

특징 강화 기법과 학습 데이터 길이 조절에 의한 Supervector Linear Kernel SVM 화자식별 개선

Improvement in Supervector Linear Kernel SVM for Speaker Identification Using Feature Enhancement and Training Length Adjustment

소 병 민, 김 경 화*, 김 민 석**, 양 일 호, 김 명 재, 유 하 진
(Byung-Min So, Kyung Wha Kim*, Min-Seok Kim**, Il-Ho Yang, Myung-Jae Kim, Ha-Jin Yu)

서울시립대학교 컴퓨터과학부, *대검찰청 음성분석실, **LG 전자기술원
(접수일자: 2011년 6월 22일; 채택일자: 2011년 7월 27일)

본 논문에서는 supervector linear kernel SVM을 사용한 화자식별 시스템의 성능을 개선하는 방법을 제안하였다. 제안한 방법은 긴 학습 데이터를 여러 개의 짧은 학습 데이터로 분할하는 것을 기본 아이디어로 하고 있다. 제안한 방법의 성능을 평가하기 위해 서로 다른 4가지 데이터베이스에 PCA, GKPCA, KMDA를 사용하여 특징 강화를 하고 실험한 뒤 결과를 분석하였다. 실험 결과 제안한 방법이 supervector linear kernel SVM을 사용한 화자 식별 성능을 향상 시키는 것을 확인하였다.

핵심용어: 화자식별, 서포트 벡터 머신, 주성분 분석, 그리디 커널 주성분 분석, 커널 다중 판별 분석

투고분야: 음성처리 분야 (2,5)

In this paper, we propose a new method to improve the performance of supervector linear kernel SVM (Support Vector Machine) for speaker identification. This method is based on splitting one training datum into several pieces of utterances. We use four different databases for evaluating performance and use PCA (Principal Component Analysis), GKPCA (Greedy Kernel PCA) and KMDA (Kernel Multimodal Discriminant Analysis) for feature enhancement. As a result, the proposed method shows improved performance for speaker identification using supervector linear kernel SVM.

Keywords: Speaker identification, SVM, PCA, GKPCA, KMDA

ASK subject classification: Speech Signal Processing (2,5)

I. 서론

최근 스마트폰과 같은 기기의 발전으로 음성 기술에 대한 관심이 증가하고 있다. 음성 기술은 기기와 음성 외에 다른 도구를 필요로 하지 않으므로 매우 편리하다. 음성 기술에는 어떤 음성이 무슨 말을 한 것인지 알아내는 음성인식 기술, 어떤 음성이 누구의 음성인지 알아내는 화자인식 기술 등이 있다. 그리고 화자인식 기술은 여러 사람 중 누구의 음성인지 알아내는 화자식별 기술과 어떤 목소리가 특정 사람의 목소리가 맞는지 확인하는 화자확인 기술로 나뉜다. 디지털 과학수사와 같은 분야에서 화자식별 기술이 사용될 수 있는데 그 결과가 법정 증빙도

구로 사용되는 등 기술의 중요성이 증가하고 있다.

최근 패턴인식 분야에서 SVM (support vector machine) [1]을 분류기로 많이 사용하고 있는데 SVM이 실험적으로 다른 분류기보다 높거나 비슷한 성능을 보여주기 때문이다. SVM은 선형 분류기이지만 선형 분류가 불가능한 상황에서 선형 분류를 하기 위해 커널함수 (kernel function) 를 사용한다. 비선형 분류를 하기 위해 입력 공간 (input space)의 샘플을 입력 공간보다 높은 차원인 특징 공간 (feature space)으로 매핑하는 것이다. Mercer의 정리를 만족하는 여러 가지의 커널함수가 있는데 그 중 linear kernel, polynomial kernel, RBF kernel가 널리 쓰인다. 그리고 화자 인식을 위해 GMM (Gaussian mixture model) [2]과 SVM을 결합한 GMM supervector linear kernel SVM 이라는 방법도 사용되고 있다 [1]. GMM-UBM (universal background model) [3]에서 각 가우시안 혼합의 평균들

책임저자: 유 하 진 (hju@uos.ac.kr)
서울특별시 동대문구 전농동 90번지 서울시립대학교 컴퓨터과학부
(전화: 02-2210-5613)

을 일렬로 나열한 것을 supervector라고 한다. UBM에 발성 (utterance)을 MAP [4] 적용한 후 얻은 supervector 값을 커널 대치로 특징 공간으로 사상한 뒤 SVM 분류를 하는 방법을 GMM supervector linear kernel SVM이라 한다. 학습에 사용된 발성의 수가 적다면 특징 공간상에 모델을 표현하는 샘플의 수도 적다고 할 수 있다. 하나의 발성을 UBM에 MAP적용하여 얻은 supervector로 모델을 학습하게 된다면 하나의 supervector가 최외각 벡터, 즉 서포트 벡터의 역할을 하게 된다. 학습된 다른 모델들 간의 분류에 모두 같은 서포트 벡터를 사용하게 된다면 분류가 제대로 되지 않을 수 있다.

본 논문에서는 모델을 표현하는 샘플들의 수를 늘리기 위해 학습 데이터를 분할하는 방법을 제안한다. 학습 데이터를 분할하게 되면 SVM 학습을 위해 사용되는 서포트 벡터의 선택에 다양성이 생기므로 성능에 좋은 영향을 줄 수 있다. 제안한 방법의 성능을 측정하기 위해 분할된 학습 데이터와 분할되지 않은 학습 데이터를 이용한 실험 결과를 비교하였다. 그리고 학습 데이터 분할의 영향이 특징 강화된 음성 DB에 대해서도 동일하게 작용하는지 알아보는 실험도 진행하였다. 특징 강화 기법으로 PCA (principal component analysis) [5], GKPCA (greedy kernel PCA) [6], KMDA (kernel multimodal discriminant analysis) [7]를 사용하였다.

본 논문은 2장에서 GMM supervector linear kernel SVM에 대해 소개한다. 다음으로 3장에서 본 논문에서 사용한 특징 강화 기법들에 대해 설명한다. 그리고 4장에서는 화자식별 실험 설계와 실험 결과에 대해 설명하고 마지막으로 5장에서 결론을 맺는다.

II. GMM supervector linear kernel SVM

2.1. SVM의 기본 아이디어

SVM은 기계 학습 알고리즘 중 하나이다. 그림 1과 같이 두 부류의 학습 데이터가 주어졌을 때 각각의 최외각에 있는 서포트 벡터를 이용하여 분류하는 최적의 초평면을 찾는다. 두 부류의 서포트 벡터 사이에 여러 개의 초평면이 존재할 수 있는데 그 중 최적의 초평면은 가장 가까운 서포트 벡터와 초평면 사이의 거리의 두 배인 마진 (margin)이 최대가 되는 초평면 (b)이다. 이 초평면 (b)은 두 부류의 중심을 이용하여 구한 초평면 (a)보다 일반화 (generalization) 능력이 뛰어나기 때문에 미지의 샘플이 들어 왔을 때 더 잘 분류할 수 있게 된다. 실제로는 선형 분류 가능한 상황보다 불가능한 상황이 많은데, 선형 분

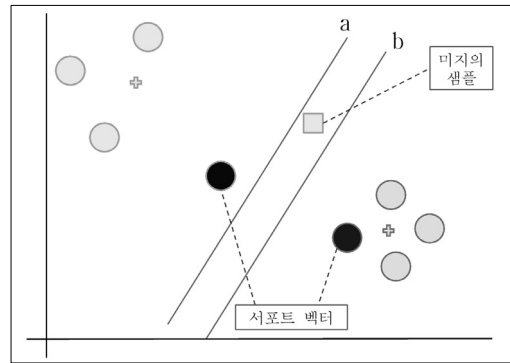


그림 1. SVM의 기본 아이디어
Fig. 1. basic idea of SVM.

류가 불가능한 상황에서는 샘플들을 입력 공간보다 더 높은 차원의 특징 공간으로 매핑하는 커널대치를 사용하여 선형 분리 불가능한 샘플들을 선형 분리 가능하게 만들 수 있다.

2.2. GMM supervector linear kernel SVM의 학습 과정

GMM supervector linear kernel SVM의 학습은 그림 2와 같은 과정에 의해 이루어진다. 먼저 UBM (universal background model) [3]생성을 위한 음성 DB에서 MFCC (mel-frequency cepstral coefficients)특징을 추출한 뒤 UBM을 생성한다. UBM은 많은 수의 화자가 발성한 음성을 모아서 하나의 모델을 학습한 것으로, 일반화된 화자의 모델이 된다. 그리고 UBM에 화자 모델 생성을 위한 음성 DB를 MAP [4] 적용한 뒤 supervector를 얻는다. 이렇게 각 음성을 적용한 뒤 생성된 supervector들에 커널 함수를 적용하여 모델을 표현하고 여러 모델을 분류하는 여러 개의 초평면을 구한다.

2.3. 제안한 방법

본 논문에서 제안한 방법은 그림 3과 같다. SVM은 기본적으로 두 부류를 분류하는 이진 분류기인데 N개의 부류를 분류하기 위해 1개의 부류와 나머지 N-1개의 부류를 분류하는 N개의 분류기를 찾는 방법을 사용한다. GMM supervector linear kernel 학습에 1개의 발성을 사용한다면 특징 공간상에 모델을 표현하는 샘플이 1개이기 때문에 N개의 분류기에 대해서 모든 서포트 벡터가 고정된다고 할 수 있다. 이 경우 분류가 제대로 되지 않을 수 있기 때문에 학습 데이터를 분할하는 방법을 사용한다. 학습 데이터를 분할하면 모델을 표현하는 샘플의 수가 늘어나 서포트 벡터 선택에 다양성이 생겨 분류 성능에 좋은 영향을 줄 수 있다.

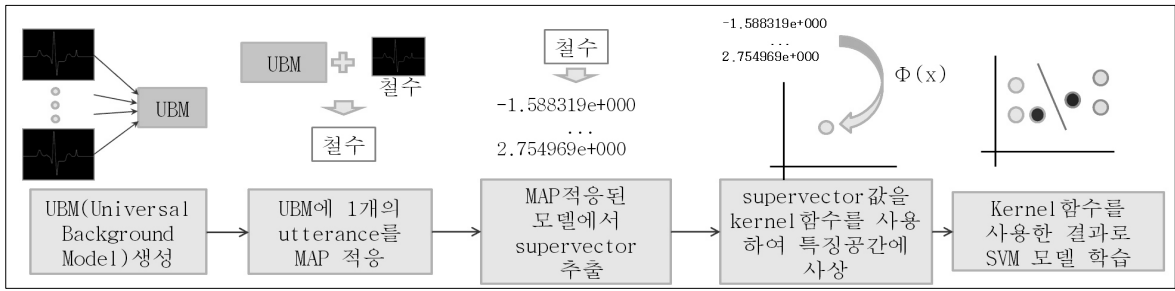


그림 2. GMM supervector linear kernel SVM의 학습 과정
 Fig. 2. training process of GMM supervector linear kernel SVM.

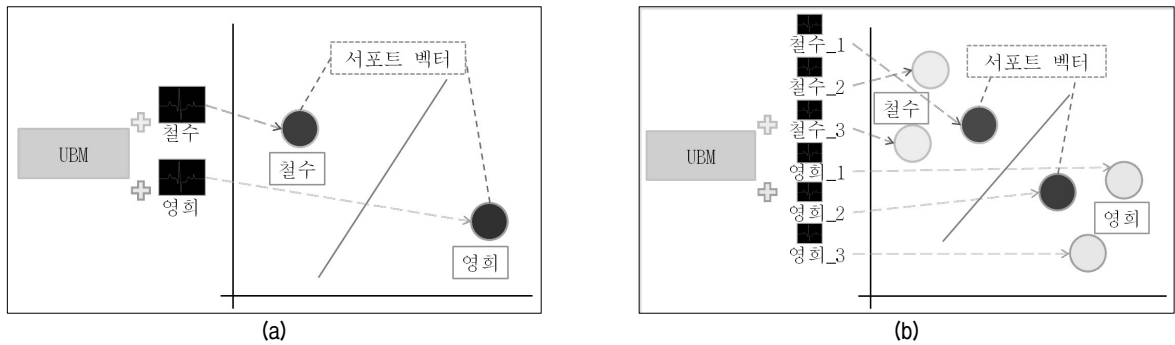


그림 3. 제안한 방법
 (a) 1개의 학습 데이터를 사용한 SVM학습 (b) 여러 개의 학습 데이터를 사용한 SVM학습
 Fig. 3. Proposed method.
 (a) SVM training using one training datum, (b) SVM training using several training data

III. 특징 강화

3.1. PCA (주성분 분석, principal component analysis)

PCA [5]는 N차원의 특징 중 정보의 손실을 최소화하는 $K(K \leq N)$ 차원의 기저를 찾는 방법이다. 고유값(eigenvalue)을 계산하고 고유값의 크기순으로 K개를 선택한 뒤 K개의 고유값과 매칭되는 고유벡터(eigenvector)로 정보의 손실을 최소화하는 K차원으로 사상하는 변환행렬을 구할 수 있다.

3.2. GKPCA (그리디 커널 주성분 분석, greedy kernel PCA)

PCA의 기본 아이디어에 커널 함수를 적용한 것이 KPCA(kernel PCA) [8]라면, KPCA에 그리디 필터링(greedy filtering)을 더한 것이 GKPCA [6]이다. 비선형 분류에서 좋은 성능을 가져오는 KPCA는 특징의 수에 비례하게 계산량이 늘어나고 요구 메모리가 커지는 단점을 갖고 있다. 이 문제를 해결하기 위해 GKPCA에서는 KPCA의 변환행렬을 찾을 때 전체 데이터가 아닌 그리디 필터링으로 선택된 일부의 부분 집합을 사용한다.

3.3. KMDA (커널 다중 판별 분석, kernel multimodal discriminant analysis)

KMDA [7]는 커널함수를 사용하여 특징 공간상으로 매핑된 데이터를 클러스터링한 뒤 각 서브 클러스터들의 중심과 전체 중심과의 거리를 최대로 하는 변환 축을 찾아 특징을 강화하는 특징 강화 기법이다. 각 화자 모델을 특징 공간상의 k개로 군집화하기 위해 커널 k-means 알고리즘을 사용한다.

IV. 화자식별 실험

본 논문에서는 2가지 방법의 실험을 진행하였다. 그 방법은 긴 학습 데이터를 여러 개의 학습 데이터로 분할하는 방법과 여러 개의 학습 데이터를 하나의 학습 데이터로 결합하는 방법이다. 그리고 데이터베이스는 2003년 NIST 화자인식 평가에서 사용된 휴대 전화 통화 DB, 2008년에 대검찰청 음성분석실에서 수집한 쌍둥이 음성 DB, 2002년 ETRI에서 수집한 한국어 증가 마이크 화자인식용 DB 그리고 TIMIT DB를 사용하였다.

4.1. 학습 데이터 길이에 따른 화자식별 성능 비교 실험

NIST DB와 쌍둥이 DB의 경우에는 기존 학습 데이터의

길이가 100초 이상이기 때문에 2초, 5초 등 다양한 길이로 분할이 가능하지만 ETRI DB와 TIMIT DB는 학습 데이터가 3초 이하의 여러 개의 파일들로 구성되어 있어 다양한 길이로 분할이 어렵다. 그래서 여러 개의 파일을 2~5개씩 묶어 역으로 실험을 진행하였다.

실험에는 KMDA를 사용한 [7]과 동일한 조건으로 실험을 하기 위해 20차 MFCC 특징에 로그 에너지를 더한 21차 특징을 사용하였다. 채널 보상 방법으로는 CMS (cepstral mean subtraction)와 각 차원의 분산을 1로 만드는 분산 정규화 기법 (variance normalization)을 사용하였다. 쌍둥이 DB를 제외한 각 음성 DB의 사일런스를 제거하기 위해 특징 추출 후 임계점 이하의 에너지를 갖는 부분을 제거 하였다. 각 음성 DB에서 사용된 화자 수, 화자 한 명당 학습에 사용된 길이, 화자 한 명당 테스트에 사용된 길이, sampling rate는 표 1과 같다.

NIST DB 실험에서는 학습에 사용되지 않은 남녀 80명을 이용하여 UBM을 생성하였다. 학습 데이터 분할이 화자 식별률에 미치는 영향을 확인하기 위해 5초, 10초, 20초, 25초, 50초, 110~130초 (분할하지 않음) 각각의 조건으로 실험을 진행하였다. 학습 데이터로부터 HTK를 이용하여 특징을 뽑은 뒤 사일런스 제거 후 분할하였다.

쌍둥이 DB 실험에서는 ETRI의 화자인식용 휴대전화 DB의 주차 데이터 남녀 100명 (남자 50명, 여자 50명)을 사용하여 UBM을 생성하였다. 학습 데이터를 0.5초, 0.75초, 1초, 2초, 3초, 4초, 6초, 12초 (분할하지 않음)로 구분하여 실험을 진행하고 결과를 확인하였다.

ETRI DB 실험에서는 월차 데이터 남녀 100명의 첫 번째 달 1회 발성 중, 단문 발화 전부를 사용하여 UBM을 생성하였다. 학습 데이터 결합이 화자 식별률에 미치는 영향을 확인하기 위해 1개 (결합하지 않음), 2개씩 연결, 5개씩 연결, 10개 전부 연결 각각의 조건으로 실험을 진행하였다. 학습 데이터 결합은 HTK에서 특징을 뽑은 뒤 사일런스 제거 후 결합하였다.

TIMIT DB 실험에서는 UBM 생성에 테스트 데이터 남녀 168명 (남자 112명, 여자 56명)을 사용하였다. 화자별 학습 데이터는 약 15초를 사용하였고 테스트 데이터도

약 15초를 사용하였다. 학습 데이터를 1개 (결합하지 않음), 2개씩 연결, 5개 전부 연결로 구분하여 실험을 진행하고 결과를 확인하였다. 학습 데이터 결합은 HTK에서 특징을 뽑은 뒤 사일런스 제거 후 결합하였다.

각각의 음성 DB에 대해 UBM 생성을 위해 사용하는 음성 DB로 특징 강화를 위한 변환 행렬을 구하였다. 그리고 변환 행렬로 UBM을 만드는 음성 DB와 학습 데이터, 그리고 테스트 데이터를 강화하여 특징 강화 실험을 진행하였다. GKPCA의 그리디 필터링으로 선택하는 샘플의 수는 50개로 하고 KMDA에서 서브 클러스터 개수인 k는 4로 설정하였다.

4.2. 화자식별 실험 결과

실험 결과에서 특징 강화를 하지 않은 것은 MFCC, PCA와 GKPCA 그리고 KMDA로 특징 강화를 한 것은 각각 PCA, GKPCA, KMDA로 표기 하였다.

NIST DB를 사용하여 실험한 화자 식별률은 표 2, 3, 4와 같다.

표 2. 채널 보상을 하지 않은 NIST DB 실험 결과
Table 2. Results of NIST DB experiment without channel compensation.

길이 (초)	MFCC	PCA	GKPCA	KMDA
5	70.50 %	66.93 %	68.23 %	69.69 %
10	70.34 %	66.12 %	68.07 %	69.69 %
20	69.36 %	64.34 %	67.58 %	69.53 %
25	68.88 %	64.18 %	66.61 %	68.39 %
50	66.12 %	62.07 %	65.64 %	66.45 %
110~130	64.34 %	61.75 %	64.82 %	64.99 %

표 3. CMS를 한 NIST DB 실험 결과
Table 3. Results of NIST DB experiment with CMS.

길이 (초)	MFCC	PCA	GKPCA	KMDA
5	71.96 %	72.12 %	70.17 %	73.74 %
10	72.60 %	69.69 %	68.88 %	74.06 %
20	70.50 %	69.53 %	69.36 %	74.06 %
25	69.36 %	68.39 %	68.88 %	71.79 %
50	66.77 %	67.26 %	69.04 %	70.50 %
110~130	64.01 %	66.45 %	66.45 %	68.07 %

표 4. CMS와 분산 정규화를 한 NIST DB 실험 결과
Table 4. Results of NIST DB experiment with CMS and variance normalization.

길이 (초)	MFCC	PCA	GKPCA	KMDA
5	67.90 %	66.28 %	69.53 %	68.55 %
10	66.28 %	65.64 %	68.39 %	68.55 %
20	64.99 %	65.15 %	66.61 %	67.74 %
25	64.18 %	65.15 %	66.77 %	67.42 %
50	64.34 %	64.34 %	66.28 %	68.39 %
110~130	63.37 %	63.53 %	64.99 %	65.96 %

표 1. DB별 실험 설계
Table 1. experiment design for each database.

	NIST	쌍둥이	ETRI	TIMIT
화자 수	100	60	100	462
학습 (초)	110~130	12	20	15
테스트 (초)	180~210	120	20	15
sampling rate	8 kHz	8 kHz	8 kHz	16 kHz

표 5. 채널 보상을 하지 않은 쌍둥이 DB 실험 결과

Table 5. Results of twins DB experiment without channel compensation.

길이 (초)	MFCC	PCA	GKPCA	KMDA
0.5	74.84 %	69.78 %	78.22 %	78.98 %
0.75	73.01 %	74.38 %	77.45 %	78.98 %
1	75 %	76.07 %	76.99 %	78.52 %
2	77.60 %	77.30 %	75 %	75.61 %
3	79.75 %	76.99 %	74.23 %	75.15 %
4	75.92 %	74.53 %	75.30 %	72.69 %
6	66.87 %	73.61 %	74.07 %	71.93 %
12	40.18 %	50.30 %	65.33 %	58.74 %

표 6. CMS를 한 쌍둥이 DB 실험 결과

Table 6. Results of twins DB experiment with CMS.

길이 (초)	MFCC	PCA	GKPCA	KMDA
0.5	80.21 %	83.58 %	92.02 %	92.02 %
0.75	83.89 %	85.58 %	91.41 %	91.56 %
1	84.96 %	86.80 %	91.71 %	90.49 %
2	87.11 %	88.65 %	91.56 %	90.49 %
3	86.65 %	89.41 %	90.33 %	89.72 %
4	80.36 %	88.49 %	87.73 %	85.88 %
6	76.53 %	86.96 %	84.04 %	83.58 %
12	55.67 %	78.37 %	70.70 %	66.71 %

표 7. CMS와 분산 정규화를 한 쌍둥이 DB 실험 결과

Table 7. Results of twins DB experiment with CMS and variance normalization.

길이 (초)	MFCC	PCA	GKPCA	KMDA
0.5	84.35 %	82.05 %	84.96 %	87.73 %
0.75	84.96 %	84.66 %	85.27 %	89.57 %
1	86.19 %	84.81 %	86.04 %	87.11 %
2	85.27 %	83.58 %	86.04 %	86.34 %
3	86.50 %	84.35 %	84.81 %	86.80 %
4	79.29 %	82.05 %	80.98 %	80.82 %
6	79.75 %	79.60 %	83.74 %	82.05 %
12	65.95 %	72.08 %	77.60 %	72.69 %

실험 결과 특징 강화를 하지 않은 실험이나 특징 강화를 한 실험에서 약 100초의 학습 데이터를 5초나 10초 길이를 갖는 여러 개의 학습 데이터로 분할한 결과가 분할하지 않았을 때의 식별률보다 높은 식별률을 보여주었다. 학습 데이터를 분할할수록 성능이 향상되는 것을 확인하여 5초 이하로 분할하는 실험을 하려 하였으나, 이 경우의 supervector의 수가 많아져서 메모리 부족으로 실험이 불가능하였다 (CPU : Inter Xeon X5650, RAM : 4 GB, OS : Windows 7 32 bit).

쌍둥이 DB를 사용하여 실험한 결과는 표 5, 6, 7과 같다.

표 8. 채널 보상을 하지 않은 ETRI DB 실험 결과

Table 8. Results of ETRI DB experiment without channel compensation.

연결 개수	MFCC	PCA	GKPCA	KMDA
1	99.1 %	99.8 %	99 %	99.2 %
2	97.5 %	98.7 %	97.9 %	97.9 %
5	88.1 %	88.9 %	86.3 %	86.3 %
10	57.2 %	62.9 %	64.2 %	67 %

표 9. CMS를 한 ETRI DB 실험 결과

Table 9. Results of ETRI DB experiment with CMS.

연결 개수	MFCC	PCA	GKPCA	KMDA
1	94.8 %	97.8 %	97 %	97.3 %
2	89.3 %	94.5 %	95.3 %	94.6 %
5	70.2 %	77 %	78 %	77.5 %
10	41.5 %	50.6 %	51.7 %	53.7 %

표 10. CMS와 분산 정규화를 한 ETRI DB 실험 결과

Table 10. Results of ETRI DB experiment with CMS and variance normalization.

연결 개수	MFCC	PCA	GKPCA	KMDA
1	95.7 %	96.2 %	97.5 %	96.9 %
2	91.4 %	94.2 %	95.4 %	95.7 %
5	76.5 %	74.3 %	77.8 %	73.9 %
10	42.4 %	57.8 %	57.6 %	57.5 %

실험 결과 특징 강화를 하지 않은 실험이나 특징 강화를 한 실험에서 12초의 학습 데이터를 여러 개의 학습 데이터로 분할한 결과가 분할하지 않은 것보다 높은 식별률을 보여주었다. 하지만 일부 실험에서 너무 짧은 단위로 학습 데이터를 분할하면 식별률이 떨어지는 결과가 나타났다. 이는 UBM에 MAP적응하는 데이터가 너무 짧아 화자 모델의 특성을 잘 표현하지 못한 것으로 판단된다. 추후 적절한 분할 길이를 찾는 방법의 연구가 필요할 것으로 보인다.

ETRI DB를 사용하여 실험한 결과는 표 8, 9, 10과 같다.

실험 결과 특징 강화를 하지 않은 실험이나 특징 강화를 한 실험에서 여러 개의 학습 데이터를 연결하지 않은 경우가 가장 높은 식별률을 보여주었다. 10개의 학습 데이터를 2개씩 연결하여 5개의 학습 데이터로 만든 경우에는 식별률이 많이 떨어지지 않았지만 5개씩 연결하여 2개의 학습 데이터로 만들거나 10개 전부를 연결하여 1개의 학습 데이터로 만든 경우에는 식별률이 많이 떨어지는 것을 확인하였다.

TIMIT DB를 사용하여 실험한 결과는 표 11, 12 13과 같다.

표 11. 채널 보상을 하지 않은 TIMIT DB 실험 결과
Table 11. Results of TIMIT DB experiment without channel compensation.

연결 개수	MFCC	PCA	GKPCA	KMDA
1	97.61 %	97.48 %	97.96 %	98.09 %
2	94.58 %	93.59 %	95.02 %	95.49 %
5	65.23 %	63.98 %	62.07 %	61.34 %

표 12. CMS를 한 TIMIT DB 실험 결과
Table 12. Results of TIMIT DB experiment with CMS.

연결 개수	MFCC	PCA	GKPCA	KMDA
1	91.25 %	92.29 %	93.76 %	95.10 %
2	84.24 %	81.68 %	84.32 %	86.62 %
5	43.20 %	39.39 %	44.45 %	44.80 %

표 13. CMS와 분산 정규화를 한 TIMIT DB 실험 결과
Table 13. Results of TIMIT DB experiment with CMS and variance normalization.

연결 개수	MFCC	PCA	GKPCA	KMDA
1	91.99 %	92.33 %	91.55 %	93.24 %
2	85.19 %	82.94 %	83.41 %	84.06 %
5	51.38 %	49.26 %	46.40 %	48.48 %

실험 결과 특징 강화를 하지 않은 실험이나 특징 강화를 한 실험에서 여러 개의 학습 데이터를 연결하지 않은 경우가 가장 높은 식별률을 보여주었다.

V. 결론

본 논문에서는 SVM 학습 데이터의 길이 조절과 특징 강화 기법이 화자 식별률에 미치는 영향에 대해 연구하였다. SVM에서 모델을 학습할 때 동일한 학습 데이터 길이를 사용하더라도 여러 개로 나뉜 것과 하나로 되어있는 것의 화자 식별률이 다를 것이라고 생각하고 연구를 진행했다. SVM에서 사용한 커널은 GMM supervector linear kernel이고 특징 강화 기법으로 PCA, GKPCA, KMDA를 사용하였다.

4가지의 서로 다른 데이터베이스를 사용한 실험 결과, 일부의 경우를 제외하고 특징 강화를 했을 때가 특징 강화를 하지 않았을 때 보다 좋은 결과를 보여주었다. 그리고 대부분의 경우에서 KMDA와 GKPCA가 PCA보다 높은 성능을 보였다. SVM의 학습 데이터의 길이 조절에 따른 실험의 결과로 여러 개로 나누어진 학습 데이터를 모델 학습에 사용한 경우가 그렇지 않은 경우보다 더 높은 식별률을 보인다는 것을 알 수 있었다. 하지만 일부 실험에서 너무 짧은 단위로 학습 데이터를 분할할 경우 오히려 식별률이 떨어지는 결과를 보였다. 향후 본 실험 결과를

바탕으로 supervector linear kernel SVM 학습에 있어 최적의 분할 길이를 찾아내는 연구를 진행하고자 한다.

감사의 글

이 논문은 2010년 대검찰청 연구용역의 지원으로 수행된 연구입니다 (과제명: 휴대폰 및 인터넷 전화음성의 화자식별률 제고 방안 연구).

참고 문헌

1. W.M. Campbell, D.E. Sturim, D.A. Reynolds, "Support Vector Machines using GMM Supervectors for Speaker Verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308-311, 2006.
2. Douglas A. Reynolds and Richard C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 1, pp. 72-83, 1995.
3. Douglas A. Reynolds, Thomas F. Quatieri and Robert B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19-41, Jan. 2000.
4. J.-L. Gauvain and C.-H. Lee, "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *IEEE Trans. Speech Audio Proc.*, vol. 2, no. 2, pp. 291-298, Apr. 1994.
5. Smith, L. I, "A tutorial on Principal Components Analysis", 2002.
6. 김민석, 양일호, 유하진, "Greedy Kernel PCA를 이용한 화자식별", *말소리*, 66호, 105-116쪽, 2008.
7. Kim, M-S., Yang, I-H., Yu, H-J., "Kernel multimodal discriminant analysis for speaker verification", *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 4498-4501, 2010.
8. B. Scholkopf, A. Smola and K.-R. Muller, "Kernel Principal Component Analysis," *In Int. Conf. on Artificial Neural Networks*, pp. 583-588, 1997.

저자 약력

•소 병 민 (Byung-Min So)



2010년: 서울시립대학교 컴퓨터과학부 (학사)
2010년 ~ 현재: 서울시립대학교 컴퓨터과학부 석사과정
※ 주관심 분야: 화자인식, 음성인식

•김 경 화 (Kyung Wha Kim)



1999년: 서울대학교 언어학과 음성학 전공 (석사)
2007년: 서울대학교 언어학과 음성학 전공 (박사)
2000년 ~ 현재: 대검찰청 과학수사담당관실 음성분석실장
※ 주관심 분야: 법음성학, 화자식별

•김 민 석 (Min-Seok Kim)



2006년: 서울시립대학교 컴퓨터과학부 (학사)
2008년: 서울시립대학교 컴퓨터과학부 (석사)
2011년: 서울시립대학교 컴퓨터과학부 (박사)
2011년 ~ 현재: LG전자 전자기술원 주임연구원
※ 주관심 분야: 화자 인식, 음성 인식

•김 명 재 (Myung-Jae Kim)



2010년: 서울시립대학교 컴퓨터과학부 (학사)
2010년 ~ 현재: 서울시립대학교 컴퓨터과학과 석사
과정
※ 주관심 분야: 화자 인식, 음성 인식

•양 일 호 (Il-Ho Yang)



2008년: 서울시립대학교 컴퓨터과학부 (학사)
2010년: 서울시립대학교 컴퓨터과학부 (석사)
2010년 ~ 현재: 서울시립대학교 컴퓨터과학과 박사
사과정
※ 주관심 분야: 화자 인식, 음성 인식

•유 하 진 (Ha-Jin Yu)



1990년: KAIST 전산학 (공학사)
1992년: KAIST 인공지능 (공학석사)
1997년: KAIST 인공지능 (공학박사)
1997년 ~ 2000년: LG전자 전자기술원 선임연구원
2000년 ~ 2002년: SL2(주) 연구소장
2002년 ~ 현재: 서울시립대학교 컴퓨터과학부 교수
※ 주관심 분야: 화자 인식, 음성 인식