

단일부류 최소제곱 서포트 벡터 머신*

우 상호⁰, 이 성환

고려대학교 인공지능연구센터 / 정보통신대학 컴퓨터학과
{shwoo⁰, swlee}@image.korea.ac.kr

One-class Least Square Support Vector Machines

Sang-Ho Woo⁰ and Seong-Whan Lee

Center for Artificial Vision Research / Dept. of Computer Science and Engineering, Korea University

요 약

서포트 벡터 머신은 얼굴인식이나 문자인식과 같은 다양한 패턴인식 문제에서 좋은 성능을 보여준다. 그러나 이러한 문제는 Quadratic Programming(QP) 문제에 관하여 몇 가지 단점을 가지고 있다. 일반적으로 대용량의 QP 문제를 해결하기 위해 많은 계산비용이 요구되며, QP 기반 시스템을 효과적으로 구현하는 것이 쉽지 않은 문제이다. 또한 대규모 데이터의 처리 시에는 입출력을 맞추기 또한 쉽지 않은 단점이 있다. 본 논문에서는 위의 단점을 극복하기 위하여 단일부류 문제를 최소제곱 서포트 벡터 머신을 기반으로 하여 해결하였다. 제안된 방법은 QP 문제를 해결하는 과정이 없이 단일부류 문제를 표현하여 최소제곱 방법을 이용하는 알고리즘이다. 제안된 방법으로 쉽고, 계산 비용을 줄이는 결과를 얻었다. 또한 서포트 벡터 영역 표식자에 확장 적용하여 선형방정식으로 구현하여, 문제를 해결하였다. 제안된 방법의 효율성을 입증하기 위하여 패턴인식 분야 중에 얼굴 인증 방법과 바이오인포매틱스 분야 중에 전립선 암 분류 문제에 적용하였다. 우리의 실험결과는 적합한 성능과 좋은 Equal Error Rate(EER)를 보여준다. 제안된 방법은 알 수 없는 물체의 분류 방법의 효율성을 증대시켰고, 실시간 응용분야에 직접적으로 적용될 수 있을 것으로 기대된다.

1. 서 론

서포트 벡터 머신 방법은 통계적인 방법에 기반한 방법으로 주어진 데이터로부터 학습을 통해서 클래스를 분류하는 방법이다[1]. 이 방법은 특징 추출에 기반한 방법에 비해 수행 속도가 빠르며, 발생 가능한 오류를 줄일 수 있는 장점을 갖고 있다. 또한, 적은 수의 학습 데이터 만으로도 효율적인 학습을 수행할 수 있고, 실용적인 응용 프로그램에 좋은 성능을 보이며, 신경망과는 달리 유일한 해를 구할 수 있는 장점이 있다. 그러나 학습 시에 항상 많은 시간이 소모되어야 하기 때문에 온라인 상에서 실험하기에는 많은 제약조건이 따른다. 이런 문제에 대한 해결책으로 Quadratic Programming(QP) 문제를 사용하는 방법들이 제시되었다. 그 중 단일부류 분류 문제는 Novelty detection이나 Outlier detection이라 불리는 방법으로 클래스들의 경계를 레이블하고 비정상적인 객체를 거부하기 위해 학습을 한다. 만약 단일 부류 분류기가 전형적인 방법의 서포트 벡터 머신처럼 학습을 시킨다면, 학습 시에 계산 복잡도가 커지기 때문에 QP 문제와 관련해서 몇 가지 불이익을 가지게 된다. 본 논문에서는 이러한 단점을 보완하기 위하여 최소제곱의 접근 방법에 의존하여, 단일 부류 문제를 선형방정식의 형태로 변환하여 해결한다. 제안된 방법의 효율성을 보여주기 위해 얼굴인증 방법과 전립선암 세포를 가진 환자의 정상적인 세포를 가진 환자의 분류 실험을 하였다[2].

2. 관련 연구

서포트 벡터 머신을 사용하여 최적의 함수 근사를 구하는 문제는 QP 문제를 푸는 것으로 볼 수 있다. 기존의

QP 문제를 풀기 위해 연구된 방법으로는 Chunking, Decomposition, Sequential Minimal Optimization(SMO) 방법이 있다[3][4][5].

2.1 Chunking 방법

Chunking 방법은 행렬 내에서 Lagrange 승수의 값이 0에 해당하는 행과 열을 제거하여 행렬의 크기를 줄이는 방법으로 대용량의 QP 문제는 작은 QP 문제로 연속적으로 분해가 가능하다. 그러나 줄여진 행렬을 저장장소에 알맞게 고정되지 못하므로, 여전히 대용량의 학습 데이터를 사용하는 문제에는 적용하기 힘들다[3].

2.2 Decomposition 방법

Decomposition 방법은 QP 문제를 매우 작은 세부 문제의 연속으로 분할하는 방법으로 적어도 Kuhn-Tucker 조건[6]을 위반하는 하나의 데이터가 있으면, 전 세부 문제의 데이터에 추가하여 전체적인 목적함수를 줄이고 제약조건의 모든 것을 준수하는 가능한 점을 유지하는 방법이다. 그러므로 적어도 하나의 제약조건을 항상 추가하는 QP 세부 문제의 결과는 점근적으로 수렴하게 된다. 이러한 방법은 항상 QP 세부 문제를 위한 상수크기의 행렬을 유지해야 하며, 모든 단계에서 하나의 데이터를 추가하거나 제거하여야 한다[4].

2.3 Sequential Minimal Optimization 방법

Sequential Minimal Optimization(SMO) 방법은 서포트 벡터 머신의 QP 문제를 풀기 위한 방법으로, 추가적인 행렬의 저장소와 각각의 서브 프로그램을 위한 반복적인 단계가 없이 빨리 해결할 수 있는 단순한 알고리즘 두 개의 알파 값을 푸는 정형적인 방법이다. 전체적으로

* 본 연구는 과학기술부 창의적연구진흥사업의 연구비 지원을 받았음.

수치적인 QP 문제를 최적화 단계를 통해서 전체적인 내부 반복단계를 피하여 적은 양의 저장공간을 필요로 하는 장점이 있다[5].

3. 제안된 단일부류 최소제곱 서포트 벡터 머신

서포트 벡터 머신은 앞에서 소개한 것처럼 일반적으로 QP 문제를 푸는 방법으로 전통적인 신경망에 비해 성능이 우수하다. 그러나 계산 복잡도가 높아 문제점으로 남아있다. 이러한 문제를 해결하기 위해 최소제곱 방법의 단일부류 서포트 벡터 머신을 제안한다.

3.1 C-SVM에 대한 최소제곱 방법

최소제곱 방법으로 단일부류 문제에 적용하였다. 제약 조건을 만족시키기 위한 α_i 값을 투영시켜서 Kuhn-Tucker 조건[6]을 사용한다.

$$\begin{aligned} \min_{w, e} F(w, e) &= \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2 \\ \text{s.t. } [w^T \varphi(x_k)] &= 1 - e_k, k=1, \dots, N \end{aligned} \quad (1)$$

여기서 제약조건 $\alpha_k > 0$ 성립해야 한다. 이런 제약조건을 가진 방정식을 푸는 것은 Lagrange 함수를 구하는 것을 의미한다. Lagrange 함수는 아래와 같이 정의한다.

$$L(w, e; \alpha) = F(w, e) - \sum_{k=1}^N \alpha_k [(w^T \varphi(x_k)) - 1 + e_k] \quad (2)$$

여기서 α_k 는 최적조건으로부터의 Lagrange 승수이다. 이때의 최적성의 조건은 다음과 같다.

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N \alpha_k \varphi(x_k) \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k, & k=1, \dots, N \\ \frac{\partial L}{\partial \alpha_k} = 0 \rightarrow [w^T \varphi(x_k)] - 1 + e_k = 0, & k=1, \dots, N \end{cases} \quad (3)$$

최적성의 조건은 다음의 선형방정식 형태로 표현한다.

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & -\mathbf{Z}^T \\ \mathbf{0} & \gamma \mathbf{I} & -\mathbf{I} & \mathbf{e} \\ \mathbf{Z} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \rho \\ \mathbf{e} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \quad (4)$$

$\mathbf{Z} = [\varphi(x_1)^T \mathbf{K} \varphi(x_N)^T]^T$, $\mathbf{1} = [1 \mathbf{K} 1]$, $\mathbf{e} = [e_1 \mathbf{K} e_N]$, $\mathbf{a} = [\alpha_1 \mathbf{K} \alpha_N]$ 일 때에 식을 다음과 같이 구성할 수 있다.

$$[\mathbf{Z}\mathbf{Z}^T + \gamma^{-1}\mathbf{I}]\mathbf{a} = \mathbf{1} \quad (5)$$

3.2 바이어스를 고려한 C-SVM의 최소제곱 방법

바이어스 ρ 를 변수로 지정하여 위에서 제시한 최소제곱 방법으로 단일부류 문제에 적용하였다. 식 (1)과 같이 제약조건을 만족시키기 위한 α_i 값을 투영시켜서 Kuhn-

Tucker 조건을 사용한다. Lagrange 함수는 다음과 같이 정의한다.

$$L(w, \rho, e; \alpha) = F(w, \rho, e) - \sum_{k=1}^N \alpha_k [(w^T \varphi(x_k)) - \rho + e_k] \quad (6)$$

α_k 는 최적조건으로부터의 랑그랑지 함수이다.

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w - \sum_{k=1}^N \alpha_k \varphi(x_k) = 0 \\ \frac{\partial L}{\partial e} = 0 \rightarrow -\gamma + \sum \alpha_k = 0, & k=1, \dots, N \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow \gamma e_k - \alpha_k = 0, & k=1, \dots, N \\ \frac{\partial L}{\partial \alpha_k} = 0 \rightarrow [w^T \varphi(x_k)] - \rho + e_k = 0, & k=1, \dots, N \end{cases} \quad (7)$$

식 (7)의 결과로 w 와 a 에 대한 식이 $w = \sum_{k=1}^N \alpha_k \varphi(x_k)$ 와 $\alpha_k = \gamma e_k$ 으로 표현되어 지며, 최적성을 위한 조건은 다음의 선형방정식의 형태로 표현한다.

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & -\mathbf{Z}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \gamma \mathbf{I} & -\mathbf{I} \\ \mathbf{Z} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \rho \\ \mathbf{e} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \quad (8)$$

$\mathbf{Z} = [\varphi(x_1)^T \mathbf{K} \varphi(x_N)^T]^T$, $\mathbf{1} = [1 \mathbf{K} 1]$, $\mathbf{e} = [e_1 \mathbf{K} e_N]$, $\mathbf{a} = [\alpha_1 \mathbf{K} \alpha_N]$ 일 때에 식 (8)을 다음과 같이 구성할 수 있다.

$$\begin{cases} \left[\sum_{k=1}^N \alpha_k \varphi(x_k) \right]^T \varphi(x_i) - \rho + e_i = 0 \\ \sum \varphi(x_i) \varphi(x_j) - \rho + \frac{1}{\gamma} \alpha_k = 0 \end{cases}$$

$$\begin{bmatrix} \mathbf{0} & -\mathbf{1}^T \\ -\mathbf{1} & \Omega + \frac{1}{\gamma} \mathbf{I} \end{bmatrix} \begin{bmatrix} \rho \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \gamma \\ \mathbf{0} \end{bmatrix} \quad \begin{cases} \Omega = [\varphi(x_i)^T \varphi(x_j)] \\ \mathbf{1} = [1 \dots 1] \\ \mathbf{a} = [\alpha_1 \dots \alpha_N] \end{cases} \quad (9)$$

식 (9)에서 아래의 식 (10)과 같이 유도된다.

$$\begin{aligned} \mathbf{a} &= [\Omega + \frac{1}{\gamma} \mathbf{I}]^{-1} \gamma \mathbf{1} \\ \rho &= (\mathbf{1}^T [\Omega + \frac{1}{\gamma} \mathbf{I}]^{-1})^{-1} \frac{\gamma}{2} \end{aligned} \quad (10)$$

본 논문에서 단일부류 최소제곱 서포트 벡터 머신 알고리즘은 기존의 QP 문제를 선형방정식의 집합을 푸는 문제로 단순화시켰다. 이를 통해 기존 방법의 서포트 벡터 머신에서의 알고리즘에서 QP 문제를 푸는 과정의 복잡도가 $O(n^3 + n)$ 이었던 것[7]을 선형방정식의 형태로 계산하여 복잡도를 $O(n^2)$ 로 기존의 QP 문제를 푸는 문제보다 줄일 수 있었다.

4. 실험 및 결과분석

본 논문에서는 얼굴 인증 실험과 암 분류 문제에 대한 다음의 두 가지의 실험을 하였다.

4.1 얼굴 인증 실험

전체적인 얼굴 인증 실험 과정은 그림과 같다.

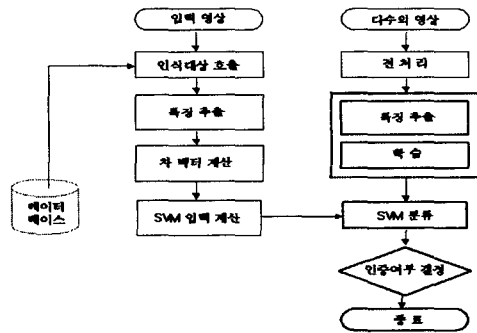


그림 1. 전체적인 얼굴 인증 시스템

서포트 벡터 머신에서의 데이터 집합은 학습데이터 집합과 테스트 데이터 집합으로 구성하였다. 학습 데이터 집합은 총 43명의 각각의 3개의 영상씩 총 126개의 영상에 대하여 학습을 시켰다. 테스트 데이터 집합에서는 일인당 한 영상을 무작위로 구성하여 실험을 하였다.

단일부류 최소제곱 서포트 벡터 머신의 성능 평가를 위해 다양한 문턱치 값에 의해 인증 거부와 잘못되어 거부한 비율이 주어지는 Receiver Operating Characteristic (ROC) 곡선을 고려하였다. 그림 2는 특징 집합에서의 학습 후의 분류된 결과를 보여준다.

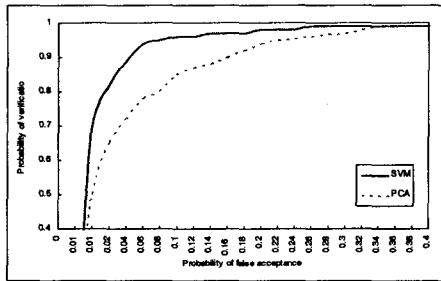


그림 2. ROC for verification.

4.2 전립선 암 분류 실험

전체적인 전립선 암 분류과정은 그림과 같다.

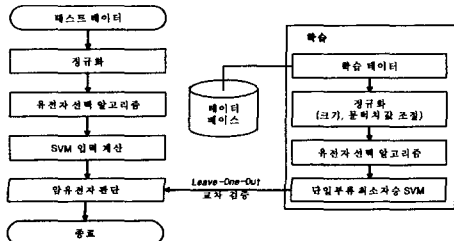


그림 3. 전립선 암분류의 블록 다이어그램

암환자 샘플과 정상 샘플의 분류를 하기 위해 34개의 테스트 데이터(25개의 암 샘플, 9개의 정상 샘플) 이용하였다. 실험의 결과를 비교하기 위해서 기존에 제시한 K-Nearest Neighbor(K-NN) 방법과 비교하였다.

실험결과, 4개의 유전자에서는 82%, 16개의 유전자에서는 92%, 50개의 유전자에서는 94%, 1000개의 유전자에서는 89%의 성능을 보였다. 이 결과는 K-NN 방법이 4개의 유전자에서는 77%, 16개의 유전자에서는 86%를 보인 것과 비교해서 우수한 결과임을 알 수 있다.

표 1. 유전자별 인식률

유전자수	K-NN	LS-SVM
4	77 %	83 %
16	86 %	92 %

실험에서의 같이 특징 정보를 추출하여 우수한 성능을 가진 분류기인 서포트 벡터 머신을 이용한 결과 우수한 성능을 보임을 알 수 있다. 따라서 본 논문에서 제안된 방법을 사용해서 실시간 감시 시스템이나 출입통제 시스템과 같은 보안시스템 및 DNA 데이터를 이용한 유전자 분류 및 각종 장치에 도입할 수 있어, 생체 인식 및 생체정보처리 분야에서 응용할 수 있다.

참고 문헌

- [1] V. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, New York, 1995.
- [2] G. Febbo et al., "Gene Expression Correlates of Clinical Prostate Cancer Behavior," Cancer Cell, pp. 203-209, March 2002.
- [3] Kaufman, "Solving the Quadratic Programming Problem Arising in Support Vector Classification," In Schölkopf, Burges and Smola, Eds., Advances in Kernel Methods, Support Vector Learning, MIT Press, pp. 147-168, 1998.
- [4] T. Joachims, "Making Large-Scale Support Vector Machine Learning Practical," In Schölkopf, Burges and Smola, Eds., Advances in Kernel Methods Support Vector Learning, Cambridge, MIT Press, pp. 169-184, 1998.
- [5] J. Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," In B. Schölkopf, C.J.C. Burges and A.J. Smola, Eds., Advances in Kernel Methods-Support Vector Learning, Cambridge, MIT Press, pp. 185-208, 1999.
- [6] R. Fletcher, Practical Methods of Optimization, Wiley, Chichester and New York, 1987.
- [7] J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Journal of Data Mining and Knowledge Discovery, Vol. 2, No. 2, pp. 121-167, 1998.