

精神障害を視野に入れたツイート行動における感情状態の検討

市村 真衣¹ 久野 雅樹¹¹ 電気通信大学大学院 情報理工学研究科 情報学専攻
i2230018@edu.cc.uec.ac.jp hisano@uec.ac.jp

概要

健全と精神障害を視点として、感情という現象の構造を解析することを目指す。本稿ではツイートを分析しており、ユーザーごとに Plutchik の 8 感情の推測値を箱ひげ図でまとめたところ、感情に個人差を確認できた。また、精神障害クラスは対照クラスより投稿量が少なく、社会から解離気味であった。コロナ前とコロナ中の年月別投稿量や感情の推測値を算出すると、両クラスともにコロナ禍以前よりネガティブな感情に寄っており、感情面において精神障害クラスは現状の精神疾患以上にコロナ禍による影響が大きいと推測できた。

1 はじめに

感情は可変であり、人にとって判断材料やコミュニケーション手段になる。学術および社会では感情の典型を追究しており、カテゴリ分類や極性を用いた感情推定方法の構築によって、現象の構造解析に取り組んでいる。精神障害の症状を気分や感情の観点からみると、うつ病はやり場のない苦しみ、双極性障害は爽快で楽しい状態と落ち込む状態の転換、統合失調症は感情の平滑化など、多種多様である。

高須ら (2022) は、ツイートには精神障害の薬剤名、プロフィール欄には精神障害の疾病名や症状の記載があるユーザーを対象に、メンタルヘルス不調に関わる表現や薬剤名を除外した一般的な会話表現から、機械学習による分類によってメンタルヘルス不調を推測した。BERT の F1 値が 0.84 であった一方、対象となる障害群 (気分障害圏、統合失調症圏及び神経症性障害、ストレス関連障害及び身体表現性障害圏) が広く、不調の種類が不明瞭である [1]。

Suhavi ら (2022) は、8 つの精神障害クラスと 1 つの対照クラス (詳細は表 2 を参照) に分類している、時間的文脈のあるユーザーレベルのツイートデータセット「Twitter-STMHD」を構築した。基本的な統計量や LIWC による分析例を示したが、クラス内外の

ユーザー数の差を含んだツイート総数でツイート行動を検討していた [2]。

本研究では、感情の個人差を NRCLex による推測値から箱ひげ図で示し、精神障害クラスや対照クラスの社会参加を投稿量の時系列比較で表した。また、コロナ前 (2018 年 1 月～2019 年 12 月) とコロナ中 (2020 年 1 月～2021 年 5 月) における、精神障害クラスや対照クラスの行動や感情を年月別投稿量や NRCLex による感情の推測値で調べた。

2 感情推測ツール「NRCLex」

以降の分析で感情分析ツール「NRCLex」を使用するために、固有の特徴を事前調査した。

2.1 手順

NRCLex とは、Plutchik の 8 感情 (joy[喜び], trust[信頼], fear[恐れ], surprise[驚き], sadness[悲しみ], disgust[嫌悪], anger[怒り], anticipation[期待]) の推測値を、0.0～1.0 の範囲で、8 つの合計値が 1.0 になるように算出する感情分析ツールである [3]。“NRC Word-Emotion Association Lexicon”[4] と NLTK ライブラリの WordNet の同義語セットに基づいた 4,463 語を含み、1 語に複数の感情ラベルを許す。例えば、“raving”には joy・fear・surprise・anger・anticipation・negative のラベルが付いている。表 1 に、ポジティブネガティブと 8 感情のクロス関係を提示した。

2.2 結果と考察

表 1 より、joy[喜び]・trust[信頼]・anticipation[期待] はポジティブの収録語数、fear[恐れ]・sadness[悲しみ]・disgust[嫌悪]・anger[怒り] はネガティブの収録語数が多かった。Jingyi ら (2020)[5] が設定したポジティブネガティブ区分に類似している。また、Jingyi ら (2020)[5] が設定したポジティブネガティブ区分によって収録語数を整理すると、ネガティブ (fear・sadness・disgust・anger) は 4,972 語に対して、ポジティブ (joy・trust・surprise・anticipation) は 3,296 語と少

表 1 NRCLex のラベルより、ポジティブネガティブ×8感情の各語数

	(なし)	positive	negative	(両方)	合計
joy	6	646	5	32	689
trust	334	826	44	27	1,231
fear	174	71	1,202	29	1,476
surprise	117	198	196	23	534
sadness	80	42	1,054	15	1,191
disgust	69	18	959	12	1,058
anger	63	38	1,125	21	1,247
anticipation	213	449	143	34	839

なかった。ネガティブな単語の収録語数が多いのは、市村ら (2023)[6] で示したように、「単語感情極性対応表」(高村ら, 2006)[7] や「日本語評価極性辞書」(小林ら, 2005)(東山ら, 2008)[8][9] と同様であった。

3 分析 1：感情の個人差

3.1 対象

データセットとして「Twitter-STMHD」(Suhavi ら, 2022)[2] を取り扱う。Twitter より、ユーザーが精神疾患 8 つのいずれかと診断されたと主張するツイートの 2 年前から 2 年後までのウィンドウで、各ユーザーのツイートを収録している。また、COVID-19 によるパンデミック期間中(2021 年 5 月まで)のツイートも、ウィンドウの期間に関係なく収集している。各ツイートログには、ツイート作成日時や内容などを含んでいる。データセット全体のツイート数は約 3 億 294 万件、ユーザー数は 34,902 人であり、大半が英語である。表 2 に収録クラスの一覧を示す。

データセットの特徴「ユーザーごとに収集期間の設定がある」により、時期によってデータセット内のユーザー数には差が生じている。投稿量を時系列的な分析によってクラス内外の行動履歴を比較するためには、ユーザー数を考慮する必要がある。

3.2 手順

精神障害クラスのユーザーと対照クラスのユーザーとの感情を比較するために、NRCLex による感情分析を行なった。まず、上記のデータセットより、2018 年 1 月～2021 年 5 月のツイートを抽出したデータセットを作成した。各クラスのユーザー数は、対照クラスは 8,144 名、注意欠如・多動症は 7,733 名、不安障害は 3,946 名、強迫性障害は 1,211 名、心的外

表 2 Twitter-STMHD の収録クラス一覧

ラベル名	説明
ADHD	attention-deficit/hyperactivity disorder, 注意欠如・多動症
Anxiety	anxiety disorder, 不安障害
OCD	obsessive-compulsive disorder, 強迫性障害
PTSD	post-traumatic stress disorder, 心的外傷後ストレス障害
Bipolar	bipolar disorder, 双極性障害
MDD	major-depressive disorder, 大うつ病性障害
PPD	post-partum depression, 産後うつ病
Depression	うつ病
Control-User	対照クラス

傷後ストレス障害は 3,237 名、双極性障害は 1,434 名、大うつ病性障害は 264 名、産後うつ病は 211 名、うつ病は 5,995 名である。ユーザーにクラス間の重複はない。ただし、1 人あたりの 3 年 5 ヶ月間のツイート総数に開きがあったため、1 人 1,000 件を目安に年月別の層化抽出を施した。次に、各ツイートの NRCLex によって感情値 8 種を算出し、ユーザーごとに平均を計算した。

3.3 分析結果と考察

図 1 は、2018 年 1 月～2021 年 5 月における、不安障害クラスのユーザー別の Plutchik の 8 感情を、箱ひげ図で示したものである。なお、箱ひげ図のひげは四分位範囲の 1.5 倍であり、箱ひげ図に重ねて示されているカラーのポイントは 1 点 1 ユーザーを表す。箱ひげ図のひげの長さが、不安障害クラス以外も含め、joy[喜び] が最大であり、surprise[驚き] が最小であった。ひげが長いのは、ユーザー間のバラツキが大きいことを示している。よって、8 感情の内では、joy は個人差が出やすい感情であり、surprise は個人差が小さい感情であると推測できる。

4 分析 2：各クラスの曜日別投稿量

4.1 手順

精神障害クラスと対照クラスとの行動を比較するために、1 時間 1 人あたりの投稿量を曜日別で調査した。データセットは、3.1 項をベースとした。まず、タイムゾーンの異なる地域がデータセット内に混在す

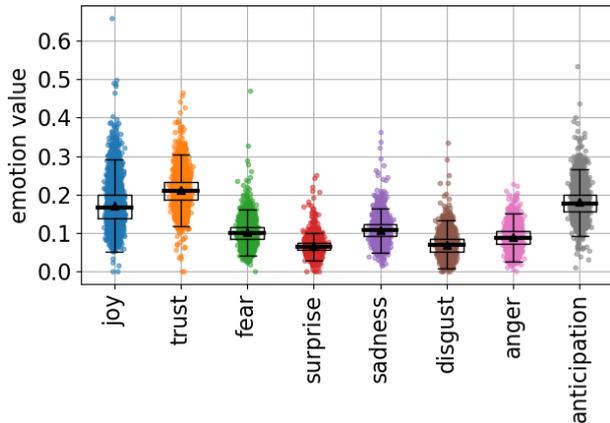


図1 2018年1月～2021年5月における、不安障害クラスのユーザー別のPlutchikの8感情

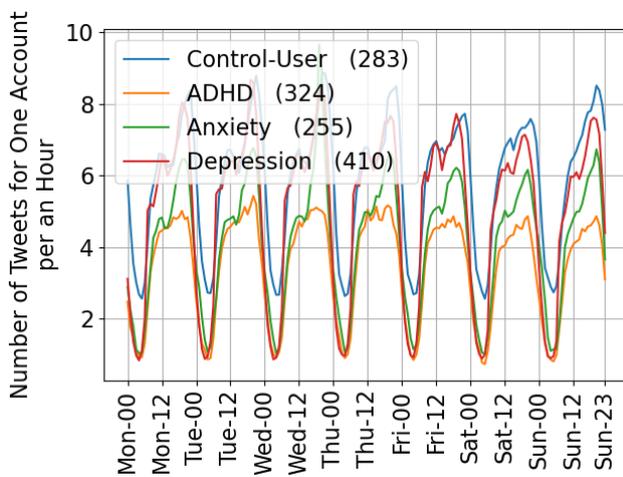


図2 1時間・1人あたりの曜日別投稿量(対照クラス, 注意欠如・多動症, 不安障害, うつ病)(横軸は「曜日-24時制」で表記)

ることを防ぐために、ユーザー情報に英国と記載のあるユーザーに限定したデータセットを作成した。各クラスのユーザー数は、対照クラスは283名、注意欠如・多動症は324名、不安障害は255名、強迫性障害は68名、心的外傷後ストレス障害は233名、双極性障害は95名、大うつ病性障害は2名、産後うつ病は2名、うつ病は410名である。ユーザーにクラス間の重複はない。次に、(曜日) × (時刻) ごとに件数をカウントし、曜日あたりの日数から1時間あたり、そして、ユーザー数から1人あたりの投稿量になるように計算した。なお、曜日は月曜日から日曜日であり、時刻は0時から23時までの分切り捨てである。

4.2 分析結果と考察

図2は、2016年1月～2021年5月の1時間・1人あたりの曜日別投稿量を、時系列変化で表した。紙面の関係上、データセット内のユーザー数が多いクラ

スである、対照クラス、注意欠如・多動症、不安障害、うつ病を作画した。全クラス共通で、火曜日と水曜日が最大であり、逆に金曜日と土曜日は最小であった。現実にある自分の世界に入り込むと、投稿量が減るのだと予想する。また、精神障害クラスは、対照クラスと比べて、何時でも投稿量が少なかった。精神障害クラスは対照クラスより、社会から解離気味であると推測できる。

5 分析3：コロナ前中の行動と感情

5.1 手順

まず、COVID-19によるパンデミック前後の行動履歴を比較するために、コロナ前である2019年1月からコロナ中の2021年5月にかけて、1ヶ月(30日間)・1人あたりの年月別投稿量を調査した。データセットは3.1項をベースとし、年月別にユーザー数、投稿量、日数を算出した。ただし、特定年月に逸脱してツイートするユーザーを外れ値をみなして、年月ごとに四分位範囲の1.5倍の外を削除した。投稿量をユーザー数で割ったのち、2月は28日(2024年は閏年のため29日)のような月別の日数差を解消するために、30日換算をした。

そして、コロナ前(2018年1月～2019年12月)とコロナ中(2020年1月～2021年5月)において、精神障害クラスと対照クラスとの感情を比較するために、期間を分けてNRCLexによる感情分析を行なった。データセットは、分析1と同様である。各ツイートのNRCLexによって感情値8種を算出し、コロナ前およびコロナ中の2期間に分けて平均を計算した。

5.2 分析結果と考察

図3は、2019年1月～2021年5月の1ヶ月(30日間)・1人あたりの年月別投稿量の時系列変化を表している。PPDを除いた精神障害クラス7つは、2020年2～3月に、Suhaviら(2022)[2]の図7では微増であったが、図3では急増しているのが分かる。新型コロナウイルスのパンデミックによる外出制限時に、Twitterを社会と繋がるプラットフォームとしてより活用していたのだと推測できる。

図4は、コロナ前(2018年1月～2019年12月)およびコロナ中(2020年1月～2021年5月)における、精神障害クラスと対照クラスとのPlutchikの8感情をレーダーチャートで示したものである。コロナ前とコロナ中に共通して言えるのは、

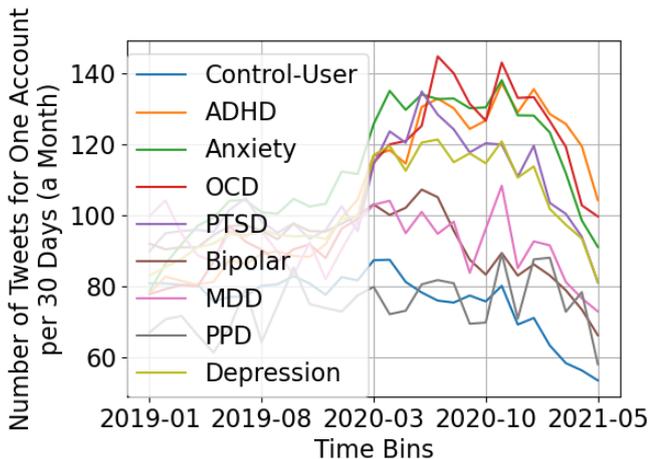


図3 1ヶ月(30日間)あたり・1人あたりの年月別投稿量(精神障害クラス8つと対照クラス1つ)

どのクラスも trust[信頼] > (anticipation[期待], joy[喜び]) > (sadness[悲しみ], fear[恐れ]) > anger[怒り] > disgust[嫌悪] > surprise[驚き]の順番で感情値が並んだことだ。精神障害クラスと対照クラスは同じく、通常時も非常時も、基本的にはポジティブな感情をより含んだツイートをしていると推測できる。また、クラス間の特徴がコロナ前では多少バラついているが、コロナ中はレーダーチャートがより重なり合っていて、クラス間の違いが小さくなっている。よって、感情面では、現状の疾患以上にコロナ禍による影響が大きくなっているといえる。変化率の絶対値の平均より、産後うつ病、うつ病、強迫性障害は対照クラスと比べて変化が大きく、非常時に対応したケアへの変更が特に必要だと考える。

コロナ前に対するコロナ中の変化率を細かく観察するために、図5を示した。コロナ前をベースとしたとき、コロナ中における精神障害クラスと対照クラスとのPlutchik8感情の変化率を、レーダーチャートで表現したものである。パーセンテージが正の値であるとコロナ中に増加傾向、負の値であるとコロナ中に減少傾向であるといえ、特に「fear[恐れ]の増加」と「joy[喜び]の減少」が大きい。Mohamedら(2021)[10]より、#covid-19付きツイートでは楽観的な傾向にあると述べているが、コロナ禍中は以前よりネガティブな感情になっていると推測できる。

6 おわりに

NRCLexの内容分析および分析1より、NRCLexは日本語の感情辞書と同じくネガティブ語の収録が多く、個人差が現れやすい/現れにくい感情が存在することが分かった。そして、分析2と分析3より、コ

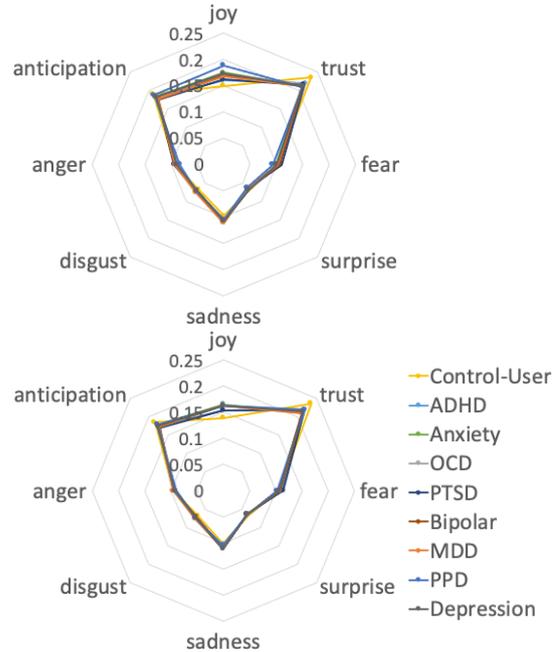


図4 コロナ前(2018年1月~2019年12月,上図)とコロナ中(2020年1月~2021年5月,下図)における、精神障害クラスと対照クラスとのPlutchikの8感情の推測値

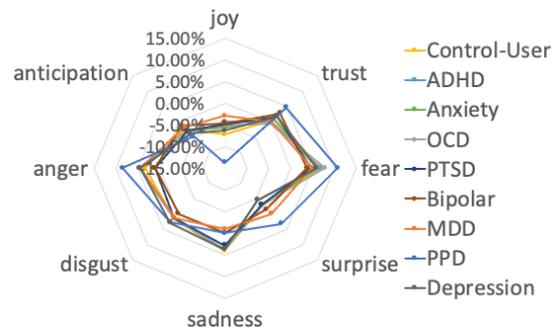


図5 コロナ中(2020年1月~2021年5月)における精神障害クラスと対照クラスとのPlutchikの8感情の推測値の、コロナ前(2018年1月~2019年12月)に対する変化率

ロナ中はコロナ前より投稿量が急増していることから、Twitterを社会のプラットフォームとして活用していた。一方で、精神障害クラスは対照クラスより曜日別投稿量が少なく、社会から解離気味であると推測した。また、分析3より、基本的にはポジティブな感情であるが、コロナ以前よりはfearの増加とjoyの減少によりネガティブな感情に寄っていた。精神障害クラスと対照クラスはコロナ中に似た感情分布をしており、感情面では現状の精神疾患以上にコロナ禍による影響が大きいと推測できた。

今回は基本的な統計処理にとどまっているため、今後はより高度な情報処理の方法を用いて、健常および精神障害を視点とした感情の構造分析をしたい。

参考文献

- [1] 高須遼, 中村啓信, 岸本泰士郎, 狩野芳伸. 大規模ツイートデータを用いたメンタルヘルス不調者の推測. 人工知能学会全国大会論文集 第 36 回, 2022.
- [2] Suhavi, Asmit Kumar Singh, Udit Arora, Somyadeep Shrivastava, Aryaveer Singh, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. Twitter-STMHD: An extensive user-level database of multiple mental health disorders. **Proceedings of the International AAI Conference on Web and Social Media**, Vol. 16, No. 1, pp. 1182–1191, May 2022.
- [3] Mark M. Bailey. NRCLex, 2019. <https://github.com/metalcorebear/NRCLex/>.
- [4] Saif Mohammad and Peter Turney. Crowdsourcing a word-emotion association lexicon. **Computational Intelligence**, Vol. 29, No. 3, pp. 436–465, 2013.
- [5] Jingyi Li, Tianming Liu, Adam Goodie, and Frederick Maier. Twitter sentiment analysis during covid-19 in florida. Master of Science (MS), University of Georgia, 12 2020.
- [6] 市村真衣, 久野雅樹. 感情カテゴリを考慮した単語極性の推定. 言語処理学会第 29 回年次大会 (NLP2023), 2023.
- [7] 高村大也, 乾孝司, 奥村学. スピンモデルによる単語の感情極性抽出. 情報処理学会論文誌ジャーナル, Vol. 47, No. 02, pp. 627–637, 2006.
- [8] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol. 12, No. 03, pp. 203–222, 2005.
- [9] 東山昌彦, 乾健太郎, 松本裕治. 述語の選択選好性に着目した名詞評価極性の獲得. 言語処理学会第 14 回年次大会論文集, pp. 584–587, 2008.
- [10] Mohamed Chiny, Marouane Chihab, Omar Bencharef, and Younes Chihab. Analysis of sentiments conveyed through twitter concerning COVID-19. **SHS Web Conf**, Vol. 119, No. 07003, 2021.