Journal of The Korea Knowledge Information Technology Society

ISSN 1975-7700

http://www.kkits.or.kr

Efficient Segmentation by Phoneme Unit using SVMs

Gwang-Seok Lee*

Department of Electronic Engineering, Gyeongnam National University

ABSTRACT

In this research, we used Support Vector Machines(SVMs) as the learning and recognition unit of the speech, one of artificial neural network, to segmented from the continuous speech into phonemes, an initial, medial, and final sound, and then, performed continuous speech recognition from it, A decision boundary of phoneme is determined by algorithm with maximum frequency in a short interval. speech recognition process is performed by Continuous Hidden Markov Model(CHMM), and we compared it with another phoneme segregated from the eye-measurement. From the simulation results, we confirmed that the method, SVMs, we proposed is more effective in an initial sound than Gaussian Mixture Models(GMMs). We plan to construct a optimum hybrid classifier of SVMs and GMMs, and apply to continuous speech recognition.

KEYWORDS: Phoneme segmentation, Pattern secognition, SVMs, CHMM, GMMs

ARTICLE INFO: Received 15 January 2014, Revised 5 February 2014, Accepted 14 February 2014.

1. 서 론

최근에 커널머신을 이용한 대표적 학습기로 SVMs이 주목을 받고 있으며 이 SVMs의 기계학습 능력을 음성분야에도 응용하려는 연구가 시도되고 있다. 한편 음성에 대한 연구가 발전함에 따라 음성신호의 여러 단위로의 분할과 라벨링된 음성 DB에 대한 필요가 증가되고 있으며 HTK와 같은 음성인식 도구에서는 발음사전을 바탕으로 자동으로 음성인식의 단위를 정렬하여 초기 음소모델을 구성하고 다이폰이나 트라이폰 모델 등으로 확장하여 사용하기도 하지만 여전히 안정된 발음

*Corresponding author is with the Department of Electronic Engineering, Gyeongnam National University, 33 Dongjin-ro Jinju Gyeongnam, 660-758, KOREA.
E-mail addresses: kslee@gntech.ac.kr

사전의 구성과 데이터의 확보에 어려움이 있다. 반면에 음성을 일정 세그먼트의 연결로 가정하고 인식단위로의 자동 분할을 통한 음성 인식시스템의 경우에는 최소 인식단위인 음소 단위로 정확하게 분할되고 라벨링된 음성 DB는 인식기의 성능에 결정적인 영향을 미치게 된다.

현재로서는 음소 단위의 음성인식이 음성의 모델 수가 가장 적음으로서 가장 많은 이점이 있는데도 불구하고 이러한 음성 DB를 만들기 위해선 많은 시간과 노력을 필요로 하며 이러한 분할구간의 자동결정은 한국어의 조음결합 현상 등을 고려할 경우 여전히 어려운 과제로 남아 있다. 음성정보의 의미 단위로 자동으로 분할하기 위한 방법[1]은 다양한 시도가 제안되어 있으나 본 연구에서는 최근 많은 연구가 활발히 이루어지고 있고 구조적 위험 최소화로 설명되는 통계적 학습이론에 기반한 일반화 성능을 보여주는 복잡하고 정교한 분류기로 알려져 있으며 높은 일반화 능력으로 다양한 패턴인식 문제에서 좋은 결과를 보여주고 있는 SVMs분류기와 전통적인 패턴인식 방법인 GMM을 이용하여 자동으로 음성을 음소단위로 분할하였다.

실험 결과, 초성의 경우 GMM보다 SVMs이 높은 인식률을 보였고, 중성, 종성의 경우는 GMM이 조금 더 높은 인식률을 보임을 알 수 있었다.

본 연구의 구성은 2장에서 선형공간과 비선형 공간에서의 SVMs 분류기[2-8]에 대하여 간략하게 설명하였으며 3장에서는 음성의 한 프레임마다 SVMs로 분류된 음성을 자동 음절로 분할하기 위한 후처리 과정을 설명하였다[9]. 실험에 사용된 DB 및 실험 결과를 4장에서 기술하였으며, 마지막으로 5장에서 결론 및 향후 과제를 제시하였다.

2. Support Vector Machine (SVMs)

2.1 선형 SVMs

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

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

여기서, ω와 b는 $f(x)=sign(w \cdot x+b)$인 판별 함수에서 적용될 파라미터들이다. $x\in R^N$ 은 입력벡터, ω는 가중치 벡터 및 b는 바이어스를 각각 나타낸다.

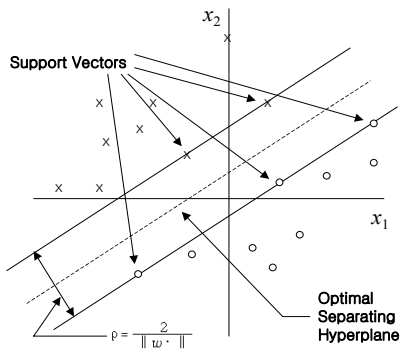

그림 1. 선형 공간에서의 분류
Figure 1. Separation in Linear Space

<그림 1>에서와 같이 2차원 공간의 경우 입력 데이터들을 분류할 수 있는 선형 분류기가 다수 존재할 수 있다. 그러나 초평면과 여기에 가장 가까운 데이터들의 거리를 최대화하는 초평면은 단 하나만 존재한다. 이러한 선형 분류기를 Optimal Separating Hyper plane(OSH)라 부르며 초평면은 $(w \cdot x)+b=0$은 다음의 조건을 만족한다.

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

식(2)의 w와 b를 적절하게 선택함으로써 다음의 제약식(3)을 만족하는 하나의 분류평면으로 수식화 할 수 있다.

$$y_i[(w \cdot x)+b] \geqq 1,\ i=1,\ \cdots,\ l \tag{3}$$

그리고 초평면을 최적으로 만드는 Cost Function은 다음의 (4)식과 같다.

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

따라서 최적화문제는 Lagrangian 곱셈기를 이용하여 등가 비 제약 최적화문제로 다시 정의되고 Lagrange함수에 의해 해결된다.

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

식 (5)를 KKT조건에 의해 정리하면 다음과 같다.

$$W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j (x_i \cdot x_j) \tag{6}$$

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

$$W_0 = \sum_{i=1}^{l} a_{oi} y_i x_i,\ b_o = -\frac{1}{2} W_o[x_r + x_s] \tag{7}$$

여기서, α는 Lagrange계수이고 x_r과 x_s는 각 클래스의 Support Vector들이며 <그림 1>에서 초평면에 각각 가장 가까운 점들이다.

2.2 비선형 SVMs

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

$$\begin{aligned} &\min \Phi(W) = \frac{\| w \|^2}{2} + C(\sum \xi_i) \\ &y_i[(w \cdot x + b)] \geqq 1 - \xi_i \\ &\textit{subject to} \text{ and } \xi_i \geqq 0,\ i=1,...,l \end{aligned} \tag{8}$$

여기서 다시 재 정의된 Lagrange함수는 다음과 같다.

$$\begin{aligned} L(w,b,a) = &\frac{\| w \|^2}{2} + C\left[\sum_{i=1}^{l}\right] - \sum_{i=1}^{l} r_i \xi_i \\ &- \sum_{i=1}^{l} a_i\{[(w \cdot x)+b]y_i - 1 + \xi_i\} \end{aligned} \tag{9}$$

여기서, r_i과 ξ_i는 식 (8)의 제약식과 관련되며 a_i의 값은 $0 \leqq a_i \leqq C$ 를 만족해야 한다. 만약 선형경계가 부적절하고 비선형 분류평면일 경우 입력 벡터 x를 고차원의 특징 공간으로 mapping할 수 있으며 그 역할은 $K(x,y)$가 담당한다.

$$W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j K(x_i \cdot x_i) \tag{10}$$

2.3 다중 클래스 분류법

SVMs의 이전 분류능력을 다중으로 확장하기 위해서는 문제에 적용하기 위한 일반적 방법인 출력 코딩 방법은 다중 클래스 문제를 풀기 위한 범용 형태로 다음과 같이 설명될 수 있다. 복잡한 다중 클래스 문제는 보다 쉬운 이진 문제들로 분할되며 분할된 이진 문제에 대한 각 이진 분류기의 출력

을 종합 하여 최종 클래스를 결정한다. 출력코딩 방법에는 OPC(One-Per-Class), All-Pairs가 대표적으로 알려져 있으나 non-sense output 때문에 이를 보완하기 위한 다양한 방법들이 모색 중에 있다. OPC에서 각 이진 분류기는 하나의 클래스와 나머지 클래스들을 구분하고 All-Pairs에서는 하나의 클래스를 또 다른 하나의 클래스와 구분하는 방식이다. 상기 두 방법은 다중 클래스 문제를 여러 개의 이진 클래스 문제로 분할하고 이들을 다시 종합하여 최종 결정을 내리는 출력코딩이라는 일반적인 방법에 속한다. 음성 인식과 같이 클래스의 수가 많고 학습 데이터가 적은 경우에는 하나의 이진 학습기가 모든 클래스를 학습하는 OPC계열의 출력코딩 방식이 적합함을 보이고 있으므로 본 연구에서는 OPC를 적용하였다. 이를 구체적으로 살펴보면 각 이진 분류기는 하나의 클래스와 나머지 다른 클래스와 구분하도록 학습된다. 만약 K개의 클래스가 있는 경우, 양의 클래스를 하나씩만 가지는 K개의 이진 분류기를 생성하게 된다. 복원은 주로 각 이진 분류기의 출력값이 가장 큰 분류기에 학습된 양의 클래스로 결정하는 방식을 따른다.

3. 음소단위로의 자동 분할

SVM을 이용한 멀티 클래스 분류를 통하여 초성 · 중성 · 종성에 대한 프레임 단위의 1차 분류를 기반으로 하여 2차 음소경계를 결정하기 위해서는 1차 분류 결과의 오 인식에 대한 보상처리가 필요하다. 이를 위해서 본 연구에서는 1차 인식결과에 단구간의 최빈값을 이용하는 후처리 알고리즘을 제안하고 적용하였다. 음소단위로의 자동 분할 처리의 전체 흐름도를 <그림 2>에 나타내었다.

Step 1 : 멀티 클래스 분류 결과를 초 · 중 · 종성으로 mapping처리

Step 2 : 시작 프레임과 종료 프레임의 전, 후에 의사 보상 값을 패딩, 패딩 프레임의 수는 짝수

Step 3 : 단구간 프레임 크기를 한 프레임씩 슬라이딩하면서 최빈 인덱스를 보상 후 인덱스로 설정

Step 4 : (패딩 프레임 수/2)의 프레임을 전후 측정

Step 5 : Step 4의 결과를 바탕으로 최종 음소 경계 결정

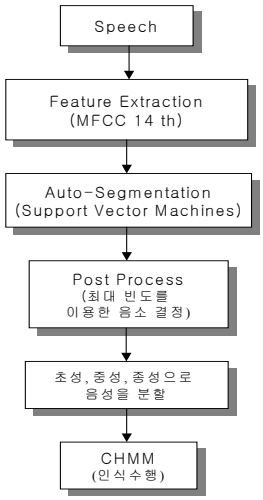

그림 2. 음소단위 자동 분할 처리 흐름도

Figure 2. Flow diagram of Automatic Segmentation by Phoneme Unit

4. 실험결과 및 고찰

4.1 시뮬레이션 조건

음소분할 실험을 위하여 CVC형 108음절 데이터베이스를 구성하였다. 본 데이터베이스는 우리말 음성의 CVC형 음절로 초성(ㅂ, ㄷ,ㄱ, ㅍ, ㅌ, ㅋ), 중성(ㅏ, ㅓ, ㅗ, ㅖ, ㅜ, ㅣ), 종성(ㄴ, ㄹ, ㅁ)으로 이루어진 120개의 유사 음절로 구성하였으며 5명의 화자가 5회 발성하여 3회분은 학습용으로 나머

지 2회분은 평가용으로 이용하였다. 목측에 의하여 초성, 중성, 종성을 분리한 음소 데이터베이스를 별도 구성하여 학습을 행하고 분할과 인식실험을 하였으며 음성 분석조건은 <표 1>과 같다.

표 1. 음성분석조건
Table 1. Conditions of speech analysis

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

4.2 시뮬레이션 및 결과

분할 방법은 SVMs, GMM 및 목측으로 각각 행하고 그 결과를 서로 비교하였다. SVMs은 초성 · 중성 · 종성으로 3개의 클래스로 하였고 GMM은 초성 · 중성 · 종성의 15개 클래스로 다중 분류를 행하고, 이를 초성 · 중성 · 종성으로 mapping하여 제안하는 후처리 알고리즘으로 그 결과를 보상하는 방법으로 음소 경계를 결정하였다.

표 2. 자동분할의 성능 (목측분할과의 편차)
Table 2. Performance of automatic segmentation (Deviation from segmentation by eye measurement)

Frame		초성	중성	종성
GMM	평균	6.32	4.94	4.68
	표준편차	6.86	6.24	6.35
SVMs	평균	3.12	5.46	5.58
	표준편차	3.32	6.48	8.30

<표 2>와 <표 3>에서 보듯이 시뮬레이션 결과로 알 수 있듯이 전반적으로는 SVMs이 양호하며 특히 초성에서 더 우수하며 중성, 종성에서는 GMM이 성능이 비교적 우수함을 확인할 수 있었다.

표 3. 음소단위 분할 후의 인식성능 (%)
Table 3. Recognition rates after segmentation by phoneme unit

Error	GMM분할	SVMs분할	목측분할
초성	23.80	20.03	13.89
중성	5.46	6.79	3.98
종성	8.15	9.07	5.37

5. 결 론

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 주목할 만한 연구과제이다. 본 연구에서는 SVMs을 이용해서 자동 음소분할을 시도하고 이를 통하여 음성인식을 행하고 GMM을 이용한 것과 그 성능을 서로 비교하였다. 시뮬레이션 결과 음성의 SVMs은 초성에서 GMM은 중성, 종성에서 각각 비교적 우수한 분할 성능을 확인했다. 현재 SVMs은 많은 연산량과 학습시간을 필요로 하지만 프로세서의 성능의 비약적 발전으로 문제가 되지 않으며 이는 무성음, 파열음, 마찰음 등의 비선형성을 가지는 음소모델에 확률 및 통계보다 나은 접근 방법으로 확인되고 있다. 그러므로 SVMs을 보다 더 최적화하기 위한 여러 가지 방법들이 연구되어야 할 것으로 생각되며 향후 SVMs과 GMM을 융합한 하이브리드 구조의 최적 음소 분할기를 구성하고 연속 음성인식에 적용할 계획이며 클래스수가 많은 경우에도 효과적으로 SVMs을 적용할 수 있도록 문제 복잡도를 최소화하면서 이진 분류기의 수를 줄일 수 있는 방향으로 연구를 계속 진행할 것이다.

References

[1] J. Ghosh, *Multiclassifier systems: back to the future*, Proceedings of the 3rd International Workshop on Multiple Classifier Systems, Lecture Note in Computer Science. Vol. 2364, pp.1-15, 2002.

[2] N. Cristianini and J. Shawe Taylor, *An introduction to support vector machine and other kernel based learning methods*, Cambridge University Press, 2000.

[3] Bernhard Scholkopf, Alexander J. Smola, "*Learning with Kernels*", The MIT Press, 2002

[4] Christopher J.C. Burges, *A tutorial on support vector machines for pattern recognition*, Bell Laboratories, Lucent Technologies. 2008.

[5] J. Weston, C. Watkins, *Multi-class support vector machines*, Technical Report, Royal Holloway, University of London, 2008.

[6] Yonas B. Dibike, Slavco Velickov, Dimitri Solomatine, Michael B. Abbott, *Model introduction with support vector machines: introduction and applications*, ASCE Journal of Computing in Civil Engineering. Vol. 15, No. 3, pp.208-216, 2007.

[7] Edgar E. Osuna, Robert Freund and Federico Girosi, *Support vector machines: training and applications*, C.B.C.L Paper, No. 144, 2007.

[8] Philip Clarkson and Pedro J. Moreno, *On the use of support vector machines for phonetic classification*, ICASSP 2005.

[9] X. Peng, *TPMSVM: A novel twin parametric margin support vector machine for pattern recognition*, pattern recognition, Vol. 44, No. 10-11, pp.2678-2692, 2011.

경남과학기술대학교 전자공학과

요 약

본 연구에서는 음성의 학습 및 인식 단위로서 연속음성을 초성, 중성, 종성의 음소단위로 분할하기 위하여 인공 신경회로망의 하나인 SVMs을 사용하였으며 분할한 음소단위의 음성으로 연속음성인식에 적용하여 그 성능을 살펴보았다. 음소경계는 단 구간에서의 최대 주파수를 가진 알고리즘에 의하여 결정되며 또한 음성인식처리는 CHMM에 의하여 이루어지며 GMM 및 목측에 의한 분할결과와도 비교하여 살펴보았다. 시뮬레이션 결과로부터 분할성능에서 전반적으로는 SVMs이 양호하며 특히 초성에서 더 우수하며 중성, 종성에서는 GMM이 보다 효율적임을 알 수 있었다. 향후 SVMs과 GMM을 결합한 하이브리드 구조의 최적 음소 분할기를 구성하고 연속 음성인식에 적용할 계획이다.

감사의 글

본 논문은 2012년도 경남과학기술대학교 연구비 지원에 의하여 연구되었음.

Gwang Seok Lee received the bachelor's degree in the Department of Electronic Engineering from the Dong-A University in 1983. He received the M.S. degree and the Ph.D. degree in the Department of Electronic Engineering from Dong-A University in 1985 and 1992, respectively. He has been a professor in the Department of Electronic Engineering at Gyeongnam National University since 1995. His current research interests include artificial intelligence, intelligent systems, biometrics. He is a life member of the KKITS.

E-mail address: kslee@gntech.ac.kr