

1. 概要

今回のシステムは主に音声を分離するフェーズと、分離した音声に読みの音声があるかないかを判定するフェーズの2つのパートから構成されている。以下にそれぞれの詳細を記す。

2. 音声の分離

音声の分離には非負値行列因子分解 (NMF) を用いる。まず、事前に公開されたすべての読みデータに通常の NMF を適用して、個別の教師基底を作成しておく。競技時に問題データを分離するときは、事前に分離した教師基底を固定の基底行列として用い、全教師あり NMF を適用してアクティベーション行列のみの最適化を行う。これにより、それぞれの教師基底に対応するアクティベーション行列を得ることができる。

3. 読みの音声のあるなし判定

分離後の音声にはノイズやエコーが乗っている。ノイズやエコーは読みの音声と同じような周波数帯であるので削除することが困難である。そこで事前に、畳み込みニューラルネットワーク (CNN) を用いて分離後のスペクトログラムを受け取り、読み音声のあるなし判定を行うモデルの学習を行う。競技時は、問題データを分離したあと、得られたすべてのスペクトログラムを判定モデルに渡し、読み音声のあるなし判定を行う。

4. 実験・開発環境

OS: Ubuntu, Windows11, MacOS

言語: Python3, C++

計算ノートブック: Jupyter lab

エディター: PyCharm, CLion, etc...