

커버곡 검색을 위한 크로마 n -gram 선택에 관한 연구

An investigation of chroma n -gram selection for cover song search

서진수,^{1†} 김정현,² 박지현²

(Jin Soo Seo,^{1†} Junghyun Kim,² and Jihyun Park²)

¹강릉원주대학교 전자공학과, ²한국전자통신연구원 콘텐츠 연구본부
(Received September 1, 2017; accepted November 29, 2017)

초 록: 음악 유사도 계산은 음악 검색 시스템 구현에 있어서 필수적인 구성 요소이다. 본 논문은 음악 검색 중에서 커버곡 검색에 대해서 다룬다. 크로마 n -gram을 이용한 커버곡 검색에 있어서 특징 DB 저장 공간을 줄이고 성능을 향상시키기 위해서 t -tab n -gram을 제안하고, n -gram 선택 방법, n -gram 집합 간 비교 방법에 관해서 연구하였다. 공개 되어 있는 커버곡 데이터셋에서 실험을 수행하여 제안된 방법이 저장 공간을 줄이면서 동시에 커버곡 검색 성능을 향상시킬 수 있음을 보였다.

핵심용어: 음악 검색, 음악 유사도, 크로마 특징, 커버곡 판별, n -gram

ABSTRACT: Computing music similarity is indispensable in constructing music retrieval system. This paper focuses on the cover song search among various music-retrieval tasks. We investigate the cover song search method based on the chroma n -gram to reduce storage for feature DB and enhance search accuracy. Specifically we propose t -tab n -gram, n -gram selection method, and n -gram set comparison method. Experiments on the widely used music dataset confirmed that the proposed method improves cover song search accuracy as well as reduces feature storage.

Keywords: Music retrieval, Music similarity, Chroma feature, Cover song identification, n -gram

PACS numbers: 43.75.Zz, 43.60.Uv

1. 서 론

디지털 음원의 사용이 확대됨에 따라 음악 데이터를 빠르고 신뢰성 있게 검색하고 제공해 줄 수 있는 오디오 정보 처리 및 검색 기술의 중요성이 증대되고 있다.^[1-3] 본 논문은 라이브 버전 및 리메이크 등을 통해서 재녹음된 음악인 커버곡을 찾는 검색 방법에 관한 것이다. 자동으로 커버곡을 찾을 수 있게 되면 유튜브 등을 통한 무단 저작권 침해에 대해서 대응할 수 있을 것으로 기대된다.

통상적으로 원곡과 커버곡을 비교하면 가수와 악기의 차이로 인한 음색, 연주 속도 및 스타일 차이로

인한 템포 및 리듬, 음악 키의 변조, 도입부와 중심부의 위치 변화 등의 구조적 변경, 가사 변경 등 다양한 차이가 존재한다.^[4] 이렇게 커버곡과 원곡의 차이가 다양하고 하나의 정의가 존재하지 않으며 구분도 모호한 경우도 자주 있으므로 난제의 하나로 남아 있다. 커버곡 검색을 위해서는 다양한 변형에도 불구하고 원곡과 커버곡 간에 보존되는 공통점을 찾아야 한다. 실제로 음원을 분석해 보면, 원곡과 커버곡 간에 가장 잘 보존되는 것은 음들의 시간적 연결을 의미하는 선율(멜로디)이다. 실제로는 상대적인 음악 키의 변조가 가능하므로 음들의 상대적인 시간 변이만이 보존된다. 따라서 기존 대부분의 커버곡 검색 방법들은 선율을 잘 표현할 수 있는 크로마(chroma) 특징 수열을 이용하여 원곡과 커버곡을 비교하였다. 크로마는 인간 청각이 옥타브 차이가 나는 주파수를

†Corresponding author: Jin Soo Seo (sseo@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)

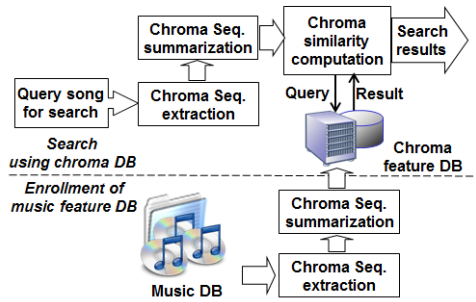


Fig. 1. Overview of the chroma-based cover song search system.

가진 두 음을 유사음으로 인지한다는 음악이론에 기반한다. 옥타브 차이나는 음악의 피치 성분들을 가산하여 음악의 전체 주파수 성분들을 하나의 옥타브 안으로 접어서 표현한 것이다. 크로마 특징은 주파수 분석에 이은 옥타브 단위로 합산하는 과정에서 가수의 목소리와 악기에 연관된 음색보다는 음악 자체의 화성적 구조에 보다 밀접하게 연관되게 된다.^[5] 따라서 크로마는 커버곡 검색에 적합한 특징이다. 다른 음악 주파수 분석 방법들과 마찬가지로 크로마 특징도 음악을 프레임 단위의 짧은 구간(수십ms ~ 수백ms)으로 나누고, 각 프레임에서 M 차 크로마 벡터 (일반적으로 $M = 12$)를 추출하게 되므로 음악 한곡으로부터 크로마 벡터 수열이 얻어지게 된다. 크로마 특징 추출에 관한 상세한 설명은 References [5]와 [6]에 있다.

본 논문에서는 Fig. 1에 주어진 바와 같이 크로마 벡터 수열을 가공하여 검색에 용이한 고정된 길이의 크로마 전곡 특징을 구하는 것에 목표를 두고 연구를 진행하였다. 일반적으로 하나의 크로마 벡터는 커버곡을 판별할 수 있을 정도의 변별력이 없으므로, 인접한 크로마 벡터들을 모아서 더 큰 길이의 벡터로 만드는 n -gram을 이용한다.^[7,8] 본 논문에서는 연속된 n 개의 크로마 벡터를 사용하지 않고, 중간에 t 개의 프레임을 건너뛰면서 n 개의 프레임을 사용하는 t -tab n -gram을 제안하였다. 또한 음악 신호의 길이가 다를 경우 크로마 벡터 수열의 길이도 달라지게 되고, n -gram 벡터의 개수도 달라지게 된다. 따라서 본 논문에서는 음악 신호로부터 얻어지는 n -gram 벡터들 중에서 중복성을 줄이고, 고정된 개수의 대표적인 n -gram들만을 선택하는 방법을 제안하였다. 두 음악신호가 커버곡인지 판별하기 위해서 음악신호

로부터 선택된 n -gram 집합들 간에 거리 계산 방법도 제안하였다.

본 논문은 n -gram 기반 커버곡 검색에 관한 연구이다. II장에서 크로마 수열로부터 t -tab n -gram을 구하는 방법, n -gram 선택 방법, n -gram 집합 간 비교 방법을 제안한다. III장에서 제안된 방법의 성능을 실험하고 결과를 비교 분석한다.

II. 크로마 수열로부터 n -gram 선택

커버곡 검색기는 Fig. 1에 주어진 바와 같이 프레임 레벨 특징을 추출하고 요약하여 DB에 저장하고, 입력 음악의 특징과 비교하여 가장 가까운 음악을 커버곡으로 판정하게 된다. 본 논문에서는 프레임 레벨 특징 요약 방법으로 n -gram을 사용하고, 한 곡의 음악신호로부터 얻어진 n -gram들 중에서 일부를 선택하여 저장하는 방법을 제안한다.

2.1 n -gram 구성 및 선택

음악 프레임에서 얻은 M 차 크로마 벡터만으로는 커버곡 검색을 할 수 있을 정도의 변별력이 없다. 따라서 연속된 n 개 프레임의 크로마 벡터들을 모아서 검색에 사용하는 것이 n -gram이다. 본 논문에서는 연속된 n 개의 크로마 벡터를 사용하지 않고, 중간에 t 개의 프레임을 건너뛰면서 n 개의 프레임을 사용하는 t -tab n -gram을 제안하였다. 음악 신호로부터 얻은 크로마 벡터들을 얻어진 시간 순으로 크로마 벡터 수열 $\mathbf{X} = (X_1, X_2, \dots, X_N)$ 로 표기하면, i 번째 t -tab n -gram 벡터 G 는 n 개의 벡터들로 이루어진 \mathbf{X} 의 부분 수열인 $(X_i, X_{i+t}, \dots, X_{i+(n-1)t})$ 로부터 하나의 nM 차 벡터를 만들어 변별력을 높인 것이다. 이웃한 프레임의 크로마 특징 벡터 간에는 일반적으로 상호연관도가 높으므로 n -gram 구성 시에 t 개의 간격을 두는 것이 더 변별력을 높게 된다. 하지만 t 값이 너무 커지게 되면 음악의 시간적 변이 특성 정보를 소실하게 되므로 적당한 t 값의 선택이 중요하다.

음악의 길이에 따라서 얻을 수 있는 n -gram 벡터의 개수는 가변적이다. 이는 음악 길이가 길어질수록 n -gram의 개수도 늘어나서 상대적으로 비교 시에 커버곡으로 선택될 확률이 늘어나게 된다. 또한 n -gram

의 개수가 많아지면, Fig. 1의 검색기 동작 시에 DB 저장 공간의 소모도 커지게 된다. 이러한 두 가지 문제점을 해결하기 위해서 본 논문에서는 음악 한곡으로부터 얻은 n -gram들 중에서 k 개의 n -gram 만을 선택하는 방법을 제안한다. 이러한 방법으로 가장 널리 사용되고 있는 방법은 k -means 군집화이다. 하지만 k -means 군집화 방법의 경우 군집 대상 벡터들이 특정 벡터 공간에 몰려 있을 경우 그 주변에서만 군집을 형성하는 제약이 있다. 따라서 본 논문에서 다루는 크로마 n -gram에 k -means 군집화를 적용할 경우 상대적으로 발현 횟수는 적지만 변별력이 높은 n -gram들이 군집화 과정에서 누락될 수 있다. 이러한 단점을 보완하기 위해서 선택된 n -gram들 간의 거리값이 최대화 되는 상호거리 최대화 샘플링 방법을 제안한다. 즉 어떤 음악 신호로부터 얻은 n -gram 벡터 수열 $\mathbf{G}=(G_1, G_2, \dots, G_M)$ 로 표기하면, \mathbf{G} 로부터 k 개의 n -gram을 선택하여 $\mathbf{S}=(S_1, S_2, \dots, S_k)$ 를 구성하되, 다음과 같이 선택된 n -gram들 간의 상호거리합인 D_S 가 최대가 되도록 한다.

$$D_S = \sum_{i=1}^k \sum_{j=i+1}^k \|S_i - S_j\|. \quad (1)$$

Table 1에 기술한 바와 같이 \mathbf{S} 의 원소를 새로운 n -gram 벡터로 바꾸는 것이 상호거리합 D_S 를 증가시키는 지 확인하여 D_S 가 증가되도록 업데이트해서 \mathbf{S}

Table 1. The pairwise-distance maximization sampling for selecting n -gram set \mathbf{S} from n -gram vector sequence \mathbf{G} .

| |
|--|
| Initialization of \mathbf{S} as the first k elements of \mathbf{G} $\mathbf{S} = (G_1, G_2, \dots, G_k)$ |
| For $t = k+1, k+2, \dots, N$ |
| 1. Sum the pairwise distances for each element of \mathbf{S} by $d_s[i] = \ S_i - G_t\ + \sum_{j=1}^k \ S_i - S_j\ $. |
| 2. Sum the pairwise distances between \mathbf{S} and G_t by $d_t = \sum_{j=1}^k \ G_t - S_j\ $. |
| 3. Calculate distance gain $d_c[i]$ of replacing S_i with G_t by $d_c[i] = d_t - d_s[i]$. |
| 4. If the maximum of the distance gain is greater than zero, update \mathbf{S} as follows: if $\max(d_c) > 0$, $S_z = G_t$ where $z = \arg \min_j d_s[j]$. |

를 구한다. 상호거리 최대화 방법은 k -means 군집화와는 달리 iteration없이 크로마 수열을 한번 스캔하여 k 개의 n -gram을 선택할 수 있는 장점이 있다. 따라서 제안된 상호거리 최대화 샘플링 방법은 k -means 군집화에 대비해서 계산량이 작고 수렴성의 문제도 없는 장점이 있다.

2.2 n -gram간 거리 기반 커버곡 검색

커버곡을 검색하고자 하는 입력 음악 신호로부터 얻은 크로마 벡터들로부터 선택된 k 개의 n -gram을 $\mathbf{Q} = (Q_1, Q_2, \dots, Q_k)$ 라고 하자. Fig. 1에 도시된 바와 같이 검색 대상 DB상의 음악도 모두 k 개의 n -gram을 선택하여 저장되어 있고 $\mathbf{A} = (A_1, A_2, \dots, A_k)$ 라고 하자. 또한 두 n -gram 집합인 \mathbf{Q} 와 \mathbf{A} 는 Reference [4]에 나온 바와 같이 전곡의 평균 크로마 벡터를 각각 구하고 평균 크로마 벡터간의 상호 상관(cross correlation)이 최대가 되는 OTI(Optimal Transposition Index)를 구하여, 각 n -gram을 OTI 만큼 순환 이동(cyclic shift) 하여 음악의 조 변화를 미리 맞추어 두었다고 가정한다. 두 n -gram 집합인 \mathbf{Q} 와 \mathbf{A} 간의 거리를 구하기 위해서, 먼저 k 행 k 열 거리 행렬 D 를 다음과 같이 각 원소 벡터 쌍 간의 유클리디안 거리로 정의한다.

$$D_{ij} = \|Q_i - A_j\| \quad \text{for } 1 \leq i, j \leq k. \quad (2)$$

각 Q_i 에 대해서 \mathbf{A} 집합에 대한 최소거리를 다음과 같이 d_{\min} 벡터에 구한다.

$$d_{\min}[i] = \min_j [D_{ij}]. \quad (3)$$

얻어진 d_{\min} 벡터의 k 개의 원소들을 올림차순으로 정렬하여 d_{sort} 벡터를 구하고 다음과 같이 \mathbf{Q} 와 \mathbf{A} 간의 거리 D_{set} 을 정의한다.

$$D_{\text{set}}(\mathbf{Q}, \mathbf{A}) = \sum_{t=1}^T d_{\text{sort}}[t]. \quad (4)$$

where $T = \lfloor rk \rfloor$

Eq. (4)에서 r 은 거리값 계산에서 조정계수이며 0과 1사이의 값을 가지며 T 는 rk 보다 작은 가장 큰 정수이다.

III. 실험 결과

본 장에서는 II장에서 제시한 바와 같이 n -gram의 길이 n , 탭간격 t , n -gram 집합 크기 k 를 가변시켜가면서 커버곡 인식 성능을 확인하였다. 커버곡 성능 검증을 위해서 음원 및 성능이 공개되어 있는 covers80 데이터셋⁹⁾을 사용하였다. 미국 콜롬비아 대학에서 커버곡 실험을 위해서 수집된 covers80 데이터셋은 원본곡과 커버곡 쌍 80개로 이루어진 것으로 모두 160곡으로 구성되어있다.

실험대상 음악들로부터 크로마 수열을 얻기 위해서 Matlab 기반의 Chroma Toolbox⁶⁾를 사용하였으며, 크로마를 얻는 방법 중 음색에 대한 불변성을 개선한 CRP(Chroma DCT-Reduced log Pitch)를 적용하였다.¹⁵⁾ 실험에 사용되는 음악 파일들을 모노로 바꾸고 22050Hz로 샘플링 주파수를 맞추고, 4410길이의 윈도우를 50%씩 겹쳐가면서 프레임의 피치를 구해서 옥타브 단위로 나누고 각 옥타브에서 12개의 크로마에 해당하는 값들을 구한다. 옥타브별로 얻어진 크로마 값을 다 더하면 최종적으로 12차수의 크로마 벡터가 얻어진다. 얻어진 크로마 수열을 리샘플링해서 0.5 s당 1개씩의 12차 크로마 벡터가 나오도록 했다. 리샘플링된 크로마 수열로부터 t -tab n -gram을 구성하고, 상호거리 최대화 방법 또는 k -means 방법을 적용하여 최종적으로 선택된 k 개의 n -gram으로 이루어진 집합을 구한다. 각 음악으로부터 선택된 n -gram 집합 간의 거리는 (4)의 수식을 이용해서 구한다. 원본 음원과 커버곡 음원간의 n -gram 집합 간 거리가 원본 음원과 다른 원본 음원들간의 n -gram 집합 간 거리에 비해서 작을 경우에 커버곡 검색이 성공한 것으로 한다. covers80 데이터셋에 있는 각 원본 음원에 대해서 데이터셋 내의 음원들과 거리 비교를 수행하고 가장 거리값이 작은 것이 커버곡이 맞은 경우의 확률을 구해서 검색 성능으로 사용하였다. 실험결과에서 k -means 방법의 경우 반복 군집화 방법으로 초기값에 따라서 선택된 n -gram이 달라지게 되므로, 본 논문의 k -means 방법의 커버곡 검색 성능은 5번 반복 수행한 평균치이다.

먼저 n -gram의 길이 n 과 n -gram 집합 크기 k 를 가변하면서 커버곡 검색 성능을 Figs. 2와 3에 도시하였

다. 크로마 n -gram 수열 축약 방법으로 상호거리 최대화 방법을 사용한 경우와 k -means 방법을 사용한 경우 모두 비슷한 경향성을 보였다. 검색 성능의 최대값은 52 % 수준으로 비슷하였으며, 검색 성능이 최대가 되는 n 값은 Figs. 2와 3에서 모두 7이었다. 실험에서 k 값을 16에서 48까지 가변하였으며 도시한 실험결과를 보면 k 값이 32일 때 성능이 가장 우수하였으며, 그 이상의 k 값을 사용하더라도 성능이 개선되지 않음을 알 수 있다. 실제적으로 커버곡 검색 시스템을 구현할 경우 각 음원별로 k 개의 n -gram을 DB에 저장해야한다. 따라서 n 과 k 에 대해서 작은 값을 사용하는 것이 DB 저장 공간을 줄이는 데 도움이 된다.

음원으로부터 얻어진 n -gram을 축약하지 않고 모두 사용할 경우인 Figs. 2와 3의 점선과 비교해 보면 축약하는 것이 더 우수한 성능을 보임을 알 수 있다. 이는 음악의 길이가 서로 다를 경우에 특징 축약을

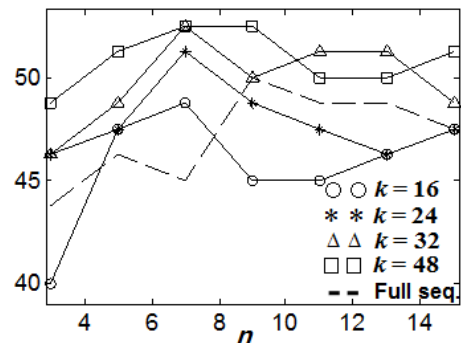


Fig. 2. Search accuracy (%) versus size of n -gram for covers80 dataset using n -gram selection with the pairwise-distance maximization.

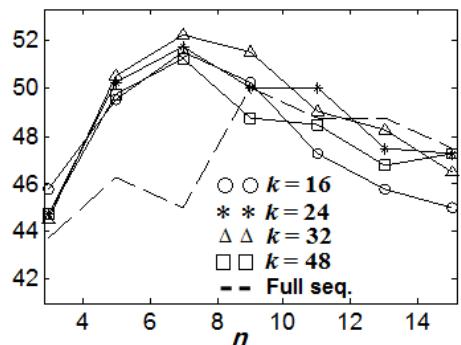


Fig. 3. Search accuracy (%) versus size of n -gram for covers80 dataset using n -gram selection with the k -means algorithm.

통해서 n -gram의 개수가 달라지는 부분을 보완할 수 있기 때문이다. 음악의 길이가 길어지면 n -gram의 개수도 늘어나므로 확률적으로 유사한 n -gram을 가지고 있을 확률이 짧은 길이의 음악보다 높아지고, 이는 커버곡 유무와 상관없이 검색 상위에 나올 확률이 높아짐을 의미한다. 따라서 음악의 길이에 무관하게 k 개의 n -gram만을 선택하는 것이 음악의 길이에 대한 정규화를 수행하여 검색 성능을 높일 수 있다.

기존 covers80 데이터셋에서 보고된 커버곡 검색 성능은 42.5%에서 67.5%이다.^[10] 제안된 방법의 최고 성능인 52.5%는 최저와 최고 성능의 중간 정도이며, 기존 방법의 베이스라인인 42.5%보다 우수하였다. 비교대상 기존 방법^[10]은 전체 크로마 수열을 모두 사용하므로 k 개의 n -gram만을 선택하는 제안된 방법이 DB 저장 공간 크기 측면에서 더 우수하다. 또한 본 논문의 결과는 템포 변형을 고려하지 않고 (2)에 유클리디안 거리를 사용하였으나 이를 템포 변형을 고려할 수 있는 DTW(Dynamic Time Warping) 등을 적용할 경우 좀 더 성능을 높일 수 있을 것으로 기대된다.

Figs. 4와 5는 Figs. 2과 3으로부터 정해진 $n = 7, k = 32$ 에 대해서 템포격 t 와 n -gram 집합 간 거리 비교 시에 사용하는 조정계수 r 를 가변하면서 커버곡 검색 성능을 도사한 것이다. 먼저 템포를 사용하지 않는 기존의 n -gram 방법인 $t=1$ 인 경우에 비해서 템포를 사용하는 것이 검색 성능을 개선함을 알 수 있다. 이는 $t=1$ 인 경우 인접한 크로마 벡터간의 상호연관성이 높으므로 n -gram 내부 중복성(redundancy)이 증가된다. 그러나 템포격으로 3보다 더 큰 값을 사용할 경우 성능 저하가 관찰되었다. 템포격이 너무 커질 경우 t -tab n -gram이 커버곡 비교에 중요한 크로마 수열의 시간적 변이 특성을 상당부분 잃어버리게 된다. 실험 결과로부터 템포격을 3으로 하는 것이 n -gram 내부 중복성은 줄여주고, 크로마 수열의 시간적 변이 특성은 보존해 줌을 알 수 있다. 상호거리 최대화 방법의 경우 조정계수 r 을 0.4에서 0.6 사이의 값을 사용하는 것이 성능을 높였으며 0.4보다 작거나 0.6보다 큰 값을 사용하면 성능의 저하가 있었다. 반면에 k -means 방법의 경우 조정계수 r 이 0.8에서 0.9 사이일 때 가장 좋은 성능을 보였다. 상호거리 최대화 방법을 통해 선택된 n -gram들은 특징공간에서 산재해 있으

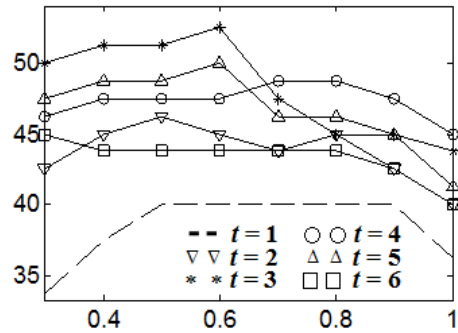


Fig. 4. Search accuracy (%) versus adjustment factor r for covers80 dataset using n -gram selection with the pairwise-distance maximization.

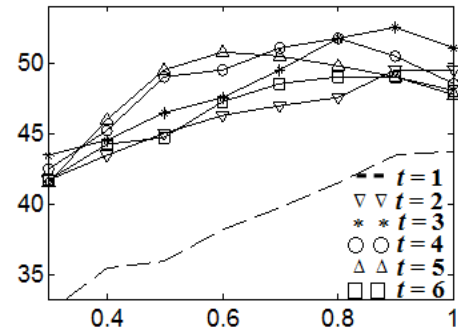


Fig. 5. Search accuracy (%) versus adjustment factor r for covers80 dataset using n -gram selection with the k -means algorithm.

로 일부만 유사하여도 커버곡으로 판별할 수 있다. 반면에 k -means 방법의 경우 선택된 n -gram들이 원특징 분포가 조밀한 곳을 중심으로 밀집한 형태를 가지는 경우가 많다. 따라서 상호거리 최대화 방법에 비해서 조정계수가 더 큰 값을 선호하게 된다. 선택된 n -gram 집합의 일부만을 사용하여 커버곡 판별을 할 수 있다면 커버곡 생성 과정에서 원곡에 대비하여 일부가 크게 변조되거나 삭제되는 변형이 발생할 경우의 커버곡 판별에 도움이 될 수 있다. 차후 다양한 변형 별로 새로운 커버곡 음악쌍을 수집해서 n -gram 방법 간의 면밀한 성능 비교가 필요할 것으로 생각된다.

IV. 결 론

본 논문은 커버곡 검색을 위한 n -gram 집합 선택 방법에 관한 논문이다. 음원별로 길이가 상이하므로 n -gram 벡터의 개수도 차이가 있어서 이를 보정하기

위해서 n -gram들 중 일부만을 선택하는 방법이 필요하다. 본 논문에서는 상호거리 최대화와 k -means 방법을 적용하여 n -gram을 선택하였고, 성능 및 장단점에 대해서 비교하였다. 널리 사용되고 있는 커버곡 데이터셋에서 실험을 수행하여 커버곡 검색을 위한 최적의 n -gram 길이, 탭간격, 집합의 크기를 찾았다. 실험을 통해서 음악으로부터 얻은 n -gram들을 모두 사용하는 것보다 고정된 길이의 n -gram 집합을 선택하여 사용하는 것이 특징 DB 저장공간을 줄이고 커버곡 검색 성능을 개선할 수 있음을 확인하였다.

감사의 글

본 연구는 문화체육관광부 및 한국저작권위원회의 2017년도 저작권기술개발사업의 연구결과로 수행되었음(연관 저작물 검색과 저작권 침해 예방을 위한 예측 탐지 기술 개발).

References

1. M. A. 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).
2. J. Lee and H. Kim, "Audio fingerprinting using a robust hash function based on the MCLT peak-pair" (in Korean), *J. Acoust. Soc. Kr.* **34**, 157-162 (2015).
3. J. Seo, J. Kim, and J. Park, "Centroid-model based music similarity with alpha divergence" (in Korean), *J. Acoust. Soc. Kr.* **35**, 83-91 (2016).
4. 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).
5. M. Muller and S. Ewert, "Towards timbre-invariant audio features for harmony-based music," *IEEE Trans. Audio Speech Lang Process.* **18**, 649-662 (2010).
6. M. Muller and S. Ewert, "Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features," *Proc. ISMIR-2011*, 215-220 (2011).
7. M. Casey, C. Rhodes, and M. Slaney, "Analysis of minimum distances in high-dimensional musical spaces," *IEEE Trans. Audio Speech Lang Process.* **16**, 1015-1028 (2008).
8. P. Grosche and M. Muller, "Toward characteristic

audio shingles for efficient cross-version music retrieval," *Proc. ICASSP-2012*, 473-476 (2012).

9. *The covers80 cover song data set, available, https://labrosa.ee.columbia.edu/projects/coversongs/covers80/, 2007*
10. D. Ellis and C. Cotton, "The 2007 LabROSA cover song detection system," in *MIREX extended abstract 2007*, (2007).

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 조교수, 부교수

▶ 김정현 (Junghyun 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 공학석사
 2010년 2월: 충남대학교 컴퓨터공학과 공학박사
 1999년 5월 ~ 현재: 한국전자통신연구원 책임연구원