

퍼지 원 클래스 서포트 벡터 머신[☆]

Fuzzy One Class Support Vector Machine

김 기 주*
KJoo Kim

최 영 식**
YoungSik Choi

요 약

OC-SVM(One Class Support Vector Machine)은 주어진 전체 데이터의 분포를 측정하는 대신에, 데이터 분포의 서포트(support)를 측정하는 기술로서 주어진 데이터를 가장 잘 설명할 수 있는 최적의 서포트 벡터(support vector)를 구하는 기술이다. OC-SVM은 데이터 분포의 표현에 아주 뛰어난 접근 방법이지만, 사람의 주관적인 중요도를 반영하는 것은 힘들다. 본 논문에서는 각 데이터에 퍼지 멤버십(fuzzy membership)을 적용하여 기존의 OC-SVM에 사용자의 주관적인 중요도를 표현할 수 있는 FOC-SVM(Fuzzy One Class Support Vector Machine)을 유도하였다.

FOC-SVM은 데이터들을 동등하게 다룬는 것이 아니라, 데이터 객체의 중요도에 따라 데이터를 다룬다. 즉, 덜 중요한 데이터의 특징 벡터는 OC-SVM의 처리과정에 덜 기여하도록 하기 위하여, 객체의 중요도에 따라 특징 벡터의 크기를 조정하였다. 이를 증명하기 위하여 가상의 데이터를 가지고 실험을 하였고, 실험 결과는 예측된 결과를 보여 주었다.

Abstract

OC-SVM(One Class Support Vector Machine) avoids solving a full density estimation problem, and instead focuses on a simpler task, estimating quantiles of a data distribution, i.e. its support. OC-SVM seeks to estimate regions where most of data resides and represents the regions as a function of the support vectors. Although OC-SVM is powerful method for data description, it is difficult to incorporate human subjective importance into its estimation process. In order to integrate the importance of each point into the OC-SVM process, we propose a fuzzy version of OC-SVM.

In FOC-SVM (Fuzzy One-Class Support Vector Machine), we do not equally treat data points and instead weight data points according to the importance measure of the corresponding objects. That is, we scale the kernel feature vector according to the importance measure of the object so that a kernel feature vector of a less important object should contribute less to the detection process of OC-SVM. We demonstrate the performance of our algorithm on several synthesized data sets. Experimental results showed the promising results.

☞ Keyword : One Class Support Vector Machine, SVM, Fuzzy One Class Support Vector Machine, Fuzzy One Class Support Vector Data Description

1. 서 론

주어진 데이터의 분포를 측정하는 것은 매우 어려운 문제이다. 이를 극복하기 위하여, OC-SVM(One Class Support Vector Machine)은 데이터

분포의 서포트(support)를 측정함으로써 문제를 단순화시킨다[1,6]. 즉, OC-SVM은 데이터가 존재하는 영역을 추출하며, 이 영역을 서포트 벡터(support vector)로써 표현한다. 표현된 영역 밖의 데이터 포인트들은 아웃라이어(outlier)로 간주된다. 또한, OC-SVM은 커널 함수 (kernel function)를 이용함으로써 데이터가 존재하는 복잡한 영역을 자연스럽게 표현할 수 있다. 즉, 비선형 영역을 처리하기 위하여 데이터 포인트들을 고차원 특징 공간(high dimensional feature space)인 커널 공간(kernel space)으로 사상(mapping) 시킨 후 커널 공간에 존

* 준 회 원 : 한국항공대학교 컴퓨터공학과 석사과정
zgeniez@hau.ac.kr(제 1저자)

** 정 회 원 : 한국항공대학교 전자정보통신공학부 조교수
choimail@hau.ac.kr(공동저자)

☆ 이 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구되었음 (KRF-2003-003-D00409)
[2003년/12/03 투고 심사 - 2005/01/31 심사완료]

재하는 데이터 영역을 측정한다. 본 논문에서는 커널 공간으로서 가우시안 커널 공간(Gaussian kernel space)을 사용한다.

OC-SVM은 “클러스터링(clustering)” [1], “novelty detection”[4,6,10], “내용기반 이미지 검색(image retrieval)” [5,9] 등과 같은 다양한 응용분야에 쓰이고 있다. [4]에서는 OC-SVM을 갑작스런 스팩트럼 변화 검출에 사용하였고, 기존의 알고리즘보다 우수한 방법이라는 것을 보였다. [6]에서는 미국 우편 서비스 데이터베이스로부터 필기체 숫자(handwritten digits)의 아웃라이어 검출에 OC-SVM을 사용하였다. [5]에서는 “사용자 의사반영 처리(relevance feedback process)”를 이용한 이미지 검색 시스템을 구축하였다.

OC-SVM의 결과로 얻어진 서포트 벡터는 전체 데이터 포인트들의 분포를 나타내는 표현 벡터(descriptive vector)로 해석할 수 있다. 즉, OC-SVM은 대부분의 데이터 포인트들이 존재하는 영역의 표면에 있는 벡터들을 선택하고, 선택되어진 서포트 벡터들을 가지고 영역을 표현한다. OC-SVM의 이러한 성질은 이미지 검색에서 사용되었다[5,9]. 즉, 서포트 벡터들이 데이터 분포의 윤곽을 묘사한다는 것이 OC-SVM의 주목할 만한 성질이다.

OC-SVM은 모든 데이터 포인트들을 동등하게 다룬다. 그러나 사전 지식으로부터 얻을 수 있는 각 데이터의 중요도 측정치에 따라 데이터 포인트들을 다루기 원하는 응용분야가 있을 수 있다. 본 논문에서는 OC-SVM에 각 데이터 포인트의 중요도에 따른 퍼지 멤버쉽(fuzzy membership)을 결합한 FOC-SVM (Fuzzy One Class Support Vector Machine)을 제안한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 먼저 표현이 다른 두 개의 OC-SVM의 이론과 성질에 대해 언급한다. 3장에서는 FOC-SVM을 유도하고 분석하고, 4장에서는 가상 데이터들을 가지고 실험 및 실험결과를 분석한다. 마지막으로 5장에서는 결론 및 향후 연구 과제를 제시한다.

2. One Class Support Vector Machine

OC-SVM은 데이터가 존재하는 영역을 추출하며, 그 영역을 서포트 벡터로 표현한다. 영역의 밖에 존재하는 데이터 포인트는 아웃라이어로 간주한다. 데이터 포인트를 고차원 특징 공간(high dimensional feature space)에서 커널 공간(kernel space)으로 매핑(mapping)하는 사상함수 ϕ 를 사용하여 비선형 영역을 처리 할 수 있다[2,3].

OC-SVM에는 데이터가 존재하는 영역을 초월구(hypersphere)로써 표현하는 것과 초월면(hyperplane)으로써 표현하는 것이 있다[6,8].

2.1 OC-SVM I : 구(hypersphere) 기반의 OC-SVM

구를 사용한 OC-SVM은 모든 데이터 포인트가 존재하는 영역을 중심점 a 와 반지름이 R 인 최소구로 표현한다. 이를 식으로 표현하면 다음과 같다[8].

$$\|\phi(x_i) - a\|^2 \leq R^2, \quad \forall i \quad (1)$$

식 (1)에 대한 최소 영역을 구하기 위해서 R^2 을 최소화 한다. 이를 목적함수로 표현하면 다음과 같다.

$$\text{minimize } R^2 + C \sum_i \xi_i, \quad (2)$$

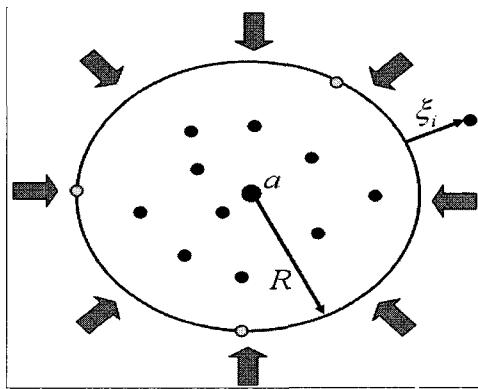
$$\text{subject to } \|\phi(x_i) - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \quad \forall i$$

여기서, C 는 상수, $C \sum_i \xi_i$ 는 penalty term,

$\|\cdot\|$ 는 유clidean 놈(Euclidean norm)이다.

식 (2)를 해결하기 위하여 라그랑지안 계수법(Lagrangian multiplier)을 적용한다.

$$L = R^2 + C \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i - \|\phi(x_i) - a\|^2) - \sum_i \gamma_i \xi_i, \quad \alpha_i \geq 0, \quad \gamma_i \geq 0 \quad (3)$$



(그림 1) 구를 사용한 OC-SVM

L 을 R , a , ξ_i 로 각각 미분하여 0 으로 놓으면 다음의 조건들이 나온다.

$$\sum_i \alpha_i = 1, \quad (4)$$

$$a = \sum_i \alpha_i \phi(x_i), \quad (5)$$

$$C - \alpha_i - \gamma_i = 0 \Rightarrow 0 \leq \alpha_i \leq C \quad (6)$$

식 (3)을 α_i 에 대해 다시 표현하면 다음과 같다.

$$\begin{aligned} & \text{maximize } W = \sum_i \alpha_i \phi(x_i) \cdot \phi(x_i) - \\ & \sum_i \sum_j \alpha_i \alpha_j \phi(x_i) \cdot \phi(x_j) \end{aligned} \quad (7)$$

$$\text{subject to } \sum_i \alpha_i = 1, \quad 0 \leq \alpha_i \leq C, \quad \forall i$$

식 (3)과 식 (7)은 KKT 충족조건(Karush-Kuhn-Tucker complementary condition)을 만족함으로써 듀얼 문제(dual problem)가 된다. KKT 충족 조건은 다음과 같다[3].

$$\gamma_i \xi_i = 0 \quad (8)$$

$$\alpha_i (R^2 + \xi_i - \|\phi(x_i) - a\|^2) = 0 \quad (9)$$

KKT 조건을 α_i 로 설명하면 1) $\alpha_i = 0$ 인 포인터 x_i 는 $\|\phi(x_i) - a\|^2 < R^2$ 의 영역에 존재 하고,

2) $0 < \alpha_i < C$ 인 포인터 x_i 는 $\|\phi(x_i) - a\|^2 = R^2$ 에 놓여 있다. 이때의 포인터 x_i 를 서포트 벡터라 한다 그리고, 3) $\alpha_i = C$ 인 포인터 x_i 는 $\|\phi(x_i) - a\|^2 > R^2$ 의 영역에 놓여 있게 되고, 이러한 데이터 포인터를 아웃라이어라 한다.

또한, 식 (7)에서 내적(dot product) $\phi(x_i) \cdot \phi(x_j)$ 은 가우시안 커널 함수(Gaussian Kernel function)를 사용하여 다시 표현 할 수 있다.

$$W = \sum_i \alpha_i \phi(x_i) \cdot \phi(x_i) - \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) \quad (10)$$

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (11)$$

σ 는 스케일(scale) 파라미터이다.

모든 데이터 포인트 x 에 대하여, 특징 공간(feature space)에서의 $\phi(x_i)$ 에서 구의 중심 a 까지의 거리 $f(x)$ 는 다음과 같다.

$$\begin{aligned} f(x) &= \|\phi(x) - a\|^2 \\ f(x) &= K(x, x) - 2 \sum_i \alpha_i K(x, x_i) + \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) \end{aligned} \quad (12)$$

구의 반지름 R 은 a 부터 서포트 벡터 x_s 까지의 거리이다. 즉, R^2 은 다음 식으로 구할 수 있다.

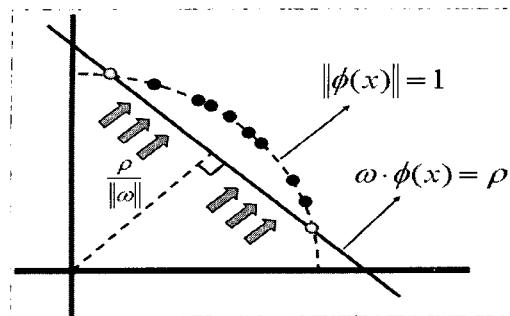
$$R^2 = f(x_s) \quad (13)$$

2.2 OC-SVMII : 평면(hyperplane) 기반의 OC-SVM

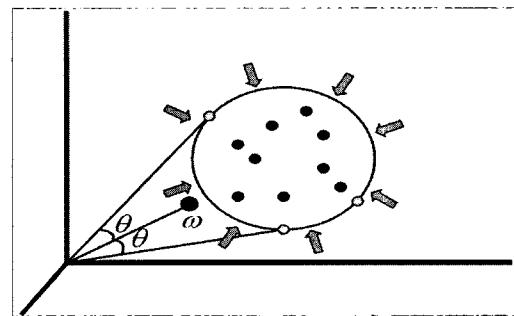
평면을 사용한 OC-SVM은 모든 데이터 포인트가 존재하는 영역을 평면 $\omega \cdot \phi(x) - \rho = 0$ 로 구분한다. 이를 식으로 표현하면 다음과 같다[6].

$$\omega \cdot \phi(x) \geq \rho, \quad \forall i \quad (14)$$

벡터의 내적은 커널 함수를 사용한다. 특히 가우시안 함수인 경우는 모든 데이터 포인트 x 에 대하여 $\|\phi(x)\| = 1$ 이다. 이것을 이용하여 $\omega \cdot \phi(x_i)$



〈그림 2〉 평면을 사용한 OC-SVM



〈그림 3〉 코사인 유사도로 해석한 OC-SVM

을 코사인 유사도(cosine similarity)로 표현 할 수 있다.

$$\begin{aligned} \omega \cdot \phi(x_i) &= \|\omega\| \cdot \|\phi(x_i)\| \cdot \cos\theta, \\ \cos\theta &= \frac{\omega \cdot \phi(x_i)}{\|\omega\|}, \\ \text{cosine similarity} &\propto \omega \cdot \phi(x_i) \end{aligned} \quad (15)$$

여기서, $\|\omega\|$ 는 모든 데이터 포인트에 대하여 동일하다. 즉, 데이터 포인트 $\phi(x)$ 와 ω 의 내적은 코사인 유사도를 의미한다. 식 14은 데이터 포인트에 대하여 중심점 ω 와의 코사인 유사도가 ρ 보다 큰 영역을 의미한다.

식 (14)로 표현한 영역을 최소화 하기 위해서는 $\|\omega\|$ 를 최소화 하고, ρ 를 최대화 해야 한다. 이를 목적함수로 표현하면 다음과 같다.

$$\text{minimize } \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i \quad (16)$$

subject to $\omega \cdot \phi(x) \geq \rho - \xi_i, \xi_i \geq 0, \forall i$

식 (16)을 풀기 위하여 라그랑지안 계수법을 적용한다.

$$\begin{aligned} L &= \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i - \rho - \\ &\sum_i \alpha_i (\omega \cdot \phi(x_i) - \rho + \xi_i) - \sum_i \gamma_i \xi_i, \\ \alpha_i &\geq 0, \gamma_i \geq 0 \end{aligned} \quad (17)$$

L 을 ω, ρ, ξ_i 로 각각 미분하여 0 으로 놓으면 다음의 조건들이 나온다.

$$\sum_i \alpha_i = 1, \quad (18)$$

$$\omega = \sum_i \alpha_i \phi(x_i), \quad (19)$$

$$C - \alpha_i - \gamma_i = 0 \Rightarrow 0 \leq \alpha_i \leq C \quad (20)$$

식 (17)를 α_i 에 대해 다시 표현하면 다음과 같다.

$$\text{maximize } W = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) \quad (21)$$

$$\text{subject to } \sum_i \alpha_i = 1, 0 \leq \alpha_i \leq C, \forall i$$

식 (16)과 (21)은 KKT 충족조건을 만족함으로써 듀얼 문제가 된다. KKT 충족 조건은 다음과 같다[3].

$$\gamma_i \xi_i = 0 \quad (22)$$

$$\alpha_i (\omega \cdot \phi(x_i) - \rho + \xi_i) = 0 \quad (23)$$

KKT 조건을 α_i 로 설명하면 1) $\alpha_i = 0$ 인 포인티 x_i 는 $\omega \cdot \phi(x) > \rho$ 의 영역에 놓여 있고, 2) $0 < \alpha_i < C$ 인 포인티 x_i 는 $\omega \cdot \phi(x) = \rho$ 에 놓여 있다. 이때의 포인티 x_i 를 서포트 벡트라 한다. 그리고, 3) $\alpha_i = C$ 인 포인티 x_i 는 $\omega \cdot \phi(x) < \rho$ 의 영역에 놓여 있게 되고, 이러한 데이터 포인티를 아웃라이어라 한다.

모든 데이터 포인트 x 에 대하여, 특징 공간에서 $\phi(x)$ 와 ω 의 유사도 $f(x)$ 는 다음과 같다.

$$f(x) = \omega \cdot \phi(x) \quad (24)$$

$$f(x) = \sum_i \alpha_i K(x, x_i)$$

ρ 는 서포트 벡터 x_s 으로 구할 수 있다.

$$\rho = f(x_s) \quad (25)$$

2.3 OC-SVM의 최적화(optimization)

식 (10)에 $K(x, x) = 1$ 과 $\sum_i \alpha_i = 1$ 을 적용 시켜 다시 쓰면 다음과 같다.

$$\text{minimize } W = \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) \quad (26)$$

또한, 식 (21)를 다시쓰면 식 (26)과 같다. 즉, 가우시안 커널을 사용한 두 OC-SVM은 목적함수와 KKT조건이 다를 뿐 같은 문제이다. 따라서, 이들에 대한 최적화 역시 같다. 식 (26)은 Quadratic programming을 이용하여 해를 구할 수 있다. 여기서는 SMO (Sequential Minimal Optimization)을 수정하여 이용한다[3,6,7].

SMO의 알고리즘은 1) 먼저 임의의 2개의 데이터 포인트를 선택하고, 2) $\sum_i \alpha_i = 1$ 과 $0 \leq \alpha_i \leq C$ 를 만족시키면서 선택된 2개의 데이터 포인트의 α_i 를 갱신한다. 3) 모든 데이터 포인트에 대해 KKT조건을 만족할 때 까지 1)과 2)를 반복한다 [3,6,7].

SMO의 알고리즘에서 선택되어진 데이터 포인트의 α 를 α_1^* 과 α_2^* 라고 가정하자. SMO는 $\alpha_1 + \alpha_2 = \alpha_1^* + \alpha_2^* = k$ 를 만족시키면서 α_1 와 α_2 를 갱신한다. 여기서 K 는 상수이고 $\alpha_1 + \alpha_2 = \alpha_1^* + \alpha_2^* = k$ 는 $\sum_i \alpha_i = 1$ 을 만족시키기 위한 조건이다. 두 개의 데이터 포인트에 대한 α 를 갱신한 후 SMO는 조건 $0 \leq \alpha_i \leq C$ 를 확인한다. 식 (26)에서 α_1 을 $k - \alpha_2$

로 치환한 후 α_2 에 대해 미분을 하여 0으로 놓음으로써 OC-SVM의 α 대한 갱신식(update equation)을 얻을 수 있다. 다음은 이를 간단화한 식이다.

$$\begin{aligned} \alpha_2 &= \alpha_2^* + \frac{\alpha_1^* - \alpha_2^*}{K_{11} + K_{22} - 2K_{12}} \\ \alpha_1 &= \alpha_1^* + \alpha_2^* - \alpha_2 \end{aligned} \quad (27)$$

여기서, $K_{ij} = K(x_i, x_j)$ 이다. 식 (27)에서 O_i 는 $\phi(x_i)$ 와 중심점 α (또는 ω)의 내적이다.

$$O_i = \sum_j \alpha_j K(x_i, x_j) \quad (28)$$

$\phi(x_i)$ 와 중심점 α (또는 ω)의 코사인 유사도를 O_i 으로 표현할 수 있다.

$$\text{sim}(\phi(x_i), a) = \frac{\phi(x_i) \cdot a}{\|\phi(x_i)\|^2 \cdot \|a\|^2} = \frac{O_i}{\|a\|^2} \quad (29)$$

여기서, $\|a\|^2$ 는 모든 데이터 포인트에 대해서 동일하다.

식 (27)에서 α_2 는 $O_1 - O_2$ 에 비례적이다. 즉, $\phi(x_2)$ 와 α (또는 ω)의 유사도가 $\phi(x_1)$ 와 α (또는 ω)의 유사도 보다 크면 α_2 는 감소하고, 반대로는 증가한다.

또한, 식 (27)의 $O_1 - O_2$ 를 중심점과의 거리에 대한 차로 다시 쓸 수 있다.

$$0.5(\|\phi(x_2) - a\|^2 - \|\phi(x_1) - a\|^2) = Q - Q \quad (30)$$

즉, 특징공간에서 데이터 포인트가 중심점에서 멀어지거나 중심점과의 유사도가 작아지면, 그 데이터에 대한 a 값은 높은 값을 가지게 되고, 서포트 벡터가 될 가능성이 커진다.

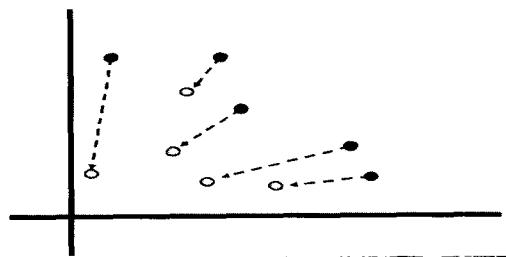
3. Fuzzy One Class Support Vector Machine

FOC-SVM(Fuzzy One Class Support Vector

Machine)은 사용자의 주관적인 중요도와 OC-SVM 을 결합한 것이다. OC-SVM의 기준은 거리(또는 유사도)이다. 여기에, 각 데이터 객체(object)에 대한 중요도 β 를 가지고 객체에 대한 거리(또는 유사도)의 크기를 조정한다.

3.1 FOC-SVM I : 구(hypersphere) 기반의 FOC-SVM

FOC-SVM I은 2.1절의 OC-SVM I에 퍼지 맴버쉽을 적용한다. 중요도 $\beta_i \in (0,1]$ 를 가지는 데이터 포인트의 집합을 $(x_1, \beta_1), (x_2, \beta_2), \dots, (x_n, \beta_n)$ 라고 가정하자. (x_i, β_i) 에서 x_i 는 객체를 표현하는 데이터 포인트이고, β_i 는 객체의 중요도를 나타낸다. 여기에서는 β_i 를 가지고 데이터 포인트의 원점에 대한 거리의 크기를 조절한다.



〈그림 4〉 원점과의 거리를 조정한 데이터 포인트들

다음은 β_i 를 적용한 목적함수이다.

$$\text{minimize } R^2 + C \sum_i \xi_i , \quad (31)$$

$$\text{subject to } \|\beta_i \phi(x_i) - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \quad \forall i$$

식 (31)을 풀기 위하여 라그랑지안 계수법을 적용한다.

$$\begin{aligned} L &= R^2 + C \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i) \\ &\quad - \|\beta_i \phi(x_i) - a\|^2 - \sum_i \gamma_i \xi_i , \end{aligned} \quad (32)$$

$$\alpha_i \geq 0, \quad \gamma_i \geq 0$$

L 을 R, a, ξ_i 로 각각 미분하여 0으로 놓으면 다음의 조건들이 나온다.

$$\sum_i \alpha_i = 1 , \quad (33)$$

$$a = \sum_i \alpha_i \beta_i \phi(x_i) , \quad (34)$$

$$C - \alpha_i - \gamma_i = 0 \Rightarrow 0 \leq \alpha_i \leq C \quad (35)$$

식 (32)을 α_1 에 대해 다시 표현하면 다음과 같다.

$$\begin{aligned} \text{maximize } W &= \sum_i \alpha_i \beta_i^2 K(x_i, x_i) - \\ &\quad \sum_i \sum_j \alpha_i \alpha_j \beta_i \beta_j K(x_i, x_j) \\ \text{subject to } \sum_i \alpha_i &= 1, \quad 0 \leq \alpha_i \leq C, \quad \forall i \end{aligned} \quad (36)$$

식 (32)과 (36)은 KKT 충족조건을 만족함으로써 듀얼 문제가 된다. KKT 충족 조건은 다음과 같다[3].

$$\gamma_i \xi_i = 0 \quad (37)$$

$$\alpha_i (R^2 + \xi_i - \|\phi(x_i) - a\|^2) = 0 \quad (38)$$

KKT 조건을 α_i 로 설명하면 1) $\alpha_i = 0$ 인 포인트 x_i 는 $\|\beta_i \phi(x_i) - a\|^2 < R^2$ 의 영역에 존재하고, 2) $0 < \alpha_i < C$ 인 포인트 x_i 는 $\|\beta_i \phi(x_i) - a\|^2 = R^2$ 에 놓여 있다. 이때의 포인트 x_i 를 서포트 벡터라 한다. 그리고, 3) $\alpha_i = C$ 인 포인트 x_i 는 $\|\beta_i \phi(x_i) - a\|^2 > R^2$ 의 영역에 놓여 있게 되고, 이러한 데이터 포인트를 아웃라이어라 한다.

모든 데이터 포인트 x 에 대하여, 특징공간에서의 $\phi(x_i)$ 에서 구의 중심 a 까지의 거리 $f(x, \beta)$ 는 다음과 같다.

$$f(x, \beta) = \|\beta \phi(x) - a\|^2 \quad (39)$$

$$f(x, \beta) = K(x, x) - 2 \sum_i \alpha_i \beta_i K(x, x_i) + \sum_i \sum_j \alpha_i \alpha_j \beta_i \beta_j K(x_i, x_j)$$

구의 반지름 R 은 a 부터 서포트 벡터 x_s 까지의 거리이다. 즉, R^2 은 다음 식으로 구할 수 있다.

$$R^2 = f(x_s, \beta_s) \quad (40)$$

최적화(Optimization)

식 (36)은 Quadratic programming을 이용하여 해를 구할 수 있다. 2.3절에서와 같이 SMO를 수정하여 이용한다[3,6,7]. FOC-SVM I에 SMO를 적용하여 a 에 대한 갱신 식을 구하면 다음과 같다.

$$\alpha_2 = \alpha_2^* + \frac{(S_1 - S_2) - 0.5(\beta_1^2 K_{11} - \beta_2^2 K_{22})}{\beta_1^2 K_{11} + \beta_2^2 K_{22} - 2\beta_1\beta_2 K_{12}} \quad (41)$$

$$\alpha_1 = \alpha_1^* + \alpha_2^* - \alpha_2$$

여기서, S_i 는 $\beta_i\phi(x_i)$ 와 중심점 a 의 내적이다.

$$S_i = \sum_j \alpha_j \beta_j \beta_j K(x_i, x_j) \quad (42)$$

식 (41)에서 $(S_1 - S_2) - 0.5(\beta_1^2 K_{11} - \beta_2^2 K_{22})$ 는 중심점과의 거리에 대한 차로 다시 쓸 수 있다.

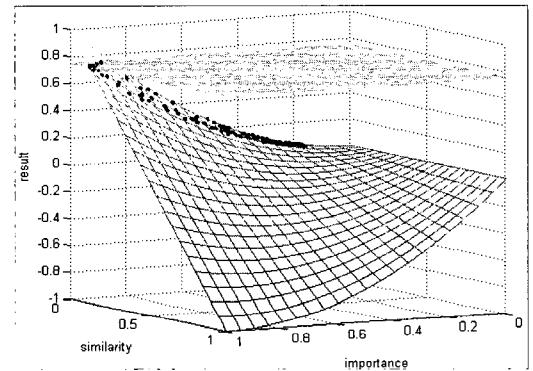
$$0.5(\|\beta_2\phi(x_2) - a\|^2 - \|\beta_1\phi(x_1) - a\|^2) = \\ (S_1 - S_2) - 0.5(\beta_1^2 K_{11} - \beta_2^2 K_{22}) \quad (43)$$

식 (41)과 식 (43)에서 FOC-SVM I은 $\beta_2\phi(x_2)$ 와 a 의 거리가 $\beta_1\phi(x_1)$ 와 a 의 거리 보다 크면 α_2 는 증가하고, 작으면 감소한다. 이것은 OC-SVM과 같다. 그러나, FOC-SVM I은 OC-SVM과 다르게 β 역시 갱신 식에 영향을 준다. 즉, FOC-SVM I에서 $\beta_i\phi(x_i)$ 이 중심점에서 멀고 β_i 이 큰 데이터 객체는 a 의 값이 높은 값을 가지게 되고, 서포트 벡터가 될 가능성이 커진다.

그림 5는 FOC-SVM I의 실행 결과 그래프이다. 그림 5에서의 곡면 그래프의 함수는 FOC-SVM I의 목적함수를 아래와 같이 다시 써서 표현된다.

$$R^2 - a^2 \geq \beta_i \times (\beta_i - 2\phi(x_i) \cdot a) \quad (44)$$

그림 5에서 ‘중요도(importance)’ β_i , ‘유사도(similarity)’는 $\phi(x_i) \cdot a$ 이다. ‘결과(result)’는 식 (44)의 우변의 값이다. 또한 상단 부분에 있는 평면은 $R^2 - a^2$ 에 대한 그래프이다. 즉, 모든 데이터 포인트들은 그림 5의 곡면 위에만 존재하고 곡면과 $R^2 - a^2$ 이 만나는 위치에 존재하는 데이터들은 서포트 벡터이고, $R^2 - a^2$ 위로 존재하는 데이터들은 아웃라이어이다.



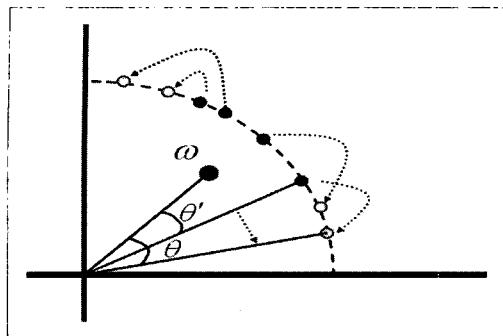
〈그림 5〉 FOC-SVM I의 실행 결과 그래프

FOC-SVM I은 주의 할 점이 있다. FOC-SVM I의 결과가 $R^2 \leq a^2$ 일 때, $\beta_i \leq \phi(x_i) \cdot a$ 인 데이터 객체는 반대로 β_i 가 작을수록 서포트 벡터가 될 가능성이 커진다.

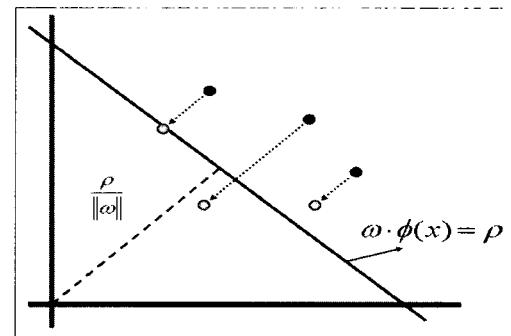
3.2 FOC-SVM II : 평면(hyperplane) 기반의 FOC-SVM

이절의 FOC-SVM II은 2.2의 OC-SVM II에 퍼지 맴버쉽을 적용한다. 중요도 $\beta_i \in (0, 1]$ 를 가지는 데이터 포인트의 집합을 $(x_1, \beta_1), (x_2, \beta_2), \dots, (x_n, \beta_n)$ 라고 가정하자. (x_i, β_i) 에서 x_i 는 객체를 표현하는 데이터 포인트이고, β_i 는 객체의 중요도를 나타낸다. 평면을 사용한 OC-SVM II의 기준은 중심점과의 코사인 유사도이다. 여기에서는 β_i 를 가지고 유사도의 크기를 조절한다.

다음은 β_i 를 적용한 목적함수이다.



〈그림 6〉 중심점과의 유사도 조정



〈그림 7〉 데이터의 위치 조정

$$\text{minimize} \frac{1}{2} \|w\|^2 + C \sum_i \xi_i - \rho \quad (45)$$

$$\text{subject to } \beta_i(\omega \cdot \phi(x_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad \forall i$$

또한, 식 (45)은 데이터 포인트를 $\omega \cdot \phi(x_i) = \rho$ 와 수직인 방향으로 평행이동시킨 것으로 해석할 수 있다.

식 (45)을 풀기 위하여 라그랑지안 계수법을 적용한다.

$$L = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i - \rho - \sum_i \alpha_i (\beta_i(\omega \cdot \phi(x_i)) - \rho + \xi_i) - \sum_i \gamma_i \xi_i, \quad (46)$$

$$\alpha_i \geq 0, \quad \gamma_i \geq 0$$

L 을 ω, ρ, ξ_i 로 각각 미분하여 0 으로 놓으면 다음의 조건들이 나온다.

$$\sum_i \alpha_i = 1, \quad (47)$$

$$\omega = \sum_i \alpha_i \beta_i \phi(x_i), \quad (48)$$

$$C - \alpha_i - \gamma_i = 0 \Rightarrow 0 \leq \alpha_i \leq C \quad (49)$$

식 (45)를 α_i 에 대해 다시 표현하면 다음과 같다.

$$\text{maximize } W = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j \beta_i \beta_j K(x_i, x_j) \quad (50)$$

$$\text{subject to } \sum_i \alpha_i = 1, \quad 0 \leq \alpha_i \leq C, \quad \forall i$$

식 (45)과 (50)은 KKT 충족조건을 만족함으로써 듀얼 문제가 된다. KKT 충족 조건은 다음과 같다[3].

$$\gamma_i \xi_i = 0 \quad (51)$$

$$\alpha_i (\beta_i(\omega \cdot \phi(x_i)) - \rho + \xi_i) = 0 \quad (52)$$

KKT 조건을 α_i 로 설명하면 1) $\alpha_i = 0$ 인 포인트 x_i 는 $\beta_i(\omega \cdot \phi(x)) > \rho$ 의 영역에 놓여 있고, 2) $0 < \alpha_i < C$ 인 포인트 x_i 는 $\beta_i(\omega \cdot \phi(x)) = \rho$ 에 놓여 있다. 이때의 포인트 x_i 를 서포트 벡터라 한다. 그리고, 3) $\alpha_i = C$ 인 포인트 x_i 는 $\beta_i(\omega \cdot \phi(x)) < \rho$ 의 영역에 놓여 있게 되고, 이러한 데이터 포인트를 아웃라이어라 한다.

모든 데이터 포인트 x 에 대하여, 특징 공간에서 $\phi(x)$ 와 ω 의 유사도 $f(x, \beta)$ 는 다음과 같다.

$$f(x, \beta) = \beta(\omega \cdot \phi(x)) \quad (53)$$

$$f(x, \beta) = \sum_i \alpha_i \beta_i K(x, x_i)$$

ρ 는 서포트 벡터 x_s 으로 구할 수 있다.

$$\rho = f(x_s, \beta_s) \quad (54)$$

최적화(Optimization)

식 (50)는 Quadratic programming을 이용하여 해를 구할 수 있다. 2.3절에서와 같이 SMO를 수정하여 이용한다[3,6,12]. FOC-SVM II에 SMO를 적

용하여 a 에 대한 개선 식을 구하면 다음과 같다.

$$\begin{aligned}\alpha_2 &= \alpha_2^* + \frac{S_1 - S_2}{\beta_1^2 K_{11} + \beta_2^2 K_{22} - 2\beta_1\beta_2 K_{12}} \\ \alpha_1 &= \alpha_1^* + \alpha_2^* - \alpha_2\end{aligned}\quad (55)$$

식 (55)에서 S_i 는 $\beta_i\phi(x_i)$ 와 중심점 ω 의 내적이다.

$$S_i = \sum_j \alpha_j \beta_i \beta_j K(x_i, x_j) \quad (56)$$

즉, S_i 는 $\phi(x_i)$ 와 중심점 ω 의 유사도를 β_i 로 조정한 것을 의미한다.

식 (55)에서 S_2 가 S_1 보다 작을수록 α_2 는 커진다. S_2 가 작은 값이 되기 위해서는 $\phi(x_i)$ 와 ω 의 유사도가 작거나 β_i 가 작은 값을 가져야 한다. 즉, FOC-SVM II에서 데이터 포인터가 서포트 벡터가 되기 위해서는 데이터 포인터와 중심점 ω 의 유사도가 낮거나 데이터 포인터의 중요도 β 가 작아야 한다.

3.3 파라미터 C 와 σ

파라미터 C 와 σ 는 서포트 벡터와 아웃라이어의 선정에 영향을 준다. 다시 말하면, 파라미터 C 와 σ 는 데이터 포인터가 존재하는 영역에 영향을 준다.

σ 은 가우시안 커널 함수의 스케일을 조절함으로써 데이터 포인터가 존재하는 영역에 영향을 준다. σ 가 작다면 각 데이터들간의 거리(또는 유사도)의 OC-SVM(또는 FOC-SVM)에 대한 영향력이 커져 데이터가 존재하는 영역을 자세하게 얻게 된다. 반대로, σ 가 크다면 각 데이터들간의 거리(또는 유사도)의 OC-SVM(또는 FOC-SVM)에 대한 영향력이 적어져 구에 가까운 영역을 얻게 된다.

파라미터 C 의 영향력은 $C=1/(nv)$ 로 표현함으로써 쉽게 알 수 있다. 여기서, n 은 주어진 데이터의 개수이고, v 는 $v \in (0,1]$ 를 갖는다. $\sum_i \alpha_i = 1$

와 $0 \leq \alpha_i \leq C$ 를 가지고 v 를 분석하면, v 는 아웃라이어 비율의 상한값이면서 서포트 벡터 비율의 하한값이다. 즉, v 를 조절하여 서포트 벡터의 수를 조절할 수 있다. 또한, $C=1$ 이면, 아웃라이어를 허용치 않고, 모든 데이터를 포함하는 영역을 서포트 벡터만으로 표현한다.

파라미터 C 와 σ 에 대한 것은 [11]과 [2]에서 심도있게 연구되었다.

3.4 퍼지 측정 결합

데이터의 중요도를 측정할 수 있는 정보 소스가 다양하게 있을 수 있다. 문제는 이러한 측정치를 어떻게 결합하는가이다. 서로 다른 소스에서 얻어진 정보를 결합하는 방법으로 퍼지 결합방법을 사용한다[12]. 퍼지결합 방법은 다음과 같이 설명된다.

다음과 같은 집합 $f^{(j)}(m_i)$, for $i = 1 \dots n$ 이 있다고 가정한다. 여기서, 각 $f^{(j)}(m_i)$ 는 서로 다른 정보 m_i 에서 얻어진 중요도를 나타낸다. 그러면 결합된 중요도는 다음과 같이 나타낼 수 있다.

$$S_j = f^{(j)}(m_1) \oplus \dots \oplus f^{(j)}(m_i) \oplus \dots \oplus f^{(j)}(m_n) \quad (57)$$

여기서, \oplus 는 결합 연산자이다.

어떠한 결합 연산자를 사용하는가에 따라서, 전체 중요도의 값이 결정된다. 예를 들어, 승산을 결합연산자로 선택한다면, 결과적으로 중요도는 다음과 같다.

$$S_j = \prod_{i=1 \dots n} f^{(j)}(m_i) \quad (58)$$

어떠한 결합연산자를 사용하는 것이 효과적인가에 대한 것은 FOC-SVM의 응용에 아주 중요한 요소가 된다.

4. 실험

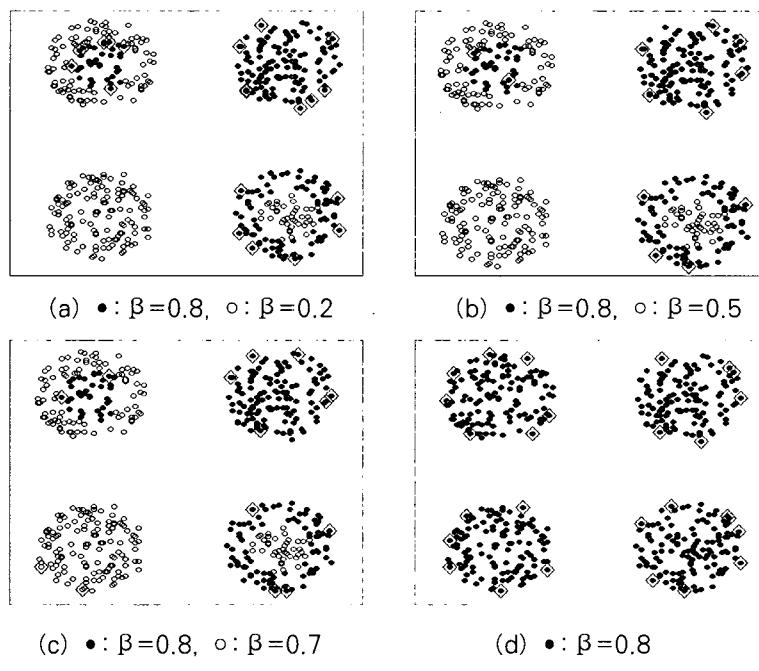
중요도와 중심점과의 거리(또는 유사도)가 FOC-

SVM에 어떠한 영향을 주는 가에 대해 알아보기 위하여 가상의 데이터를 가지고 실험을 하였다. 데이터 집합은 네 개의 군집으로 이루어져 있는 비선형 데이터이다. 각 군집들은 각각 1) 낮은 중요도를 가진 데이터, 2) 높은 중요도를 가진 데이터, 3) 바깥쪽은 낮은 중요도, 안쪽은 높은 중요도를 가진 데이터, 4) 바깥쪽은 높은 중요도, 안쪽은 낮은 중요도를 가진 데이터로 되어 있다. FOC-SVM의 파라미터는 $C=1$, $\sigma = sd \times 0.5$ 로 설정하였다. sd 는 데이터의 표준편차(standard deviation)이다. 그림 8와 그림 10은 FOC-SVM의 비선형 데이터 집합에 대한 실행 결과이다. 흑점은 중요도 $\beta_i = 0.8$ 로 주고 흰점은 β_i 를 (a) 0.2 (b) 0.5 (c) 0.7 (d) 0.8로 설정하였다. 마름모의 데이터는 실행 결과로써 서포트 벡터를 의미한다.

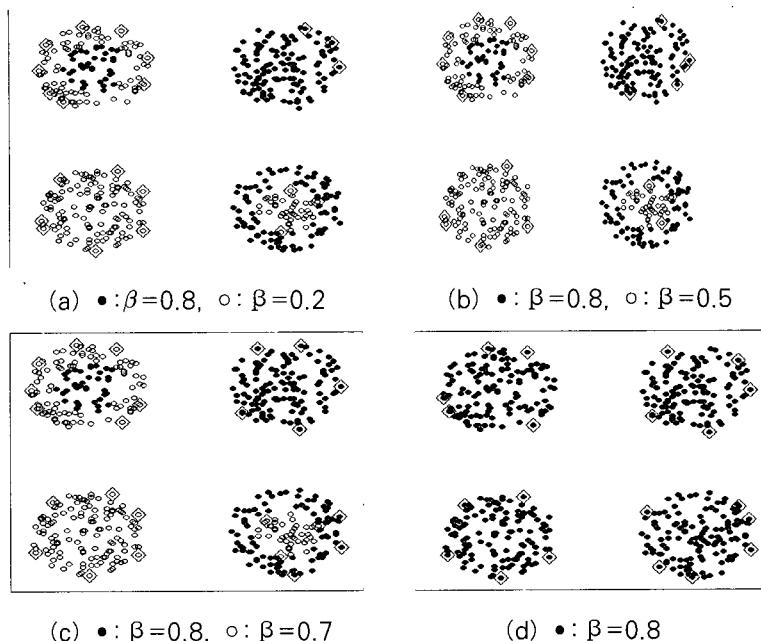
그림 8는 2.1절의 FOC-SVM I에 대한 결과이다. FOC-SVM I의 서포트 벡터 선정 기준은 중요도가 높으면서 중심점과의 거리가 큰 것이다. (a)와 (b)는 흑점에서만 서포트 벡터가 추출되었다. 흑점의 데이터와 흰점의 데이터의 중요도 차이가

크기 때문에 중심점의 거리 보다는 중요도에 편중되게 서포트 벡터가 선정된 것이다. (c)는 좌측 하단의 데이터 군에서 흰점의 데이터에서도 서포트 벡터가 추출되었다. 중요도 차이가 줄어들어, 중요도는 낮지만 중심점과의 거리가 큰 데이터에서 서포트 벡터가 추출된 것이다. (d)는 모든 데이터의 중요도가 같다. 따라서 중심점과의 거리만으로 서포트 벡터가 선정되었다. (d)는 OC-SVM과 같은 결과이다.

그림 9은 FOC-SVM II의 실행 결과이다. FOC-SVM II의 서포트 벡터 선정 기준은 중요도 값이 낮거나 중심점과의 유사도가 낮은 것이다. (a)는 중요도의 차이가 커 대체적으로 중요도 값이 낮은 데이터에서 서포트 벡터로써 많이 추출되었다. 그러나, FOC-SVM I과는 다르게 중요도가 높은 흑점이지만 유사도가 아주 낮은 데이터에서도 서포트 벡터가 추출되는 것을 볼 수 있다. (b)와 (c)로 갈수록 중요도의 차이가 줄어들면서 서포트 벡터가 흰점에서 많이 추출되었던 것이 흑점에서 추출되는 양이 늘어나는 것을 볼 수 있다. (d)는 그림



〈그림 8〉 FOC-SVM I의 결과



〈그림 9〉 FOC-SVMII의 결과

8의 (d)처럼 OC-SVM과 같은 결과이다.

는 것이다.

5. 결 론

OC-SVM은 모든 데이터 포인트들을 동등하게 다룬다. 그러나 사전지식으로부터 얻을 수 있는 각 데이터의 중요도 측정치에 따라 데이터 포인트들을 처리해야 하는 응용분야가 있을 수 있다. 기존의 OC-SVM 방법으로는 이것을 해결하기 어렵다. 본 논문에서는 이러한 문제를 해결하기 위하여 퍼지 멤버쉽을 사용하였다. 즉, 모든 데이터들을 동등하게 다루는 것이 아니라, 각 데이터의 중요도 측정치에 따라 데이터 포인트들을 다루는 새로운 접근 방법인 FOC-SVM(Fuzzy One Class SVM)을 제안하였다. FOC-SVM은 OC-SVM의 응용분야에 적용되어 더욱 훌륭한 결과를 얻을 수 있을 것이라 기대한다.

향후 연구 방향은 FOC-SVM의 적합한 응용분야를 더욱 넓히는 것과 그에 따른 데이터의 중요도, 즉 퍼지 멤버쉽 함수를 더욱 다양하게 연구하

참 고 문 헌

- [1] Asa Ben-Hur, David Horn, Hava T. Siegelmann, and Vladimir Vapnik, "Support Vector Clustering", Journal of Machine Learning Research 2, pp.125-137, 2001.
- [2] Klaus Robert Muller, Sebastian Mika, Gunnar Ratsch, Koji Tsuda, and Bernhard Scholkopf, "An Introduction to Kernel-Based Learning Algorithms", IEEE Transactions on Neural Networks, Vol. 12, No.2, pp.181-202, March, 2001.
- [3] Nello Cristianini and John Shawe-Taylor, "An Introduction to Support Vector Machines and other kernel-based learning methos", Cambridge University Press, 2000.
- [4] Manuel Davy and simon Godsill, "Detection

- of abrupt spectral changes using support vector machines: an application to audio signal segmentation", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, 12-17, pp. 1313-1316, May 2002.
- [5] Yunqiang Chen, Xiang Sean Zhou, and T.S. Huang, "One-class SVM for learning in image retrieval", Proceedings of International Conference on Image Processing, Vol. 1, pp.34-37, 2001.
- [6] B. Scholkopf, J.C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution", Microsoft Research Technical Report, MSR-TR-99-87, 1999.
- [7] J. C. Platt "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", Microsoft Research Technical Report, MST-TR-98-14, 1998.
- [8] D. M. J. Tax and R. P. W. Duin, "Support vector domain description", Pattern Recognition Letters, 20(11-13):1191-- 1199, December 1999.
- [9] Tax, D. M. J. and Duin, R. P. W., "Image database retrieval with Support vector data description", Proceedings of the Sixth Annual Conference of the Advanced School for Computing and Imaging, ASCI, Delft, June 2000.
- [10] P. Hayton, B. Scholkopf, L. Tarassenko, and P. Anuzis, "Support vector novelty detection applied to jet engine vibration spectra", in NIPS'2000, 2000.
- [11] Tax, D. M. J. and Duin, R. P. W., "Outliers and data descriptions", Proceedings of the Seventh Annual Conference of the Advanced School for Computing and Imaging, ASCI, Delft, June 2001.
- [12] J. M. Keller, et. al., "Evidence Aggregation Networks for Fuzzy Logic Inference", IEEE Transactions on Neural Networks, Vol. 3, No. 5, pp. 762-769, 1992

❶ 저자 소개 ❶



김기주 (KiJoo Kim)

2004년 한국항공대학교 컴퓨터공학과 졸업(학사)
2004년~현재 한국항공대학교 대학원 컴퓨터공학과 석사과정
관심분야 : 웹데이터마이닝, 영상처리, etc.
E-mail : zgeniez@hau.ac.kr



최영식 (YoungSik Choi)

1985년 연세대학교 전자공학과 졸업(학사)
1987년 연세대학교 대학원 전자공학과 졸업(석사)
1996년 미국 Missouri주 Columbia대학교 대학원 Computer Engineering 졸업(박사)
2000년~2002년 KT(주) 멀티미디어연구소 연구실장
2002년~현재 한국항공대학교 항공전자정보통신공학부 조교수
관심분야 : 퍼지, 웹데이터마이닝, 영상처리, 이미지검색, etc.
E-mail : choimail@hau.ac.kr