

알파 다이버전스를 이용한 무게중심 모델 기반 음악 유사도

Centroid-model based music similarity with alpha divergence

서진수,[†] 김정현,* 박지현*

(Jin Soo Seo,[†] Jeonghyun Kim,* and Jihyun Park*)

강릉원주대학교 전자공학과, *한국전자통신연구원 콘텐츠 연구본부

(Received August 20, 2015; revised November 2, 2015; accepted December 18, 2015)

초 록: 음악 유사도 계산은 음악 검색 및 분류 등의 정보 처리 시스템 구현에 있어서 가장 중요한 부분이다. 본 논문은 최근 제안된 무게중심 모델을 이용한 음악 검색 방법에 대해서 살펴보고, 무게중심 모델의 확률 분포 유사도를 이용하여 음악 검색을 수행하고 성능을 평가하였다. 확률 분포간의 거리는 주어진 두 개의 확률 분포가 특정 기준에서 얼마나 가까운 지를 계산하는 것으로 다이버전스라고 불리기도 한다. 본 논문에서는 무게중심 모델에서 확률 분포 간의 거리 비교 시에 알파 다이버전스를 활용하였다. 알파 다이버전스는 알파 값에 따라 다양한 형태를 가지며, 널리 사용되고 있는 KLD(Kullback-Leibler)와 BD(Bhattacharyya Distance)를 포함한다. 음악 장르와 가수 데이터셋에서 검색 실험을 수행했고, 확률 분포 거리 기반 유사도와 벡터 거리 기반 유사도의 음악 검색 성능을 비교하였다. 알파 다이버전스를 통해서 무게중심 모델 기반 음악 검색 성능을 개선시킬 수 있음을 보였다.

핵심용어: 음악 검색, 음악 유사도, 알파 다이버전스, Renyi 다이버전스, KL 다이버전스, Bhattacharyya distance

ABSTRACT: Music-similarity computation is crucial in developing music information retrieval systems for browsing and classification. This paper overviews the recently-proposed centroid-model based music retrieval method and applies the distributional similarity measures to the model for retrieval-performance evaluation. Probabilistic distance measures (also called divergence) compute the distance between two probability distributions in a certain sense. In this paper, we consider the alpha divergence in computing distance between two centroid models for music retrieval. The alpha divergence includes the widely-used Kullback-Leibler divergence and Bhattacharyya distance depending on the values of alpha. Experiments were conducted on both genre and singer datasets. We compare the music-retrieval performance of the distributional similarity with that of the vector distances. The experimental results show that the alpha divergence improves the performance of the centroid-model based music retrieval.

Keywords: Music retrieval, Music similarity, Alpha divergence, Renyi divergence, KL divergence, Bhattacharyya distance

PACS numbers: 43.75.Zz, 43.60.Uv

1. 서 론

정보 처리 기기 및 네트워크 기술의 발달에 따라 음악 데이터를 빠르고 신뢰성 있게 검색하고 제공할 수 있는 오디오 정보 처리 및 검색 기술의 중요성이 증대되고 있다.^[1,2] 음악 정보 처리 서비스는 사용

하는 유사도 기준에 따라서 다양하며, 핑거프린팅과 같이 입력 음악과 정확히 일치하는 아카이브상의 음악을 찾는 경우도 있고,^[3,4] 장르 분류^[5,6] 및 유사음악 검색^[7-10]과 같이 특정한 성질을 공유하는 다수의 결과를 출력하는 경우도 있다. 본 논문에서는 특정한 성질을 공유하는 유사음악 검색에 대해서 다룬다. 본 논문에서는 특히 Fig. 1에 주어진 바와 같이 검색 대상 음악 파일들로부터 음악 신호를 처리하여 특정

[†]Corresponding author: Jin Soo Seo (jsseo@gwnu.ac.kr)
Department of Electronic Engineering Gangneung-Wonju National University, 7 Jukhun-gil, Gangneung, Gangwon 25457, Republic of Korea.
(Tel: 82-33-640-2428, Fax: 82-33-656-0740)

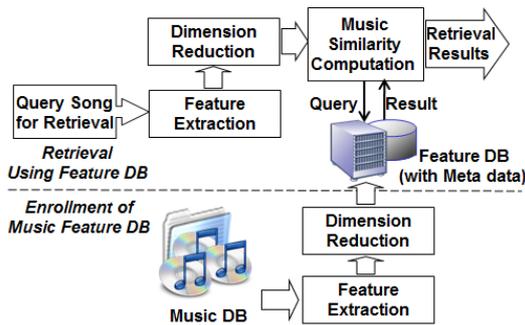


Fig. 1. Overview of the content-based music retrieval system.

을 추출하고 DB(Data Base)에 저장하고, 입력 음악에 대해서 DB 상의 특징과의 거리 비교를 통해서 음악의 유사도를 구해서 음악 검색에 활용하는 특징 기반 유사 음악 검색 시스템에 관한 연구이다.

유사도 기반 음악 검색의 어려움은 유사도 판정의 근거가 주관적이고 정량적으로 표현하기 어렵기 때문이다. 음악 유사도 판정을 위해서 음악의 템포, 리듬, 멜로디, 감정 등을 이용하는 시도도 있었으나, 아직 이러한 고수준의 음악 속성들에 대한 정확한 추정기가 어려우므로 아직까지는 상대적으로 분석이 쉬운 저수준의 스펙트럼 특징이 널리 사용되고 있다. 이러한 스펙트럼 특징들은 음악의 음색을 잘 표현하며, 스펙트럼 특징에는 스펙트럼 무게중심, 스펙트럼 롤 오프 주파수(spectral roll-off frequency), 밴드 에너지 비, 차분 스펙트럼 크기, 스펙트럼 선속(spectral flux) 등 다양하지만 MFCC(Mel-frequency Cepstral Coefficients) 특징이 가장 널리 사용되고 있다.^[5-10] 초기의 특징 기반 유사도 검색은 주로 어떤 노래의 음악 신호 전체에서 얻은 MFCC 벡터들을 하나의 GMM(Gaussian Mixture Model) 또는 k -means 군집을 통해서 모델링하였다. 이렇게 각 노래에서 얻은 특징 모델들간의 거리를 EMD(Earth-Mover Distance)^[7]를 통해서 구하였다. 음악 특징 군집화 및 군집간의 거리를 비교하는 방법은 군집을 구할 때 수렴성의 문제가 발생할 수 있고, 군집간의 거리를 구하는 것이 닫힌 해가 존재하지 않는 경우가 많고 일반적으로 계산량이 많은 단점이 있다. 따라서 이러한 단점을 보완한 방법으로 UBM(Universal Background Model)^[11]을 이용한 방법들이 최근 관심을 받고 있다.^[8-10] 일반적으로 UBM

은 미리 수집된 다양한 음악 신호로부터 특징벡터들을 얻고 이를 GMM을 통해서 모델링한 것이다. 특히 GMM을 UBM으로 이용하고 SV(Super Vector)^[12]개념을 적용한 방법들이 음악 검색에 성공적으로 적용되었다.^[8,9] 최근 GMM이 아닌 k -means 군집을 UBM으로 이용하는 무게중심 모델에 기반한 음악 검색 방법^[10]이 제안되었다. 기존 GMM을 이용하는 SV와 비교하여 계산량은 줄어들지만 성능에 있어서 큰 차이를 보이지 않았다.^[10] 본 논문은 최근에 제안된 무게중심 모델로부터 SV를 구하는 음악 검색 방법에 확률 분포간 거리를 적용하여 음악 유사도를 구하고 성능을 평가하였다. 일반적으로 SV간의 거리 비교는 유클리디안이나 코사인 거리 등의 벡터 거리가 주로 사용되어 왔다.^[8,9,12] 하지만 벡터 거리는 단순히 특징간의 차이를 구하는 것으로, 음악 신호와 같이 다수의 특징 벡터들의 군집으로 주어질 경우에는 특징의 분포간의 차이를 구하는 것이 더 타당하다. 특히 무게중심 모델을 이용할 경우 특징의 평균과 표준편차 정보를 구할 수 있고 이로부터 특징을 다변량 정규 분포로 모델링 할 수 있다.^[10] 즉 각 음악 파일을 단순한 SV가 아니라 정규 분포들의 군집으로 나타낼 수 있다. 두 음악 파일에서 각각 얻어진 정규 분포들간의 거리를 구하여 음악 유사도로 사용할 수 있다. 확률 분포간의 거리는 주어진 두 개의 확률 분포가 특정 기준에서 얼마나 가까운 지를 계산하는 것으로 다이버전스(divergence)라고 불리기도 한다.^[13-15] 본 논문에서는 무게중심 모델에서 확률 분포간의 거리 비교의 일반적인 형태인 알파 다이버전스를 적용하였다. 알파 다이버전스는 Renyi 다이버전스라고 불리기도 한다.^[16] 알파 다이버전스는 알파 값에 따라 여러 형태의 다이버전스를 나타낼 수 있으며, 널리 사용되고 있는 KLD(Kullback-Leibler Divergence)와 BD(Bhattacharyya Distance, Hellinger distance라고도 불림)를 특수 경우로 포함한다.^[17] 기존 연구^[10]에서도 대칭형 KLD를 사용한 결과가 있지만, 본 논문에서는 일반적인 형태의 알파 다이버전스를 무게중심 모델에 적용하여 그 장단점을 분석하였다. 장르와 가수 데이터셋에서 실험을 수행하여 각 거리 비교 방법들의 음악 검색 성능을 평가하였다.

본 논문은 무게중심 모델 특징의 확률 분포 거리

에 기반한 음악 검색에 관한 연구이다. II장에서 무계중심 모델을 이용한 음악 검색 방법 및 확률 분포 거리 기반 유사도에 대해서 살펴보고, 특히 알파 다이버전스를 무계중심 모델 기반 음악 유사도에 적용한다. III장에서 제안된 방법의 성능을 실험하고 결과를 비교 분석한다.

II. 알파 다이버전스를 이용한 무계중심 모델 기반 음악 유사도

기존의 GMM-UBM^[11]에 기반한 SV 방법^[12]은 음악 신호에서 얻어지는 특징들을 모아서 차수가 크긴 하지만 하나의 벡터로 만들어 줌으로써 검색에 쉽게 활용할 수 있는 장점이 있어 큰 관심을 받아왔다.^[8,9] 최근 SV 계산 시에 GMM의 계산량을 줄일 수 있는 *k*-means 군집화를 대신 이용하는 무계중심 모델에 기반한 음악 검색 방법이 제안되었다.^[10] 무계중심 모델은 화자 인증^[18]과 화자 변화 감지^[19] 등에도 성공적으로 적용되었다. 본 논문은 알파 다이버전스를 이용하여 무계중심 모델 기반 음악 검색의 성능을 향상시키는 것을 목표로 한다.

2.1 다이버전스를 이용한 확률 분포 간 거리

확률 분포 거리는 다이버전스라고 불리며,^[13-15] 통계적 추론을 통한 가설검증,^[20] 압축,^[21] 분류 및 인식^[22,23] 등의 다양한 문제들에 활용되어 왔다. 알파 다이버전스는 Renyi가 1961년에 제안한 것으로 Renyi 다이버전스라고도 불리며, 임의의 확률 분포 P 와 Q 간의 차수 알파(α)인 알파 다이버전스 D_α 는 다음과 같이 정의된다.^[13-17]

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1} \ln \int p(x)^\alpha q(x)^{1-\alpha} dx. \quad (1)$$

알파 다이버전스에서 차수 알파는 0과 1 사이의 값을 가져야 한다. 특히 알파 값이 0.5인 경우는 BD이다. 특히 BD는 P 와 Q 에 대해서 대칭적인 형태를 가지며, 신호 분류 및 인식 등에서 우수한 성능을 보였다.^[22] 또한 Eq.(1)의 알파 다이버전스 정의에서 알파 값을 1에 한없이 가깝게 하면, 다음과 같이 KLD

로 수렴한다.^[17]

$$\begin{aligned} \lim_{\alpha \rightarrow 1} D_\alpha(P\|Q) &= D_{KL}(P\|Q) \\ &= \int p(x) \ln \frac{p(x)}{q(x)} dx. \end{aligned} \quad (2)$$

위와 같이 알파 다이버전스는 확률 통계 및 패턴 인식 등에 널리 사용되고 있는 다양한 다이버전스들의 일반적인 형태이다.

본 논문에서는 무계중심 모델로부터 얻어지는 특징들의 평균과 표준편차를 이용하여 음악 신호를 다변량 정규 분포들의 집합으로 모델링한다. 다변량 정규 분포에 대해서 알파 다이버전스는 닫힌 해를 가지며, 평균과 공분산이 각각 (μ, Σ) 과 (μ_2, Σ_2) 로 주어진 두 다변량 정규 분포 N_1 과 N_2 사이의 알파 다이버전스는 $\Sigma = \alpha \Sigma_2 + (1-\alpha)\Sigma_1$ 라고 할 때 다음과 같다.^[14,15]

$$\begin{aligned} D_\alpha(N_1\|N_2) &= \frac{(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2)}{2} \\ &\quad - \frac{1}{2\alpha(\alpha-1)} \ln \frac{\det(\Sigma)}{\det(\Sigma_1)^{1-\alpha} \det(\Sigma_2)^\alpha}. \end{aligned} \quad (3)$$

참고로 다변량 정규 분포에 대한 KLD는 다음과 같다.

$$\begin{aligned} D_{KL}(N_1\|N_2) &= \frac{(\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2)}{2} \\ &\quad + \frac{1}{2} \left[\text{Tr}(\Sigma_2^{-1} \Sigma_1 - I) + \ln \frac{\det(\Sigma_2)}{\det(\Sigma_1)} \right]. \end{aligned} \quad (4)$$

위 Eq.(4)에서 $\text{Tr}()$ 은 행렬의 대각합을 의미한다. 무계중심 모델^[10]에서는 공분산이 대각행렬로 주어지므로, Eqs.(3)과 (4)에 주어진 알파 다이버전스 및 KLD의 계산이 용이하다. 본 논문에서는 알파 값을 변화시켜가면서 알파 다이버전스의 음악 유사도 성능을 확인하였다.

2.2 무계중심 모델

무계중심 모델 기반 음악 유사도^[10] 비교 방법은 *k*-means 군집의 각 클러스터의 중심점과 음악 신호의 매 프레임의 특징 벡터의 차이인 무계중심-편향

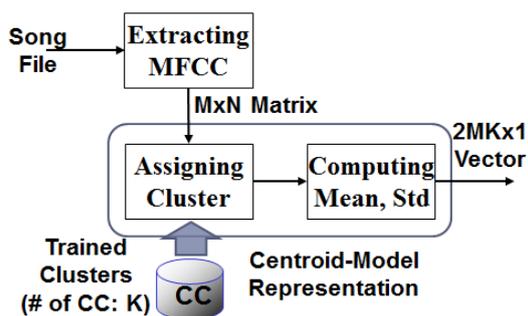


Fig. 2. Block diagram of the feature extraction using the centroid model.^[10]

벡터와 각 클러스터에 소속된 음악 특징 벡터의 개수를 이용한다. Fig. 2에 주어진 바와 같이 음악을 길이 L 인 프레임(본 논문에서는 23.2ms)으로 나누고, 각 프레임에서 M 차 MFCC 벡터를 구하였다(본 논문에서는 $M=19$). 미리 학습 음악 데이터들로부터 k -means 알고리즘을 이용하여 K 개의 클러스터의 중심점 $\mathbf{V}=(V_1, V_2, \dots, V_K)$ 를 구한다. 클러스터 중심점들인 \mathbf{V} 를 이용한 특징 표현 방법을 무계중심 모델^[10]이라고 하며, 본 논문의 SV 계산 시에 UBM으로 사용된다. 입력 음악 신호가 N 개의 프레임을 가지고 있다고 하고, 얻어진 MFCC 벡터들을 집합 $\mathbf{X}=(X_1, X_2, \dots, X_N)$ 로 표기한다. 음악 신호의 i 번째 프레임의 특징 벡터 X_i 가 소속된 클러스터의 인덱스 z_i 는 클러스터 중심점들인 \mathbf{V} 중에서 X_i 와 유클리디안 거리가 가장 가까운 점으로 다음과 같이 주어진다.

$$z_i = \arg \min_{1 \leq k \leq K} \|X_i - V_k\|^2. \quad (5)$$

또한 k 번째 클러스터에 속하는 MFCC 벡터들의 인덱스를 모아서 다음과 같이 클러스터 멤버십 집합 C_k 를 정의한다.

$$C_k = \{i | z_i = k, 1 \leq i \leq N\}. \quad (6)$$

무계중심-편향 벡터 r_i 는 X_i 와 소속 클러스터의 중심점과의 차이로 다음과 같다.

$$r_i = X_i - V_{z_i}. \quad (7)$$

각 클러스터의 중심점 V_k 는 특정 음색 그룹을 대표한다고 생각할 수 있고, 무계중심-편향 벡터 r_i 는 입력 음악 파일이 그 음색 그룹에서 떨어진 정도의 정보를 가지고 있다고 생각할 수 있다. 이 때 k 번째 클러스터에 속한 무계중심-편향 벡터들의 평균 μ_k 과 표준편차 σ_k 를 각각 다음과 같이 정의한다.

$$\mu_k = \frac{1}{|C_k|} \sum_{i \in C_k} r_i, \quad (8)$$

$$\sigma_k^2 = \frac{1}{|C_k|} \sum_{i \in C_k} (r_i - \mu_k)^2. \quad (9)$$

위 Eqs.(8)과 (9)에서 $|C|$ 는 집합 C 의 원소수이다. 기존 SV^[8,9] 방법은 GMM 평균의 변화량만을 고려하는 벡터적 특징이라면, 무계중심 모델은 특징 벡터의 평균과 표준편차 정보를 모두 이용하는 분포적 특징이다.

2.3 무계중심 모델 기반 음악 유사도

무계중심 모델에서는 각 클러스터의 무계중심-편향 벡터들의 평균과 표준편차를 이용하여 음악검색을 수행하며, 유사도를 비교하고자 하는 두 음악 신호를 A 와 B 라고 하고, A 와 B 의 k 번째 클러스터 멤버십 집합을 각각 $C_{A[k]}$ 와 $C_{B[k]}$ 라 하고, 무계중심-편향 벡터들의 평균을 각각 $\mu_{A[k]}$ 와 $\mu_{B[k]}$ 라 하고, 표준편차를 각각 $\sigma_{A[k]}$ 와 $\sigma_{B[k]}$ 라 하자. 무계중심 모델에서 음악 유사도 비교를 위한 각 클러스터의 가중치는 클러스터 멤버십 집합 원소 수를 비교하여 다음과 같이 정한다.

$$w_k = \frac{\min(|C_{A[k]}|/N_A, |C_{B[k]}|/N_B)}{\sum_{p=1}^K \min(|C_{A[p]}|/N_A, |C_{B[p]}|/N_B)}. \quad (10)$$

위 Eq.(10)에서 N_A 와 N_B 는 각각 음악 신호 A 와 B 의 프레임의 개수이다. 위 Eq.(10)과 같이 음악 신호의 각 프레임에서 얻어진 MFCC 벡터들이 각 클러스터에 속할 확률을 가중치로 사용하면 검색의 정확도를 더 높일 수 있다.

무계중심 모델을 이용한 음악 유사도 비교 방법으로 먼저 벡터 거리인 유클리디안 거리를 이용한 방법들을 살펴보겠다. 무계중심-편향 벡터들의 평균을 이용한 가중 유클리디안 거리 비교 $D_M(A, B)$ 는 다음과 같이 주어진다.

$$D_M(A, B) = \sum_{k=1}^K w_k \|\mu_{A[k]} - \mu_{B[k]}\|^2. \quad (11)$$

무계중심-편향 벡터들의 평균과 표준편차를 모두 이용한 가중 유클리디안 거리 비교 $D_{MS}(A, B)$ 는 다음과 같이 주어진다.

$$D_{MS}(A, B) = D_M(A, B) + \sum_{k=1}^K w_k \|\sigma_{A[k]} - \sigma_{B[k]}\|^2. \quad (12)$$

비록 D_{MS} 는 무계중심-편향 벡터들의 평균과 표준편차를 모두 활용하지만, 그 분포적 특성에 대한 고려 없이 단순히 차이를 구하므로 유사도 비교 성능 개선에 제약점이 있다. 계산량 저감 외에도 기존의 GMM-UBM 기반 SV^[8,9]와 비교하여 무계중심 모델^[10]을 이용할 경우의 장점은 무계중심-편향 벡터의 평균과 표준편차를 계산할 수 있다는 것이며, 이로부터 무계중심-편향 벡터가 공분산이 대각행렬인 다변량 정규분포를 따른다고 근사할 수 있다. 즉, 각 클러스터에 속하는 특징벡터들의 편향 정도를 다변량 정규분포로 모델링한 것이다. 이렇게 확률분포로 모델링할 경우 두 음악 신호 A 와 B 의 유사도를 비교할 때, Eqs.(11)과 (12)에서와 같이 단순한 벡터 거리가 아니라 확률 분포 거리인 Eqs.(3)과 (4)를 사용하는 것이 타당하다. 두 음악 A 와 B 의 k 번째 클러스터의 정규분포가 각각 $N_{A[k]}$ 와 $N_{B[k]}$ 라고 하면, 알파 다이버전스를 이용한 음악 거리 비교인 $D_\alpha(A, B)$ 는 다음과 같이 주어진다.

$$D_\alpha(A, B) = \sum_{k=1}^K w_k D_\alpha(N_{A[k]}, N_{B[k]}). \quad (13)$$

Eq.(13)에서 $N_{A[k]}$ 는 평균이 $\mu_{A[k]}$ 이고 표준편차가 $\sigma_{A[k]}$

인 정규분포이고, $N_{B[k]}$ 는 평균이 $\mu_{B[k]}$ 이고 표준편차가 $\sigma_{B[k]}$ 인 정규분포이다. 기존의 무계중심 모델^[10]에서도 확률 분포 거리인 대칭 KLD를 유사도 비교에 활용하여 의미 있는 성능 개선을 보였다. 본 논문에서는 무계중심 모델에 비대칭 KLD와 BD를 포함하는 일반적인 형태의 확률분포 간 거리 비교 방법인 알파 다이버전스를 적용하여 무계중심 모델 기반 음악 검색의 성능을 개선하고자 한다.

III. 실험 결과

본 장에서는 II장에서 살펴본 무계중심 모델을 이용한 음악 검색기에 제안한 바와 같이 알파 다이버전스를 적용하고, 기존의 음악 검색 방법들과 성능을 비교하였다. 음악 검색의 성능을 비교하는 것은 음악 유사도에 대한 검증된 기준값이 존재하지 않으므로 상당히 어려운 문제이다. 또한 음악 유사도는 주관적인 면이 많고 수치화하기 어려우므로,^[24] 대부분의 선행 음악 검색 연구들에서는 같은 장르 또는 가수의 음악들이 다른 장르 또는 가수의 음악들에 비해서 서로 인간 지각적으로 유사하다는 가정에 바탕을 두고 음악 검색 성능을 평가한다.^[5,6,8,9] 본 논문에서도 장르와 가수 데이터셋 두 가지 데이터셋을 사용하여 성능을 실험하였다. 본 논문에서도 같은 가정을 통해서 장르와 가수 기준으로 제안된 음악 검색 방법의 정확도가 얼마나 높을 지를 구해 음악 검색 성능 지표로 사용한다. 선행 연구와 같은 방식의 성능 실험을 함으로써, 선행 연구들의 결과와도 직접적으로 비교할 수 있는 장점도 있다. 음악 장르 데이터셋인 GTZAN 데이터셋^[5]은 블루스, 클래식, 컨츄리, 디스코, 힙합, 재즈, 메탈, 팝, 레게, 락의 10개의 장르에 각각 100곡씩 30s 길이의 1000개의 음악파일로 이루어져있다. 가수 데이터셋^[10]은 남녀 각각 17명씩 총 34명의 가수 별로 20곡씩 680곡으로 이루어져있다. 데이터셋의 음악들로 특징 DB를 만들고, 데이터셋의 각 음악을 질의 음악으로 만들어진 특징 DB에 대해 음악 검색을 수행하였다.

실험에 사용되는 음악 파일들을 모노로 바꾸고 22050 Hz로 샘플링 주파수를 맞춘 후, 512 길이(23.2 ms)의 해닝(Hanning) 윈도우를 50 %씩 겹쳐가면서

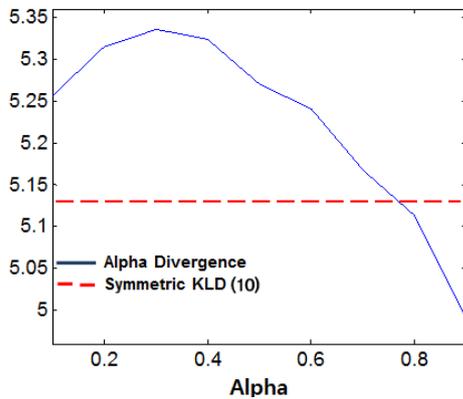


Fig. 3. Average number of correctly retrieved songs among the closest 10 songs with the criterion of the same genre versus alpha of $D_{\alpha}(A, B)$ on the genre dataset with 1000 songs.

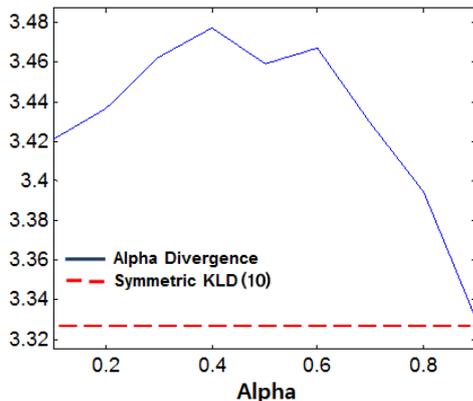


Fig. 4. Average number of correctly retrieved songs among the closest 10 songs with the criterion of the same singer versus alpha of $D_{\alpha}(A, B)$ on the singer dataset with 680 songs.

적용하고 FFT(Fast Fourier Transform)를 가한다. 이렇게 주파수 도메인으로 신호를 변환해서 얻은 각 프레임의 스펙트럼으로부터 19차 MFCC를 계산하였다. 본 실험의 무계중심 모델에 사용되는 k -means 클러스터는 다양한 장르의 156곡에서 얻어진 MFCC 벡터들로부터 학습되었으며, 클러스터의 개수인 K 값은 12를 사용하였다. 학습된 클러스터의 중심점들을 이용하여, 각 실험대상 음악을 Eqs.(8)과 (9)에 주어진 바와 같이 평균과 표준편차를 각각 구한다. 구해진 평균과 표준편차를 이용하여 Eq.(13)의 알파 다이버전스에서 알파 값을 변화시켜가면서 음악 검색을 수행하였다. 본 실험 결과에서는 Eqs.(11)~(13)에서 A 를 입력 질의 음악이라 가정하고, B 를 검색 대상

DB상의 음악이라 가정한다.

장르 데이터셋에 대한 음악 검색 실험결과는 Fig. 3에 주어져 있다. 알파 값을 변화시켜가면서, 검색 결과 중 가까운 10곡 (closest 10)에서 질의 음악과 장르가 일치하는 결과의 개수를 그렸다. 기존 무계중심 모델 논문^[10]에서 가장 좋은 성능을 보였던 대칭형 KLD에 비해서 알파 다이버전스의 성능이 더 우수함을 알 수 있다. 알파값이 증가함에 따라 완만히 검색 성능이 향상되다가 알파값이 0.6을 넘어서면 성능의 큰 감소가 관찰된다. Fig. 3의 결과를 분석하기 위해서 장르 데이터셋에 있는 1000곡의 음악 각각에 대해서 알파값을 0.1에서 0.9 까지 변화시키면서 검색 성공률의 변화 양상을 확인하여, 1000곡의 음악을 다음의 5가지 형태로 분류하였다.

- 비감소형: 알파값의 증가에 따라 검색 성공률이 감소하지 않음(즉, 검색 성공률이 알파값에 대해 비감소 함수로 증가 또는 값을 유지, 단 알파값에 대해 상수함수인 경우는 제외).
- 비증가형: 알파값의 증가에 따라 검색 성공률이 증가하지 않음(즉, 검색 성공률이 알파값에 대해 비증가 함수로 감소 또는 값을 유지, 단 알파값에 대해 상수함수인 경우는 제외).
- 무변화형: 알파값 변화에 대해서 검색 성공률의 차이가 없음(즉, 검색 성공률이 알파값에 대해 상수함수임).
- 최대형: 양 끝단이 아닌 알파값에 대해서 검색 성공률이 최대인 알파값이 존재함.
- 기타형: 알파값에 대해서 증가와 감소가 혼재함.

실험 대상 1000곡의 음악들 중 비감소형이 198곡, 비증가형 277곡, 무변화형 144곡, 최대형 261곡, 기타형 120곡이었다. 최대형의 음악들에서 검색 결과가 최대가 되는 알파값은 골고루 분포하였다. 이러한 결과를 토대로 분석해보면 Fig. 3의 형태가 각 1000곡의 특성들이 평균화되어서 나타난 것으로 분석 가능하다. 즉, 알파값에 따라서 음악 별로 다양한 형태로 성능이 변화하므로, 중간 정도의 알파값에서 평균적으로 가장 우수한 성능을 보임을 알 수 있다. 음악별로 위 5가지 형태 중 어느 곳에 속하는 지 미리 예측

할 수 있다면, 질의 음악 별로 최적의 알파값을 산정할 수 있게 되어서 추가적인 검색 성능 향상도 기대해 볼 수 있을 것으로 생각되지만, 본 실험에서는 질의 음악의 특징과 최적 알파값 사이에 큰 상관관계를 발견하지 못했다. 질의 음악 특성뿐만 아니라 검색 대상 음악들의 구성과도 연관된 문제라 예측이 어려운 것으로 생각된다.

가수 데이터셋에 대해서도 같은 실험을 수행하여, 역시 음악 검색 결과 중 가까운 10곡(closest 10)에 대한 가수 일치도를 Fig. 4에 나타내었다. 장르 데이터셋의 결과인 Fig. 3과 유사한 결과를 보였고, 다이버전스 계산에서 알파 값을 0.2에서 0.6 사이로 하는 것

이 음악 검색 성능에 도움이 됨을 확인하였다. 역시 가수 데이터셋의 680곡의 음악들을 5가지 유형으로 분류하였다. 비감소형이 151곡, 비증가형 175곡, 무변화형 173곡, 최대형 118곡, 기타형 63곡이었다. Fig. 4도 Fig. 3과 마찬가지로 여러 유형의 음악들의 성능이 결합되어 중간 정도의 알파값에서 평균적으로 가장 우수한 성능을 보임을 알 수 있다. 서로 겹치지 않는 두 데이터셋에서 서로 다른 기준으로 검색 결과를 평가했을 때, 알파값에 따라서 유사하게 검색 성공률이 변화하는 것을 Figs. 3과 4를 통해서 관찰할 수 있으며, 이를 토대로 본 논문의 결과가 실제 음악 서비스의 대규모 데이터셋에 적용했을 때도 유사하

Table 1. Average number of closest songs correctly retrieved with the criterion of the same genre using the GTZAN genre dataset with 1000 songs. A query song denoted by A , and a DB song is denoted by B .

Types of Features	Type of Distance	Distance Measure	Average number of correctly-retrieved songs		
			Closest 5	Closest 10	Closest 20
Centroid-Model	Euclidean	$D_M(A, B)$	2.634	4.797	8.344
		$D_{MS}(A, B)$	2.844	5.165	8.972
	KL divergence (in case of $\alpha = 1$)	$D_{KL}(A, B)$	2.711	4.797	8.280
		$D_{KL}(B, A)$	2.838	5.095	8.780
		$D_{KL}(A, B) + D_{KL}(B, A)$	2.834	5.131	8.998
	Alpha divergence	$D_\alpha(A, B)$ (in case of $\alpha = 0.3$)	2.974	5.336	9.291
$D_\alpha(A, B)$ (in case of $\alpha = 0.4$)		2.969	5.324	9.286	
$D_\alpha(A, B)$ (in case of $\alpha = 0.5$)		2.946	5.271	9.219	
$SV^{[9]}$	Euclidean	Euclidean	2.618	4.661	8.072
	Cosine	UBM-normalized Cosine	2.866	5.167	9.024
Logan's Method ^[7]	EMD	EMD	2.743	4.801	8.384
Random Selection			0.5	1.0	2.0

Table 2. Average number of closest songs correctly retrieved with the criterion of the same singer using a singer dataset with 680 songs. A query song denoted by A , and a DB song is denoted by B .

Types of Features	Type of Distance	Distance Measure	Average number of correctly-retrieved songs		
			Closest 5	Closest 10	Closest 20
Centroid-Model	Euclidean	$D_M(A, B)$	1.922	3.134	4.671
		$D_{MS}(A, B)$	1.979	3.222	4.753
	KL divergence (in case of $\alpha = 1$)	$D_{KL}(A, B)$	2.021	3.223	4.740
		$D_{KL}(B, A)$	2.073	3.278	4.807
		$D_{KL}(A, B) + D_{KL}(B, A)$	2.087	3.325	4.860
	Alpha divergence	$D_\alpha(A, B)$ (in case of $\alpha = 0.3$)	2.160	3.462	5.102
$D_\alpha(A, B)$ (in case of $\alpha = 0.4$)		2.168	3.477	5.102	
$D_\alpha(A, B)$ (in case of $\alpha = 0.5$)		2.161	3.459	5.120	
$SV^{[9]}$ (UBM-Normalized)	Euclidean	Euclidean	1.834	2.954	4.386
	Cosine	UBM-normalized Cosine	2.073	3.340	4.970
Logan's Method ^[7]	EMD	EMD	1.743	2.776	4.044
Random Selection			0.147	0.294	0.588

계 동작할 것을 기대할 수 있다. 또한 Figs. 3과 4의 결과를 보면 장르와 가수 데이터셋 둘 다에서 0.2와 0.6 사이 구간에서 안정적인 성능을 보이므로 실제 적용 시에 알파값을 정하는 것이 용이함을 알 수 있다.

기존의 음악 유사도 비교 방법들을 이용한 검색 성능과 다이버전스를 이용한 검색 성능을 Tables 1과 2에서 정리하였다. 음악 검색 결과 질의 음악과 가장 가까운 5곡, 10곡, 20곡 중에서 질의 음악과 같은 장르 또는 가수의 음악이 몇 곡이나 포함되어 있는지를 음악 검색의 정확도로 사용하였다. 벡터거리를 이용하는 방법들에 대비하여, 다이버전스를 이용하는 방법들이 더욱 우수한 성능을 보임을 알 수 있다. 또한 알파 다이버전스를 이용하면 벡터 거리에 비해서 10% 정도, KLD에 비해서 5% 내외로 검색 성능을 향상시킬 수 있음을 알 수 있다. 기존 방법들 중 SV⁹⁾와 비교하면, 정규화하지 않은 SV에 대해서는 10% 이상, UBM-표준화한 SV에 대해서도 3% 정도 성능 향상을 보인다. UBM-표준화는 얻어진 SV와 UBM-SV와의 차이를 이용하는 방법으로, 본 논문에서 다룬 무계중심 모델에서 Eq.(7)의 무계중심-편향 벡터를 구하는 과정은 기존 SV의 성능을 개선하기 위해 제안된 표준화 과정인 SV와 UBM-SV간의 차를 구하는 방법⁹⁾과 개념적으로 유사점이 있다. 이와 함께 무계중심모델에서는 기존의 SV에서 고려하지 않은 분산 정보를 추가로 활용했기 때문에 성능을 향상시킨 것으로 생각된다. 실험 결과로부터 같은 확률분포간 거리인 KLD와 비교해서도 본 논문에서 제안한 알파 다이버전스가 무계중심모델 기반 음악 검색 성능에 더욱 효과적임을 알 수 있다.

IV. 결 론

본 논문에서는 무계중심 모델을 이용한 음악 검색에 알파 다이버전스에 기반한 유사도를 적용하고 기존 방법들과 성능을 비교 분석하였다. 무계중심 모델에서는 무계중심-편향 벡터의 평균과 표준편차를 구할 수 있으므로, 기존의 SV와는 다르게 단순한 벡터 거리 비교 방식 보다는 확률 분포 거리를 활용하는 것이 타당하다. 특히 본 논문에서는 확률 분포 거리들 중에서 일반적인 형태를 가지는 알파 다이버전

스를 적용했다. 알파 다이버전스는 알파 값에 따라서 KLD와 BD 등의 다양한 거리 비교 방법을 특수 경우로 포함하며, 다변량 정규 분포 간의 다이버전스 계산 시에 닫힌 해가 존재한다. 장르와 가수 데이터셋 상의 음악 검색 실험을 통해서 검색 성능을 구하였으며, 검색 성능을 최대화 하는 알파값을 찾았다. 음악 검색을 위한 무계중심 모델간 거리 비교에서 확률 분포 거리인 알파 다이버전스를 이용하여 성능 개선을 이룰 수 있음을 확인하였다.

감사의 글

본 연구는 문화체육관광부 및 한국저작권위원회의 2015년도 저작권기술개발사업의 연구결과로 수행되었음(2015-related-9500, 연관 저작물 검색과 저작권 침해 예방을 위한 예측 탐지 기술 개발).

References

1. Z. Fu, G. Lu, K.M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," *IEEE Trans. Multimedia* **13**, 303-319 (2011).
2. M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: current directions and future challenges," *Proceedings of the IEEE* **96**, 668-696 (2008).
3. J. Seo, "A robust audio fingerprinting method based on segmentation boundaries" (in Korean) *J. Acoust. Soc. Kr.* **31**, 260-265 (2012).
4. C. Park, M. Park, S. Kim, and H. Kim, "Music identification using pitch histogram and MFCC-VQ dynamic pattern" (in Korean), *J. Acoust. Soc. Kr.* **24**, 178-185 (2005).
5. G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech and Audio Process.* **10**, 293-302 (2002).
6. J. Seo, "A musical genre classification method based on the octave-band order statistics" (in Korean), *J. Acoust. Soc. Kr.* **33**, 81-86 (2014).
7. B. Logan and A. Salomon, "A music similarity function based on signal analysis," in *Proc. ICME-2001*, 745-748 (2001).
8. C. Cao and M. Li, "Thinkit's submissions for MIREX2009 audio music classification and similarity tasks," in *Mirex abstracts of ISMIR-2009*, (2009).
9. C. Charbuillet, D. Tardieu, and G. Peeters, "GMMsupervector for content based music similarity," in *Proc. DAFx-11*,

- 425-428 (2011).
10. J. Seo, "A music similarity function based on the centroid model," *IEICE Trans. Information and Systems* **96**, 1573-1576 (2013).
 11. D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing* **10**, 19-41 (2000).
 12. W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters* **13**, 308-311 (2006).
 13. F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Trans. Information Theory* **52**, 4394-4412 (2006).
 14. L. Pardo, "Statistical inference based on divergence measures," CRC press 2005.
 15. M. Gil, F. Alajaji, and T. Linder, "Renyi divergence measures for commonly used univariate continuous distributions," *Information Sciences* **249**, 124-131 (2013).
 16. A. Renyi, "On measures of entropy and information," in *Proc. Berkeley Symp. Probability Theory and Mathematical Statist.*, 547-561 (1961).
 17. T. V. Erven and P. Harremoës, "Renyi divergence and Kullback-Leibler divergence," *IEEE Trans. Information Theory* **60**, 3797-3820 (2014).
 18. V. Hautamaki, T. Kinnunen, I. Karkkainen, J. Saastamoinen, M. Tuononen, and P. Franti, "Maximum a posteriori adaptation of the centroid model for speaker verification," *IEEE Signal Process. Letters* **15**, 162-165 (2008).
 19. J. Seo, "A speaker change detection method based on a weighted distance measure over the centroid model," *IEICE Trans. Information and Systems* **95**, 1543-1546 (2012).
 20. F. Alajaji, P. N. Chen, and Z. Rached, "Csiszar's cutoff rates for the general hypothesis testing problem," *IEEE Trans. Information Theory* **50**, 663-678 (2004).
 21. P. Harremoës, "Interpretations of Renyi entropies and divergences," *Physica A: Statistical Mechanics and its Applications* **365**, 57-62 (2006).
 22. A. O. Hero, B. Ma, O. Michel, and J. Gorman, "Alpha-divergence for classification, indexing and retrieval," Tech. rep., University of Michigan, (2001).
 23. H. G. Kim, and D. Shin, "Speaker verification using SVM kernel with GMM-supervector based on the Mahalanobis distance" (in Korean), *J. Acoust. Soc. Kr.* **29**, 216-221 (2010).
 24. J. Lee, "How similar is too similar?: Exploring users' perceptions of similarity in playlist evaluation," in *Proc. ISMIR*, 109-114 (2011).

저자 약력

▶ 서진수 (Jin Soo Seo)



1998년 2월: KAIST 전기 및 전자공학과
공학사
2000년 2월: KAIST 전기 및 전자공학과
공학석사
2005년 2월: KAIST 전기 및 전자공학과
공학박사
2006년 3월 ~ 2008년 2월: 한국전자통신
연구원 선임연구원
2008년 3월 ~ 현재: 강릉원주대학교 전자
공학과 조교수, 부교수

▶ 김정현 (Jeonghyun Kim)



1999년 2월: 전남대학교 전산학과 공학사
2001년 2월: 전남대학교 전산학과 공학석사
2001년 3월 ~ 현재: 한국전자통신연구원
선임연구원

▶ 박지현 (Jihyun Park)



1997년 2월: 서강대학교 전산학과 공학사
1999년 2월: 서강대학교 컴퓨터학과 공학
석사
2010년 2월: 충남대학교 컴퓨터공학과 공학
박사
1999년 5월 ~ 현재: 한국전자통신연구원
책임연구원