

# 음성인식기 구현을 위한 SVM과 독립성분분석 기법의 적용

박정원, 김평환, 김창근, 허강인  
 동아대학교 전자공학과

전화 : 051-200-6961 / 핸드폰 : 016-597-1976

## Adoption of Support Vector Machine and Independent Component Analysis for Implementation of Speech Recognizer

Jeong-Won Park, Pyoung-Hwan Kim, Chang-Keun Kim, Kang-In Hur  
 Dept. of Electronic Engineering, Dong-A University  
 E-mail : jwpark@donga.ac.kr

### Abstract

In this paper we propose effective speech recognizer through recognition experiments for three feature parameters(PCA, ICA and MFCC) using SVM(Support Vector Machine) classifier. In general, SVM is classification method which classify two class set by finding voluntary nonlinear boundary in vector space and possesses high classification performance under few training data number. In this paper we compare recognition result for each feature parameter and propose ICA feature as the most effective parameter.

### I. 서론

음성인식성능에 있어서 고려되어야할 사항은 효과적인 인식기 및 특징 파라미터의 선택이다. 현재 대부분의 음성인식 시스템은 음성의 시간적 변화를 통계적으로 모델링 할 수 있는 HMM(hidden markov model)을 주로 사용하고 있으나 이는 불충분한 학습 데이터에 대해 인식 성능이 급격히 저하되는 단점이 있다. 최근 Machine Learning에 관한 연구가 활발히 진행되면서 SVM은 비선형 패턴 분류에 있어서 좋은 분류 성능을 나타내며 적은 학습 데이터에 대해서도 인식 성능이 뛰어나다고 알려져 있다. Machine Learning에 관련된 알고리즘들의 분류 성능은 특징 공간 내 입력 패턴들의 분포도가 얼마나 선형적 분리가 가능하냐에 달려있다.

본 논문에서는 학습 데이터량을 변화시키면서 HMM과 SVM의 인식 성능을 비교 분석하고, SVM 인식기에 대해 입력 특징 파라미터로 MFCC(mel frequency cepstrum coefficient)외에 독립성분분석(ICA)과 주성분해석(PCA)에 의한 MFCC의 특징공간을 변화시킨 파라미터를 사용하여 인식 실험을 수행하고 각각의 특징에 대한 인식 성능과 특성을 비교 분석 하였다.

### II. 본론

#### 2.1 Principal Component Analysis

주성분 해석기법(PCA)은 입력의 선형성과 특성 식별을 이용하여 입력 데이터에서 변동량이 큰 주축(Principal Axis)을 순서대로 찾을 수 있기 때문에 데이터의 차원을 효과적으로 축소할 수 있다. zero-mean 특성이 있는  $n$ 차원 신호  $x$ 에 대해 차원 축소는 식(1)과 같으며 원신호로의 복원은 식(2)와 같다.

$$a = [x^T q_1, x^T q_2, \dots, x^T q_l] = Q^T x \quad (1)$$

$$\hat{x} = Qa = \sum_{j=1}^l a_j q_j \quad (2)$$

여기서  $n > l$  이며  $a$  는  $l$ 차원으로 축소된 신호이다.

식(1)(2)에서  $a$  와  $Q$  는  $x$ 의 공분산 행렬(covariance matrix)  $R$ 의 고유치와 고유벡터의 일부로,  $R$ 를 구하는 방법은 식(3)에, 고유벡터와 고유치를 구하는 방법은 식(4)에 나타내었다.

$$R = \frac{1}{n} \sum_{i=1}^n (x x^T) \quad (3)$$

$$R \mathbf{q}_i = \mathbf{a}_i \mathbf{q}_i, \quad i = 1, 2, \dots, n \quad (4)$$

차원 축소된 신호  $\mathbf{a}$ 는 중요도 순으로 나열된  $n$ 개의 고유벡터  $\mathbf{Q}$ 중에  $l$ 개만 사용하여 얻을 수 있다

본 논문에서는 GHA (Generalized Hebbian Learning)을 이용하여 구현하였다. GHA는 자율학습특성이 있으며 학습된 가중벡터  $\mathbf{W}$ 는 중요도 순에 의해 내림차순으로 정렬되는 특성을 가지고 있다. GHA의 학습은 식(5)와 같다.

$$\Delta w_{ji}(n) = \eta [y_j(n) x_i(n) - y_j(n) v(n)] \quad (5)$$

$$y_j(n) = \sum_{i=0}^{l-1} w_{ji}(n) x_i(n), \quad v(n) = \sum_{k=0}^l w_{kj}(n) y_k(n)$$

학습된 가중벡터는 입력벡터  $\mathbf{x}$ 의 고유벡터 특성을 가지고 있으며 식(6)에 의해서 얻어진 계수  $\mathbf{c}$ 는 고유치 특성을 가진다. 본 논문에서는 고유치 계수  $\mathbf{c}$ 를 PCA를 이용한 특징 파라미터로 사용하였다.

$$\mathbf{c}_j = \sum_{i=0}^{l-1} w_{ji}(n) x_i(n), \quad \mathbf{c} = \{c_0, c_1, \dots, c_{l-1}\} \quad (6)$$

## 2.2 Independent Component Analysis

독립성분분석(ICA)은 특정신호의 생성단계에서 선형적으로 혼합되어 있는 독립신호원들을 관측데이터에서 분리해내는 방법이다. 임의의 신호는 몇 개의 확률적으로 독립인 신호들에 가중치가 곱해진 다음 혼합되어 생성된 것이라 가정하고 정보이론(Information Theory), 고차통계특성(High Order Statistics), 추정이론(Estimation Theory) 등을 이용하여 독립신호원간의 통계적인 의존성 최소가 되는 가중벡터를 추정하여 임의의 신호가 생성되게 하는 독립신호원을 구할 수 있다.

음성 신호는 통계적으로 독립인 고차 신호 특성들로 구성되어 있으며 고차 신호 특성들은 음성 신호의 주파수와 위상 스펙트럼을 나타내는 가중벡터를 통해서 추출될 수 있다. 음성 신호로부터 독립 고차 신호 특성들을 분리해내기 위한 가중벡터는 상호정보량(mutual information)을 최소화하거나 또는 비정규성(nongaussianity)을 최대화 하는 방법으로 학습시킬 수 있다. 이렇게 학습된 가중벡터는 음성 신호의 특징 주파수 대역에 민감한 특성을 보인다.

관측벡터  $\mathbf{x}$ 는 식(7)과 같이 서로 확률적으로 독립인 벡터  $\mathbf{y}$ 의 선형결합형태로 표현할 수 있다.

$$\mathbf{x} = \mathbf{A} \mathbf{y} \quad (7)$$

$$\mathbf{y} = \mathbf{W} \mathbf{x} \quad \mathbf{W} = \mathbf{A}^{-1} \quad (8)$$

여기서,  $\mathbf{A}$ 는 선형혼합행렬이라 하고 독립벡터  $\mathbf{y}$ 는  $\mathbf{A}$ 의 역행렬인 식(8)의 가중벡터  $\mathbf{W}$ 를 구하여 얻을 수 있다. 추정된 독립성분들의 통계적인 독립성은 상호정보량(mutual information)으로 정의한다. 관측벡터의 상호정보량은 추정된 독립성분들의 결합엔트로피와 각각의 엔트로피의 차로 계산을 하거나 Kullback-Leibler 발산정리에 의해 다음과 같이 정의한다.

$$J(\mathbf{W}) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{i=1}^N p_i(y_i)} d\mathbf{y} \\ = -H(\mathbf{y}) + \sum_{i=1}^N H(y_i) \quad (9)$$

상호정보량  $J(\mathbf{W})$ 은 항상 양수이고, 또한 각 성분들이 독립적으로 분리되었을 때는 0 이 된다.

독립성분분석의 목적은 상호정보량  $J(\mathbf{W})$ 을 최소화 하는  $\mathbf{W}$ 를 추정하여 식(8)에 의해 독립벡터  $\mathbf{y}$ 를 구하는 것이다. 학습방법은 식(10)의 자연 감소법(natural gradient)을 사용하여 가중벡터  $\mathbf{W}$ 를 추정할 수 있다.

$$\Delta \mathbf{W} = -\eta \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} [\mathbf{W}^T \mathbf{W}] \\ = \eta [ \mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}) \mathbf{y}^T ] \mathbf{W} \quad (10)$$

여기서  $\mathbf{I}$ 는 단위행렬,  $\boldsymbol{\varphi}(\mathbf{y}) = [\varphi_1(y_1) \dots \varphi_N(y_N)]^T$ , 그리고  $\varphi_i(y_i) = -\frac{\partial \log p(y_i)}{\partial y_i}$  이다.

본 실험에서는 추정된  $\mathbf{W}$ 에 의해 얻어진 독립벡터  $\mathbf{y}$ 를 특징 파라미터로 사용하였다.

## 2.3 Support Vector Machine

SVM은 SRM(Structural Risk Minimization)이론으로부터 발전한 비선형 이진 패턴 분류기이다.

선형 SVM에 있어서 두 개의 클래스를 구분할 수 있는 초평면(Hyperplane)은 무수히 많으나 그림 1에서 처럼 두 클래스 간 점들의 거리  $\rho$ 를 최대화하도록 하면 OSH(Optimal Separating Hyperplane)는 유일한 해로 존재한다.

다시 말해, 식(11)의 OSH는 거리  $\rho$ (식(12))를 최대화 하는 최적의 가중벡터  $\mathbf{w}_*$ 와 바이아스  $b_*$ 로서 얻어질 수 있다.

$$\mathbf{w}_*^T \cdot \mathbf{x}_i + b_* = 0 \quad (11)$$

$$\rho = 2 / \|\mathbf{w}_*\| \quad (12)$$

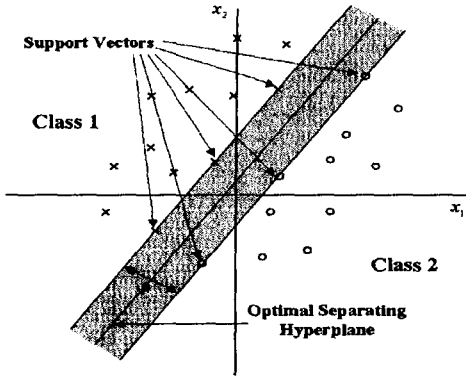


그림 1 OSH for Linear Separable Patterns.

$w_*$ 와  $b_*$ 는 다음의 목적식(식(13))과 제약식(식(14))을 갖는 QP(Quadratic Programming)에 의해 목적식을 최소화하는 방법으로 구한다.

$$\Phi(w) = 1/2 w^T w \quad (13)$$

$$d_i(w^T x_i + b) \geq 1 \text{ for } i=1, 2, \dots, N \quad (14)$$

QP은 식(15)의 Lagrangian 함수의 안장점(saddle point)을 찾는 문제로써, 편미분과 Kuhn-Tucker 조건으로서 식(16)을 최대화 하는  $\alpha$ (Lagrange Multiplier)를 찾는 문제로 바꿀 수 있다.

$$J(w, b, \alpha) = 1/2 w^T w - \sum_{i=1}^N \alpha_i [d_i(w^T x_i + b) - 1] \quad (15)$$

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - 1/2 \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j \quad (16)$$

식(16)을 최대화 하는  $\alpha_*$ 를 구한다면 식(17)(18)에 의해  $w_*$ ,  $b_*$ 를 구한다.

$$w_* = \sum_{i=1}^N \alpha_* d_i x_i \quad (17)$$

$$b_* = 1 - w_*^T x^{(s)} \text{ for } d^{(s)} = 1 \quad (18)$$

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

결론적으로 임의의 입력 패턴  $x$ 는 식(19)에 의해 분류된다.

$$f(x) = \text{sign}(w_*^T x + b_*) \quad (19)$$

입력 패턴이 선형 분리가 불가능한 경우는 비선형 SVM을 사용하여 Soft Margin  $\xi_i$ 과 커널(kernel) 함수  $K(x, x_i)$ 에 의해 선형분류가 가능한 고차원 공간으로 확장하여 OSH를 찾는다.

$$d_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \text{ for } i=1, 2, \dots, N \quad (20)$$

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - 1/2 \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (21)$$

커널(Kernel) 함수로는 보통 다항식, RBFN(radial-basis function networks) 커널이 주로 사용되며 본 실험에서는 RBFN 커널을 사용하였고 OPC(One-Per-Class)를 사용하여 멀티 클래스로 확장하였다.

### III. 실험 및 비교 분석

#### 3.1 인식 실험 1(HMM, SVM)

인식 실험에 사용된 음성데이터는 ETRI Samdori 데이터(20명의 화자가 각 숫자음 4회 발성)로 분석조건은 표 1과 같다.

표 1. 분석 조건

A/D convert	16kHz, 16bit, pre-emphasis
window	hamming(16ms, 256samples)
shifting period	3.75ms(60samples)
feature parameter	10th MFCC

인식실험은 학습 데이터량에 따른 각각의 인식 알고리즘에 대한 인식을 변화를 관찰하였다. 결과는 그림2와 같다. 그림 2의 결과에서 학습데이터수가 높을 때는 HMM이 SVM보다 인식성능에서 약간의 우수함이 나타나지만 학습데이터가 적을수록 HMM에서는 인식을 저하가 현저히 나타나지만 SVM에서는 인식률의 변화가 거의 없었다. 위 실험의 결과로서 SVM이 적은 학습데이터에 대해서도 HMM에 비해 신용 있는 인식률을 나타냄을 알 수 있었다.

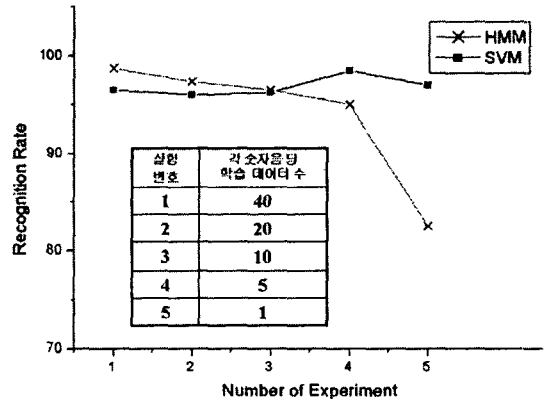


그림 2 학습 데이터량에 따른 인식을 비교

#### 3.2 인식 실험 2(PCA, ICA, MFCC)

사용한 음성데이터는 인식실험 1 과 동일한 ETRI Samdori 데이터를 사용하였다. 20명 화자가 2회 발성한 숫자음 400개를 학습데이터로 사용하였고 인식데이터로는 학습데이터를 포함한 800개의 숫자음을 사용하였다. 각 파라미터의 특징 추출에 대한 블록도는 그림 3과 같다.

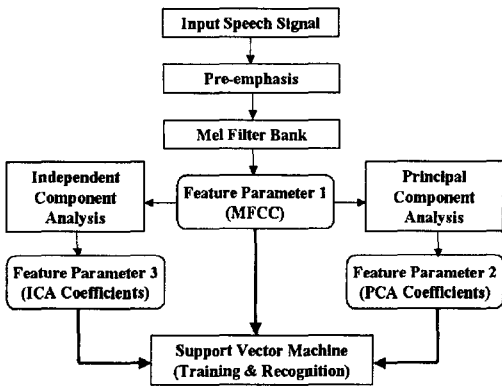


그림 3. Block diagram of Feature Extraction

표 2. 각 파라메터에 대한 인식성능

Feature	차수	인식률
MFCC	10	96.5%
PCA Coefficient	10	92%
ICA Coefficient	10	98%

동일한 특징 차원에 대해 각 파라메터의 인식 성능은 ICA를 통한 파라메터의 성능이 가장 우수하였다.

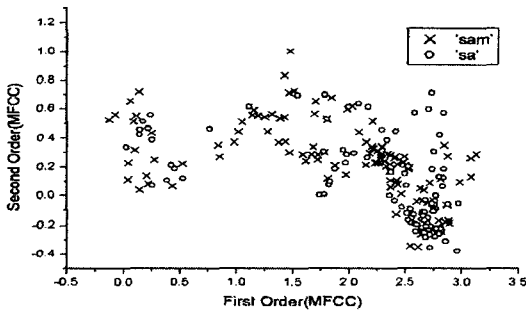


그림 4. Distribution of 1st & 2nd Coeff.(MFCC)

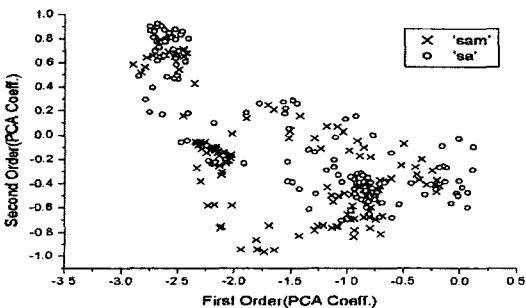


그림 5. Distribution of 1st & 2nd Coeff.(PCA)

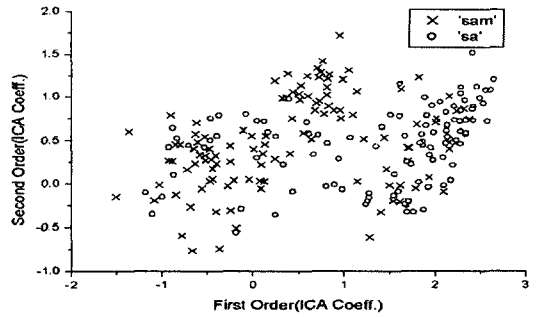


그림 6. Distribution of 1st & 2nd Coeff.(ICA)

그림 4, 5, 6은 특성이 유사한 숫자음 '삼' 과 '사'에 대한 각각의 파라메터 1차, 2차 성분만을 취하여 분포도로 나타내었다. 그림 6의 ICA에 의한 파라메터가 특징 공간에서 선형 분별성이 가장 높다는 것을 알 수 있었고 이 결과가 SVM인식기에서 높은 성능으로 나타났다.

#### IV. 결론

본 연구의 결과 제안된 음성인식 시스템은 ICA와 SVM을 사용한 시스템으로써 SVM은 학습 데이터량이 적은 상황에서도 좋은 인식성능을 보여주었고 이는 많은 학습데이터로 인한 속도 저하가 문제가 되는 응용분야에서 좋은 성능을 발휘 할 수 있을 것으로 사료된다. 그리고 독립성분분석에 의한 특징 파라메터는 MFCC나 주성분해석에 의한 특징 파라메터 보다 특징 공간상에서 패턴의 분류능력이 우수함을 확인할 수 있었다. 차후 더 정확한 효율성 검증을 위해 화자인식, 화자 적응화 등 다양한 음성 신호 처리 분야에 적용해 보아야 할 것이다.

#### 참고 문헌

- [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.