바이올린과 첼로 연주 데이터를 이용한 분류 알고리즘의 성능 비교

정회원 김 재 천*, 종신회원 곽 경 섭**

Performance Comparison of Classification Algorithms in Music Recognition using Violin and Cello Sound Files

Jae Chun Kim*, Kyung sup Kwak** *Regular Members*

요 약

음악인식에 주로 사용되는 세 가지 알고리즘의 성능을 비교하였다. 다양한 분류알고리즘을 소개하고 그 중 베이지안법, 최근접이웃법과 k-최근접이웃법을 이용하여 악기를 분류하였다. 악기 샘플파일에서 영교차율, 평균, 분산, 평균피크레벨의 4가지 특성값을 추출하여 분류시스템의 데이터로 사용하였다. 사용된 악기 샘플은 바이올린, 바로크 바이올린, 바로크 첼로이다. 실험결과 최근접이웃 알고리즘이 악기 분류에 있어서 가장 좋은 성능을 보여 주었다. 최근접이웃 알고리즘은 단순하면서도 빠른 계산결과를 보여 악기 분류에 적절한 알고리즘으로 판단되었다.

Key Words : music, spectral analysis, audio signal, pattern recognition, data mining.

ABSTRACT

Three classification algorithms are tested using musical instruments. Several classification algorithms are introduced and among them, Bayes rule, NN and k-NN performances evaluated. ZCR, mean, variance and average peak level feature vectors are extracted from instruments sample file and used as data set to classification system. Used musical instruments are Violin, baroque violin and baroque cello. Results of experiment show that the performance of NN algorithm excels other algorithms in musical instruments classification.

Ⅰ. 서 론

특성 값(feature vector)은 대상을 특정한 범주로 분류하는데 있어서 데이터로 사용된다. 음악인식시스템에서 분류단계는 특성추출 다음 단계에 해당한다. 다양한 분류기법이 존재하고 있으며 본 논문에 적절하다고 판단되는 기법에 대한 개념, 수학적 이론에 대하여 논할 것이다.

분류이론은 판단이론(decision theory)의 하위 분야이다. 일정한 패턴이 특정한 군(class)에 포함된다는 가정에 기반 한다. 개별적인 패턴의 차이에도 불구하고 패턴으로부터 추출한 일련의 특성들이 일정한 범주에 포함된다. 이러한 특성들을 판단하여 범주에 대한 소속 여부를 결정 한다. 음악의 경우 같은 장르에 포함된 곡이라 할지라도 많은 부분에서 차이를 보인다. 이런 상황에서도 인간은 쉽게 장르 판단을 할 수 있는데 이는 동일 장르는 공유할 수 있는 기본특성을 가지고 있다는 것을 증명한다.

분류는 기하학적인 접근을 통하여 접근하는 것이 용이하다. 전술한 바와 같이 특성값은 특성공간의

* 인하대학교 전자공학과 통신공학연구실 (milgaroo@naver.com),
** 인하대학교 정보통신대학원 (kskwak@inha.ac.kr)
논문번호 : KICS2005-01-009, 접수일자 : 2005년 1월 4일

점들로 위치한다. 분류시스템의 목적은 특정범주에 속한다고 판단할 수 있는 특성공간상의 판단경계선을 찾는 것이다. 그러므로 입력된 데이터는 공간상의 위치를 판단하여 범주를 결정할 수 있다[1,2].

Ⅱ. 분류시스템의 구성요소

특성 값 추출 시스템은 특정한 상황에 적합하도록 설계되어야 한다. 기본적으로 영역(domain)에 종속적이다. 반면 분류기법은 영역으로부터 독립적인 작업이다. 이는 특성 값 추출은 일종의 압축단계로서 영역에 종속적인 데이터를 일반적인 수치표현으로 옮기는 작업으로 이해할 수 있다. 이렇게 변환된 일반적인 수치는 분류시스템의 작업 대상이 된다. 특성 값 추출은 음악, 음향심리학, 신호 처리 등 다양한 분야에 대한 이해가 필수적이다. 음성인식의 경우 기존 연구가 많이 축적되어 있어 음악인식에 필요한 다양한 도구를 제공할 수 있다.

■ 설계요소

분류시스템의 설계에 있어서 고려해야 할 요소는 특성 값의 선택, 특성 값의 수, 데이터거리 측정, 시스템의 신뢰성 등이다. 분류시스템은 가공되지 않은 데이터(raw data)를 다루지 않는다. 특성 값만을 분류대상으로 사용하므로 적절한 특성 값의 추출은 분류시스템 설계에 있어 필수적이다.

■ 해결해야할 과제

두 가지 요인이 같은 범주에 속하는 패턴의 차이를 발생 시킨다. 첫 번째, 시스템에 사용된 모델에서 기인할 경우이다. 모델이 단순할 경우 정제되지 않은 데이터를 출력하고 이 모델에 대한 정보가 없는 사용자는 큰 혼란을 겪게 된다. 두 번째, 의도하지 않은 잡음에 의한 차이발생이다. 어떤 원인에 의하여 차이가 발생했는지 안다면 이에 대처할 수 있다.

■ 시스템의 훈련 및 학습

분류시스템의 형태의 결정과 학습을 통한 파라미터 값의 설정은 특정분류시스템 설계의 핵심이다. 여기서 학습이라 함은 샘플데이터를 이용하여 미지의 파라미터 값을 찾는 과정이다. 분류는 감독학습(supervised learning)으로 볼 수 있다. 상태를 알고 있는 특성 값을 사용하여 분류시스템을 학습시킨다. 이렇게 학습된 분류시스템은 미지의 특성 값을 입력 받아 분류결과를 출력한다. 비감독학습(unsupervised learning; clustering)의 경우, 미리 정해진 상태를 사용하지 않고 시스템 자신이 결정한 기준으로 분류한다.

■ 판단경계의 일반화

학습 과정에 있어서 판단경계의 일반화 적절한 타협점을 찾는 과정이다. 만일 분류시스템이 트레이닝 데이터에 최적화된다면 미지의 데이터에 대한 분류 정확도는 감소할 것이다. 적절한 타협점을 찾음으로서 시스템의 성능을 높일 수 있다. 학습 데이터에 의해 결정된 경계는 적절한 타협점으로 일반화되어야 한다.

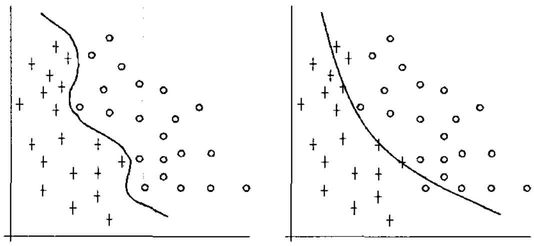

그림 1. (a)는 학습 데이터의 경우 완벽한 성공률을 보이지만 미지의 데이터의 경우 실패율이 높아진다. (b)는 학습 데이터에 최적화 되지는 않았지만 전체적인 성공률은 높아진다.

Ⅲ. 분류 알고리즘의 종류

3.1 베이지언 추정이론

분류시스템의 목표는 판단성공률을 높이는 것이다. 베이지언(Bayesian) 추정이론은 이러한 접근방법의 하나이다. 확률이론에 근거하여 타협점(trade-off)를 정량화한다.

x는 개체의 상태이며 P(x)는 개체의 상태가 x일 확률이다. ω_i는 군(class)으로 서 전체 군을 몇 가지 상태로 구분한 것이다. 예를 들면 ω_1=재즈, ω_2=국악, ω_3=클래식 이라 하면 $P(\omega_1)+P(\omega_2)+P(\omega_3)=1$이다. $P(\omega_i|x)$는 x 상태일 때 ω_i 일 조건부확률이며 $P(x|\omega_i)$는 ω_i 군에 속할 때 x 상태일 확률이다.

$$P(x|w_i) = P(x, w_i)/P(w_i) \tag{1}$$

여기서 $P(x, \omega_i)$는 두 사건 동시에 일어날 확률이다.

$$P(w_i|x) = P(w_i, x)/P(x) \tag{2}$$

$P(w_i, x) = P(x, w_i)$이므로 식(2)를 식(3)으로 정리한다.

$$P(x, w_i) = P(w_i|x)P(x) \tag{3}$$

식(3)을 식(1)에 대입하면

$$P(x|w_i) = P(w_i|x)P(x)/P(w_i) \tag{4}$$

식(4)를 사후확률(posterior probability)이라 한다.

전체 군을 ω_1, ω_2, ω_3으로 나누었다면 $P(w_1|x) + P(w_2|x) + P(w_3) = 1$이 된다. 예를 들어 $P(w_1|x) = 0.8$, $P(w_2|x) = 0.1$ 그리고 $P(w_3|x) = 0.8$이면 미지의 x를 만난다면 시스템은 x를 w_1으로 분류하게 된다. 따라서 어떤 개체가 x라는 상태에 있으면 $P(w_i|x)$ 값이 가장 큰 w_i에 x가 속할 것이라고 예측한다.

$$P(w_1|x) < P(w_2|x) \tag{5}$$

라 가정하면 어떤 객체가 x라는 상태일 때 ω_2 군에 속한다고 예측한다. 식(5)의 좌우변을 모두 치환하면

$$P(x|w_1)P(w_1)/P(x) < P(x|w_2)P(w_2)/P(x) \tag{6}$$

식(5)를 정리하면

$$P(x|w_1)/P(x|w_2) < P(w_2)/P(w_1) \tag{7}$$

식(7)을 베이즈 추정법(Bayes' decision rule) 이라 한다. 식(5, 6, 7)에서 왼쪽 항이 더 크다면 어떤 객체 x는 ω_1에 속할 것이고 오른쪽 항이 더 크다면 객체 x는 ω_1에 속할 것이다. 양쪽 항 모두 학습을 통해 알 수 있는 값이다. 왼쪽 항은 우도비(likelihood ratio), 오른쪽 항은 문턱값(threshold)라 한다.

3.1.1 판별함수

판별함수(discriminant function)는 c개의 $g_i(x)$ 판별함수를 계산하여 가장 큰 값으로 분류하는 네트워크로 표현할 수 있다. $g_i(x) > g_j(x)$ $(i \neq j)$를 만족하면 특성값 x를 w_i 군으로 분류한다.

각각의 판별함수는 범주 w_i와 관계되고 gi(x)가 최대가 되는 범주로 분류된다. 베이즈 분류시스템이 최소 에러비(error ratio)를 가질 경우, 판별함수의 최대값은 최대사후확률에 해당한다. 따라서 gi(x)= $g_i(x) = P(w_i|x)$이며 판별함수를 식(10)과 같이 나타낼 수 있다.

$$g_i(x) = P(x|w_i)P(w_i) \tag{8}$$

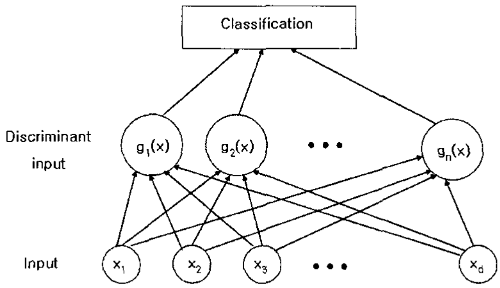

그림 2. d개의 입력 xi와 c개의 판별함수 gi(x)로 구성된 분류시스템의 구조

3.1.2 정규밀도 분포

베이즈 분류시스템은 사전확률 $P(w_i)$와 조건부 확률밀도 함수 $P(x|w_i)$ 의해 결정된다. 정규밀도분포는 가장 일반적인 밀도분포이다. 정규밀도분포함수는 w_j 군의 특성벡터 x가 정형(prototype) 벡터 μ_j에 기반 할 경우를 모델링한 것이다. d차원의 정규밀도 함수는 식(9)과 같이 표현할 수 있다.

$$P(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \sigma^{-1}(x-\mu)} \tag{9}$$

여기서 x는 d 차원의 열벡터이고 μ는 d 차원의 평균(mean) 벡터이다. σ는 d-by-d 공분산(covariance) 행렬이다. 식(9)의 단순화된 표현은 식(10)과 같다.

$$P(x) \sim N(\mu, \sigma) \tag{10}$$

기대값 ε은 식(13)과 같다.

$$\varepsilon[x] = \int x p(x) dx \equiv \mu \tag{11}$$

공분산(covariance) 행렬은 식(12)와 같다.

$$\varepsilon[(x-\mu)(x-\mu)^T] = \int (x-\mu)(x-\mu)^T P(x) dx \equiv \sigma \tag{12}$$

만일 데이터가 정규분포를 이루고 평균과 공분산 μ, σ 값을 알고 있으면 군판별함수의 조건부밀도함

수인 $P(x|w_i)$은 식(9)에 정의된 정규밀도함수로 대체할 수 있다. 정규밀도함수는 강력한 도구로서 분류객체가 정규분포성질을 가지고 있을 때에만 적용할 수 있다. 그러나 자연 상태에서 정규분포는 그리 흔한 경우는 아니다. 실제 분류시스템에서의 전형적인 문제는 다음과 같다.

완전한 확률분포구조에 대하여 알 수 없으므로 정규밀도와 같은 표준적인 분포모델을 따른다고 가정한다. 이런 경우, 평균벡터 μ와 같이 조건부확률밀도함수의 계수(parameter)들을 결정해야 한다.

실제 데이터의 구조는 미리 알 수 없으므로 분류시스템은 학습 데이터를 통해서 정보를 얻어야 한다.

대부분의 분류상황에서 확률의 밀도분포는 알 수 없으므로 일반적인 분포형태에 대한 가정을 해야 한다. 일반적인 형태의 분포를 결정하는 것은 함수의 계수만 결정하면 되므로 분류시스템구현이 용이해진다. 미지의 확률분포는 샘플데이터를 이용한 학습을 통해 구해진다. 예를 들면, $P(x|w_i)$를 정규분포라 가정하면 평균 μ와 공분산 σ 값만 찾아내면 된다.

3.2 은닉 마코브모델

은닉 마코브모델(Hidden Markov Models; HMMs)은 일련의 판단을 내리는 문제에 대한 해법을 제시한다[3]. 분류시스템에 있어서 일련의 상태를 거치는데 시간 t의 상태는 이전 상태의 영향을 받은 결과이다. HMMs은 시간 t의 상태는 시간(t-1)의 영향만 받는 다고 가정한다. 음성은 연속적인 음의 발생이라는 측면에서 HMMs의 주요 응용분야중 하나이다. 마코브 모델의 주요개념은 전이확률(transition probability)이다.

전이확률 $a_{ij}=P(w_j(t+i)|w_i(t))$은 시간 t의 상태 w_i 일 때 시간(t+1)의 상태 w_i일 확률이다. 전이 확률은 대칭적일 필요는 없고 같은 상태를 한번 이상 거쳐 갈 수 있다. 상태 $w(t)$에 접근할 수 없을 경우 심볼만이 관측가능하다. 상태 $w(t)$에 있는 시스템은 심볼 $v(t)$를 방출한다. 확률 $b_{jk}=P(v_k(t)|w_j(t))$는 상태 $w_j(t)$에 있는 시스템이 심볼 $v_k(t)$를 방출 할 확률이다. HMMs에서 의미하는 은닉은 직접 접근할 수 없고 숨겨져 있다는 의미이다.

HMMs의 구조(상태와 심볼의 개수)와 학습샘플의 개수를 알면 확률 a_{ij}와 b_{jk}를 알 수 있다. 이를 마코브 모델 트레이닝(Markov model training)이라 한다.

3.3 최근접이웃법

사후확률을 예측하는데 가장 많이 사용하는 기법은 최근접이웃법(nearest-neighbour rule; NN)이다. 어떤 x를 분류하고자 할 때 인접한 기준점(prototype point)이 군 $D=(x_1,\ldots,x_n)$에 속한다면 x도 군 D에 속한다고 판단한다. k-최근접이웃법(k-nearest neigbor rule; k-NN)은 최근접이웃법의 확장판이다. 어떤 x의 k개수 인접 점들의 레이블(label)을 조사한다. 특정 레이블의 상대적인 빈도는 $P(w_i|x)$로 표시한다. $P(w_i|x)$ 값이 가장 큰 값일 때 x는 w_i로 분류된다. 여기서 최적의 k값을 찾는 것이 핵심이다. 예측에 사용된 인접점들은 $P(w_i|x')\approx P(w_i|x)$를 만족할 만큼 가까워야한다. k는 신뢰성 있는 값을 보장할 수 있을 만큼 커야 한다.

최근접이웃법과 같은 알고리즘은 특성 공간상에서 특성벡터 사이의 거리를 측정한다. 일반적인 거리측정은 민코브스키 측정법(Minkowski metric) 식(14)를 사용한다.

연속적인 값의 거리를 측정하는 데 가장 많이 쓰이는 것은 유클리디안 거리(Euclidean distance)이다. 이는 2차원 또는 3차원 상의 거리를 측정하는데 흔히 사용된다. 민코브스키 측정법에서 p=2인 경우가 유클리디안 거리이다.

$$d_p(x_i,x_j)=\sqrt[p]{\sum_{k=1}^{d}(|x_{i,k}-x_{j,k}|)^p}=\|x_i-x_j\|_p \tag{14}$$

$$d_2(x_i,x_j)=\sqrt[p]{\sum_{k=1}^{d}(|x_{i,k}-x_{j,k}|)^p}=\|x_i-x_j\|_2 \tag{15}$$

3.4 k-최근접이웃법

k-최근접이웃(k-nearest neighbor)법은 주어진 데이터 셋 (x_i,y_i)을 이용하여 x에 대한 y의 값을 예측한다. x에 가까운 x_i의 k값을 이용하여 y값을 예측한다. 여기서 k는 일종의 계수로서 가장 영향이 큰 요인에 가중치를 두어 예측할 수 있다. 예측과정은 다음과 같다. 학습 데이터 셋은 분류대상의 분류에 사용될 계수를 생성하는데 사용한다. 즉, 분류시스템에는 분류대상(국악, jazz, classical)이 있고 이를 예측하기 위한 각 음악장르의 특징(음의 질감, 에너지, 리듬 등)들이 존재한다. 분류대상과 가장 가까운 학습 데이터 셋의 k 계수를 구한다. 그리고 유클리디안 거리 측정법을 사용하여 분류대상과 학습

데이터의 거리를 구한다. 이러한 과정을 나머지 분류대상 데이터에 대하여 반복한다. k 계수가 많아질수록 계산시간은 길어지지만 비이상적인 데이터의 악영향을 감소시킬 수 있다. 실제로는 수십 개 이하의 k 값을 사용한다.

NFL(Nearest Feature Line)은 최근접이웃법과 유사한 비계수 분류기법이다. 최근접이웃법은 개별적으로 x에 인접한 기준점 x'와 비교하지만 NFL은 한 번에 여러 개의 인접 기준점 정보를 이용한다. NFL은 군내의 두개의 기준점 x_i, x_j를 통과하는 선이다. 각각의 군내에 가능한 기준점 쌍을 통과하는 NFL이 생성된다. 따라서 점 x와 NFL의 최소거리를 사용하여 분류작업을 한다[4].

Ⅳ. 실 험

같은 데이터그룹을 이용하여 베이즈법, k-최근접이웃법, 최근접이웃법의 분류성공률을 비교하였다. 데이터로는 바로크 바이올린, 바로크 첼로, 바이올린 표본을 사용하였다. 바로크 악기는 18세기 이전에 사용되던 악기로서 현대악기의 전신이다. 바로크 바이올린과 현대 바이올린의 차이점은 현의 재질이 현대악기는 금속이고 바로크악기는 동물의 내장을 꼬아서 만든 것이다. 외관상구조적인 모습은 유사하다. 바로크악기는 현대악기보다 음량이 작지만 섬세한 음색을 발성한다[5, 6]. 표본은 44.1k, 15초 길이의 96개 데이터를 준비했다. 각각의 악기 표본은 32개로 구성되어 있다. 한 개의 표본마다 5개의 특성값을 추출하였다. 분류시스템은 10진 교차유효화(ten-fold cross-validation)를 사용하여 훈련하였다. 교차유효화(cross-validation)는 패턴인식에서 사용하는 일반적인 훈련기법으로 데이터집합은 같은 크기의 n부분으로 나눠지고 (n-1)부분은 분류 시스템을 학습시키기 위해 사용한다. 나머지 n번째 부분은 학습된 시스템을 시험하기 위해 사용한다[7].

4.1 분류에 사용된 특성값

■ 영교차율

영교차율(Zero Crossing Rate; ZCR)은 시간영역의 특성 값으로 처리시간이 빠르고 시간에 따른 변화를 분석하고 통계적 특성을 이용하여 음성(voiced) 신호와 비음성 신호를 구분하는데 주로 사용되었다. 식(16)과 같이 표시하며 음성신호는 비음성신호보다 작은 영교차율값을 갖는다.

$$Z_n = \sum_m |sgn[x(m)] - sgn[x(m-1)]| w(n-m) \tag{16}$$

여기서

$$sgn[x(n)] = \begin{cases} 1 & , x(n) \ge 0 \\ -1 & , x(n) < 0 \end{cases} \tag{17}$$

그리고

$$w(n) = \begin{cases} 1/2 & , 0 \le n \le N-1 \\ 0 & , otherwise \end{cases} \tag{18}$$

m은 함수의 윈도우크기이다. 음악은 일정시간 동안 안정적인 경향이 있기 때문에 음성신호보다 작은 영점교차율 분산을 갖는다. 발자국 소리, 새의 지저귐 소리도 음성신호와 유사한 특성을 갖으며 같은 악기 중에서도 피아노와 같은 단속적인 소리는 현악기보다 음성과 유사한 특성을 보인다[8].

■ 평균, 분산

신호의 평균값은 일반적인 인식시스템에서는 사용하지 않는다. 그러나 본 실험에서는 한정된 데이터집합을 이용해 분류시스템을 실험하는 것이므로 유용하게 사용된다.

■ 평균 피크레벨

신호의 피크레벨의 평균값이다. 일정한 임계값을 초과하는 신호는 피크로 간주하고 피크의 빈도와 세기를 측정한다. 음압 또는 음량에 따른 악기의 특성을 수치화한다.

$$AVpeaklevel = \sum \frac{x(t)}{f_p} \tag{19}$$

x(t)는 측정된 피크의 크기이고 f_p는 피크의 빈도이다.

■ 군

학습데이터가 소속된 범주를 말한다. 본 분류시스템은 감독학습(supervised learning)의 일종으로 데이터의 군을 이용하여 학습한다.

4.2 실험결과

표 1,2,3은 실제 입력군과 출력결과의 비교표이다. 구분의 세로항은 실제 입력군을 나타내고 가로항은 분류결과이다. 악기별로 32개의 샘플을 사용하여 전

체 96개의 샘플을 사용하였다. 따라서 표 1의 경우 현대바이올린의 입력과 출력이 만나는 칸의 값 32는 현대바이올린 32개의 샘플이 정확하게 분류된 것을 나타낸다. 바로크첼로의 경우 입력과 출력이 만나는 칸에 31개가 분류되었고 바로크바이올린으로 1개의 샘플이 분류되었다. 바로크바이올린의 경우 21개의 샘플이 모던바이올린으로 분류되어 가장 낮은 성공률을 보였다. 표 3은 모던바이올린의 경우 5개의 샘플이 바로크바이올린으로 분류되었고 바로크첼로는 31개가 정확히 분류되었고 나머지 1개는 분류되지 못하였다. 바로크바이올린은 3개의 샘플이 현대바이올린으로 분류되었고 1개는 분류되지 못하였다. 표 3에서 일반화된 분류경계선상의 샘플 2개는 최종적으로 분류되지 못하였다.

표 1. 베이즈법 실험결과, 분류성공률 77.0833%.

구분	modern-violin 출력	baroque-cello 출력	baroque-violin 출력
modern-violin입력	32	0	0
baroque-cello입력	0	31	1
baroque-violin입력	21	0	11

표 2. 최근접이웃법 실험결과, 분류성공률 88.5417%.

구분	modern-violin 출력	baroque-cello 출력	baroque-violin출력
modern-violin입력	26	0	6
baroque-cello 입력	0	32	0
baroque-violin 입력	3	2	27

표 3. k-최근접이웃법 시험결과, 분류성공률 89.5833%, 2개의 표본데이터는 분류되지 못 함.

구분	modern-violin 출력	baroque-cello 출력	baroque-violin출력
modern-violin입력	27	0	5
baroque-cello입력	0	31	0
baroque-violin입력	3	0	28

표 4, 5, 6의 TP rate(True Positive)는 특정군의 요소로서 해당 군으로 분류된 것을 말한다.

FP rate(False Positive)는 특정군의 요소가 아니면서 특정군으로 분류된 것을 말한다. FN rate (False Negative)는 특정군의 요소로서 다른 군으로 분류된 것을 나타낸다. Precision은 특정클래스로 분류된 모든 요소들 중에서 정확히 분류된 요소들의 비를 나타낸다. Recall은 특정군에 속하는 모든 요소들 중에서 정확히 분류된 요소들의 비를 나타낸다.

$$Precision = \frac{|TP|}{|TP| + |FP|} \quad (20)$$

$$Recall = \frac{|TP|}{|TPRIGHT| + |FNRIGHT|} \quad (21)$$

F-measure는 Precision과 Recall을 통합한 파라미터로서 Recall과 Precision이 모두 필요할 때 사용한다.

$$F = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} = \frac{2|TP|}{2|TP| + |FP| + |FN|} \quad (22)$$

표 4. 베이즈법, 군별 세부 정확도.

구 분	TP rate	FP rate	Precision	Recall	F-Measure
modern-violin	1	0.328	0.604	1	0.753
baroque-cello	0.969	0	1	0.969	0.984
baroque-violin	0.344	0.016	0.917	0.344	0.5

표 5. 최근접이웃법, 군별 세부 정확도.

구 분	TP rate	FP rate	Precision	Recall	F-Measure
modern-violin	0.813	0.047	0.897	0.813	0.852
baroque-cello	1	0.031	0.941	1	0.97
baroque-violin	0.844	0.094	0.818	0.844	0.831

표 6. k-최근접이웃법, 군별 세부 정확도

구 분	TP rate	FP rate	Precision	Recall	F-Measure
modern-violin	0.844	0.048	0.9	0.844	0.871
baroque-cello	1	0	1	1	1
baroque-violin	0.903	0.079	0.848	0.903	0.875

표 7의 카파통계(kappa statistics)는 만일 입력데이터 각각의 군별 소속을 알고 있다면 분류결과의 출력물과 비교한 성공률을 나타낸다. Kappa=1이면 분류결과의 완벽한 일치를 나타내고 Kappa=0이면 전혀 일치하지 않음을 나타낸다. 음의 Kappa값은 임의의 분류확률보다 낮게 일치하였음을 나타낸다. 일반적으로 카파통계값이 0.9 이상이면 우수하다고 판단하며 0.7이하는 개선이 필요함을 나타낸다. 평균자승오차(mean squared error)는 실제값과 예측값의 차이를 측정한다.

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(a_i-p_i)^2 \tag{23}$$

표 7. 전체 정확도

구 분	Bayes rule	NN	k-NN
Correctly Classified Instances	74	85	86
Incorrectly Classified Instances	22	11	8
Kappa statistic	0.6563	0.8281	0.8724
Mean absolute error	0.1382	0.0764	0.0587
Root mean squared error	0.3335	0.2764	0.2344
Relative absolute error	31.0835%	17.1754%	13.4738%
Root relative squared error	70.6988%	58.5868%	50.2108%
Total Number of Instances	96	96	96

V. 결 론

소리신호 분류에 가장 많이 사용되는 분류알고리즘을 비교하였다. 표 1, 2, 3에서와 같이 k-최근접이웃 알고리즘이 분류성공률 89.5833%로서 악기인식에 있어 빠른 계산과 가장 좋은 성공률을 보였다. 베이즈법은 77.0833% 성공률로 가장 낮은 결과를 보였다. 분류 오류는 바로크 바이올린과 현대 바이올린서 발생하였다. 두 악기는 바이올린으로서 음역대와 음색이 유사하여 오류가 발생한 것으로 판단된다. 반면 바로크 첼로는 음역대가 낮아 상대적으로 낮은 오류율을 보였다. 전체적으로 벡터거리 알고리즘을 사용한 k-최근접이웃법과 최근접이웃법이 좋은 성능을 보였다. 음악 파일로부터 추출한 특성값은 2차원 벡터상의 위치이므로 벡터거리를 이용한 알고리즘의 분류결과가 우수한 것으로 확인되었다.

참 고 문 헌

〔1〕 Richard O. Duda, Peter E. Hart, and David G. Stork, "Pattern classification," *John Wily & Sons*, 1999.

〔2〕 Anil K. Jain, Robert P.W. Duin, and Jianchang Mao, "Statastical pattern recognition: A review," IEEE *Transactions On Pattern Analysis and Machine Intelligence*, 22(1):4-37, 1999.

〔3〕 L. R. Rabiner, and B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, 3(1):4-16, January 1986.

〔4〕 Stan Z. Li, "Content-based classification and retrieval of audio using the nearest feature line method," *IEEE Transactions on speech and audio processing*, 9(5):619-615, September, 2000.

〔5〕 박초연, "바로크시대의 운궁악기의 조건과 주법에 관한 연구," *서울대학교 대학원*, 1987.

〔6〕 조선우, "바로크시대의 연주실제에 관한 고찰," *음악이론연구*, Vol.5, No.0, 247, 2000.

〔7〕 Ian H. Witten, Eibe Frank, "Data mining: practical machine learning tools and techniques with java implementations," *Morgan Kaufmann*, 1999.

〔8〕 T. Zhang, and C. Kuo, "Content-based classification and retrieval of audio," *In SPIE's 43rd Annual meeting-Conference of advanced signal processing agorithms, architectures, and implementations VIII*, San Diego, July, 1998.

김 재 천(Jae Chun Kim) 정회원

1996년 2월 인하대학교 전자공학과 학사 졸업

1999년 2월 인하대학교 첨단정밀공학과 석사 졸업

2003년 3월 인하대학교 전자공학과 박사 수료

<관심분야> 통신이론, DSP, Audio Analysis, Data Retrieving

곽 경 섭(Kyung sup Kwak) 종신회원

1977년 2월 인하대학교 전기공학과 졸업

1979년 2월 인하대학교 전기공학과 석사

1981년 12월 미국 USC 전기공학과 석사

1988년 2월 미국 UCSD 통신이론 및 시스템 박사

1988년 2월~1989년 2월 미국 Hughes Network Systems 연구원

1989년 2월~1990년 3월 미국 IBM Network Analysis Center 연구원

1995년 1월~1999년 12월 IEEE Seoul Section 총무이사

2000년 3월~2002년 2월 인하대학교 정보통신전문대학원 원장

2003년~현재 World Scientific Journal Assoc. Editor

2003년~현재 인하대학교 초광대역 무선통신 연구센터장

2004년 1월~2004년 12월 한국통신학회 감사

2005년 1월~현재 한국통신학회 수석 부회장

<관심분야> 이동통신, 멀티미디어통신, UWB, 무선네트워크