第 33 回競技部門：登録番号 30008

チーム名：過学習

学校名：有明工業高等専門学校

1. システムの概要

本大会で使用される問題データを再現した音声を作成し、使用された音声の特徴を捉えるために機械学習を行う。音声のまま機械学習を行うことは困難であると判断したため音声を画像に変換し、畳み込みニューラルネットワーク(以下CNNとする)を使用する。これにより学習をより効率的に進めていく。

2 章のデータセットと学習用画像を作成するプログラム、および 3 章の機械学習を行うプログラムを用いて問題解決を目指す。また、4 章のデータの送受信を行うプログラムを用いて試合情報や問題情報、分割データの取得や解答の送信を行う。

2. データセットと学習用画像の作成

機械学習を行うにあたって、入力する画像と正解のラベルが入ったデータセットが必要となる。このデータセットを作成するために音声の合成やずらし、冒頭と末尾の削除などをして問題データを再現した大量の音声を作成する。

音声を配列として扱うことで問題データにより近い音声の再現や音声同士の足し合わせとずらしを容易に表現することが可能となる。正解となるラベルは音声とともに生成する。これは、読みのデータの総数と同じ 88 個の数値が入った配列であり、数値は 1 または 0 である。使用された音声が該当する場所のみ 1 とする。

フーリエ変換によって、周波数と時間を軸にとるスペクトログラムに変換させ(図1-①)、これにより音声から画像への変換を実現する。また、画像化する際に画像ごとに縮尺が変わることを避けるため音声の長さを 3 秒に揃え、足りない部分は空白の音声で補う処理を施す。

3. 機械学習

ニューラルネットワークの中でも短期間で良い結果が期待できる CNN を使用する。CNN は画像の処理を行うものである。入力はピクセル数個あり、各入力値には輝度が入っている。

コンボリューションと呼ばれる画像に様々なフィルタをかける操作(図 1-②)によって特徴を見つけやすくすることで効率的な学習が可能になる。問題データには読みデータが 3~20 個含まれ、試合開始前にいくつ含まれているかがわかる。そのため、含まれている読みデータ数に対応した 18 個のニューラルネットワークを構成し、正確性が上がるようにする。

機械の予測値の出力は正解となるラベルと同様に88個の配列として出力する(図1-③)。配列内の各数値は0~1であり、該当する箇所の音声が使用された可能性が高いほど1に近くなり、低いほど0に近くなる。この出力を正解となるラベルと共に損失関数に渡し、予測値と正解がどれだけ離れているか誤差をとる。誤差が小さいほどより正確に予測ができていると判断する。また、1つのデータセットを何度も繰り返し学習させることで正解となるラベルに近い予測値を出力できるように実装する。大会本番ではある値で閾値をとり、0あるいは1の整数値に整える。

4. データの送受信

PythonのRequestsライブラリを使用する。試合情報と問題情報を取得する関数、取得する分割データの指定とそのwavファイルを取得する関数、解答の送信をする関数を一つのプログラムにまとめ、モジュールとして実装する。これにより作成したコンテンツをより直感的に解答として送信できるようにする。

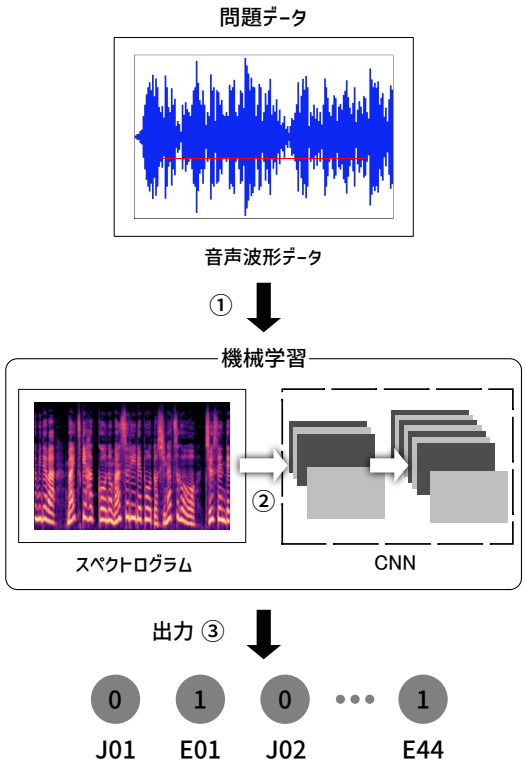

図 1. システムの概要