

ICISA 2020 での研究発表の報告

北 迫 祐 樹

Yuki KITASAKO

情報メディア学専攻修士課程 3年

1. はじめに

私は、2020年12月16日から18日に開催されたICISA 2020 (International Conference on Information Science and Applications 2020) で「Proposal of situation estimation system using AI speaker」の研究について発表を行った。

2. 研究概要

2.1 研究背景

この先、各家庭に音声 AI を備えた機器が増えていく中で、音声 AI が取得する音をつかいユーザーの状況を推測し、共有することが可能になれば、ユーザーの状況に合わせたサービスの提供や、室内見守りシステムの構築など幅広い分野への応用が期待出来ると考え、実現するシステムの提案を行う。本システムでは、あらかじめ、音声 AI に命令する中で毎日繰り返し命令が行われるような比較的頻度の高い命令を想定し、その時のユーザーの命令より前の環境音を分析する。特徴的な環境音を数種類シンボル音として分類しておく。ユーザーの家に設置された複数台の音声 AI が搭載された機器をつかい、命令や一定以上の確度のシンボル音が発生するたびに、命令やシンボル音の種類と発生した時間とを記録し収集する。収集したデータを用いることで、特定のシンボル音の並びが類似した時間間隔で発生する度に特定の命令が行われているといったパターンを解析することが可能になる。データ解析あるいは機械学習により判明した特定のシンボル音の発生パターンと類似した発生パターンを認識したとき、特定の命令を行うかユーザーに提案する。提案システムがユーザーの状況を推測するために行うパターンマッチングで使用するシンボル音の識別に関

する実験として①何をシンボル音とすることが出来るか、②それぞれ異なるシンボル音同士の識別が可能であるか検証を行うために2種類の実験を行った。

2.2 実験

実験1は①について検証するため、録音した日常の環境音の中からシンボル音の候補となる特徴音を調べた。3週間、1日6時間以上の録音を行った。録音した環境音を実際に聞くことで環境音の中からシンボル音の候補となる特徴音をリストにまとめた。実験2は②について検証するため、実験1で調べたシンボル音の候補の中から機械学習を行い、識別が正しく行えているかの検証を行った。外側から鍵を開ける音、ドアの開閉音、内側から鍵を閉める音の3つの特徴音を選び、3つの特徴音をそれぞれ120個ずつ録音したデータを、図1に示す画像のような縦軸が周波数、横軸が時間のスペクトログラムになるように短時間フーリエ変換を行い画像データに変換した。鍵を開ける音、ドアの開閉音、鍵を閉める音を画像データ毎にラベリングし、それぞれ120枚のうち80枚を学習用データに20枚を検証用データにしてCNNモデルを構築し学習を行った。このときCNNの学習結果として識別の精度と損失

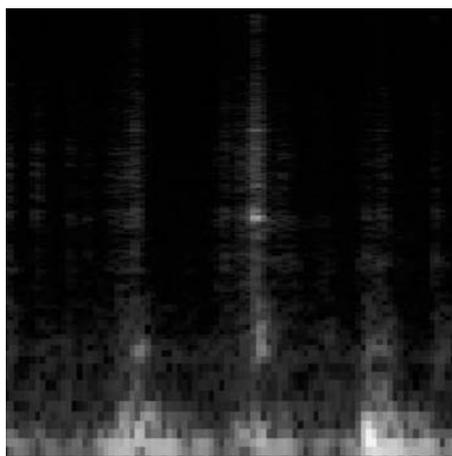


図1 縦軸が周波数、横軸が時間のスペクトログラム

値をグラフにした。学習の済んだモデルで未確認データを各 20 個ずつの識別を行いその予測精度から識別が可能か検証した。

2.3 実験結果

実験 2 の結果を図 2, 図 3 に示す。予測精度の平均は約 97.8% だった。

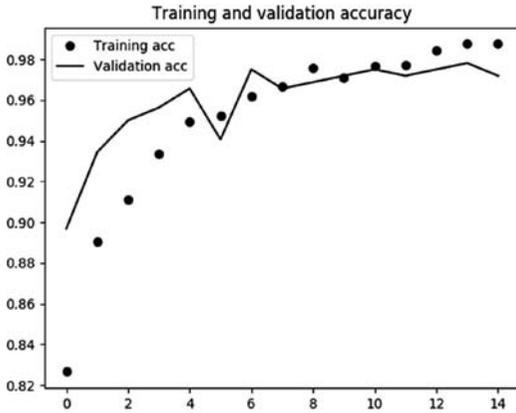


図 2 精度を表すグラフ

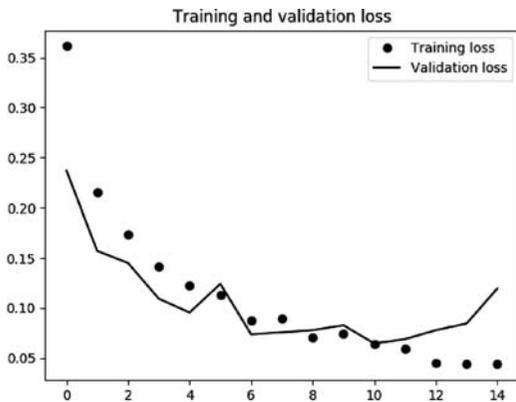


図 3 損失を表すグラフ

2.4 考察

多数のシンボル音の候補を見つけることが出来た上に平均 97.8% と高い精度でシンボル音同士の識別が可能であることがわかった。今後は、個別のシンボル音として識別が可能であることがわかったため、シンボル音をシンボルデータの列として扱うことでデータとして軽いものにし、多くのシンボルの中からユーザーの状況を推測することが可能になるシンボルのパターンを見つけることが出来ないかパターン解析を行う。また、今回のスペクトログラムは時間の長さが異なる音も含まれていたため、音データをオーバーラップで短く区切り画像に変換したもので同様の識別が行えるか確かめる実験についても検討している。

3. おわりに

最後に、今回の発表を行うにあたってご指導をいただいた三好力教授に深く感謝いたします。

参考文献

- [1] Shinagawa Yoshimitsu, Kishimoto Toshio, Ota Shigeru: "Home Behavior Monitoring Using Microphone Sensors", Kawasaki Medical and Welfare Society, Vol.15, No.2(2006) : 615-620
- [2] A. Mesaros, T. Heittola et al. "Detection and classification of acoustic scenes and events : Outcome of the DCASE 2016 challenge," IEEE/ACM Transactions on ASLP, vol.26, no.2(2018) : 379-393.