

## 거리 함수 학습을 활용하여 장르 분류를 위한 특징 셋의 간소화 방법 연구

장달원 신사임 이종설 장세진 임태범

전자부품연구원

## Feature reduction based on distance metric learning for musical genre classification

Jang, Dalwon Shin, Saim Lee, JongSeol Jang, Sei-Jin Lim, Tae-Beom

Korea Electronics Technology Institute

## 요약

음악 장르 분류 분야에서는 다양한 특징을 모아서 특징 벡터를 만들고 이를 support vector machine (SVM)와 같은 분류기에 입력하는 시스템이 주로 사용되고 있다. 이 논문에서는 거리 함수 학습을 음악 장르 분류를 위한 특징 벡터의 간소화에 적용하였다. 여러 거리 함수 학습 방법 중 하나의 방법을 선택하고, 기존의 논문들에서 사용되었던 특징 셋을 활용하여 기존 특징 셋에 대해서 성능을 떨어뜨리지 않으면서 특징 셋의 길이를 줄일 수 있는지 살펴본다. 우리의 실험에서는 168차원의 특징 셋을 10차원까지 줄였는데, 이 경우 분류 정확도가 2% 이내로 저하되었다.

## 1. 서론

디지털 음악의 사용이 늘어나면서 대용량의 음악 데이터베이스에 저장된 정보들을 효율적으로 관리하고, 검색하는 기술에 대한 수요가 존재했다. 따라서 음악 정보 검색에 대한 다양한 연구가 시도되었으며, 음악 장르 분류 분야는 그 중 하나라 할 수 있다[1-7]. 음악 장르 분류는 정보를 알 수 없는 하나의 음악 입력이 있을 때 그것을 분석하고, 미리 정해진 여러 개의 장르 중 하나의 장르를 출력해 주는 기술이다. 음악의 장르는 주요한 음악 정보 중 하나로 음악 데이터베이스 관리에 꼭 필요한 정보라 할 수 있고, 자동적인 분류가 가능하다면 판매처, 방송국 등 대용량의 음악 데이터베이스를 보유하는 업체 뿐만 아니라 개인에게도 많은 도움이 될 것이다. 음악장르 분류 분야는 2000년대 이후로 많이 연구되었으며 특징(feature)들을 추출하는 과정과 분류기를 이용해서 결과를 얻는 과정으로 구성된다[1-7]. 현재는 Mel-Frequency Cepstral Coefficients (MFCC)를 비롯한 여러 가지 특징들을 추출하고, 이를 합쳐서 하나의 특징 벡터를 만들어서 사용하고 있다. 때에 따라서는 특징 벡터를 간소화 과정을 거치기도 한다. 분류기 중에서는 support vector machine (SVM)이 일반적으로 좋은 성능을 보이고 있다.

본 논문에서는 거리 함수 학습 중 한 방법인 maximally collapsing metric learning (MCML)을 사용하여 특징 벡터 간소화를 하였을 때 음악 장르 분류 성능을 살펴보았다. 이전 연구에서 MCML을 분류기에 이용하였을 때 음악 장르 분류 시스템의 성능향상이 되는 지에 대해서 살펴보았고[8,9], 본 논문에서는 앞선 연구를 기반으로 기존의 특징 셋의 차원을 줄여가면서 성능이 어떻게 되는지 알아보았다.

## 2. MCML 방법 [10]

일반적으로 거리 함수 학습 방법은  $N$ 차원의 벡터 사이의 거리를 정의하는데 있어서  $N \times N$  크기의 어떤 positive semi-definite (PSD) 행렬  $A$ 를 파라미터로 이용한다. 파라미터의 값에 따라서 거리 함수가 결정되고, 그 파라미터는 학습을 통해서 결정된다.  $N$ 차원의 벡터  $\vec{x}_i$ 와  $\vec{x}_j$ 에 대해서 다음과 같은 거리를 정의한다.

$$d_A(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)^T A (\vec{x}_i - \vec{x}_j)$$

MCML의 기본 원리는 하나의 클래스에 속한 모든 특징 벡터들은 하나의 점에 모여 있고, 다른 클래스에 속한 특징 벡터끼리는 무한히 먼 위치에 있다는 이상적인 가정 하에 행렬  $A$ 를 정한다[10]. 이런 가정 하에 각각의 훈련 데이터마다 정의되는 조건부 확률 분포는 이상적인 상황에 최대한 가깝게 되도록 학습한다. 어떤 벡터  $\vec{x}_i$ 에 대한 조건부 확률을 아래와 같이  $\vec{x}_i$ 와 다른 벡터와의 거리를 기준으로 정해진다.

$$p^A(j|i) = \frac{1}{Z_i} e^{-d_A(\vec{x}_i, \vec{x}_j)} = \frac{e^{-d_A(\vec{x}_i, \vec{x}_j)}}{\sum_{k \neq i} e^{-d_A(\vec{x}_i, \vec{x}_k)}}$$

위와 같이 정의된 조건부 확률 분포가 이상적인 확률분포와 가깝게 만들기 위해서 두 확률 분포 사이의 KL divergence를 최소화시키는 방법을 사용한다. 이 때, 이상적인 확률 분포는  $\vec{x}_i$ 와 같은 클래스에 속한 벡터들 사이의 거리는 모두 0이므로 조건부 확률값은 어떤 상수를 가지고,  $\vec{x}_i$ 와 다른 클래스에 속한 벡터와의 거리는 무한대이므로 조건부 확률 값이 0이 되도록 정한다. 훈련

데이터에 속하는 각각의 벡터  $\vec{x}_i$ 에 대해서 계산된 확률 분포와 이상 확률 분포와의 KL divergence를 구하고 이를 전체에 대해서 더한 합을 최소화하는 행렬  $A$ 를 구한다. 행렬  $A$ 를 구하는 자세한 과정은 [10]에서 찾아볼 수 있다.

행렬  $A$ 를 이용해서 정의된 거리 함수는  $\sqrt{A}$ 로 선형 투영 (linear projection) 후 유클리디언 거리를 측정하는 것으로 해석할 수 있다. 하지만 이 투영을 통해서 벡터의 차원을 줄일 수 없고, 행렬  $A$ 를 기반으로 저차로 변환하는 투영 행렬을 구해야 벡터 간소화가 가능하다. 행렬  $A$ 를 고유값 분해(eigen decomposition)하여  $N$ 개의 고유값  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_N$ 과 고유벡터,  $\vec{v}_i$ 를 얻었을 때, 투영행렬은 다음과 같이 구한다.

$$W = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_q}) [\vec{v}_1, \dots, \vec{v}_q] \quad (\text{단, } q \leq N)$$

위 투영 행렬을 이용해서 특징벡터를 간소화한다.

### 3. 실험

실험에서는 GTZAN 데이터 셋을 사용하였는데[1], 이 데이터는 10가지 장르에 대해서 각각 100곡씩 데이터를 가지고 있다. 우리는 이 데이터 셋에 대한 분류 정확도를 10 fold cross-validation을 이용해서 계산하였다. 따라서, 900곡을 훈련 데이터로 사용하고, 100곡을 테스트 데이터로 사용하였으며 훈련 데이터와 테스트 데이터를 바꿔가면서 10번의 과정을 거치고, 전체 평균을 구한다. 우리는 [6] 논문을 참고하여 168차원을 특징 벡터를 구성하였다. 이 논문에서 사용한 특징 벡터는 MFCC, DFB, OSC에 대한 평균과 분산, 그리고 세 가지에 대한 특징기반 모듈레이션 스펙트럼을 구하였다[6]. MCML을 이용해서 거리 함수를 결정할 때, 행렬  $A$ 의 초기값을 정하고, 여러 번의 반복 과정을 거치게 되는데, 우리는 이 실험에서 초기값을 identity 행렬로 정하고, 경우에 따라서 100번~200번의 반복과정을 거치게 하였다. 벡터 간소화한 후 길이인  $q$ 를 다양하게 하면서 실험하였고, 간소화한 결과를 SVM과 k-nearest neighbor(k-NN) 분류기를 사용해서 분류 결과를 얻었다. 실험에서는 linear 커널을 사용하는 SVM 분류기와 유클리디언 거리를 사용하는 일반적인 k-NN 분류기를 사용하였다.

아래 표 1에서 다양한 길이로 특징벡터를 간소화했을 때의 성능을 나타내었다. 특징 벡터를 그대로 사용했을 때 성능이 각각 75.1, 76.0, 80.3이기에, MCML이 성능 향상을 이끌어냈다는 것을 알 수 있고[8], 특징 벡터를 아주 짧게 줄여도 성능에 2%이내의 차이만 있었다. MCML을 통한 특징벡터 간소화가 성공적이었다고 할 수 있다.

벡터길이( $q$ )	분류기 별 성능		
	k-NN(k=3)	k-NN(k=5)	SVM
168	81.0	82.4	80.1
100	80.7	82.4	89.3
50	80.5	82.4	80
30	80.8	82.3	79.2
20	80.0	82.5	79.4
10	79.7	81.3	79.8

표 1. 벡터 간소화했을 경우의 분류 정확도

### 4. 결론

앞선 연구들에 이어서 거리 함수 학습 방법의 하나인 MCML 방법을 음악 장르 분류에 활용해 보았다. 총 168차원의 특징 벡터를 10차원까지 줄였지만, 성능의 큰 저하가 나타나지 않았다. 논문에서 사용한 시스템에서는 좋은 성과를 얻었지만, 더 다양한 특징 셋에 대한 성능검증 및 타 알고리즘과의 비교가 추가적으로 연구되어야 할 것 이다.

### 5. 감사의 글

이 논문은 2014년도 정부(산업자원부)의 재원으로 산업원천기반 구축개발사업의 지원을 받아 수행된 연구임(No. 10037244 ).

### 6. 참고문헌

- [1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Trans. Speech Audio Process., vol. 10, no. 5, pp. 293-302, 2002.
- [2] T. Li , M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification" Proc. ACM Con! on Research and Development in Information Retrieval, pp. 282-289, 2003
- [3] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai, "Music type classification by spectral contrast feature," Proc. ICME 02, vol. 1, pp. 113-116, 2002.
- [4] C-H. Lee, J-L. Shih, K-M. Yu, and J-M Su, "Automatic music genre classification using modulation spectral contrast feature," Proc. ICME 07, 2007
- [5] D. Jang, M. Jin, and C. D. Yoo, "MUSIC GENRE CLASSIFICATION USING NOVEL FEATURES AND A WEIGHTED VOTING METHOD," in Proc. ICME 08, June 23 2008-April 26 2008.
- [6] S.-C. Lim, J.-S. Lee, S.-j. Jang, S.-P. Lee, and M. Y. Kim, "Music-Genre Classification System based on Spectro-Temporal Features and Feature Selection," IEEE Trans on Consumer Electronics, Vol. 58, No. 4m Nov. 2012.
- [7] S.-C. Lim, S.-J. Jang, S.-P. Lee, and M. Y. Kim, "Music genre/mood classification using a feature-based modulation spectrum," in Proc. IEEE Int. Conf. Mobile IT Convergence, 2011, pp. 133-136.
- [8] 장달원, 신사임, 이종설, 장세진, 임태범, "거리 학습과 k-nearest neighbor 분류기를 이용한 음악 장르 분류," 정보과학회 하계학술대회, 2013
- [9] 장달원, 신사임, 이종설, 장세진, 임태범, "장르 분류 성능 향상을 위한 거리함수 학습의 활용," 한국방송공학회 하계학술대회, 2013
- [10] A. Globerson and S. Roweis, "Metric learning by collapsing classes," Adv. Neural Inf. Process. Syst., vol. 18, pp. 451 - 458, 2006.