

事前学習を用いた心拍生体認証における効率的な学習モデルの構築

上條 元彌[†] 藤原 章大[‡] 馬 建華[‡]法政大学大学院 情報科学研究科[†] 法政大学 情報科学部[‡]

1. まえがき

近年ウェアラブルデバイスは半導体技術の進歩により低コスト化と製品化が進み、広く普及するようになった。ウェアラブルデバイスは生体情報を継続的に取得することができ、デバイスを利用した生体認証技術の応用研究が数多くなされている。例えば、ウェアラブルデバイスで取得可能である心拍データは個人特有の生体情報であり、情報量が少なくユニークな性質を持つ。しかしデバイスで取得するデータは、装着部位の制約や体動の影響を受け、高品質かつ多様なデータを取得することが困難である。

本研究はウェアラブルデバイスで取得した心拍データより生体認証モデルを構築する。そしてファインチューニングに基づきデバイスで取得したデータ(心拍データ)に相関のあるオープンデータセット(音声データ)で事前学習し、心拍データの学習モデルに転用することで、少量のデータで汎化性能を持つモデルを構築する。結果、事前学習により、心拍データにおけるモデルの精度と学習の収束速度が改善されることを確認した。さらに、学習に利用するデータ数を減らした場合の学習精度と収束速度を検証する。

2. 関連研究

Ekizら[1]は、スマートバンドのPPGセンサから心拍変動(HRV)を取得し、心拍データの特徴から人物認証を行った。28名の健康な男女(男性16名、女性12名)は、デバイスを装着した状態で、平常時とストレス誘発時でデータの取得を行った。そして実験で得たデータを用いて機械学習分類器より識別を行う。評価指標の等価エラー率(ERR)は、ランダムフォレストの分類器で10.08%を示し、心拍データによる人物認証が可能であることを示した。また別環境で取得したデータに対する識別精度は未確認であり、多様な状況で高い識別精度を示すモデルの構築には、データセットの大規模化を主張している。

汎化性能を持つ高精度な深層学習モデルを構築するためには、データセットの規模を大きくし、多様な環境でデータ取得を行う方法が挙げられる。しかしデータセット作成の高コスト化が想定される。またユーザーの登録に時間と手間を要し、実用性に欠けることが考えられる。

3. 提案

本研究ではファインチューニングに基づき、音声の一般公開データセットによる学習済みの話者認識モデルを利用することで上記の問題の解決を図る。多様かつ大規模なデータを事前学習させたモデルを利用することで、少量のデータで学習可能、かつ汎化性能を持った心拍認証モデルを構築することを目的とする。以下に提案モデルの構築について順次論述する。

3.1. 心拍データの取得実験

データ取得の実験には20代男性5名が参加した。心拍データの取得にはPolar H10心拍センサ(© Polar Electro)を使用する。センサを胸部に装着することで、心電図(ECG)を取得可能としている。そして各人物10分間のデータ取得を2セット行う。1セット目は椅子に座った状態でデータを取得し、2セット目は床で横になった状態でデータを取得する。取得したECGをCSVファイルに145Hzで心電図波形として保存する。

3.2. 一般公開データセットのデータ処理

本研究ではECGの持つ周期性に着目し、同じく周期性があり、個人の生体情報である音声データのコーパスを利用する。JVSコーパス[2]はスタジオ収録の声優・俳優などの音声を含む。また、話者ごとに複数の発話方法を含んだ150個の音声ファイルで構成される。

データ処理はPythonパッケージより行う。はじめに各音声データを音声波形へと変換する。音声データは24kHzの周波数で構成され、取得した波形に4000点ごとにハミング窓を掛け、フーリエ変換する。各ファイルの音声波形にこの操作を、150点ずらしながら繰り返し、スペクトログラムを保存する(図1左)。また、音声データ

と同じ条件で心拍データの処理も行い、スペクトログラムを保存する(図1右). データ処理の結果、心拍のスペクトログラムは5人で4531枚(各人物で約900枚)であるのに対して音声のスペクトログラムは142994枚となった.

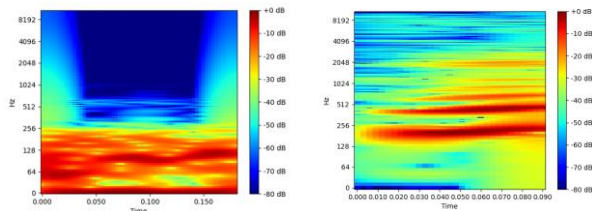


図1. 心拍(左)・音声(右)のスペクトログラム

3.3. ファインチューニングによるモデル構築

音声のスペクトログラムは各人物ごとに7:3の割合で訓練データと評価データに分割し、深層学習モデルの入力とした。モデルの構築にはPythonパッケージのTensorFlowを利用した。モデルの中間層は二次元畳み込みニューラルネットワークで、5クラス分類で99.95%の精度を示す話者認識モデルを構築した。そして提案モデルは、この学習済みモデルの出力層を削除し、新たに全結合層と出力層を追加したモデルとなる。本論ではモデル①として扱う。モデル①は、事前学習していない2つのモデルと比較することで有用性を示す。モデル①と同じ構造を持ち、かつ事前学習を行っていないものをモデル②(入力層→畳み込み層→全結合層→追加した全結合層→出力層)、モデル②の始めの全結合層を削除したものをモデル③(入力層→畳み込み層→追加した全結合層→出力層)とする。即ちモデル①とモデル②は重みの数が等しく、モデル①とモデル③は全結合層の層と重みの数が等しくなる。

4. 結果と考察

3つのモデルより、全ての心拍スペクトログラム(4531枚)で学習を行った。学習曲線とモデルの精度を図2、表1に示す。学習は100エポックで行った。提案モデル(モデル①)が訓練・評価共に収束速度が最も速く、最も高い精度を得た。

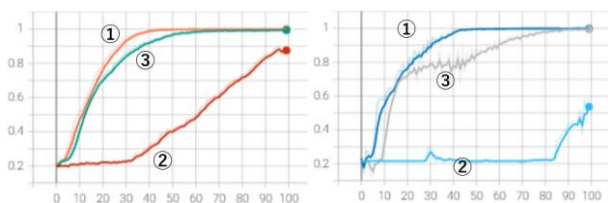


図2. 各モデルの学習曲線(左:訓練 右:評価)

次に学習に使うデータを210枚に減らし、少量のデータに対する精度と収束速度を確認する。3

つのモデルの学習曲線とモデルの精度を図3、表2に示す。提案モデル(モデル①)が訓練データに対し収束速度が最も速いことが検証された。

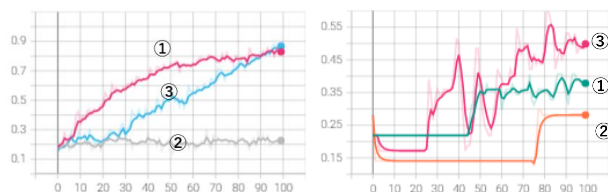


図3. データ数を減らした場合の各モデルの学習曲線(左:訓練 右:評価)

表1. 各モデルの識別精度

訓練データ数:3171 評価データ数:1360	Accuracy	Loss
モデル①	99.93	0.014
モデル②	59.49	1.171
モデル③	99.49	0.483

表2. データ数を減らした場合の各モデルの識別精度

訓練データ数:147 評価データ数:63	Accuracy	Loss
モデル①	37.82	1.241
モデル②	28.13	1.558
モデル③	49.86	1.269

図2、図3より、提案モデルの学習曲線が大きな変動をせずに上昇していることから、学習済みモデルの重みが大幅な更新をしていないと考えられ、さらに心拍と音声のデータに含まれる個人の特徴には相関があると考えられる。またデータ数を減らした場合、提案モデルは他のモデルと比較して収束速度が速く、少ない学習回数で高い精度を得た。認証精度を更に向上させることで、少量の心拍データから深層学習による生体認識モデルを構築する場合、事前学習モデルを利用することは有効であると考えられる。

5. まとめ

心拍情報の生体認識モデルは、学習済みの音声認識モデルをファインチューニングした結果、通常の深層学習モデルより収束速度が速く、高精度なモデルを構築することができた。

文献

- [1] Ekiz, D., Can, Y. S., Dardagan, Y. C., Ersoy, C. Can a smartband be used for continuous implicit authentication in real life. IEEE Access, 8, 2020, pp. 59402-59411.
- [2] The JVS (Japanese Versatile Speech) Corpus, "https://sites.google.com/site/shinnosuketakamichi/re-searchtopics/jvs_corpus"