

# 앙상블 기법을 활용한 FMCC 정확도 향상 연구

이남길 김동희  
인천대학교 컴퓨터공학부  
lyscroix@naver.com, d0\_0h@inu.ac.kr

## A study on improvement of FMCC using Ensemble techniques

Nam Gil LEE, Dong Hee KIM  
Incheon National University, Computer Science and Engineering

### 요 약

FMCC(Female-male Classification Challenge)는 남녀 각각 20명이 3-5음절로 이루어진 한국어 단어 발화 음성 데이터를 통해 남녀를 구분하는 문제이다. 이 문제에서는 MFCC를 이용해 음성 특징을 추출하여 모델의 입력에 사용하였다. 본 연구에서는 보다 향상된 남녀 구분 정확도를 위해 앙상블 기법 중 하나인 Voting Classifier 기법을 적용하였다. 전처리 과정을 거친 데이터를 통해 각 분류기 모델을 학습시키고, 각 모델의 예측 결과 중 가장 많은 표를 얻은 예측 결과를 최종 결과로 사용한다.

### 1. 서 론

본 연구는 잡음 환경에서의 남녀 음성을 분류하는 문제에서의 앙상블 기법을 통한 분류기 성능 향상에 관한 연구이다. 문제를 해결하기 위해 MFCC, Scaling 등의 데이터 전처리 과정을 거친 다음, 앙상블 기법을 통해 분류 작업을 진행한다. 앙상블 기법은 Voting Classifier를 사용하며, 앙상블에 사용되는 분류기는 SVM, K-NN, RF이다. 분류기 모델들의 하이퍼 파라미터 최적화는 Optuna를 통해 실행하였다.

### 2. 데이터 전처리

#### 2.1 잡음 제어

남녀 각각 20명이 발화한 8000개의 오디오 데이터는 ~5dB SNR 정도의 TV 방송 배경 잡음에 오염되어 있다. 잡음 환경에서 남녀 분류 문제를 수행하면 잡음에 노출된 데이터로 학습을 진행하기에 분류 정확도가 떨어진다. 정확도 향상을 위해서 입력 오디오 데이터의 잡음을 줄여주는 noisereduce 라이브러리를 활용하였다.

#### 2.2 MFCC(Mel-Frequency Cepstral Coefficients)<sup>[1]</sup>

2.1의 잡음 제어를 통과한 오디오 데이터에 MFCC 기법을 적용하여 오디오 특징을 생성하였다. MFCC는 입력 오디오에 대해 pre-emphasis, windowing, FFT, Mel scale Filter Bank, Log tranform, DCT 과정을 통해 추출할 수

있다. 일반적인 음성 최대 주파수는 여성 기준 250Hz에 존재하기에 MFCC의 최대 주파수를 400Hz로 설정하였으며 MFCC의 계수는 20으로 설정하였다.

#### 2.3 입력 차원 통일 및 정규화

MFCC를 통해 얻어낸 특징은 각 데이터마다 다른 길이를 가지고 있다. 모델에 입력하기 위해서 특징의 길이를 같은 크기로 통일해야 하는 작업이 필요하다. MFCC의 평균 길이인 34에 10% 여유를 둔 37의 값으로 입력 데이터의 길이를 맞췄다. 또한 입력 데이터의 스케일을 맞추기 위해 정규화를 진행하였다.

### 3. 분류기 모델

#### 3.1 SVM (Support Vector Machine)<sup>[2]</sup>

Support Vector Machine 분류기는 이진 분류 작업에서 두 클래스 간의 마진을 최대화하는 최적의 초평면을 찾도록 학습된다.

#### 3.2 K-NN (K-Nearest Neighbors)<sup>[3]</sup>

K-Nearest Neighbors는 특징 공간에 미리 레이블이 지정된 k개의 가장 가까운 훈련 샘플 중 다수 클래스를 기반으로 새로운 샘플을 분류한다.

#### 3.3 RF (Random Forest)<sup>[4]</sup>

Random Forest는 다수의 의사 결정 트리의 예측 결과를 종합하여 최종 결과를 결정하는 앙상블 기법이다.

### 3.4 Voting Classifier<sup>[5]</sup>

분류기로 입력한 모델들의 예측 결과를 집계하여 가장 많은 표를 얻은 클래스를 최종 예측 결과로 사용한다.

## 4. 실험

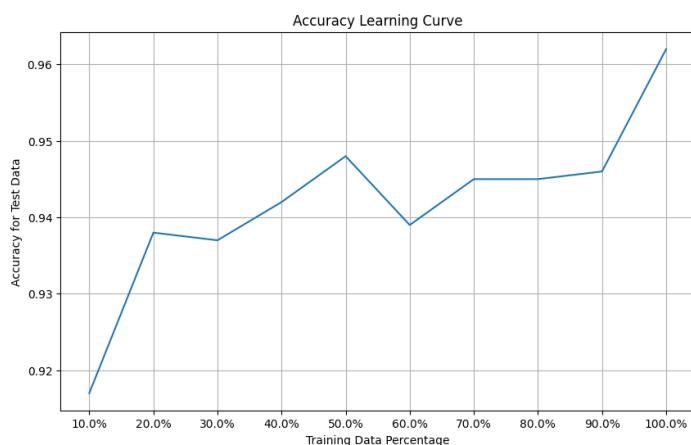
3절에서 언급한 SVM, K-NN, RF를 앙상블 기법 중 하나인 Voting Classifier를 사용하여 오디오 남녀 구별 문제를 진행하였다. 입력 데이터는 2절의 데이터 전처리 과정을 거쳤다.

각 분류기 모델에서의 최적의 하이퍼 파라미터를 찾기 위해 Optuna<sup>[6]</sup>를 사용하였다. Optuna는 트리 기반과 Pruning Mechanism을 통해 효율적으로 하이퍼 파라미터 최적화를 가능하게 한다.

## 5. 실험 결과

모델	정확도
SVM	94.3%
KNN	93.9%
RF	92.6%
<b>Voting Classifier</b>	<b>96.2%</b>

위 표는 앙상블 기법에 사용한 분류기 모델들을 각각 학습시켰을 때의 정확도를 나타낸다. 모델을 하나만 사용했을 때에 비해 확연한 정확도 향상을 관찰할 수 있다.



딥러닝 모델이 아니기 때문에 Epoch 을 통한 학습 곡선을 그릴 수 없어 학습 데이터를 분리해 10%부터 100%까

지 학습시킨 모델로 테스트 데이터에 대한 정확도를 측정하였다.

```
===== Results Analysis =====  
Test: 과탐_test_results.txt  
True: fmcc_test_ref.txt  
Accuracy: 96.20%  
Hit: 962, Total: 1000  
=====
```

테스트 결과

모델 사이즈: 72.4MB

가상환경 크기: 544MB

## 6. 실험 환경

OS: Windows 10

Python version: 3.10

## 7. 참고 문헌

[1] MFCC

<https://librosa.org/doc/main/generated/librosa.feature.mfcc.html>

[2] Support Vector Machine

<https://scikit-learn.org/stable/modules/svm.html>

[3] K Nearest Neighbors

<https://scikit-learn.org/stable/modules/neighbors.html>

[4] Random Forest Classifier

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

[5] Voting Classifier

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html>

[6] Optuna

[https://hyunicecream.tistory.com/75#google\\_vignette](https://hyunicecream.tistory.com/75#google_vignette)

[7] Chat GPT

[8] Github Copilot