2005년도 대한전기학회 하계학술대회 논문집 2005. 7. 18-20

다중 클래스 분류를 위한 FSVM

이 선 영, 김 성 수
충북 대학교 전기공학과

FSVM for Multi Class Classification

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

Abstract - Support vector machine(SVM)은 입력 데이터를 두개의 다른 클래스로 구별하는 결정면을 학습과정을 통하여 구한다. 기존의 SVM은 단지 이차 클래스에 대하여 적용되어지나, 많은 응용분야에서 입력 데이터들은 몇 개의 다중 클래스로 분류해야 한다. 다중 클래스 분류 문제는 기존의 SVM을 사용할 수 있는 일반적으로 몇 개의 2차 문제로 분해하여 풀 수 있다. 실례로 one-against-all 방법을 적용하면, n 클래스 문제는 n개의 두 클래스 문제로 변환 하여 풀 수 있다. 본 논문에서는 입력 패턴들을 다중 클래스로 분류 할 때 퍼지 소속도를 응용한 소프트 마진 알고리즘의 상한 경계값을 각 클래스에 따라 다르게 적용함으로써 기존의 SVM 보다 더 우수한 학습 능력을 가짐을 보였다.

1. 서 론

1995년에 Vapnik에의 소개된 새로운 분류 기법인 써포트 벡터 머신(Support Vector Machine : SVM)의 학습 알고리즘은 훈련 패턴들을 서로 다른 두개의 클래스로 이원 분류 할 때 기준이 되는 결정 함수를 찾는다. 이 SVM의 기본 원리는 훈련 패턴들을 고차원 특징공간으로 사상 시킨 후 두 분류 사이의 여유를 최대화 시키는 결정 함수 즉 분리 초평면을 찾는 것이다. 최적 초평면의 해답은 써포트 벡터라고 불리는 몇 개의 입력 데이터들의 결합으로 나타낸다[1-2]. 입력 데이터들은 정확히 두 클래스 중의 어느 하나로 지정되지 않을지도 모르며, 어떤 것은 한개 클래스로 완전히 지정되는 것이 더 중요할 수도 있다. 어떤 잡음에 의해 훼손된 데이터들은 무의미 하여 학습 시 훼손된 데이터를 버리는 것이 나을지도 모른다. 그러나 기존의 SVM은 이러한 능력이 부족하다. 이러한 부족함을 해결하기 위해 퍼지 소속도를 SVM에 적용하여, 소프트-마진 알고리즘을 확장하였다[3]. 본 논문에서는 두클래스를 가지는 경우와 세 개 이상의 다중 클래스의 경우 퍼지 소속도를 가진 SVM을 제안한다. 입력 벡터들을 다중 클래스로 분류 할 때 퍼지 소속도를 응용한 소프트 마진 알고리즘의 상한 경계값을 각 클래스에 따라 다르게 적용함으로써 기존의 SVM보다 더 우수한 학습 능력을 가짐을 보인다. 2장에서 SVM의 이론을 설명하고, 3장에서는 제안된 FSVM(Fuzzy Support Vector Machine)이론과 가중치 FSVM을 살펴본다. 그리고 4장에서는 FSVM을 기준데이터에 적용해 보고, 끝으로 5장에서는 결론을 맺는다.

2. 써포트 벡터 머신

이 장에서는 SVM의 기본적인 이론에 대해 설명한다. 훈련 데이터들이 식(1)과 같이 주어진다.

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\cdots,(\mathbf{x}_l,y_l),\ \mathbf{x}\in R^n,\quad y\in\{+1,-1\} \tag{1}$$

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.

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

$$(C - \overline{\alpha_i})\overline{\xi_i} = 0, \qquad i = 1, \cdots, l \quad (9)$$

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 써포트 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.

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

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

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

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

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

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

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

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. 훈련 데이터들은 다른 소속도를 가지고 하나 이상의 클래스에 속할 수 있다. 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 적용해 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$\mathbf{x}_i \in \mathrm{R}^N$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.

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

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

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

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 라그랑시안을 구성한다.

$$L(\mathbf{w}, \mathrm{b}, \xi, \alpha, \beta) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + \mathrm{C} \sum_{i=1}^{l} \mathrm{s_i} \xi_\mathrm{i} \quad (17)$$

$$- \sum_{i=1}^{l} \alpha_i (y_i(\mathbf{w} \cdot \mathbf{z}_i + b) - 1 + \xi_i) - \sum_{i=1}^{l} \beta_i \xi_i$$

그리고 $L(\mathbf{w}, b, \xi, \alpha, \beta)$의 안장점을 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.

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

$$(s_i C - \overline{\alpha_i})\overline{\xi_i} = 0, \qquad i = 1, \cdots, l \quad (23)$$

$\overline{\alpha}_i > 0$에 대응하는 데이터 $\mathbf{x}_i$는 써포트 벡터라 한다. $0 < \overline{\alpha_i} < s_i C$에 해당하는 써포트 벡터는 초평면의 마진에 놓여 있고, $\overline{\alpha}_i = s_i C$에 대응하는 것은 잘못 분류된 써포트 벡터이다. SVM과 FSVM사이의 중요한 차이점은 $\overline{\alpha_i}$의 같은 값을 가지는 데이터가 FSVM에서 s_i에 따라 다른 형태의 써포트 벡터를 가리키게 된다는 점이다.

4. 실험 및 고찰

이 절에서는 퍼지 SVM을 이용한 기준 데이터의 클래스 분류를 살펴본다. 하나의 클래스를 다른 나머지 클래스로부터 분류할 때, 분류하고자 하는 클래스에 나머지 클래스의 데이터들보다 더 큰 소속도를 부여한다.

$$s_1 = 1,\ s_2 = 0.1 \qquad (24)$$

그림 1은 기존의 SVM을 이용한 학습결과를 보여주며, 그림 2는 식(24)의 소속도를 각 데이터에 부여하여 FSVM 학습한 결과를 보여준다. 그림 2는 'x'로 표현되는 데이터에 가중치를 주어 'x'클래스가 완전히 분리 될 수 있게 하였다.

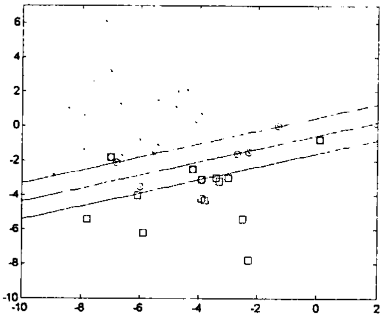

그림1. SVM의 학습 결과

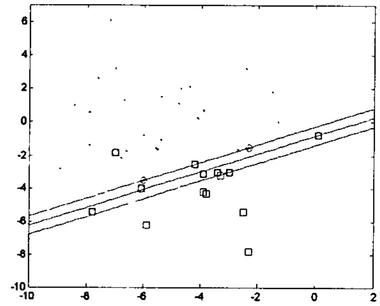

그림2. FSVM의 학습 결과

그림2에서 보는 것과 같이 'x'의 클래스는 에러 없이 분류 되었다.

다음으로 다중 클래스의 예를 살펴본다. 세 개의 클래스를 갖는 분꽃 데이터와 breast cancer의 경우 기존의 SVM에 대한 학습 능력과 본 문에서 제안한 퍼지 써포트 벡터 머신을 이용하여 분류한 결과를 비교해 본다. 세 개의 클래스로 분류되는 분꽃 데이터 150개 중 반인 75개는 학습에 이용하고 나머지 75개는 검증에 이용한다. 퍼지 소속도 s_1는 분류하고자 하는 클래스에 부여하고, s_2는 나머지 클래스들에 부여한다. FSVM의 퍼지 소속도는 $s_1 = 1$, $s_2 = 0.4$의 값을 갖는다. 표1은 두 기준 데이터의 학습 후 검증 데이터를 통한 정확도를 보여준다. 분꽃 데이터의 경우 SVM을 이용할 경우 94%의 정확도를 가졌고, FSVM의 경우는 96%의 정확도를 보였다. Breast cancer의 경우 SVM은 68.83%이고 FSVM의 경우 71.43%로 향상된 학습결과를 보여주었다.

표 1. SVM과 FSVM

Dara	SVM	FSVM
Iris	94%	96%
Breast Cancer	68.83%	71.43%

표 1은 두 가지 기준 데이터들의 경우 기존의 SVM보다 FSVM을 이용한 패턴 분류의 경우 더 나은 정확도를 가짐을 보여준다.

5. 결 론

패턴 분류의 많은 응용분야에서 특정한 한 클래스의 분류 정확도가 매우 중요할 수 있다. 나머지 클래스의 분류 정확도가 낮아지더라도 전체 분류 정확도를 높이고자 하는 것이다. 이러한 경우는 특정한 한 클래스에 속하는 패턴을 오 인식할 경우에 많은 비용과 치명적인 결과가 발생하여 이를 방지 하고자 할 때 적용될 수 있다. 본 논문에서는 SVM이 이러한 문제에 적용 될 수 있도록 두 그룹의 상한 경계 값에 벌칙 함수를 다르게 적용하는 방법을 제안한다. 즉, 입력 데이터에 퍼지 소속도를 적용하여, 결정면 학습 시 입력 데이터의 기여를 다르게 하는 퍼지 SVM을 제안하였다. 4장에서 기존의 SVM과 본 논문에서 제안한 가중치 FSVM의 학습 능력을 보였다. 두 클래스의 분류의 경우와 기준 데이터의 다중 클래스 분류의 경우 기존의 SVM을 이용한 분류보다 FSVM을 이용한 분류가 더 나은 결과를 가짐을 확인하였다.

[참 고 문 헌]

[1] C. Cortes and V. Vapnik, "Support-vector Networks", Machine Learning, vol. 20, pp. 273-297, 1995.

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

[3] 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.

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

[5] 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.

[6] A. K. Jain, R. P. W. Duin and J. Mao, "Statistical pattern recognition: A review", IEEE Trans. Pattern Anal. and Machine Intell, Vol.PAM I22, no.1, pp.4-37, Jan. 2000.