ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 20-04-15
102

j.inst.Korean.electr.electron.eng.Vol.24,No.4,1022~1027,December 2020
http://dx.doi.org/10.7471/ikeee.2020.24.4.1022

GMM-supervector를 사용한 SVM 기반 화자분류에 대한 연구

A Study on SVM-Based Speaker Classification Using GMM-supervector

이 경 록*★

Kyong-Rok Lee*★

Abstract

In this paper, SVM-based speaker classification is experimented with GMM-supervector. To create a speaker cluster, conventional speaker change detection is performed with the KL distance using the SNR-based weighting function. SVM-based speaker classification consists of two steps. In the first step, SVM-based classification between UBM and speaker models is performed, speaker information is indexed in each cluster, and then grouped by speaker. In the second step, the SVM-based classification between UBM and speaker models is performed by inputting the speaker cluster group. Linear and RBF are applied as kernel functions for SVM-based classification. As a result, in the first step, the case of applying the linear kernel showed better performance than RBF with 148 speaker clusters, MDR 0, FAR 47.3, and ER 50.7. The second step experiment result also showed the best performance with 109 speaker clusters, MDR 1.3, FAR 28.4, and ER 32.1 when the linear kernel was applied.

요 약

본 논문에서는 GMM-supervector를 특징 파라미터로 하는 SVM 기반 화자 분류에 대해서 실험하였다. 실험을 위한 화자 클러스터를 생성하기 위해서 기존의 SNR 기반 가중치를 반영한 KL거리 기반 화자변화검출을 실행하였다. SVM 기반 화자 분류는 2단계로 이루어져있다. 1단계는 UBM과 화자 모델들간의 SVM 기반 분류를 시행하여 각 클러스터에 화자 정보를 인덱싱한 다음 화자별로 그룹핑한다. 2단계는 화자 클러스터 그룹에 UBM과 화자모델들간의 SVM 기반 분류를 시행한다. SVM의 커널 함수로는 Linear와 RBF를 사용하였다. 실험결과, 1단계에서는 Linear 커널이 화자 클러스터 148개, MDR 0, FAR 47.3, ER 50.7로 좋은 성능으로 보였다. 2단계 실험결과도 Linear 커널이 화자 클러스터 109개, MDR 1.3, FAR 28.4, ER 32.1로 좋은 성능을 보였다.

Key words : Speaker Classification, SVM, GMM-supervectgor, Linear, RBF

* Professor, Dept. of IT Engineering, Nambu University
★ Corresponding author
E-mail : krlee@nambu.ac.kr, Tel : +82-62-970-0123
Manuscript received Nov. 16, 2020; revised Dec. 25, 2020; accepted Dec. 29, 2020.

Ⅰ. 서론

멀티미디어 검색은 정보검색 시스템의 발전에 힘입어 나날이 영향력이 증대되고 있다. 그러나, 과거의 데이터의 경우 관련 정보의 부족으로 인하여 검색의 사각지대에 놓여있는 것이 현실이다. 이러한 과거 데이터 중 가장 활용도가 높은 것으로 예측되는 뉴스를 대상으로 음성 인식 기반 뉴스 검색을 위한 전처리인 화자 단위 분할 알고리즘은 크게 세

가지로 분류된다. 첫 번째, 디코더 기반 분할은 입력 음성신호에 대하여 인식기 기반 분석을 실시하여 화자의 변화를 검출한다. 두 번째, 모델 기반 분할은 사전에 훈련된 화자모델 기반 화자인식을 사용하여 화자의 변화를 검출한다. 세 번째, 매트릭스 기반 분할은 시간에 따라 변화하는 두 개의 인접 분석윈도우를 비교하여 화자의 변화를 검출한다[1].

논문 [2]에서는 별도의 인식기가 없고 사전에 화자정보가 없는 조건에서 화자변화를 검출하기 위해서 매트릭스 기반 분할을 사용하여 1차 화자분할을 실시한 다음, 실제 뉴스 환경의 다양한 환경소음을 보상하기 위해서 SNR(Signal to Noise Ratio) 기반 가중 KL(Kullback Leibler) 거리를 활용하여 검출된 화자 변화 지점을 검증하였다.

본 논문에서는 음성 인식 기반 뉴스 검색 전처리 2단계로 논문 [2]의 화자변화 검출 시스템에서 검출된 화자변화지점 정보를 바탕으로 화자별 분류에 대해서 연구하였다. 논문 [2]에서는 MDR(Missed Detect Rate)를 0으로 고정한 반작용으로 높은 FAR (False Alarm Rate)을 보여주었다. 이에 본 논문에서는 높은 FAR를 해결하면서 화자에 대한 사전 지식이 없는 무감독 화자분류를 위해 최근 관련 분야에서 높은 성능을 보이고 있는 SVM(Support Vector Machine)을 적용하였다.

II. 기존의 화자변화 검출 실험

논문 [2]의 화자변화 검출 시스템을 Fig. 1과 같다. 먼저 묵음을 기준으로 음향을 분리하여 음향 클러스터를 만든다. 다음으로 각 음향 클러스터를 매트릭스 거리 기반 화자변화 검출 방법 중 양호한 성능을 보이는 BIC(Bayesian Information Criterion)을 적용하여 10초 길이의 분석 윈도우 X, Y를 0.5초씩 이동하면서 화자변화지점을 검출한다. 이 때 각 분석 윈도우는 GMM(Gaussian Mixture Model)으로 모델링된다. 검출된 화자변화 지점은 $\triangle BIC$를 사용하여 1차로 검증한다. 1차 검증된 화자변화 지점은 SNR 기반 가중 함수 w_m을 적용한 KL 거리 D_s를 사용해서 2차 검증하였다[2], [3].

실험을 위한 데이터베이스는 논문 [2]에서 화자변화 검출 실험을 위하여 구성된 데이터베이스를 확장하여, 실제 뉴스 데이터 4회분을 대상으로 잡음환경, 화자정보 등을 고려한 선별기준을 적용하여 구축하였다. 실험을 위한 데이터베이스 구성은 Table 1과 같으며, 총 화자의 수는 19명(남자 12명, 여자 7명)이다. 화자변화 검출 실험에서는 논문 [2]에서 제안한 SNR 기반 가중치 함수 w_m을 반영한 KL 거리 D_s를 적용하였다. D_s는 수식 (1)과 같다. Σ_{ux_s}, Σ_{uy_s}는 분석 윈도우 X, Y에서 w_m을 적용한 GMM-UBM(Universal Background Model)의 공분산이다.

$$D_s = \frac{1}{2} tr\left[\left(\Sigma_{ux_s} - \Sigma_{uy_s}\right)\left(\Sigma_{uy_s}^{-1} - \Sigma_{ux_s}^{-1}\right)\right] \quad (1)$$

실험결과는 수식 (2)과 같이 MDR, FAR를 기준으로 분석하였다.

$$MDR = \frac{N_{scd} - N_{cd}}{N_{scd}} \times 100 \quad (2)$$

$$FAR = \frac{N_{all} - N_{cd}}{N_{all}} \times 100$$

N_{scd}는 실제 화자 변화 지점의 수, N_{cd}는 실제 화자 변화 지점 중 검출된 수, N_{all}는 검출된 전체 화자 변화 지점의 수이다.

논문 [2]에서 가장 높은 성능을 보인 threshold가 0.024이고, β가 0.42일 때 MDR 0, FAR 66.5의 성능을 보였다. 여기서 검출된 화자변화 지점은 본

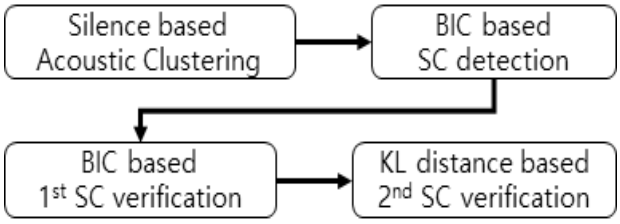

Fig. 1. A brief block diagram of the conventional speaker change detection system.
그림 1. 기존 화자변화검출 시스템의 블록 다이어그램

Table 1. Description of database.
표 1. 데이터베이스 세부사항

Aspects of Noise	Number of speaker change	Pattern of speaker change
Low SNR	34	Male ↔ Male
	26	Male ↔ Female
High SNR	10	Male ↔ Male
	8	Male ↔ Female

논문에서 화자분류를 위한 화자 클러스터 생성의 기준이 된다.

III. SVM 기반 화자 분류 시스템

본 논문에서는 기존의 화자변화 검출 시스템의 결과를 바탕으로 화자 클러스터를 생성하고 각 화자 클러스터를 대상으로 화자분류를 실시하였다. 뉴스 데이터에 대한 화자정보가 없다는 전제 하에서 실험하였기 때문에 해당 클러스터에 몇 명의 화자가 있는지 어떤 화자가 있는지 모르기 때문에 무감독으로 화자분류를 실시하였다. 화자분류를 위한 알고리즘은 강력한 분류 방법인 SVM을 이용하였다.

1. GMM-supervector

기존 GMM 기반 화자 인식 연구에서 화자 및 채널 변동성을 보상하기 위해 잠재요소 분석을 사용한다. 이 변동은 잠재요인을 사용한 MAP(Maximum A Posteriori) 적용된 GMM의 평균을 모델링하는 방법으로 나타낼 수 있다. 이러한 접근의 핵심은 GMM의 mixture 성분의 누적된 평균으로 구성되어 있는 GMM-supervector를 사용하는 것이다. 즉, GMM의 평균벡터들을 하나의 벡터로 누적한 벡터를 supervector라 한다. 각 화자별 클러스터들은 GMM-UBM에 대한 MAP 적응을 통해 얻어진다. 본 논문에서 SVM의 특징 파라미터로 GMM-supervector를 사용하였다. 다음은 GMM-supervector에 대한 개념을 나타낸 것이다[4], [5].

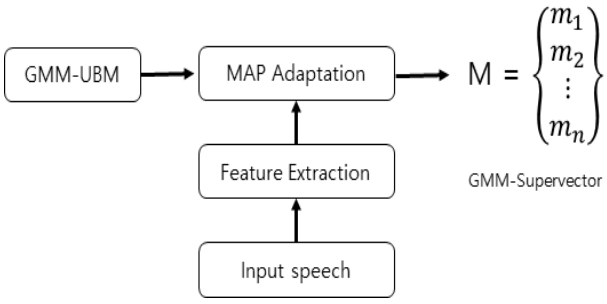

Fig. 2. GMM-supervector concept.
그림 2. GMM-supervector 개념

2. SVM

SVM은 두 클래스 사이의 최대 여백으로 분리되는 하이퍼 평면을 식별하는 이진 분류기이다. 데이터가 2차원 분포를 보인다면 SVM의 결정경계는 선이고, 데이터가 3차원 분포를 보인다면 SVM의 결정경계는 평면이다. SVM의 목적은 최대 마진을 가지는 결정경계를 찾는 것이다. 여기서 마진은 가장 가까운 훈련 데이터에서 결정 경계까지의 거리로 정의된다. 기본 SVM 분류기는 주어진 두 클래스로 데이터를 분류하는데 사용된다. SVM은 1 대 나머지 또는 1 대 1 분류를 적용해서 복수개의 클래스로 쉽게 확장을 할 수 있다. 기본 SVM 분류기는 선형으로 분리 가능한 데이터를 분리하는데 사용할 수 있는 선형 분류기로 사용할 수 있다[6], [7].

SVM은 입력데이터들의 공간을 고차원 공간으로 변환하여 비선형 결정 경계를 검출할 수 있다. 입력 데이터의 원래 공간과 고차원 공간 사이의 관계는 본질적으로 비선형이기 때문에 SVM의 목표는 비선형 결정 경계를 결정하는 것이다. 이렇듯, 데이터를 고차원 공간 즉, 비선형 공간으로 변환하는데 사용되는 것이 커널함수이다. SVM의 목적은 최적의 분리 초평면(hyper plane)을 결정하는 것이다. 최적의 초평면은 마진에서 최대 거리를 갖는 평면이다. 이 최적의 초평면은 훈련되지 않은 패턴에서 다른 방법보다 양호한 성능을 보인다. 훈련 데이터 $\{x_i, y_i\}_{i=1}^{n}$가 주어지면 SVM의 목표는 최대 마진으로 데이터를 분리할 수 있는 초평면을 결정하는 것이다. 이 때, $\{x_i\}_{i=1}^{n}$는 데이터이고, $y_i \in \{-1, 1\}$는 x_i의 레이블이다. 이 초평면은 두 클래스 간에 최대거리를 갖는 방정식 $W^T X+b$ 로 정의할 수 있다. 양의 클래스(class1)에 속하는 $W^T X+b \geq 1$인 데이터는 양의 클래스 샘플로 분류되고, 음의 클래스(class2)에 속하는 $W^T X+b \leq -1$인 데이터는 음의 클래스 샘플로 분류된다. $W^T X+b = 0$인 데이터는 마진 안의 값이다. 최적화 함수는 두 개의 초평면 사이의 거리가 최대가 되도록 설계되었으며, 이는 $\frac{1}{\| w \|}$를 최대화하는 것이며, $\| w \|$를 최소화하는 것이다[6].

SVM의 목적함수는 다음과 같다.

$$L_p = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{n} \alpha_i \{ y_i (w_i^t x_i + w_0) - 1 \} \quad (3)$$

위 함수의 쌍대를 구하면 최적화 문제에 대한 해를 쉽게 계산할 수 있다. 그러나, 데이터가 선형적으로 분리 불가능하거나 겹치는 경우 복잡성이 발

생한다. 대부분의 데이터는 간단하지 않으며 사용 가능한 예제(학습 사례)를 기반으로 새로운 예제(테스트 사례)를 정확하기 분류하기 위해서는 더 복잡한 경계가 필요한 경우가 많다. 이에 제안된 소프트 마진 SVM은 비선형적으로 분리 가능한 데이터와 겹치는 데이터에 대해 매우 강력하고 효율적이다[6], [8]-[10].

본 논문의 SVM 모듈은 Chih-Chung Chang and Chih-Jen Lin의 LIBSVM(A Library for Support Vector Machines) ver 3.0(Matlab 버전)을 사용하였다.

3. Kernel 함수

SVM은 서로 다른 클래스의 데이터가 분리 평면의 양쪽에 위치하도록 분리하는 초평면을 얻는 것이 목적이다. 데이터는 분류 가능 여부에 따라 선형으로 분리 가능, 비선형으로 분리 가능, 중첩으로 분류할 수 있다. 선형 분리 가능 데이터는 두 클래스의 데이터를 선형 경계를 사용하여 분리할 수 있다. 비선형 분리 가능 데이터는 두 클래스의 데이터를 비선형 경계를 사용하여 분리할 수 있다. 중첩 데이터는 선형 경계나 비선형 경계로 분리할 수 없다. 데이터가 입력 공간에서 선형으로 분리 될 수 없는 경우 소프트 마진 SVM은 잘못 분류 된 데이터 수를 최소화하고 일반화되는 강력한 분리 초평면을 찾을 수 없다. 이를 위해 커널을 사용하여 데이터를 선형적으로 분리 할 수 있는 커널 공간이라고 하는 더 높은 차원의 공간으로 변환할 수 있다. 따라서 커널 공간에서 선형 초평면은 입력 데이터 공간의 데이터를 분리하는 작업에서 고차 분리 초평면을 구하는 것 대신 서로 다른 클래스들을 분리하는데 사용될 수 있다. 커널 선택은 데이터 특성에 크게 좌우된다. 선형 커널은 클래스들 간의 경계가 선형일 때 사용된다. 다항식 커널은 이미지 처리에 널리 사용된다. Gaussian 및 RBF (Radial Basis Function)는 대부분 사전 지식이 없을 때 적용되는 범용 커널이다. Table 2는 자주 사용되는 커널에 대해서 정리한 것이다. 다항식 커널과 RBF 커널, Sigmoidal 커널은 비선형 커널로서 클래스들 간의 경계가 비선형이거나 중첩일 때 사용된다. 본 논문에서는 대중적으로 널리 사용되는 Linear와 사전 지식이 없을 때 적용되는 RBF를 적용하였다[6], [11], [12].

Table 2. types of widely used kernels.
표 2. 자주 사용되는 kernel들

Kernel Type	$K(X_i, X_j)$
Linear	$(X_i^T . X_j)$
Polynomial	$(X_i^T . X_j + 1)^h$
RBF	$e^{-\|X_i - X_j\|^2/2\sigma^2}$
sigmoidal	$\tanh(kX_i . X_j - \delta)$

4. SVM 기반 화자분류

본 논문에서는 화자에 대한 사전 지식이 없는 무감독 화자분류에 SVM을 적용하고자 한다. 일반적으로 SVM은 분류하고자 하는 대상을 사전에 훈련해서 모델링하고 신규 데이터가 어떤 모델에 적합한지를 분류하는데 사용된다. 이를 무감독 화자분류에 적용하기 위해서 UBM을 제외한 화자모델이 전혀 없는 상황에서 시작하는 화자분류 2단계 시스템을 구상하였다.

SVM 기반 화자분류 1단계는 1번째 화자 클러스터를 훈련하여 최초 화자모델을 구축한다. 다음 화자 클러스터들은 UBM과 구축된 화자모델 그룹들 간에 SVM 기반 화자 분류를 한다. UBM으로 분류되면 새로운 화자모델로 훈련하여 등록한다. 기존의 화자 모델로 분류되면 해당 화자로 인덱싱한다. 이러한 과정을 전체 화자 클러스터에 대해서 실시한다. 분류 결과를 바탕으로 인접한 화자 클러스터가 동일 화자로 인덱싱되는 경우에는 하나의 클러스터로 통합하였다.

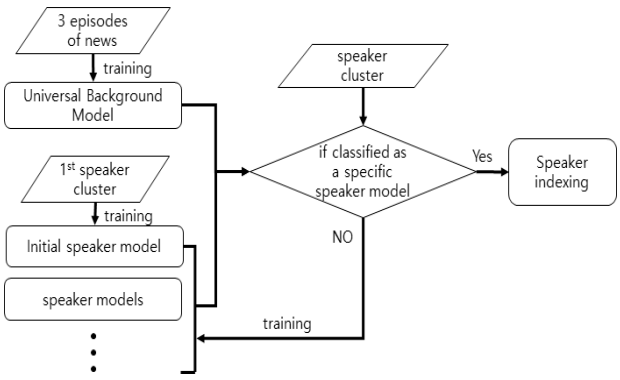

Fig. 3. The first step in SVM based speaker classification.
그림 3. SVM 기반 화자분류 1단계

SVM 기반 화자본류 2단계는 1단계의 결과를 바탕으로 화자 클러스터를 각 화자 인덱싱에 따라 그

룹핑한 다음 첫 번째 화자 클러스터 그룹을 훈련하여 첫 번째 화자 모델로 구축한다. 다음 화자 클러스터 그룹에 대하여 1단계와 동일하게 UBM과 구축된 화자모델 그룹간에 SVM 기반 화자분류를 실시한다. UBM으로 분류되면 새로운 화자모델로 훈련하여 등록하고 기존의 화자모델로 분류되면 해당 화자로 인덱싱한다.

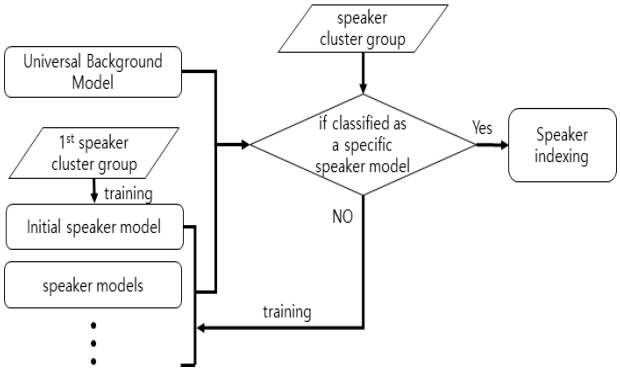

Fig. 4. The second step in SVM based speaker classification (verification).
그림 4. SVM 기반 화자분류 2단계

Ⅳ. 실험결과

2절의 화자변화 검출 실험 결과물인 233개의 화자 클러스터를 대상으로 Linear와 RBF 커널함수를 적용하여 화자분류 실험을 실시하였다.

SVM 기반 화자분류 실험을 위하여 ER(Error Rate)을 제안한다. N_{sc}는 총 화자 클러스터 수이고, N_{fc}는 다른 화자로 잘 못 분류된 화자 클러스터의 수이다.

$$ER = \frac{N_{fc}}{N_{sc}} \times 100 \quad (4)$$

SVM기반 화자분류 1단계 실험 결과 각 화자 클러스터들이 화자별로 그룹핑되면서 화자 클러스터 수와 FAR이 화자변화 검출 실험에 비해 감소하였다. 특히, Linear 커널을 적용한 경우가 더 좋은 결과를 나타내었으며, RBF 커널을 적용한 경우에서는 실제 화자변화 지점이 동일 화자가 오판정되면서 MDR이 발생하였다.

Table 3. Results of the first step experiment of Svm-based speaker classification.
표 3. SVM 기반 화자 분류 실험 1단계의 결과

Kernel function	number of Speaker Cluster	MDR	FAR	ER
Linear	148	0.0	47.3	50.7
RBF	162	1.3	51.9	56.2

SVM기반 화자분류 2단계 실험은 1단계 실험결과를 입력으로 하여 실험하였다. 실험결과 화자별 모델의 훈련 데이터가 보완되면서 보다 견인해지고 이에 따라 MDR을 제외한 나머지 수치가 1단계에 비해서 감소하였다. 다만, 오판정이 Linear 커널과 RBF 커널에서 공히 발생하여 MDR이 증가하였다.

Table 4. Results of the second step experiment of Svm-based speaker classification.
표 4. SVM 기반 화자분류 실험 2단계의 결과

Kernel function	number of Speaker Cluster	MDR	FAR	ER
Linear	109	1.3	28.4	32.1
RBF	127	3.8	38.6	42.5

Ⅴ. 결론

본 논문에서는 GMM-supervector를 특징 파라미터로 하는 SVM 기반 화자 분류에 대해서 실험하였다. 실험을 위한 화자 클러스터를 생성하기 위해서 기존의 SNR 기반 가중치를 반영한 KL거리를 적용한 화자변화 검출을 실시하였다. 화자변화 검출은 threshold 0.024일 때 233개의 화자 클러스터가 생성되었고, MDR 0, FAR 66.5였다.

무감독 SVM 기반 화자분류는 2단계로 이루어져 있다. 1단계는 UBM과 화자모델들 간의 SVM 기반 분류를 시행하여 각 클러스터에 화자 정보를 인덱싱한다. 2단계는 1단계의 화자 클러스터 그룹을 입력으로 하여 1단계와 같이 UBM과 화자모델들 간의 SVM 기반 분류를 시행한다. SVM 기반 분류를 위한 커널 함수로는 Linear와 RBF를 사용하였다. 실험결과 1단계에서는 Linear 커널을 적용한 경우가 화자 클러스터 수 148개, MDR 0, FAR 47.3, ER 50.7로 RBF에 비해 나은 성능으로 보였다. 2단계 실험결과도 Linear 커널을 적용한 경우가 화자 클러스터 수 109개, MDR 1.3, FAR 28.4, ER 32.1로 가장 좋은 성능을 보였다. 이는 화자변화 검출 실험에 비해서 화자 클러스터 수는 53.2 포인트 감소하였고, FAR은 38.1 포인트 감소한 수치

이며, 1단계 실험에 비해서 MDR은 1.3 포인트 증가, FAR은 18.9 포인트 감소, ER은 18.6포인트 감소한 결과이다.

References

[1] Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S. Parthasarathy, and Qian Huang, "Speaker detection in broadcast news data," *Proc. ICSLP '98*, vol.4, pp.1339-1343, 1998.

[2] Joon-Beom Cho, Ji-eun Lee, Kyong-Rok Lee, "The Study on Speaker Change Verification Using SNR based weighted KL distancem," *Journal of Convergence for Information Technology*, vol.7, no.6, pp.159-166, 2017.
DOI: 10.22156/CS4SMB.2017.7.6.159

[3] Joon-Beom Cho, Ji-eun Lee, Kyong-Rok Lee, "The Study on the verification of Speaker Change using GMM-UBM based KL distance," *Journal of Convergence Society for SMB*, vol.6, no.1, pp. 71-77, 2016. DOI: 10.22156/CS4SMB.2016.6.4.071

[4] 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.
DOI: 10.1109/LSP.2006.870086

[5] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, P. A. Torres-Carrasquillo, "Support vector machines for speaker and language recognitionm," *Computer Speech & Language*, vol.20, no.2, pp. 210-229, 2006.

[6] Deepika Kancherla, Jyostna Devi Bodapati, Veeranjaneyulu N, "Effect of Different Kernels on the Performance of an SVM Based Classification," *IJRTE*, vol.5S4, no.7, pp.2277-3878, 2019.

[7] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, vol.20, No.3, pp. 273-297, 1995. DOI: 10.1023/A%3A1022627411411

[8] D. DeCoste, K. Wagstaff, "Alpha seeding for support vector machines," *Proceedings of the 6th ACM SIGKDD international conf. on Knowledge discovery and data mining*. ACM, pp.345-349, 2000. DOI: 10.1145/347090.347165

[9] Ngoc Nam BUI, Jin Young KIM, Tan Dat TRINH, "A Non-linear GMM KL and GUMI Kernel for SVM Using GMM-UBM Supervector in Home Acoustic Event Classification," *IEICE TRANS. FUNDAMENTALS,* vol.E97.A, no.8, pp. 1791-1794, 2014. DOI: 10.1587/transfun.E97.A.1791

[10] *S. Theodoridis, K. Koutroumbas, "Support Vector Machines: The Non-linear Case, Pattern Recognition,"* Academic Press, pp.198-200, 2008.

[11] *Mariette Awad, Rahul Khanna. "Efficient Learning Machines,"* Apress, pp.39-66, 2015.

[12] Mehmet G¨onen, Ethem Alpaydın, "Multiple Kernel Learning Algorithms," *Journal of Machine Learning Research*, vol.12 pp.2211-2268, 2011.

BIOGRAPHY

Kyong-rok Lee (Member)

1997 : BS degree in Electronics Engineering, Honam University.
2001 : MS degree in Electronics Engineering, Chonnam National University.
2006 : PhD degree in Electronics Engineering, Chonnam National University.
2008~ : Associate Professor, Nambu University