

커버곡 검색 정확도 향상을 위한 적합도 기반 크로마그램 쌍별 유사도

A relevance-based pairwise chromagram similarity for improving cover song retrieval accuracy

서진수[†]

(Jin Soo Seo^{1†})

¹강릉원주대학교 전자공학과

(Received October 11, 2023; accepted January 9, 2024)

초록: 음악 유사도 계산은 음악 검색 서비스 구현에서 중요한 구성 요소 중 하나이다. 본 논문은 커버곡 검색 성능을 제고하기 위해서, 크로마그램 벡터 별로 커버곡 검색 적합도를 구하여 음악 유사도 계산 시 가중치로 활용한다. 커버곡 검색 적합도는 확률 적합도 모델을 이용하여 구한다. 커버곡 검색에 도움이 될 수 있는 분별도가 높은 벡터에 높은 가중치를 부여하고, 흔하게 존재하여 분별도가 떨어지는 벡터에 낮은 가중치를 부여하는 방식으로 음악 유사도 함수를 유도한다. 두 가지 커버곡 실험 데이터셋에서 성능 비교를 수행하여 제안한 음악 유사도 함수가 커버곡 검색 성능을 개선시킬 수 있음을 보였다.

핵심용어: 커버곡 검색, 음악 유사도, 확률 적합도 모델, 검색 가중치

ABSTRACT: Computing music similarity is an indispensable component in developing music search service. This paper proposes a relevance weight of each chromagram vector for cover song identification in computing a music similarity function in order to boost identification accuracy. We derive a music similarity function using the relevance weight based on the probabilistic relevance model, where higher relevance weights are assigned to less frequently-occurring discriminant chromagram vectors while lower weights to more frequently-occurring ones. Experimental results performed on two cover music datasets show that the proposed music similarity improves the cover song identification performance.

Keywords: Cover song identification, Music similarity, Probabilistic relevance model, Search weight

PACS numbers: 43.75.Zz, 43.60.Uv

1. 서론

컴퓨터 및 데이터 저장장치, 통신 네트워크, 정보 처리 기기의 발달에 따라서 온라인 음악 유통 시장의 규모가 커지고 있다. 따라서 대규모 음원 데이터베이스에서 사용자의 요구에 맞추어 음악을 빠르고 신뢰성 있게 찾아서 제공해 줄 수 있는 검색 기술의 필요성이 커지고 있으며, 관련 연구들이 활발히 수

행되어 왔다. 본 논문은 다양한 음악 검색 문제 중에서 커버곡 검색의 성능을 높이기 위한 음악 유사도 비교 방법에 관해서 다룬다. 커버곡은 콘서트 현장에서 라이브 녹음, 편집이나 리메이크 등을 통해서 재녹음된 음악을 가리킨다.^[1-3] 커버곡 검색 기술은 웹하드 및 유튜브 등 데이터 공유 서비스에서 저작권 보호, 중복된 음원을 가진 음악 아카이브 정리 등에 활용될 수 있을 것으로 기대된다.

[†]Corresponding author: Jin Soo Seo (jsseo@gwnu.ac.kr)

Department of Electronic Engineering, Gangneung-Wonju National University, 7 Jukhun-gil, Gangneung, Gangwon-Do 25457, Republic of Korea

(Tel: 82-33-640-2428, Fax: 82-33-656-0740)



Copyright©2024 The Acoustical Society of Korea. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

음악 검색은 목적에 따라서 다양한 유사도 기준을 적용하게 되어 사용하는 특징과 음악 유사도 비교 방법이 달라진다. 다양한 음악 검색 문제 중에서 본 논문은 커버곡 검색을 다루며, 특히 커버곡 검색을 위한 음악 유사도 비교 방법에 관해서 연구한다. 커버곡 검색을 위한 음악 유사도 함수를 찾기 위해서는 원곡과 커버곡 간의 공유되는 특성을 찾아야 한다. 커버곡을 만드는 편집이나 녹음 과정에서 가수와 악기의 차이로 인한 음색 변화, 연주 속도 및 스타일 차이로 인한 템포, 리듬, 음악 키 변조 등 다양한 종류의 변형이 발생하므로, 커버곡 검색을 위한 음악 특징을 찾고 유사도 함수를 정의하는 것은 여전히 어려운 문제로 남아있다.

현재 가장 널리 사용되고 있는 원곡과 커버곡 간의 공통점은 음들의 시간적 연결을 의미하는 선율(멜로디)이다. 선율을 표현하는 스펙트럼 특징에는 대표적으로 크로마그램이 있다. 크로마그램은 음악 신호를 짧은 길이의 프레임 단위로 나누고 각 프레임에서 스펙트럼을 구하고 옥타브 차이나는 음악 스펙트럼 성분들을 가산하여 음악의 전체 스펙트럼 성분들을 하나의 옥타브 안으로 접어서 표현한 것이다. 일반적으로 커버곡 검색은 전곡 단위 입력에 대해서 이루어지며, 음악 신호로부터 얻어지는 특징인 크로마그램 수열을 직접 비교하여 시간축 상에서 정합하는 수열 직접 비교 방법^[2]과 특징 벡터 수열을 가공하여 검색에 용이한 고정된 길이의 전곡 특징^[4]을 구하는 방법으로 나눌 수 있다. 본 논문은 커버곡 검색 방법들 중에서 음악 신호로부터 얻어지는 크로마그램 수열을 쌍별로 직접 비교하여 시간축 상에서 정합을 구하는 수열 직접 비교 방법에 대해서 다룬다. 이런 수열 직접 비교 방법은 음성 인식과 deoxyribonucleic acid(DNA) 수열 분석에서 사용해왔던 dynamic time warping^[5]이나 Smith-Waterman(SW) 알고리즘^[5] 등을 활용한다. 수열 직접 비교에서 가장 핵심이 되는 부분은 쌍별 크로마그램 유사도를 정의하는 것인데, 기존 연구들에서는 쌍별 크로마그램 유사도를 정합과 부정합 시에 각각 양과 음의 상수로 미리 정해진 고정값을 사용하였다.^[2] 최근 커버곡 검색 성능을 개선하기 위해서 개별 크로마그램 벡터의 허브 지수를 고려하여 쌍별 유사도를 정규화하는 방법이 제안

되었고, 커버곡 검색 성능을 높일 수 있음을 보였다.^[6] 하지만 허브 지수가 임의로 정의되어 있어서 실험적으로는 좋은 성능을 보였으나, 이론적인 근거가 부족한 단점이 있었다. 본 논문에서는 확률 적합도 모델^[7-9]에 기반하여 각 크로마그램 벡터의 커버곡 검색 적합도를 구하고, 크로마그램 쌍별 유사도가중치로 사용하는 방법을 제안한다. 또한 본 논문에서 제안한 확률 적합도 기반 가중치와 허브 지수 기반 가중치^[6]를 특수 경우로 포함함을 보였다. 쌍별 크로마그램 유사도에 기존 고정값을 사용하는 방법과 가변 가중치를 사용하는 제안한 방법의 커버곡 검색 성능을 실험을 통해서 비교하였다.

본 논문은 커버곡 검색을 위한 크로마그램 쌍별 유사도 함수에 관한 연구이다. II장에서 크로마그램 특징 추출 및 쌍별 비교 방법을 살펴보고, 적합도를 기반으로 유사도 함수를 제안한다. III장에서 제안된 방법의 성능을 실험하고 결과를 비교 분석한다.

II. 커버곡 검색 적합도 기반 크로마그램 쌍별 유사도

크로마그램 쌍별 유사도를 이용한 커버곡 검색 방법으로 SW에 기반한 수열 직접 비교 방법을 살펴보고, 쌍별 유사도를 고정된 값으로 사용하는 기존 방법을 개선하기 위해서 쌍별 유사도 가중치를 확률 적합도 모델에 기반하여 유도하는 방법을 제안한다.

2.1 크로마그램 쌍별 유사도를 이용한 SW 기반 커버곡 검색

커버곡 검색은 음악 신호 전체에서 얻은 특징에 대해서 이루어지며, 수열 직접 비교 방법은 Fig. 1에 도시한 바와 같이 특징 벡터 수열 간의 쌍별 유사도

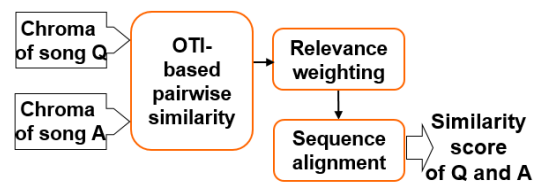


Fig. 1. (Color available online) Overview of the music similarity computation for cover song identification based on the proposed relevance weight.

를 구하고 SW를 통해서 최적 정합 경로를 찾는다.^[2] 커버곡 유무를 판단하고자 하는 두 음악으로부터 추출된 크로마그램 벡터 수열을 추출하고, 각 크로마그램 벡터 간 비교를 통해서 크로마그램 쌍별 유사도를 구한다.

커버곡 생성 과정에서 다양한 변형을 겪게 되고, 특히 음악의 조변화가 있을 경우 크로마그램 쌍별 유사도를 직접 유클리디안 또는 코사인 거리 등으로 비교하는 것이 불가능하다. 음악 조변화는 크로마그램 상에서 크로마추 방향으로 순환 이동으로 나타나기 때문이다. 이러한 특성을 반영하여 전곡 크로마그램의 조변화도를 구하고, 각 프레임의 크로마그램 비교 시에 음악의 조변화도가 전곡과 일치하면 정합으로, 일치하지 않으면 부정합으로 판정하는 Optimal Transposition Index(OTI)를 이용한 방법이 좋은 성능을 보였다.^[2] OTI는 음악의 조변화에 대해서 불변성을 가지므로 OTI 기반 쌍별 유사도도 조변화에 불변성을 가지게 된다. OTI를 이용한 쌍별 유사도는 다음과 같이 구할 수 있다. Fig. 1에서 커버곡을 찾고자하는 질의 음악을 Q, 음원 데이터베이스상의 검색 대상 음악을 A라고 하자. 두 음악 Q, A의 프레임 개수가 각각 M과 N일 때, 크로마그램 수열을 얻어진 시간 순으로 각각 Q, A로 표기하면, 다음과 같이 q_m 과 a_n 은 L차 크로마그램 벡터가 된다(일반적으로 L = 12).

$$\begin{aligned} Q &= \{q_1, q_2, \dots, q_M\}, \\ A &= \{a_1, a_2, \dots, a_N\}. \end{aligned} \quad (1)$$

음악이 조변화되면 크로마그램 벡터는 순환 이동하게 된다. 벡터 q_m 이 주어졌을 때, 벡터 간 유사도가 최대가 되도록 a_n 이 순환 이동해야하는 빈의 개수인 $OTI_{m,n}$ 은 벡터간 내적 \langle, \rangle 으로 다음과 같이 정의된다.

$$OTI_{m,n} = \arg\max_l \langle q_m, cshift(a_n, l) \rangle. \quad (2)$$

Eq. (2)에서 $cshift(a_n, l)$ 는 a_n 벡터를 l만큼 순환 이동한 벡터를 가리키며, l을 0에서 L-1까지 가변하면서 최대가 되는 l값을 찾는다. 전곡 평균 크로마그램 벡터를 각각 g_Q 와 g_A 라고 하면, OTI_g 는 다음과 같이 정의된다.

$$OTI_g = \arg\max_l \langle g_Q, cshift(g_A, l) \rangle. \quad (3)$$

$OTI_{m,n}$ 과 OTI_g 를 비교하여, Q와 A의 프레임 특징 벡터 간 쌍별 유사도 행렬 S를 다음과 같이 얻는다.

$$S_{m,n} = \begin{cases} \mu_+ & \text{if } OTI_{m,n} = OTI_g \\ \mu_- & \text{otherwise} \end{cases}. \quad (4)$$

Eq. (4)에서 μ_+ 와 μ_- 는 각각 정합과 부정합을 나타내는 상수로 논문^[2]과 같이 $\mu_+ = 1, \mu_- = -0.9$ 를 사용하였다.

쌍별 유사도는 프레임 특징 간 유사도 이므로 입력 음악 간의 관계가 커버곡인지 여부를 판단하기 위해서는 전곡 유사도를 구해야 한다. 쌍별 유사도 상에서 SW를 통해서 최적 정합 경로를 찾아서 전곡 유사도를 유도한다.^[2] SW를 통해서 최적 정합 경로를 찾게 되면, 커버곡을 만드는 과정에서 자주 발생하는 음악 연주 속도 변화에 강인하게 되는 장점이 있다. SW 알고리즘은 쌍별 유사도 행렬 S의 시간축 방향 연속성을 고려하여, 다음과 같이 시간축 정합 행렬 H를 구한다.

$$H_{m,n} = \max \begin{cases} H_{m-1, n-1} + S_p - \delta(S_{m-2, n-2}, S_p) \\ H_{m-2, n-1} + S_p - \delta(S_{m-3, n-2}, S_p) \\ H_{m-1, n-2} + S_p - \delta(S_{m-2, n-3}, S_p) \\ 0 \end{cases}. \quad (5)$$

Eq. (5)에서 $S_p = S_{m-1, n-1}$ 이며, 정합 제약 조건 함수 $\delta()$ 는 다음과 같이 주어진다.

$$\delta(p, q) = \begin{cases} 0 & \text{if } q > 0 \\ 0.5 & \text{if } q \leq 0, \text{ and } p > 0 \\ 0.7 & \text{if } q \leq 0, \text{ and } p \leq q \end{cases}. \quad (6)$$

정합값 $H_{m,n}$ 은 Q수열의 m, A 수열의 n번 위치까지의 시간축 정합의 최대값을 의미한다. 커버곡 검색을 위해서 수열 Q와 A 간의 거리는 H의 최대값의 역수와 두 수열의 길이에 따라 다음과 같이 구한다.

$$d(A, B) = \frac{M+N}{\max(H)}. \quad (7)$$

2.2 적합도 기반 크로마그램 쌍별 유사도

OPI 기반 쌍별 유사도 계산에서, 질의 음악의 크로마그램 벡터 q_m 별로 OPI 기준으로 정합되는 상대 크로마그램 벡터의 개수에 차이가 있다. 많은 수의 크로마그램 벡터와 정합이 되는 q_m 은 그만큼 변별력이 떨어진다고 볼 수 있으며, 기존 연구^[6]에서는 이를 허브 지수로 표현하고 허브 지수를 기반으로 Eq. (4)의 쌍별 유사도를 정규화 하였다. 본 논문에서는 확률 적합도 모델^[7]을 기반으로 q_m 의 커버곡 인식 적합도를 구하여 쌍별 유사도의 가중치로 사용한다. 적합도를 구하기 위해서 q_m 에 대해서 두 가지 확률 λ_m 과 β_m 을 다음과 같이 정의한다.

$$\lambda_m = P(A \ni q_m | R_{AQ} = 1),$$

$$\beta_m = P(A \ni q_m | R_{AQ} = 0).$$

먼저 λ_m 은 노래 A가 Q의 커버곡일 때 ($R_{AQ} = 1$), A의 특징 수열 A가 q_m 을 포함하고 있을 확률이다. 다음으로 β_m 은 노래 A가 Q의 커버곡이 아닐 때 ($R_{AQ} = 0$), A가 q_m 을 포함하고 있을 확률이다.

두 확률 λ_m 과 β_m 을 이용하여 q_m 과 정합되는 것이 A가 Q의 커버곡인 지 여부를 판단하는 데 유용한 정도인 적합도를 구한다. 확률 적합도 모델에 따르면, q_m 의 커버곡 인식에 대한 적합도는 q_m 존재 여부에 따른 로그 우도의 차이로 다음과 같이 주어진다.^[8]

$$W(q_m) = \log \frac{P(q_m \in A | R_{AQ} = 1)}{P(q_m \in A | R_{AQ} = 0)}$$

$$- \log \frac{P(q_m \notin A | R_{AQ} = 1)}{P(q_m \notin A | R_{AQ} = 0)}$$

$$= \log \frac{\lambda_m (1 - \beta_m)}{\beta_m (1 - \lambda_m)}. \quad (8)$$

위 적합도 $W(q_m)$ 은 각 q_m 들이 커버곡 판별에 확률적으로 독립적이라는 가정 하에 유도된다. 위 적합도를 계산하기 위해서는 λ_m 과 β_m 을 추정해야하며, 일반적으로 커버곡을 만드는 방법이 다양하여 λ_m 을 추정하는 것은 어렵다. 커버곡을 만드는 방식에 따라서 λ_m 이 정해지므로 이를 미리 확률분포로 유도할 수는 없다. 일반적으로 대부분의 경우 A와 Q의 관계가 커

버곡이 아니므로, β_m 을 다음과 같이 Kronecker delta (KD) 함수를 이용하여 추정한다.

$$\beta_m = \frac{n_m}{N}. \quad (9)$$

where $n_m = \sum_{i=1}^N \text{KD}(OTI_{m,i}, OTI_g)$

Eq. (9)의 β_m 을 Eq. (8)에 대입하고 분모와 분자에 확률 적합도 모델^[8]에 따라 0.5를 더하여 다음과 같이 q_m 의 커버곡 인식 1형 적합도 W_1 을 구한다.

$$W_1(q_m) = \log \frac{\lambda_m}{(1 - \lambda_m)} + \log \frac{(1 - \beta_m)}{\beta_m}$$

$$= C + \log \frac{N - n_m + 0.5}{n_m + 0.5}. \quad (10)$$

Eq. (10)에서 n_m 이 $N/2$ 보다 클 경우 log를 취하면 음수가 되고, 이를 방지하기 위해서 분자의 n_m 항을 제거하는 방법이 제안^[9]되었고, 이를 적용하여 커버곡 인식 2형 적합도 W_2 를 다음과 같이 구한다.

$$W_2(q_m) = C + \log \frac{N + 0.5}{n_m + 0.5}. \quad (11)$$

Eqs. (10)과(11)에서 상수 C는 λ_m 과 연관되어 있어서 추정이 어려우므로 보통 0으로 고정한다. 확률 적합도 모델, 적합도 유도 과정, 확률 분포 추정은 References [7] ~ [9]에 자세히 기술되어 있다. 마지막으로 얻어진 적합도를 다음과 같이 로지스틱 함수를 통해서 0과 1사이로 정규화한다. 로지스틱 함수의 파라미터 α, κ 값은 III장에서 실험적으로 정한다.

$$NW(q_m) = \frac{1}{1 + \kappa e^{-\alpha W(q_m)}}. \quad (12)$$

정규화된 적합도 $NW(q_m)$ 를 Eq. (4)의 유사도 행렬 S와 Eq. (6)의 정합 제약 조건 함수 $\delta(\cdot)$ 에 곱한 후에, Eq. (5)의 시간축 정합 행렬 H를 구한다. 얻어진 H를 이용하여 Eq. (7)을 이용하여 기존 방법과 동일하게 커버곡 검색을 수행한다. Eq. (12)에서 $W(q_m)$ 을 Eq. (11)의 W_2 로 사용하되 분자와 분모에 0.5가 더해진 항을

제거하고 α 값을 1로 할 경우 이전 연구^[6]의 허브 지수와 같아진다.

III. 실험 결과

본 장에서는 제안한 적합도 기반 가변 쌍별 유사도와 기존 상수 쌍별 유사도의 커버곡 검색 성능을 비교하였다. 커버곡 성능 비교를 위해서 음원 및 성능이 공개되어 있는 covers80 데이터셋과 자체적으로 수집한 covers330 데이터셋을 사용하였다. 미국 콜롬비아 대학에서 커버곡 실험을 위해서 수집된 covers80 데이터셋은 원본곡과 커버곡 쌍 80개로 이루어진 것으로 모두 160곡으로 구성되어있다.^[10] 자체적으로 수집한 covers330 데이터셋은 1000곡으로 이루어져 있으며, 330곡은 커버곡 실험용으로 30개의 원곡과 각 원곡당 10가지 커버곡 버전으로 구성된다. 나머지 670곡은 검색 성능을 평가하기 위해서 사칭자(imposter)로 삽입되었다.

커버곡 검색 성능 지표로 covers80 데이터셋에서는 각 커버곡을 80곡의 원곡 데이터셋과 비교하여 가장 거리가 가까운 것이 입력 커버곡의 원곡이 맞을 경우의 확률인 P@1을 구하였다. covers330 데이터셋에 대해서는 커버곡 실험용 330곡 각각을 1000곡 데이터셋 전체에 대해서 비교를 수행하여 자신을 제외하고 거리가 가까운 10곡 중 커버곡이 맞는 곡의 개수인 MNCI₁₀을 구하였다. 두 데이터셋 모두에 대해서 검색에 첫 번째로 성공한 커버곡의 검색 순위를 평균한 Rank₁과 Mean of Average Precision(MAP)를 구하였다.

실험대상 음악들로부터 크로마그램 수열을 얻는 다양한 방법들 중에서 최근 우수한 성능을 보이는 것으로 알려진 Convolutional and Recurrent Estimators for Music Analysis(CREMA)^[11] 특징을 사용하였다. CREMA는 음악 코드 분석을 위해서 만들어진 딥러닝 모델로부터 얻어진다. 실험에 사용되는 음악 파일들을 모노로 바꾸고 22,050 Hz로 샘플링 주파수를 맞춘 후, 4,410길이의 윈도우를 50%씩 겹쳐 가면서 프레임의 피치를 구해서 옥타브 단위로 나누고 각 옥타브에서 12개의 크로마에 해당하는 값들을 구한다. 옥타브별로 얻어진 크로마 값을 다 더하면 최종

적으로 12차수의 크로마그램 벡터가 얻어진다. 얻어진 크로마그램 수열을 리샘플링해서 0.5s당 1개씩의 12차 크로마그램벡터가 나오도록 했다.

음원데이터셋 covers80과 covers330에서 Eq. (12)에서 적합도를 정규화하는 로지스틱 함수의 파라미터 값을 가변시켜가면서 커버곡 검색 성능을 확인하고 데이터셋 별로 Tables 1과 2에 각각 정리하였다. 로지스틱 함수 파라미터 κ 는 10에서 30까지 가변하였으며, α 는 0.25에서 1까지 가변하였다. Tables 1과 2에서 고려한 파라미터 값 변이 영역 밖에서는 성능이 감소하였다. covers80 데이터셋에서는 κ 가 큰 값일 때

Table 1. Cover song identification performance of the covers80 dataset. Accuracy measures are the average rank of the first correctly identified cover, Rank₁, precision at one, P@1, and the mean of average precision, MAP.

Method	κ	α	Rank ₁	P@1	MAP
OTI-SW based on relevance weight W_1	10	0.25	13.51	0.688	0.734
	10	0.5	13.83	0.688	0.729
	10	0.75	14.51	0.681	0.722
	10	1	14.82	0.681	0.715
	20	0.25	13.35	0.688	0.734
	20	0.5	13.69	0.688	0.729
	20	0.75	14.09	0.694	0.729
	20	1	13.99	0.675	0.717
	30	0.25	13.39	0.688	0.734
	30	0.5	13.63	0.694	0.733
	30	0.75	13.82	0.706	0.736
	30	1	13.82	0.681	0.722
OTI-SW based on relevance weight W_2	10	0.25	13.59	0.694	0.736
	10	0.5	13.06	0.700	0.745
	10	0.75	13.01	0.694	0.740
	10	1	12.88	0.694	0.737
	20	0.25	13.50	0.694	0.738
	20	0.5	12.95	0.706	0.746
	20	0.75	12.60	0.694	0.738
	20	1	12.54	0.700	0.739
	30	0.25	13.51	0.694	0.738
	30	0.5	12.84	0.706	0.747
	30	0.75	12.42	0.706	0.744
	30	1	12.18	0.700	0.740
OTI-SW (Hubness) ^[6]			11.94	0.706	0.744
OTI-SW (Original) ^[2]			14.50	0.656	0.712

Table 2. Cover song identification performance of the covers330 dataset. Accuracy measures are the average rank of the first correctly identified cover, Rank₁, the mean number of covers identified within the ten first answers, MNCl₁₀, and the mean of average precision, MAP.

Method	κ	α	Rank ₁	MNCl ₁₀	MAP
OTI-SW based on relevance weight W_1	10	0.25	6.23	7.848	0.811
	10	0.5	6.61	7.897	0.816
	10	0.75	6.54	7.839	0.814
	10	1	6.47	7.745	0.807
	20	0.25	6.23	7.861	0.812
	20	0.5	6.64	7.864	0.814
	20	0.75	6.13	7.782	0.807
	20	1	5.84	7.612	0.791
	30	0.25	6.23	7.864	0.812
	30	0.5	6.52	7.833	0.812
	30	0.75	5.79	7.688	0.801
	30	1	5.17	7.467	0.779
OTI-SW based on relevance weight W_2	10	0.25	6.42	7.797	0.806
	10	0.5	6.52	7.909	0.815
	10	0.75	6.57	7.918	0.818
	10	1	6.48	7.924	0.819
	20	0.25	6.47	7.812	0.807
	20	0.5	6.72	7.888	0.815
	20	0.75	6.67	7.876	0.816
	20	1	6.45	7.852	0.815
	30	0.25	6.45	7.827	0.807
	30	0.5	6.64	7.879	0.814
	30	0.75	6.41	7.830	0.813
	30	1	6.01	7.800	0.809
OTI-SW (Hubness) ^[6]			6.49	7.924	0.818
OTI-SW (Original) ^[2]			8.03	7.512	0.781

성능이 더 개선되었으나, covers330 데이터셋에서는 작은 값일 때 더 좋은 성능을 보였다. 다만 고려한 파라미터 변이 영역 내에서 큰 성능 차이는 보이지 않았으므로, 제안된 방법을 실제 적용 시 다른 데이터셋을 사용하더라도 고려한 영역 내의 로지스틱 함수 파라미터 값 중 선택하여 사용하면 될 것으로 보인다. 2.2절에서 언급한 바와 같이 기존 허브 지수 기반 방법은 W_2 와 유사한 수식 형태를 가지게 되어 비슷한 성능을 보이는 것을 확인할 수 있다. 가중치를 구하는 방법 간의 비교를 보면, 커버곡 인식 2형 적합도

인 W_2 를 가중치로 사용하는 것이 커버곡 인식 1형 적합도 W_1 을 사용하는 것에 비해서 조금 더 우수한 검색 성능을 보였다. 제안한 가중치 기반 크로마그램 쌍별 유사도는 MAP를 기준으로 두 데이터셋 모두에서 기존 고정값 기반 유사도에 비해서 커버곡 검색 성능을 4.9% 개선시키는 것을 관찰하였다.

IV. 결 론

커버곡 검색을 위한 크로마그램 쌍별 유사도를 확률 적합도 모델로부터 구한 가중치를 이용하여 구하였다. 커버곡 검색에 도움이 될 수 있는 분별도가 높은 벡터에 높은 가중치를 부여하고, 혼하게 존재하여 분별도가 떨어지는 벡터에 낮은 가중치를 부여한다. 두 가지 가중치 계산 방법을 제안하였으며, 가중치 값은 로지스틱 함수를 통해서 정규화하였다. 두 커버곡 데이터셋에서 성능 비교 실험을 수행하여, 제안한 커버곡 검색 적합도 기반 음악 유사도가 커버곡 검색 성능을 향상시킬 수 있음을 보였다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2023년도 문화기술 연구개발 사업으로 수행되었음(과제명: 딥러닝을 활용한 고속 음악 탐색 기술 개발, 과제번호: CR202104004)

References

1. F. Yesiler, G. Doras, R. M. Bittner, C. J. Tralie, and J. Serra, "Audio-based musical version identification: Elements and challenges," *IEEE Signal Proc. Mag.* **38**, 115-136 (2021).
2. J. Serra, E. Gomez, P. Herrera, and X. Serra, "Chroma binary similarity and local alignment applied to cover song identification," *IEEE Trans. Audio Speech Lang. Process.* **16**, 1138-1151 (2008).
3. J. Seo, "A code-based chromagram similarity for cover song identification" (in Korean), *J. Acoust. Soc. Kr.* **38**, 314-319 (2019).
4. F. Yesiler, J. Serra, and E. Gomez, "Accurate and scalable version identification using musically-motivated embeddings," *Proc. ICASSP*, 21-25 (2020).

5. T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *J. Mol. Biol.* **147**, 195-197 (1981).
6. J. Seo, "Pairwise similarity normalization based on a hubness score for improving cover song retrieval accuracy," *IEICE Trans. Information and Systems*, **105**, 1130-1134 (2022).
7. S. Robertson and K. S. Jones, "Relevance weighting of search terms," *J. Am. Soc. Inf. Sci.* **27**, 129-146 (1976).
8. S. Robertson, "Understanding inverse document frequency: on theoretical arguments for IDF," *J. Documentation*, **60**, 503-520 (2004).
9. S. Robertson and S. Walker, "On relevance weights with little relevance information," *Proc. SIGIR*, 16-24 (1997).
10. *Covers80 Cover Song Data Set, Available*, <https://labrosa.ee.columbia.edu/projects/coversongs/covers80/>, (Last viewed March 12, 2024).
11. B. McFee and J. P. Bello, "Structured training for large-vocabulary chord recognition," *Proc. ISMIR*, 188-194 (2017).

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 : 한국전자통신연구원 선임연구원
 2008년 3월 ~ 현재: 강릉원주대학교 전자공학과 교수