회귀문제를 위한 비선형 특징 추출 방법

*김성민 **곽노준
아주대학교 전자공학과
*koolland@ajou.ac.kr **nojunk@ajou.ac.kr

Nonlinear feature extraction for regression problems

*Kim, Seongmin **Kwak, Nojun
Division of Electrical & Computer Engineering, Ajou University

요약

본 논문에서는 회귀문제를 위한 비선형 특징 추출방법을 제안하고 분류문제에 적용한다. 이 방법은 이미 제안된 선형판별분석법을 회귀문제에 적용한 회귀선형판별분석법(Linear Discriminant Analysis for regression:LDAr)을 비선형 문제에 대해 확장한 것이다. 본 논문에서는 이를 위해 커널함수를 이용하여 비선형 문제로 확장하였다. 기본적인 아이디어는 입력 특징 공간을 커널 함수를 이용하여 새로운 고차원의 특징 공간으로 확장을 한 후, 샘플 간의 거리가 큰 것과 작은 것의 비율을 최대화하는 것이다. 일반적으로 얼굴 인식과 같은 응용 분야에서 얼굴의 크기, 회전과 같은 것들은 회귀문제에 있어서 비선형적이며 복잡한 문제로 인식되고 있다. 본 논문에서는 회귀 문제에 대한 간단한 실험을 수행하였으며 회귀선형판별분석법(LDAr)을 이용한 결과보다 향상된 결과를 얻을 수 있었다.

1. 서론

최근 패턴 인식에서 고차원의 입력 데이터에 대한 변환들에 대한 알고리즘의 성능이 향상되고 있다. 일반적으로 특징 추출을 위한 차원 감소와 관련된 방법들은 선형 공간에서 수행하게 되는데, 특히 얼굴 인식과 같은 응용 분야의 경우 대부분의 인식 대상이 실제로는 조명, 포즈, 크기, 표정 변화 등에 의해 비선형의 특징을 가지게 된다. 이러한 비선형 데이터에 대하여 선형 공간에서 특징 추출을 하는 데에는 제한된 결과를 얻을 수밖에 없다. 이러한 문제를 해결하고자 비선형 패턴의 데이터를 선형을 이룰 수 있는 임의의 특징 공간으로 사상함으로써 선형 변환과 동일한 효과를 얻도록 한다. 이렇게 사상된 새로운 공간은 원래의 입력 공간에 비해 고차원의 공간으로 비선형의 특징을 가진 입력 데이터들을 선형 특징으로 이룰 수 있는 공간을 의미한다. 이러한 특징 공간은 무한대 차원으로 확장될 수 있을 만큼 고차원이기 때문에 입력 공간에서 보다 많은 정보를 갖는 축이 상당히 많아짐을 의미한다. 그러나 이러한 무한대 차원에 가까운 특징 공간에서 사상 함수(Φ)를 찾는 것은 거의 불가능하다. 보통 데이터를 직접 특징 공간으로 사상하는 대신 특징공간에서 데이터 간 내적 연산을 통한 커널 함수를 이용하여 고차원의 공간으로 사상하게 된다.

패턴 인식, 기계 학습, 통계학 분야에서 회귀문제는 지속적인 관심의 대상이다. 회귀는 유한한 잡음들의 표본에서 정보가 있는 데이터의 함수를 예측하기 위한 방법이다. 데이터 분류에 있어서 회귀는 지도 학습에 해당한다. 지도학습은 훈련 데이터로부터 함수를 만들어내는 기계학습의 기술이다. 훈련 데이터는 입력 대상의 쌍과 원하는 출력으로 구성된다. 이 논문에서는 입력 대상과 원하는 출력에 대해 각각 입력 특징과 목표변수라 명하였다. 본 논문에 앞서 선형판별분석법을 이용하여 회귀문제에 적용한 방법(LDAr)[1]을 간단히 살펴보면 다음과 같다. N개의 입력특징과 원하는 목표변수의 쌍 $\{x_i, y_i\}$에 대해 특징값들의 집합 $f's(=w_i^T x)$을 구한다. 여기서 입력 데이터 x의 선형변환은 목표변수 y의 정보들을 반영하게 된다. 이렇게 구성된 특징값들의 집합에 대하여 Fisher's criterion[2]을 회귀문제에 적용할 수 있도록 변경을 한 후 고유값 문제를 풀어서 새로운 특징값을 추출한다.

2. 회귀문제를 위한 비선형 특징 추출(Generalized Discriminent Analysis for regression:GDAr)

LDAr을 비선형 공간으로 확장하기 위해서는 주어진 N개의 입력 특징과 원하는 목표변수들의 쌍 $\{x_i, y_i\} \in R^m$에 대하여 입력특징 $x_i(\in R^m)$을 비선형 사상 함수를 이용하여 고차원차원의 공간 (R^F)으로 사상한다.[3] 즉,

$$\Phi(x): R^m \rightarrow R^F,\ F > m \tag{1}$$

여기서 입력특징은 표준화(평균이 0이고, 분산이 1)[4] 되었다고 가정

을 한다. 여기서 주목할 것은 새로운 특징 공간의 차원은 무한대일 수 도 있다. 이렇게 새로운 고차원의 공간으로 사상된 입력특징을 기존 LDA에서와 같이 클래스간 분산과 클래스내 분산을 각각 다음과 같이 구한다.

$$S_{wr}^{\Phi}=\frac{1}{n_b}\sum_{(i,j)\in A_{br}} f(y_i-y_j)\left[\Phi(x_i)-\Phi(x_j)\right]\left[\Phi(x_i)-\Phi(x_j)\right]^T$$

(2)

$$S_{br}^{\Phi}=\frac{1}{n_b}\sum_{(i,j)\in A_{br}} f(y_i-y_j)\left[\Phi(x_i)-\Phi(x_j)\right]\left[\Phi(x_i)-\Phi(x_j)\right]^T$$

(3)

위 식에서 A_{br} 과 A_{wr} 은 목표변수들 간의 거리에 의해 정의[1]가 된다. 즉,

$$A_{br}=\{(i,j)|\ |y_i-y_j|\geq\tau, i<j\} \quad (4)$$
$$A_{wr}=\{(i,j)|\ |y_i-y_j|<\tau, i<j\} \quad (5)$$

일반적으로 분류 문제에 있어서 목표변수들은 이산값을 가지게 된다. 하지만, 회귀 문제에서는 목표변수들이 연속값으로 존재한다. 그러기 때문에 이산 목표변수들을 회귀 문제에 적용하기 위해서 soft class의 개념과 가중치 함수 $f(\bullet)$ 이용하여 목표변수들의 거리를 이용하여 가까운 거리에 속하는 입력특징은 동일한 클래스로 먼 거리에 속하는 입력특징은 다른 클래스로 구분을 한다[1]. 하지만, 본 논문에서는 가중치 함수를 목표변수들 간의 거리를 이용하지 않고, 데이터의 인덱스간 거리의 절대값을 τ와 비교하여 클래스를 결정하였다.

$$A_{br}=\{(i,j)|\ |i-j|\geq\tau\} \quad (6)$$
$$A_{wr}=\{(i,j)|\ |i-j|<\tau\} \quad (7)$$

또한 식(2), (3)을 정리하기 위해 커널 함수를 다음과 같이 정의 한다.

$$K=\left[k(x_i,x_j)\right]=\Phi(x_i)^T\Phi(x_j) \quad (8)$$

즉, $N\times N$ 크기의 K 행렬을 얻을 수 있다. 식(8)을 이용하여 고차원의 특징 공간에서의 클래스 내 분산과 클래스간 분산을 다음과 같이 정리를 한다.

$$S_{wr}^{\Phi}=\frac{2}{n_w}\Phi(X)(D_w-W_w)\Phi(X)^T \quad (9)$$

$$S_{br}^{\Phi}=\frac{2}{n_b}\Phi(X)(D_b-W_b)\Phi(X)^T \quad (10)$$

위 식에서 W_w, W_b, D_w, 그리고 D_b 는 각각 다음과 같이 얻게 된다.

$$W_w=[w_{ij}]=\begin{cases} f(i-j),(i,j)\in A_{wr} \\ 0, otherwise\end{cases} \in R^{N\times N} \quad (11)$$

$$D_w=diag(d_1,\cdots,d_N), d_i=\sum_{j=1}^{N} w_{ij} \quad (12)$$

$$W_b=[v_{ij}]=\begin{cases} f(i-j),(i,j)\in A_{br} \\ 0, otherwise\end{cases} \in R^{N\times N} \quad (13)$$

$$D_b=diag(b_1,\cdots b_N), b_i=\sum_{j=1}^{N} v_{ij} \quad (14)$$

이렇게 고차원 공간에서 정의된 클래스내 분산과 클래스간 분산을 이용하여 Fisher's criterion을 다음과 같이 정리할 수 있다.

$$J(V)=\frac{|V^T S_{br}^{\Phi} V|}{|V^T S_{wr}^{\Phi} V|} \quad (15)$$

일반적으로 특징공간에서의 투영행렬 V는 학습 데이터의 선형 변환을 의미한다. 즉,

$$V=\sum_{i=1}^{N}\alpha_i\Phi(x_i)=\Phi(X)\alpha\ , \alpha\in R^{N\times 1} \quad (16)$$

결과적으로 식(13)의 분모, 분자를 각각 다음과 같이 정리를 할 수 있다..

$$V^T S_{br}^{\Phi} V=\alpha^T K(D_b-W_b)K^T\alpha \quad (17)$$
$$V^T S_{wr}^{\Phi} V=\alpha^T K(D_w-W_w)K^T\alpha \quad (18)$$

여기서 $S_B^K=K(D_b-W_b)K^T$와 $S_W^K=K(D_w-W_w)K^T$로 치환을 한 후, 각각에 대해서 일반화된 특이값 분해(generalized singular value decomposition)을 하여 α를 구한다.

3. 실험 결과 및 분석

본 논문에서 제안하는 알고리즘을 이용한 실험을 위하여 몇 가지 고려해야할 것들이 있다. 우선, 입력특징들을 고차원의 공간으로 사상하기 위해서 커널 함수를 정의를 해야 한다. 본 논문에서는 여러 가지 커널 함수 중 가우시안 RBF(Gaussian Radial Basis Function)[5]를 선택하였다.

$$k(x,y)=\exp\left(\frac{-|x-y|^2}{c}\right) \quad (19)$$

커널 함수에서 커널 상수(c)는 경험적인 방법을 통하여서 적절한 값을 찾게 된다. 다음으로 입력특징들의 클래스를 나누기 위해서 사용되는 경계값(τ)은 입력특징의 총수*0.05로 이 값 역시 경험적인 방법을 통해서 설정을 하였다. 마지막으로 회귀문제를 풀기 위하여 5-최근접 이웃 알고리즘(5-nearest neighbor)를 이용하였다. 본 논문에서는 제안하는 방법을 실험하기 위해 임의의 1000개의 데이터를 생성하였다. 이 데이터에 대하여 선형, 비선형 회귀문제(식 18, 19)를 비선형 특징 추출 방법에 적용하였으며 이미 제안된 LDAr과의 결과를 비교하였다.

$$y = 2x_1 + x_2 \quad (20)$$

$$y = 4(x_1 - 2x_2)^2 + (2x_1 + x_2)^2 \quad (21)$$

우선, 제안한 비선형 특징 추출방법에 대하여 커널 상수값과 추출한 특징의 개수에 따른 입력특징들의 평균 제곱합 제곱근 오차(root mean square error)를 비교하였다. 선형 데이터들에 대한 평균 제곱합 제곱근 오차는 특징수의 개수와 커널 상수에 상관없이 일정한 값(≈ 0.15)이 나옴을 확인하였다. 이와는 다르게 비선형 데이터에 대한 평균 제곱합 제곱근 오차는 커널 상수가 5일 때 0.12로 가장 작게 나옴을 확인하였다.

또한 본 논문에서는 기존의 LDAr을 이용한 특징 추출의 개수에 따른 평균 제곱합 제곱근 오차를 GDAr의 결과와 각각 비교하였다.

특징수	1	2	3	4	5
LDAr	0.15	0.17	0.18	0.20	0.20
GDAr	0.15	0.15	0.15	0.15	0.15

표 1. 선형데이터에 대한 평균제곱합 제곱근 오차

특징수	1	2	3	4	5
LDAr	0.47	0.44	0.37	0.38	0.44
GDAr	0.12	0.12	0.12	0.12	0.12

표 2. 비선형데이터에 대한 평균제곱합 제곱근 오차

표 1, 2는 선형데이터와 비선형데이터에 대한 결과를 각각 나타냈으며, GDAr의 결과는 선형 데이터와 비선형 데이터에 대하여 커널 상수를 각각 1, 5일 때의 결과이다. 위 표에서 확인 할 수 있듯이 커널 상수의 값을 적절히 찾았을 경우 GDAr의 성능이 향상됨을 확인하였다.

4. 결론

회귀문제에 대한 비선형 특징 추출 방법은 선형 특징 추출 방법의 한계를 보완할 수 있다는 사실에 기반한다. 특 패턴 인식에서 얼굴 인식과 같은 응용 분야에서 얼굴의 크기, 회전, 등은 비선형의 특징을 가지기 때문에 선형 특징 추출 방법을 적용하기가 어렵다. 또한 클래스 정보를 특징 추출에 이용함으로써 회귀 문제를 좀 더 정확히 풀수 있음을 보였다. 본 논문에서 제안한 방법을 얼굴 인식과 같은 분야로 확장하여 인식률을 향상시킬 수 있을 것으로 기대한다.

참고 문헌

[1] Nojun Kwak, Jung-Won Lee, "Feature Extraction based on Subspace Methods for Regression Problems," Neurocomputing, vol. 73, issues 10-12, pp. 1740-1751, 2010.

[2] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd Edition, Academic Press, 1990.

[3] Ming-Hsuan Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods" Preceedings of the Fifth international conference on Automatic Face and Gesture Recognition, pp. 215-220, 2002.

[4] Bernhard Schölkopf, Alexander Smola, Klaus-Robert Müller, "Nonlinear Component Analysis as a Kernel Eignevalue Problem" Neural Computation MIT. July 1, 1998, Vol. 10, No. 5, pp. 1299-1319

[5] Klaus-Robert Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, Bernhard Schölkopf, " An Introduction to Kernel-Based Learning Algorithms" IEEE Transactions on Neural networks. pp. 181-201, 2001.