|  |  |  |  |
| --- | --- | --- | --- |
| 部　　　門 | **競　技　部　門** | No.1 登録番号 | 30008 |



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.2** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | ４月 | | | ５月 | | | ６月 | | | ７月 | | | ８月 | | | ９月 | | | １０月 | | | | 問題分析 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 設計 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 実装 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 試用・トレーニング |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |   1) 予定開発期間：５ヶ月  2) 予定開発人数：３人 |
| **No.3** | 実現方法 |
| 1) 音声の解析アルゴリズム |
| 〇全体の方針  ・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。  ・音声の波形データを画像化（音声の前処理）を行い、画像による機械学習を行う。  〇音声データの前処理  合成された音声の波形データを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　（図1)。 画像化することによって、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、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。    図１  図２  〇機械学習の方針  ・問題データは88種類の単体の音声(サンプル)データをランダム(3～20種)で合成して作成し、分割やずらしなども表現する。その際にどの音声データが合成されたのかを表す教師データも作成する(図2)。  ・画像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から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(図3)。  図３  〇本戦での出力  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個の配列である。これを1と0のみの出力に変換する。取り札が44種類のため、日本語・英語のうちどちらかが1の場合は、その札を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** | |