

호흡기 소리 분석을 통한 질병유무 판단과, 질병분류

김기홍[○], 김동익, 이가경, 이상민

세종대학교 컴퓨터공학과, 무인이동체공학과

kh9342@naver.com, diakes@naver.com, eeuw95@naver.com, diasy3277@naver.com

요 약

본 논문에서는 청진소리를 이용하여 질병유무 판단과 질병분류를 위한 딥러닝 기반 기법을 소개한다. 청진소리 학습 방법으로 음성특징을 스펙트럼 이미지로 추출하여 CNN 을 이용하여 학습하는 방식과, 음성특징을 데이터로 추출하여 SVM 을 이용하여 학습하는 방식을 제안한다. 각각의 방식을 이용하여 질병의 유무 판단과 질병이 분류되는 것을 확인하였다.

1. 서론.

1.1 배경

최근 4 년간 미세먼지와 황사로 인하여 호흡기 질환 환자수가 2 배 급증하였다. 이와 같은 경상의료비는 꾸준히 증가하고 있으며 이와 관련하여 의료비 부담 또한 증가하고 있다. 해당 문제를 해결하고자 디지털 헬스케어가 각광받고 있으며 본 프로젝트 또한 호흡기 소리를 통하여 질병을 분류함으로써 디지털 헬스케어 시장의 성장에 도움이 되고자 본 주제를 선택하였다.

1.2 데이터 셋

ICBHI 2017 Challenge Respiratory Sound 데이터를 이용하였다. 데이터는 ESSUA 연구실과, D. Pedro, Aveiro 병원에서 측정하였으며 총 126 명, 7 개질환과 건강한 사람 청진소리 920 개를 측정하였으며, 각 데이터에서 2750 개의 wheezes 과 crackles 를 측정하였다.

1.3 기존연구와 차별성

기존연구에서 청진소리의 wheezes 와 crackles 부분의 특징 이미지추출 후 CNN 을 이용하여 학습하였다. 질병유무 판단 정확도는 84% 질병분류 wheezes, crackles, wheezes 와 crackles 가 모두 있는 경우, 모두 존재하지 않는 경우 총 4 가지의 경우로 분류한 경우 약 70%의 정확도를 보였다.

본 프로젝트에서 기존 연구와 같이 질병의 유무 판단과 질병분류를 진행하였다. 하지만 질병 분류에

있어 기존연구와 다르게 질병에서 발생하는 특징인 wheezes 와 crackles 를 판단하지 않고 신호 자체를 하나의 질병으로 지정하여 질병분류를 진행하였다. 이와 같이 진행한 이유로는 호흡기 질환에서 발생하는 소리는 wheezes 와 crackles 뿐만 아니라, rhonchi, pleural friction 과 같은 소리도 발생하기 때문에 질병과 데이터를 일대일 매핑하였다.

2. 특징 이미지를 이용한 CNN 학습방법

2.1 특징 추출

특징추출의 방식으로 청진데이터의 스펙트럼 이미지를 생성하였다. 방식으로 STFT, MFCC, MFCC Delta 방식중 MFCC 와 MFCC Delta 방식을 선택하였다. STFT(Short Time Fourier Transform)는 음성파일의 전체길이에 비례하여 타임프레임을 설정하여 이미지의 크기가 다르게 생성되기 때문에 특징추출방식에서 제외하였다.

MFCC(Mel Frequency Cepstral Coefficient)는 총 6 개의 단계를 거쳐서 음성신호를 스펙트럼 이미지로 변환한다

첫번째 단계는 음성 신호를 일정 구간을 나눠준다. FT(Fourier Transform)적용시 시간정보가 사라지는 것을 방지하기 위해서 일정 구간인 프레임으로 음성신호를 나눠준다.

두번째로 프레임으로 나눠진 음성신호를 주파수의 분포를 추출하기 위해 푸리에 변환을 한다.

세번째 사람의 청각의 특징, 저주파 영역에 민감하고 고주파 영역에 둔감한 특징을 이용한 Mel Filter 를 적용한다.

네번째 Mel Filter 를 적용한 데이터에 로그를 취

함으로써 소리의 크기로 인한 에너지 변화를 줄일 수 있다.

다섯 번째, DCT(Discrete Cosine Transform)를 적용하여 Mel Filter 적용에 있어 발생한 상관관계를 제거한다.

마지막으로 26 log filter bank 를 적용하여 음성 인식을 못하는 12~13 개의 부분을 제거함으로써 소음을 제거한다.

MFCC Delta 는 MFCC 적용 후 MFCC Delta, 미분을 함으로써 음성의 변화율을 스펙트럼 이미지를 생성한다. Delta 를 두 번 적용하는 경우 음성 데이터 변화의 가속도를 스펙트럼 이미지를 추출할 수 있다. 청진 데이터는 일정한 간격으로 반복되는 데이터이기 때문에 Delta 를 한번만 적용하였다.

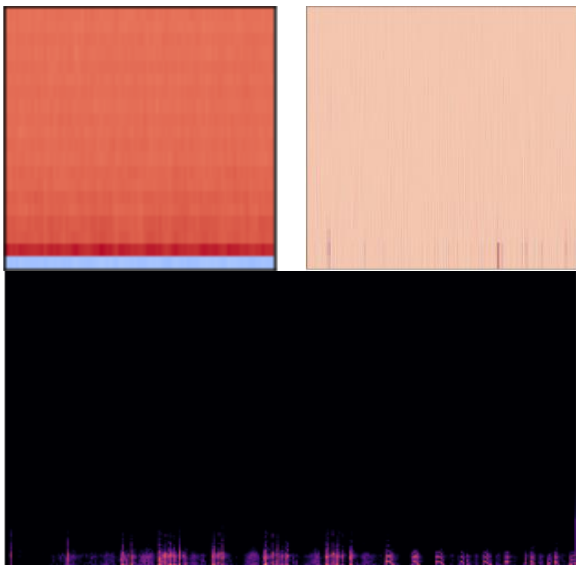


그림 1. MFCC, MFCC Delta, STFT 를 적용한 스펙트럼 이미지

2.2 CNN 학습

CNN(convolutional Neural Network)는 convolution 을 이용해서, 컨볼루션을 이미지 패치를 원본이미지로 부터 커널 사이즈만큼 가져와서 이미지 패치와 컨볼루션 커널을 연산한다. 커널은 점점 더 주어진 이미지를 필터링해 주게 됨. Cnn 을 3 개의 layer 계층과 1 개의 fc 를 사용해서 구현함. 특징 추출된 training_rate=0.8 인 112X112, 56X56, 28X28, 14X14 사이즈 데이터를 batch_size=100 learning_rate=0.001 을 적용하여 질병 유무와 질병 분류를 시도하였다.

질병 유무 판단을 라벨링 하여 0~1 까지 총 2 개의 class 으로 구분하였고, 질병 분류 0~7 까지 총 8 개의 class 으로 구분하였다

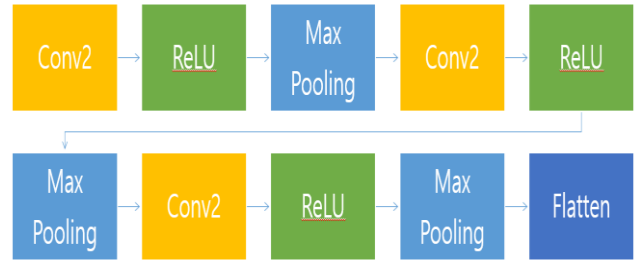


그림 2. CNN 레이어 구성도

3. 데이터를 이용한 SVM 학습방법

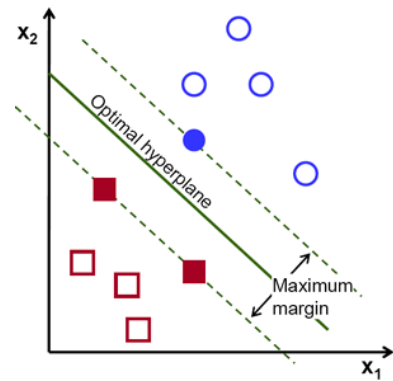


그림 3. SVM 학습과정

지도 학습 방식의 대표 분류 기법인 SVM 을 이용하여 질병 유무 판단, 분류 작업을 진행하였다. SVM 은 데이터를 벡터공간으로 표현한 후 서포터 벡터간 거리를 최대화하는 방식으로 데이터를 분류한다. 선형분류와 더불어 비선형 분류에도 자주 사용되며 비선형 분류를 하기 위해 데이터를 고차원 특징 공간으로 만드는 작업이 요구된다. 이러한 작업은 kernel function 을 통해 특징 공간을 접어버리거나 꼬아버려 비선형 데이터를 선형으로 분류할 수 있게 한다. 학습법의 장점으로는 차원수가 데이터 수보다 많을 때 효과적이다. 단점으로는 데이터가 너무 많으면 속도가 느리고 메모리 소모가 크다.

3.1 특징 추출

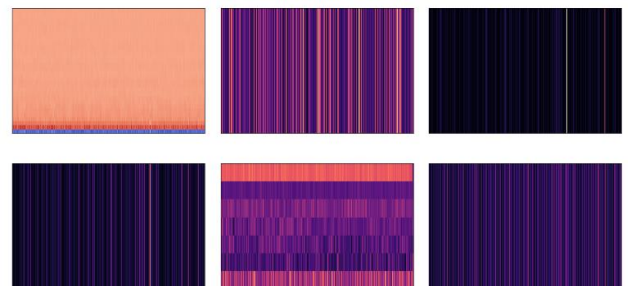


그림 4. SVM 학습을 위한 특징 추출 이미지(zro)

crossing rate, spectral roll off, spectral centroid, spectral contrast, spectral bandwidth)

CNN 학습에서 MFCC 를 통해서만 특징을 추출한 것과 달리 SVM 에서는 zero crossing rate, spectral roll off, spectral centroid, spectral contrast, spectral bandwidth 등 6 가지 방식이 사용됐다.

Zero crossing rate 는 음성 신호의 smoothness 를 측정하여 신호를 구별하는 방법이다. 소리에 따라 파형의 형태가 다르기 때문에 smoothness 를 측정하여 소리를 구분할 수 있다.

Spectral roll off 는 스펙트럼 magnitude 분포의 80%가 집중되어 있는 주파수를 나타낸다. 스펙트럼의 형태와 낮은 주파수 영역에 신호의 에너지가 얼마나 집중되어 있는지 보여준다.

Spectral Centroid 는 STFT 의 magnitude 스펙트럼의 중심을 뜻한다. Centroid 는 스펙트럼의 형태를 측정하는 방법 중의 하나이다.

Spectral Contrast 매 프레임마다 6 개의 구역으로 나누어 스펙트럼의 peak, valley 의 차이점을 계산한다.

Spectral bandwidth 는 주파수의 대역폭을 측정하는 방법 중 하나이다.

각 음원파일마다 6 가지 특징 값을 정규화 후 정렬과 통합과정을 진행하여 특징을 추출하였다.

3.2 차원축소

특징이 많으면 기계학습 모델이 잘 훈련되지 않거나 과적합을 일으키고, 훈련된 모델을 해석하여 용이한 정보를 얻기 어렵다. 또한 고차원 데이터는 시각화가 불가능하기에 분석결과를 공유하는 것 또한 쉽지 않다. 따라서 특징추출의 보편적인 방법인 주성분분석(PCA)방법을 사용하여 주어진 특징들을 조합하여 새로운 특징 값을 계산하도록 하였다

3.3 RBF 커널 사용

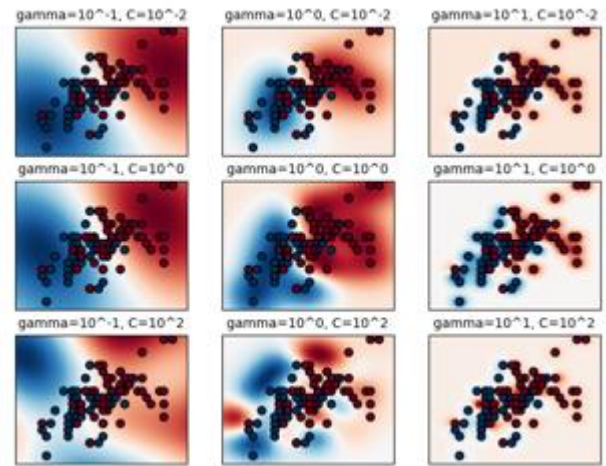


그림 5. Cost, Gamma 적용

선형으로 구분하지 못하는 구조를 RBF 커널을 통해 데이터를 변환하여 분류할 수 있다. RBF 커널은 Cost, Gamma 값을 parameter 로 받으며 Overfitting 을 막기 위해서는 위 값들의 조절이 필요하다. C 의 값이 작을 때는 제약이 큰 모델을 만들고 각 벡터포인트의 영향력을 적게 받는다. 반대로 C 값이 증가할수록 각 벡터포인트들은 모델에 큰 영향을 주며 결정 경계를 휘어서 정확하게 분류한다. Gamma 값은 커질수록 하나의 벡터포인트에 더 민감하게 반응한다.

3.4 SVM 학습

6 가지 특징 추출로 뽑아낸 데이터를 음원별로 1D 벡터에 데이터를 갖도록 했다. 총 210 개 특징을 추출하여 모든 음원에 대한 정보를 담은 테이블을 만들고 train 과 test 에 비율을 8:2 로 나눠 학습, 검증을 진행했다. 이후 train 과 test 데이터에 대한 정규화 과정을 거치고 PCA 차원축소 과정을 거쳤다. 차원축소가 적용된 데이터를 가지고 RBF 커널을 통해 비선형 데이터를 선형분류 했다. 커널의 C, G 값을 찾아주는 방법으로는 grid 기능을 이용하여 가장 높은 성능을 보여주는 C, G 값을 바로 찾고 이를 적용하도록 했다.

4. 실험 결과 및 분석

4.1 CNN 결과

920 개의 데이터셋을 기준으로 training rate=0.8, batch size=100, learning rate=0.001 을 설정하여 이미지 사이즈, 특징추출 방식별 CNN 학습 질병유무 판단, 질병분류 판단 정확도를 측정하

였다.

표 1. CNN 질병유무 판단 결과

| size | mfcc | mfcc_delta | avg |
|---------|------|------------|-------|
| 112X112 | 97.8 | 95.9 | 96.85 |
| 56X56 | 96.7 | 93.9 | 95.3 |
| 28X28 | 95.1 | 98.3 | 96.7 |
| 14X14 | 97.2 | 96.1 | 96.65 |

표 2. CNN 질병분류 판단 결과

| size | mfcc | mfcc_delta | avg |
|---------|------|------------|-------|
| 112X112 | 82.2 | 86.8 | 84.5 |
| 56X56 | 84.1 | 88 | 86.05 |
| 28X28 | 84.7 | 86.4 | 85.55 |
| 14X14 | 82.6 | 86.4 | 84.5 |

질병유무 판단의 경우 96.38%의 정확도를 보였고 질병분류의 경우 85.15%의 정확도를 보였다. 질병분류 정확도 향상을 위하여 training rate, batch size, learning rate, training, test 데이터셋 구성의 변화를 적용하여 보았다.

표 3. CNN 질병분류 Training_rate변경 결과 평균

| size | 0.7 | 0.8 | 0.9 | avg |
|---------|-------|-------|---------|-------|
| 112X112 | 84.9 | 84.5 | 84.2 | 84.53 |
| 56X56 | 85.45 | 86.05 | 89.1 | 86.87 |
| 28X28 | 85.95 | 85.55 | 82 | 84.5 |
| 14X14 | 86.3 | 84.5 | 83.75 | 84.85 |
| avg | 85.65 | 85.15 | 84.7625 | |

이미지 사이즈와 training rate 에 따른 변경 값을 비교하여 56x56 이미지, training rate 0.8 이 가장 큰 값을 보였다.

표 4. CNN 질병분류 Learning_rate변경

mfcc, mfcc delta 결과 평균

| size | 0.01 | 0.001 | 0.0001 | Avg |
|------|------|-------|--------|-----|
|------|------|-------|--------|-----|

| | | | | |
|-------|------|------|------|------|
| 56X56 | 85.5 | 85.5 | 85.5 | 85.5 |
|-------|------|------|------|------|

56x56 이미지와 training rate=0.8 을 고정후, Learning rate 의 값을 변경하였으나 큰 변화는 확인하지 못했다.

표 5. CNN 질병분류 Batch_size 변경

mfcc, mfcc delta 결과 평균

| size | 100 | 200 | 300 | 400 | avg |
|-------|------|------|------|------|------|
| 56X56 | 84.6 | 84.6 | 84.6 | 84.6 | 84.6 |

56x56 이미지와 training rate=0.8 learning rate = 0.001 고정 후 batch size 값을 변경하였으나 큰 변화는 확인하지 못했다.

표 6. CNN 질병분류 train, test set 변경 결과

| | Case1 | Case2 | Case3 | avg |
|-------|-------|-------|-------|-------|
| 56X56 | 88 | 85.8 | 86.35 | 86.72 |

56x56 이미지, training rate=0.8, learning rate=0.001, batch size=100 고정후 test, train 의 데이터 셋 구성을 변경하였다. 변경방식은 random 함수를 이용하여 구성하였다. 평균 86.72% 정확도 향상을 위한 조건변경 전 평균 85.15%로 약 1.57%향상을 보였다.

4.2 SVM 결과.

질병유무 판단의 경우 Train data 개수를 100 개, 200 개, 300 개, 400 개로 바뀌가며 정확도를 측정하였습니다. 질병분류 문제의 경우 Train data 를 100 개 ~ 800 개까지 100 개씩 늘려가며 정확도를 측정하였습니다.

표 7. SVM질병유무 판단 결과

| Train Data 수 | 정확도 |
|--------------|-------|
| 100 개 | 0.95% |
| 200 개 | 0.95% |
| 300 개 | 0.98% |
| 400 개 | 0.97% |

표 8. SVM질병분류 판단 결과

| Train Data 수 | 정확도 |
|--------------|-------|
| 100 개 | 0.85% |
| 200 개 | 0.95% |
| 300 개 | 0.95% |
| 400 개 | 0.93% |
| 500 개 | 0.89% |
| 600 개 | 0.93% |
| 700 개 | 0.87% |
| 800 개 | 0.87% |

SVM 의 질병 유무 판단의 결과를 보았을 때 약간의 오차는 있지만 일반적으로 train data 의 개수가 올라갈수록 더 높은 정확도를 보이는 것으로 나타났다. 하지만 질병 분류 작업의 경우 train data 의 개수가 올라가도 정확도가 감소하는 등 변화가 나타난다. 이는 train data 와 test data 의 비율을 8:2 로 잡았는데 라벨링되어 있는 데이터들이 균일하지 않게 분포하다 보니 test 데이터를 어떻게 잡느냐에 따라 정확도에 영향을 미친 것으로 추측된다. 따라서 800 개를 train 한 모델이 가장 신뢰도 높은 모델이라 할 수 있으며 이러한 문제는 총 데이터가 많으면 해결될 것이라고 판단한다.

5. 결론

일률적인 학습법의 비교는 어렵지만 질병 유무 판단의 경우 CNN 에서 920 개 데이터를 학습했을 때 기준 96%의 정확도를, SVM 에서 100 개 학습기준 95%의 정확도를 보였다. 또한 질병 분류 판단의 문제에서는 CNN 이 920 개의 데이터를 학습했을 때 85%의 정확도를, SVM 에서는 100 개 기준 85%, 200 개 기준 95% 등의 결과를 얻었다. 하지만 모델이 갖는 신뢰도를 따져보았을 때 SVM 은 질병분류 문제에서 800 개 학습 시 도출한 87% 정확도를 갖는다고 말하는 것이 적절하며 이는 CNN 대비 2% 정도 더 나은 결과를 도출했다고 판단한다.

감사의 글

본 프로젝트는 인공지능 팀 프로젝트를 목표로 지능기전공학부 최유경 교수님의 지도를 받아 수행되었다

참고문헌

[1] "A respiratory sound database for the development of automated classification" B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A.Oliveira, C. Jácome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho, N. Maglaveras

[2] <https://www.kaggle.com/eatmygoose/cnn-detection-of-wheezes-and-crackles>

[3] http://news.khan.co.kr/kh_news/khan_art_view.html?art_id=201903061006001

[4] "Spectrum Analysis the short-time Fourier spectrum" 조원주 2009, MARTE class

[5] <http://www.econovill.com/news/articleView.html?idxno=350496>

[6] <http://www.rapportian.com/news/articleView.html?idxno=117040>

[7] <https://m.blog.naver.com/mylogic/220988857132>

[8] <https://www.webmd.com/lung/lung-sounds#2>

[9] <https://www.msmanuals.com/professional/pulmonary-disorders/bronchiectasis-and-atelectasis/bronchiectasis>

[10] https://www.emedicinehealth.com/upper_respiratory_infection/article_em.htm

[11] Motoda, Hiroshi, and Huan Liu. "Feature selection, extraction and construction." Communication of IICM (Institute of Information and Computing Machinery, Taiwan) Vol 5 (2002): 67-72.

[12] https://en.wikipedia.org/wiki/Principal_component_analysis

[13] https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding

[14] https://en.wikipedia.org/wiki/Partial_least_squares_regression

[15] <https://scikit-learn.org/>

[16] <https://www.kaggle.com/nirajvermafcg/support-vector-machine-detail-analysis>

[17] <https://www.kaggle.com/anmour/svm-using-mfcc-features/notebook>

[18] "SVM 커널 함수 최적화를 이용한 표적 식별" 정상훈, 이강인, 김종만, 정용식, 정현교 2018.10, 대한전기학회

[19] "SVM 에 기반한 음악 장르 분류를 위한 특징 벡터 정규화 방법", 임신철, 장세진, 이석필, 김무영, 2011.9, 대한전자공학회

[20] <http://banana-media-lab.tistory.com/entry/Librosa-python>