on the K-nearest neighborhood, which requires the full reference set for testing and is in practice very expensive, we propose to use a new local density degree based on the so-called RSDE (reduced set density estimator) [11], which is of a sparse representation in the weighting coefficients, instead of the K-nearest neighborhood. Next, we observe that the dual representation of both the conventional D-SVDD and its improved version proposed here can be converted into the LMI-based optimization called the SDP (semi-definite programming) problems, and propose to solve the D-SVDD problems after converting them into the SDP problems. Since the SDP problems can be efficiently solved by reliable and efficient convex optimization techniques [12], the conversion into the form of the SDP is of great practical value.

The remaining parts of this paper are organized as follows: In Section 2, preliminaries are provided regarding SVDD, D-SVDD, LMI, SDP, and RDSE. Our main results on improving the D-SVDD and the LMI-based formulation are presented in Section 3 together with experimental

Manuscript received Feb. 8, 2007; revised Mar. 5, 2007.

illustrations. Finally, in Section 4, concluding remarks are given.

2. Preliminaries

2.1 Support vector data description

The SVDD method, which approximates the support of objects belonging to normal class, is derived as follows [3, 4]: Consider a ball B with the center $a \in \mathcal{R}^d$ and the radius R, and the training data set D consisting of objects $x_i \in \mathcal{R}^d$, $i = 1, \cdots, N$. Note that since the training data may be prone to noise, some part of the training data could be abnormal objects. The main idea of the SVDD is to find a ball that can achieve two conflicting goals simultaneously. First, it should be as small as possible, and with equal importance, it should contain as many training data as possible. Obviously, satisfactory balls satisfying these objectives can be obtained by solving the following optimization problem:

$$\begin{array}{ll} \min & L_0(R^2, a, \xi) = R^2 + C\sum_{i=1}^N \xi_i \\ \text{s. t.} & \|x_i - a\|^2 \leq R^2 + \xi_i, \ \xi_i \geq 0, \quad i = 1, \cdots, N. \end{array} \tag{1}$$

Here, the slack variable ξ_i represents the penalty associated with the deviation of the i-th training pattern outside the ball. The objective function of (1) consists of the two conflicting terms, *i.e.*, the square of radius, R^2, and the total penalty $\sum_{i=1}^N \xi_i$. The constant C controls relative importance of each term; thus called the trade-off constant. Note that the dual problem of (1) is:

$$\begin{array}{ll} \max_\alpha & \sum_{i=1}^N \alpha_i \langle x_i, x_i\rangle - \sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_j \langle x_i, x_j\rangle \\ \text{s. t.} & \sum_{i=1}^N \alpha_i = 1, \ \alpha_i \in [0, C], \quad i = 1, \cdots, N. \end{array} \tag{2}$$

From the Kuhn-Tucker condition one can express the center of the SVDD ball as $a = \sum_{i=1}^N \alpha_i x_i$, and can compute the radius R utilizing the distance between a and any support vector x_i on the ball boundary. After the training phase is over, one may decide whether a given test point $x \in \mathcal{R}^d$ belongs to the normal class utilizing the following criterion: $f(x) \triangleq R^2 - \|x - a\|^2 \geq 0$. In order to express more complex decision regions in $\mathcal{R}^d$, one can use the so-called feature map $\phi : \mathcal{R}^d \rightarrow F$ and balls defined on the feature space F. Proceeding similarly as the above and utilizing the kernel trick $\langle\phi(x), \phi(z)\rangle = k(x, z)$, one can find the corresponding feature-space SVDD ball B_F in F, whose center and radius are a_F and R_F, respectively. If the Gaussian function

$$k(x, z) = \exp(-\|x - z\|^2/\sigma^2) \tag{3}$$

is chosen for the kernel, one has $k(x, x) = 1$ for each $x \in \mathcal{R}^d$, which is assumed throughout this paper. Finally, note that in this case, the SVDD formulation is equivalent to

$$\begin{array}{ll} \min_\alpha & \sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_j k(x_i, x_j) \\ \text{s. t.} & \sum_{i=1}^N \alpha_i = 1, \ \alpha_i \in [0, C], \quad i = 1, \cdots, N, \end{array} \tag{4}$$

and the resulting criterion for the normality is represented by

$$\begin{array}{rcl} f_F(x) & \triangleq & R_F^2 - \|\phi(x) - a_F\|^2 \\ & = & R_F^2 - 1 + 2\sum_{i=1}^N \alpha_i k(x_i, x) \\ & & \quad - \sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_j k(x_i, x_j) \\ & \geq & 0. \end{array} \tag{5}$$

2.2 Density-induced support vector data description

Despite its usefulness, the SVDD does not have an explicit mechanism that can reflect individual significance of the data points separately. However, there are many cases in real world problems in which each data point deserves different degree of significance. Recently, an extension of the SVDD that can assign distinctive significance for each data point according to the local density of data was introduced by Lee *et al.* [6], and the extension, which is called the D-SVDD, utilizes a density-induced distance measure based on the notion of the local density degree so that the data having higher density degree can attract more significance than other data having lower density degree. The essence of the D-SVDD can be summarized as follows [6]:

- The local density degree $\rho : \mathcal{R}^d \rightarrow [0, \infty)$ is defined for each data point $x_i \in D$ as follows:

$$\rho(x_i) = \exp(\omega \times \frac{\text{MEAN}_K}{d(x_i, x_i^K)}), \ i = 1, \cdots, N, \tag{6}$$

where x_i^K is the K-th nearest neighborhood of x_i with respect to the distance measure

$$d(x_i, x_i^K) \triangleq \| x_i - x_i^K \|, \tag{7}$$

MEAN_K is the mean distance of the K-th nearest neighborhoods of all data, *i.e.*,

$$\text{MEAN}_K \triangleq \frac{1}{N}\sum_{i=1}^N d(x_i, x_i^K), \tag{8}$$

and $\omega \in [0, 1]$ is the parameter controlling the sharpness of ρ. Note that ρ yields higher value for data in a higher density region.

- The density-induced distance from each data point $x_i \in D$ to the D-SVDD ball center $a_D \in F$ is defined as follows:

$$\delta(x_i) = \{\rho(x_i)\}^{1/2} \parallel \phi(x_i) - a_D \parallel . \quad (9)$$

- Proceeding similar to the SVDD formulation utilizing the above density-induced distance instead of the conventional feature-space distance, the problem of approximating the support of the normal data with feature-space balls is represented as the following optimization problem:

$$\begin{array}{ll} \min & \tilde{L}_0(R_D^2, a_D, \xi) = R_D^2 + C\sum_{i=1}^N \xi_i \\ \text{s. t.} & \delta^2(x_i) \le R_D^2 + \xi_i, \\ & \xi_i \ge 0, \quad i = 1, \cdots, N. \end{array} \quad (10)$$

Utilizing the Lagrangian optimization theory and the kernel trick along with the Gaussian kernel, one can show that the center of the D-SVDD ball is expressed as

$$a_D = (\sum_{i=1}^N \alpha_i \rho(x_i)\phi(x_i))/(\sum_{i=1}^N \alpha_i \rho(x_i)), \quad (11)$$

and the dual representation of (10) becomes equivalent to the following optimization problem [6]:

$$\begin{array}{ll} \min_\alpha & \frac{\sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_j\rho(x_i)\rho(x_j)k(x_i,x_j)}{\sum_{i=1}^N \alpha_i\rho(x_i)} \\ & -\sum_{i=1}^N \alpha_i\rho(x_i) \\ \text{s. t.} & \sum_{i=1}^N \alpha_i = 1, \\ & \alpha_i \in [0, C], \quad i = 1, \cdots, N. \end{array} \quad (12)$$

Note that if ρ satisfies $\rho(x_i) = 1$ for each i, which is the case observed when $\omega = 0$ in (6), the above D-SVDD problem can be reduced to a problem equivalent to the conventional SVDD problem (4). Also note that in [6], the following is used as the criterion for testing the normality of a testing data point $x \in \mathcal{R}^d$:

$$\begin{array}{l} f_D(x) \triangleq R_D^2 - \|\phi(x) - a_D\|^2 \\ = R_D^2 - 1 + \frac{2}{T}\sum_{i=1}^N \alpha_i\rho(x_i)k(x_i, x) \\ -\frac{1}{T^2}\sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_j\rho(x_i)\rho(x_j)k(x_i, x_j) \\ \ge 0, \end{array} \quad (13)$$

where $T \triangleq \sum_{i=1}^N \alpha_i\rho(x_i)$.

2.3 Linear matrix inequalities and SDP problems

Among the important tools of this paper are the linear matrix inequalities, which mean the inequality constraints of the form [12]

$$A(x) \triangleq A_0 + x_1A_1 + \cdots + x_mA_m < 0, \quad (14)$$

where $x \triangleq (x_1, \cdots, x_m)$ is the decision variables, $A_0, \cdots, A_m$ are given symmetric matrices and "$<$" stands for "negative definite". Note that since $A(y) < 0$ and $A(z) < 0$ imply that $A(\lambda y + (1-\lambda)z) < 0$ for any $\lambda \in [0, 1]$, the LMI (14) is a convex constraint on the variable x. Also note that since multiple LMIs $A^{(1)}(x) < 0, \cdots, A^{(p)}(x) < 0$ can be expressed as the single LMI $\text{diag}(A^{(1)}(x), \cdots, A^{(p)}(x)) < 0$, there is no distinction between a set of LMIs and a single LMI. It is well-known that many of LMI-based optimization problems can be solved in polynomial time [12], and a toolbox of Matlab for convex problems involving LMIs is readily available [13]. The class of optimization problems having a linear objective under LMI constrains is called the SDP problems, and will play an important role here in this paper. Note that the SDP problems are given in the following form, and the built-in function of the Matlab toolbox solving the class is "mincx":

$$\begin{array}{ll} \min & c^T x \\ \text{s. t.} & A(x) < B(x), \end{array} \quad (15)$$

where $A(x) < B(x)$ above is a short hand notation for an LMI inequality as in (14).

2.4 Reduced set density estimator

Based on a data sample $D = \{x_1, \cdots, x_N\} \subset \mathcal{R}^d$ drawn from the density $p(x)$, the general form of a kernel density estimator is given as

$$\hat{p}_h(x; \gamma) = \sum_{i=1}^N \gamma_i \, \mathcal{G}_h(x, x_i), \quad (16)$$

where

$$\mathcal{G}_h(x, x_i) \triangleq \frac{1}{h^d}\mathcal{G}(\frac{x - x_i}{h}), \quad (17)$$

and as for the density $\mathcal{G}$, we consider the Gaussian probability density function $\mathcal{G}(\cdot) \triangleq \frac{1}{(2\pi)^{d/2}}\exp(-\frac{1}{2}\|\cdot\|^2)$ in this paper. As shown in [11], the γ_i of the kernel density estimator (16) minimizing the ISE (integrated squared error) criterion defined as

$$\int_{\mathcal{R}^d} |p(x) - \hat{p}_h(x; \gamma)|^2 dx \quad (18)$$

can be obtained by the following quadratic programming problem for $\gamma \triangleq [\gamma_1 \cdots \gamma_N]^T$:

$$\begin{array}{ll} \min & \gamma^T C_h \gamma/2 - \gamma^T G_h 1_N \\ \text{s. t.} & \sum_{j=1}^N \gamma_j = 1, \ \gamma_i \ge 0, \quad i = 1, \cdots, N, \end{array} \quad (19)$$

where C_h is the $N \times N$ matrix whose (i,j)-th entry is

$$C_h(x_i, x_j) \triangleq \int_{\mathcal{R}^d} \mathcal{G}_h(x, x_i)\mathcal{G}_h(x, x_j)dx, \quad (20)$$

G_h is the $N \times N$ matrix whose (i,j)-th entry is $\mathcal{G}_h(x_i, x_j)$, and 1_N is the $N \times 1$ vector whose entries are all $1/N$. The kernel density estimator resultant from the above minimization is a sparse representation in the weighting coefficients [11], thus called the reduced set density estimator or the RSDE.

3. Main Results

The D-SVDD is a generalization of the SVDD that can assign different significance to each data point as mentioned above, and was shown to be able to yield better performance than the conventional SVDD [6]. In this section, we improve the D-SVDD in two respects: First, we note that the D-SVDD utilizes the notion of the local density degree which is computed based on the K-nearest neighborhood, which requires the full reference set for testing and is in practice very expensive. Thus, we propose to use a new local density degree based on the RSDE, which is of a sparse representation in the weighting coefficients, instead of the K-nearest neighborhood. An obvious advantage in using the RSDE instead is that it does not need the full reference set for testing. A possible inconvenience in obtaining the RSDE via (19) is that there is no explicit mechanism to control the degree of sparseness. To endow this mechanism, here we slightly modify the original formulation for the RSDE by utilizing the pre-specified constant $\nu \in (0,1)$ in the following way:

$$\begin{array}{ll} \min & \gamma^T C_h \gamma/2 - \gamma^T G_h 1_N \\ \text{s. t.} & \sum_{j=1}^N \gamma_j = 1, \ \gamma_i \in [0, 1/N\nu], \quad i = 1, \cdots, N. \end{array} \quad (21)$$

The use of ν in (21) is the strategy borrowed from the so-called nu SVMs [1]. Note that from

$$\begin{array}{l} 1 = \gamma_1 + \cdots + \gamma_N \\ \leq (\text{Number of the non-zero } \gamma_i) \times \frac{1}{N\nu} \end{array} \quad (22)$$

and

$$\begin{array}{l} (\text{Number of the } \gamma_i \text{ having the value } 1/N\nu) \times \frac{1}{N\nu} \\ \leq \gamma_1 + \cdots + \gamma_N = 1, \end{array} \quad (23)$$

we have

$$\nu \leq \frac{(\text{Number of the non-zero } \gamma_i)}{N}, \quad (24)$$

and

$$\frac{(\text{Number of the } \gamma_i \text{ having the value } 1/N\nu)}{N} \leq \nu. \quad (25)$$

Hence, one can see that it pre-specifies a lower bound for the ratio of the strictly positive γ_i and simultaneously an upper bound for the ratio of the γ_i reaching $1/N\nu$ with the value of ν. In this paper, we propose to use a new local density degree function based on the RSDE obtained by (21). More specifically, the local density degree function $\rho : \mathcal{R}^d \to [0, \infty)$ is now defined for each training data point $x_i \in D$ as follows:

$$\rho(x_i) = \exp(\omega \times \frac{\hat{p}_h(x_i;\gamma)}{\text{MEAN}_p})), \ i = 1, \cdots, N, \quad (26)$$

where $\hat{p}_h(x_i;\gamma)$ is the value of the RSDE (16) at x_i, MEAN_p is the mean of the $\hat{p}_h$ values over all training data, *i.e.*,

$$\text{MEAN}_p \triangleq \frac{1}{N}\sum_{i=1}^N \hat{p}_h(x_i;\gamma), \quad (27)$$

and $\omega \in [0,1]$ is the parameter controlling the sharpness of ρ. Note that in [14], the ρ was defined similarly utilizing the parzen-window-based $\hat{p}_h$. An advantage of the use of (26) based on the RSDE is that it requires only a partial reference set for testing according to the sparseness of the RSDE, while the use of the K-nearest neighborhood or the parzen window for computing the local density degree as in [6] or [14] requires the full reference set for testing, which is very expensive in practice.

As the next issue, we here observe that the dual representation of the D-SVDD (12) which has a nonlinear objective with linear constraints can be converted into the SDP form, which can be efficiently solved by interior point methods [12]. Note that in [6], it was simply noted that the dual representation (12) is a linearly constrained optimization problem without much details on how to solve it. In order to convert (12) into an SDP, we introduce an additional variable β satisfying

$$\begin{array}{l} \frac{1}{\sum_{i=1}^N \alpha_i \rho(x_i)} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \rho(x_i)\rho(x_j)k(x_i,x_j) \\ \quad - \sum_{i=1}^N \alpha_i \rho(x_i) < \beta. \end{array} \quad (28)$$

Note that to minimize the β satisfying (28) is equivalent to minimize the original objective function of (12). Also, note that from the standard results on the Schur complement [12], the inequality (28) is equivalent to the following LMI:

$$\begin{bmatrix} \beta + \sum_{i=1}^N \alpha_i \rho(x_i) & \alpha^T K_\rho^{1/2} \\ K_\rho^{1/2}\alpha & \sum_{i=1}^N \alpha_i \rho(x_i) I \end{bmatrix} > 0, \quad (29)$$

where α is the $N \times 1$ vector consisting of the α_i, I is the $N \times N$ identity matrix, and $K_\rho^{1/2}$ is the square-root of the $N \times N$ matrix K_ρ defined by $K_\rho(i,j) = \rho(x_i)\rho(x_j)k(x_i,x_j)$. As a result of the above conversion processes, the problem of finding the dual variable vector α in (12) can now be reformulated as the following SDP:

$$\min \quad \beta$$
$$\text{s.t.} \quad \begin{cases} \begin{bmatrix} \beta + \sum_{i=1}^{N} \alpha_i \rho(x_i) & \alpha^T K_\rho^{1/2} \\ K_\rho^{1/2} \alpha & \sum_{i=1}^{N} \alpha_i \rho(x_i) I \end{bmatrix} > 0 \\ \sum_{i=1}^{N} \alpha_i = 1 \\ 0 \le \alpha_i \le C, \quad i = 1, \cdots, N \end{cases} \tag{30}$$

Here, note that with simple change of variables, the equality condition $\sum_{i=1}^{N} \alpha_i = 1$ can be easily eliminated from the constraint part of the above problem, which transforms the above into the canonical form (14). The steps for eliminating equalities from the above constraints are well explained in [12]. Mathematical formulation in the form of the SDP is of great practical value because they can be solved by reliable and efficient convex optimization techniques [12], *e.g.*, the LMI Control Toolbox for use with Matlab [13]. Note that all experimental results reported in this paper are obtained utilizing the built-in function "mincx" of the LMI Control Toolbox.

To illustrate the proposed method and compare its performance with the conventional SVDD and D-SVDD, simulations were performed on the wine recognition data from the UCI KDD archive [15]. The wine recognition data consist of 178 data points belonging to three classes, and each data point is represented by 13 attribute values. In each simulation, the false negative rate was estimated from ten independent runs of two-fold cross-validation for the chosen normal class, and the data in other classes were used as the testing data for the outlier class to evaluate the false positive rate. All the data were utilized after re-scaling so that the training data points should have the spread of the unit length along each attribute direction. The trade-off parameter C and the kernel width parameter σ were selected via the 10-fold cross-validation to yield the best performance for the SVDD, and the same parameters were also used for the D-SVDD of [6] and the version proposed in this paper. For the density width parameter h in the probability density (17) of the RSDE, $h = \sqrt{\text{trace}(\text{cov}(x_i))}$ was used as in [14]. The average error rates (*i.e.*, the means of the false positive rate and the false negative rate) computed from the simulations were summarized as follows (Here in each case, the best value of ω was found by exhaustive search over $\{0.1, 0.2, \cdots, 0.7\}$):

Normal class	1	2	3
SVDD	7.67	18.40	8.73
D-SVDD of [6]	2.77	16.67	3.81
Proposed method(nu=0.5)	1.59	15.18	2.89

From comparing the error rates, one can see that the proposed D-SVDD with $\nu = 0.5$ yields significantly better results than the SVDD, and shows prediction accuracies comparable to the D-SVDD of [6] even with smaller reference data sets.

4. Concluding Remarks

In this paper, we addressed the problem of improving the D-SVDD method, which was recently proposed incorporating the idea that the data in high density region are more significant than those in a lower density region in describing a target class data set. Two issues were addressed for the improvement here. First, we observed that the decision function of the conventional D-SVDD is dependent on the use of the K-nearest neighborhood, which is expensive owing to reference to full data set, and proposed a new way of computing local density degree based on the RSDE, by which one can compute the local density degrees less expensively with only a partial reference set. Next, we observed that the dual representation of the conventional D-SVDD and the proposed version of this paper can both be solved utilizing the SDP formulation. This observation can provide an efficient way to solve the one class problems reliably and efficiently, and indeed, we conveniently used the MATLAB LMI Control Toolbox in obtaining all the experimental results reported in this paper. Simulations performed for the wine recognition data showed that the proposed method can yield better results than the SVDD, and comparable results with the conventional D-SVDD even with smaller reference data sets. Further investigations yet to be done include extensive comparative studies, which will reveal the strengths and weaknesses of the proposed method, and in-depth analysis on why the D-SVDD can yield better performance than the conventional SVDD.

References

[1] Schölkopf, B., Smola, A.J.,: Learning with Kernels. MIT Press, Cambridge, MA (2002)

[2] Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge, UK (2004)

[3] Tax, D.M.J., Duin, R.P.W.: Support vector domain description. Pattern Recognition Letters **20** (1999) 1191–1199

[4] Tax, D.M.J., Duin, R.P.W.: Support vector data description. Machine Learning **54** (2004) 45–66

[5] Tax, D.M.J.: One-Class Classification. Ph.D. Thesis, Delft University of Technology (2001)

[6] Lee, K, Kim, D.-W., Lee, D., Lee, K.H.: Improving support vector data description using local density degree. Pattern Recognition **38** (2005) 1768–1771

[7] Rätsch, G., Mika, S., Schölkopf, B., Müller, K.-R.: Constructing boosting algorithms from SVMs: An application to one-class classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 1184–1199

[8] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., Williamson, R. C.: Estimating the support of a high-dimensional distribution. Neural Computation **13** (2001) 1443–1471

[9] Lanckriet, G. R. G., El Ghaoui, L., Jordan, M. I.: Robust novelty detection with single-class MPM. Advances in Neural Information Processing Systems **15** (2003) 905–912

[10] Campbell, C., Bennett, K. P.: A linear programming approach to novelty detection. Advances in Neural Information Processing Systems **13** (2001) 395–401

[11] Girolami, M., He, C.: Probabilistic density estimation from optimally condensed data samples. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 1253–1264

[12] Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V.: Linear Matrix Inequalities in Systems and Control Theory. SIAM, Philadelphia, PA (1994)

[13] Gahinet, P., Nemirovski, A., Laub, A.J., Chilali, M.: LMI Control Toolbox. MathWorks Inc., Natick, MA (1995)

[14] Lee, K, Lee, D., Lee, K.H.: A new data description method based on the density-induced support vector information. In: Proceedings of the 6th International Symposium on Advanced Intelligent Systems (2005) 616–621

[15] Blake, C. L., Merz, C. J., UCI Repository of Machine Learning Database, www.ics.uci.edu/~mlearn/MLRepository.html, 1998

Jooyoung Park
Professor, Dept. of Control & Instrumentation Engineering, Korea University
Research Area: Neural networks, Learning theory, Nonlinear systems
E-mail : parkj@korea.ac.kr

Daesung Kang
Graduate Student, Dept. of Control & Instrumentation Engineering, Korea University
Research Area: Support vector machines, Reinforcement learning, Robotics
E-mail : mpkds@korea.ac.kr