

「心の健康」促進に向けた精神疾患支援モデルの構築

井上佳花¹, 大場春佳², 水野信也¹

キーワード: 精神疾患リスク推定, 自然言語処理, 深層学習, SNS テキスト分析, 特徴語辞書構築

1. 研究背景と目的

現代社会においては、精神疾患や心理的ストレスを抱える人々の数が年々増加しており、それに伴って、適切な医療的支援を必要とするケースに十分対応できていないという深刻な課題が顕在化している[1]。とりわけ重症度の高い患者であっても、精神科医療機関における予約の長期化や、精神疾患に対する根強い社会的偏見、さらには診断結果を受け入れることへの心理的抵抗感などが複合的に作用し、受診の遅れを招く事例が少なくない。その結果、治療開始までに時間を要し、症状の悪化や社会的孤立の進行といった二次的リスクが高まる傾向が指摘されている。

先行研究においても、SNS データや日常的な言語表現から精神的健康状態を推定する試みが行われている[2,3]。こうした背景を踏まえ、本研究では、自然言語処理 (Natural Language Processing; NLP) および機械学習技術を活用し、日常的な言語表現から精神的リスクを早期に検出し、必要に応じて医療機関との連携を促す新たな支援モデルの構築を目的とする (図 1)。具体的には、SNS 投稿、日記、アンケートなどのオープンなテキストデータを分析対象とし、心理状態を反映する言語的特徴を抽出・数値化することで、精神疾患の有無や重症度傾向を推定するアルゴリズムの開発を行う。

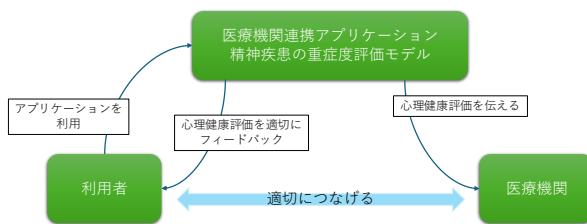


図 1 医療機関との連携を促す新たな支援モデル

2. 分析方法および結果

2.1. データセット選定

本研究では、精神疾患の有無を分類する機械学習モデルの構築を目的とする。自殺リスク検出モデルの本構築に先立ち、複数の公開データセット[4-6]から予備的に抽出したデータを用いて、BiLSTM による予備実験を実施し、各デ

ータセットの分類性能および学習特性を比較した。その結果、Kaggle Suicide Detection は最も高い分類精度を示し、学習過程も安定して収束した。一方、Dreaddit ではデータ規模の影響により過学習の傾向が確認され、SDCNL ではデータ件数が少ないとから汎化性能に課題が残った。これらの結果を踏まえ、本研究では Kaggle Suicide Detection データセット[4]を主要データセットとして採用することとした。

2.2. 精神疾患判定モデルの構築

本節では、Kaggle Suicide Detection データセットを用いて構築した SNS テキストに基づくリスク推定モデルの性能評価結果を示す。ロジスティック回帰、ランダムフォレスト、線形 SVC、ナイーブベイズの 4 種類の機械学習モデル、LSTM、BiLSTM、CNN の 3 種類の深層学習モデル、および事前学習済み言語モデルである BERT の計 8 モデルを比較した結果、機械学習モデルではロジスティック回帰と線形 SVC がそれぞれ AUC 0.9309, 0.9305 と高い性能を示した一方、ナイーブベイズおよびランダムフォレストはやや低い性能にとどまった。深層学習モデルはいずれも AUC 0.98 以上と非常に高い性能を示し、これは大規模かつ高品質でクラスバランスの取れたデータにより安定した表現学習が可能であったことに起因すると考えられる。また、小規模データセットにおいては CNN および BiLSTM が比較的安定した性能を示したのに対し、LSTM は性能低下が顕著であり、データ規模への依存性が確認された。さらに、BERT モデルは AUC 0.9981, F1 Score 0.98, Accuracy 0.98 とすべての指標で最高性能を達成し、文脈理解能力に基づく高度な特徴抽出が本タスクにおいて極めて有効であることが示された。

表 1 モデル性能比較

モデル	分類手法	AUC Score	その他指標
ロジスティック回帰	機械学習	0.9309	F1: 0.87 / Acc: 0.86
線形 SVC	機械学習	0.9305	F1: 0.90 / Acc: 0.87
ナイーブベイズ	機械学習	0.9185	F1: 0.85 / Acc: 0.83
ランダムフォレスト	機械学習	0.9135	F1: 0.85 / Acc: 0.83
BiLSTM	深層学習	0.9852	F1: 0.97 / Acc: 0.96
CNN	深層学習	0.9849	F1: 0.96 / Acc: 0.95
LSTM	深層学習	0.9848	F1: 0.97 / Acc: 0.95
BERT	深層学習	0.9981	F1: 0.98 / Acc: 0.98

1 順天堂大学

2 中央大学

2.3. 精神疾患特徴語辞書の構築

BiLSTM モデルによる予測結果および誤分類データの詳細分析を基に、精神疾患リスクを示唆する語彙を体系的に整理した精神疾患特徴語辞書を構築した。まず、non-suicide および suicide テキストに対する頻出語分析、TF-IDF 分析、Word Cloud、可視化、共起ネットワーク分析、Word2Vec クラスタリングを通じて、suicide テキストでは「kill」「suicide」「die」「worthless」「hopeless」など自己破壊的・絶望的語彙が顕著である一方、non-suicide テキストでは日常生活や対人関係に関する中立的・肯定的語彙が中心となることを確認した(図 2)。さらに、誤分類データの分析から、皮肉表現や比喩表現、多義語、助けを求める表現に含まれる中立・肯定語彙などが誤判定の主要因であることが明らかとなった。これらの知見を基に、TF-IDF 上位語と共起語情報を統合した Negative 特徴語辞書(約 500~600 語)と、False Negative データに特化した補助的特徴語辞書(約 200~300 語)を構築し、単語単体ではなく文脈情報を考慮した評価を可能とした。その結果、構築した辞書は約 95% の網羅性と高い判別力を有し、BiLSTM モデルの判定精度向上および解釈可能性の強化に有効であることが示された。

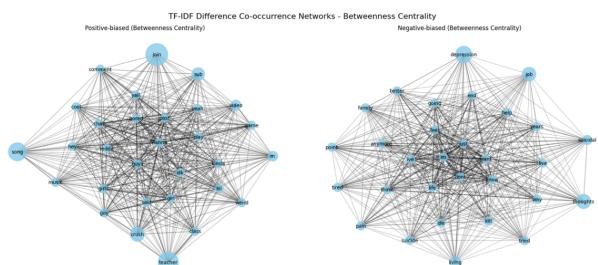


図 2 TF-IDF 差分共起ネットワーク(媒介中心性)

3. 考察

本研究の結果から、BiLSTM、CNN、BERT といった深層学習モデルは、従来の機械学習モデルと比較して精神疾患リスク推定において顕著に高い性能を示した。これは、精神疾患に関する言語表現が強い文脈依存性を持つという特性と、深層学習モデルの文脈理解能力および階層的特徴抽出能力が適合したためであると考えられる。特に BiLSTM は、双方向の文脈情報を同時に考慮できる点で、自殺関連表現に含まれる多義性や意味の揺らぎを適切に捉えることが可能であり、高い識別性能につながった。また、特徴語分析の結果から、自己否定、孤立、絶望、無力感、自殺願望といった語彙が体系的なクラスターを形成していることが確認され、精神疾患リスクが特定の語彙集合と文脈構造によって表現されていることが示唆された。これらの知見は、単語の出現頻度だけでなく、語彙間の関係性や意味的まとまりを考慮することの重要性を示している。

一方で、誤分類分析から、比喩表現、婉曲表現、多義語、および助けを求める表現に含まれる中立・肯定的語彙が、自殺リスク検出を困難にする主要因であることが明らかと

なった。この課題に対し、本研究では精神疾患特徴語辞書を構築し、文脈依存性や共起関係を考慮した語彙的知識の導入がモデルの信頼性と解釈可能性を高める可能性を示した。

4. 結論と今後の展望

本研究では、自然言語処理および機械学習・深層学習手法を用いて、SNS 投稿テキストから精神疾患リスクを高精度に推定するモデルの構築と分析を行った。複数の公開データセットと多様なモデルを比較した結果、規模およびラベル品質の観点から Kaggle Suicide Detection データセットが最適であることを明らかにし、特に BERT および BiLSTM が高い分類性能を示すことを確認した。さらに、TF-IDF 分析、頻出語分析、共起ネットワーク分析、Word2Vec クラスタリングを通じて精神疾患リスクを示す特徴語とその関連構造を体系的に抽出し、文脈情報や共起関係を含む精神疾患特徴語辞書を構築したことで、モデル性能の向上と判定根拠の可視化を可能とした。これらの結果は、実用レベルの精神疾患支援システムの実現可能性を示すものである。

今後は、多言語対応モデルの構築や対話型システムへの実装、時系列的な精神状態変化の分析、および医療専門家との連携を視野に入れた支援体制の整備が重要な課題となる。本研究で提案した手法は、精神疾患の兆候を早期に捉え、適切な支援へつなぐ一次スクリーニングとして有効であり、医療現場の負担軽減や相談の心理的障壁の低減にも寄与する可能性がある。SNS など日常的なテキストデータを活用した継続的なメンタルヘルス支援の新たなアプローチとして、本研究の成果は高い社会的意義を有すると結論づけられる。

参考文献

- [1]赤松達也、大槻克文、秋山敏夫、斎藤裕、矢内原巧、木村武彦、早川達郎、富山三雄、亀井雄一. 婦人科外来における精神疾患合併更年期障害患者の取り扱い. 日本産科婦人科学会雑誌. 1996;48(9):806-812.
- [2]木口恵美子. SNS 投稿テキストを用いた精神状態推定に関する研究. 情報処理学会論文誌. 2022;63(4):1234-1245.
- [3]木下真理、岸本泰士郎. テキストマイニングを用いたメンタルヘルス評価の可能性. 日本精神保健学会誌. 2020;29(1):78-89.
- [4]Komati NE. Suicide and Depression Detection Dataset. Kaggle; 2020. Available from:<https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch> [Accessed 2024].
- [5]Turcan E, McKeown K. Dreaddit: A Reddit Dataset for Stress Analysis in Social Media. arXiv preprint arXiv:1911.00133. 2019.
- [6]Zhaque A. SDCNL: Suicide vs Depression Classification using NLP. GitHub; 2020. Available from:<https://github.com/ayaanzhaque/SDCNL> [Accessed 2024].