

다중 클래스 분류를 위한 강인한 SVM 설계 방법

- 생체 인식 데이터에의 적용 -

조민국^o 박혜영
경북대학교 컴퓨터과학과
fourdentist@hotmail.com^o, hjpark@knu.ac.kr

Robust SVM Design for Multi-Class Classification

- Application to Biometric data -

Minkook Cho^o Hyeyoung Park
Dept of Computer Science, Kyungpook National University

요 약

Support vector machine(SVM)은 좋은 일반화 능력을 가진 학습시스템으로, 최근 다양한 패턴 인식 분야에서 적용되고 있다. SVM은 기본적으로 이진 분류기이므로 두 개 이상의 클래스를 분류하기 위해서는 다중 클래스 분류가 가능한 형태의 설계 방법이 필요하다. 이를 위해 각 클래스별로 독립적인 SVM들을 만들어 결과를 병합하는 방식이 주로 사용되어 왔다. 그러나 이러한 방법은 클래스의 수는 많고 한 클래스 내의 데이터의 수가 많지 않은 경우에는 SVM의 일반화 성능을 저하시키고 노이즈에 민감해지는 문제점을 가지고 있다. 이를 해결하기 위해 본 논문에서는 각 클래스내의 데이터간의 유사도 측정을 위한 통계적 정보를 안정적으로 추출하기 위해 두 데이터의 쌍을 입력으로 받는 새로운 SVM 설계 방법을 제시한다. 제안한 방법을 실제 생체인식 데이터에 적용한 실험에서 기존의 방법보다 우수한 분류 성능을 보임을 확인할 수 있었다.

1. 서 론

최근 다양한 패턴 인식 문제에 Support Vector Machine(SVM)을 많이 적용하고 있다[1]. SVM은 통계적 학습이론에 기반하여 결정 경계와 학습 샘플과의 거리를 최대화하도록 분류기를 학습함으로써 뛰어난 일반화 성능을 보여준다[2][3]. 하지만 다른 학습 시스템과 같이 좋은 성능을 보이기 위해선 SVM을 학습시킬 충분히 많은 데이터가 필요하다. 도메인의 성격상 데이터의 수가 상당히 제한적 일 때(예를 들면, 의료 데이터)는 좋은 성능을 보장하기 어렵다. 그리고 SVM은 기본적으로 이진 분류기이므로 다중 분류를 위해선 다중 클래스 분류기로 확장 되어야 한다.

다중 클래스 분류를 위한 기준에 잘 알려진 설계 방법은 크게 2가지로 나뉘 볼 수 있다. 첫째, 일대다(one_vs_all;OVA) 방법으로 각 클래스 별로 해당 클래스와 그 이외 나머지를 분류하는 방법이다[4]. 이는 클래스의 수만큼의 분류기가 있어야 한다. 즉, 클래스의 수가 k개이라면, k개의 분류기가 있어야 하고 데이터가 어떤 클래스인지 분류하기 위해선 k개의 분류기 중에서 출력 값이 가장 높은 SVM을 선택하는 방식으로 비교가 이

루어진다[4]. 둘째, 전부-쌍(all-pairs:AP) 방법으로 각각의 클래스가 쌍을 이루도록 분류기를 만든다. 만약, 클래스의 수가 k개이라면, 클래스의 조합만큼 즉, $k(k-1)/2$ 개의 분류기가 있어야 한다. 어떤 클래스인지 분류하기 위해선 트리를 이용하여 k-1회의 토너먼트식 비교가 이루어진다. 전자의 경우, SVM을 학습시키기 위한 다른 클래스의 샘플을 정하는 것이 모호하다는 단점이 있고 후자의 경우, 학습해야할 SVM의 수가 $\alpha(k^2)$ 인 단점이 있다. 또한 두 방법 모두 한 클래스에 속하는 데이터의 수가 적을 때는 노이즈에 강인한 결과를 얻기 힘들다.

이러한 문제점을 해결하기 위해 본 논문에서는 두 데이터 쌍 간의 유사도를 결정하는 함수를 SVM으로 학습시켜 다중 클래스의 분류에 이용하는 방법을 제안한다. 제안한 방법은 1개의 SVM으로 모든 클래스를 분류할 수 있으며, SVM을 학습할 때 사용하는 데이터의 수가 늘어나므로 학습 셋으로부터 분류 시 필요한 통계적 정보를 최대한 활용하는 시스템을 만들 수 있다.

2. SVM을 이용한 다중 분류기

SVM은 두 개의 클래스에 속한 샘플들을 선형 분리하는 고차평면(hyperplane)을 찾는다. 이때 찾아진 고차평면을 결정 경계라 하며 그 결정 경계에 가장 가까운 샘플들이 서포트 벡터(Support Vector)가 된다. 이 고차평면은 서포트 벡터와의 거리가 최대가 되도록 최적화 된

이 논문은 2004년도 경북대학교 학술 진흥 연구비에 의해 연구되었음.

다. 일반적으로 주어진 두 샘플이 선형 분리가 된다면, 수식(1)에서 w 가 결정 경계를 이루는 벡터가 되고, 학습에 의해 찾아지는 계수 α_i 가 0이 아닌 x_i 들이 서포트 벡터가 된다.

$$wx + b = 0, \quad w = \sum_{i=1}^m \alpha_i y_i x_i \quad (1)$$

하지만 주어진 두 샘플이 선형 분리가 안 된다면, 이 두 샘플을 커널 함수 $K(x_i, x_j)$ 를 통해 더 높은 차원으로 매핑시켜 선형 분리가 되도록 만든 다음 결정 경계를 찾는다.

이와 같이 SVM은 두 클래스의 서포트 벡터와 결정 경계의 거리를 최대화 하도록 학습되어 일반화 성능을 높이는 전략을 취한다. 하지만 SVM을 학습시키기 위한 데이터의 수가 충분하지 않다면, 다른 통계적 학습 시스템과 마찬가지로 노이즈에 강인한 분류를 하지 못 할 것이다. 또한 기본적인 SVM은 이진 분류기이므로 다중 클래스 분류를 위해서는 복수개의 SVM을 병합하는 등의 적절한 설계가 이루어져야 한다.

다중 클래스 분류를 위해 SVM을 설계 방법으로 일대다(OVA) 방법과 전부-쌍(all-pairs:AP) 방법이 있다. 그중 가장 널리 쓰이는 방법이 일대다 방법(OVA)으로, 각 클래스 별로 해당 클래스에 속한 원소와 그 외 전부에 속하는 원소들로 이루어진 데이터로 SVM을 학습시킨다. 이 방법의 경우 k개의 클래스를 분류하기 위해 k개의 SVM을 생성해야 한다. 따라서 클래스의 수가 많은 다중 분류 문제의 경우 SVM의 수가 많아져 계산량이 증가하게 된다. 뿐만 아니라, 한 클래스에 속하는 데이터가 많지 않은 경우, 하나의 SVM 학습 시에 분류 대상이 되는 두 그룹(해당 클래스에 속한 데이터 그룹과 나머지 데이터 그룹) 간의 샘플 수가 비정상적인 불균형을 이루어 학습에 나쁜 영향을 미치게 되고, 그 SVM에 해당하는 클래스에 속하는 적은 수의 데이터가 가지는 노이즈에 민감해져 일반화 성능을 저하 시킬 위험을 가진다. 전부-쌍(all-pairs:AP)의 경우도 마찬가지로 클래스의 수가 많고, 클래스내의 데이터 수가 많지 않은 경우는 계산량과 일반화 성능에 있어서 문제점을 가진다.

따라서 본 논문에서는 의료 데이터, 생체인식 데이터, 생물정보 데이터 등과 같이 클래스 수에 비해 상대적으로 한 클래스 내의 데이터 수가 많지 않은 문제에서도 좋은 일반화 성능을 낼 수 있는 노이즈에 강인한 SVM 다중 분류기를 설계하는 방법을 제안한다.

3. 제안하는 방법

본 논문에서 제안하는 방법에서는 먼저 그림1에서와 같이 SVM에 주어질 새로운 입력 데이터를 만든다. 학습 집합 내의 모든 데이터는 자신의 클래스를 포함한 모든 데이터와 쌍을 이루어 두 데이터의 순서쌍 형태로 하나의 새로운 데이터를 만든다. 즉, 새로운 입력 데이터 z_i 는 수식 (2)와 같이 주어진다(k는 클래스의 수).

$$z_i = (x_i, x_j) \quad 1 \leq i, j \leq k \quad (2)$$

이 새로운 입력 z_i 에 대한 출력 값 y_i 는 다음 (3)과 같

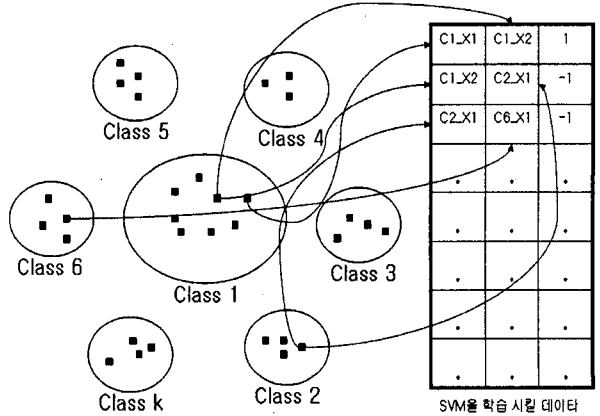


그림 1. 입력 데이터 생성 방법이 할당한다.

$$y_{ij} = \begin{cases} 1 & i=j \\ -1 & i \neq j \end{cases} \quad (3)$$

만약 학습 셋에 k개의 클래스가 존재하고 한 클래스 당 n개의 데이터가 존재한다면 제안하는 방법에서 새롭게 생성된 데이터 셋의 원소의 수는 $kn(kn - 1)$ 개 이다. 이는 기존의 OVA 방법을 쓰는 경우 하나의 SVM에 주어지는 학습 데이터의 수가 kn개인 것을 고려할 때 월등히 많은 수임을 알 수 있다. 따라서 보다 통계적으로 안정적인 학습이 가능해지고, 노이즈에 강인한 분류기를 만들 수 있으리라 기대된다. 또한 생체 데이터와 같이 한 클래스내의 데이터 수 n이 적은 경우 OVA 방법에서 사용하는 SVM이 가지는 일반화 성능 저하의 문제도 극복할 수 있다.

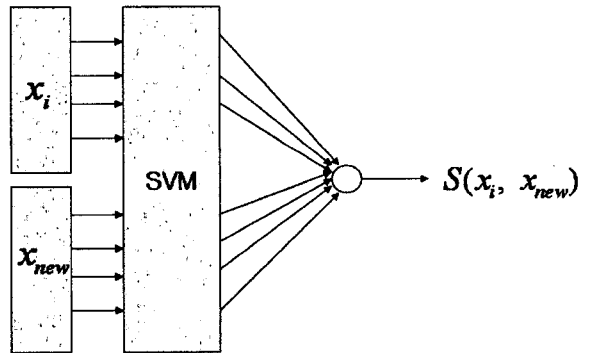


그림 2. 데이터 쌍을 입력으로 받은 SVM

위에서 만든 데이터로 학습한 SVM은 주어진 입력이 같은 클래스 내에 속하는 두 데이터로부터 온 것인지를 판단하는 정보를 제공한다고 볼 수 있다. 즉, 같은 클래스 내의 두 데이터 쌍으로 이루어진 입력은 목표 출력 값이 1이고 각각 다른 클래스로 부터의 데이터 쌍으로 이루어진 입력은 목표 출력 값이 -1이다. 따라서, 입력 $z_i = (x_i, x_j)$ 에 대한 출력 값 $S(x_i, x_j)$ 는 x_i 와 x_j 가 같은 클래스에 속하는지를 판단하는 일종의 유사도 값으로 간주할 수 있다.

이렇게 학습된 SVM이 가지는 또 다른 특징은, SVM이 내는 결과를 유사도 값으로 하여 K-근접이웃방식

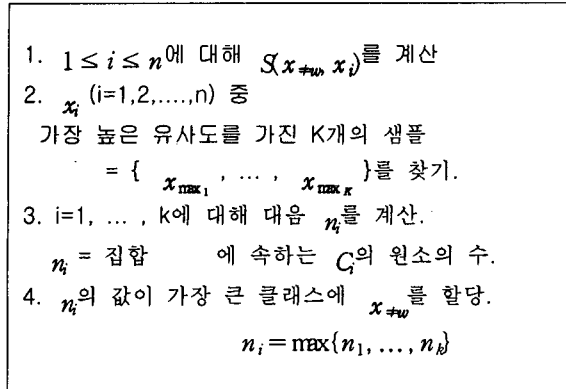


그림 3. 제안하는 SVM을 이용한 분류 방법

(K-NN)과 같은 분류 방법을 접합하여 사용할 수 있다는 점이다. 이때 K 값을 적절히 선택함으로써 분류 성능의 향상을 기대할 수 있다. 이 SVM과 K-NN을 결합한 분류 방법을 정리하면 그림 3과 같다.

4. 실험 결과

제안한 방법이 기존에 알려진 다중 클래스 분류 방법보다 얼마나 일반화 성능을 향상 시키는지를 알아보기 위해 실제 생체 인식 데이터에 대해서 비교 실험을 수행하였다. 생체 인식 데이터는 분류 대상이 되는 개체(즉, 클래스)의 수는 많으나 개체 당 데이터 획득의 비용이 높아 클래스 내의 데이터 수가 상대적으로 적은 문제의 전형적인 예라 할 수 있다. 실험 데이터로는 우선, 14명의 서로 다른 사람으로부터 260개의 홍채 영상을 획득하였다. 그 후, 각 사람의 데이터로부터 무작위로 5개씩을 뽑아 70개를 학습 데이터로 사용하고 나머지 190개는 테스트 데이터로 사용하였다. 이때 SVM에 주어진 입력은 홍채 영상을 그대로 사용하지 않고 주성분 분석(PCA)[5] 방법을 통해 차원을 70 차원으로 줄인 것을 사용하였다[6].

표1은 기존의 OVA 방법과 제안하는 방법의 인식 성능을 비교한 결과이다. SVM의 커널로는 RBF과 다항식(polynomial)을 사용하였으며[4], 커널의 RBF 파라미터인 σ^2 은 20인 경우에 대하여, 또 다항식 파라미터는 2에 대하여 비교 실험하였다.

K=1에서 RBF 커널에서는 두 방법 모두 97.37%로 같은 인식률을 보였다. 그러나 K값이 커짐에 따라 OVA 방법은 인식률이 감소하는데 반해, 제안하는 방법의 경우 K=4인 경우에 최대 98.95%의 인식률을 보였다. 또한 다항식 커널을 사용한 경우에는 K=1인 경우에도 제안하는 방법이 좋은 인식률을 보였고, 두 방법 모두 최고 인식률을 보인 K=2인 경우에 제안하는 방법이 약 5% 높은 인식 성능을 보였다. 이는 제안하는 방법에서 같은 클래스 내의 데이터들 간의 유사도를 높게 측정하도록 SVM이 학습된 것에 기인하는 것으로 볼 수 있다.

표 1. 실험 결과 인식률 표

방법	커널 (파라미터)	K=1	K=2	K=3	K=4	K=5
OVA	Poly(2)	93.16 %	93.16 %	83.68 %	83.68 %	83.68 %
	RBF ($\sigma^2=20$)	97.37 %	97.37 %	87.89 %	87.89 %	87.89 %
제안하는 방법	Poly(2)	97.37 %	98.95 %	98.42 %	98.42 %	98.42 %
	RBF ($\sigma^2=20$)	97.37 %	97.89 %	97.89 %	98.95 %	98.42 %

5. 결론

본 논문에서는 분류 대상 클래스의 수는 많고 클래스 내의 데이터 수가 적은 경우에도 노이즈에 강인한 일반화 능력을 보이는 SVM 다중 클래스 분류기의 설계 방법을 제안하였다. 실험에서 기존의 OVA 보다 적당한 K 값에서 높은 일반화 성능을 보인 것은 데이터의 수가 제한적이어서 SVM을 학습시키기에 불충분할 때, 데이터가 가진 정보를 극대로 활용하여 SVM을 학습시켰기 때문이라고 해석할 수 있다. 본 방법은 생체인식 데이터뿐만 아니라 의료 데이터나 생물정보(Bioinformation) 데이터와 같이 데이터의 수가 제한적인 도메인에서도 마찬가지로 효과적으로 활용될 수 있을 것이다.

참고 문헌

[1] Colin Campbell, An Introduction to Kernel Methods, Radial Basis Function Networks: Design and Applications, Springer Verlag, Berlin, 2000
 [2] Vladimir N. Vapnik, An Overview of Statistical Learning Theory, IEEE Trans. on Neural Networks, Vol 10, No. 5, pp.988-999, 1999
 [3] P.L. Bartlett and J. Shawe-Taylor. Generalization performance of support vector machines and other pattern classifiers. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, Advances in Kernel Methods - Support Vector Learning, pages 43-54, Cambridge, MA, 1999. MIT Press
 [4] Bernhard Schölkopf, Alexander J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning), MIT Press, 2001
 [5] Duda, R.O., P.E. Hart, Pattern Classification and Scene Analysis, New York: Wiley, 1973
 [6] Kwanyong Lee, Hyeyoung Park, A New Similarity Measure Based on Intra-class Statistics for Biometric Systems, ETRI Journal, Vol 25, No. 5, pp.401-406, 2003