

딥러닝 기반 모창가수 구분에서 특징 추출 방법 - 히든싱어 컨테스트

채다인, 서다빈, 이영경, 최대선°

공주대학교 의료정보학과

Feature Extraction Method for Deep Learning-based Singer Identification - Hidden Singer Contest

Chae Da-in, Seo Dabin, Young-kyoung Lee, Daeseon Choi°

Department of Medical Information, Kongju University

sunchoi@kongju.ac.kr

요약

본 논문에서는 원가수와 모창가수를 구분하는 방법과 여기에 사용되는 특징 추출 방법을 제안한다. 구분 방법은 컴퓨터에게 원가수의 원곡 음원을 학습시킨 후 원가수와 모창가수 음원 5개를 비교하여 가장 유사도가 높은 것을 원가수로 분류한다. 특징 추출은 MFCC를 사용하는데 파라메터인 주파수 대역과 윈도우 사이즈에 따라 성능이 달라지므로, 파라메터 선택을 위해 30곡의 테스트 셋을 구성하여 실험을 실시하였다.

1. 서론

히든싱어는 원가수(original singer)와 모창자(mocking singer)를 구분하는 내용을 담은 JTBC 방송사의 TV 프로그램이다[1]. 1 명의 원가수와 4 명의 모창자가 등장하여 블라인드로 1 곡의 노래를 나눠 부르고 패널들이 모창자를 구분하는 투표를 하는데 모창자들이 잘 훈련되어 있기 때문에, 패널들도 모창자와 원가수를 구분하지 못하는 경우가 많다.

최근 이세돌 9 단과 알파고의 바둑 대결 등 많은 분야에서 인공지능이 인간의 능력을 능가하는 경우가 속속 등장하고 있다. 특히, 패턴인식 분야에서 얼굴인식의 경우 인공지능이 인간의 정확률을 능가하는 97.5%의 정확률(accuracy)을 달성한 것으로 보고되고 있다[2].

본 논문에서는 원 가수와 모창 가수를 구분하는 방법과 여기에 사용되는 특징추출 방법을 제안한다. 특징추출은 MFCC의 주파수 대역과 윈도우 사이즈를 변경해가며 이에 따른 성능을 알아본다.

구분방법은 컴퓨터에게 원가수의 원곡 음원을 학습시킨 후 원가수와 모창가수 음원 5 개와 비교하여 가장 유사도가 높은 것을 원가수로 분류하도록 한다. 특징 추출시 파라메터 선택은 30 곡의 테스트 셋을 구성하여 최적의 파라메터를 선택한다.

본 논문은 2 장에서는 관련연구, 3 장에서는 제안방법, 4 장에서 실험 구성과 결과를 제시하고 5 장에서 결론을 맺는다.

2. 관련연구

가수 인식(Artist Identification)은 장르 분류, 무드 분류, 악기 인식, 자동 주석 등과 함께 음악 검색(Music Information Retrieval)의 한 분야로 오랫동안 연구되어 왔다[3]. 기존 연구의 정확률은 75% 수준이다[4]. 그러나 우리가 아는 범위에서는 모창자와 원가수를 구분하는 연구는 없었다.

원가수의 음원을 학습시킨 후 모창자의 음원이 포함된 총 5 개에 음원에 대하여 원가수를 찾는 문제는 화자인증과 유사하다. 그러나 화자인증 연구 분야에서도 흉내낸 목소리를 구분하는 연구는 없었다. 한편, 노래를 대상으로 하는 가수인식(singer identification) 문제는 화자인식(speaker recognition) 보다 훨씬 어렵다. 그 이유는 원가수의 목소리 외에 노래의 반주 음악이 포함되어 있기 때문이다[3].

3. 제안방법

본 논문에서 제안하는 모창가수 구분을 위한 특징 추출 방법과 이를 적용한 원가수 구분 과정이 그림 1에 나타나 있다. 구분은 학습과정(Training Phase)과 테스트과정(Test Phase)으로 구성된다. 학습과정은 다음과 단계로 구성된다.

- 음원 추출 : 원가수의 원곡 음원에서 모창음원과 비교할 부분을 추출한다. 이 때 공백을 제거 한다. 이때 사용되는 원곡음원은 모창 프로그램

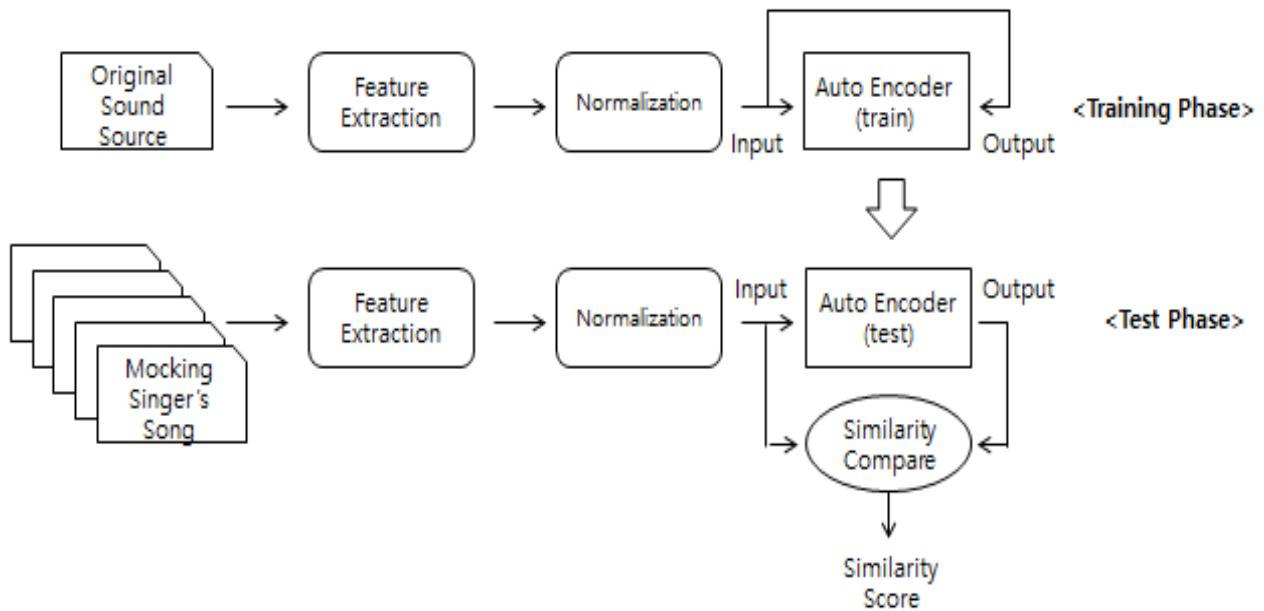


그림 1. 학습 및 테스트 구조

에 등장하는 원가수의 것이 아닌 기준에 발표된 원곡을 의미한다.

- **특징 추출 (Feature Extraction)** : MFCC 기법을 이용하여 각 음원에서 특징 값을 추출한다. MFCC는 음성인식과 가수인식에서 많이 사용되는 특징으로 수십 ms 정도의 음성신호에 대한 주파수 대역 별 에너지 값을 의미한다. 따라서, 특징 벡터가 추출된다. MFCC를 적용할 때 가수의 성별에 따라 주파수 대역을 다르게 필터링하여 정확도를 향상시킬 수 있다.
- **정규화 (Normalization)** : 특징 벡터가 추출되면 정규화를 수행한다. 각 음원마다 노래음량이 다르기 때문에 정규화가 필요하다. 그런데, 음성인식 분야에서는 일반적으로 사용하는 Cepstral Mean Normalize 기법을 적용하면 이는 MFCC 이후 단계에서 이루어진다. Cepstral Mean Normalize 방법은 평균(mean)을 뺀 후 분산(variance)으로 나누는 것이다.
- **학습 (Train):** 정규화된 특징 벡터를 이용해 Auto-encoder[5]를 학습시킨다. Auto-encoder는 신경망의 일종으로 입력값에 따라 출력값을 재생산 할 수 있도록 신경망을 학습시키는 기법이다. 그림 2는 Auto-encoder 구조를 나타낸다. 그림처럼 encoder 과 decoder 를 붙여 놓으면 주어진 입력값을 그대로 재현하는 기능을 수행한다. 이때 학습데이터에 따라 값의 재현도가 달라진다. 즉 학습된 데이터와 유사한 패턴의 입력 값일수록 출력 값의 재현도도 높아지게 된다. 제안방법에서는 학습은 원가수의 원곡 음원에 대해서만 이루어진다. 즉, 컴퓨터에게는 원가수의 노래 소리만을 학습시킨다.

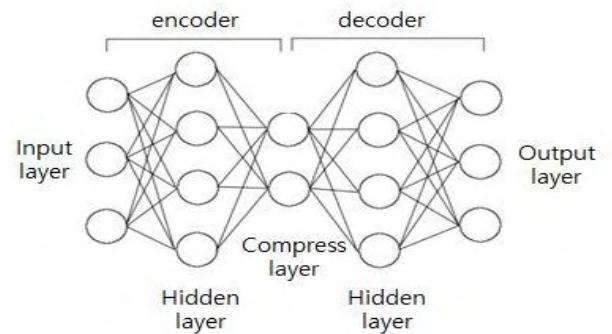


그림 2. Auto-encoder 구조

학습과정을 거친 Auto-encoder를 이용해 모창음원을 입력하여 원가수를 찾는 테스트 과정은 다음과 같다.

- **음원추출 :** 원가수가 포함된 5 개의 히든싱어 모창에서 노래 부분을 추출한다.
- **특징추출 및 정규화 단계 :** 학습과정과 동일하다.
- **구분 (Test):** 구분 단계에서는 원 가수와 모창자를 포함한 총 5 명의 모창음원의 특징 벡터를 Auto-encoder에 입력한다. 그러면 Auto-encoder는 출력값을 도출하는데 이 출력값과 입력값의 차이를 계산하는 방법으로 유사도를 구한다. 차이 계산은 Mean Square Error (MSE)를 사용한다. 최종적으로 이 차이가 가장 적은 음원을 원가수의 음원으로 선택하게 된다.

4. 실험

4.1 실험 구성

제안 방법의 특징 추출 파라미터를 결정하고, 성능을 측정하기 위한 실험 구성을이다.

- **테스트 셋**

히든싱어 방송 프로그램 30 편을 이용하여 30 개로 구성했다. 각기 원가수 원곡 1 개와 원가수와 모창자를 포함한 총 5 개의 모창 음원이 포함된다. 원가수 원곡은 기존에 발표된 음원에서 추출하였고 모창은 30 명의 가수에 대한 각 5 명의 모창 음원을 방송프로그램으로부터 추출하였다.

- **특징 추출 - 파라미터 설정**

MFCC 파라미터의 두 가지를 변경해가며 실험하였다. 첫 번째 파라미터는 남녀 성별에 따른 로우 주파수 커팅 지점이다. 남성 가수와 여성가수의 노래 주파수 대역이 다르고, 반주 등에 의한 구분을 막기 위해 주파수를 커팅한다.

두 번째 파라미터는 윈도우 사이즈이다. 윈도우 사이즈 란, 수십 ms 동안의 음원의 특징을 추출함에 있어 그 지속시간을 의미한다. 음원의 sampling rate 이 44.1KHz 이므로 윈도우 사이즈가 256 인 경우 0.5ms 를 의미한다.

- **Auto-encoder**

Auto-encoder 의 노드갯수는 다음과 같다. 입/출력 벡터 사이즈가 43 인 상태에서 hidden layer 의 수는 25 개 compress layer 수는 30 개로 하였다..

- **실험환경**

- 운영체제 : Ubuntu 15.04
- 언어 : Python 2.7
- deep learning 라이브러리: Keras over theano
- GPGPU : Nvidia GTX Titan X (GPU 3,072 개)

한편, 인간과의 비교를 통해 컴퓨터 능력의 수준을 판단하기 위해, 지원자들을 대상으로 원가수 원곡을 들려준 후 5 개의 모창을 들려주고 이중 원가수를 찾는 퀴즈를 열었다. 퀴즈는 1등 5 만원 등, 상금을 걸었고 대학생 지원자 24 명이 참가하였다.

4.2 실험 결과

남녀 주파수 필터 별 구분 성능 실험결과는 그림 3 와 같다. 남자와 여자의 목소리의 주파수영역을 조절하며 최대의 정확률의 조건을 찾아낸 것이다. 남자는 50 Hz, 75 Hz, 100 Hz, 125 Hz 의 4 개 주파수 커팅지점을 시험했고, 여성가수는 125 Hz, 150Hz, 175 Hz 3 개의 주파수 커팅 지점을 실험하여 남자는 총 18 명의 지원자 중 12 개를 맞춘 75Hz 가 66%로

최고의 정확률을 보였으며, 여자는 총 12 명의 지원자 중 6 개를 맞춘 150Hz 에서 50%의 최고치가 나왔다.

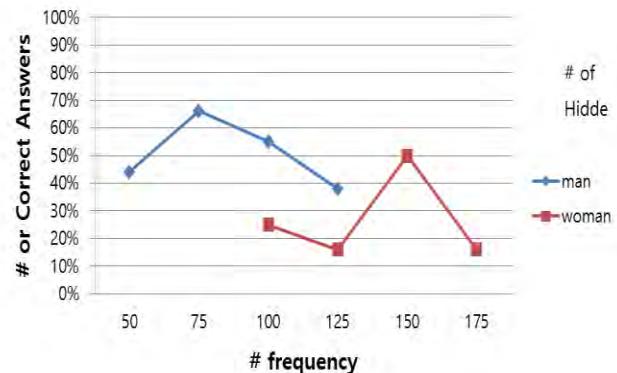


그림 3. 남녀 주파수 필터 별 정답율

윈도우 사이즈 실험결과는 그림 4 에 나와있다. 윈도우 사이즈 256, 512, 768, 1024 4 개의 값을 시험했는데, MFCC 라이브러리의 한계로 실험 가능한 최소값은 256 이었다. 실험결과 결과는 256 에서 최고의 정확률이 나오고 점차 감소하는 추세가 나타남을 알 수 있다.

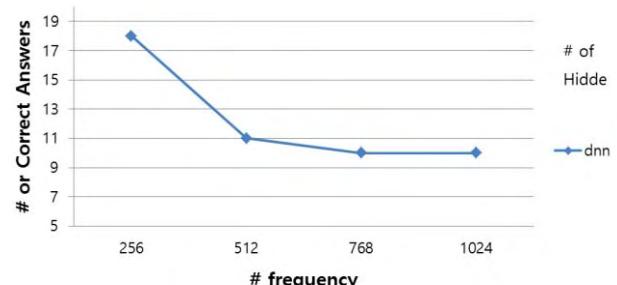


그림 4. MFCC 윈도우 사이즈 별 성능

인간과 컴퓨터의 성능 비교를 위한 퀴즈 결과, 지원자 24 명 중 최고 성적을 거둔 참가자는 컴퓨터와 동일한 18 개를 정답으로 맞추어 60%의 정확률을 보였다. 참가자들의 평균 성적은 10.5 였다.

인간 우승자와 컴퓨터의 답안 결과표를 부록으로 첨부하였다.

5. 결론

Auto-encoder 기반의 컴퓨터는 제안 방법에 따라 주파수 영역과 윈도우 사이즈를 변경해 실험한 결과 최고 60%의 정확률을 보였는데, 이는 24 명의 퀴즈 참가자들 중 최고 성적과 동일하며 퀴즈 참가자의 평균값 10.5 보다는 훨씬 높은 성능이다.

향후 연구 방향으로, 학습 데이터 확장을 들 수 있다. 현재 학습 방법은 원가수의 원곡 1 곡만을 학습하는 것인데 사람은 원가수의 다른 노래도 학습이 되어있는 경우가 많아, 컴퓨터에게도 원가수의 다른 곡을 학습시킴으로써 컴퓨터의 성능이 향상될 것으로 기대된다.

본 연구는 고도로 훈련한 모창자를 구분하는 것으로, 화자인증의 보안성을 높이기, 음성 대조를 통한 포렌식 등에 널리 활용될 수 있을 것이다.

6. 참고문헌

- [1] 히든싱어, JTBC,
http://home.jtbc.joins.com/Event/Event.aspx?prog_id=PR10010362&menu_id=PM10032984.
- [2] ExtreamTech, “Facebook’s facial recognition software is now as accurate as the human brain, but what now?”, 2014
- [3] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, “A Survey of Audio-Based Music Classification and Annotation”, in IEEE Trans. on Multimedia, vol.13, no.2, 2011
- [4] J. Shen, J. Shepherd, B. Cui, KL TAN, “A Novel Framework for Efficient Automated Singer Identification in Large Music Databases”, ACM Trans. Inf. Syst., vol. 27, no. 3, pp.1-31, 2009
- [5] Autoencoder,
<https://en.wikipedia.org/wiki/Autoencoder>

[부록] 가수 별 사람 1 위와 컴퓨터의 결과

가수	가수 성별	인간 우승자	컴퓨터
박정현	여	O	O
성시경	남	X	X
조관우	남	X	O
이수영	여	X	X
장윤정	여	O	X
박상민	남	O	O
김종국	남	O	O
윤민수	남	O	O
김건모	남	X	X
이은미	여	O	X
조성모	남	O	O
김범수	남	O	O
주현미	여	X	X
아이유	여	X	O
박진영	남	X	X
김광석	남	O	O
이선희	남	O	X
환희	남	X	X
태진아	남	O	O
인순이	여	O	O
윤종신	남	O	O
김태우	남	O	X
보아	여	X	O
민경훈	남	X	O
소찬휘	여	X	X
김정민	남	X	X
김연우	남	O	O
거미	여	O	O
변진섭	남	O	O
백지영	여	O	O