

장르 분류 성능 향상을 위한 거리함수 학습의 활용

장달원 신사임 이종설 장세진 임태범

전자부품연구원

dalwon@keti.re.kr

Applying distance metric learning for improvement of 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 사용하는 방법과 support vector machine (SVM) 분류기가 주로 사용되고 있다. 이 논문에서는 거리 함수 학습을 음악 장르 분류에 적용하여 성능 향상을 꾀한다. 여러 거리 함수 학습 방법 중 하나의 방법을 선택하고, 일반적으로 많이 사용되는 특징 셋을 활용하여 다양한 특징 셋에 대해서 적용하였을 때, 실제 성능 향상이 있는지를 알아본다. 세 종류의 특징 셋을 사용하여 실험한 결과 두 가지 특징이 같이 있는 특징 셋에 대해서만 성능 향상이 있었으며, SVM보다 높은 성능을 보이지 못했다.

1. 서론

디지털 음악의 사용이 늘어나면서 대용량의 음악 데이터베이스에 저장된 정보들을 효율적으로 관리하고, 검색하는 기술에 대한 수요가 생겨왔다. 그리하여 음악 정보 검색과 관련된 다양한 연구가 있어왔으며, 그 중 하나로 음악 장르 분류 분야가 있다[1-7]. 음악 장르 분류는 입력 음악에 대해서 그것을 분석하고, 미리 정해진 여러 개의 장르 중 하나의 장르를 입력에 대한 것으로 출력해 주는 기술이다. 음악의 장르는 음악 정보의 관리에 꼭 필요한 정보로 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 셋을 만들어서 사용하고 있다. 여러 가지 분류기 중 support vector machine (SVM)이 일반적으로 좋은 성능을 보이고 있다.

이 논문에서는 거리 함수 학습 방법이 음악 장르 분류 시스템의 성능 향상에 도움을 줄 수 있는지에 대해서 알아본다. 거리 함수 학습 방법은 2000년대 중반 이후 다양한 방법들이 개발되었고 [8-11], 다양한 분류 분야에 적용되었다. 이 논문에서는 그 중 maximally collapsing metric learning (MCML) 방법을 사용하였다 [11]. 거리 함수 학습 방법을 사용할 경우, 분류기는 k-nearest neighbor (k-NN) 분류기를 사용하게 된다. MCML을 일반적으로 많이 사용되는 음악장르 분류 테스트 데이터와 세 종류의 특징셋을 사용하는 음악 장르 분류 시스템에 적용하였을 때 성능의 변화를 알아보고, 현재 많이 사용되는 SVM 분류기와의 성능을 비교한다.

2. MCML 방법 [11]

일반적으로 거리 함수 학습 방법은 N 차원의 벡터에 대해서 $N \times N$ 크기의 어떤 positive semi-definite (PSD) 행렬 A 를 파라미터로 정의하고, 파라미터 행렬 A 를 이용해서 다음과 같은 거리를 정의한다.

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

여기에서 \vec{x}_i 와 \vec{x}_j 는 N 차원의 어떤 벡터를 나타낸다. 파라미터가 어떤 값으로 결정되느냐에 따라서 거리 함수가 결정되고, 그 파라미터는 학습을 통해서 결정된다.

MCML의 기본 원리는 하나의 클래스에 속한 모든 특징 벡터들은 하나의 점에 모여 있고, 다른 클래스에 속한 특징 벡터끼리는 무한히 먼 위치에 있다는 이상적인 상황을 가정하는 것이다 [11]. 이런 가정에 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. 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 방법은 [11]에서 찾아볼 수 있다.

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.

세 가지 특징 셋을 구성하였다: 특징 셋 1) MFCC의 평균(13차원), 특징 셋 2) MFCC의 평균과 분산(26차원), 특징 셋 3) MFCC, Octave-band spectral contrast (OSC) [3] 의 평균과 분산(58차원). 이 특징들을 오랜 연구를 통해서 음악 장르 분류 분야에서 검증된 특징이며, 논문으로 발표된 다양한 시스템에서 많이 사용되기에 [4-7] 이것들로 결정하였다. 각 프레임마다, 각각 13차원의 MFCC와 16차원의 OSC 특징이 추출되는데, 이 특징들은 여러 프레임간의 평균, 분산을 구한 것이다. 사용한 세 가지 특징 셋 자체가 높은 성능을 보장할 수 없지만, 다양한 시스템에서 많이 사용되는 특징들이기 때문에, 일반적인 경우에 MCML이 성능 향상에 영향을 미치는지에 대해 알아보기 적합할 것으로 생각된다.

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. 실험에서는 linear 커널을 사용하는 SVM 분류기와 유클리디안 거리를 사용하는 일반적인 k-NN 분류기를 비교대상으로 삼는다.

아래의 표 1.에 다양한 경우의 분류 정확도를 기술하였다. 세 가지 특징 셋에 대해서 특징 셋 3에 대해서만 1~2% 정도 성능이 향상됨을 확인할 수 있었다. 하지만 기존의 k-NN 분류기에 비해서는 성능이 향상되었지만, SVM보다는 낮은 성능을 보였다. 그리고, 특징 셋 1, 2에 대해서는 오히려 성능이 떨어졌다.

분류기	분류 정확도		
	특징 셋 1	특징 셋 2	특징 셋 3
k-NN (k=3)	45.6	34.9	26.6
k-NN (k=5)	43.4	33.4	25
SVM	46.9	32.8	22.9
MCML + k-NN (k=3)	50.7	35.7	24.6
MCML + k-NN (k=5)	47.6	36.2	23.4

표 1. 다양한 경우의 분류 정확도

4. 결론

현재 음악 장르 분류 분야에 많이 사용되는 k-NN 분류기의 성능을 향상해 보기 위해서 거리 함수 학습 방법을 적용하고, 이것이 일반적인 상황에서 성능 향상을 이끌 수 있는지 알아보기 위해서 일반적인 특징 셋 3가지를 정하고 이에 대해서 성능을 알아보았다. MFCC만을 사용한 단순한 특징 셋일 경우에는 학습과정이 오히려 성능을 떨어뜨리는 결과를 보였으나, 두 가지 특징이 섞인 특징 셋을 사용하니, 성능이 좋아지는 모습을 보였다. 본 논문에 제시된 결과만으로는, 우리 실

험에서 사용한 MCML 방법이 음악 장르 분류에 도움을 주는지에 대한 명쾌한 해답을 얻을 수 없었다. 더욱 복잡한 형태의 실제적인 특징 셋에 대해서 어떤 결과를 보이는지 알아봐야 할 것이다.

5. 감사의 글

이 논문은 2013년도 정부(지식경제부)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] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," Adv. Neural Inf. Process. Syst. vol. 14, pp. 521 - 528, 2002.
- [9] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning a Mahalanobis metric from equivalence constraints," J. Mach. Learning Res., vol. 6, pp. 937 - 965, 2005.
- [10] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighborhood components analysis," Adv. Neural Inf. Process. Syst., vol. 17, pp. 513 - 520, 2005.
- [11] A. Globerson and S. Roweis, "Metric learning by collapsing classes," Adv. Neural Inf. Process. Syst., vol. 18, pp. 451 - 458, 2006.