International Journal of Fuzzy Logic and Intelligent Systems, vol. 3, no. 2, December 2003 pp. 227-232

Semiparametric Kernel Fisher Discriminant Approach for Regression Problems

Jooyoung Park, Won-Hee Cho, and Young-Il Kim

Department of Control and Instrumentation Engineering, Korea University

Abstract

Recently, support vector learning attracts an enormous amount of interest in the areas of function approximation, pattern classification, and novelty detection. One of the main reasons for the success of the support vector machines(SVMs) seems to be the availability of global and sparse solutions. Among the approaches sharing the same reasons for success and exhibiting a similarly good performance, we have KFD(kernel Fisher discriminant) approach. In this paper, we consider the problem of function approximation utilizing both predetermined basis functions and the KFD approach for regression. After reviewing support vector regression, semi-parametric approach for including predetermined basis functions, and the KFD regression, this paper presents an extension of the conventional KFD approach for regression toward the direction that can utilize predetermined basis functions. The applicability of the presented method is illustrated via a regression example.

Key words : Function approximation, KFD regression, basis functions

I. Introduction

Recently, the theory and practice about the SVMs(support vector machines) have been well studied[1-2], and as a result, support vector learning has become one of the most important tools in the area of intelligent systems. In general, support vector machines have been derived utilizing the concepts such as the maximal margin[3], regularization-based function approximation on the RKHS(reproducing kernel Hilbert space)[4], or the Bayesian approach[5]. Due to their inherent advantages such as good generalization capabilities, SVMs now enjoy strong popularity in the fields of pattern recognition[6], function approximation[7], and novelty detection[8-10].

Support vector learning can be used to train MLPs(multi-layer perceptrons) with single layer of hidden nodes and RBFNs(radial basis function networks). These networks, when trained by the support vector learning, have the following features:

(1) Unlike conventional training algorithms such as the error back-propagation method, the number of hidden nodes of the networks can be determined automatically in the process of learning.

(2) They do not suffer from the problem of convergence to local minimum, thus can yield globally optimal solutions for the given training data.

In particular, the problem that each different initial condition ultimately converges to a different solution is not an issue in the area of support vector learning.

(3) The solutions obtained via support vector learning can have a certain level of generalization capabilities which can be explained with the help of statistical learning theory.

In addition to the above, another of the main reasons for the success of the support vector machines seems to be the availability of a sparse solution. Among the approaches sharing the same reasons for success and exhibiting a similarly good performance as SVMs, we have KFD(kernel Fisher discriminant) approach[11-14]. In this paper, we consider the problem of function approximation utilizing both predetermined basis functions and the KFD approach for regression. After reviewing support vector regression, semi-parametric approach[15] for including predetermined basis functions, and the KFD regression, this paper presents an extension of the conventional KFD approach for regression toward the direction that can utilize predetermined basis functions. Moreover, the applicability of the presented method is illustrated via a regression example.

The remaining part of this paper is organized as follows: In Section II, we review the problems of support vector regression and kernel Fisher discriminant regression. In Section III, we present our main result on the KFD regression with predetermined basis functions included. Section IV considers an example to show the applicability of the presented method. Finally, in Section V, concluding remarks are given together with possible topics for future works.

II. Preliminaries

2.1. Epsilon support vector regression

The problem of function approximation utilizing support vector learning is often called SVR(support vector regression)[7]. In the following, we briefly review the so-called epsilon SVR method.

Given the set of training data $\{(x_i, y_i)\}_{i=1}^{m}$, the objective of the epsilon SVR is to find a smooth approximator

$$f(x) = \langle w, \psi(x) \rangle + b$$

which minimizes the total sum of epsilon-insensitive errors[7]

$$|y_i - f(x_i)|_\varepsilon \triangleq \max\{|y_i - f(x_i)| - \varepsilon, 0\}.$$

Note that $\phi(x)$ is the feature vector in the feature space F, which is in general of much higher dimensional than the input space for x. The objective of the epsilon SVR can be achieved by solving the following optimization problem[2]:

$$\min \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \tag{1}$$

$$\begin{aligned} s.t. \quad & y_i - (\langle w, \phi(x_i)\rangle + b) \le \varepsilon + \xi_i \\ & (\langle w, \phi(x_i)\rangle + b) - y_i \le \varepsilon + \xi_i^* \\ & \xi_i, \ \xi_i^* \ge 0, \ \forall i \end{aligned}$$

Here, C is a positive trade-off constant which determines relative importance of the error values ξ_i, ξ_i^* compared to $\|w\|^2/2$ which may indicate the smoothness of approximating function f.

The Lagrange function associated with the above optimization problem is as follows:

$$\begin{aligned} L = & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_{i^*}) \\ & - \sum_{i=1}^{m}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, \phi(x_i)\rangle + b) \\ & - \sum_{i=1}^{m}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, \phi(x_i)\rangle - b) \\ & - \sum_{i=1}^{m}(\eta_i\xi_i + \eta_i^*\xi_i^*) \end{aligned}$$

(Here, we have constraints $\alpha_i^{(*)}, \eta_i^{(*)} \ge 0$)

Note that w, b, ξ_i, ξ_i^* are primal variables, while the $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ are dual variables which are introduced as Lagrange multipliers. Since the optimal solution of problem (2) is the saddle point in the augmented space consisting of coordinates for both primal and dual variables[1], the following should hold at the optimum..

$$\frac{\partial L}{\partial w} = 0 \Leftrightarrow w - \sum_{i=1}^{m}\alpha_i\phi(x_i) + \sum_{i=1}^{m}\alpha_i^*\phi(x_i) = 0 \tag{3}$$

$$\therefore \ w = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i)$$

$$\frac{\partial L}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = 0 \Leftrightarrow \alpha_i^{(*)} + \eta_i^{(*)} = C$$

$$\therefore \ \alpha_i^{(*)} \in [0, C], \ \forall i$$

(Here, by the notation (*), we mean that the cases with and without * both hold.)

By plugging the results of (3) into the Lagrange function L, we obtain the following dual problem:

$$\begin{aligned} \max \ D = & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle \phi(x_i), \phi(x_j)\rangle \\ & + \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^{m}(\alpha_i + \alpha_i^*)\varepsilon \end{aligned} \tag{4}$$

$$\begin{aligned} s.t. \quad & \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i^{(*)} \in [0, C], \ \forall i \end{aligned}$$

Here, if the feature map $\phi: R^n \to F$ is chosen to be the one corresponding to the widely used gaussian radial basis function kernel, then the following kernel trick[2] holds:

$$\langle \phi(x), \phi(y)\rangle = k(x, y) = \exp(-\|x - y\|^2/2\sigma^2)$$

Also, it follows that the dual problem (4) is equivalent to the following:

$$\begin{aligned} \max D = & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ & + \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^{m}(\alpha_i + \alpha_i^*)\varepsilon \end{aligned} \tag{5}$$

$$\begin{aligned} s.t. \quad & \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i^{(*)} \in [0, C], \ \forall i \end{aligned}$$

Note that problem (5) is in the form of QP(quadratic programming). By solving (5), the optimal α_i and α_i^* can be obtained.. Moreover, the optimal value of bias term b is obtained by utilizing the Kuhn-Tucker condition[1]. Then, based on the equation

$$w = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i),$$

the approximating function f can be represented as follows:

$$f(x) = \langle w, \phi(x)\rangle + b = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{6}$$

2.2. Kernel Fisher discriminant approach

For the clarity and convenience of description, we first consider the binary classification problem[11]. Let $\{x_i \mid i = 1, \cdots, m\}$ be training data set, and $y \in \{-1, 1\}^m$ be the vector of corresponding labels. Also, define $1 \in R^m$ be the vector of all ones, $1_1, 1_2 \in R^m$ as binary (0, 1) vectors corresponding to the class labels. Furthermore, let I, I_1, and I_2 be index sets, respectively and we denote the cardinality of each I_i as l_i (i.e., $l_i = |I_i|$). As is well-known, in the linear case, Fisher's discriminant is computed by maximizing

$$J(w) = (w^T S_B w) / (w^T S_W w),$$

where $S_B = (m_2 - m_1)(m_2 - m_1)^T$,

$$S_W = \sum_{k=1}^{2}\sum_{i \in I_k}(x_i - m_k)(x_i - m_k)^T,$$

and m_k is the sample mean for class k.

As was shown in [14], with an expansion for w in feature space $w = \sum_{i \in I}\alpha_i\phi(x_i)$, the kernel version of the Fisher discriminant problem can be reduced to minimize the following:

$$J(\alpha) = \frac{(\alpha^T\mu)^2}{\alpha^T N\alpha} = \frac{\alpha^T M\alpha}{\alpha^T N\alpha},$$

where $\mu_i = \frac{1}{l_i}K1_i, \ i = 1, 2,$

$$N = KK^T - \sum_{i=1}^{2} l_i\mu_i\mu_i^T, \quad \mu = \mu_2 - \mu_1,$$

$M=\mu\mu^T$, and $K_{ij}=\langle \phi(x_i), \phi(x_j)\rangle = k(x_i, x_j)$, $i=1,\cdots, m$, $j=1,\cdots,m$.

In [11], it was shown that how the kernel Fisher discriminant problem can be cast as the following convex quadratic programming problem:

$$\min_{\alpha}\ \alpha^T N\alpha + CP(\alpha) \qquad (7)$$
$$\text{s.t. } \alpha^T(\mu_2-\mu_1)=2,$$

where P is a regularization operator, and C is a regularization constant. Furthermore, in Proposition 1 of [11], it was shown that for given $C>0$, any optimal solution α to the optimization problem (7) is also optimal for the following quadratic program and vice versa:

$$\min_{\alpha, b, \xi}\ \|\xi\|^2 + CP(\alpha)$$
$$\text{s.t. } K\alpha + b = y+\xi$$
$$1_i^T\xi = 0 \text{ for } i=1,2.$$

where $\alpha, \xi\in R^m$, $b\in R$, $C>0$.

One of the best parts of the above formulation is that it is now straightforward to obtain the kernel Fisher discriminant approach for regression. As was described in [11], instead of ±1 outputs y, we now have real-valued y's. Also, instead of two classes, there is only one class left. This recipe yields the following formulation for the KFD regression: Given the training data $\{(x_i, y_i)\in R^r\times R \mid i=1,\cdots,m\}$, the function $f(x)=\sum_{i=1}^m \alpha_i k(x_i, x)+b$ which is an optimal approximator in the sense of kernel Fisher discriminant can be found by solving the following quadratic programming problem:

$$\min_{\alpha, b, \xi}\ \|\xi\|^2 + CP(\alpha) \qquad (8)$$
$$\text{s.t. } K\alpha + 1b = y+\xi,$$
$$1^T\xi = 0,$$

where $\alpha, \xi\in R^m$, $b\in R$, $C>0$, $K_{ij}=k(x_i, x_j)$, $i=1,\cdots,m$, $j, 1, \cdots, m$.

Here, C is a regularization constant that will be chosen by the user, and $P(\alpha)$ is the operator for regularization. Throughout this paper, we consider the l_1 norm for the regularization operator, i.e.,

$$P(\alpha)=\|\alpha\|_1 = |\alpha_1|+\cdots+|\alpha_m|$$

Note that the l_1 norm is one of the most popular choices for the regularization operator in the related studies. Also note that this regularization operator plays a central role in yielding a sparse solution. Based on the obvious connection between (8) and support vector regression, we will mean, by the support vectors, the vectors x_i whose coefficient α_i is non-zero in the final solution $f(x)=\sum_{i=1}^m \alpha_i k(x_i, x)+b$.

III. KFD regression with predetermined basis functions included

The main objective of this paper is to find an approximator consisting of two parts: the first part is the usual neural approximator such as RBFN or MLP with a single hidden layer, while the second part is a linear combination of independent basis functions $\{\phi_1(\cdot), \ldots, \phi_n(\cdot)\}$ which are provided based on the domain knowledge. If we use the support vector learning, the structure of the above kind of approximators can be represented in the following form[15]:

$$f(x)=\langle w, \phi(x)\rangle + \sum_{i=1}^n \beta_i\phi_i(x) \qquad (9)$$

For fitting the given data

$$\{(x_i, y_i)\in R^r\times R \mid i=1,\cdots,m\}$$

with functions in the form of (9), the authors of [15] studied about a semi-parametric approach that generalizes the method of the epsilon SVR method. Since the KFD regression and the epsilon support vector regression share many similarities, we take the strategy of first reviewing the semi-parametric approach of [15] and then following similar steps to establish a semi-parametric approach for KFD regression. The mathematical formulation of [15] for approximating the training data with function (9) is as follows:

$$\min\ \frac{1}{2}\|w\|^2 + C\sum_{i=m}^{l}(\xi_i+\xi_i^*) \qquad (10)$$
$$s.t.\ \langle w, \phi(x_i)\rangle + \sum_{j=1}^n \beta_j\phi_j(x_i) - y_i \le \varepsilon + \xi_i^*$$
$$y_i - \langle w, \phi(x_i)\rangle - \sum_{j=1}^n \beta_j\phi_j(x_i) \le \varepsilon + \xi_i$$
$$\xi_i, \xi_i^* \ge 0,\ \forall i$$

Based on the above objective function and constraints, the corresponding Lagrange function has the following form:

$$\begin{aligned} L =\ & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^m(\xi_i+\xi_i^*) \\ &+ \sum_{i=1}^m \alpha_i\Big\{y_i - \langle w, \phi(x_i)\rangle - \sum_{j=1}^n\beta_j\phi_j(x_i) - \varepsilon - \xi_i\Big\} \\ &+ \sum_{i=1}^m \alpha_i^*\Big\{\langle w, \phi(x_i)\rangle + \sum_{j=1}^n\beta_j\phi_j(x_i) - y_i - \varepsilon - \xi_i^*\Big\} \\ &+ \sum_{i=1}^m \eta_i(-\xi_i) + \sum_{i=1}^m \eta_i^*(-\xi_i^*) = 0 \end{aligned} \qquad (11)$$

$$(\text{Here, } \alpha_i^{(*)}, \eta_i^{(*)} \ge 0)$$

By utilizing the saddle point condition, one can eliminate the primal variables, which yields the following dual problem:

$$\begin{aligned} \max\ & -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)k(x_i, x_j) \\ & -\varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i-\alpha_i^*) \end{aligned} \qquad (12)$$
$$s.t.\ \sum_{i=1}^m(\alpha_i-\alpha_i^*)\phi_j = 0,\ \forall j$$
$$\alpha_i, \alpha_i^* \in [0, C],\ \forall i$$

The solution α_i and α_i^* of (12), together with the β_j

which are obtained by applying the Kuhn-Tucker condition to support vectors, provides the coefficients of the following approximator:

$$f(x) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x_i, x) + \sum_{j=1}^{n}\beta_j\phi_j(x) \quad (13)$$

Now based on the above review, we choose (13) as the mathematical form of an approximator for the semi-parametric KFD regression. Note that in this equation, each entry of α_i and α_i^* should be non-negative. Then, the remaining problem is how to find the coefficients α_i, α_i^*, and β_i in the context of the semi-parametric KFD approach. Considering the optimization problem (8) for the KFD regression together with the formulation (10) for the semi-parametric support vector regression, we can come up with the following mathematical formulation for the semi-parametric KFD regression:

$$\min_{\alpha,\alpha^*,\beta,\xi} \ \|\xi\|^2 + C\|\alpha - \alpha^*\|_1 \quad (14)$$
$$\text{s.t. } K(\alpha - \alpha^*) + \Phi\beta = y + \xi,$$
$$1^T\xi = 0,$$

where $\alpha, \alpha^*, \xi \in R^m$, $\beta \in R^n$, $C > 0$,
$K_{ij} = k(x_i, x_j)$, $i = 1, \cdots, m$, $j, 1, \cdots, m$,

$$\Phi = \begin{bmatrix} \phi_1(x_1) & \cdots & \phi_n(x_1) \\ \vdots & \vdots & \vdots \\ \phi_1(x_m) & \cdots & \phi_n(x_m) \end{bmatrix} \in R^{m\times n},$$

and $\alpha_i^{(*)} \geq 0$ for $\forall i \in \{1, \cdots, m\}$.
Here note that minimizing the l_1 norm of $\alpha - \alpha^*$ under the constraints $\alpha_i^{(*)} \geq 0$ for $\forall i \in \{1, \cdots, m\}$ leads to the condition

$$\alpha_i\alpha_i^* = 0 \ \text{for} \ \forall \ i \in \{1, \cdots, m\}. \quad (15)$$

Here, we note that under the condition (15), we can conclude that the problem (14) is equivalent to the following quadratic programming problem:

$$\min_{\alpha,\alpha^*,\beta,\xi} \ \sum_i \xi_i^2 + C\left(\sum_i \alpha_i + \sum_i \alpha_i^*\right) \quad (16)$$
$$\text{s.t. } \begin{bmatrix} K & -K & -I & \Phi \\ 0 & 0 & 1^T & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ \alpha^* \\ \xi \\ \beta \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix},$$
$$\begin{bmatrix} -I & 0 & 0 & 0 \\ 0 & -I & 0 & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ \alpha^* \\ \xi \\ \beta \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Finally, note that the KFD regression problem (8) with l_1 regularizer is essentially a special case of the semi-parametric KFD regression problems formulated above.

IV. A simulation example

To evaluate the applicability of the proposed method, we considered an approximation problem for the data generated from the following[15]:

$$f(x) = \sin x + sinc(2\pi(x-5))$$

In Fig. 1, we showed the above function together with each component. For training data, we first considered samples $\{(x_i, y_i) \mid y_i = f(x_i)\}$, where the x_i are the equally spaced 100 samples on the interval [-0,1]. As the set of the independent basis functions, we chose $\{\sin x, \cos x, 1\}$. This means that at the outset of curve-fitting, we had not only the training data but also the additional information that the sinusoidal functions $\sin x$, $\cos x$, and/or the constant functions could play an important role in describing the given data. For the neural approximator part, we used the gaussian radial basis functions with the width $\sigma = 0.25$. The simulation result for this noise-free case was shown in Fig. 2. The trade-off constant C in (16) was chosen as 0.01 in this case. As a different data set, we next considered samples $\{(x_i, y_i) \mid y_i = f(x_i) + \xi_i\}$, where ξ_i is the gaussian noise with zero mean and variance 0.2. With the same set of parameters σ and C applied to the second set of data, we obtained the simulation result shown in Fig. 3. Finally, we observed how the trade-off constant C was related with the

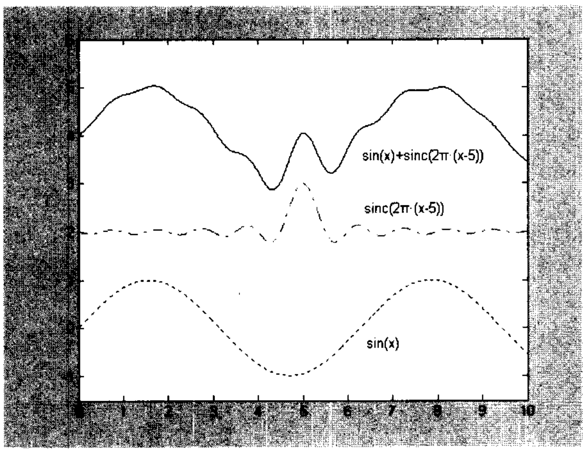


Fig. 1. Functions considered in the example

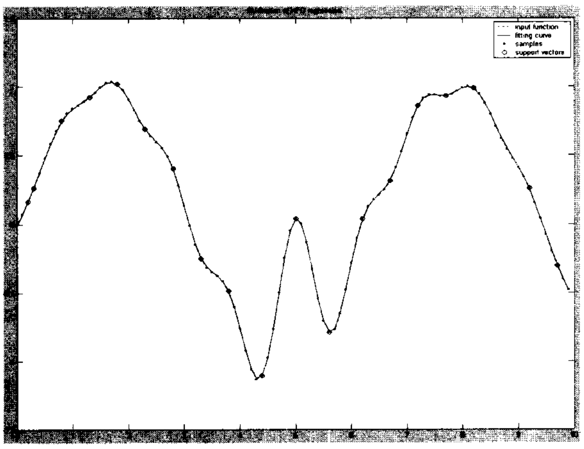


Fig. 2. The curve-fitting result for the data generated without noise

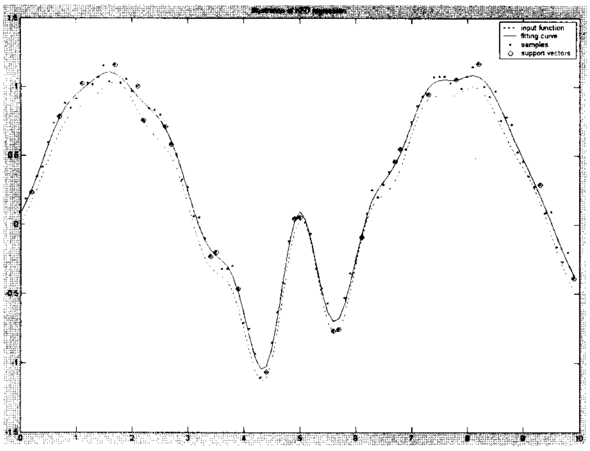


Fig. 3. The curve-fitting result for the data generated with noise

Table 1. A summary on the relation between the trade-off constant C and the number of support vectors

Value of the trade-off constant C	Number of support vectors	
	Without noise	With noise
0.001	26 (26.0%)	29 (29.0%)
0.01	19 (19.0%)	24 (24.0%)
0.1	11 (11.0%)	16 (16.0%)
1	3 (3.0%)	3 (3.0%)

number of support vectors for both set of training data, and results were summarized in the table below Fig. 3. The simulation results reported in this section seem to be reasonably well, and lead us to expectation that the proposed semi-parametric KFD method could be a good choice for a certain class of regression problems.

V. Concluding remarks

In this paper, we first reviewed the support vector regression, KFD approach for regression, and the semi-parametric support vector regression, then proposed a quadratic programming-based method for the semi-parametric KFD regression. We expect that the proposed method could be a promising approximation tool when hints on what kind of independent basis functions are probable are given in addition to the training data. The future works that need to be done to improve completeness of the present work include the comparative studies based on a wide range of simulation examples and the refinement of the theoretical aspects of the proposed method.

References

[1] N. Cristianini and J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, Cambridge University Press, 2000.

[2] B. Schölkopf and A. J. Smola, Learning with kernels, MIT Press, 2002.

[3] K.-R. Müller, 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.

[4] F. Girosi, "An equivalence between sparse approximation and support vector machines," Neural Computation, vol. 10, no. 6, pp. 1455-1480, 1998.

[5] J. T. Kwok, "The evidence framework applied to support vector machines," IEEE Transactions on Neural Networks, vol. 11, no. 5, pp. 1162-1173, 2000.

[6] C. Burges, "A tutorial on support vector machines for pattern recognition," Knowledge Discovery and Data Mining, vol. 2, no. 2, pp. 121-167, 1998.

[7] A. Smola and B. Schölkopf, "A tutorial on support vector regression," Neuro Colt Technical Report NC-TR-1998-030, Royal Holloway College, University of London, UK, 1998.

[8] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, and A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," Neural Computation, vol. 13, pp. 1443-1471, 2001.

[9] G. Rätsch, S. Mika, B. Schölkopf, and K.-R. Müller, "Constructing boosting algorithms from SVMs: An application to one-class classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 1-15, 2002.

[10] C. Campbell and K. P. Bennett, "A linear programming approach to novelty detection," Advances of NIPs 2000, pp. 395-401, 2000.

[11] S. Mika, G. Rätsch, and K.-R. Müller, "A mathematical programming approach to the kernel Fisher algorithm," In T.K. Leen, T.G. Dietterich, and V. Tresp, editors, Advances in Neural Information Processing Systems 13, pages 591-597. MIT Press, 2001.

[12] G. Baudat and F. Anouar, "Generalized discriminnt analysis using a kernel approach," Neural Computation, vol. 12, no. 10, pp. 2385-2404, 2000.

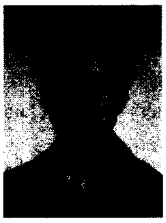
[13] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," In Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, editors, Neural Networks for Signal Processing IX, pp. 41-48, IEEE, 1999.

[14] S. Mika, G. Rätsch, J. Weston, and B. Schölkopf. "Invariant feature extraction and classification in kernel spaces," In S.A. Solla, T.K. Leen, and K.-R. Müller, editors, Advances in Neural Information Processing Systems 12, pp. 526-532, MIT Press, 2000.

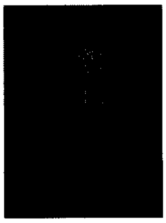
[15] A. Smola, T. Friess, and B. Schölkopf, "Semiparametric Support Vector and Linear Programming Machines," Nuero COLT Technical Report Series NC2-TR-1998-

024, August, 1998.

Jooyoung Park received his BS degree in electrical engineering from Seoul National University, Korea, in 1983, and his PhD degree in electrical and computer engineering from the University of Texas at Austin, USA, in 1992. He joined Korea University in 1993, where he is currently a professor in the Department of Control and Instrumentation Engineering. His recent reseach interests include SVMs and neural networks.



Won-Hee Cho received his BS degree in control and instrumentation engineering from Korea University, Korea, in 2002. He is now pursuing his Master degree in the Department of Control and Instrumentation Engineering, Korea University. His research interests include SVMs and neural networks.



Young-Il Kim received his BS degree in control and instrumentation engineering from Korea University, Korea, in 2002. He is now pursuing his Master degree in the Department of Control and Instrumentation Engineering, Korea University. His research interests include SVMs and fuzzy control.