

음성 인식을 위한 다양한 특징 파라미터의 SVM 인식 성능 비교

김평환*, 박정원*, 김창근*, 이광석**, 허강인*
 동아대학교 전자공학과*
 진주산업대학교 전자공학과**

Recognition Performance Comparison to Various Features for Speech Recognizer Using Support Vector Machine

Pyoung-Hwan Kim*, Jeong-Won Park*, Chang-Keun Kim*, Gwang-Seok Lee**, Kang-In Hur*

Dept. of Electronic Engineering, Dong-A University*

Dept. of Electronic Engineering, Jinju National University**

e-mail : jwpark@donga.ac.kr

요약

본 논문은 SVM(support vector machine)을 이용한 음성인식에 대해 효과적인 특징 파라미터를 제안한다.

SVM은 특징 공간에서 비선형 경계를 찾아 분류하는 방법으로 적은 학습 데이터에서도 좋은 분류 성능을 나타낸다고 알려져 있으며 최적의 특징 파라미터를 선택하기 위해 본 논문에서는 SVM을 이용한 음성인식기를 사용하여 PCA(principal component analysis), ICA(independent component analysis) 알고리즘을 적용하여 MFCC(mel frequency cepstrum coefficient)의 특징 공간을 변화시키면서 각각의 인식 성능을 비교 검토하였다. 실험 결과 ICA에 의한 특징 파라미터가 가장 우수한 성능을 나타내었으며 특징 공간에서 각 클래스의 분포도 또한 ICA가 가장 높은 선형 분별성을 나타내었다.

I. 서론

음성인식성능에 있어서 고려되어야 할 사항은 효과적인 인식기 및 특징 파라미터의 선택이다.

현재 대부분의 음성인식 시스템에서는 인식 알고리즘으로 HMM를 사용하고 있다. HMM은 시계열에서 변화하는 데이터에 대해 통계적으로 모델링 함으로써 대부분의 응용에서 좋은 성능을 나타내고 있다. 그러나 학습데이터가 불충분 할 경우 인식 성능이 급격히 저하되는 문제점이 있다. 최근 활발히 연구가 진행되고 있는 분별적인 비선형 패턴 분류방법의 일종인 SVM(Support Vector Machine)은 학습 데이터가 적은 경우에서도 인식성능의

저하가 없이 빠르게 패턴 분류가 가능하다.

본 논문에서는 SVM 인식기에 대해 입력 특징 파라미터로 일반적으로 사용되고 있는 MFCC외에 독립성분분석과 주성분분석에 의한 MFCC의 특징공간을 변화시킨 파라미터를 가지고 인식 실험을 하여 각각의 특징에 대한 인식 성능과 특징을 비교 분석하였다.

본 논문의 구성은 2장에서 4장까지 SVM, ICA, PCA 각각에 대한 알고리즘을 설명하였고 5장에서는 실험결과를 기술하였으며 6장에서 결론을 맺는다.

II. Support Vector Machine

1. 선형 SVM

선형 SVM은 선형분리가 가능한 입력 패턴에 대해 두 개의 클래스를 구분할 수 있는 초평면(Hyperplane)은 무수히 많으나 그림 1에서처럼 두 클래스간 점들의 거리 ρ 를 최대화하는 OSH(Optimal Separating Hyperplane)는 유일한 해로 존재한다. 다시 말해, OSH는 거리 ρ 를 최대화하며 식(1)(2)를 만족하는 최적의 가중벡터 w_* 와 바이아스 b_* 로서 얻어질 수 있다.

$$w^T \cdot x_i + b \geq 1 \quad \text{for } d_i = 1 \quad (1)$$

$$w^T \cdot x_i + b \leq -1 \quad \text{for } d_i = -1 \quad (2)$$

w 의 비용함수(cost function, $1/2 w^T w$)에 대한 제한된 최소화법(constrained minimization)을 해결하기 위해

QP(quadratic programming)로 w_* 와 b_* 를 구할 수 있다.

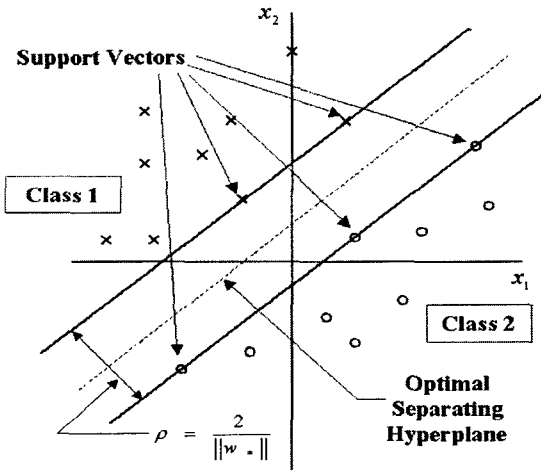


그림 3. OSH for Linear SVM

QP는 식(3)의 Lagrangian 함수의 안장점(saddle point)를 찾는 문제로 편미분과 Kuhn-Tucker 조건에 의해 식(4)를 최대화하는 α (Lagrange multiplier)를 찾는 문제로 바꿀 수 있다.

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [d_i (w^T x_i + b) - 1] \quad (3)$$

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j \quad (4)$$

식(4)를 최대화하는 α_* 를 구한 다음 식(5)와 (6)에 의해 w_* , b_* 를 구한다.

$$w_* = \sum_{i=1}^N \alpha_{*,i} d_i x_i \quad (5)$$

$$b_* = 1 - w_*^T x^{(s)} \text{ for } d^{(s)} = 1 \quad (6)$$

여기서 $x^{(s)}$ 는 그림 1의 Support Vectors이다.

결론적으로 임의의 입력 패턴 x 는 식(7)의 $f(x)$ 에 의해 분류된다.

$$f(x) = \text{sign}(w_*^T x + b_*) \quad (7)$$

2. 비선형 SVM

입력 패턴이 선형 분리가 불가능한 비선형 SVM의 경우에는 Soft Margin ξ_i 과 커널(kernel) 함수 $K(x, x_j)$ 에 의해 선형분류가 가능한 고차원 공간으로 확장하여 OSH를 찾는다.

$$d_i (w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \text{ for } i = 1, 2, \dots, N \quad (8)$$

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (9)$$

커널 함수로는 보통 다항식, RBFN 커널이 주로 사용되며 본 논문에서는 RBFN 커널을 사용하였고 OPC(One-Per-Class)를 사용하여 멀티 클래스로 확장하였다.

III. Principal Component Analysis

특징추출(feature extraction)은 데이터 분류(data classification) 또는 패턴 인식(pattern recognition)의 중요한 문제이다. PCA(Principal Component Analysis)는 입력의 선형성과 특성 식별을 이용하여 데이터의 차원을 축소하며 특징을 추출하는 방법 중 널리 이용되어지는 것 중 하나로, 패턴인식이나 영상처리에서 Karhunen-Loeve 변환으로 잘 알려져 있다.

zero-mean 특성이 있는 n 차원 신호 x 에 대해 차원 축소는 식(10)과 같다.

$$\begin{aligned} a &= [a_1, a_2, \dots, a_l]^T \\ &= [x^T q_1, x^T q_2, \dots, x^T q_l] = Q^T x \quad (10) \end{aligned}$$

여기서 $n > l$ 이며 a 는 x 에 대해 l 차원으로 축소된 신호이다. 그리고 축소된 신호에서 원 신호로의 복원은 식(11)과 같다.

$$\hat{x} = Qa = \sum_{j=1}^l a_j q_j \quad (11)$$

식(10)(11)에서 a 와 Q 는 x 의 공분산 행렬(covariance matrix) R 의 고유치와 고유벡터이다.

차원 축소된 신호 $a = [a_1, a_2, \dots, a_l]$ 는 중요도 순으로 나열된 n 개의 고유벡터 $Q = [q_1, q_2, \dots, q_n]$ 중에 l 개만 사용하여 얻을 수 있다

본 논문에서는 PCA를 신경망으로 구현한 GHA(Generalized Hebbian Learning)를 이용하였다. GHA는 순방향 단층구조의 신경망으로 자율학습특성이 있으며 학습된 가중벡터 W 는 중요도 순에 의해 내림차순으로 정렬되는 특성을 가지고 있다. GHA의 학습은 식(12)와 같다.

$$\Delta w_{ji}(n) = \eta [y_j(n) x_i(n) - y_j(n) v(n)] \quad (12)$$

$$y_j(n) = \sum_{i=0}^{l-1} w_{ji}(n) x_i(n)$$

$$v(n) = \sum_{k=0}^j w_{kj}(n) y_k(n)$$

학습된 가중벡터는 입력벡터 x 의 고유벡터 특성을 가지고 있으며 식(13)에 의해서 얻어진 계수 c 는 고유치 특성을 가진다. 본 논문에서는 고유치 계수 c 를 PCA를 이용한 특징 파라미터로 사용하였다.

$$c_j = \sum_{i=0}^{l-1} w_{ji}(n) x_i(n), \quad c = \{c_0, c_1, \dots, c_{l-1}\} \quad (13)$$

IV. Independent Component Analysis

독립성분분석은 관측신호에 선형적으로 혼합되어 있는 독립신호원들을 사전에 어떠한 정보도 없이 고차 통계특성(High Order Statistics), 정보이론(Information Theory), 추정이론(Estimation Theory)등을 이용하여 분리하는 BBS(Blind Source Separation)문제에 대한 하나의 기법이다.

식(14)에서 관측신호 $x(t)$ 를 몇 개의 확률적으로 독립인 신호 $s(t)$ 들에 가중벡터가 곱해진 다음 혼합되어 생성된 것이라 가정하고 독립신호원간의 통계적인 의존성을 정의한 다음 의존성이 최소가 되는 가중치 W 를 추정하여 식(15)를 통해 관측신호에 대한 독립신호를 구할 수 있다.

$$x(t) = A s(t) \tag{14}$$

$$s(t) = W x(t) , W = A^{-1} \tag{15}$$

독립신호 $s(t)$ 의 합으로 나타나는 관측신호 $x(t)$ 가 정규(gaussian)분포를 가진다면 $x(t)$ 를 구성하는 각각의 독립신호는 비정규(nongaussian)분포를 가지게 된다. 따라서, 독립성분 $s(t)$ 의 비정규성(nongaussianity)을 최대화하도록 가중벡터 W 를 학습함으로써 얻을 수 있다. 비정규성은 고차 통계이론에서의 첨도(kurtosis)와 정보이론의 negentropy에 의해 측정될 수 있다. $y = s(t)$ 라 할 때 첨도는 식(16)과 같다.

$$kurt(y) = E[y^4] - 3 \tag{16}$$

첨도는 데이터의 outlier에 민감한 단점을 가지고 있기 때문에 보통 비정규성 측정에서는 negentropy를 주로 사용한다.

정보이론에서 엔트로피는 식(17)과 같이 정의된다.

$$H(y) = - \int p_y(\eta) \log p_y(\eta) d\eta \tag{17}$$

엔트로피는 정규분포를 가지는 변수에 대해 큰 값을 가지며 여기서 $p_y(\eta)$ 는 y 의 확률밀도 함수이다.

negentropy는 엔트로피를 정규화한 것으로 식(18)과 같이 정의된다.

$$J(y) = H(y_{gauss}) - H(y) \tag{18}$$

여기서 y_{gauss} 는 상관과 공분산이 같은 정규 랜덤변수이다. negentropy는 계산의 복잡성 때문에 고차 통계특성과 nonquadratic 함수를 이용하여 식(19)과 같이 근사화할 수 있다.

$$J(y) \approx \{G(y) - G(\nu)\}^2 \tag{19}$$

여기서 ν 는 표준 정규분포를 따르는 랜덤변수이다. G 는

Hartley의 엔트로피 함수의 세 가지 조건(base, monotonicity, additivity)을 만족하며, 이런 함수를 대비함수(Contrast Function)라 하고 다음과 같은 함수들이 많이 사용된다.

$$G_1(y) = 1/a_1 \log \cosh a_1 y \tag{20}$$

$$G_2(y) = -\exp(-y^2/2), 1 \leq a_1 \leq 2 \tag{21}$$

학습은 Gradient 알고리즘을 사용하여 다음과 같이 수행된다.

$$\Delta w \propto Y z g(w^T z)$$

$$w \leftarrow w / \|w\|$$

$$Y = E[G(w^T z)] - E[G(\nu)] \tag{22}$$

여기서 $g = G'$ 이고 z 는 관측신호 x 를 zero-mean, unit-variance로 whiten한 신호이다.

V. 실험 결과

인식 실험은 SVM을 사용하여 그림 2와 같이 세 가지 특징 파라미터에 대한 인식성능을 비교하였다. 실험에서 사용된 음성데이터는 ETRI Samdori 데이터(20명의 화자가 4회 발성)를 사용하였다. 20명화자가 2회 발성한 숫자음 400개를 학습데이터로 사용하였고 인식데이터로는 학습데이터를 포함한 20명화자의 4회 발성분 800개의 숫자음을 사용하였다. 각 파라미터의 특징 추출에 대한 블록도는 그림 2와 같다.

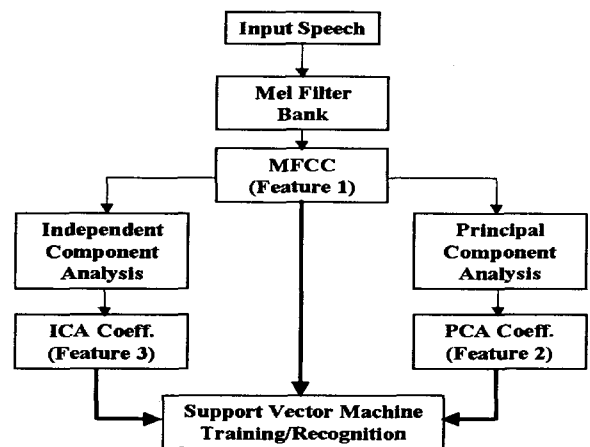


그림 4. Feature Extraction

그림 3에서 동일한 특징 차원에 대해 각 파라미터의 인식 성능은 ICA분석을 통한 파라미터의 성능이 가장 우수하였다. 그림 4, 5, 6은 특성이 유사한 숫자음 '삼'과 '사'에 대한 각각의 파라미터 1차, 2차 성분만을 취하여 분포도로 나타내었다. 그 결과 그림 6의 ICA에 의한 파라미터의 분포도가 선형 분별성이 가장 높다는 것을 알 수 있으며 이는 SVM에서 가장 좋은 인식률로 나타났다.

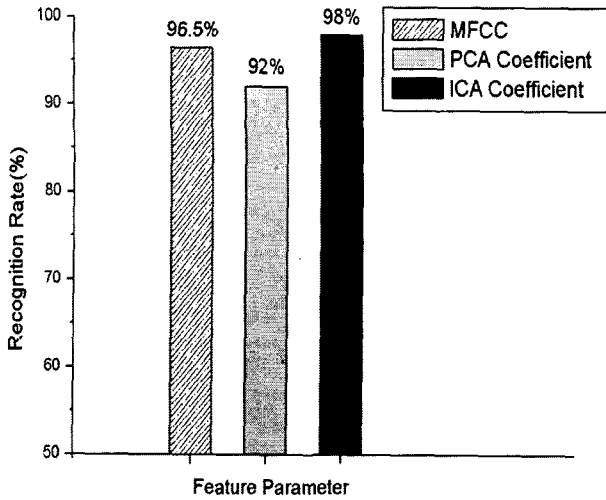


그림 5. Recognition Rate for Each Parameter

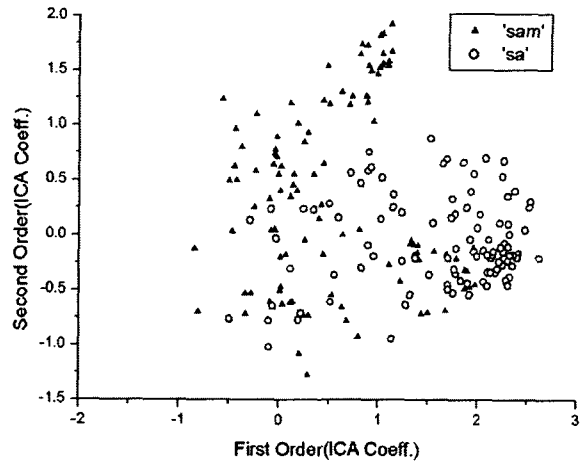


그림 8. Distribution of 1st & 2nd Coeff.(ICA)

VI. 결론

본 연구의 결과 독립성분분석에 의한 특징 파라미터가 SVM 음성인식기에서 가장 높은 인식 성능을 보여주었다. 이는 비선형 패턴 분류에 강인한 특성을 지닌 SVM을 인식기로 사용할 때 독립성분분석은 특성이 다른 음소로 이루어진 숫자음에 대해 기존의 다른 특징 파라미터보다 특징공간에서 좀더 분류하기 쉬운 형태로 존재하기 때문이다. 따라서 본 연구에서 제안하는 효과적인 음성인식 시스템은 ICA 특징 파라미터를 사용한 SVM 인식기로 이것은 많은 학습데이터로 인해 속도 저하가 문제가 되는 응용분야에서 적은 학습데이터로 높은 인식 성능 나타낼 수 있을 것이라 사료된다. 차후 더 정확한 효율성 검증을 위해 화자인식, 화자 적응화 등 다양한 음성 신호 처리 분야에 적용해보아야 할 것이다.

참고문헌

- [1] A.Hyvärinen, J.Karhunen, E.Oja, "Independent Component Analysis", John Wiley & Sons, 2001
- [2] S. Haykin, "Neural Networks", Prentice hall, 1999
- [3] Scholkopf, Smola "Learning with Kernels", The MIT Press, 2002
- [4] S.Choi, A.Cichocki, S.Amari, "Flexible independent component analysis", IEICE Trans. Fundamentals, vol.E83-A, no.12, pp.2715-2722, 2000
- [5] G.J.Jang, Y.H.Oh, "Text-independent speaker identification using support vector machines", Proceeding of 11'th KISS fall conference, pp.45-48, 1999
- [6] C.K.Kim, S.B.Kim, S.H.Kim, K.I.Hur, "Performance Improvement of Speech Recognition Based on Independent Component Analysis", ICSP2001, Vol.2 of 2, pp.663-666.

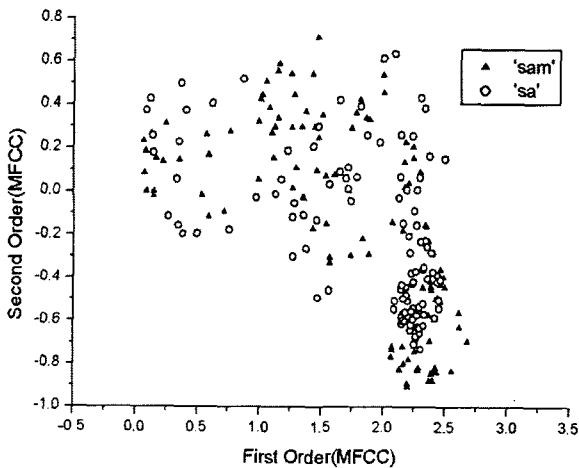


그림 6. Distribution of 1st & 2nd Coeff.(MFCC)

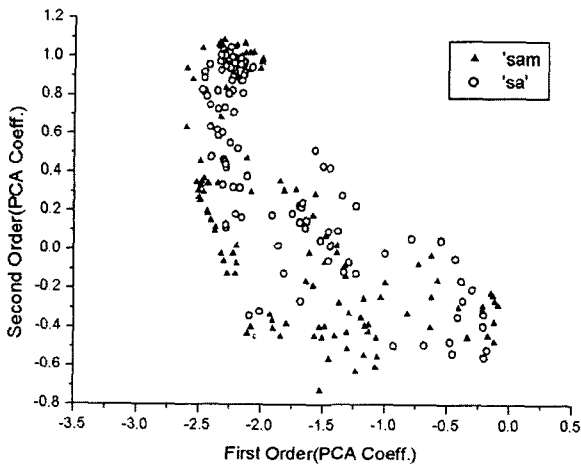


그림 7. Distribution of 1st & 2nd Coeff.(PCA)