

脳波データの圧縮を考慮したスケーラブルな 電極チャネルと特徴量の組み合わせ

鶴ヶ崎 裕真[†] 田邊 俊[‡] 田谷 昭仁[†] 戸辺 義人[†]

青山学院大学理工学部情報テクノロジー学科[†]

青山学院大学大学院理工学研究科理工学専攻知能情報コース[‡]

1. はじめに

近年、2025年問題による影響の1つの、患者数の増加・医師不足の影響が懸念されている。これは、医師が身近にいないことから、診断されるべき患者への対応が難しくなることを意味する。この問題を解決するために、サーバやクラウドに脳波データを転送することで、遠隔患者に対する診断の実現が期待される。脳波データをそのまま送るとサイズが大きいという問題があり、データ圧縮が望ましい。更に、データ圧縮をしても診断の精度が下がらないことが要求される。本研究では、要求される診断の正解率を満たす、特徴量の数と電極数の指標を作成する。

2. 関連研究

脳波から、うつ病の診断を可能にした関研究がある¹⁾。この研究では、90%を超える精度で診断することを可能にしている。

また、脳波と視線の関連を調査した研究がある²⁾。この研究は、学習時の特徴量を決め打ちで作成しているという問題がある。本研究では関連研究²⁾のデータセットを使用する。

3. 設計

3.1 学習モデルの選定

作成されたデータの学習に先立ち学習モデルを決定した。具体的には、ベイズ最適化によりパラメータチューニングを行い、正解率が最も高くなった学習モデルを使用した。

表 3.1. 各モデルの正解率

モデル	正解率
勾配ブースティング	79.0
ナイーブベイズ	78.6
Ada ブースト	78.6
決定木	78.6
SVM	78.6
k 近傍法	78.5
ランダムフォレスト	78.4
ロジスティック回帰	74.8

Scalable Combination of EEG Channels and Features for Data Compression/ Yuma TSURUGASAKI, shun TANABE, Akihito TAYA, Yoshito TOBE/ Aoyama Gakuin University

ベイズ最適化によるパラメータチューニングの結果、表 3.1 のようになった。このことから、本研究では、勾配ブースティングを使用する。

3.2 特徴量の作成

特徴量を作成するにあたり、tsfresh³⁾を使用した。tsfreshとは、時系列データから特徴量を作成するライブラリである。このライブラリを使用することで、最大値・最小値・中央値・平均値など、合計714個の考えられる全ての特徴量を作成する。

時系列データから特徴量を作成するアルゴリズムは、以下のようになっている。

- ① $H_0^{\phi} = \{\text{特徴量 } X_{\phi} \text{ がラベル } Y \text{ に関係ない}\}$
- ② 帰無仮説の X_{ϕ} の P 値が 0.05 以下のものを選ぶ

3.3 特徴量削減

tsfreshにより作成されたデータを圧縮するにあたり、3種類の特徴量削減手法を使用した。

3.3.1 ANOVA 分析の F 値

ANOVA 分析とは、3つ以上の平均値の相違を検討する場合の方法であり、単変量選択を行うことで、特徴量を選択する。具体的には、ラベルと特徴量間での ANOVA 分析の F 値を指標にし、順位付けを行う。その順位に基づき、指定した数の特徴量を選択する。

3.3.2 ランダムフォレストと勾配ブースティング

モデルを使用し、特徴量の重要度を表す feature_importances_ の値を取得する。具体的には、本研究でのモデルは、ランダムフォレストと勾配ブースティングの、2種類のモデルを使用する。その際に、それぞれの特徴量の重要度を取得する。構築されたモデルの中で、最も重要度の低い特徴量を削減する。同様の処理を、指定した特徴量数になるまで、再帰的に繰り返す。

4. 圧縮の効果

特徴量・電極数を変化させることにより、それぞれのときの正解率を求める。具体的には、

tsfresh により作成された特徴量を, ANOVA 分析の F 値・ランダムフォレスト・勾配ブースティングの 3 手法により削減する. その後, 勾配ブースティングにより, 各削減率での精度を求める.

4.1 結果

電極数が 1 本, 圧縮率が 0~95% のときの正解率を図 4.1 に示す.

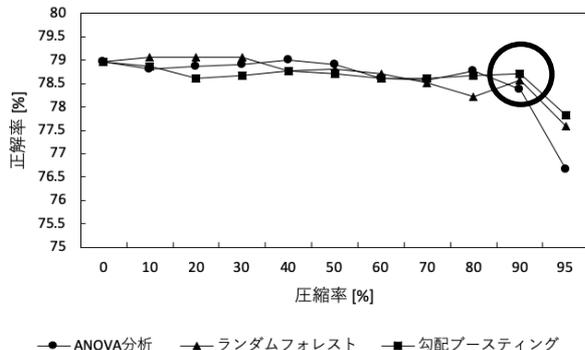


図 4.1. 電極 1 本の時の圧縮率に対する正解率

電極数が 2 本, 圧縮率が 0~95% のときの正解率を図 4.2 に示す.

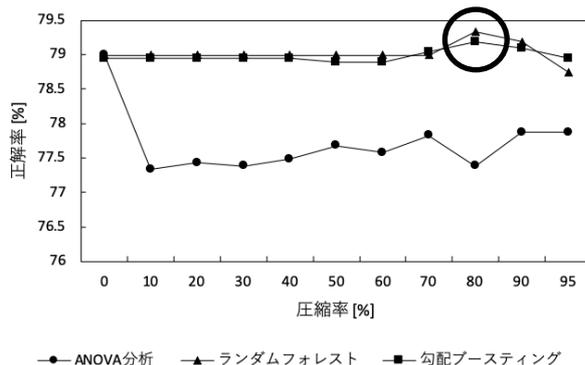


図 4.2. 電極 2 本の時の圧縮率に対する正解率

電極数が 3 本, 圧縮率が 0~99% のときの正解率を図 4.3 に示す.

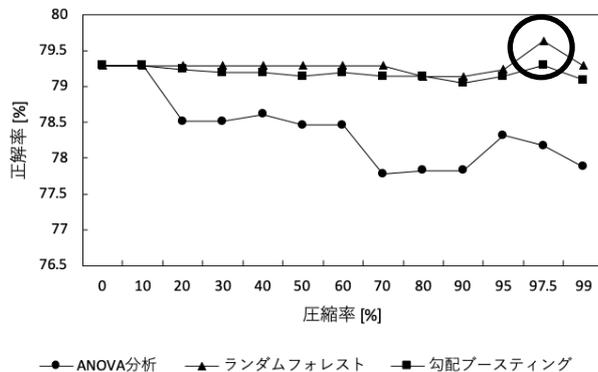


図 4.3. 電極 3 本の時の圧縮率に対する正解率

図 4.1~4.3 から, 特徴量の圧縮率が大きくなっても図中の円記号で示した箇所までは正解率

はほぼ一定であることがわかる. 図 4.1 では, さらに圧縮率を大きくすると正解率が急激に減少する. 図 4.2~図 4.3 では, 図中の円記号で示した箇所の正解率が最も高くなっている. これは, 相関の高い特徴量が圧縮されたことにより, 有効な特徴量のみが使用されたからと考えられる. 更に, 図 4.2~図 4.3 より, ANOVA 分析とランダムフォレスト・勾配ブースティングを比較した場合, 後者の精度が高いことがわかる.

5. 評価

関連研究²⁾の電極数 1~3 本のときの正解率との比較をしたときの, 結果を図 5.1 に示す. 本研究のすべての条件において, 関連研究²⁾の精度を上回っていた.

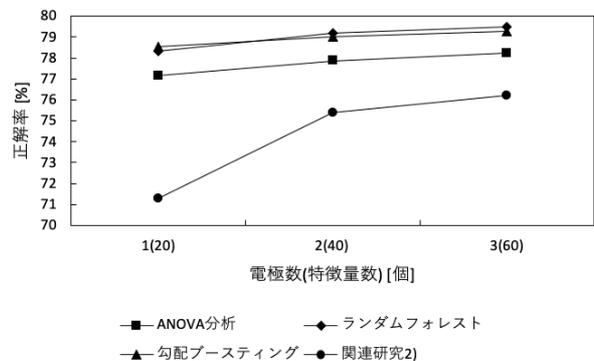


図 5.1 関連研究との比較

6. むすび

本研究で, 要求される診断の正解率を満たす特徴量の数と電極数の指標を作成した. 具体的には, tsfresh により得られた特徴量の数・電極数を変化させることにより, ある割合まで, 特徴量を圧縮することが可能という結果となった.

参考文献

- 1) Acharya, U. R., Shu, L. O., Yuki, H., et al.: Automated EEG-based screening of depression using deep convolutional neural network, Computer Methods and Programs in Biomedicine, Vol. 161, pp. 103-113 (2018).
- 2) 下田功一, 田邊俊, 中山悠 ほか: e ラーニング集中計測に向けた視線と脳波の関連性調査, 電子情報通信学会総合大会, D-9-2, 東京, (2019).
- 3) Maximilian, C., Nils, B., Julius, N., et al.: Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh - A Python package), Neurocomputing, Vol. 307, pp. 72-77 (2018).