

うつ傾向推定に向けたツイート内容の解析法についての一検討

玉井 森彦[†] 米澤 (深谷) 拓吾[†] 川西 直[†] 長谷川 晃朗[†]
 武内 良男[†] 田近 亜蘭[‡] 小川 雄右[‡] 古川 壽亮[‡]
[†]株式会社国際電気通信基礎技術研究所 [‡]京都大学大学院医学研究科
 {morihit, tfukaya, river24, ahase, takeuchi}@atr.jp

1 はじめに

WHOが発表した統計 [1]によると, うつ病はDALY (disability-adjusted life year: 病気によって死亡することによる損失と病気を抱えながら生きる苦しみの損失とを合算したもの) という指標において, 先進国の国民にとって大きな負担であるとされ, その対策は重要な課題となっている.

我々は, うつ病の治療に有効な療法の一つである認知行動療法に着目し, その実践を支援するためのICT基盤の構築を進めている. 具体的には, 認知行動療法において重要とされるユーザ自身による行動の振り返りに関して, スマートフォンで取得可能なライフログ情報 (ある場所への滞留時間など) を用いて行動の記録を半自動的に行う方式や, 容易かつ直感的に記録を入力するためのユーザインタフェースについての研究開発を行なっている [2].

また, ライフログ情報の活用の一貫として, ユーザがTwitterへ投稿した内容から, ユーザのうつ度合いを推定する方式についても研究を進めている [3]. Twitterへの投稿内容は, ユーザ自身の, ある瞬間における気持ちや考えを含んでおり, その内容からうつ度合いを推定できる可能性がある. これにより, 認知行動療法における臨床家 (医師, 看護師, 心理士など) によるユーザのうつ度合いの客観的判断や, うつ度合いの高まりの検出に基づく早期介入によるうつ病の再発防止などに役立てることができると考えている.

ユーザのうつ度合いの推定に向け, 本稿では, うつ病のユーザのツイート内容に表れる特徴的な語を抽出し, 投稿内容でのそれらの語の出現状況から, そのユーザがうつ群であるか, または非うつ群であるか分類を行う方式について検討した結果を述べる. 本稿の内容の初期検討の結果については, 文献 [3] に述べられているが, 本稿では, データセットの規模を拡大した上で, 素性に用いる語の選択が分類性能へ与える影響などについて, より詳細な分析を行った結果を述べる.

2 関連研究

うつ傾向の早期発見を目的とし, 松本ら [4], Wangら [5] はブログへの投稿内容を対象として, うつ傾向を検出する手法を提案している. また, 菊池ら [6], Choudhuryら [7] はTwitterへの投稿内容からうつ傾向を検出する手法を提案している.

これらの研究では, うつ傾向のユーザの投稿内容をどのように収集するかが大きな課題となる. 例えば菊池ら [6] は, 数十名から数百名の実験協力者を募り, CES-D (Center for Epidemiologic Studies Depression Scale) などのうつ傾向判別のための主観テストを実施し, 実験協力者を「うつ群」と「非うつ群」に分けている. これらの主観テストは, 疫学的な観点から判別性能が検証されたものであり, うつ傾向のユーザのデータを高い信頼性で収集することが可能な一方で, データの収集にはコストがかかるという問題がある.

本研究では, 低コストで収集した大量のデータを対象にうつ傾向の推定を行う方法として, ユーザ自身による「抗うつ剤を飲んだ」というような服薬の投稿に着目し, そのような投稿のあるユーザを「うつ群」としてデータ収集を行うことで, 各ユーザをうつ群と非うつ群に分類可能かについて検討する. これにより, Twitterから収集可能な大規模なデータセットを対象に分類モデルの構築が可能となるため, うつ傾向の推定に有効な語の抽出などによる分類性能の向上を低コストで実現できる可能性がある.

3 データセット

本研究で使用したデータセットは, 2015年2月~11月において, Twitter APIを用いてダウンロードしたツイートの集合からなる. うつ群のユーザのツイート, 及び非うつ群のユーザのツイートは, 以降で述べる方法によりそれぞれ収集した. なお, 以下で収集したツイートにおいて, 他ユーザの発言を再投稿

する retweet, および特定の他ユーザへの発言である mention は対象外とした。

3.1 うつ群のデータセットの収集

Twitter の検索 API を用いて, うつ病の患者が服用する可能性のある薬名について, その薬名を含むツイートを投稿したユーザを発見し, 各ユーザのツイート履歴を収集した。検索に使用した薬名を表 1 に示す。なお, 薬名によるツイートの検索の際には, ツイートの言語設定が日本語 (ja) のもののみを対象とした。最終的に, 合計 2000 ユーザのツイート履歴 (履歴に含まれるツイート数の平均値は 7538) を収集した。

表 1: うつ群のデータセットの収集に使用した薬名

アミトリプチリン, イミプラミン, スルモンチール, アナフラニール, アモキシサン, アンプリット, プロチアデン, ノリトレン, ルジオミール, クロンモリン, ノイオミール, マプロミール, テトラミド, テシパール, セチプチリンマレイン, レスリン, デジレル, アンデブレ, デプロメル, ルボックス, フルボキサミンマレイン, パキシル, パロキセチン, ジェイゾロフト, レキサプロ, ミルナシプラン, トレドミン, サインバルタ, リフレックス, レメロン, スルピリド, ドグマチール, ミラドル, アビリット, マーゲノール, ベタナミン
--

3.2 非うつ群のデータセットの収集

Twitter の Streaming API を用いて, Twitter に投稿されたツイートからランダムサンプリングしたツイートから, それを投稿したユーザ名を抽出し, 各ユーザのツイート履歴を収集した。なお, ツイートのサンプリングの際には, ツイートの言語設定が日本語 (ja) のもののみを対象とした。最終的に, 合計 2000 ユーザのツイート履歴 (履歴に含まれるツイート数の平均値は 7451) を収集した。

4 うつ群と非うつ群への分類

本節では, 各ユーザのツイート履歴に基づき, 各ユーザをうつ群, または非うつ群に分類する方法について述べる。

まず, 各ユーザのツイート履歴を形態素解析し, ツイート履歴に含まれるそれぞれの語の出現回数をカウントする。次に, ツイート集合に出現した語の集合からユーザをうつ群と非うつ群に分類するのに貢献する特徴的な語の抽出を行う。そのため, まず, 語の重要度を表す指標として広く用いられている tf-idf[8] に基づき各ユーザのツイート履歴に出現する各語の重み付

けを行う。以下では, ユーザ i のツイート履歴における語 j の tf-idf 値による重みを w_{ij} と表記する。

そして, 各語 j について, うつ群のユーザでの重みの平均値と, 非うつ群のユーザでの重みの平均値を求め, さらに 2 つの平均値の差 d_j を求める。すなわち, うつ群のユーザの集合を U , 非うつ群のユーザの集合を V としたとき, d_j を以下により求める:

$$d_j = \frac{\sum_{i \in U} w_{ij}}{|U|} - \frac{\sum_{i \in V} w_{ij}}{|V|}. \quad (1)$$

d_j の値は, うつ群のユーザのツイート履歴において特徴的な語である場合, 大きな正の値となり, 一方非うつ群のユーザのツイート履歴において特徴的な語である場合は, 大きな負の値となる。また両群のツイート履歴において共通して特徴的な語である場合や, 共通して特徴のない語である場合は, 0 に近い値となる。

求めた d_j の値に基づき語の集合をソートし, 上位 n 語, 及び下位 n 語を選択する。選択された語の集合を W とすると, ユーザ i について, W に属す各語 j の重みの列を素性ベクトルとする。

最後に, 機械学習に基づき, 各ユーザの素性ベクトルをうつ群または非うつ群へ分類する。

5 実験

前節において求めた素性ベクトルに基づき, ユーザをうつ群と非うつ群に分類する際の分類性能を求める実験を行った。

本実験では, 形態素解析を行うため, mecab¹を用いた。なお, 使用した mecab のバージョンは 0.996 であり, 辞書として IPA 辞書²を用いた。また, 分類器としてサポートベクターマシン (SVM) を用い, その実装として, R 言語における実装の一つである kernlab³ (バージョンは 0.9-22) の ksvm 関数 (パラメタ値はデフォルト設定) を用いた。

うつ群に属すユーザの集合 U と非うつ群に属すユーザの集合 V に対し, 分類性能を次のように求めた。まず, U と V をそれぞれ訓練集合とテスト集合に 2 分割する。そして, 訓練集合に属す U と V の各ユーザの素性ベクトルを求め, SVM による学習を行う。その後, 学習後のモデルを用いてテスト集合に属す U と V の各ユーザの素性ベクトルをうつ群, または非うつ群に分類し, 実際のクラスとの比較を行った。分類性能の指標として, 正解率 (accuracy), 精度 (precision), 再現率 (recall) を用いた。また, 全ての実験におい

¹<http://taku910.github.io/mecab/>

²mecab-ipadic-2.7.0-20070801

³<https://cran.r-project.org/web/packages/kernlab/index.html>

表 2: ユーザ数と分類性能

ユーザ数	正解率 (%)	精度 (%)	再現率 (%)
200	67.0	63.9	78.0
800	71.4	70.3	74.0
1400	72.3	74.0	70.0
2000	73.4	74.9	70.4

表 3: 品詞の種類と分類性能

品詞	正解率 (%)	精度 (%)	再現率 (%)
動詞	73.6	73.8	73.1
形容詞	70.1	70.3	69.5
副詞	70.9	70.0	73.0
3 品詞	73.4	74.9	70.4

て、 U に属すユーザ数と V に属すユーザ数は同数とした。また、特に述べない場合、式 (1) に基づき選択する語数は 1000 (上位 500 個と下位 500 個) とする。

5.1 データセットの規模と分類性能

データセットの規模が大きくなるにつれ、分類性能が変化するかどうかを調査した。データセットに含まれるうつ群、及び非うつ群のユーザ数を 200~2000 まで増加させながら、分類性能の変化を求めた。結果を表 2 に示す。なお、表 2 の「ユーザ数」は、例えば 200 の場合、 $|U| = |V| = 200$ であることを意味する。

表 2 より、ユーザ数が 200 の場合、他の場合に比べ若干分類性能が悪いが、800 以上のユーザ数に対しては一定の分類性能を維持していることが分かる。

5.2 素性に用いる語の品詞と分類性能

次に、素性として用いる語の品詞の違いが分類性能に与える影響について調査した。品詞として、動詞、形容詞、副詞、のそれぞれについて、各品詞のみを素性に用いる場合と、動詞、形容詞、副詞の 3 つを用いる場合の計 4 つの場合について、ユーザ数が 2000 の場合 ($|U| = |V| = 2000$) の分類性能を求めた。結果を表 3 に示す。

表 3 より、品詞の違いが分類性能に与える影響は大きくないことが分かる。ただし、動詞については他の品詞に比べ多少分類性能の向上が見られた。また、3 つの品詞を用いた場合でも、動詞のみを用いた場合とほぼ同等の分類性能であることが分かる。

各品詞についてどのような語が特徴的な語として選択されたかについて、表 4 に示す。これは、式 (1) の値に基づき語をソートした場合の、上位 15 個及び下位 15 個の語を示したものである。表 4 より、うつ群には「死ぬ」、「悪い」、「怖い」などネガティブな印象の語がみられ、一方、非うつ群には、「頑張る」、「楽し

表 4: 各品詞におけるうつ群と非うつ群に特徴的な語

品詞	うつ群	非うつ群
動詞	する, なる, てる, いる, れる, ある, 思う, 飲む, 言う, しまう, 死ぬ, 出る, られる, 考える, 買う	ほる, 会える, 返す, 探す, 絡む, ゆう, 会う, 呼ぶ, 待つ, すぎる, 頑張る, くれる, く, くださる, ふる
形容詞	ない, 良い, 無い, いい, 悪い, 多い, 高い, 痛い, 辛い, 面白い, っぽい, 怖い, おかしい, 酷い, 強い	ねむい, あやい, おそい, たのしい, 優しい, うれしい, はやい, かつこよい, おもしろい, 仲良い, かわいい, 楽しい, やばい, 嬉しい, 可愛い
副詞	どう, もう, そう, よく, まだ, ちよつと, なぜ, 少し, 全く, やはり, まあ, やっぱり, 本当に, 実際, とても	ゆめ, さ, スグ, 是非, ぼん, るる, いやいや, どー, りん, すぐ, ほんとに, 仲良く, これから, よろしく, めっちゃ

い」、「仲良く」などポジティブな印象の語が見られる。ただし、直感的にはうつ群と非うつ群の分類に貢献しそうでない語も多く含まれていることが分かる。

5.3 素性に用いる語の数と分類性能

次に、素性に用いる語数の違いに対する分類性能の変化を調査した。形態素解析に基づき抽出される語として、動詞、形容詞、副詞の 3 つを対象とした。提案方式に基づき、抽出された語の集合を式 (1) に基づきソートした後、上位 n 個と下位 n 個の語を選択して、その tf-idf 値の列を素性ベクトルとした。また、ユーザ数は 2000 ($|U| = |V| = 2000$) とした。選択する語数を 2~5000 まで変化させた場合の各分類性能を表 5 に示す。なお、表の「語数」は、例えば 100 の場合、上位 50 個と下位 50 個の語を選択したことを意味する。

また、提案方式との比較のため、形態素解析に基づき抽出された語から、ランダムに n 個の語を選択し、その tf-idf 値の列を素性ベクトルとした場合の分類性能を表 6 に示す。また参考のため、図 1 に形態素解析に基づき抽出された語 (計 15454 個) の d_j の値のヒストグラムを示す。なお、図 1 では見やすさのため表示する d_j の値の範囲を狭めているが、 d_j の値の最大値は 0.0815、最小値は -0.0107 であった。

表 5 より、選択する語数が少ない場合は、語数が増加するにつれ分類性能が向上するが、100 語程度あれば、それ以上語数を増やしても性能の向上はほとんど見られないことが分かる。

また表 6 より、語数が少ない場合、素性に用いる語をランダムに選択する方法では、提案方式に比べ分類性能が大きく悪化していることが分かる。これは、図 1 より分かるように、 d_j の値はほとんどが 0 付近に

表 5: 素性に用いる語数と分類性能 (提案方式)

語数	正解率 (%)	精度 (%)	再現率 (%)
2	62.4	60.7	70.6
10	67.3	67.1	67.8
100	72.7	74.1	69.7
1000	73.4	74.9	70.4
5000	73.3	73.9	72.1

表 6: 素性に用いる語数と分類性能 (素性に用いる語をランダムに選択)

語数	正解率 (%)	精度 (%)	再現率 (%)
2	51.2	60.9	6.7
10	59.4	62.2	47.8
100	59.7	59.3	61.5
1000	67.3	66.9	68.6
5000	71.7	71.6	71.7

分布しており、語をランダムに選択すると、多くが 0 付近に分布する語を選択してしまうことによる。従って、式 (1) に基づき分類に貢献する語を選択することの有効性が示されたと言える。また、語数が 5000 程度あれば、ランダムに選択した場合でも、提案方式で 100 語用いる場合と同等の性能が得られることが分かるが、語数が増えることで計算に必要なリソースが増加する。提案方式により選択する語数を絞り込むことで、SVM によるモデルの作成などに必要な計算リソースが削減でき、より大規模なデータセットに対して効率的に分類を実行することができる。

6 おわりに

本稿では、Twitter への投稿履歴から、ユーザをうつ群と非うつ群に分類する方式を提案し、その分類性能について調査した結果を述べた。実験により、ユーザ数が 2000 の場合、動詞、形容詞、副詞を対象に合計 1000 個の語に対する tf-idf 値の列を素性ベクトルとして SVM によるツイート履歴の分類を行うことで、正解率 73%、精度 75%、再現率 70% でうつ群か非うつ群かの分類が可能であることが分かった。また、素性に用いる語を、うつ群と非うつ群の tf-idf 値の平均値の差に基づき選択することで、ランダムに選択する場合に比べ、少数の語で比較的高い分類性能を得ることが分かった。

本稿で述べた方式では、ユーザごとにうつ群と非うつ群への分類を行なっているが、ユーザのうつ傾向の変化に基づくうつ病の再発防止などに役立つためには、各ユーザの時間経過に伴ううつ傾向の変化を検出することが求められる。今後、そのような時間的に粒度の細かい単位でうつ傾向を推定する方式について検討を進める予定である。

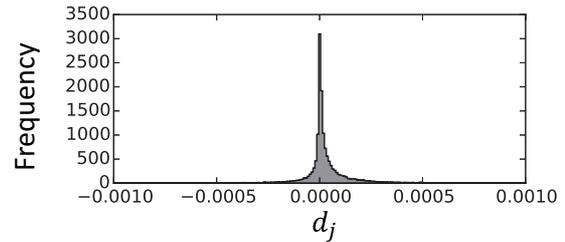


図 1: d_j の値の分布

謝辞

本研究成果は、国立研究開発法人情報通信研究機構 (NICT) の委託研究「ソーシャル・ビッグデータ活用・基盤技術の研究開発」により得られたものである。

参考文献

- [1] The global burden of disease: 2004 update, World Health Organization, 2008.
- [2] N. Kawanishi, M. Tamai, A. Hasegawa, Y. Takeuchi, A. Tajika, Y. Ogawa, and T. Furukawa: "LifeLog-based Estimation of Activity Diary for Cognitive Behavioral Therapy," *Adjunct Proc. of the 2015 ACM Int'l Joint Conf. on Pervasive and Ubiquitous Computing and the 2015 ACM Int'l Symposium on Wearable Computers*, pp. 1251–1256, 2015.
- [3] 深谷拓吾, 川西直, 長谷川晃朗, 田近亜蘭, 小川雄右, 堀越勝, 古川壽亮, 武内良男: "うつ傾向推定に向けた抗うつ剤服用の投稿を起点とした Twitter 解析の初期検討," 言語処理学会第 21 回年次大会発表論文集, pp. 445–448, 2015.
- [4] 松本和幸, 吉岡伸浩, 北研二, 任福継: "うつキーフレーズと感情変動に基づくブログからのうつ検出手法," 言語処理学会第 18 回年次大会発表論文集, pp. 1126–1129, 2012.
- [5] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao: "A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network," *Trends and Applications in Knowledge Discovery and Data Mining*, Vol. 7867, pp. 201–213, 2013.
- [6] 菊池佑介, 津川翔, 岸野文郎, 中島康祐, 伊藤雄一, 大崎博之: "Twitter におけるユーザの単語使用頻度を用いたうつ傾向の推定," 電子情報通信学会総合大会, p. 168, 2014.
- [7] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz: "Predicting Depression via Social Media," *Proc. of the 7th Int'l AAAI Conf. on Weblogs and Social Media (ICWSM'13)*, pp. 349–350, 2013.
- [8] P. Fung and L. Y. Yee: "An IR Approach for Translating New Words from Nonparallel, Comparable Texts," *Proc. of ACL-COLING'98*, pp. 414–420, 1998.