

平成 28 年 5 月 24 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26870076

研究課題名（和文）ソーシャルメディアにおける大規模行動履歴を活用したユーザのうつ傾向推定

研究課題名（英文）Recognizing depression of social media users from their records of activities

研究代表者

津川 翔（Tsugawa, Sho）

筑波大学・システム情報系・助教

研究者番号：40632732

交付決定額（研究期間全体）：（直接経費） 3,000,000 円

研究成果の概要（和文）：近年、うつ病が世界的に重大な社会問題となっている。うつ病の適切な治療のためには、まず、個人のうつ傾向を把握することが重要である。本研究では、人が日常的に利用するソーシャルメディアにおけるユーザの活動履歴から、ユーザのうつ傾向を推定するモデルを構築した。評価実験により、ソーシャルメディアにおけるユーザの活動履歴がうつ傾向を推定する上で有用であることを示した。さらに、モデルの構築に有用な特徴量や学習期間の長さとうつ傾向の推定精度の関係についても明らかにした。

研究成果の概要（英文）：In recent years, depression is one of significant problems in the world. For the effective treatment of depression, it is important to recognize depressive tendencies of individuals. This study constructed several models to estimate depressive tendency of social media users from the features obtained from user's activities. Through experiments, we showed that using the records of user's activities is effective for recognizing depression. We also showed effective features for recognizing depression of social media users, and the relation between the length of the training period for constructing the models and their estimation accuracy.

研究分野：ネットワーク科学

キーワード：ソーシャルメディア うつ 機械学習 Twitter

1. 研究開始当初の背景

近年、うつ病が世界的に重大な社会問題となっている。世界保健機関 (WHO) は、2012 年の時点において、世界で 3 億 5 千万人のうつ病患者が存在すると推計している [WHO2012]。また、日本の厚生労働省によって 3 年ごとに全国の医療施設に対して行われている患者調査の結果によると、1996 年に約 43 万人であった国内のうつ病患者数は、2008 年には、約 104 万人と、著しく増加している。うつ病は、最悪の場合、自殺の原因にもなるため、うつ病患者の増加は非常に深刻な問題である。

うつ病の適切な治療のためには、まず、個人のうつ傾向を把握することが重要である [WHO2012]。WHO の報告では、うつ病患者の半数以上は、自身がうつであることを認識せず、必要な治療を受けていないと述べられている。また一次医療においても、うつが認識されずに、その結果適切な治療を受けられないことが問題とされている [Tiemens1999]。このことから、うつ病の適切な治療のためには、まず、患者自身、あるいは周囲の人が、その患者のうつ傾向を把握することが重要であると考えられる。

日常生活における人の活動履歴などの客観的な情報から、個人のうつ傾向を推定することができれば、うつが認識されていなかった患者に適切な治療を受けさせることが可能となると期待される。しかしながら、うつ傾向の推定にはアンケート結果のような主観的な情報を用いるのが一般的であり、個人の活動履歴のような客観的な情報からうつ傾向を推定する手法は確立されていない。

代表者は、研究開始当初の時点において、うつ傾向を推定するための情報として、ソーシャルメディアに蓄積された大規模なユーザの活動履歴に着目し、その有効性を国内外の学会で報告していた [Tsugawa2013]。ソーシャルメディアは、ユーザが日頃の感情を表現するために利用することも多いため、ソーシャルメディアにおけるユーザの活動履歴が、そのユーザのうつ傾向の推定に有用な情報ではないかとの着想に至った。

2. 研究の目的

本研究では、これまでの研究成果を発展させ、人が日常的に利用するソーシャルメディアにおけるユーザの活動履歴から、ユーザのうつ傾向を推定するモデルを構築することを目的とする。これまでの研究では、数十名程度の小規模なユーザの Twitter における投稿履歴を用いて、投稿に用いられる単語の出現頻度と、ユーザのうつ傾向との相関を分析するにとどまっていた。それに対して本研究では、大規模なユーザの活動履歴とそのユーザのうつ傾向のデータを収集し、これらの

大量のデータを教師データとする教師あり機械学習のアプローチで高精度なうつ傾向推定モデルの構築を目指す。これを達成するために、次の研究課題を設定し、遂行する。

- (1) 大規模教師データの収集
- (2) 機械学習によるうつ傾向推定モデルの構築
- (3) うつ傾向推定モデルの評価

3. 研究の方法

- (1) アンケート収集 Web サイトの構築による教師データの収集

まず、うつ傾向を推定するモデルを構築するための教師データとして、代表的なソーシャルメディアである Twitter ユーザの活動履歴と、そのユーザのうつ傾向を表す指標 (以下ではうつ病尺度と呼ぶ) を収集した。うつ病尺度を得るためのアンケート収集 Web サイトを構築・運用することで、モデルの構築に必要な教師データを収集した。構築した Web サイトでは、うつ傾向を推定するための代表的な尺度である Center for Epidemiologic Studies Depression Scale (CES-D) を得るためのアンケートを実施した。さらに、アンケート結果の信頼性を確保するため、Beck's Depression Scale (BDI) と呼ばれるうつ病尺度のアンケートも実施し、BDI のスコアと CES-D のスコアの相関が著しく低い回答は教師データから除くようにした。

アンケート収集のための Web サイトでは、アンケートに回答してもらうための工夫として、回答してもらったユーザに対して、ユーザのソーシャルメディアにおける活動履歴を解析することによって得られるココロスコアと呼ばれる称号を提示した (図 1)。また、Web サイトでは、ココロスコアの結果を Twitter においてつぶやけるようにし、口コミによってアンケート収集 Web サイトの情報を広める機能も追加した。

ココロスコアはかったー

① 登録情報入力 ② スコア推定アンケート ③ スコア推定アンケート ④ スコア推定アンケート ⑤ スコア推定アンケート ⑥ ココロスコア 発表結果



図 1:構築した Web サイトのスクリーンショット

(2) 特徴量の抽出とうつ傾向推定モデルの構築

次に、収集した Twitter ユーザの活動履歴からうつ傾向の推定に有用な特徴量を設計・抽出した。ユーザの投稿で使用される単語の出現頻度、トピックモデルによって推定した投稿のトピック比率、ポジティブな単語およびネガティブな単語の投稿における出現頻度、ツイートの投稿の頻度、投稿時間、フォロワー数やフォロー数などを含む様々な特徴量を抽出した。

抽出したユーザの特徴量と、アンケートから得られたユーザのうつ傾向を推定する尺度 CES-D のスコアを教師データとする教師あり機械学習によって、うつ傾向を推定するモデルを構築した。ここでは、教師あり機械学習の代表的な手法として Support Vector Machine を用いた。

(3) うつ傾向推定モデルの評価

様々な特徴量の組み合わせでうつ傾向を推定するモデルを構築し、構築したモデルによってどの程度の精度でユーザのうつ傾向を推定できるか評価した。収集したデータの一部をモデルの構築のための訓練データとして用いて、残りのデータを評価用のデータとして用いた。評価指標として、構築した各モデルについて、適合率、再現率、F 値、精度を計測した。これらの評価指標をモデルごとに比較することで、どの特徴量がうつ傾向の推定に有用であるかについても分析した。

4. 研究成果

本研究では、(1) ソーシャルメディアにおけるユーザの活動履歴からユーザのうつ傾向を推定する上で有用な特徴量を明らかにした。(2) それらの特徴量を用いてユーザのうつ傾向を推定するモデルを構築し、その有効性を定量的に明らかにした。さらに、(3) モデルの構築に必要な学習期間の長さを明らかにした。(4) また、予測に有用な特徴量を分析する過程で、ソーシャルメディアにおいてネガティブな投稿はポジティブな投稿やニュートラルな投稿と比較して拡散力が強いということを発見した。以下では、上記の 4 つの成果について、より詳細に報告する。

(1) うつ傾向の推定に有用な特徴量の分析

教師データ収集のために構築した Web サイトを運用し、のべ 214 ユーザの Twitter における活動履歴とうつ病尺度のアンケート結果を収集した。まったくツイートを行っていないユーザおよび、うつ病尺度のアンケートへの回答時間が著しく短いユーザのデータを除き、209 名のユーザのデー

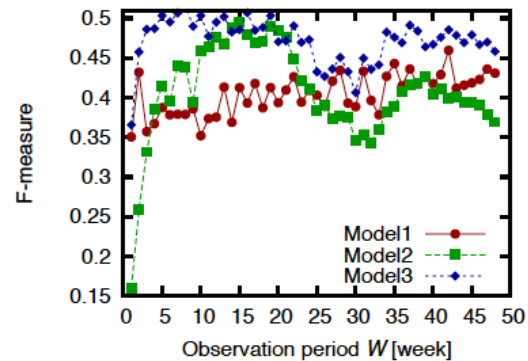


図 2: 学習期間と推定精度の関係

タを分析に用いた。209 名のユーザをうつ傾向の有無によって、分類したところ、81 名はうつ傾向あり、他の 128 名はうつ傾向なしと推定された。

収集したユーザの活動履歴から 3 章で説明した特徴量を抽出し、うつ傾向ありのユーザとなしのユーザの特徴量の違いを分析した。分析の結果、うつ傾向ありのユーザとなしのユーザの間で、ネガティブな単語の使用頻度、ツイートの投稿頻度、リツイートの頻度、などの特徴量に有意差が存在することを示した。この分析の結果から、これらの特徴量がユーザのうつ傾向の推定に有用であることが示唆された。

(2) うつ傾向推定モデルの構築とその定量的評価

収集したデータから 3 章で述べた特徴量を抽出し、Support Vector Machine による教師あり機械学習で、ユーザのうつ傾向を推定するモデルを構築した。

10 分割交差検証によって、構築したモデルのうつ傾向の推定精度を評価した。その結果、特徴量として、トピックモデルを用いて推定した投稿のトピック比率、投稿におけるポジティブな単語の使用頻度、ネガティブな単語の使用頻度、投稿の頻度、リツイートの頻度、URL を含む投稿の頻度、フォロワー数、およびフォロー数を用いたモデルによって、適合率 0.61、再現率 0.37、F 値 0.46、精度 66% で、ユーザのうつ傾向の有無を推定できることを示した。

なお、様々な特徴量の組み合わせによってうつ傾向の推定モデルを構築し、それらの推定精度を比較することで、どのような特徴量がうつ傾向を推定する上で有用であるかも分析した。その結果特に、ユーザの投稿のトピックやユーザの投稿に含まれるネガティブな単語の出現頻度が、有用な特徴量であることも示した。

(3) モデルの構築に必要な学習期間の分析

さらに、ソーシャルメディアにおけるユーザの活動履歴からうつ傾向を推定する際に、どの程度の期間の活動履歴が必要であるかを分析した。モデルを構築するのに用いる活動履歴の観測期間（学習期間）の長さを変化させ、構築したモデルの推定精度を計測した。図 2 に学習期間の長さ、推定精度を表す F 値の関係を示す。この結果から、うつ傾向の推定には、2 ヶ月程度の学習期間が必要であることが分かった。また、同様に学習期間を長くすることで、かえって推定精度が低下してしまう場合もあることが分かった。

(4) ソーシャルメディアにおける投稿の拡散現象の分析

研究成果 (1) ～ (3) で収集したデータとは別の約 400 万件の大規模なツイートデータを利用し、ソーシャルメディアにおける投稿の拡散現象の分析を行った。ツイートに含まれる単語を分析することにより、ツイートを、ポジティブ、ネガティブ、ニュートラルの 3 つのクラスに分類した。クラスによって、ツイートの拡散範囲および拡散速度がどの程度異なるかを調査した。拡散範囲を測る指標としてリツイート数を、拡散速度を測る指標として、ツイートが N 回リツイートされるまでの時間 N リツイート時間を用いた。

分析の結果、ネガティブなツイートは拡散力が強く、ポジティブなツイートおよびニュートラルなツイートと比較してリツイート数が多く、N リツイート時間が短いことなどを示した。人がネガティブな事象に対して強く反応する Negativity Bias と呼ばれるバイアスが存在することはこれまで知られている。本研究の結果は、ソーシャルメディアにおいてもこのようなバイアスが存在することを示すものであると考えている。

今後、このような投稿の拡散のダイナミクスを考慮することで、他のユーザの投稿や、ユーザ間の交流の履歴から、うつ傾向の推定に有用な特徴を抽出できるのではないかと考えている。

世界的に問題となっているうつ病の治療のためには、まず個人のうつ傾向を把握することが重要であり、本研究におけるこれらの成果はうつ病の治療に貢献するものと期待される。これらの研究成果は、本分野における最も重要な会議の一つである ACM CHI2015 にフルペーパーとして採択されている（採択率 23%）。このことは、本研究の成果が学会において高く評価されていることを示している。

また本研究の成果は、周辺分野の研究動向も含めて、電子情報通信学会の研究会において招待講演としても発表した。

<引用文献>

[WHO2012] M. Marcus et al. "Depression: A global public health concern," WHO Department of Mental Health and Substance Abuse, Tech. Rep., 2012

[Tiemens1999] B. Tiemens et al. "Training primary-care physicians to recognize, diagnose and manage depression: does it improve patient outcomes?" , Psychological Medicine 29,04,pp. 833-845. Jul. 1999,

[Tsugawa2013] S. Tsugawa, et al. "On estimating depressive tendency of Twitter users utilizing their tweet data," in Proceedings of the 2nd International Workshop on Ambient Information Technologies (AMBIT 2013), pp. 29-32, Mar. 2013.

5 . 主な発表論文等

〔雑誌論文〕(計 3 件)

S. Tsugawa, and H. Ohsaki, "Negative messages spread rapidly and widely on social media," in Proceedings of the 3rd ACM Conference on Online Social Networks (COSN 2015), pp. 151-160, Nov. 2015. doi:

10.1145/2817946.2817962 (査読有)

S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, "Recognizing depression from Twitter activity," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI 2015), pp. 3187-3196, Apr. 2015. doi: 10.1145/2702123.2702280 (査読有)

S. Tsugawa, and H. Ohsaki, "Analysis of the robustness of degree centrality against random errors in graphs," in Proceedings of the 6th Workshop on Complex Networks (CompleNet 2015), pp. 25-36, Mar. 2015. doi:

10.1007/978-3-319-16112-9_3 (査読有)

〔学会発表〕(計 6 件)

津川 翔, "通信行動データを利用したコミュニティ計測に関する研究の動向," 電子情報通信学会 コミュニケーションクオリティ研究会, 2016 年 1 月 22 日, 筑波大学 (茨城県つくば市)(招待講演)

津川 翔, 大崎 博之, "ソーシャルメディアにおける投稿の感情がその拡散力に与える影響," コミュニケーションクオリティ研究会, 2015 年 9 月 1 日, 京都工芸繊維大学 (京都府京都市)

森 久幸, 津川 翔, 大崎 博之, "ソー

シャルメディアにおける投稿の感情がその拡散力に与える影響の分析," 電子情報通信学会 通信行動工学時限研専第 8 回研究会, 2015 年 3 月 9 日, 大阪大学 (大阪府茨木市)

津川 翔, 大崎 博之, "ネットワーク分析における中心性指標のロバスト性に関する一考察," 第 57 回自動制御連合講演会, 2014 年 11 月 11 日, ホテル天坊 (群馬県渋川市)

森 久幸, 津川 翔, 大崎 博之, "ソーシャルメディアにおける投稿のセンチメントがその拡散範囲に与える影響の分析," 電子情報通信学会ソサイエティ大会, 2014 年 9 月 25 日, 徳島大学 (徳島県徳島市)

津川 翔, 大崎 博之, "グラフにおけるランダム誤りに対する次数中心性のロバスト性解析," 電子情報通信学会 コミュニケーションクオリティ研究会, 2014 年 7 月 10 日, 大阪大学 (大阪府豊中市)

6. 研究組織

(1) 研究代表者

津川 翔 (TSUGAWA, Sho)

筑波大学・システム情報系・助教

研究者番号: 40632732