논문 2011-48SP-5-13

음성/음악 분류 향상을 위한 2차 조건 사후 최대 확률기법 기반 SVM

(Improving SVM with Second-Order Conditional MAP for Speech/Music Classification)

임 정 수*, 장 준 혁**

(Chungsoo Lim and Joon-Hyuk Chang)

요 약

Support vector machine (SVM)은 패턴인식 분야에 많이 사용되어지고 있고 그 한 예로서 3GPP2 selectable mode vocoder (SMV)와 같은 규격화된 코덱에 쓰여 코덱의 음성/음악 분류 성능을 향상시킬 수 있다. 본 논문에서는 SVM을 개선시켜 음성/음악의 분류성능을 더욱 향상시키는 새로운 방법을 제안한다. 음성/음악신호의 각 프레임들은 서로 강한 상관관계를 가지고 있는데, 이를 바탕으로 2차 조건 사후 최대 확률기법을 SVM에 적용하여 음성/음악 분류성능을 향상시킨다. 또한 SVM을 학습시킬 때 적용되는 기존의 기법들과는 달리 제안되는 기법은 SVM이 패턴분류를 행할 때 사용된다. 그렇기 때문에 기존의 기법들과 독립적으로 개발되고 사용될 수 있고, 따라서 패턴분류의 성능을 한층 더 향상시킬 수 있다. 실험을 통해 제안된 기법의 독립성과 성능향상을 기존의 기법들과 비교하여 증명하였다.

Abstract

Support vector machines are well known for their outstanding performance in pattern recognition fields. One example of their applications is music/speech classification for a standardized codec such as 3GPP2 selectable mode vocoder. In this paper, we propose a novel scheme that improves the speech/music classification of support vector machines based on the second-order conditional maximum a priori. While conventional support vector machine optimization techniques apply during training phase, the proposed technique can be adopted in classification phase. In this regard, the proposed approach can be developed and employed in parallel with conventional optimizations, resulting in synergistic boost in classification performance. According to experimental results, the proposed algorithm shows its compatibility and potential for improving the performance of support vector machines.

Keywords : Second-order Conditional Maximum a posteriori (Second-order CMAP), Support Vector Machine (SVM), Selectable Mode Vocoder (SMV), Speech/Music Classification Algorithm

* 정회원, 목포대학교
(Mokpo National University)
** 정회원, 한양대학교 융합전자공학부
(Dep. of Electronic Engineering, Hanyang University)
※ 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT핵심기술개발사업의 일환으로 수행하였음. [KI001824, 장애인 및 고령자를 위한 Digital Guardian 기술개발]
※ 또한 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2009-0085162)
접수일자: 2011년2월25일, 수정완료일: 2011년6월14일

Ⅰ. 서 론

최근 IT기술의 발전으로 이동통신 단말기나 차량 네비게이션 등의 무선통신기기를 이용한 멀티미디어 서비스가 늘어나고 이를 이용하는 고객의 수도 빠르게 늘고 있다. 다양한 멀티미디어 서비스는 일반적으로 많은 데이터 전송을 요구하는데 이를 제한된 주파수 대역폭으로 감당하기에는 어려움이 있다. 이를 극복하기 위한 방편으로 제한된 주파수 대역의 효율적인 활용이 활발

히 연구되어지고 있고 가변적인 전송률을 가지는 다양한 음성 코덱이 개발 되었다[1~2]. 실제로 음성신호의 유형에 따라 다른 비트수를 할당하는 것은 음성의 음질에 영향을 미치기 때문에 정확한 신호분류 기술의 핵심기술로서 다루어지고 있다. 이런 음성코덱 중의 하나인 ETSI의 3GPP2 selectable mode vocoder (SMV)는 전송률을 4단계로 나누어 부호화 하는 방식을 채택하고 있다[1].

SMV의 음성/음악 분류를 개선하기 위하여 support vector machine (SVM)을 이용한 방법들이 최근에 제안되었다[3~5]. SVM은 패턴인식에 우수함을 인정받아 많이 이용 되고 있는 machine learning 기법의 하나로써 데이터 마이닝 분야는 물론, 얼굴인식, 생체인식, 문자인식, 그리고 음성인식 등 다양한 분야에 응용되고 있다[6~7]. SVM은 SMV의 음성/음악의 분류 성능을 많이 향상시킬 수 있는데, SVM을 이용한 방법 중 가장 최근에 발표된 연구에서는 음성/음악 프레임간의 상호연관성을 바탕으로, 과거 프레임의 SVM 판정결과를 고려하여 현재 프레임의 클래스를 추정하고 이에 따라 커널 파라미터를 조정하여 음성/음악 분류의 정확도를 높이는 방법을 소개하였다[5]. 이 방법은 성능의 향상뿐만 아니라 다른 기법과 병행하여 사용할 수 있다는 장점이 있다. 그러나 이 방법에서는 과거 프레임의 SVM 분류결과를 음성과 음악, 두 클래스의 예측에만 사용하고, 이에 따라 고정된 값이 커널 파라미터에 더해지는 다소 유연성이 부족한 단점이 있다. 또한 커널 파라미터의 조정으로 SVM의 판정을 조절하는 것은 사용된 데이터에 의존된 결과를 보일 수 있고, radial basis function (RBF)을 커널함수로 사용한 경우에 제한되어 있다는 단점도 있다.

이 기법의 장점은 유지하며 단점을 극복하기 위하여 본 논문에서는 2차 조건 MAP (maximum a posteriori)을 SVM에 적용하는 기법을 제안한다. 보통 SVM의 판별식에서는 하나의 고정된 문턱 값을 사용하게 되는데 제안하는 방법에서는 과거 프레임의 SVM 분류결과를 기초로 하여 4가지의 문턱 값 중 하나를 선택하게 된다. 두 개의 클래스로 예측하여 파라미터를 조정하던 것에 비하여 좀 더 세밀하게 과거의 정보를 사용할 수 있게 되었고 또한 문턱 값을 조정함으로써 SVM 분류를 보다 확실하게 그리고 커널함수의 종류에 관계없이 조정할 수 있게 되었다.

본 논문은 다음과 같이 구성된다. Ⅱ장에서는 SMV와 SVM에 대해서 간략히 설명하고 Ⅲ장에서는 2차 MAP을 적용한 SVM을 소개한다. Ⅳ장에서는 실험 설정과 실험결과를 보이고, Ⅴ장에서 본 논문을 끝맺는다.

Ⅱ. SMV와 SVM의 개요

SMV는 ETSI의 3GPP2 표준 코덱으로서 extended code excited linear prediction (ex-CELP) 기반의 압축방식을 사용하는데, 사람의 청각 특성에 최적화된 모델을 사용하여 음성을 저 전송률로 압축하는데 효율적이다[8~9]. 또한, 한정된 주파수 대역을 효율적으로 활용하기 위해 프레임 단위로 4가지의 가변 전송률을 제공하며 이동국과 기지국 사이의 통신망 채널에 따라 동적으로 변환되는 4가지 모드를 지원한다. 이러한 다양한 평균 전송률을 제공하기 때문에 시스템의 효율성과 음질간의 균형을 선택적으로 조절 할 수 있다.

SMV에서의 음악 분류 과정은 먼저 음성 검출기 (voice activity detection, VAD)에서 입력 신호가 음성과 묵음 또는 주변 잡음으로 나뉜 후 음성으로 판별된 경우에만 거치게 되며 음성/음악 분류에 사용되는 파라미터들은 다음과 같다.

1. 이동 평균 에너지 $\overline{E}$

$$\overline{E} = 0.75 \cdot \overline{E} + 0.25 \cdot E \quad (1)$$

E는 프레임 에너지 이다.

2. 잡음/묵음의 이동 평균 반사계수 $\overline{k_N(i)}$

$$\overline{k_N(i)} = 0.75 \cdot \overline{k_N(i)} + 0.25 \cdot k_1(i)$$

$$i = 1,...,10 \quad (2)$$

3. 부분적 잔류 에너지의 이동 평균 $\overline{E_N^{res}}$

$$\overline{E_N^{res}} = 0.9 \cdot \overline{E_N^{res}} + 0.1 \cdot E^{res} \quad (3)$$

$\overline{E_N^{res}}$는 $\overline{k_N}$에 따라서 값이 새로워진다.

4. 정규화 된 피치 상관도의 이동 평균 $\overline{corr_P}$

$$\overline{corr_P} = 0.8 \cdot \overline{corr_P} + 0.2 \cdot \left(\frac{1}{5} \cdot \sum_{i=1}^{5} corr_p^B(i)\right) \quad (4)$$

$corr^{B_P}(i)$는 이전 프레임의 피치 상관도이다.

5. 주기적 계수 $\overline{c_{pr}}$

$$\overline{c_{pr}} = \omega \cdot \overline{c_{pr}} + (1-\omega) \cdot c_{pr} \quad (5)$$

ω는 c_{pr}에 따라 값을 바꿔주는 정해진 가중치이다.

6. 음악 연속 계수의 이동 평균 $\overline{c_M}$

$$\overline{c_M} = 0.9 \cdot \overline{c_M} + 0.1 \cdot c_M \quad (6)$$

SMV의 VAD에서는 식 (1)~(5)로부터 나온 결과를 정해진 문턱 값과 비교하여 음성의 유무를 판단하며 음악의 분류는 $\overline{c_{pr}} \geq 18$ 또는 $\overline{c_M} > 200$이면 음악으로 판단한다.

SVM은 이진 패턴 분류에 뛰어난 성능을 보이는데 알려지지 않은 확률분포를 갖는 데이터에 대하여 잘못 분류할 확률을 최소화하는 구조적인 위험 최소화 (structural risk minimization)를 바탕으로 하고 있다. 선형적으로 분류가 가능한 경우, 두 개의 다른 클래스를 가르면서 가장 근접한 벡터들과의 거리가 최대화가 되는 초평면을 구한다[10]. 이런 초평면은 가중벡터 (weight vector)와 바이어스로 나타내어지는데 이것들은 2차계획법 (quadratic programming)을 풀어서 구해지게 되고 구해진 벡터와 바이어스를 가지고 SVM의 판별식을 나타내면 다음과 같다.

$$f(\boldsymbol{X}) = \sum_{i=1}^{l} \alpha_i^* y_i (\boldsymbol{X_i^*} \cdot \boldsymbol{X}) + b^* \quad (7)$$

$\boldsymbol{X}$는 입력벡터이고 $\boldsymbol{X_I^*}$는 학습에 의하여 구해진 support vector, α_i^*는 학습에 의해 구해진 라그랑제 승수 (Lagrange multiplier), 그리고 b^*는 바이어스 이다. 이 식은 입력벡터들이 선형으로 분류가 가능한 경우의 식이고 선형으로 분류가 불가능 한 경우 커널함수를 사용하는데 이 경우의 SVM의 판별함수는 다음과 같다.

$$f(\boldsymbol{X}) = \sum_{i=1}^{M} \alpha_i^* y_i K(\boldsymbol{X_i^*}, \boldsymbol{X}) + b^* \quad (8)$$

$K(\boldsymbol{X_i^*}, \boldsymbol{X})$는 커널함수로 RBF (radial basis function)이나 polynomial등이 널리 쓰이고 있다. 판별식의 결과는 정해진 문턱 값과 비교되어 클래스를 분류하게 되는데 주로 이 문턱 값으로는 0이 쓰인다. 즉 판별식의 값이 0보다 크면 음성으로 분류를 하고, 작으면 음악으로 분류를 한다.

Ⅲ. 2차 조건 MAP (maximum a posteriori)를 이용한 향상된 SVM

이번 장에서는 음성/음악 프레임 간의 상호관계를 바탕으로 2차조건 MAP을 이용하여 SVM의 분류성능을 향상시키는 기법을 소개한다. SVM을 이용한 음성/음악의 분류기법 중 이렇게 2차조건 MAP을 사용한 것은 없었으며, 2차조건 MAP을 이용해 음성향상을 제안한 논문은 근래에 발표되었다[11]. 이 논문과 본 논문의 차이점은 음성향상의 경우 음성의 존재/비존재를 구별하는 것이고 본 논문에서는 음성/음악을 구별한다는 것이다.

SVM 판별식의 값은 확률 값이 아니므로 확률로 매핑하기 위하여 다음과 같은 시그모이드 (sigmoid) 모델을 사용한다[12].

$$P(H(n) = H_1 | f_n) = \frac{1}{1 + \exp(Af_n + B)} \quad (9)$$

n은 프레임의 번호이고 H_0와 H_1은 각기 음악과 음성을 나타내는 기본가설이고 $H(n)$은 프레임 n에 대한 바른 가설이다. f_n은 n번째 프레임의 판별식 값이고 A와 B는 maximum likelihood estimation을 통해 학습되어지는 파라미터 이다. 위의 식은 현재 프레임이 음성일 확률이고 음악일 확률은 $1 - P(H(n) = H_1 | f_n)$로 표현될 수 있다. 이렇게 매핑된 확률 값을 가지고 음성과 음악을 분류하는 조건식은 다음과 같이 표현된다.

$$\frac{P(H(n) = H_1 | f)}{P(H(n) = H_0 | f)} \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \alpha \quad (10)$$

일반적으로 음성/음악 신호는 음성구간, 음악구간, 그리고 무음구간으로 구성되어 지는데, 각 구간은 어느 정도의 길이를 가지며 반복된다. 그러므로 각 구간은 대체로 많은 수의 프레임으로 구성된다. 그러므로 주어진 구간 안에서는 강력한 상관성으로부터 다음 프레임이 현재 프레임과 같은 종류일 확률이 아주 높다고 할 수 있다. 예를 들어 과거의 몇 입력 프레임이 음악 프레임 이라면 현재 프레임도 음악 프레임일 확률이 아주 높다는 것이다. 이런 강한 상관성으로부터 다음과 같은

식을 얻을 수 있다.

$$P(H(n)=H_0|H(n-1)=H_0, H(n-2)=H_0) > P(H(n)=H_0) \quad (11)$$

이러한 상호 연관성을 바탕으로 현재의 판별식 값 뿐 아니라 과거의 두 프레임의 판별결과를 고려한 새로운 판별방법을 다음 식과 같이 나타낼 수 있다.

$$\frac{P(H(n)=H_1|f_n, H(n-1)=H_i, H(n-2)=H_j)}{P(H(n)=H_0|f_n, H(n-1)=H_i, H(n-2)=H_j)} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \quad i=0, 1\ j=0, 1. \quad (12)$$

식 (12)를 Bayes′ rule을 이용하여 바꾸면 다음식이 된다.

$$\frac{P(f_n|H(n)=H_1, H(n-1)=H_i, H(n-2)=H_j)}{P(f_n|H(n)=H_0, H(n-1)=H_i, H(n-2)=H_j)} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha'_{ij} \quad i=0, 1 \quad j=0, 1 \quad (13)$$

여기서 α'_{ij}는 다음과 같다.

$$\alpha'_{ij} = \alpha \frac{P(H(n)=H_0|H(n-1)=H_i, H(n-2)=H_j)}{P(H(n)=H_1|H(n-1)=H_i, H(n-2)=H_j)} \quad (14)$$

현재 프레임의 SVM의 판별 값 은 현재 프레임의 클래스에 의해서 가장 많이 좌우되므로 식 (13)은 다음과 같이 간략화 될 수 있다[13].

$$\frac{P(f_n|H(n)=H_1)}{P(f_n|H(n)=H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha'_{ij} \quad i=0, 1 \quad j=0, 1. \quad (15)$$

다시 Bayes′ rule을 사용하여 변형시키면,

$$\frac{P(H(n)=H_1|f_n)}{P(H(n)=H_0|f_n)} \underset{H_0}{\overset{H_1}{\gtrless}} \beta_{ij}, \ i=0, 1 \quad j=0, 1 \quad (16)$$

이 된다. β_{ij}는 $\alpha'_{ij}P(H(n)=H_1)/P(H(n)=H_0)$ 이다. 각 문턱 값들은 과거 프레임들의 클래스에 따라 선택적으로 사용된다. 예를 들어 바로 전 프레임이 음성 (H_1) 프레임이고 그 전 프레임도 음성 (H_1)인 경우는 문턱값 β_{11}을 사용하고, 바로 전 프레임이 음악 (H_0) 프레임이고 그 전 프레임은 음성 (H_1)이라면 문턱값 β_{01}을 사용한다. 이렇게 과거 프레임들의 클래스 분류를 바탕으로 네 개의 문턱값을 사용함으로써 기존의 방법들보다 더 정밀하게 현재 프레임을 예측할 수 있게 되었다. 이 4개의 문턱 값 모두를 하나의 식에서 포함하는 판별조건은 다음과 같다.

$$f_n \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (17)$$

여기서 η는 4개의 문턱 값으로 다음과 같이 나타내질 수 있다.

$$\begin{aligned}\eta = {} & \beta_{00}P(H(n-2)=H_0|f_{n-2})P(H(n-1)=H_0|f_{n-1}) \\ & +\beta_{01}P(H(n-2)=H_1|f_{n-2})P(H(n-1)=H_0|f_{n-1}) \\ & +\beta_{10}P(H(n-2)=H_0|f_{n-2})P(H(n-1)=H_1|f_{n-1}) \\ & +\beta_{11}P(H(n-2)=H_1|f_{n-2})P(H(n-1)=H_1|f_{n-1}) \quad (18)\end{aligned}$$

이 식은 각각의 문턱값에 그에 해당하는 확률을 곱해준 형태로, 예를들어 β_{00}의 경우 이전의 두 프레임이 음악일 때 선택되는 것이므로 이전 두 프레임이 각각 음악일 확률을 두 확률의 곱으로 표현하여 적용한 것이다.

지금까지의 식들은 과거의 두 프레임의 SVM 분류결과를 바탕으로 하고 있다. 그러나 SVM의 분류는 완전하지가 않고 또한 잘못된 분류를 가지고 문턱 값을 조정함으로써 SVM의 또 다른 잘못된 분류를 유도할 수도 있다. 이를 막기 위하여 과거의 두 프레임만 보는 것이 아니라 두 개의 프레임 집합을 고려한다. 하나의 프레임 집합에는 여러 개의 프레임이 속해있고 각 집합은 같은 수의 프레임을 포함한다. 과거 두 프레임의 클래스를 고려하였듯이 두 집합의 클래스를 고려하야 하는데 각 집합의 클래스는 그 집합 안에 속한 프레임들의 판별값에 따라 결정된다. 즉, 한 집합 내 프레임들의 판별식 값의 평균이 음악 클래스에 속한다면 그 집합의 클래스는 음악이 된다. 예를 들어 과거의 두 프레임의 판별결과를 가지고 적합한 β_{ij}를 고를 때, 현재의 프레임이 n번째 프레임이라고 한다면, f_{n-1}과 f_{n-2}를 고려하게 된다. 그러나 과거의 두 프레임이 아니라 프레

임의 집합을 고려한다면 f_{n-1}과 f_{n-2} 대신에 각 집합의 평균값 (f'_{n-1}, f'_{n-2})을 가지고 적당한 문턱 값을 결정하게 된다.

$$f'_{n-1} = \frac{1}{N}\sum_{i=1}^{N} f_{n-i} \tag{19}$$

$$f'_{n-2} = \frac{1}{N}\sum_{i=1}^{N} f_{n-N-i} \tag{20}$$

N은 한 집합에 속하는 프레임의 개수로 본 논문에서는 20을 사용하였다.

Ⅳ. 실 험

1. 실험 설정

본 실험을 위해서 음성 데이터베이스로 8kHz로 샘플링 된 약 6 sec 정도의 깨끗한 음성으로 326명의 남자와 138명의 여자 화자에 의해서 화자마다 10개의 파일이 발음된 TIMIT 데이터베이스가 사용되었다[14]. 음악 데이터베이스는 CD로부터 다섯 가지 장르의 음악을 모바일 폰을 통해서 녹음하였고, 8kHz로 다운 샘플링 하여 사용하였으며, 각기 약 5분 정도의 길이를 가진다. 학습으로는 음성파일 4200개와 음악파일 50개 (블루스 10개, 클래식 10개, 힙합 10개, 재즈 10개, 메탈 10개)가 사용되었다.

객관적인 평가를 위해 10-fold 교차검증을 수행하였으며 각 테스트 파일은 5개의 음성부분 (6~12초), 하나의 음악장르로 구성된 5개의 음악부분(28~32초), 10개 무음부분 (3~15초)으로 되어있다. 트레이닝 파일의 음악부분은 모든 장르의 음악이 혼합되었다. 성능 평가를 위해 테스트 파일의 20ms마다 실제 결과를 음성, 음악, 무음으로 분류하여 저장하고 SVM의 분류 결과와 비교하였다.

실험에 사용된 특징벡터로는 Ⅱ장에서 소개된 6가지의 파라미터를 벡터로 구성해 사용하였고, 제안된 알고리즘의 문턱값은 β_{00}=0.808, β_{01}=0.708, β_{10}=0.630, β_{11}=0.193로 설정하였다. 제안된 알고리즘과 비교대상인 적응 커널파라미터 (adaptive kernel parameter) 기법[5]의 파라미터 조정량은 음성으로 추정되는 프레임에게는 -0.06을 그리고 음악으로 추정되는 프레임에게는 +0.06으로 정하였다.

2. 실험 결과

제안된 알고리즘을 검증하기 위해서 제안된 알고리즘과 기존의 알고리즘[3, 5]의 음성/음악 분류성능을 비교하였고 표 1에 그 결과를 나타내었다.

P_d는 각 음성과 음악이 정확하게 분류될 확률이고 P_e는 (1-P_d)로서 음성과 음악을 합친 error probability이다. SVM은 아무 개선을 시키지 않은 순수한 support vector machine이고 AKP는 과거 프레임의 분류를 바탕으로 현재 프레임의 클래스를 추정하고 그 추정에 근거하여 커널파라미터를 조정해 주는 기법이다. CMAP은 본 논문에서 제안한 기법을 나타낸다. AKP와 CMAP은 모두 SVM을 개선함으로써 분류성능을 향상시키는 기법으로 과거 분류결과를 바탕으로 현재 분류에 영향을 준다는 것은 동일하다. 그러나 근본적으로 두 가지의 차이점이 있는데, 첫째는 현재 분류에 영향을 주는 방법이다. AKP에서는 커널함수 파라미터를 조정하였지만 본 논문에서는 판별 문턱값을 조정한다. 둘째는 기법의 유연성이다. AKP는 커널파라미터에 더해주는 값이 고정된 두 개 이지만 제안된 기법에서는 네 가지 다른 값을 문턱값으로 사용하여. 보다 개선된 유

표 1. 제안된 기법과 기존 기법들과의 음성/음악 분류 성능 비교

Table 1. Comparison with a conventional support vector machine and a support vector machine enhanced by adaptive kernel parameter in terms of speech/music detection probability Pd and total error probability P_e.

Class	Algorithm	Speech P_d	Music P_d	Overall P_e
Blues	SVM[3]	0.85	0.87	0.13
	AKP[5]	0.94	0.89	0.10
	CMAP	0.93	0.92	0.08
Classic	SVM	0.74	0.66	0.33
	AKP	0.81	0.69	0.29
	CMAP	0.81	0.79	0.21
Hiphop	SVM	0.781	0.894	0.12
	AKP	0.85	0.94	0.08
	CMAP	0.85	0.98	0.04
Jazz	SVM	0.75	0.91	0.12
	AKP	0.85	0.94	0.08
	CMAP	0.84	0.96	0.07
Metal	SVM	0.76	0.85	0.17
	AKP	0.85	0.87	0.14
	CMAP	0.83	0.94	0.08
Avg	SVM	0.79	0.84	0.17
	AKP	0.87	0.87	0.14
	CMAP	0.86	0.92	0.10

연성을 가진다.

표에서 알 수 있듯이 제안된 기법은 기존의 SVM과 비교하였을 때 많은 성능향상을 보인다. 또한 제안된 기법과 동일하게 과거 프레임의 분류결과를 바탕으로 한 기법 (AKP)과 비교하여도 보다 나은 전체적인 성능을 보인다. 성능 뿐 아니라 AKP의 경우 알고리즘의 수정없이는 RBF를 커널함수로 사용하지 않은 SVM에는 사용할 수 없는 반면, 제안된 기법은 아무런 제약 없이 모든 SVM에 적용할 수 있다는 장점이 있다. 표에는 나오지 않았지만 변별적 가중치 학습을 이용해 SVM의 음성/음악 분류성능을 향상시킨 기법[4]과 비교해 본 결과, 성능면에서 제안된 기법이 더 우수하였고 또한 이 기법과 제안된 기법은 같이 병용될 수 있음을 알게 되었다.

V. 결 론

본 논문에서는 SVM의 음악/음성 분류성능을 향상시키기 위해 패턴 판별 시에 인접 프레임간의 강한 상호연관성을 바탕으로 2차조건 MAP을 이용하는 방법을 제안하였고 ETSI의 3GPP2 표준코덱인 SMV의 실시간 음성/음악 분류에 적용하여 보았다. 이 기법은 SVM의 성능을 향상시킬 뿐 아니라 다른 기법들과도 병용할 수 있다는 장점도 가지고 있다. 실험을 통하여 기존의 기법들과 비교한 결과, 기존의 기법보다 나은 SMV의 음성/음악 분류 성능을 보였다.

앞으로의 연구과제로는 라디오방송을 녹음하고 실험 데이터로 사용하여 제안된 기법의 실제적 응용 가능성을 가늠해 볼 계획이다.

감사의 글

본 연구는 지식경제부 및 한국산업기술평가관리원의 IT핵심기술개발사업의 일환으로 수행하였음. [KI001824, 장애인 및 고령자를 위한 Digital Guardian 기술개발]

또한 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2009-0085162)

참 고 문 헌

[1] 3GPP2 Spec., "Source-controlled variable-rate multimedia wideband speech codec (VMR-WB), service option 62 and 63 for spread spectrum systems," *3GPP2-C.S0052-A*, vol. 1.0, April. 2005.

[2] Y. Gao, E. Shlomot, A. Benyassine, J. Hyssen, Huan-yu Su, and C. Murgia, "The SMV algorithm selected by TIA and 3GPP2 for CDMA appications," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 709-712, May 2001.

[3] S. -K. Kim and J. -H. Chang, "Speech/music classification enhancement for 3GPP2 SMV codec based on support vector machine," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E92-A, no. 2, pp. 630-632, February 2009.

[4] S. -K. Kim and J. -H. Chang, "Discriminative weight training for support vector machine-based speech/music classification in 3GPP2 SMV codec," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E93-A, no. 1, pp. 316-319, January 2010.

[5] 임정수, 송지현, 장준혁, "SVM의 미세조정을 통한 음성/음악 분류 성능향상," 전자공학회 논문지 SP편 48권 2호, 141-148쪽, 2011년 3월

[6] X. Wang, J. Chen, P Wang, Z. Huang, "Infrared human face auto locating based on SVM and a smart thermal biometrics system," in *Proc. Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)*, vol. 2, pp. 1066-1072, October 2006.

[7] A. Ganapathiraju, J. E. Hamaker, J. Picone, "Applications of support vector machines to speech recognition," *IEEE Trans. Signal Processing*, vol. 52, pp. 2348-2355, August 2004.

[8] S. C. Greer, and A. Dejaco, "Standardization of the selectable mode vocoder," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 953-956, May 2001.

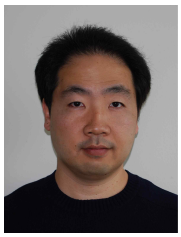
[9] C. V. Goudar, P. Rabha, M. Deshpande, and A. Rao, "SMVLite: reduced complexity selectable mode vocoder," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 701-704, May 2006.

[10] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 988-999, 1999.

[11] J. -M. Kum and J. -H. Chang, "Speech enhancement based on minima controlled

recursive averaging incorporating second-order conditional MAP criterion," *IEEE Signal Processing Letters*, Vol. 16, no. 7, pp. 624-627, July 2009.
[12] John C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*, MIT Press, pp. 61-74, 1999.
[13] J. W. Shin, H. J. Kwon, S. H. Jin, and N. S. Kim, "Voice activity detection based on conditional map criterion," *IEEE Signal Processing Letters*, vol. 15, no. 2, pp. 257-260, February. 2008.
[14] W. M. Fisher, G. R. Doddington and K. M. Goudie-Marshall, "The DARPA speech recognition research database: Specifications and status," in *Proc. DARPA Workshop Speech Recognition*, pp. 93-99, February 1986.

저 자 소 개



임 정 수(정회원)
1996년 인하학교 전기공학과 학사
2004년 University of Maryland ECE 석사.
2009년 North Carolina State University ECE 박사
2010년 인하대학교 박사후 연구원
2011년 목포대학교 연구교수
<주관심분야 : 컴퓨터 구조, 임베디드 시스템, 신호처리, 인공지능>

장 준 혁(정회원)
1998년 경북대하교 전자공학과 학사.
2000년 서울대학교 전기공학부 석사.
2004년 서울대학교 전기컴퓨터 공학부 박사.
2000년～2005년 (주)넷더스 연구소장
2004년～2005년 캘리포니아 주립대학, 산타바바라(UCSB) 박사후연구원
2005년 한국과학기술연구원(KIST) 연구원
2005년～2011년 인하대학교 전자공학부 조교수
2011년～현재 한양대학교 융합전자공학부 부교수
<주관심분야 : 음성신호처리, 오디오 신호처리, 통신 신호처리, 휴먼/컴퓨터 인터페이스>