

Face image classification by SVM

박혜정¹⁾, 심주용²⁾, 김문태³⁾, 오광식⁴⁾, 김대학⁵⁾

요 약

최근 들어 SVM(support vector machines)은 기계학습의 분야에서 많은 응용이 이루어지고 있으며 특히 분류(classification)나 회귀(regression)분석의 영역에서 많은 연구가 진행중이다. 본 논문에서는 SVM을 이용하여 입력영상자료(image data)를 분류하고자 한다. RGB 컬러 영상자료가 입력되면 이미지 크기에 관계없이 이미지 자체를 입력패턴으로 인식하고 SVM을 통한 훈련(training)을 거친 결과(weight 들과 bias 추정치)를 이용하여 입력영상자료가 사람인가를 분류할 수 있는 문제를 다룬다. 제안된 방법의 타당성은 152개의 영상자료에 적용하여 분석되었다.

주요용어 : SVM, Image data, Classification, Training.

1. 서론

과거 20년 이상 computer vision 영역에서 수많은 얼굴 인식에 관한 논문들이 출판되었다. 수많은 실세계 응용들, 즉 감시(surveillance), 보안접근(secure access), 인간/컴퓨터 환경(human/computer interface)과 칩과 하드웨어의 발전 또한 상업용 얼굴 인식 시스템의 발전을 유도했다. 제한된 시나리오 속에서 이들 시스템들은 다소 성공은 했지만 여전히 얼굴 인식의 중요 문제인 조명, 앞면 얼굴의 표정과 포즈에 관한 변화에 대해서는 여전히 인식하지 못하는 문제로 남아 있다.

본 논문에서는 입력 영상자료가 조명이나 얼굴의 표정과 포즈, 이미지의 크기와는 상관없이 일단 사람인가를 분류해내는 문제를 다루고자 한다. 분류 문제를 위해 기계학습 분야에서 많은 응용이 이루어지고 있으며 특히 Gunn(1988)이 설명한 바 있는 분류(classification)나 회귀(regression)분석의 분야에서 많이 활용되고 있는 SVM방법을 적용하였다.

2. 서포트 벡터 머신(support vector machines)

입력자료에 대한 출력값으로 $\{-1,+1\}$ 과 같이 이진값만이 존재한다면, 자료를 두 그룹으로 분리 가능하다. 즉 두 그룹으로 구분 짓는 초평면(hyperplane)이 존재하게 되며, 다음과 같이 표현할 수 있다.

$$\langle \mathbf{w}, \boldsymbol{\varphi}(\mathbf{x}) \rangle + b = 0 \quad (1)$$

여기서, \mathbf{x} 는 입력 벡터, \mathbf{w} 는 가중치 벡터, b 는 상수, $\boldsymbol{\varphi}(\cdot)$ 은 $\boldsymbol{\varphi}(\mathbf{x}) = \{\varphi_1(\mathbf{x}), \dots, \varphi_{N+1}(\mathbf{x})\}$

-
- 1) 대구가톨릭대학교 전산통계학과 박사과정, 경북 경산시 하양읍 금락리 330
 - 2) 대구가톨릭대학교 정보통계학전공 겸임교수, 경북 경산시 하양읍 금락리 330
 - 3) 대구가톨릭대학교 전산통계학과 박사과정, 경북 경산시 하양읍 금락리 330
 - 4) 대구가톨릭대학교 정보통계학전공 교수, 경북 경산시 하양읍 금락리 330
 - 5) 대구가톨릭대학교 정보통계학전공 교수, 경북 경산시 하양읍 금락리 330

인 커널함수이다. 그러면 SVM방법을 이용한 최적화 문제는 다음과 같은 라그랑즈 함수로 표현할 수 있다.

$$L(\mathbf{w}, b, \xi, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum \xi_i - \sum \alpha_i \{y_i (\langle \mathbf{w}, \varphi(\mathbf{x}_i) \rangle + b) - 1 + \xi_i\} \quad (2)$$

여기서, $\alpha_i > 0$ 는 라그랑즈 배수(lagrange multiplier)이다. 이를 이용하여 Vapnik(1995,1998)은 다음의 쌍대문제(dual problem)

$$\max_{\alpha \geq 0} \left\{ \min_{\mathbf{w}, b, \xi} L(\mathbf{w}, b, \xi, \alpha) \right\} \quad (3)$$

로 최적화의 해를 구하였다. $\min_{\mathbf{w}, b, \xi} L(\mathbf{w}, b, \xi, \alpha)$ 이 최적값을 가질 조건은 편미분을 하여 구한 값을 식(3)에 대입하면 다음과 같은식으로 나타낼 수 있다..

$$\max_{\alpha} \left\{ \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle \right\}, \sum_{i=1}^{\ell} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C \quad (4)$$

따라서, 해 α 에 대응하는 최적의 \mathbf{w}^*, b^* 는 다음과 같이 계산할 수 있다.

$$\mathbf{w}^* = \sum_{i=1}^{\ell} \alpha_i^* y_i \varphi(\mathbf{x}_i), b^* = y_i - \sum_{i=1}^{\ell} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}_o) \quad (5)$$

단, 임의의 서포트 벡터 \mathbf{x}_i 에 대해 $0 < \alpha_o^* < C$ 이며 b^* 를 계산하는데는 Karush-Kuhn-Tucker(KKT, Smola와 Scholkopf(1998))조건이 사용된다. KKT조건에 따르면 임의의 0이 아닌 α_o^* 에 대해서 다음 식이 성립한다.

$$\alpha_o^* [y_i (\langle \mathbf{w}^*, \varphi(\mathbf{x}_o) \rangle + b^*) - 1 + \xi_i^*] = 0, i=1, 2, \dots, \ell \quad (6)$$

만일, $\alpha_o^* \leq C$ 이면 ξ_i^* 는 0이 된다. 따라서 $0 \leq \alpha_o^* \leq C$ 인 \mathbf{x}_o 를 선택하면 위 조건에 따라

$$y_i (\langle \mathbf{w}^*, \varphi(\mathbf{x}_o) \rangle + b^*) - 1 = 0 \quad (7)$$

이 된다. 따라서 b^* 는 위의 식처럼 구할 수 있게 된다. 따라서 입력자료에 대한 분류는 다음 함수를 이용하며 처리된다.

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^{\ell} y_i \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b^* \right) \quad (8)$$

여기서 $\text{sign}(\cdot)$ 는 \cdot 이 양수이면 1, 그렇지 않으면 -1을 가지는 함수이다.

고차원 문제 공간으로의 변환에 이용되는 커널 함수는 Mercer의 이론을 만족하는 함수를 이용하여 해결할 수 있다. 그런 함수들로는 자유도 q 를 가지는 다항식, 가우시안, 이계층 퍼셉트론등이 있다. 본 실험에서는 자유도 2인 다항식 커널 함수를 이용하였다. Osuna, Freund와 Girosi(1997)의 연구에 의하면 식(4)을 최대화하는 것은 고차원의 특징공간에서 단지 그들의 내적인 $\langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$ 의 크기만 필요로 하게 되어 고차원일 경우의 계산상의 어려움을 덜 수 있다. 따라서 그 내적을 표현하는 함수를 커널 함수 $K(\mathbf{x}_i, \mathbf{x}_j)$ 로 표현하면 다음과 같다.

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &\equiv \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle = \langle \varphi(\mathbf{x}_j), \varphi(\mathbf{x}_i) \rangle \\ &= \sum_{k=1}^{\infty} \alpha_k^2 \langle \varphi_k(\mathbf{x}_i), \varphi_k(\mathbf{x}_j) \rangle \end{aligned} \quad (9)$$

3. 영상자료에서의 분류

얼굴 인식 방법은 크게 두가지 방법으로 분류될 수 있다. 한 방법은 영상자료에서 얼굴을 인

식할 수 있는 특징들만 뽑아서 분류하는 특징 기반 접근(component-based approach)과 다른 방법은 영상자료 전체를 입력으로 받아 들여서 분류하는 전역 접근(global approach)방법이다.

전역 접근방법은 단일 특징 벡터 즉 얼굴 전체가 분류자(classifier)의 입력자료이다. 몇몇 분류방법들은 최소 거리 분류법(minimum distance classification in the eigenspace, Sirovitch와 Kirby(1987)), 피셔의 판별 분석(Fisher's discriminant analysis, Belhumeur등(1997)), 신경망(neural network, Fleming과 Cottrell(1990))등이 있다. 전역 접근방법은 정면 얼굴에 대해서는 매우 잘 분류를 한다. 그러나, 위의 분류방법들은 전역 특징들이 얼굴의 변화와 찍는 각도에 너무나 민감하기 때문에 포즈의 변화에 로버스트하지 않다. 이런 문제에 대하여 기계학습 분야에서 많은 연구가 이루어지고 있다. 특히 분류(classification)나 회귀(regression)분석에 많이 적용되고 있는 SVM방법을 영상자료의 분류에 활용하고자 한다.

SVM에 영상자료를 적용하기 전에 조명이나 얼굴의 표정과 포즈, 이미지의 크기와는 무관히 사람인가를 분류해내는 문제를 다루기 위해 RGB 컬러 영상자료를 입력으로 받아들였다. 영상자료를 입력자료로 받아들일 때 영상자료의 크기에 구애를 받지 않기 위해 영상자료를 입력 받은 후 이미지의 규격을 통일되게 규격화를 시켰다. 규격화된 영상의 크기는 입력된 영상자료의 정보를 40×40 화소(pixel) 영상자료로 축소(image resize)를 하여 통일시켰다.

이미지를 확대, 축소를 하기 위해서 사용되는 여러 가지 보간법 중에 본 논문에서는 가장 쉬운 인접 화소로 이미지를 보간하는 최근접화소보간법(nearest neighbor interpolation: 최근접법)이 사용하여 입력 영상자료를 축소시켰다.

축소된 영상자료를 다시 최소 분산 양자화(minimum variance quantization)를 사용하여 영상자료의 각 픽셀(화소)의 농도를 이산적인 정수 값으로 변환하였다. 이 정수변환을 통해 3차원의 RGB 컬러 영상자료를 2차원의 index color로 영상자료를 변환을 하여 영상의 정보를 256색 컬러 정수형으로 변환을 하였다. 이 정수형으로 변환된 영상자료는 분류를 위해 축약을 하였다.

축약된 이미지를 그레이 스케일 레벨의 영상자료로 변환을 하였으며, 이 변환된 영상자료는 전역접근방법의 입력 패턴으로 만들기 위해서 각 영상자료를 하나의 행벡터(1600)로 변환 하여 하나의 입력패턴으로 만들었다. 이 변환된 영상자료를 SVM방법을 적용하여 훈련시켰다.

4. 영상자료의 분류 모의실험

모의실험을 위해 RGB 컬러 영상자료 152개를 준비하였다. 152개의 영상자료 중에서 훈련용(training)으로 76개(50%)를 테스트용으로 76개(50%)의 영상으로 분할하였다. RGB 컬러 영상자료들을 SVM방법으로 적용하기 전에 세 단계의 선행처리과정을 고려하였다. 첫 번째로 입력으로 들어온 임의의 크기의 RGB 컬러 영상자료들에 대해서 규격화를 위해 영상자료의 정보를 40×40 화소 영상자료로 축소(image resize) 한 후 다시 RGB 3차원 영상자료를 2차원으로 즉 256색 이미지로 다시 축약하였다. 두 번째로 칼라 영상들을 그레이 스케일 레벨 색상의 영상으로 변환을 하였으며, 세 번째로 그레이 스케일 레벨 영상을 다시 이미지 전체를 하나의 패턴으로 인식이 되도록 행벡터로 변환을 하였다. 여기서 훈련용으로 준비한 영상자료를 이용하여 선행처리단계를 거친 다음에 이 변환된 영상자료를 SVM의 입력자료로 읽어들이고 후 SVM방법으로 훈련하여 결과(weight 들과 bias 추정치)를 구하였다. 이 구한 결과들을 이용하여 테스트용으로 준비한 영상자료들에 적용하여 테스트 영상자료들이 사람인가를 분류하였다. 실험을 위해 준비한 자료의 예는 다음 [그림1]과 같다.



[그림1] 영상자료의 일부분

[그림1]에서 위의 그림들은 훈련용 영상자료들의 예제들이며, 아래쪽의 그림들은 테스트를 위해 준비한 영상자료들의 예제들이다. SVM을 위해 사용된 커널은 일반 영상 처리를 위해 자주 사용되는 다항 커널함수(polynomial kernel)를 이용하였으며, 커널에 사용된 모수는 $d=2$, $p=1$ 이다. SVM에 사용될 일반화 모수 C 값은 10 fold 교차타당성(cross-validation) 방법을 이용하여 최적 값으로 $C=10$ 을 얻었으며, 그 결과 SV(Support Vector)로서 36개(훈련용 자료의 47.4%)가 선택되었으며, 이 선택된 SV에 의해 가중치 값들이 구해 졌다. 아래의 표는 훈련 후 구해진 값들을 이용하여 test 영상자료에 적용한 결과이다.

[표1] 영상 자료의 분류 결과

	정분류	오분류	전체 영상자료 수
test 자료	72	4	76
비율	94.74 %	5.26 %	

[표1]에서 볼 수 있듯이 영상전체를 하나의 패턴으로 인식하여 분류한 결과 분류율이 94.74%로 분류가 잘 되었다. 여기서 오분류된 영상자료들을 살펴보면 대부분이 꽃에 대한 영상자료로서 잘 분류가 되지 않았다. 꽃에 대한 종류들이 다양하여 전체적으로 꽃에 대한 패턴은 인식이 되지 않은 것으로 보인다. 그 외에 다른 영상 자료들에 대해서는 분류가 잘 되었다. 훈련을 통해서 얻어진 가중치값들을 이용하여 다시 훈련 자료를 대상으로 분류해 본 결과 오분류된 비율은 0%이다.

5. 결론 및 토의

본 논문에서는 영상자료가 조명이나 얼굴의 표정과 포즈, 이미지의 크기에 구애받지 않기 위해 RGB 컬러 영상자료를 입력으로 받아 이미지 전체를 입력패턴으로 인식하고 SVM을 이용하여 분류하는 방법을 연구하였다. 그 방법으로 SVM을 통하여 얻어진 가중치들과 편의(bias)의 추정값들을 이용하여 분류를 결정 짓는 초평면을 구하고 모의 영상자료에 적용해 본 결과 상당히 좋은 결과를 나타내었다. 얼굴 인식의 영역에 SVM방법을 적용하는 연구가 기대된다.

참고문헌

Belhumeur, P., Hespanha, P., and Kriegman, D.(1997), Eigenfaces vs fisherfaces: recognition

- using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19. 7., 711-720
- Fleming, M., and Cottrell, G.(1990), Categorization of faces using unsupervised feature extraction, In Proc, *IEEE IJCNN International Joint Conference on Neural Networks*, pages 65-70.
- Gunn, S.(1998), Support Vector Machines for Classification and Regression, *ISIS Technical Report, Univ. of southampton*.
- Osuna, E. E., Freund, R. and Girosi, R.(1997), Support vector machines: Training and applications, *Massachusetts Institutes of Technology*, A.I. Memo No. 1602 C.B.C.L Paper No. 144
- Sirovitch, L., and Kirby, M.(1987), Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*. 2. 519-524
- Smola, A., and Scholkopf, B.(1998), On a Kernel-Based Method for Pattern Recognition, Regression, Approximation and Operator Inversion. *Algorithmica*, 22, 211-231.
- Vapnik. V.(1995), *The Nature of Statistical Learning Theory*, Springer, N.Y., ISBN 0-387-94559-8.
- Vapnik. V.(1998), *Statistical Learning Theory*, Springer, N.Y.