

# 계층구조의 분류를 통한 서포트벡터머신 기반의 음성/음악 분류기의 실용도 향상기법

최석환, 조용욱, 조지우, 임정수, 이연우, 이성로

목포대학교

srlee@mokpo.ac.kr

## A Technique to Improve the Practicality of SVM-based Speech/Music Classifiers Through Hierarchical Classification

Seokhwan Choi, Youngok Cho, Jiu Cho, Chungsoo Lim, Yeonwoo Lee, Seong Ro Lee

Mokpo Nat. Univ.

### 요 약

본 논문은 제한된 대역폭의 효율적인 활용을 위한 가변 전송률 코덱을 목표로 제안된 서포트벡터머신 기반의 음성/음악 분류기의 실용도를 높이기 위한 기법을 제안한다. 서포트벡터머신 기반의 음성/음악 분류기는 높은 분류능력을 가지고 있지만 많은 계산량을 요구하기 때문에 실시간으로 사용하기에는 부적합한 면이 있다. 따라서 계층적 분류를 통해 서포트벡터머신 기반의 음성/음악 분류기의 실용성을 향상시키는 기법을 제안한다.

### I. 서 론

최근 무선통신기기를 통한 멀티미디어 서비스의 대중화로 제한된 주파수 대역을 효율적으로 활용하는 기법이 활발하게 연구되어지고 있다. 제한된 통신망을 효율적으로 사용하기 위해 가변적인 전송률을 가지는 다양한 음성 코덱이 개발 되었는데[1], 음성신호의 종류에 따라 다른 전송률을 할당하기 위해서는 우선 음성신호의 종류를 정확하게 분별하는 것이 필요하다. 특히 근래에는 기계학습의 한 기법인 서포트벡터머신 (SVM)을 사용하여 3GPP2 표준 코덱인 selectable mode vocoder (SMV) [1]의 음성/음악 분류성능을 향상시키는 기법 [2]이 제안되었다. 이 기법은 별도의 특징추출과정 없이 기존의SMV 파라미터들을 사용하며 코덱내의 기존의 음성/음악 분류성능을 크게 향상시킨다. 그러나 서포트벡터머신은 많은 연산량을 요구하므로 실시간으로 사용하기는 무리가 있다. 따라서 본 논문에서는 계층적 분류를 통해서 서포트벡터머신의 분류 부담을 줄여 음성/음악 분류의 정확도는 유지하면서 전체적으로 실행시간과 에너지소비를 줄이는 기법을 소개한다.

### II. 본론

이 장에서는 RBF를 커널함수로 사용하는 SVM을 복습하

고 계산 량과 메모리 크기에 영향을 미치는 서포트벡터의 개수와 차원이 어떻게 실행시간과 에너지소비량을 변화시키는지 분석하여 본다. 입력벡터 X가 선형으로 분류가 가능한 경우, 서포트벡터머신의 판별함수는 다음 식과 같다.

$$f(X) = \sum_{i=1}^M \alpha_i^* z_i \langle X_i^*, X \rangle + b^* \quad (1)$$

$X_i^*$ 는 학습에 의해 구해진 M개의 서포트 벡터 중 i번째 벡터이다. 최적화 바이어스 (optimization bias)  $b^*$ 와 라그랑제 승수 (Lagrange Multiplier)  $\alpha^*$ 는 학습에 의해 구해지는 quadratic programming problem의 해이다. 위의 식에서 알 수 있듯이 서포트벡터의 개수와 차원은 총 계산량에 직접적인 관계를 가지고 있는데, 복잡 데이터의 분류일수록 다수의 고차원서포트벡터를 요구한다. 따라서 높은 계산량을 요구하게되고 특히 실시간 애플리케이션의 경우 계산량의 감소 없이는 실제로 적용하기가 어렵다.

본논문에서 소개하는 계층식 분류기는 계산량이 적은 간단한 분류기가 첫 단계에 위치하고 서포트벡터머신의 분류기는 두 번째 단계에 위치한다. 첫단계의 분류기가 분류를 수행하는 경우 두번째 단계에 있는 서포트벡터머신 기반 분류기는 사용하지 않게 된다. 이렇게하여 계층적 구조의 분류기는 기존의 서포트벡터머신 분류기에 비하여 적은 연산량을 요구하게 되어 분류시간과 에너지소비의 감소를 얻을 수 있다.

첫번째 단계에 위치하는 분류기는 다음과 같은 세 가지의 조건을 만족해야 한다. 첫째, SMV 코덱에 별 무리 없이 내장되어야 한다. 둘째, 서포트벡터기반의 분류기보다 훨씬 간단해야 한다. 셋째, 최소한 한가지 클래스에 대한 분류성능이 좋아야 한다. 첫번째 단계의 분류기의 후보로 가장 쉽게 생각할 수 있는 SMV 기존의 음성/음악 분류기를 살펴보면, 첫번째와 두번째 조건은 쉽게 만족시키는 것을 알 수 있다. 원래 SMV 코덱안에서 동작하는 매우 간단한 분류기이기 때문이다. 이 분류기를 살펴보면, 코덱내에서 만들어지는 두가지 파라미터 (running mean of the periodicity counter, music continuity counter)를 미리 정의된 문턱값과 비교하여 음성/음악을 분류하는 간단한 구조로 되어있다. 그러나 세번째 조건은 잘 만족시키지 못한다. 거의 모든 음성프레임은 정확하게 분류가 되지만 음악프레임의 경우 17%정도만이 음악으로 분류된다. 그러나 음악으로 분류된 프레임 중 90%가 실제로 음악프레임인 것을 관찰할 수가 있었다. 즉 17%의 음악프레임이 90%의 정확도를 가지고 분류가 되었다는 것이고 첫 단계의 분류기로 사용이 가능하다는 것이다.

그러나 기존의 SMV 음성/음악 분류기를 그대로 사용하게 된다면 20%남짓의 입력프레임만이 첫번째 분류기에 의해 음악으로 분류가 되고 서포트벡터 기반 분류기는 사용하지 않게 된다. 그러므로 첫번째 분류기의 사용도를 20%보다 높이기 위하여 기존의 문턱값을 줄여서 사용한다. 문턱값으로는 계층구조의 분류기의 분류성능은 유지하면서 첫번째 분류기의 사용도를 최대화할 수 있는 값을 사용하였다.

### III. 실험

실험을 위하여 [2]에서 사용된 음성/음악 데이터를 이용하였고 또한 제안된 서포트벡터머신 기반 분류기를 비교대상으로 삼았다. 또한 분류시간과 에너지소비를 측정하기 위해 사이클 (cycle)단위로 프로세서의 행동과 에너지소비를 모델한 시뮬레이터인 sim-panalyzer [3]를 사용하였다. 사용된 시뮬레이터의 파라미터는 표 1에 정리하였다.

표 1. 프로세서 시뮬레이터 파라미터

ISA	ARMv4T
Clock Freq	800Mhz
Pipeline	5 stages, in-order, dual issue
Branch	Dynamic prediction (bimodal)
Cache	64KB separate Inst/Data cache, 4way, 32B line, 1 cycle latency
Memory	64bit bus, 100 cycle latency, 1 port
FPU	1 adder (2 cycle, pipelined) 1 Multiplier (4 cycle, pipelined)

제안된 계층구조의 분류기를 검증하기 위하여 기존의 서포트벡터머신 기반의 분류기 [2]와 비교하였다. 그림 1은 기존의 분류기와 계층적 구조의 분류기의 분류정확도, 한개의 프레임을 분류하는데 걸리는 평균시간, 그리고 한개

의 프레임을 분류하는데 소비되는 평균에너지를 나타내었다. 세가지 측정값 모두 서포트벡터머신 기반의 분류기의 측정값으로 정규화 되었다. 그림에서 볼 수 있듯이 분류의 정확도는 유지되면서 실행시간은 21%, 그리고 에너지 소비는 17%가 감소하였다. 즉 계층적 분류를 사용하여 서포트벡터머신 기반의 분류기의 사용횟수를 줄임으로써 실행시간과 에너지소비를 감소시킬 수 있음을 증명하였다.

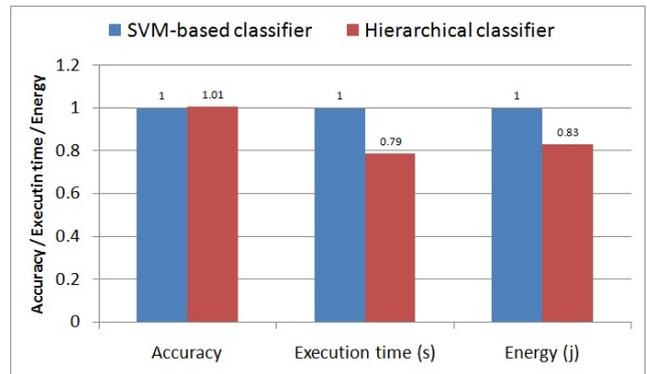


그림 2. 계층구조의 분류기와 서포트벡터 기반 분류기의 비교.

### IV. 결론

본 논문에서는 서포트벡터머신 기반 음성/음악 분류기의 실용성을 높이기 위하여 계층적 구조의 분류기를 제안하였다. 서포트벡터머신 기반의 분류기 앞에 간단한 SMV 기존의 분류기를 변형하여 필터로 사용함으로써 서포트벡터머신 기반의 분류기의 사용도를 줄여 전체적으로 분류성능은 유지하면서 실행시간과 소모 에너지를 감소시켰다.

### ACKNOWLEDGMENT

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임(2011-0022980)

본 연구는 지식경제부 및 정보통신산업진흥원의 IT융합 고급인력과정 지원사업의 연구결과로 수행되었음”(NIPA-2012-H0402-12-1001)

### 참고 문헌

[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, Apr. 2005.

[2] 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, Feb. 2009.

[3] T. Austin, T. Mudge, and D. Grunwald, Sim-panalyzer. (<http://www.eecs.umich.edu/~panalyzer/>).