部 門

競技部門

No.1 登録番号

30008

予定開発期間:5ヶ月
予定開発人数:3人

No.2

	4月		5月			6月			7月			8月			9月			10月			
問題分析				1								٧									
設計						†				\											
実装										†					\						
試用・トレーニング													4						\		

実現方法

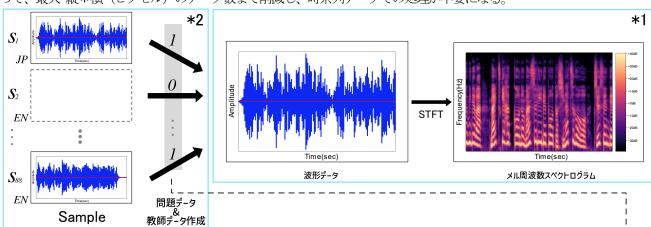
1) 音声の解析アルゴリズム

全体の方針

合成された音声データを学習した機械学習のプログラムにかけて、どの音声が合成されたのかを出力する。 機械学習を行う前に、波形データを画像化する音声の前処理を行い、画像による機械学習を行う。

・音声データの前処理

合成された音声データ(波形データ)をSTFT(短時間フーリエ変換)を用いてメル周波数スペクトログラム(画像)に変換する。(*1) 本来必要であった時系列ごとの処理や、入力データである48,000*t(s) 個のデータを、画像化することによって、最大 縦*横(ピクセル)のデータ数まで削減し、時系列データでの処理が不要になる。

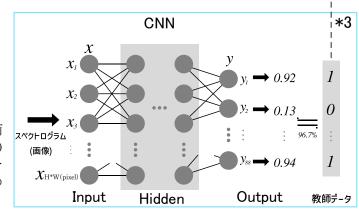


sNo.3

・機械学習の方針

問題データは単体のサンプルデータ(88種)をランダム(3~20種)で合成して作成し、分割やずらしなども表現する。その際にどの音声が合成されたかを表す教師データも作成する。(*2)

スペクトログラム化された問題データを入力データとして渡し、CNN(畳み込みニューラルネットワーク)で画像の機械学習を行い、どの音声が使われたのかを表す0~1の88個の配列を出力する。その後、問題データを作成した際に生成された教師データと出力データとの差で誤差をとり、学習を行う。(*3)



・本番(本戦)での出力

CNNの出力は、使われた音声が 1, 使われていない音声が 0 に近い値になっている 88 種類の配列である。これを 1, 0 に近いほうに変換し、取り札が 44 種類のため、日本語・英語のどちらかが 1 だと、その音声に対応した札を 1 として出力する。

2) その他 (独創的なところ)

- ・音声を波形データとして扱わず、画像化して処理を行うこと。
- ・スペクトログラムにメル周波数スペクトログラムを用いることで、人間の聞き取れる周波数をスペクトログラム化する。 これにより、音声の特徴などを見つけやすくなるのではないかと考えた。

No.4

開発言語:Python

開発環境: Anaconda, Visual Studio Code, TensorFlow, FFmpeg, Neural Network Console, Git Hub

第33回 全国高等専門学校 プログラミングコンテスト: 群馬大会