

내용 기반 음악의 감성 추출 연구 동향 및 평가치 기반 청중 기호 분석

이중형, 김민욱, 진영영, 윤경로
 건국대학교 컴퓨터공학부
 {ljh0716, minuk, yychen, yoonk}@konkuk.ac.kr

Research Trends of Content-based Music's Emotion Extraction and Analysis of Audience's Preference based on Rating Information

Jonghyung Lee, Min-Uk Kim, Yingying Chen, Kyoungro Yoon
 School of Computer Science and Engineering, Konkuk University

요 약

본 논문에서는 내용 기반 음악의 감성추출 연구의 동향을 살펴보고, 평가치를 기반으로 청중의 기호를 분석해 본다. 청중의 기호 분석은 가능한 많은 사람이 공감 할 수 있는 데이터를 사용하고자, 최근 이슈가 되고 있는 「나는 가수다」 방송 프로그램의 음원 분석을 통해 이루어졌다. 데이터베이스는 총 7 번의 공연/경연마다 7 곡씩 순위가 매겨지므로 총 49 곡으로 이루어진다. 청중 평가단은 총 500 명의 엄격히 선발된 10대에서 50대, 각각 100명 씩으로 구성되었다. 특징값은 오픈소스인 MIRtoolbox를 이용해서, 총 376개의 특징값을 추출하여 평가치와의 상관관계를 구해 청중 평가단의 기호를 분석한다. 실험결과 총 376개의 특징값 중 10개의 특징값이 청중평가단의 기호와 상관관계가 있다는 것이 나타났다. 마지막으로 내용 기반 음악의 감성 추출 연구에서 보이는 감성의 주관적 인식성 및 심리학적 설명의 난해함을 줄이고자, 향후 과제로 내용 기반 및 평가치 기반의 시스템을 결합한 감성 기반 음악 추천 시스템을 제안한다.

1. 서론

음악은 인류가 존재할 때부터 지금까지 종교적, 정치적, 사회적 도구로 광범위하게 사용되어 왔다. 관습적으로 음악에 대한 정보는 제목, 작곡가, 가수, 발매 날짜 등으로 이루어져 왔으며, 대부분 이런 사전적 정보를 이용해서 음악을 추천하고 검색해왔다. 현대에는 물질적 풍요로움과 더불어 다양한 형태의 미디어가 폭넓게 보급되면서 콘텐츠 서비스 분야의 변화를 예고하고 있다. 이에 따른 기술 개발의 관점에서도 변화의 필요성이 증가되면서 감성을 접목시킨 연구들이 각광을 받고 있다[1].

음악의 중요하고 기본적인 기능 중 하나는 감성을 유발하는 것이다. 음악은 청자에게 웃기, 울기, 전율, 가슴 벅참 등 매우 다양한 정서적 반응을 일으키며, 대부분의 사람들은 경험적으로 이에 공감할 것이다. 최근에는 음악의 “무드”라는 새로운 타입의 메타데이터가 등장했으나, 감성 인식의 주관성 및 심리학적 설명이 난해하기 때문에 객관적인 데이터베이스 구축이 어렵고, 그나마 진행되었던 연구도 제각기 다른 감성 모델 및 분류기를 사용하여 비교 및 평가가 힘든 상황이다.

본 논문의 구성은 다음과 같다. 2장에서 내용 기반 음악의 감성 추출 연구의 동향을 살펴보고, 3장에서는 최근 이슈가 되고 있는 「나는 가수다」 방송 프로그램 음원의 평가치를 기반으로 청중 평가단의 기호를 분석해 본다. 4장에서는 내용 기반 및 평가치 기반의 시스템을 결합한 감성 기반 음악 추천 시스템을 제안하면서 결론을 맺는다.

2. 내용 기반 음악의 감성 추출 연구의 동향

내용 기반 음악의 감성 추출 연구는 감성 인식의 주관성 및 심리학적 설명의 난해함 때문에 객관적인 비교 및 평가가 힘들다. 그럼에도 현재 많은 관련 연구가 진행되고 있으며, 향후 인간의 감성 메커니즘이 밝혀진다면 더욱 향상된 결과를 얻을 수 있을 것이다. 내용 기반의 음악 감성 추출 연구는 보통 다음과 같은 단계를 거친다[2].

1. 감성 모델 도출
2. 데이터베이스 확보 및 분류
3. 특징 추출 및 선택
4. 패턴 인식의 과정

그림 1은 음악의 감성 추출 연구의 단계를 나타낸다. 본 논문에서는 「나는 가수다」 방송 프로그램[3] 음원의 평가치를 통해 어떤 저-수준의 특징 값이 청중의 기호/선호도에 영향을 미치는지를 알아보고자 했고, 10대부터 50대까지의 다양한 연령대 500명이 느낀 결과는 보편성이 있다고 가정하였다.

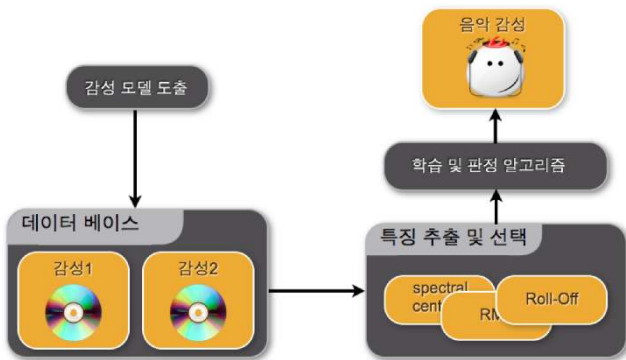


그림 1. 음악의 감성 추출 연구의 과정

2.1 무드와 감성

감성 기반 음악 연구들을 살펴보면 무드(Mood)와 감성(Emotion)이 엄밀한 구분 없이 혼란스럽게 섞여 쓰이곤 한다. 하지만 무드는 보편적으로 길게 지속되는 느낌인 반면, 감성은 비교적 짧은 시간 동안 느껴지는 것이다. 또한 감성은 쉬지 않고 변하지만, 무드에 비해 영향을 많이 주는 차이점이 있다. 이러한 개념적 차이로 심리학 분야에서는 인간의 반응에 초점을 두어 감성이란 단어를 주로 사용하고, 음악 정보 검색(MIR) 분야에서는 음악에 초점을 둔 메타데이터를 위한 무드라는 단어를 주로 사용한다[4]. 본 논문은 음악의 영향으로 사람이 받는 감성에 초점을 맞추어 감성이란 단어를 사용한다.

무드와 감성은 상당한 동질성으로 인해 서로 깊은 관련성을 갖지만, 그 차이점은 분명히 존재하며, 각각의 개념이 적합하게 사용되기 위해 구별의 명확함이 수반되어야 할 것이다.

2.2 감성 모델

기본 감성에 대한 견해는 학자들마다 의견이 다르다. Ekman 은 행복, 슬픔, 화, 두려움, 혐오, 놀람이 발견된다고 주장하고[5], Plutchik 은 분노, 두려움, 혐오, 놀람, 기쁨, 슬픔, 수용, 기대감 등을[6], Watson 은 두려움, 분노, 사랑의 3 가지 정서가 기본적이라고 보았다[7].

이 외에도 기본 감성에 대한 많은 학설이 있지만 음악 정보 검색 분야에서는 Hevner[8]와 Russell[9]의 모델을 주축으로 연구가 진행되어왔다. 그러나 최근에는 Hevner 모델이 오래 전에 구축되어 현대의 정서와 맞지 않는다는 이유로 잘 사용되지 않고, 음악 감성을 비교적 잘 군집화하고 있는 Russell 모델을 개선한 Thayer 의 감성 모델이 가장 널리 쓰이고 있다[10].

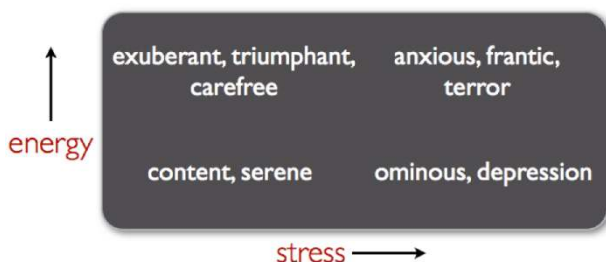


그림 2. Thayer 의 감성 모델

그림 2 는 Thayer 의 감성 모델이다. Thayer 모델은 세로(energy)축과 가로(stress)축으로 이루어져 음악 정보 검색 분야에서 특징 값들을 감성 모델에 사상(事象)하기 쉬운 2 차원 형태를 갖고 있다.

2.3 데이터베이스의 분류와 특징 추출 및 선택

IT 기술의 발전으로 콘텐츠의 양이 증가함에 따라 메타 데이터를 통해 관리하는 수동적인 방법은 많은 인적, 시간적 비용을 요구하고 있어 비효율적이다. 이를 개선하기 위해 메타데이터가 아닌 콘텐츠의 내용을 기반으로 특성을 파악하여 자동으로 분류, 추천 및 검색을 할 수 있는 연구가 많이 진행되고 있다[11]. 오디오의 특징 값은 저-수준의 음향적 특징 계층, 이를 바탕으로 추출한 중간-수준의 오디오 신호 계층, 그리고 특징 값을 의미적으로 통합한 고-수준의 모델로 표현 할 수 있다[12].

많은 특징 값 중 데이터베이스의 분류와 관련된 특징 값을 선택하는 방법은 또 다른 연구 분야로써 수 많은 알고리즘이 개발 및 적용되고 있다[13]. 데이터베이스의 분류에 따라 선택되는 특징 값은 달라지게 되는데, 사실 음악의 감성 추출 연구에서 손꼽히는 난관 중 하나는 객관적인 데이터베이스의 분류이다.

3. 「나는 가수다」 음원 분석을 통한 청중 평가단의 기호 분석

음악의 감성 추출 연구에서 신뢰성 있는 특징 값의 선택을 위한 방법으로 평가치 기반의 기호 분석 방법을 제안한다. 최종 목표를 위한 실험의 일환으로 「나는 가수다」 방송 프로그램의 음원 분석을 통해 청중 평가단의 기호를 분석해 본다.

3.1 데이터 베이스

본 실험에서는 감성의 주관적 인식성 및 심리학적 설명의 난해함을 줄이고 가능한 많은 사람이 공감할 수 있는 데이터를 사용하고자 「나는 가수다」 방송 프로그램의 음원을 데이터베이스로 사용하였다. 한번의 공연/경연에 7 명의 가수가 노래하고, 총 7 번의 공연/경연이 있었으므로 데이터 베이스는 모두 49 곡으로 이루어져 있다.

평가는 공연/경연 때마다 청중 평가단에 의해 순위가 매겨지는데 청중 평가단은 총 500 명의 엄격히 선발된 10 대, 20 대, 30 대, 40 대, 50 대 각각 100 명씩으로 구성되었다. 표 1 은 최근 경연의 곡명, 가수, 그리고 평가치를 나타낸다.

본 논문에서는 이러한 다양한 연령대의 500 명이 평가한 평가치를 통해 도출된 저-수준의 특징 값들은 보편적이고 신뢰할 수 있는 특징 값이라고 가정하였다.

표 1. 「나는 가수다」 최근 경연 결과

곡명	가수	평가
천일동안	옥주현	1 등
해야	윤도현	2 등

그대 내 품에	박정현	3 등
비상	JK 김동욱	4 등
주먹이 운다	이소라	5 등
네버 엔딩 스토리	김범수	6 등
편지	BMK	7 등

3.2 특징 추출

음원의 특징 추출은 오픈 소스인 MIRtoolbox 1.3.2[14] 를 사용하였다. 특징 값은 리듬(Rhythm), 음조(Tonal), 음색(Timbre), 강약(Dynamics), 변동(Fluctuation), 그리고 스펙트럼 형(Spectral)의 총 6 개의 고-수준 단계로 구성되었으며, 이에 따른 중간-수준 특징 값은 RMS, Peak, Centroid, Tempo, Zero-cross 등의 28 개의 단계로 구성하여 총 376 개의 특징 값을 추출하여 사용하였다. 그림 3 은 중간-수준과 고-수준의 특징 값을 나타낸 구성도를 나타낸다.

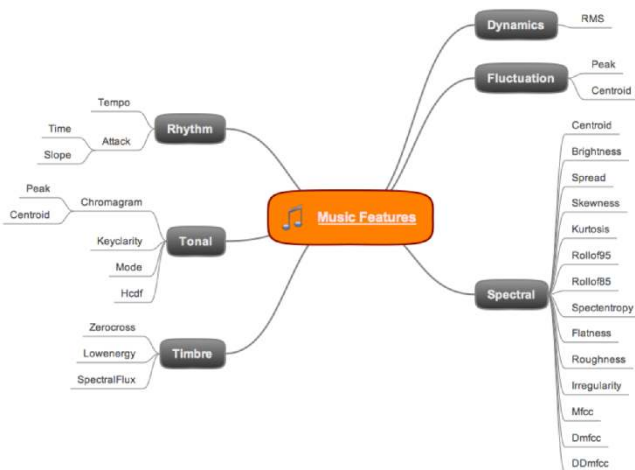


그림 3. 중간-수준 및 고-수준 특징 값 구성도

3.3 청중 평가단 기호 분석

음원에서 추출한 특징 값을 정규화(normalize)한 후 청중 평가단에 의해 매겨진 순위와의 다중상관분석(Correlation Analysis)을 실시한다. 상관분석은 확률론과 통계학에서 두 변수간에 어떤 선형적 관계를 갖고 있는지를 분석하는 방법이다. 이때 두 변수간의 관계의 강도를 상관계수(Correlation Coefficient)라 한다. X 와 Y 의 상관계수 ρ는 다음과 같이 구할 수 있다.

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

본 논문에서는 MIRtoolbox 의 mirmap 이라는 함수를 사용하여, 어떤 특징 값들과 평가치와의 상관관계를 분석했으며,

mirmap 의 함수의 사전 작업은 특징 값을 정규화(normalize)하는 것이다. 이 후 반복적인 작업을 거쳐 상관관계가 적은 특징 값들을 제거하게 된다[15].

그림 4 는 선별된 특징 값을 상관 관계가 높은 순서로 그래프로 표현한 것이며, X 축은 상관 관계를, Y 축은 p-value 를 나타낸다.

X 축을 나타내는 상관 관계 값은 양수와 음수로 나타날 수 있는데, 양수인 경우는 등수가 높아질수록 해당 특징 값도 증가하는 경향을, 음수인 경우는 등수가 높아질수록 해당 특징 값이 감소하는 경향이 있다는 것을 뜻한다.

Y 축을 나타내는 p-value 값은 유의확률 또는 유의수준이라고도 하며, 모평균이 실제로 같지만 잘못해서 기각할 확률을 나타낸다. 보통 1%, 5%, 10% 를 주로 사용하며, 논문에서는 5% 값을 사용하였다. 따라서 그림 4 에서 파란색 일수록, 즉 p-value 값이 작을수록 특징 값을 신뢰할 수 있다고 할 수 있고, 빨간색 일수록 그 반대의 성향을 가진다고 할 수 있다. 그림 4 에서 볼 수 있는 것처럼 X 축의 절대값이 작아질 수록, p-value 는 증가한다.

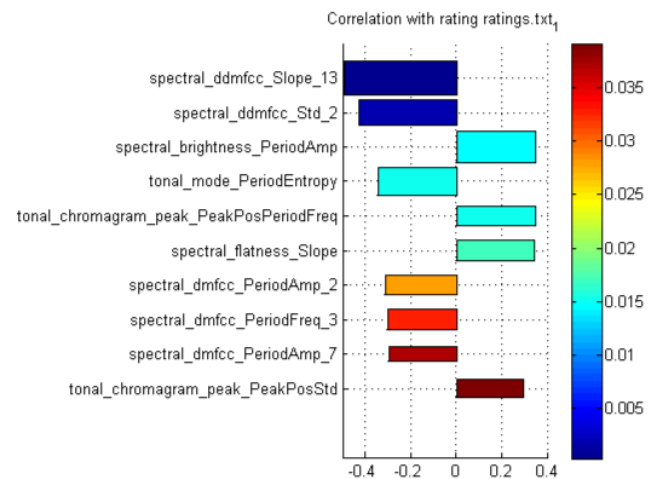


그림 4. 특징 값과 평가치 상관관계

예를 들어, 그림 4 의 Y 축에서 가장 아래 쪽에 위치하는 tonal_chromagram_peak_PeakPosStd 특징 값은 피크 값의 위치에 대한 표준편차를 나타내는데, 특징 값과 평가치 상관 관계(X 축)의 절대값은 비교적 낮고, 또한 p-value 는 가장 높은 값을 갖는다. 따라서 이 특징 값은 1 등부터 7 등 사이에서 변하는 특징 값의 상관 관계가 적고 신뢰성도 낮다고 볼 수 있다.

4. 결론 및 향후 과제

본 논문에서는 내용 기반 음악의 감성 추출 연구의 동향을 살펴보고, 「나는 가수다」 방송 프로그램의 음원 분석을 통해 청중 평가단의 기호를 분석하여 총 376 개의 특징 값 중 평가치와 관련 있는 10 개의 특징 값을 선별하였다.

물론 청중 평가단 모두가 눈을 감고 음악을 청취하고 평가를 내리진 않는다. 하지만 본 논문에서는 500 명의 다양한

연령대의 청중 평가단이 내린 평가치는 보편적으로 감성에 영향을 주는 특징 값을 추출하는데 도움을 줄 수 있다고 가정하였다.

또한 mirmap 을 통한 특징 값과 평가치의 상관관계는 선형적으로 변하는 관계만 나타내주고 있다. 따라서 앞서 살펴본 10 개의 특징 값들은 청중 평가단이 느끼는 선호도/기호에 따라 선형적으로 감소 또는 증가하는 특징 값들이라고 할 수 있다.

사실 사람의 감정 메커니즘은 매우 복잡하여 음악을 청취함에 있어 딱 하나의 감정 만을 표상하지는 않는다. 슬픔과 공포를 동시에 느낄 수도 있고, 행복과 평화로움을 동시에 느낄 수도 있다. 심지어는 2 개 이상의 감정을 동시에 느낄 수도 있다. 하지만 위에서 언급한 음악 감정 연구의 방법에 따라 이런 결론을 도출하기란 매우 어렵다. 설령 실험자가 음악을 듣고 3~4 개의 감정에 해당 수치를 적어 넣는다 해도 쉬운 일은 아닐 것이다.

따라서 향후 과제로 음원을 분류하고 특징 값을 선정하는 방법에 본 논문에서 제안한 평가치 기반 청중 기호 분석 방법을 적용한 실험을 진행할 예정이다. 각 감정마다 평가치의 상관관계를 기반으로한 특징 값을 추출 할 수 있으면, 감정과 연관된 특징 값의 경향 분석이 가능할 것이다. 그러면 하나의 음원에서 각 감정의 특징 값 경향을 기반으로 한 복합적인 감정 사상이 가능할 것이라 예상된다.

사사

본 연구는 서울시 산학연 협력사업(SS100006)의 지원에 의해 수행되었습니다.

참고 문헌

- [1] Y.-H. Yang , C.-C. Liu and H. H. Chen "Music emotion classification: A fuzzy approach", Proc. ACM Multimedia, p.81 , 2006.
- [2] Han B, Rhon S, Jun S, Hwang E (2010) Music emotion classification and context-based music recommendation. Multimed Tools Appl (MTAP) 47 (3):433-460
- [3] MBC '우리들의 일밤-나는 가수다' homepage, www.imbc.com/broad/tv/ent/sundaynight/
- [4] Hu, X. 2010. Music and mood: where theory and reality meet. In Proceedings of iConference 2010.
- [5] Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6, 169-200.
- [6] Plutchik, R. (1962). The emotions: Facts, theories, and a new model. New York: Random House.
- [7] Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.
- [8] Hevner, K. 1936. Experimental studies of the elements of expression in music. American Journal of Psychology, 48:246-68.
- [9] Russell, J. A. 1980. A circumplex model of affect. Journal

of Personality and Social Psychology, 39: 1161-1178.

- [10] R. E. Thayer, The Biopsychology of Mood and Arousal, 1989.
- [11] E. M. Schmidt, D. Turnbull, and Y. E. Kim, "Feature selection for content-based, time-varying musical emotion regression," in MIR '10: Proc. of the Intl. Conf. on Multimedia Information Retrieval, Philadelphia, PA, 2010, pp. 267-274.
- [12] 임재덕, 콘텐츠 분류를 위한 오디오 신호 특징 추출 기술, 전자통신동향분석 제 24 권 제 6 호 2009 년 12 월, pp. 121-132.
- [13] S. B. Kotsiantis, "Feature selection for machine learning classification problems: a recent overview," Artificial Intelligence Review, May. 2011.
- [14] MIRtoolbox, <http://users.jyu.fi/~lartillo/mirtoolbox/>
- [15] O. Lartillot, T. Eerola, P. Toivainen, and J. Fornari. Multi-feature modeling of pulse clarity: Design, validation, and optimization. In Proceedings of the International Symposium on Music Information Retrieval, 2008.