

ツイートデータを用いた精神状態と生活習慣の可視化システムの提案

木村 幸[†] 宮治 裕[‡]
 青山学院大学[†] 青山学院大学[‡]
 社会情報学部 社会情報学部

1. はじめに

近年では精神疾患のために医療機関に通っている患者数は大幅に増加しており、平成 26 年は 392 万人、平成 29 年では 400 万人を超えている [1]。また、昨今のコロナウイルスの流行のための外出自粛によって、ストレスを感じている人が多くなっている [2]。一般にストレスは精神疾患の原因とされており、今後「コロナ疲れ」による精神疾患患者の増加が懸念される。精神疾患は疾患によっては症状や現れ方が多彩であるため、発症初期の症状では罹患していることに気づくことが難しい。そこで、本研究では精神疾患に罹患することを未然に防ぐことを目的としたシステムの構築を目指す。

2. 関連研究

関連する研究として治部ら [3] の研究がある。この研究では保育士を対象に職場で精神的または身体的にもっともストレスを感じていることについて具体的に自由に記述してもらい、得られたテキストを分析した結果、職場の人間関係や給与への不満が読み取れるような共起グラフや WordCloud が出力された。また、片山ら [4] の研究によると生活習慣が望ましい状態にある者は、気分状態が安定し、活動性が高いことが報告されている。

これらの結果より、テキストから精神状態や生活習慣を読み取ることができる情報を抽出・可視化し、精神状態や生活習慣を管理することで精神疾患に罹患することを予防できるのではないかと考え、本研究をおこなうことにした。本研究では、ツイートデータを用いてテキストマイニングをおこなうことで、システム利用者に定期

的な自由記述や日記を書かせる等の手間減らし、システムを長期的に使ってもらえるようにする。

3. システム構成

本システムは、可視化したいツイートデータをアップロードするアップロード部、可視化のために必要な前処理やデータの加工をおこなうテキストマイニング部、加工された情報を可視化する可視化部の 3 つで構成される。

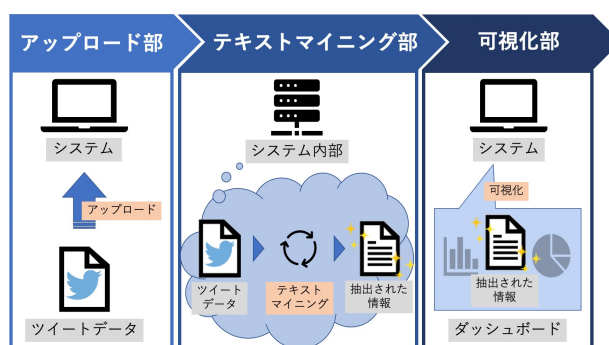


図 1: システム構成図

アップロード部では、あらかじめダウンロードしたツイートデータをシステムにアップロードする。

テキストマイニング部ではツイートデータをクレンジング（余計な情報を省く処理）し、形態素解析をおこなったあとに頻度や感情ラベルを付与することで精神状態や生活習慣を把握するために有益な情報を抽出する。まず、Amazon Comprehend の感情分析 API と単語感情極性対応表 [5] を用いてツイートと単語の感情分析をおこなった。次に、Pandas の集計関数を用いて日毎、週毎、月毎にデータの集計をおこなった。

可視化部では加工されたツイートデータを「よく使う単語セクション」、「生活習慣セクション」、「ツイート数と感情スコアの推移セクション」の 3 つのセクションに分けてダッシュボード上に表示する。「よく使う単

Proposal for mental state and life Habit visualization system with tweet data

[†]Miyuki Kimura, Aoyama Gakuin University

[‡]Yutaka Miyaji, Aoyama Gakuin University

語セクション」では棒グラフと WordCloud を用いて単語の出現頻度を可視化する。「生活習慣セクション」では、時間帯ごとのツイート数と時間帯ごとの感情スコアを可視化する。「ツイート数と感情スコアの推移セクション」では感情スコアとツイート数の時系列推移を可視化する。

4. 検証実験

4.1. 目的および手法

本実験では、本システムで可視化された情報の有効性や本システムで提示した感情分析の精度、システム全体の有効性を検証することを目的とする。実験には 20 代の男女 12 人に協力してもらい、実際にシステムを利用したあとでアンケートに回答していただいた。

4.1.1. 本システムで可視化された情報の有効性の検証

3 つのセクションごとに期待される効果が得られたかどうかを 4 件法のアンケートで検証する。

4.1.2. 本システムで提示した感情分析の精度の検証

システム内で用いている単語感情極性対応表を用いて感情ラベルを付与した単語の例と、システム内で用いている感情分析 API を用いて感情ラベルを付与したツイートの例を提示し、その結果について 4 件法のアンケートで評価をおこなった。

4.1.3. システム全体の有効性を検証

システムの使いやすさやシステムの効果について、4 件法のアンケートで評価する。

4.2. 結果および考察

4.2.1. 本システムで可視化された情報の有効性の検証

よく使う単語セクションについては、関連する質問の平均値が 2.75 と 3 未満であったことから、期待される効果が十分に得られなかったといえる。自由記述の回答から、単語のネガティブ・ポジティブの分類の精度や可視化の方法を見直すことで、改善できると推測できる。その他のセクションについては、関連する全ての質問で平均値が 3.0 以上であったことから、期待する効果が得られたといえる。

4.2.2. 本システムで提示した感情分析の精度の検証

ツイートの感情分析の精度についての質問では、平均値は 3.42 であり、ツイートの感情分析の精度は高いといえる。単語の感情分析の精度についての質問では、平均値が 2.17 であり、単語の感情分析の精度は低かったといえる。自由記述の回答の中でも、単語の感情分析の精度について改善を求めるコメントが多くあり、今後システムを改善していく上で重要な課題であるといえる。

4.2.3. その他システム全体の評価

システム全体の評価についての質問では、全ての質問で平均値が 3.0 以上になった。よって、提案するシステムは有効であるといえる。当システムで自動でのツイート収集が可能になった場合、システムを継続して使いたいかという質問では、平均値が 3.58 であった。そのことから、ツイートデータの取得を自動化することで、システムの有効性を向上させることができると推察できる。また、システムの操作のしやすさについての質問では評価が高かったユーザー群と評価が低かったユーザー群で、平均値の差が他の質問に比べて大きかった。そのことから、システムのインターフェイスを改善することで、ユーザーの評価を改善することができるといえる。

5. おわりに

本研究では精神疾患にかかることを予防することを目的とし、テキストから得られた精神状態や生活習慣を読み取ることができる可視化を行うことでメンタルを管理するシステムの構築をおこなった。システムはアップロード部、テキストマイニング部、可視化部に分けられる。

また、構築したシステムが精神状態や生活習慣を管理する効果があるかを検証した。実験の結果から、提案するシステムは有効であるといえた。今後の課題としては、単語の感情分析の精度の改善、システムのインターフェイスの改善、ツイートの取得を自動化する機能の追加の 3 つがある。これらを実現することで、よりシステムの有効性を向上させることができると推察できる。

参考文献

- [1] 厚生労働省. 知ることからはじめよう みんなのメンタルヘルス精神疾患のデータ. <https://www.mhlw.go.jp/kokoro/speciality/data.html>.
- [2] BrainPad. ブレインパッドが独自調査、新型コロナウイルスによる自粛「ストレス増加」の 1 位専業主婦 72 <https://www.brainpad.co.jp/news/2020/05/12/11385>.
- [3] 哲也治部, 秀之小山. 保育士のストレスに関する質的研究 : テキスト・マイニングを用いた職場のストレスの分析. 関西福祉科学大学 EAP 研究所紀要, No. 14, pp. 25-32, mar 2020.
- [4] 片山友子, 水野 (松本) 由子, 稲田紘. 大学生の生活習慣とメンタルヘルスの関連性. 総合健診, Vol. 41, No. 2, pp. 283-293, 2014.
- [5] 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, pp. 627-637, 2006.