

SVM을 이용한 자동 음소분할에 관한 연구

권 호 민, 한 학 용, 김 창 근, 허 강 인

동아대학교 전자공학과

전화 : 051-200-6961 / 핸드폰 : 018-561-5621

Research about auto-segmentation via SVM

Ho-Min Kwon, Hag-Yong Han, Chang-Keun Kim, Kang-In Hur

Dept. of Electronics, Dong-A University

E-mail : hmkwon@donga.ac.kr

Abstract

In this paper we used Support Vector Machines (SVMs) recently proposed as the learning method, one of Artificial Neural Network, to divide continuous speech into phonemes, an initial, medial, and final sound, and then, performed continuous speech recognition from it. Decision boundary of phoneme is determined by algorithm with maximum frequency in a short interval. Recognition process is performed by Continuous Hidden Markov Model(CHMM), and we compared it with another phoneme divided by eye-measurement. From experiment we confirmed that the method, SVMs, we proposed is more effective in an initial sound than Gaussian Mixture Models(GMMs).

I. 서론

음성에 대한 연구가 발전함에 따라 분할과 라벨링된 음성 DB에 대한 필요가 증가되고 있다. HTK와 같은 음성인식도구에서는 발음사전을 바탕으로 자동으로 음성인식의 단위를 정렬하여 초기 음소 모델을 구성하고 다이폰이나 트라이폰 모델 등으로 확장하여 사용하기도 하지만 안정된 발음사전의 구성과 데이터의

확보에 어려움이 있다. 반면에 음성을 일정 세그먼트의 연결로 가정하고 인식단위로의 자동 분할을 통한 음성인식 시스템의 경우, 최소 인식단위인 음소 단위로 정확하게 분할되고 라벨링된 음성 DB는 인식기의 성능에 결정적인 영향을 미치게 된다. 현재, 음소 단위의 음성인식이 음성의 모델 수가 적어 많은 이점이 있는데도 불구하고 이러한 음성 DB를 만들기 위해선 많은 시간과 노력을 필요로 하며 이러한 분할구간의 자동결정은 한국어의 조음결합 현상 등을 고려할 경우 어려운 과제로 남아있다. 또한 자음의 인식을 저하도 실제 적용상의 문제점으로 지적된다.

음성정보의 의미 단위로 자동으로 분할하기 위한 방법은 다양한 방법이 제안되어져 있다. 본 논문에서는 최근 많은 연구가 활발히 이루어지고 있는 SVM 분류기와 전통적인 패턴인식 방법인 GMM을 이용하여 자동으로 음성을 분할을 행하고 이를 통하여 인식을 행하였다. 실험결과 초성의 경우 낮은 인식률을 보였지만 중성과 종성의 경우, 목측으로 분할한 데이터와 비교한 결과 상당한 성능을 확인할 수 있었다.

본 논문의 구성은 2장에서 선형공간과 비선형 공간에서의 SVMs 분류기에 대해서 간단하게 설명하였고, 3장에서는 음성의 한 프레임마다 SVMs로 분류된 음성을 자동음절분할 학기 위한 후처리 과정을 설명하였다. 실험에 사용된 DB 및 실험 결과를 4장에서 기술하였으며, 마지막으로 5장에서 결론 및 향후 과제를 제시하였다.

II. Support Vector Machines(SVMs)

2.1 Linear SVMs

$y_i = \{-1, +1\}$ 인 두 개의 클래스에 속하도록 $(x_1, y_1), \dots, (x_l, y_l) \in R^N$ 학습 벡터를 분류한다고 고려하자. 여기서 우리는 초평면을 이용해서 두 개의 클래스로 분류할 것이다.

$$(\mathbf{w} \cdot \mathbf{x}) + b = 0, \quad \mathbf{w} \in R^N \text{ and } b \in R \quad (1)$$

여기서 \mathbf{w} 와 b 는 $f(x) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b)$ 인 판별 함수에서 적용될 파라미터들이다.

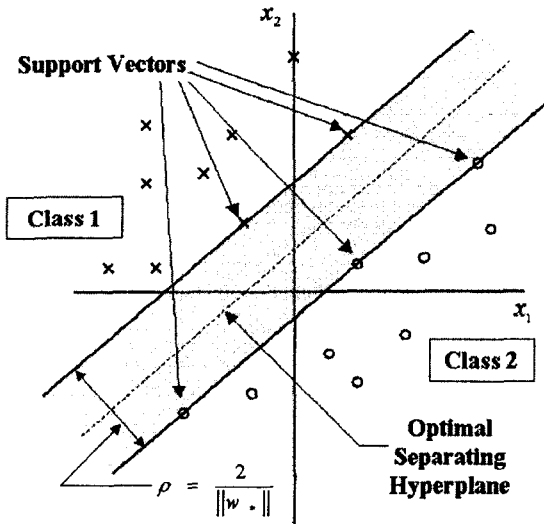


그림 1 선형 공간에서의 분류

그림 1에서와 같이 2차원 공간의 경우 입력 데이터들을 분류할 수 있는 선형 분류기가 많이 존재할 수 있다. 그러나 초평면과 여기에 가장 가까운 데이터들의 거리를 최소화하는 초평면은 단 하나만 존재한다. 이런 선형 분류기를 Optimal Separating Hyperplane(OSH)라 부른다. 초평면은 $(\mathbf{w} \cdot \mathbf{x}) + b = 0$ 은 다음 조건을 만족한다.

$$\begin{aligned} (\mathbf{w} \cdot \mathbf{x}) + b > 0 & \text{ if } y_i = 1 \\ (\mathbf{w} \cdot \mathbf{x}) + b < 0 & \text{ if } y_i = -1 \end{aligned} \quad (2)$$

식 (2)의 \mathbf{w} 와 b 를 적절하게 선택함으로써 다음의

제약식을 만족하는 하나의 분류평면으로 수식화 할 수 있다.

$$y_i [(\mathbf{w} \cdot \mathbf{x}) + b] \geq 1, \quad i = 1, \dots, l \quad (3)$$

그리고 초평면을 최적으로 만드는 cost function은 다음과 같다.

$$\Phi(\mathbf{w}) = \frac{\|\mathbf{w}\|^2}{2} \quad (4)$$

따라서 최적화문제는 라그랑지안 승수(Lagrangian Multiplier)를 이용하여 등가 비제약 최적화문제로 다시 정의되고 라그랑제 함수(Lagrange function)의 안장점을 구함으로써 해결된다.

$$L(\mathbf{w}, b, a) = \frac{\|\mathbf{w}\|^2}{2} - \sum_{i=1}^l a_i \{[(\mathbf{w} \cdot \mathbf{x}) + b]y_i - 1\} \quad (5)$$

식 (5)를 KKT조건에 의해 정리하면 다음과 같다.

$$W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (6)$$

식 (6)의 해답을 찾기 위해선 2차 계획법(Quadratic Programming)과 대수적 방법을 필요로 하며 OSH는 다음에 의해 구해진다.

$$\mathbf{W}_0 = \sum_{i=1}^l a_{0,i} d_i \mathbf{x}_i, \quad b_0 = -\frac{1}{2} \mathbf{w}_0 \cdot [\mathbf{x}_r + \mathbf{x}_s] \quad (7)$$

여기서 \mathbf{x}_r 과 \mathbf{x}_s 는 각 클래스의 support vector들이고 그림 1에서 초평면에 각각 가장 가까운 점들이다.

2.2 Nonlinear SVMs

비선형 데이터 공간에서는 제약식을 위반하는 양을 평가하는 새로운 변수 ξ_i 를 이용함으로써 마진을 최대화한다.

$$\begin{aligned} \text{Minimise } \Phi(\mathbf{W}) &= \frac{\|\mathbf{w}\|^2}{2} + C(\sum \xi_i) \\ y_i [(\mathbf{w} \cdot \mathbf{x} + b)] &\geq 1 - \xi_i \\ \text{Subject to } \text{and } \xi_i &\geq 0 \quad i = 1, \dots, l \end{aligned} \quad (8)$$

재 정의된 라그랑제 함수는 다음과 같다.

$$L(\mathbf{w}, b, a) = \frac{\|\mathbf{w}\|^2}{2} + C \left[\sum_{i=1}^n \alpha_i \right] - \sum_{i=1}^n r_i \xi_i - \sum_{i=1}^n \alpha_i \{[(\mathbf{w} \cdot \mathbf{x}) + b]y_i - 1 + \xi_i\} \quad (9)$$

여기서 r_i 과 ξ_i 는 식 (8)의 제약식과 관련되며 α_i 의 값은 $0 \leq \alpha_i \leq C$ 를 만족해야 한다. 만약 선형경계가 부적절하고 비선형 분류평면일 경우 입력 벡터 \mathbf{x} 를 고차원의 특징 공간으로 맵핑(mapping)할 수 있고 그 역할은 $K(\mathbf{x}, \mathbf{y})$ 가 담당한다.

$$\mathbf{w}(a) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \quad (10)$$

이 때, 커널 K 는 적당하게 선택될 수 있으며, 본 논문에서는 비교적 우수하다고 알려진 RBFN 커널을 사용하였다. 따라서, 비선형 데이터 공간에서 OSH를 결정짓는 분류기는 다음과 같다.

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^n y_i \alpha_i K(\mathbf{x} \cdot \mathbf{x}_i) + b_o \right) \quad (11)$$

2.2 사용된 다중 클래스 분류 기법

SVMs의 이진 분류능력을 다중으로 확장하기 위해서 OPC(One-Per-Class), PWC(PaireWise Coupling)이 일반적으로 알려져 있으며 non-sense output 때문에 이를 보완하기 위한 다양한 방법들이 모색 중이며 본 논문에서는 OPC를 적용하였다.

III. 후처리 과정

SVM을 이용한 멀티 클래스 분류를 통하여 초·중·중성에 대한 프레임 단위의 1차 분류를 기반으로 하여 2차 음소경계를 결정하기 위해서는 1차 분류 결과의 오인식에 대한 보상처리가 필요하다. 이를 위해서 본 논문에서는 1차 인식결과에 단구간의 최빈값을 이용하는 후처리 알고리즘을 제안하고 적용하였다.

step1 멀티 클래스 분류 결과를 초·중·중성으로 맵핑

step2 시작 프레임과 종료 프레임의 전, 후에 의사 보상 값을 패딩, 패딩 프레임의 수는 짝수

step3 단구간 프레임 크기를 한 프레임씩 슬라이딩하면서 최빈 인덱스를 보정한 후 인덱스로 설정

step4 (패딩 프레임 수/2)의 프레임을 전후 제거

step5 step4의 결과를 바탕으로 최종 음소 경계 결정

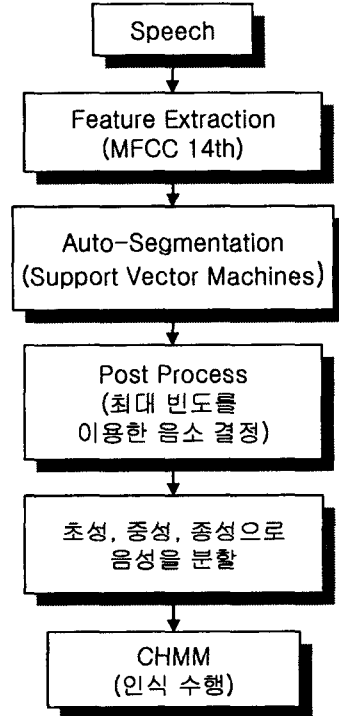


그림 2 전체적 구현 블록도

IV. 실험결과 및 고찰

4.1 실험조건

음소분할 실험을 위하여 CVC형 108음절 DB를 구성하였다. 본 DB는 우리말 음성의 CVC형 음절로 초성(비 디 기 교 트 쿨), 중성(나 키 고 게 기 |), 종성(니 리 리)으로 이루어진 108개의 유사 음절로 구성되어 있으며 5명의 화자가 5회 발성하여 3회분은 학습용으로 나머지 2회분은 평가용으로 이용하였다. 목적에 의하여 초성, 중성, 종성을 분리한 음소 DB를 별도로 구성하여 학습을 행하고 분할과 인식실험을 행하였다. 음성의 분석조건은 표 1과 같다.

표 1. 음성 데이터의 분석조건

A/D	16kHz, 16bit
Filtering	LPF , 7 kHz
Step Size	60 point
Window Length	256 point
Feature Parameter	MFCC 14th

교적 우수한 분할 성능을 확인했고 CHMM을 통한 분할된 음소 인식에서 이 결과를 확인했다. 현재 SVM은 많은 연산량과 학습시간을 필요로 하지만, 무성음, 파열음, 마찰음 등의 비선형성을 가지는 음소 모델에선 확률 및 통계보다 나은 접근 방법으로 사료된다. 따라서 SVM을 최적화시키기 위한 많은 방법들이 연구되어야 할 것이며, 향후 SVM을 통한 최적의 음소 인식기를 구현할 계획이다.

4.2 실험방법 및 결과

분할 방법은 SVM과 GMM으로 각각 행하고 그 결과를 서로 비교하였다. SVM과 GMM 모두 다중 분류를 행하고 이를 초·중·종성으로 맵핑하여 제안하는 후처리 알고리즘으로 그 결과를 보상하는 방법으로 음소 경계를 결정하였다.

표2 자동분할의 성능(목측분할과의 편차)

Frame		초성	중성	종성
GMM	평균	6.33	4.97	4.69
	표준 편차	6.88	6.29	6.36
SVM	평균	3.45	7.19	6.59
	표준 편차	3.53	8.32	8.90

표3 자동음소분할을 통한 음소단위 인식 결과

Error(%)	GMM분할	SVM분할	목측분할
초성	23.80	20.93	13.89
중성	5.46	7.79	3.98
종성	8.15	9.07	5.37

참고문헌

- [1] Bernhard Scholkopf, aAlexander J. Smola, "Learning with Kernels", The MIT Press, 2002
- [2] Christopher J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Bell Laboratories, Lucent Technologies.
- [3] J. Weston, C. Watkins, "Multi-class Support Vector Machines, Technical Report, Royal Holloway, University of London.
- [4] Yonas B. Dibike, Slavco Velickov, Dimitri Solomatine, Michael B. Abbott, "Model Induction with Support Vector Machines: Introduction and Applications", ASCE Journal of Computing in Civil Engineering, 2001, vol. 15, No. 3, pp. 208-216
- [5] Edgar E. Osuna, Robert Freund and Federico Girosi, "Support Vector Machines: Training and Applications", C.B.C.L Paper No. 144, 1997.
- [6] Philip Clarkson and Pedro J. Moreno, "On the use of Support Vector Machines for Phonetic Classification", ICASSP99
- [7] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and selected Applications in Speech Recognition", 1989 IEEE
- [8] ftp://ftp.idiap.ch/pub/learning/SVMTorch.tgz

V. 결론 및 향후과제

우리말은 외국어와 달리 초성, 중성, 종성이 합해져서 음절을 이루고 이 음절이 단어와 문장을 이루기 때문에 인식단위, 특히 음소와 같은 최소 단위로의 안정된 분할은 연속음성인식을 위한 주목할만한 연구과제이다. 본 논문에서는 SVM을 이용해서 자동 음소분할을 시도하고 이를 통하여 음성인식을 행하고 GMM을 이용한 것과 그 성능을 서로 비교하였다. 실험결과 음성의 SVM은 초성에서, GMM은 중성, 종성에서 각각 비