

# Kernel Principal Component Analysis의 앙상블을 이용한 화자 식별

양 일 호\*, 김 민 석\*\*, 소 병 민\*, 김 명 재\*, 유 하 진\*  
\*서울시립대학교 컴퓨터과학부  
\*\*LG전자 전자기술원

## Speaker Identification using Ensemble of Kernel Principal Component Analysis

IL-Ho Yang\*, Min-Seok Kim\*\*, Byung-Min So\*, Myung-Jae Kim\*, Ha-Jin Yu\*  
\*School of Computer Science, University of Seoul  
\*\*Advances Research Institute, LG Electronics  
heisco@hanmail.net, ms@uos.ac.kr, sbm1210@naver.com, arthmody@naver.com,  
hjyu@uos.ac.kr

### Abstract

In this paper, we propose a new approach of speaker identification technique using KPCA(kernel principal component analysis). This approach uses ensemble of multiple classifiers(speaker identifiers). KPCA enhances features for each classifier. To reduce processing time and memory requirements, we select limited number of samples randomly which used as estimation set of each KPCA basis. The experimental result shows proposed approach better than GKPCA(greedy KPCA).

### 1. 서론

음성 및 화자 인식 시스템의 인식률은 채널 · 잡음 등의 환경 조건에 따라 악영향을 받을 수 있다. 이를 개선하기 위해 MFCCs 등의 화자 특징을 적합한 기저로 사상하여 강화하는 방법들이 쓰인다.

PCA(principal component analysis)는 널리 사용되는 방법이지만 특징이 비선형으로 분포되어 있을 때에는 적합한 변환 기저를 찾지 못할 수 있다. 비선형 특징을 다룰 수 있는 KPCA(kernel PCA)는 변환 기저를 추정하는데 쓰이는 특징의 수에 비례하여 계산량 및 메모리 요구량이 크게 증가하므로 짧은 발성에서도 많은 수의 특징을 추출하는 음성 및 화자 인식 분야에 그대로 적용하기는 어렵다. GKPCA(greedy KPCA)는 KPCA의 계산량 · 메모리 요구량 문제를 해결하기 위

해 전체 특징을 잘 표현하는 적은 수의 대표 샘플을 greedy filtering으로 취하여 변환 기저를 추정한다[1].

본 연구에서는 GKPCA와 유사하게 전체 특징으로부터 한정된 수의 대표 샘플을 취하여 KPCA 변환 기저를 추정한다. 이 때, 대표 샘플을 랜덤하게 선택하되 이러한 과정을 여러 번 거쳐 복수의 KPCA 변환 기저 및 이것으로 강화한 특징으로 학습한 서로 다른 분류기들을 얻는다. 그리고 이 분류기들로 앙상블을 구성하여 각 분류기의 인식 결과를 종합한다.

### 2. 제안한 방법

GKPCA는  $t$ 개의 전체 특징 샘플로부터  $n$ 개의 대표 샘플을 선택하여 KPCA 기저를 추정한다( $n \ll t$ ). 하지만 한정된 샘플만을 이용하므로 최적의 기저를 찾지 못할 수 있다(그림 1).

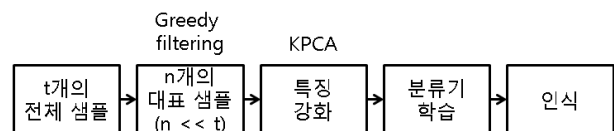


그림 1. GKPCA 수행 과정

본 연구에서는 greedy filtering으로  $n$ 개의 대표 샘플을 취하는 방법 대신, 랜덤하게  $n$ 개의 대표 샘플을  $m$ 회 취하여 각각에 대해 KPCA를 수행한 뒤 서로 다른  $m$ 개의 분류기를 학습하여 앙상블 결합하는 방법을 제안한다(그림 2).

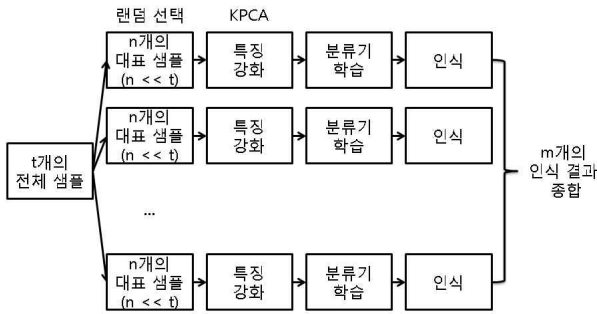


그림 2. m개 KPCA-분류기의 앙상블 결합 수행 과정

이는 GKPCA보다 m배 많은 대표 샘플을 이용하므로 식별률 향상 효과가 더 높을 것으로 기대하였다. 이 때, 랜덤 선택한 대표 샘플 집합들을 GA(genetic algorithm)[3]를 사용하여 개선하였다.

### 3. 실험 설계 및 결과

#### 3.1 실험 설계

ETRI 화자인식용 휴대전화 DB의 문장 발성을 이용하여 GMM-UBM[2] 화자 식별 실험을 수행하였다. 월차 화자 101명의 문장 발성 10개(1월차 1번째 발성)로 UBM(universal background model)을 구성(총 1010개 발성)하고 주차 화자 104명에 대해 화자 모델 학습(1주차 1번째 발성, 총 1040개) 및 테스트(1주차 3번째 발성, 총 1040개)를 수행하였다. 화자 모델을 학습하기 위해 가우시안 혼합 수 32개에서 10회 반복 학습한 UBM에 대해 1회 MAP 적용하였다( $\tau=1$ ).

#### 3.2 특징 추출

20차 MFCCs(mel-frequency cepstral coefficients)에 로그 에너지를 더하여 총 21차 특징을 추출하였다(window size = 25ms, shift = 10ms). 에너지 기반으로 사일런스를 제거한 뒤, 발성별로 CMS(cepstral mean subtraction)를 수행하였다.

#### 3.3 특징 강화

PCA · GKPCA · 제안한 방법을 이용하여 특징을 강화하였다. 이 때 UBM 학습 데이터로부터 변환 기저를 추정하여 모든 특징(UBM 학습 · 화자 모델 학습 및 테스트 데이터)을 강화하였다.

KPCA 수행에는 가우시안 커널 함수를 이용하였다( $\sigma=21$ ). greedy filtering 및 랜덤 선택시 대표 샘플 수는 100개로 고정하였다.

#### 3.4 랜덤 선택한 대표 샘플 집합의 개선

제안한 방법에서는 대표 샘플을 25회 랜덤 선택하고 이것을 1세대 개체 집합으로 하는 GA 연산으로 10회 개선하였다. 각 개체의 적합도는 화자 모델 학습 데이

터에 대한 식별률로 계산하되, 이 때의 UBM은 혼합 수 4개에서 5회 반복 학습하였다. 이전 세대에서 가장 높은 적합도를 지닌 1개를 유지(elitism 적용)하고 나머지는 룰렛 선택을 통해 교차(cross over) 생성한 뒤 각 샘플을 5% 확률로 랜덤 변경하는 돌연변이(mutation)를 적용하였다.

#### 3.5 앙상블 결합 방법

각 분류기에서 가장 높은 유사도(likelihood)를 갖는 부류(class)에 1표를 더하는 방식으로 총 25개의 분류기에 대해 다수 투표를 적용하였다.

표 1. 실험 결과

특징	차원	화자 식별률 (%)
MFCCs	21	79.13
PCA		83.46
GKPCA		84.04
제안한 방법		<b>84.90</b>

실험 결과 제안한 방법이 PCA 및 GKPCA보다 높은 식별률을 보이는 것을 확인하였다.

### 4. 결론

본 논문에서는 계산량 및 메모리 요구량이 많은 KPCA를 화자 인식 분야에 적용하기 위해, 한정된 수의 대표 샘플을 랜덤 선택해 복수 개의 분류기를 생성하여 앙상블 결합하는 방법을 제안하였다. 실험 결과 제안한 방법은 PCA 및 GKPCA에 비해 높은 화자 식별률을 보였다.

### Acknowledgement

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0024047)

### 참고문헌

[1] 김민석, 양일호, 유하진, "Greedy Kernel PCA를 이용한 화자식별," 말소리, 제66호, pp.105-116, 2008.  
 [2] Douglas A. Reynolds, Thomas F. Quatieri and Robert B. Dunn. "Speaker Verification Using Adapted Gaussian Mixture Models," Digital Signal Processing., Vol.10, pp.19-41, Jan. 2000.  
 [3] Melanie Mitchell, "An Introduction to Genetic Algorithms," The MIT Press, 1998.