

発話音声の音響特徴量を用いた抑うつ度推定手法

Depression score estimating method using acoustic features of speech utterances

森 浩貴*¹
MORI Koki

目良 和也*²
MERA Kazuya

黒澤 義明*²
KUROSAWA Yoshiaki

竹澤 寿幸*²
TAKEZAWA Toshiyuki

*¹ 広島市立大学情報科学部
School of Information Sciences
Hiroshima City University

*² 広島市立大学大学院情報科学研究科
Graduate School of Information Sciences
Hiroshima City University

In this paper, we propose a method to estimate speaker's depression score using acoustic features of his/her speech. 150 speech utterances that 15 subjects read 10 types of sentences were recorded as training data, and the depression scores of the subjects were calculated by Beck Depression Inventory (BDI) just after the recording. Acoustic features are calculated by using openSMILE or Surfboard, and Support Vector Regression or LightGBM are used for machine learning procedure. The experimental results showed that the estimated depression scores obtained a correlate efficient of 0.932 with the correct answer.

1. はじめに

新型コロナウイルス感染症 (COVID-19) の感染拡大を防ぐためのロックダウンや外出制限は、多くの人の心の健康に多大な影響をもたらしており、日本においてもうつ病の人の割合は2013年から2020年にかけて2倍以上に増加している[OECD 21]. うつ病は早期発見・早期治療により状態を改善しやすくなるが、自分自身でうつ病であることに気づくのは難しい。そこでこれまで発話音声を用いてうつ状態の有無を判定する技術[宗16]が開発されてきたが、うつ状態の程度までは推定することができていない。そこでうつ状態の程度を推定できれば、図1に示すように病状に適した処置を提案することが可能となる。

本論文では、入力発話から算出した音響特徴量を用いて話者の抑うつの程度を推定する手法を提案する。使用する音響特徴量として、感情推定に有効とされる openSMILE[Eyben 10]と、疾病に起因する音声変化を捉えられるとされる Surfboard [Lenain 20]について比較検討を行う。

2. 発話音声からの抑うつ度推定手法

2.1 提案手法の構成

提案手法の構成を図2に示す。発話音声の音響特徴量から話者の抑うつ度を推定する機械学習器を構築するため、学習フェーズでは、説明変数として発話音声から算出された音響特徴量、目的変数として発話者の抑うつ度を学習させる。音響特徴量の算出には openSMILE の eGeMAPsv02 feature set と Surfboard を用いる。抑うつ度の指標には抑うつ度を客観的に測る自己評価表の一つであるベックうつ病調査表を用いる。そして推定フェーズでは、発話者の音声の音響特徴量から、発話者の抑うつ度を推定する。

2.2 ベックうつ病調査表

ベックうつ病調査表 (Beck Depression Inventory: BDI) [Beck 61] は、抑うつ度を客観的に測る自己評価表の一つである。定期的にこのテストを行うことによって、自分自身の気分の傾向を数値として測定することが可能となる。

BDI の使い方としては、各項目をよく読み、最近 2, 3 日の気

連絡先: 目良和也, 広島市立大学大学院情報科学研究科,
mera@hiroshima-cu.ac.jp

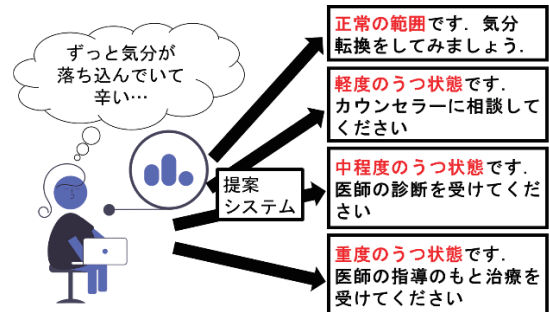


図1 推定された抑うつ度に基づくアドバイスの提示

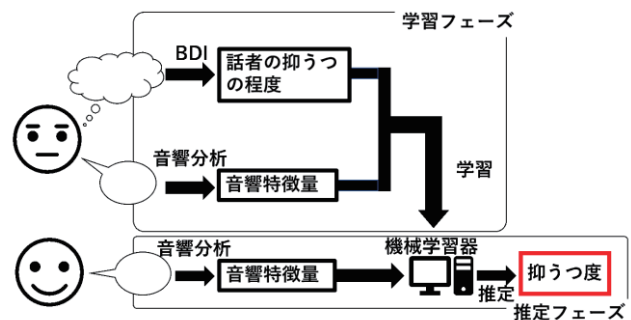


図2 提案手法の構成

分に一番近い答えを選ぶ。各項目の選択肢には 0~3 の点数が与えられている。21 個すべての項目に答え、その合計点数が対象者の現在の抑うつ度となる。表 1 に点数ごとの抑うつ状態および処置を示す[Burns 90].

2.3 Surfboard

Surfboard[Lenain 20]は、元々医療分野への応用として開発されたオープンソースの Python ライブラリである。Surfboard は「成分」を単一の値 (例: ラウドネス) または時系列 (例: MFCC) として抽出する。後者の場合、時系列から統計量 (例: 標準偏差) を抽出し、時間に依存しない特徴を作成することが可能である。

疾病に起因する音声の変化は多くの疾患で報告されており、MDD (大うつ病障害) においてもそのような音声の変化が確認

表 1 点数ごとの抑うつ状態と適した処置

点数	状態および処置
0~10	正常範囲. ただの憂鬱な状態.
11~16	ノローゼ気味. 軽いうつ状態.
17~20	臨床的な意味でのうつ状態との境界. 専門家の治療が必要.
21~30	中程度のうつ状態. 専門家の治療が必要.
31~40	重いうつ状態. 専門家の治療が必要.
40 以上	極度のうつ状態. 専門家の治療が必要.

表 2 MDD(大うつ病障害)に関する音響特徴量

特徴量名	意味	変化
MFCCs (13 種)	13 分割されたメル周波数ケプストラム係数	↓
F0 contour †	基本周波数の時間平均値	↓
F0 SD	基本周波数の標準偏差	↑
Intensity †	音響インテンシティの時間平均値	↓
Intensity SD	音響インテンシティの標準偏差	↓
Jitter variants (5 種)	ジッター(音の高さのゆらぎ)の変異型	↑
Shimmer variant (5 種)	シマー(音の大きさのゆらぎ)の変異型	↑
Formant F1	第 1 フォルマント	↔
HNR	倍音とノイズの比率	↑

されている[Low 20]. Surfboard ではさまざまな疾病に関連する音声変化を捉えるための音響特徴量を扱っており, MDD との臨床的関連性が実証された特徴量についても抽出されている. 表 2 に MDD に関する特徴量と特徴の変化を示す. 表中の↑は健常者の音声と比較して特徴が増加, ↓は健常者の音声と比較して特徴が減少, ↔ は特徴に違いはみられるがどのように変化するかは不明であることを示している. また, 表内のダガー(†)は時系列であることを示し, 基準値は時間平均値である.

本手法では, 表 2 に示した 29 種類の音響特徴量を抑うつ度推定に用いる.

3. 抑うつ度推定実験

3.1 実験の手順

2 種類の音響分析ツール(openSMILE, Surfboard)で算出した音響特徴量セットを説明変数とする抑うつ度推定手法について比較評価を行う. また, それらを gain に基づいて絞り込んだセットを説明変数とする手法についても比較評価を行う. openSMILE では複数の音響特徴量セットを算出可能だが, 本実験では感情推定に有効とされる eGeMAPsv02 feature set を用いる. gain による絞り込みは LightGBM で特徴量の重要度を算出し, leave-one-out と leave-one-person-out のどちらの交差検証でも重要度が 0.05 以上と算出された特徴量のみを使用する方法で行う. openSMILE, Surfboard で算出した音響特徴量の中から絞り込んだ音響特徴量を表 3 および表 4 に示す. 表 4 において MFCC の低周波数帯域の成分に特徴量を限定している. これは, MFCC の低周波成分は声道の音響特性を表していることから, 健常者と抑うつ状態の話者の間で発話時の口の開き方や舌の動かし方など口腔の形状に影響が出る可能性を示唆している.

また, 機械学習器には Support Vector Regression (SVR) と LightGBM の 2 種類を使用し, leave-one-out と leave-one-person-

表 3 絞り込んだ音響特徴量 (openSMILE)

波形*	説明
HNRdBACF_sma3nz_a mean	27.5Hz を基準とした基本周波数の算術平均
F0semitoneFrom27.5Hz_sma3nz_pctlrange0-2	調波成分と雑音成分の音響エネルギー比の percentile20.0 と percentile80.0 の値の幅
F3frequency_sma3nz_stddevNorm	第 3 フォルマントの中心周波数の変動係数(標準偏差÷算術平均)

*)接尾辞_sma3 は, 時間窓 3 の移動平均フィルタが適用されたデータ(_sma3nz は非ゼロ条件付き)を意味する.

表 4 絞り込んだ音響特徴量 (Surfboard)

波形	説明
mfcc1~mfcc8	13 分割されたメル周波数ケプストラム係数のうち低周波数側の 8 成分
HNR	倍音とノイズの比率

表 5 推定値と正解値の相関係数と決定係数

相関係数 (決定係数)	leave-one-out		leave-one-person-out	
	SVR	LightGBM	SVR	LightGBM
Surfboard	-0.442 (0.195)	0.932 (0.868)	-0.932 (0.872)	0.260 (0.067)
Surfboard (絞り込み)	-0.149 (0.022)	0.852 (0.727)	-0.933 (0.871)	0.344 (0.119)
openSMILE	0.802 (0.643)	0.903 (0.815)	0.080 (0.006)	0.227 (0.051)
openSMILE (絞り込み)	0.465 (0.217)	0.766 (0.586)	0.138 (0.019)	0.255 (0.065)

out によって交差検証を行う.

実験用音声データは, 15 人の実験参加者に同じ 10 文を音読してもらうことで収録した. そして音声収録の直後, 実験参加者にベックうつ病調査表(BDI)に回答してもらうことで各実験参加者の発話時の抑うつ度を算出している.

3.2 実験結果

表 5 に各実験における相関係数と決定係数を示す. leave-one-out と leave-one-person-out の両実験において, SVR より LightGBM の方がより正解値と相関が高い推定結果が得られた. leave-one-out において最も正の相関が強かったのは, “学習器 = LightGBM, 音響特徴量 = Surfboard” の組み合わせによる手法であった. 一方, leave-one-person-out では, “学習器 = LightGBM, 音響特徴量 = Surfboard (絞り込み)” の組み合わせによる手法で最も相関が高かった.

次に, 学習器に LightGBM を用いた抑うつ度推定手法の結果の散布図を図 3 および図 4 に示す. 図 3 は leave-one-out, 図 4 は leave-one-person-out の結果である. leave-one-person-out は leave-one-out に比べ, 同一人物の推定値のばらつきは少ないが, 抑うつ度が極端に高いまたは低い実験参加者の抑うつ度推定がうまくいっていないように見える. また, 図 4(d) では抑うつ度が中程度(4 以上 11 以下)のデータに限定すれば高い正の相関がみられた.

3.3 考察

3.2 節の結果より, leave-one-out 交差検証では十分な抑うつ度推定性能が確認できたが, leave-one-person-out 交差検証における推定値の相関は十分ではなかった. そこで本節では leave-one-person-out による推定精度の向上について考察する. 最も相関係数が高かった “学習器 = LightGBM, 音響特徴量 =

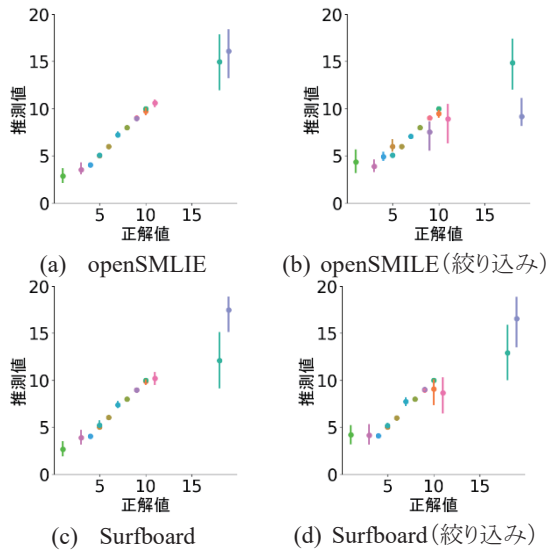


図3 leave-one-out 交差検証による抑うつ度推定結果

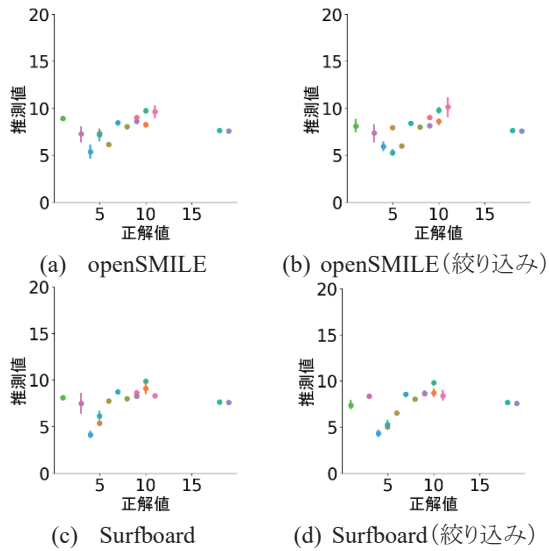


図4 leave-one-person-out 交差検証による抑うつ度推定結果

Surfboard(絞込み)”の組み合わせによる実験結果(図4(d)参照)において、正解抑うつ度が低いデータ(1, 3)と極端に高いデータ(18, 19)を除くと強い正の相関があるように見える。

そこで、正解抑うつ度が3以下または18以上である実験参加者のデータを除外し、LightGBMを用いて3.2節と同条件で再度抑うつ度推定実験を行った。表6に学習データごとの相関係数と決定係数、図5に散布図を示す。

実験の結果、いずれの音響特徴量を用いた実験においても抑うつ度が中程度のデータに限定した場合に大幅な相関係数の向上がみられた。特に、音響特徴量にSurfboardを用いた学習器において相関係数0.798、決定係数0.637という最も良い結果が得られた。このことより、学習に用いていない人物についても、中程度の抑うつ度であれば提案手法で推定可能であることが示された。これは中程度の抑うつ度の実験参加者が多いため、データを限定させた場合に相関係数の向上がみられたと考えられる。

表6 中程度の抑うつ度に対する推定結果の相関

相関係数 (決定係数)	openSMILE	openSMILE (絞込み)	Surfboard	Surfboard (絞込み)
抑うつ度 4~11に限定	0.569 (0.435)	0.665 (0.443)	0.798 (0.637)	0.777 (0.603)
全データ	0.227 (0.051)	0.255 (0.065)	0.260 (0.067)	0.344 (0.119)

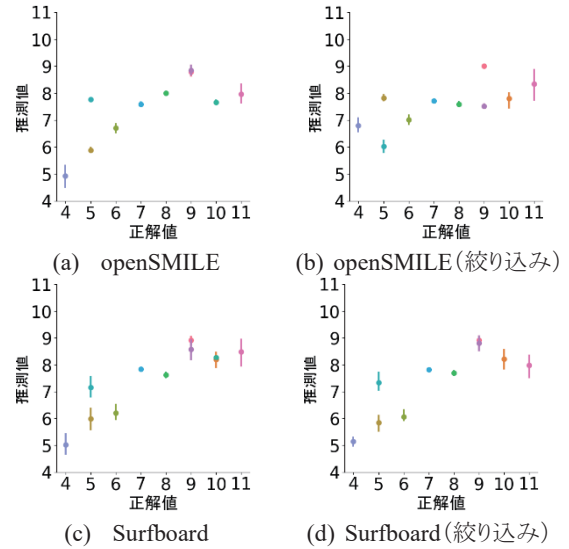


図5 中程度の抑うつ度に限定した抑うつ度推定結果

4. まとめ

本研究では、入力発話から得られた音響特徴量を用いて話者の抑うつ度を推定する手法を提案した。学習フェーズでは、説明変数として発話音声から算出された音響特徴量、目的変数として発話者の抑うつ度を学習させた。そして推定フェーズでは、発話者の音声の音響特徴量から、発話者の抑うつ度を推定した。

2種類の音響分析ツール(openSMILE, Surfboard)で算出した音響特徴量セットを用いた機械学習実験を行った結果、leave-one-outにおいて最も正の相関が強かったのは、“学習器=LightGBM、音響特徴量=Surfboard”の組み合わせによる手法であった(相関係数=0.932)。一方、leave-one-person-outでは、“学習器=LightGBM、音響特徴量=Surfboard(絞込み)”の組み合わせによる手法で最も相関が高かった(相関係数=0.798)。

今後はデータ数や実験参加者の年齢層を広げた実験を行う予定である。また音読してもらったセリフ数やセリフの種類を変えた実験を行い、leave-one-person-outによる推定精度の向上を目指す予定である。

謝辞

本研究の一部は国立研究開発法人科学技術振興機構(JST)の「COIプログラム令和4年度加速支援(Grant番号JPMJCA2208)」の支援によって行われた。

参考文献

- [Beck 61] Beck, A. T.; Ward, C. H.; Mendelson, M.; Mock, J.; Erabaugh, J. An inventory for measuring depression. *Psychiatry*. 1961.
- [Burns 90] Burns, D, D. いやな気分よさようなら～自分で学ぶ「抑うつ」克服法. 星和書店, 1990.
- [Eyben 10] Eyben, F.; Wöllmer, M.; Schuller, B. openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor. *ACM Multimedia Conference – MM*. 2010, pp.1459-1462.
- [Lenain 20] Raphael L.; Jack W.; Abhishek S.; Emil F. Surfboard: Audio Feature Extraction for Modern Machine Learning. *Proceedings of INTERSPEECH2020*. 2020, pp.2917-2921. doi: 10.21437/Interspeech.2020-2879.
- [Low 20] D. M. Low.; K. H. Bentley.; S. S. Ghosh. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*. 2020, vol. 5, no. 1, pp. 96–116.
- [OECD 21] OECD. Tackling the mental health impact of the COVID-19 crisis: An integrated, whole-of-society response. *OECD Policy Responses to Coronavirus (COVID-19)*. OECD Publishing. 2021. <https://doi.org/10.1787/0ccafa0b-en>.
- [宗 16] 宗未来, 竹林由武, 関沢洋一, 下地貴明. “声”だけで, うつ病はどこまで診断可能か? ～音声感情認識技術にアンサンブル型機械学習モデルを応用したうつ病スクリーニング機能に関する精度の検証. *経済産業研究所ディスカッション・ペーパー*. 2016, 16-J-054.