

소셜 태그를 이용한 강인한 음악 분류 기법

이재성[○], 김대원^{*}

^{○*}중앙대학교 컴퓨터공학부

e-mail: jslee.cau@gmail.com[○], dwkim@cau.ac.kr^{*}

Robust Music Categorization Method using Social Tags

Jaesung Lee[○], Dae-Won Kim^{*}

^{○*}Dept. of Computer Science and Engineering, Chung-Ang University

● 요약 ●

음악 검색에 있어 소셜 태그 정보는 사용자 하여금 음악의 내재적 의미를 빠르게 파악할 수 있도록 한다. 음악의 소셜 태그 정보는 음악 추천 시스템을 활용하는 사용자(청취자)에 의해 점진적으로 완성되기 때문에 초기에 완전한 태그 정보를 수집하는 것은 어렵다. 본 논문에서는 음악의 일부 태그가 누락되어 있는 상황에서 음악 정보 검색을 자동으로 수행할 수 있는 클래스 분류 알고리즘을 제안하고자 한다.

키워드: 음악 소셜 태그(Music Social Tag), 음악 정보 검색(Music Information Retrieval), 음악 분류(Music Categorization)

I. 서론

최근 대량으로 유통되는 음악 속에서 사용자가 원하는 음악을 찾기 위한 음악 정보 검색 기술의 중요성이 강조되고 있다 [1]. 음악의 내재적 특성을 활용한 음악 분류 기법(Content-based Music Categorization)은 새롭게 공개되어 분류 정보가 없는 음악이 주어졌을 때, 학습된 사상 함수를 활용하여 자동으로 분류 정보를 추론할 수 있다 [2].

사상 함수를 학습시키기 위한 분류 정보를 취득하는 것이 많은 비용을 유발한다고 알려져 있다 [2,3]. 이러한 문제점을 극복하기 위해 최근 SNS를 통해 부여한 음악 태그를 활용하는 연구들이 주목을 받고 있다 [4,5]. 그러나 익명의 여러 사용자들이 동시에 하나의 음악에 태그를 부여하지 않으므로, 특정 시점에 수집된 음악 태그들이 불완전할 수 있다. 본 논문에서는 음악의 태그가 불완전하게 부여되어 있는 상황에서 강인하게 음악 분류를 수행하는 알고리즘을 제안하고자 한다.

II. 제안하는 알고리즘

주어진 음악 집합 $X \subset R^d$ 에 포함된 하나의 음악 $x \in X$ 가 d 개의 음향 정보로 표현된다고 하자. 이 때, x 에 대해 연관성이 높은 태그 집합 y 에 사상시키는 함수 $f(x) \rightarrow y$ 를 찾으면, $f(\cdot)$ 함수를 이용하여 새로운 음악에 대해 연관성이 높은 태그 집합을 찾을 수 있다. 일반적으로 하나의 음악은 최대 q 개의 태그 중 1개 이상의 태그를 가질 수 있으므로 $y \in \{0,1\}^q$ 로 표현되며, y 의 j 번째

값($1 \leq j \leq q$)이 1이라고 하였을 때 j 번째 태그가 x 에 부여됨을 의미하고 0은 그 반대를 나타낸다 [1,3,7].

일반적인 학습 과정에서 부여되지 않은 태그들은 모두 음성으로 간주하고 학습을 진행하기 때문에, 불완전한 태그 정보는 클래스 불균형 문제를 발생시킨다 [6,8]. 각 음악에 대해 양성으로 부여된 음악들의 R^d 상에서의 분포가 서로 다를 수 있으므로, 각 태그 정보에 대해 서로 다른 사상함수 $f_j(\cdot)$ 를 학습해야 한다. j 번째 태그가 양성으로 부여된 음악들의 집합을 tr_j 라고 하자. 그리고 태그가 부여되지 않은 새로운 음악 x 에 대해서 tr_j 에 포함된 최근린 음악을 $NN_j^{tr}(x)$ 라고 하자. 이 때, x 와 $NN_j^{tr}(x)$ 사이의 비유사도 d_1 은 다음과 같이 계산할 수 있다.

$$d_1 = \|x - NN_j^{tr}(x)\| \quad (1)$$

다음으로 $NN_j^{tr}(x)$ 와 tr_j 에 포함되어 있는 음악 중 $NN_j^{tr}(x)$ 의 최근린 음악을 $NN_j^{tr}(NN_j^{tr}(x))$ 라고 하자. 이 때, 학습 데이터에서 양성으로 판단된 음악들 사이의 비유사도 d_2 를 다음과 같이 계산할 수 있다.

$$d_2 = \|NN_j^{tr}(x) - NN_j^{tr}(NN_j^{tr}(x))\| \quad (2)$$

새로운 음악 x 가 현재 학습 중인 태그에 대해 양성으로 판정되기 위해서는 적어도 x 의 비유사도가 양성으로 판정된 음악들 사이의 비유사도를 초과하지 않아야 한다. 따라서 x 의 비유사도 초과 비율을

다음과 같이 구할 수 있다.

$$P_j(x) = d_1/d_2 \quad (3)$$

본 논문에서는 $P_j(x)$ 가 1보다 작으면 양성으로 그 반대의 경우는 음성으로 태그를 부여하였다. 위의 과정을 j 개의 레이블에 대해 수행한다.

III. 실험

제안하는 알고리즘의 성능을 검증하기 위해서 음악 정보 검색 분야의 공개 데이터 셋인 CAL500 [2], Emotions [7]를 활용하였다. 비교 대상 알고리즘으로는 Multi-label Nearest Neighbor (MLNN)를 사용하였다 [9]. 클래스 분류 성능을 비교하기 위해 다중 레이블 분류 정확도(Multi-label accuracy)를 활용하였다. 태그가 부족한 상황을 시뮬레이션하기 위해 부여된 태그 중 일부를 무작위로 삭제하고, 삭제된 비율에 따라 태그 불완전 비율(tag incompleteness)을 각 실험 결과에 표시하였다.

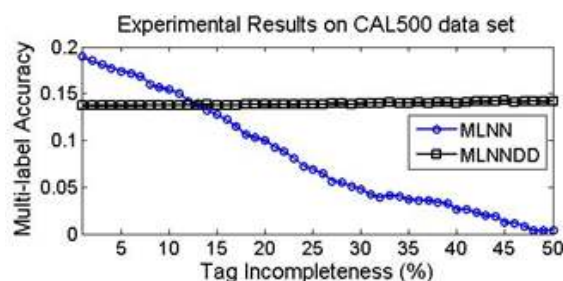
음악의 소셜 태그가 부족한 상황에 따른 다중 레이블 클래스 분류기의 성능을 그림 1에 나타내었다. 그림 1(a)와 1(b)에서 볼 수 있듯이, 제안하는 알고리즘은 태그 제거 비율에 거의 영향을 받지 않고 일정한 다중 레이블 분류 정확도를 보여줌을 알 수 있다. 실험 결과를 보았을 때, CAL500 데이터 셋에서는 85-90% 미만의 태그가 주어진 상황에서, Emotions 데이터 셋에서는 90-95% 미만의 태그가 주어진 상황에서 제안하는 알고리즘이 MLNN에 비해 더욱 좋은 성능을 보이는 것을 알 수 있다.

IV. 결론

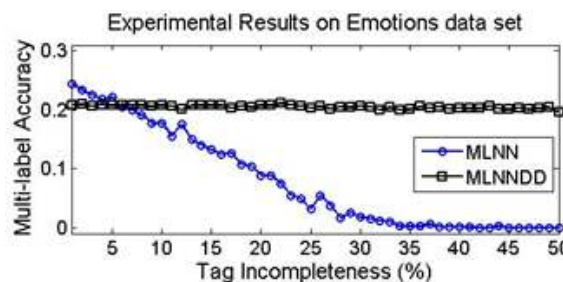
본 논문에서는 불완전한 소셜 태그가 부여된 상태에서의 음악 검색 및 추천을 위한 음악 태그의 예측 성능을 높이는 방법을 제안하였다.

사 사

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2014년도 문화기술연구개발지원사업의 연구결과로 수행되었음.



(a) CAL500 데이터 셋의 다중 레이블 분류 정확도



(b) Emotions 데이터 셋의 다중 레이블 분류 정확도

그림 1. CAL500 데이터 셋과 Emotions 데이터셋에서 태그 불완전 비율에 따른 MLNN과 제안하는 MLNDD의 성능 비교 도표
Fig. 1. Comparison results of MLNN and Proposed method in terms of Multi-label accuracy on CAL500 and Emotions datasets.

참고문헌

- [1] Z. Fu et al., "A Survey of Audio-based Music Classification and Annotation," IEEE Trans. Multimedia, 2011.
- [2] D. Turnbull et al., "Semantic Annotation and Retrieval of Music and Sound Effects," IEEE Trans. Audio Speech Lang. Process., 2008.
- [3] Z. Zhao et al., "Large-scale Music Tag Recommendation with Explicit Multiple Attributes," In Int. Conf. Multimedia, 2010.
- [4] F. Font et al., "Folksonomy-based Tag Recommendation for Online Audio Clip Sharing," In ISMIR, 2012.
- [5] Y. Song et al., "Do Online Social Tags Predict Perceived or Induced Emotional Responses to Music?," In ISMIR, 2013.
- [6] Y. Lin et al., "Exploiting Online Music Tags for Music Emotion Classification," ACM Trans. Multimedia Comput. Commun. Appl., 2011.
- [7] K. Trohidis et al., "Multi-Label Classification of Music into Emotions," In ISMIR, 2008.
- [8] D. Tax, "Data Description in Subspaces," In Int. Conf. Pattern Recognit., 2000.
- [9] M. Zhang and Z. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," Pattern Recognit., 2007.