

SNS データを使った「メンヘラ」の予測

伊藤 智彦^{1,a)} 中里 秀則¹

概要: 愛情が暴走して他者を傷つけてしまったり、SNSで「死にたい」とツイートするような人は「メンヘラ」と呼ばれ、社会問題化している。このような人が自分自身や周囲の人を傷つけ、大きな問題になる前に適切にフォローする必要がある。本稿では、このメンヘラに該当する人を見つけることを目指し、ツイートデータの学習による探索を行った結果を報告する。特にメンヘラカルチャーを代表する「#病み垢さんと繋がりたい」というハッシュタグに注目し、このハッシュタグに投稿する人が、メンヘラの性向をもっていると仮定し、そのハッシュタグをもつツイートを学習データとして、一般のツイートから同様の性向を探索した。

Predicting “Menhera” Using Social Network Data

1. はじめに

近年、「メンヘラ」という単語がある。この単語は2010年頃に掲示板「2ちゃんねる」(現「5ch」)[8]のメンタルヘルス板から始まるスラングであり、まだ厳密な定義はない。

一般的には「メンヘラ」という単語には否定的な印象を抱くことが多く、それはメンヘラにまつわるネガティブな話題が多いからだろう。特に2019年5月に発生した好きすぎて彼氏を刺した事件[5]は記憶に新しいだろう。この女性は「好きで好きでしかたなかった」「一緒にいるためには殺すしかないと思った」と語っており、支配欲が暴走した結果だと考えられている。この事件はまさに「行きすぎたメンヘラ」の結果だと捉えた人は多く、メンヘラな人と言えば否定的な印象を抱くことに拍車がかかったように思う。

この事件はあまりにも極端だが、ニュースに取り上げられるような派手な事件ではなくとも、日常的に問題は起こっており、メンヘラは社会問題化している。現在、Twitterなどのマイクロブログやその他のSNSには様々な情報が発信されているが、このSNSを起点にした問題が特に多い。例えば、Twitterで「つらい」「寂しい」とツイートしているアカウントに片っ端からDMを送り、優しい人のふりをして心の隙間につけ込もうとするユーザがいる。他にも

「家出したい」「死にたい」といったツイートをする、悩んでいたり寂しかったりする若者が狙われる傾向にある[1]。

「つらい」や「死にたい」などの精神的に病んでいるようなツイートをすることで日頃の愚痴や悩みを吐き出すためのアカウントは「病み垢」と呼ばれる。もちろん「病み垢」自体は悪いものではなく、自身の悩みをツイートしたり、他のユーザから共感されたり同じ悩みを抱える人に相談することで救われる人がたくさんいることも確かである。SNSによってはネガティブな内容を発信しようとするサービス側から発言を規制・制限をされたり、運営からメッセージが届くものがある。例えばTwitterでは特定のキーワードで検索をすると検索欄の一番上に自殺防止センターへの連絡先が表示されるという仕様がある。このような機能や仕様は一定の成果をあげてはいるものの、拾いきれていないユーザ層もいる。

これらの社会的な背景から、SNSを禁止したり利用に制限をかけたりすることなく、このような問題や事件を未然に防ぐ必要があると考える。そこで、本研究ではまずはトラブルに巻き込まれ得る「病み垢」と病み垢カルチャーを代表する「#病み垢さんと繋がりたい」というハッシュタグに注目し、Twitterのデータからツイート主がメンヘラかどうかを予測した。

2. メンヘラとは

「メンヘラ」という単語はスラングであり、まだ厳密な定義はない。本章ではその「メンヘラ」についてまとめた。

¹ 早稲田大学
Waseda University

^{a)} thomasito@akane.waseda.jp

2.1 語源

「メンヘラ」という単語の由来は2010年頃に掲示板「2ちゃんねる」(現「5ch」)[8]のメンタルヘルス板から始まる。メンタルヘルス板の正式名称は「メンタルヘルス@2ch 掲示板」で、精神疾患・精神科に関する専門的な話題を扱う掲示板である。当事者同士が意見交換するなかでメンタルヘルス(精神衛生)の略語として「メンヘル」が広まりはじめた。そして、このような掲示板を利用する人を英語の接尾辞 -er を加えて「メンヘルな人」という意味で「メンヘラー」という言葉が生まれた。これが後に「メンヘラ」へ変化したとされている[10]。

2.2 現代の「メンヘラ」の解釈

語源から精神疾患という印象を抱きやすいこの単語だが、使われていくうちに「メンヘラ」が指すものが変わってきた。特に、若い世代では比較的ライトで恋愛に関連するイメージで使われている。実際、近年では「誰かに愛情を向けてもらいたいという傾向を持つ」[34]、「ニュアンスの曖昧な言葉」、「否定的な言葉とは言い切れない側面も持ち合わせています」[6]という解釈がなされている。この「否定的な言葉とは言い切れない側面」というのは、ファッション要素としての「メンヘラ」があるからである。代表的な例として「病みかわいい」呼ばれるファッションのジャンルがある。この言葉は江崎びす子によって生まれたもので、「『ゆめかわいい』という言葉に『病み』の要素がたされた派生語」であり、「病みの要素とは『自殺・殺人』『暴力・暴言』『リスカ・自傷』『薬』など」だという[11]。他にも「地雷メイク」などもあり、ここにおいて、「メンヘラ」は一つの自己表現として使われており、恋愛文脈とは違った独自の意味合いで使用されている。さらに、異性にかまってもらいたいためにメンヘラを装う「ファッションメンヘラ」という言葉もあり、「メンヘラ」はある種ブランド化し、メンヘラであることをステータスにすることもある。[4]ではファッションとして病んでいる風な演出をすることを“Beautiful Suffering”と表現し、問題視している。

メンヘラの特徴は境界型パーソナリティ障害と似ているため、精神疾患と混同されることもあるが、ここまで見てきた通り語源が精神疾患の文脈から来るだけであり、「メンヘラ」という単語はなんらかの精神疾患を表すものではない。

ここまですべてをまとめると、現代の「メンヘラ」とは「他者との関わり、特に異性から向けられる愛情を強く求めており、恋愛面での人間関係が原因で病みやすい人」という解釈ができる。

2.3 メンヘラに関連する研究

メンヘラについての研究としては松崎良美の研究[34]、[35]、[36]が挙げられる。メンヘラに関連する用語を“メ

ンタルヘルス・スラング(MHS)”と定義し、MHSの使用について考察されている。[34]では「メンヘラ」は

「自分という存在がそばを向かれるのは何より恐れるが、自分以外の他人がどうなっても知ったことではない、一言で言えば『愛してくれないならここで死んでやる!』』という感情を持ち合わせているとされる点で特徴的

と説明されており、調査対象者の90.8%が「メンヘラ」という単語を目や耳にしたことがあるという結果が得られている。また、[35]では

困難な状況に直面した本人が、MHSによってセルフ・ラベリングすることで、一定の意味を見出していた

という結論を得ている。

吉田純也の研究[32]では2017年に作成されたメンヘラ.jp[9]というメンタルヘルスに問題を抱えた方がより良い生活をおくることを支援するためのWebサイトの文章をクラスタ分析し、自殺統計データとの類似点を指摘している。分析の結果、メンヘラの抱える悩みは大きくわけて「薬」「心身の症状」「家庭」「学校」「仕事」「人間関係」「死生」であると結論付け、さらに「人間関係」の中でも「現実世界の人間関係」や「非現実世界の人間関係」の問題など、より詳細な原因を明らかにした。

メンヘラという単語は使用していないが、メンヘラの特徴と似たものを扱っている研究も紹介する。[37]が作成した恋人分離不安尺度では「自分のことを考えていて欲しいと思う」「相手を独占したいと思う」などが項目として挙げられている。メンヘラと似た属性としてあるバンギャ(ヴィジュアル系ロックバンドを熱狂的に支持する十代半ばから二十代後半の女性達)のアイデンティティ構築についての研究では、自傷行為は

敢えて自己の苦痛と悲しみと叫びとを身体を通じて再確認しようとし、時に、自傷行為すら厭わず、そこに生きる自己を確認しようとする

行為であると考察されている[33]。さらに、自傷行為の中でもリストカットに焦点を当てた研究[14]ではリストカットを「皮膚を切りつける、引っ掻くなどして傷つけ、血液が流れて傷跡が残るようにする意図的な行為」と定義している。リストカットをする人を「カッター」と表現し、自傷行為は心理的苦痛を抱える人には比較的受け入れられた習慣であると述べている。

「メンヘラ」という属性は英語では意味がやや広い“alternative teenagers”, “stigmatized groups”と表現されている。[31]では自殺意図のない自傷行為は心理的苦痛を抱えるときに他者からの助けを求める有効なコミュニケーション手段と見なされていることが示唆されている。[18]ではうつ病を抱える人はうつ病であることを自己のアイデンティティにする傾向にあるという結果が得られている。

2.4 メンヘラを取り巻く問題

メンヘラを取り巻く問題を一言で表すなら「病んだ状態にうまく対処できないメンヘラは自分自身や周囲の人を傷つけてしまう」ことである。周囲の人を傷つけてしまうというのは、まず思い浮かぶのは2019年5月に発生した好きすぎて彼氏を刺した事件?だろう。この女性は「好きで好きでしかたなかった」「一緒にいるためには殺すしかないと思った」と語っており、支配欲が暴走した結果だと考えられている。この事件はあまりにも極端だが、直接傷つけてしまう以外の例もある。[37]では恋人を支配する行動の項目に暴力的な支配行動と束縛的な支配行動を複数項目挙げている。自分自身を傷つけるのは自殺以外にも、前述した自殺意図のない自傷行為(NSSI)がある。病み垢のツイートには自傷行為をした画像を載せたものが少なくないが、これはNSSIはコミュニケーション手段だから[31]である。自傷行為のアピールをすることでに他者からの助けを求めたり、自らのアイデンティティを確保するため[6]である。しかし、[22]が指摘するように、SNSで共有された自傷行為の様子は見た者の好奇心をあおり、自分も自傷行為を試みようとする気持ちを増大させる可能性がある。

もう一つ、病んだ状態に対処するための手段としてNSSIと並んでメジャーな選択肢がSNSで話し相手を探すということがある。メンヘラと呼ばれる人が頻繁に使用するSNSやウェブサービスは複数あり、代表的なものとしてTwitterが挙げられる。Twitterでメンヘラであることを前提にして他者と交流する目的のアカウントは「病み垢」と呼ばれる。「病み垢」を作る理由として

- 不満や愚痴を吐き出したい
- 吐き出した不満を共有したい
- 他にも似た不満やつらさを抱えている人と共感したい
- 病んでいる自分に浸りたい

などが挙げられる。これらの理由をまとめると、「病み垢」の役割には大きく分けて自己表現としての機能と社会的な機能の2つがある。[18]にもある通り、自らのアイデンティティを確保し、コミュニティに属することが目的である。ハッシュタグを含めたツイートをすることで他の病み垢と繋がるという方法がよく使われる。この際に使用されるハッシュタグとして代表的なものとして「#病み垢さんと繋がりたい」「#メンヘラさんと繋がりたい」「#病み垢」「#リスカ」などがある。これらのハッシュタグによって他のアカウントと繋がり話し相手を探すことが多いが、そこからトラブルに巻き込まれることも少なくない。例えば、「泊め男」と呼ばれる、家出少女を自宅に泊める男性がいる。[7]で取り上げられている事件を簡潔に述べると、男性が女子中学生を4日間自宅に寝泊まりさせ、容疑として未成年者誘拐の疑いになったという事件である。この際に被害者は「監禁ではない」、容疑者は「誘拐ではない」と供述

しており、

一方ともう一方が、加害意識、被害意識もなく結びついてしまった

と表現されている。また、SNSで繋がるきっかけとして「#家出」「#家出少女」などのハッシュタグが紹介されている[7]。

ニュースとして取り上げられた表立った事例はまだ少ないが、病み垢カルチャーとは馴染みのない人が見えないところで、病み垢を持つユーザーは前述した「#病み垢さんと繋がりたい」「#メンヘラさんと繋がりたい」などのハッシュタグを通して日常的に「大丈夫?」「話聞くよ?」などの連絡を受けている。その中には性的な目的のユーザーからの連絡も多いが、かまってもらうためには仕方なく応じるしかないという構図が出来上がっている。このようなやりとりを通して病み垢ユーザーは傷つき、また話を聞いてくれる人を求めて他のユーザーと繋がるというループが生まれる。「病み垢」にはつらい気持ちを吐き出して気持ちを整理する、共感できる人と繋がれるなど、病んでいる人の救いになる一方で、このような負の側面もある。

ここまで見てきたことをまとめると、本研究ではメンヘラを「人間関係、特に恋愛的理由から病みやすく、病んだ状態にうまく対処できないあまり自分自身や周囲の人を傷つけてしまう人」と定義する。そして、このような人たちが先に見たような大きな問題が発生する前に適切なフォローをできるよう、このような問題を抱えた人を見つけることが本研究のモチベーションである。

その中でもメンヘラカルチャーを代表する「#病み垢さんと繋がりたい」タグに着目し、このハッシュタグに投稿する人がメンヘラの性向をもっていると仮定した。本研究では「#病み垢さんと繋がりたい」に投稿する人のツイートを学習データとして、一般のツイートから同様の性向を探索した。

3. 先行研究

本研究ではメンヘラを「人間関係、特に恋愛的理由から病みやすく、病んだ状態にうまく対処できないあまり自分自身や周囲の人を傷つけてしまう人」と定義し、特に病み垢やハッシュタグを使って話し相手を探したり、自傷行為の様子をSNSに載せる人に注目した。しかし、このようなハッシュタグを使いそうな人を予測する先行研究はまだないため、ここでは「メンヘラ」の語源となった精神疾患を予測する先行研究を紹介する。

オンラインで精神疾患を予測する研究は多数存在する。採用される手法は[28]など、古くからSVMがよく使われてきた。他にも、不完全なテキストデータに立ち向かうためにn-gram, LDA, トピックモデリングが使われてきた[17],[15],[26]。[19]ではWord2VecやGloVe Word Clustersが精神疾患になりやすいユーザーの判別に効くことが研

究されている。

うつ病以外の事例だと PTSD を予測する研究 [15],[16],[17],[27]、不安障害や強迫性障害を予測する研究 [16],[21]、境界性パーソナリティ障害や双極性障害を予測する研究 [15],[16],[20] など存在する。

Twitter 以外の Web サービスも使用している研究事例としては、馬場隆寛らの研究 [30] では悩み相談掲示板のココオル [12] をソースとしている。ココオルはメンタルヘルスの相談を専門家に相談することや、ユーザー同士で会話ができる Web サービスである。この研究では手法には n-gram の Bag-of-Words と SVM を採用し、テキストからうつ病が陽性か陰性かを予測している。また、特徴のある単語も分析しており、日常的に発信されるテキストからうつ病の陽性を予測できると結論づけている。他にも Twitter のデータを使用して学習したモデルを使って中国のミニブログサイトである Weibo[13] 上のデータでうつ病を予測する研究 [29] が存在する。

4. 手法

4.1 前処理

Twitter からはツイート本文や画像に加え、ツイートの時刻や所在地、さらにユーザーのプロフィール文などのデータも取得することができる。本論では文章のみ、具体的にはツイート本文とユーザーのプロフィール文を入力する。そのため、収集したツイートに次の前処理を行った。まずツイートから URL と画像を除去し、文章のみにした。文章からは記号と数字を取り除き、英単語は大文字をすべて小文字に変換した。後述する手法では単語ベースのアプローチも採用するため、文章を単語単位に分割する。本研究では文書を単語に分割する形態素解析器として、京都大学情報学研究所共同研究ユニットプロジェクトを通じて開発されたオープンソースの形態素解析エンジン MeCab[2] を使用した。MeCab には初期から導入されている Mecab-ipadic というシステム辞書が使用されているが、この辞書は固有表現や新語に弱いという特徴があり、固有名詞を適切に分割できないことがある。例えばうつ病などを扱う上で頻出する「希死念慮」という単語は Mecab-ipadic を使用すると「希死」と「念慮」に分割されてしまう。このままではメンヘラ特有の単語を適切に分割できないため、辞書には新語・固有表現に強い mecab-ipadic-NEologd[3] を採用した。この辞書は定期的に新しい固有表現を追加しており、これにより初期から導入されている辞書では適切に分割できなかった単語も扱えるようになる。単語に分割された文章から、名詞、動詞、形容詞のみを取り出し、動詞と形容詞は終止形に変換した。

ストップワードと呼ばれる、分類に寄与しない一般的な単語は、出現頻度が高い割に役に立たず予測精度に悪影響

を与えるため、取り除く必要がある。英語では "a", "the", "but", "what" など、日本語では「は」「が」「を」などのひらがな一文字から成る助詞や指示語、句読点などが代表例である。前処理の最後にストップワードを除去した。

4.2 特徴量の抽出

特徴量の抽出の手法を 3 つ比較した。

4.2.1 uni-gram

uni-gram とは 1 文字単位の N-gram であり、文書を文字単位に分割してベクトルで表現することで特徴量を得る手法である。データセットに出現した各文字の出現回数を数え上げたものがその文書のベクトルとなる。このため、文字の出現順序は考慮されない。

4.2.2 Bag of Words

Bag of Words(BoW) は uni-gram を単語レベルにしたものである。文書を単語に分割し、文書中に出現する単語を数えたものがその文書のベクトルとなる。BoW では uni-gram 同様、単語が出現する順番は考慮されない。

4.2.3 分散表現

分散表現、または埋め込み (word embedding) とは主に単語や複合語などの自然言語の構成要素にベクトルを割り当て、空間内の 1 つの点を 1 つの単語や複合語と見なす手法である。近年、単語分散表現である Word2Vec[23] [24] [25] などが注目を集めている。単語分散表現は通常数十から数百次元のベクトルになり、前の 2 つの手法に比べて低次元で密なベクトルになるので表現力が高いということが強みである。分散表現は単語の使用方法に基づいて学習される。これにより、同様の方法で使用される単語同士は近いベクトルになり、単語同士の意味が似ているということをつかえることができる。

4.3 分類の手法

分類の手法を 3 つ比較した。

4.3.1 Support Vector Machine

Support Vector Machine(SVM) は教師あり学習のアルゴリズムで、分類と回帰の両方の課題に使用できる機械学習分類器である。SVM の目的は n 次元空間 (n は特徴量の次元数) に 2 クラスのデータをプロットした際に、各点との距離が最大となるマージン最大化超平面を求めることである。シングルカーネル SVM は、ソーシャルメディアを含むさまざまなドメインのデータ分析に広く使用されており、線形 SVM はテキスト分類の最もパフォーマンスの高い方法の 1 つであることが古くから知られている。

4.3.2 Naive Bayes

Naive Bayes の元となる確率モデルは独立性仮定と共にベイズの定理 (式 1) を適用することに基づいており、教師あり学習の設定で用いられ、多くの実用例では単純ベイズ分類器のパラメータ推定には最尤法が使われる。

$$P(\text{label} | \text{features}) = P(\text{label}) \times \frac{P(\text{features} | \text{label})}{P(\text{features})} \quad (1)$$

Naive Bayes は Bag of Words と相性が良く、単語の共起情報を用いて予測することに向いており、SVM と並んでテキスト分類で広く使われる手法である。

4.4 Random Forest

Random Forest は決定木を弱学習器とするアンサンブル学習アルゴリズムであり、分類、回帰、クラスタリングに用いられる。名前の由来はランダムサンプリングされた教師データによって学習した多数の決定木を使用することによる。学習過程は、まずブートストラップ法によるランダムサンプリングによって訓練データから B 組のサブサンプルを生成し、次に各サブサンプルを訓練データとして B 本の決定木を作成する。その後、これらの決定木の出力を、分類ならば多数決を、回帰ならば平均値を Random Forest の出力とする。

4.5 評価指標

精度の評価には accuracy, precision, recall, F-measure を使用し、以下のように定義される。Accuracy は正と予測したもののうち、実際に正であったもの、または負と予測したもののうち、実際に負であるものの割合である。Precision は正と予測したデータのうち、実際に正であるものの割合、Recall は実際に正であるもののうち、正であると予測されたものの割合である。F-measure は Precision と Recall の調和平均である。

5. データ分析

5.1 使用したデータ

Twitter から収集したデータを使用する。ツイートからツイート主が「#病み垢さんと繋がりたい」を使ったことがある人と同じような問題を抱えている人を見つけることが本研究の目的である。

そこで、今回は 2020 年 6 月 18 日から 2020 年 10 月 22 日までの間にハッシュタグ「#病み垢さんと繋がりたい」を使用したことのあるアカウントをメンヘラとみなし、メンヘラ側の学習データとして 84804 件のツイートを収集した。対する非メンヘラのデータとしては日本語のツイートをランダムに 84804 件収集した。期間は同様に 2020 年 6 月 18 日から 2020 年 10 月 22 日の間のツイートである。どちらもプライブは含まず、ハッシュタグ入りのツイートも除外してある。入力する文書はツイートのみのもので、ツイートとそのユーザのプロフィール文を足し合わせたものの 2 種類を使用して比較する。ユーザのプロフィールを含めることで「どんな人が」「どんな文章を書くか」を考慮することができる。

表 1 両データの頻出語の比較

Table 1 Comparison of Frequent Words.

メンヘラの頻出語		非メンヘラの頻出語	
単語	出現回数	単語	出現回数
垢	20578	好き	15334
病み垢	19082	垢	10532
dm	13780	フォロー	9767
くれる	10447	推す	7503
フォロー	9369	成人	6353
好き	8178	bot	6148
依存	7533	見る	6072
フォロバ	7106	ゲーム	5574
病む	6869	言う	5196
通話	6442	大好き	4850
絡む	5835	済	4655
いい	5823	くん	4524
愚痴	5480	くれる	4479
エロ	5153	いい	4476
言う	5080	呟く	4164
死ぬ	4661	描く	4043
寂しい	4599	無言	3985
成人	4289	絵	3982
生きる	4212	最近	3926
line	4107	今日	3719

5.2 頻出語の比較

まず、テキストデータの違いについて分析した。両データの頻出語の上位 20 単語を比較すると以下の表 1 のような結果になった。

非メンヘラの頻出語からはユーザが自分の好きなものや近況を共有する意図の単語が頻出している。対するメンヘラの文章には「病み垢」「依存」「病む」「死ぬ」「生きる」など、メンヘラ特有の単語が頻出している。また、「dm」「通話」「絡む」「寂しい」「line」などが頻出することから積極的に他者と繋がってみたい、話してみたいという欲求が現れていると考えられる。

5.3 願望の表現を含む単語

次に、メンヘラ側のツイートから、願望の表現を含む単語を抽出し、頻出した単語の上位 15 個をまとめたものが表 2 である。

最も多い「死にたい」という単語が圧倒的に多く、2 番目に多いものと大きく差をつけている。「愛されたい」「必要とされたい」「依存したい」などの表現は [34] が述べている「自分という存在がそっぽを向かれるのは何より恐れる」という特徴に当てはまっている。

また、「死にたい」以外にも「消えたい」や、その裏返しである「生きたい」は [34] が述べている「『愛してくれないならここで死んでやる!』という感情を持ち合わせている」という特徴の現れであると考えられる。

「死にたい」の次に頻出している「繋がりたい」や、それ

表 2 願望の表現を含む単語
Table 2 An Example of Table.

単語	出現回数
死にたい	4306
繋がりたい	855
愛されたい	802
話したい	746
消えたい	609
されたい	599
生きたい	599
行きたい	549
必要とされたい	478
絡みたい	410
認めてほしい	362
食べたい	343
依存したい	335
仲良くしたい	334
やりたい	298
いきたい	293

と似ている「話したい」「絡みたい」「仲良くしたい」などの単語は病み垢の役割の1つである社会的な機能と結びついていると解釈でき、ここでも積極的に他者と繋がりたい、話してきたいという欲求が現れていると考えられる。先述したように、ハッシュタグは前処理で取り除いているため、「繋がりたい」をはじめとした願望の表現は、「#病み垢さんと繋がりたい」などのハッシュタグに含まれるものではなく、地の文で書かれたものである。ここからも社会的な機能の強さがうかがえる。

他にも上位15個には含まれなかったが「リスカしたい」「切りたい」などのNSSIを表す単語もあり、第2章で述べたメンヘラの特徴がよく現れていると言える。

6. 評価

6.1 学習と予測の結果

本研究では「#病み垢さんと繋がりたい」タグを過去に使用したことがあるアカウントのツイートをモデルに学習させ、入力したツイートが「#病み垢さんと繋がりたい」タグを過去に使用したことがあるアカウントかどうかを予測することでモデルの精度を評価した。手法としては uni-gram と SVM の組み合わせ、Bag of Words と Naive Bayes の組み合わせ、単語分散表現と Random Forest の組み合わせを使用した。

表3がツイートのみを入力した際の予測結果の精度である。

次に、ツイートに加え、ツイートしたユーザーのプロフィール文を含めた文章を入力した際の予測結果の精度をまとめたものが表4である。

表3, 表4の予測結果にある通り、Accuracyで最大87.0%の精度を達成した。この結果からメンヘラ特有の表現をう

まく抽出できていることを示した。ツイートとユーザーのプロフィールを入力し、単語分散表現を特徴量とし、Random Forest で予測することが最も精度が高いことがわかる。

6.2 議論

本研究で学習に使用したデータは、「#病み垢さんと繋がりたい」タグを過去に使用したことがある人がメンヘラの性向をもっていると仮定し、ツイートを収集したが、ここで論点になるのが教師データのアノテーションについてである。

「#病み垢さんと繋がりたい」タグを過去に使用したことがあるアカウントのツイートをメンヘラとみなして教師データに使ってよいかどうかである。もちろん病み垢も常に病んだツイートをしているわけではなく、日常的なツイートをすることもあろう。そのため、特に病んだ内容ではないツイートもメンヘラ側のデータとして扱われてしまった。これの対策としてツイートの文章に合わせてプロフィール文も入力することでノイズに対処した。

本研究ではメンヘラを「人間関係、特に恋愛的理由から病みやすく、病んだ状態にうまく対処できないあまり自分自身や周囲の人を傷つけてしまう人」と定義し、自分自身や他者を大きく傷つけて深刻な問題に発展する前に適切なフォローをできるようにすることが必要であると述べてきた。その中でも特にSNSで「病み垢」というアカウントを持ち、「#病み垢さんと繋がりたい」というハッシュタグを使って話し相手を探す人に注目してきた。実験の結果、ハッシュタグを使う人の書く文章の特徴を見つけることができた。もちろん全ての「病み垢」ユーザーがトラブルに巻き込まれるわけではないため、「病み垢」を使う人のうち、特にどのような文章を書く人がSNSを通じてトラブルに巻き込まれやすいかという特徴はまだ見つかっていない。これを実現するには究極的には実際にトラブルに巻き込まれた又は巻き込まれうる状況だったアカウントのツイートを教師データにすることが理想である。しかし、実際にこのデータを集めるのは困難であるため、本研究では「#病み垢さんと繋がりたい」タグを使ったことがある人のツイートと似た文章を書く人を見つけることを目的とした。本研究だけでは病んだ状態にうまく対処できずに自分自身や周囲の人を傷つけてしまう前にフォローが必要な人すべてを見つけることはまだできていない。こちらは今後の展望としたい。

7. おわりに

本稿ではメンヘラに該当する人を見つけることを目指し、メンヘラカルチャーを代表する「病み垢さんと繋がりたい」というハッシュタグに注目し、このハッシュタグに投稿する人が、メンヘラの性向をもっていると仮定し、そのハッシュタグをもつツイートを学習データとして、一般の

表 3 予測結果 ツイートのみの場合
Table 3 Prediction Results of Only Tweets

手法	Class	Accuracy	Precision	Recall	F-measure
unigram	Overall	0.70861779	—	—	—
	positive	—	0.68355879	0.75819483	0.71894495
	negative	—	0.73865857	0.66069104	0.69750275
BoW	Overall	0.71816789	—	—	—
	positive	—	0.70610758	0.73080179	0.71824249
	negative	—	0.73065671	0.70595455	0.71809326
単語分散表現	Overall	0.71045102	—	—	—
	positive	—	0.72084828	0.68744104	0.70374842
	negative	—	0.70096463	0.73348681	0.70374842

表 4 予測結果 プロフィール文を含めた場合
Table 4 Prediction Results of Tweets With User Bio.

手法	Class	Accuracy	Precision	Recall	F-measure
unigram	Overall	0.80655547	—	—	—
	positive	—	0.83252941	0.76584151	0.79779427
	negative	—	0.78457352	0.84699242	0.81458899
BoW	Overall	0.85865012	—	—	—
	positive	—	0.87843395	0.83138720	0.85426332
	negative	—	0.84099251	0.88572754	0.86278054
単語分散表現	Overall	0.87019062	—	—	—
	positive	—	0.90282486	0.82801360	0.86380246
	negative	—	0.84284091	0.91188886	0.87600638

ツイートから同様の性向を探索した。頻出する単語の分析からメンヘラの特徴が現れていることがわかる。予測精度は単語分散表現の和を文書の特徴量として Random Forest で学習することで accuracy が 87.0% であった。この結果から一般のツイートからメンヘラの性向を見つかることができると言える。

メンヘラであることは問題ではなく、病んだ状態にうまく対処できずに自分自身や周囲の人を傷つけてしまったりトラブルに巻き込まれてしまうことが問題である。さらなる研究が進み、このような人たちが抱える課題を未然に防げることを願う。

謝辞 A4 横型に対するガイドを基に、本稿を作成した。クラスファイルの作成においては、京都大学の中島 浩氏にさまざまなご教示を頂き、さらに BiBTeX 関連ファイルの利用についても快諾頂いたことを深謝する。また、A4 横型に対するガイドを作成された当時の編集委員会の担当者に深謝する。

参考文献

[1] : ASCII.jp : 女児誘拐事件からもあらためて知っておきたい、子どもによる Twitter のリスクと対策, <https://ascii.jp/elem/000/001/986/1986733/>. (Accessed on 12/15/2020).
[2] : MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <https://taku910.github.io/mecab/>. (Accessed on 01/20/2021).

[3] : neologd/mecab-ipadic-neologd: Neologism dictionary based on the language resources on the Web for mecab-ipadic, <https://github.com/neologd/mecab-ipadic-neologd>. (Accessed on 01/20/2021).
[4] : We need to stop making mental illness look cool on social media - i-D, https://i-d.vice.com/en_uk/article/a35de4/we-need-to-stop-making-mental-illness-look-cool-on-social-media. (Accessed on 01/04/2021).
[5] : 《東京・新宿》好きすぎて彼氏を刺した女の、猟奇的な“血まみれ素顔”(2019年6月3日) - エキサイトニュース, https://www.excite.co.jp/news/article/Jprime_15273/. (Accessed on 12/15/2020).
[6] : 「メンヘラ」という言葉はどう変化してきたか。精神科医が解説(熊代 亨) — 現代ビジネス — 講談社(1/3), <https://gendai.ismedia.jp/articles/-/54238>. (Accessed on 12/16/2020).
[7] : 「#家出少女」に群がる”泊め男”の恐ろしい実態 — 週刊女性 PRIME — 東洋経済オンライン — 経済ニュースの新基準, <https://toyokeizai.net/articles/-/319675>. (Accessed on 01/20/2021).
[8] : メンタルヘルス - 5ちゃんねる 掲示板, <https://mevius.5ch.net/utu/>. (Accessed on 12/09/2020).
[9] : メンヘラ.jp, <https://menhera.jp/>. (Accessed on 01/20/2021).
[10] : メンヘラとは(メンヘラとは) [単語記事] - ニコニコ大百科, <https://dic.nicovideo.jp/a/%E3%83%A1%E3%83%B3%E3%83%98%E3%83%A9>. (Accessed on 12/15/2020).
[11] : 江崎びす子たん 公式ブログ - 『病みかわいい』とは何か - Powered by LINE, <https://lineblog.me/ezakibisuko/archives/12337428.html>. (Accessed on 12/16/2020).
[12] : 悩み相談掲示板のココオル, <https://cocooru.com/>.

- (Accessed on 01/21/2021).
- [13] : 微博-随时随地新事, <https://weibo.com>. (Accessed on 01/21/2021).
- [14] Anna Gradin Franzén, L. G.: The beauty of blood? Self-injury and ambivalence in an Internet community, *Journal of Youth Studies*, No. 14, pp. 279–294 (2011).
- [15] Coppersmith, G., Dredze, M. and Harman, C.: Quantifying Mental Health Signals in Twitter, *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Baltimore, Maryland, USA, Association for Computational Linguistics, pp. 51–60 (online), DOI: 10.3115/v1/W14-3207 (2014).
- [16] Coppersmith, G., Dredze, M., Harman, C. and Hollingshead, K.: From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses, *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, Association for Computational Linguistics, pp. 1–10 (online), DOI: 10.3115/v1/W15-1201 (2015).
- [17] Coppersmith G, Harman C, D. M.: Measuring Post Traumatic Stress Disorder in Twitter, *ICWSM* (2014).
- [18] Cruwys T, G. S.: “Depression is who I am” : Mental illness identity, stigma and wellbeing, *BMC Psychiatry*, Vol. 49, No. 14, p. 137 (2014).
- [19] Daniel Preotiuc-Pietro, M. S. H. A. S. and Ungar, L.: Mental Illness Detection at the World Well-Being Project for the CLPsych 2015 Shared Task, *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Association for Computational Linguistics, pp. 40–45 (online), DOI: 10.3115/v1/W15-1205 (2015).
- [20] Elvis Saravia, Chun-Hao Chang, R. J. D. L. Y.-S. C.: MIDAS: Mental illness detection and analysis via social media., *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1418–1421 (online), DOI: 10.1109/ASONAM.2016.7752434 (2016).
- [21] Hao, B., Li, L., Li, A. and Zhu, T.: Predicting Mental Health Status on Social Media, *Cross-Cultural Design. Cultural Differences in Everyday Life* (Rau, P. L. P., ed.), Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 101–110 (2013).
- [22] 松本俊彦: 自傷行為の理解と援助: 「故意に自分の健康を害する」若者たち, 日本評論者 (2009).
- [23] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings* (Bengio, Y. and LeCun, Y., eds.) (2013).
- [24] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems* (Burgess, C. J. C., Bottou, L., Welling, M., Ghahramani, Z. and Weinberger, K. Q., eds.), Vol. 26, Curran Associates, Inc., pp. 3111–3119 (online), available from (<http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>) (2013).
- [25] Mikolov, T., Yih, W.-t. and Zweig, G.: Linguistic Regularities in Continuous Space Word Representations, *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, Association for Computational Linguistics, pp. 746–751 (online), available from (<https://www.aclweb.org/anthology/N13-1090>) (2013).
- [26] Mitchell, M., Hollingshead, K. and Coppersmith, G.: Quantifying the Language of Schizophrenia in Social Media, *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, Association for Computational Linguistics, pp. 11–20 (2015).
- [27] Pedersen, T.: Screening Twitter Users for Depression and PTSD with Lexical Decision Lists, *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, Association for Computational Linguistics, pp. 46–53 (2015).
- [28] Schwartz, H. A., Eichstaedt, J., Kern, M. L., Park, G., Sap, M., Stillwell, D., Kosinski, M. and Ungar, L.: Towards Assessing Changes in Degree of Depression through Facebook, *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Association for Computational Linguistics, pp. 118–125 (2014).
- [29] Shen, T., Jia, J., Shen, G., Feng, F., He, X., Luan, H., Tang, J., Tiropanis, T., Chua, T.-S. and Hall, W.: Cross-Domain Depression Detection via Harvesting Social Media, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, International Joint Conferences on Artificial Intelligence Organization, pp. 1611–1617 (online), DOI: 10.24963/ijcai.2018/223 (2018).
- [30] Takahiro Baba, Kensuke Baba, D. I.: Detecting Mental Health Illness Using Short Comments, *AINA 2019: Advanced Information Networking and Applications*, pp. 265–271 (2019).
- [31] Young, R., S. N. G. R. e. a.: Why alternative teenagers self-harm: exploring the link between non-suicidal self-injury, attempted suicide and adolescent identity., *BMC Psychiatry*, Vol. 49, No. 14, p. 137 (2014).
- [32] 吉田純也: メンヘラサポートサイトの文章のクラスター解析による“メンヘラ”の分類と考察, 名古屋大学 (2020).
- [33] 吉田光宏: ヴィジュアル系ロックバンドに心酔する女子達のアイデンティティ構築, 神田外語大学日本研究所紀要, No. 9, pp. 73–116 (2017).
- [34] 松崎良美: “メンタルヘルス・スラング”を定義する – 都内女子大生を対象とした横断研究より –, 津田塾大学紀要, Vol. 49, pp. 197–216 (2017).
- [35] 松崎良美: 女子大学生の“メンタルヘルス・スラング”使用と首尾一貫感覚 (SOC : Sense of Coherence), 博士論文, 津田塾大学大学院 (2018).
- [36] 松崎良美: 女子大学生における “メンタルヘルス・スラング” の使用 – 健康生成論の発想からの考察 –, 総合人間学, Vol. 13, pp. 131–144 (2019).
- [37] 園田直子片岡祥: 恋人への分離不安と愛情及び交際期間が恋人支配行動に及ぼす影響, パーソナリティ研究, Vol. 23, No. 1, pp. 13–28 (オンライン), DOI: 10.2132/personality.23.13 (2014).