2004년도 대한전기학회 하계학술대회 논문집 2004.7.14-16

퍼지 써포트 벡터 머신을 이용한 패턴 분류

이 선 영, 김 성 수
충북 대학교 전기공학과

Pattern Classification using Fuzzy Suppot Vector machine

Sun-Young Lee, Sung-Soo Kim
Chungbuk National University, School of Electrical Engineering

Abstract - 일반적으로 support vector machine (SVM)은 입력 데이터를 두개의 다른 클래스로 구별하는 결정면을 학습을 통하여 구한다. 특히 비분류 문제, 비선형 분류 문제들과 같은 두-클래스 문제를 해결하기 위해 데이터를 고차원의 특징 공간에서 다룬다. 많은 응용분야에서, 각 입력 데이터들은 이 두개의 클래스 중의 하나로 완전히 정의되지 않을 수도 있다. 이러한 문제를 해결하기 위해 우리는 본 논문에서 FSVM(fuzzy support vector machine)을 적용한다. 각 입력 데이터에 퍼지 맴버십(fuzzy membership)을 적용하여 결정면의 학습과정에 입력 데이터들이 다른 기여(contribution)를 할 수 있게 한다. 본 논문에서는 기준 데이터 집합에 대해 제안된 방법을 실험하고, FSVM이 기존의 SVM보다 더 나음을 보인다.

1. 서 론

Support vector machines(SVMs)은 Dr. Vapnik, 과 그의 연구 그룹에 의해 발전된 새로운 분류 기법이며, 최근 많이 연구되고 있다[1]. 두 클래스 사이의 최대 마진을 가지는 최적 초평면(hyperplane)에서 발전된 SVM이론은 구조 위험 최소화(Structural risk minimization)의 개념에 기초한다. 많은 응용에서 SVM은 전통적인 학습 기구들 보다 높은 수행성을 보여줬고 분류 문제를 해결하는 강력한 수단으로 소개되어 왔다. SVM은 입력 데이터들을 고차원의 특징 공간으로 사영(mapping)하고, 이 공간에서 두 클래스 사이의 마진이 최대가 되는 분리 초평면을 찾는다. 마진을 최대화하는 것은 이차 방정식 문제(Quadratic programming)이고, 문제를 해결하기 위해 라그랑제 계수(Lagrangian multiplier)를 이용한다. 사영에 대한 어떤 정보 없이 SVM은 특징 공간에서 커널(kernel) 함수의 내적으로 최적 초평면을 찾는다. 최적 초평면의 답은 support vector라고 불리는 몇 개의 입력 데이터들의 조합으로 쓰여진다. SVM 기법은 점점 더 많은 응용분야에서 사용된다. 그러나 어떤 입력 데이터들은 두 클래스 중의 어느 한 클래스로 정확히 지정되지 않을지도 모른다[2]. SVM이 더 정확히 데이터들을 분리 할 수 있도록 한개 클래스로 완전히 지정하거나, 잡음에 의해 훼손된 데이터들은 무의미하므로 이 훼손된 데이터들은 제거하는 것이 나을 수 있다. 그러나 SVM은 이런 능력이 부족하다. 그래서 본 논문에서는 SVM의 입력 데이터에 퍼지 소속도를 적용하여, 결정면의 학습에 대한 입력 데이터의 기여를 다르게 하는 퍼지 SVM을 제안한다. 2장에서는 SVM의 이론을 설명하고 3장에서는 퍼지 소속도를 적용한 FSVM에 대해 설명한다. 그리고 4장에서는 실험을 통한 FSVM의 성능을 보여주고 마지막으로 5장에서는 결론을 갖는다.

2. 본 론

이 장에서는 분류 문제에서 SVM의 기본적인 이론에 대해 설명한다. 식(1)과 같이 지정된 훈련 데이터의 집합이 주어졌다고 가정한다.

$$(y_1,\ \mathbf{x}_1),\cdots,(y_l,\ \mathbf{x}_l) \tag{1}$$

각 훈련 데이터 $\mathbf{x}_i \in R^N$는 두 클래스 중 하나의 클래스에 속하고 라벨 $y_i \in -1, 1,\ i=1,\cdots,l$이 주어진다. 대부분의 실제 응용분야에서 입력 공간에서 적당한 초평면을 찾는 것은 제한적이다. 이 제한성에 대한 해결방법은 입력 공간을 고차원의 특징 공간으로 사영하고, 이 특징 공간에서 최적의 초평면을 찾는 것이다. $\mathbf{z}=\varphi(\mathbf{x})$ 이 사영함수 φ을 가지고 R^N으로부터 특징 공간 Z로 사영된 특징 공간 벡터로 정의한다. 우리는 $(\mathbf{w}, b)$의 쌍으로 정의된 식(2)와 같은 초평면을 찾고 식(3)의 함수 $f(\mathbf{x}_i)$로 $\mathbf{x}_i$를 분리 할 수 있다.

$$\mathbf{w}\cdot\mathbf{x}+b=0 \tag{2}$$

$$f(\mathbf{x}_i)=\mathrm{sign}(\mathbf{w}\cdot\mathbf{z}_i+b)=\begin{cases}1, & \text{if } y_i=1\\ -1, & \text{if } y_i=1\end{cases} \tag{3}$$

여기서 $\mathbf{w}\in Z$ 이고, $b\in R$이다.

만일 $(\mathbf{w}, b)$이 존재하여 식(4)의 부등식이 집합 S의 모든 원소에 유효하면, 집합 S는 선형 분리된다고 말한다.

$$\begin{cases}(\mathbf{w}\cdot\mathbf{z}_i+b)\ge 1 & \text{if } y_i=1\\ (\mathbf{w}\cdot\mathbf{z}_i+b)\le -1 & \text{if } y_i=-1\end{cases}\quad i=1,\cdots,l \tag{4}$$

선형 분리 집합 S에서 두개의 다른 클래스의 사영된 훈련 데이터들 사이의 마진에 대해 유일한 최적 초평면을 찾을 수 있다. 만일 집합 S가 선형 분리가 아니면 SVM형태에서 잘못된 분류가 얻어질 수 있어, 사전 처리로 음이 아닌 어떤 변수 $\xi\ge 0$를 적용하여 식(4)를 다음과 같이 변환한다.

$$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\ge 1-\xi_i,\quad i=1,\cdots,l \tag{5}$$

식(5)의 영이 아닌 ξ_i는 식(4)를 만족하지 않는 점 $\mathbf{x}_i$에 대응하여, $\Sigma_{i=1}^{l}\xi_i$는 잘못된 분류에 대한 값을 나타낸다고 생각 할 수 있다. 최적 초평면 문제는 다음 식의 답을 구하는 것으로 생각한다.

$$\begin{aligned}&\text{minimize}\ \ \frac{1}{2}\mathbf{w}\cdot\mathbf{w}+C\sum_{i=1}^{l}\xi_i\\ &\text{subject to}\ \ y_i(\mathbf{w}\cdot\mathbf{z}_i+b)\ge 1-\xi_i,\quad i=1,\cdots,l\\ &\xi_i\ge 0,\ \ i=1,\cdots,l\end{aligned} \tag{6}$$

여기서 C는 상수이며 SVM식에서 유일한 자유 매개변수이다. 이 변수의 조정은 마진의 최대화와 분류 위반 사이의 관계를 조정할 수 있다. 식(6)에서 초평면을 찾는 것은 QP 문제이다. 이것은 Lagrangian 구성으로 풀 수 있고 식(7)처럼 변형 할 수 있다.

$$\text{maximize}\ \ W(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \mathbf{z}_i\cdot\mathbf{z}_j$$

$$\text{subject to } \sum_{i=1}^{l} y_i \alpha_i = 0 \quad 0 \le \alpha_i \le C, \quad i=1,\cdots,l \quad (7)$$

여기서 $\alpha = (\alpha_1, \cdots, \alpha_l)$는 제약 조건 식(5)에 대한 음이 아닌 Lagrange 계수의 벡터이다.
Kuhn-Tucker 정리는 SVM에서 중요한 역할을 한다. 이 정리에 따라 식(7)의 답 $\overline{\alpha}_i$는 다음을 만족한다.

$$\overline{\alpha}_i (y_i (\overline{\mathbf{w}} \cdot \mathbf{z}_i + \overline{b}) - 1 + \overline{\xi}_i) = 0, \quad i=1,\cdots,l \quad (8)$$

$$(C - \overline{\alpha}_i)\overline{\xi}_i = 0, \quad i=1,\cdots,l \quad (9)$$

이 등식으로부터 식(8)에서 영이 아닌 값 $\overline{\alpha}_i$는 제약 식(5)의 등호를 만족하는 것들이다. $\overline{\alpha}_i > 0$에 대응하는 점 $\mathbf{x}_i$를 서포트 벡터(support vector)라고 부른다.
최적 초평면 $\overline{\mathbf{w}} \cdot \mathbf{z} + \overline{b}$을 구하기 위해 식(10)과 같이 쓰고 $\overline{b}$는 Kuhn-Tucker조건 식(8)으로 구할 수 있다.

$$\overline{\mathbf{w}} = \sum_{i=1}^{l} \overline{\alpha}_i y_i \mathbf{z}_i \quad (10)$$

결정 함수는 식(3)과 식(10)으로부터 일반화된다.

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{z} + b) = \text{sign}\left(\sum_{i=1}^{l} \alpha_i y_i \mathbf{z}_i \cdot \mathbf{z} + b\right) \quad (11)$$

φ에 대한 정보 없이 식(7)과 식(11)를 계산하는 것은 불가능하다. 그러나 SVM은 φ 에 관한 정보 없이 단지 특징 공간 Z에서 데이터 점들의 내적으로 계산할 수 있는 커널 함수 $K(\cdot, \cdot)$만 필요하다.

$$\mathbf{z}_i \cdot \mathbf{z}_j = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \quad (12)$$

$$\text{maximize } W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subject to } \sum_{i=1}^{l} y_i \alpha_i = 0 \quad 0 \le \alpha_i \le C, \quad i=1,\cdots,l \quad (13)$$

그래서 식(13)의 답을 구해 비선형 분리 초평면을 구할 수 있고 결정함수는 식(14)와 같다.

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{z} + b) = \text{sign}\left(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i \cdot \mathbf{x} + b)\right) (14)$$

3. FSVM

이 장에서는 FSVM에 대해 설명한다. SVM은 분류 문제를 푸는데 있어 강력한 툴이지만 약간의 제한들이 있다. 많은 실제 응용에서 훈련 데이터들이 미치는 영향은 다르다. 종종 어떤 훈련 데이터들은 분류 문제에서 다른 데이터들 보다 더 중요하다. 우리는 의미 있는 학습 데이터들이 정확하게 분류되어지는 것을 원하고 점들이 잘못 분류되던 되지 않던 노이즈와 같은 학습 데이터들은 고려하지 않는다. 훈련 데이터들은 다른 소속도를 가지고 하나 이상의 클래스에 속할 수 있다. 이 개념은 전통적인 SVM기법에서는 고려되어지지 않고 단지 엄밀한 주어진 정보만을 고려한다. 퍼지 소속도 $0 < s_i \le 1$의 값으로 각 훈련 데이터 $\mathbf{x}_i$를 할당한다. 이 퍼지 소속도 s_i는 분류 문제에서 한 클래스에 기여하는 훈련 데이터의 크기로 나타내어 질 수 있고 $(1 - s_i)$는 무의미함의 크기로 나타내어 질 수 있다. 여기서 SVM의 개념을 퍼지 소속도를 이용해 확장하고 FSVM을 만들었다[2].
훈련 데이터들의 집합 S가 아래 식과 같이 퍼지 소속도를 가지고 주어졌다고 가정하자.

$$(y_1, \mathbf{x}_1, s_1), \cdots, (y_l, \mathbf{x}_l, s_l) \quad (15)$$

각 훈련 데이터 $\mathbf{x}_i \in R^N$는 라벨 $y_i \in -1, 1$와 퍼지 소속도 $\sigma \le s_i \le 1$이 주어진다. 여기서 $i=1,\cdots,l$이고 $\sigma > 0$이다. 퍼지 소속도 s_i는 한 클래스에 기여하는 점 $\mathbf{x}_i$의 크기이고 파라미터 ξ_i는 SVM에서 에러의 크기이므로, $s_i \xi_i$는 다른 가중치를 가지는 에러의 값이다. 최적 초평면 문제는 다음 식과 같이 답을 구한다.

$$\text{maximize } \frac{1}{2} \mathbf{w} \cdot \mathbf{w} \; C \sum_{i=1}^{l} s_i \xi_i$$

$$\text{subject to } y_i(\mathbf{w} \cdot \mathbf{z}_i + b) \ge 1 - \xi_i, \quad i=1,\cdots,l$$

$$\xi_i \ge 0, \quad i=1,\cdots,l \quad (16)$$

여기서 C는 상수이다. 식 (16)에서 s_i가 작을수록 파라미터 ξ_i의 영향을 줄이고 대응하는 점 $\mathbf{x}_i$는 덜 중요하게 다루어진다.
이 최적화 문제를 풀기 위해 우리는 Lagrangian을 구성한다.

$$L(\mathbf{w}, \mathbf{b}, \xi, \alpha, \beta) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{l} s_i \xi_i - \sum_{i=1}^{l} \alpha_i (y_i (\mathbf{w} \cdot \mathbf{z}_i + b) - 1 + \xi_i) - \sum_{i=1}^{l} \beta_i \xi_i \quad (17)$$

그리고 $L(\mathbf{w}, b, \xi, \alpha, \beta)$의 말안장점(saddle point)을 찾는다. 파라미터들은 다음의 조건들을 만족해야 한다.

$$\frac{\partial L(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i \mathbf{z}_i = 0 \quad (18)$$

$$\frac{\partial L(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial b} = -\sum_{i=1}^{l} \alpha_i y_i = 0 \quad (19)$$

$$\frac{\partial L(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial \xi_i} = s_i C - \alpha_i - \beta_i = 0 \quad (20)$$

식(17)에 위의 조건들을 적용하면 식(16)은 다음과 같이 다시 쓸 수 있다.

$$\text{maximize } W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subject to } \sum_{i=1}^{l} y_i \alpha_i = 0 \quad 0 \le \alpha_i \le s_i C, \quad i=1,\cdots,l \quad (21)$$

그리고 Kuhn-Tucker 조건은 다음과 같이 정의된다.

$$\overline{\alpha}_i (y_i (\overline{\mathbf{w}} \cdot \mathbf{z}_i + \overline{b}) - 1 + \overline{\xi}_i) = 0, \quad i=1,\cdots,l \quad (22)$$

$$(s_i C - \overline{\alpha}_i)\overline{\xi}_i = 0, \quad i=1,\cdots,l \quad (23)$$

$\overline{\alpha}_i > 0$에 대응하는 데이터 $\mathbf{x}_i$는 서포트 벡터라고 부르고 두 가지의 형태가 있다. $0 < \overline{\alpha}_i < s_i C$에 해당하는 서포트 벡터는 초평면의 마진에 놓여 있고, $\overline{\alpha}_i = s_i C$에 대응하는 것은 잘못 분류된 서포트 벡터이다. SVM과 FSVM사이의 중요한 차이점은 $\overline{\alpha}_i$의 같은 값을 가지는 데이터가 FSVM에서 s_i에 따라 서포트 벡터의 다른 형태를 가리키게 된다는 점이다.

4. 실험 및 고찰

주어진 문제에 알맞은 퍼지 소속도를 정하려면 우선 퍼지 소속도의 하한선을 정하고 데이터 집합의 주요 특징을 고려하여 특징과 퍼지 소속도 사이의 연관성을 만든다. 본 논문에서는 임의로 주어진 데이터를 두 클래스로 분류하려고 한다. 먼저 순차적으로 입력되는 데이터를 고려하여 퍼지 소속도 s_i를 시간에 대한 함수로 만든다. 입력 데이터들은 과거의 입력보다 최근의 입력 데이터가 더 중요하다. $t_1 \le \cdots \le t_l$은 시스템에 입력이 들어오는 시간이고 $s_1 = \sigma \le \cdots \le s_l = 1$로 소속도는 다음과 같다.

$$s_i = f(t_i) = (1-\sigma)\left(\frac{t_i - t_1}{t_l - t_1}\right)^2 + \sigma \qquad (24)$$

그림 1과 그림 2는 각각 SVM과 FSVM의 결과를 보여준다. 'x' 가 한 클래스이고 '□'이 다른 클래스이다. 숫자들은 입력 데이터의 순서를 나타낸다. 숫자가 작을 수록 오래된 데이터이다. 그림2의 마지막 10개 데이터에서 FSVM을 이용한 결과가 SVM에서 보다 높은 정확성을 보여준다.

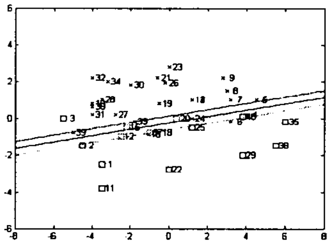

그림 1. 시간 특성을 가지는 데이터의 SVM 학습 결과

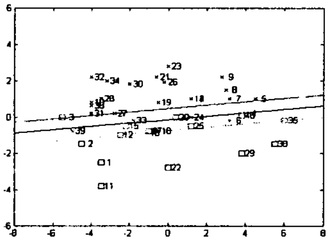

그림 2. 시간 특성을 가지는 데이터의 FSVM 학습결과

다음으로 주어진 데이터를 각 클래스의 중심 값을 이용하여 중심과의 거리관계로 퍼지 소속도를 부여했다[3]. +1 클래스의 평균값 $\mathbf{x}_+$, 다른 클래스와의 최소거리는 r_+로 나타내고, -1 클래스의 평균은 $\mathbf{x}_-$, 최소거리는 r_-로 나타내면 소속도는 다음과 같다.

$$s_i = \begin{cases} 1 - r_+ / |\mathbf{x}_- - \mathbf{x}_i| + \sigma, & \text{if} \quad y_i = 1 \\ 1 - r_- / |\mathbf{x}_+ - \mathbf{x}_i| + \sigma, & \text{if} \quad y_i = -1 \end{cases} \qquad (25)$$

그림3은 SVM을 이용한 결과이고 그림 4는 FSVM을 적용한 결과이다. SVM을 이용한 분류에서 잘못된 분류 데이터는 네 개인 반면 FSVM을 이용한 분류에서는 두 개의 오차를 가진다.

두개의 실험을 통해 퍼지 소속도를 적용한 FSVM의 분류가 더 잘 이루어짐을 알 수 있다.

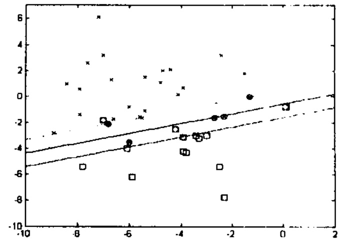

그림 3. SVM을 이용한 분류

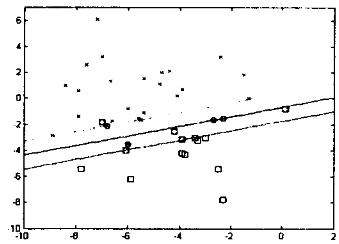

그림 4. FSVM을 이용한 분류

5. 결 론

어떤 입력 데이터들은 두 클래스 중의 어느 한 클래스로 정확히 분류되지 않을 수 있다. 정확히 데이터들을 분리 할 수 있도록 한개 클래스로 완전히 지정하거나, 잡음에 의해 훼손된 데이터들은 무의미하므로 제거하는 것이 나을 수 있다. 그러나 SVM은 이런 능력이 부족하여 본 논문에서는 SVM의 입력 데이터에 퍼지 소속도를 적용하여, 결정면 학습시 입력 데이터의 기여를 다르게 하는 퍼지 SVM을 제안하였다. 4장에서 살펴보았듯이 기존의 SVM을 이용한 분류보다 FSVM을 이용한 분류가 더 나은 결과를 가져왔다.

[참 고 문 헌]

[1] C.Burges, "A tutorial on support vector machines for pattern recognition," Data Mining and knowledge discovery, Vol. 2, No. 2, 1988.

[2] Chun Fu Lin and Sheng De Wang, "Fuzzy support vector machines", IEEE Transactions on Neural Networks Vol.13. No. 2, pp.464-471, March 2002.

[3] X.Zhang, "Using class-center vectors to build support vector machines," in Proc. IEEE NNSP'99, pp.3-11, 1999.

[4] T.Inoue, and S.Abe, "Fuzzy support vector machines for pattern classification", International Joint Conference on Neural networks, Vol.2, 15-19 pp.1449-1454, July 2001