

Twitter 利用者の感情・行動表現の時系列的変動

発表者: I 類 メディア情報学 プログラム 学籍番号 1710673 吉田周平

指導教員: 久野雅樹 教授

1 研究背景・目的

SNS の一つである Twitter には、リアルタイムに多くの人々が感情や行動を反映させていると考えられる。API を用いるとこれらの情報を入手することができる。これまで Twitter における感情や行動の表現の時系列変動をテーマとした研究はあったが、その多くは対象が限定的である。

そこで、日本の Twitter における使用アプリや表現の時系列変動を観察し、傾向や周期性などの普遍的な法則を発見することを目指す。大衆の心理状態や行動との関係を考察できれば、マーケティング等の分野に役立てられると考えた。ゆえに、本研究ではより基礎的で普遍的な結果を出すことを目指す。先行研究をもとに、日本の Twitter においても分析的や実存的といった思考の周期があり、それが感情や行動表現として整合して現れていると仮説を立てた。

2 先行研究

2.2 英国での研究例[1]

2010 年から 14 年までに英国の 54 都市から Twitter に投稿された約 8 億件のツイートと約 70 億の単語を分析し、英国の人々の心理状態が 1 日を通してどのように変化するかを明らかにした。LIWC[2]と呼ばれる語彙を抽象化してカテゴリ化するためのツールを使用している。心理的特徴の日周リズムを、24 時間周期の主成分分析により見出した。主成分の 2 つで全体の分散の 85% が説明できることがわかった。第一の要因は、午前 5 時から午前 6 時までをピークとする分析的思考(あの出来事が起きたのはこういう原因があったはずなどと分析する思考)、第二の要因は午前 3 時から午前 4 時をピークとする実存的思考(自分自身の存在意義はなんだ、なぜあれは存在するのだといった思考)であった。

3 コーパス構築

3.1 対象

TwitterAPI[3]を用いて、2020/12/03-2021/02/03 の間につぶやかれた、言語設定が日本語であるツイートを 29251028 個収集した。使用した属性はツイート時刻、ツイートに使用されたアプリ、テキスト本文、言語の 4 つである。

3.2 手法

ツイート本文テキストに対して形態素解析器 MeCab[4]で形態素解析を行い、単語分割を行ったのち、形容詞と動詞の原型の抽出を行った。個々の単語に対して出現数、または個々のツイートソースに対して 1 分間隔で 1 週間分の長さの時系列ベクトルにした。前後 5 分で移動平均を取り、標準化を行った。

4 実験

4.1 対象

ツイート数が 10,000 以上のツイートソース 27 種類、出現数が 10000 以上の形容詞 131 個、出現数が 50000 以上の動詞 159 個の時系列ベクトルに対して日周期または週周期でクラスタリングを行った。

4.2 手法

エルボー法で最適クラスター数を決めた。クラスターに属する個々の単語の時系列ベクトルとクラスター中心の時系列ベクトルの平均二乗誤差を求め、クラスター中心に近い単語(平均二乗誤差が小さい単語)でクラスターの解釈を行った。

4.3 結果と考察

4.3.1 ツイートソース

一時間毎の周期が強いソースと、そうでないソースとに分かれた。周期が強い単語は bot や定期投稿によるものが多く、これらはリアルタイムに利用者の感情や行動を反映するとは考えづらい。したがって、以降の分析ではここで1時間毎の周期がない方のツイートソースであるものに限って行った。また、クラスター数を増やすとゲームシェアをするためのソースは夜に増えるなど、利用者の行動を表す結果が得られた。

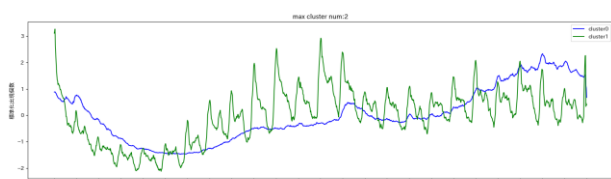


図 1: ツイートソースでのクラスタリング(ツイート数)

4.3.2 動詞

クラスター数 4 とした。「ある」や「いる」などの決まった状態や移動を表す動詞が含まれ、比較的朝と昼に多いクラスター 1、決まった時間帯に増え、それ以外の時間帯は少ない動詞「起きる」「寝る」「答える」「あける」が含まれるクラスター 2、「置く」「当たる」「届く」を含み、日付の変わり目に急に増えるクラスター 3、「わかる」「終わる」「始める」などの瞬間的な状態の変化を表す動詞が含まれ、比較的夜に増えるクラスター 4 に分かれた。週周期でも同様の分かれ方をした。動詞分類については書籍[6]を参考にした。一日の中で継続/状態的な行動表現から瞬間的な行動表現への時系列変動が見られる。夜の方がよりリアルタイムに行動したことなどをつぶやけることが原因と考えた。週変動に関しても同様のクラスターの分かれ方をした。

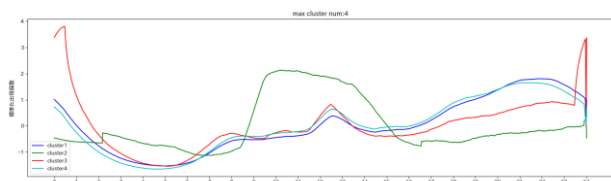


図 2: 動詞のクラスタリング(単語出現数)

4.3.3 形容詞

朝に非常に多く「ねむい」「眠い」「寒い」「さむい」「暖かい」(気候に対する分析と睡魔に関係する単語)が含まれるクラス

ター 1、朝から昼にかけて多く、「ない」「多い」「高い」など比較的客観的な単語が含まれるクラスター 2、夜に多く、「すごい」「かわいい」「良い」など比較的主観的でポジティブな単語が多く含まれるクラスター 3 に分かれた。この結果は先行研究の分析的な思考の変化に対応していると考えられる。

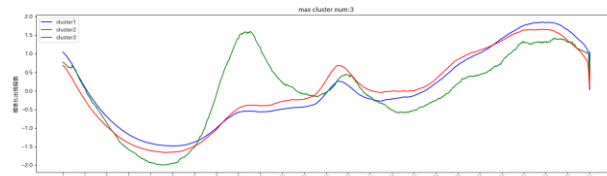


図 2: 形容詞のクラスタリング(単語出現数)

4 まとめ

感情や行動の表現に一定の日周期と傾向が見られ、人々の行動と整合している。これらの傾向の原因には日毎の思考の周期が影響していると考えられる。また、日周期は週の周期に強い影響を与えている。

より一般的な結果を得るために単語の出現数の下限を上げた結果、分析使える単語の種類が減ってしまった。さらに多くのツイートを収集する必要がある。

参考文献

- [1] Fabon Dzogang, Stafford Lightman, Nello Cristianini (2018) Diurnal variations of psychometric indicators in Twitter content, *PLOS ONE (Open Access Journal)*
- [2] LIWC, <http://www.liwc.net/>
- [3] ツイートをリアルタイムでストリーミング、Twitter 開発者チュートリアル
<https://developer.twitter.com/en/docs/tutorials/stream-tweets-in-real-time>
- [4] 工藤拓 (2013) MeCab, <http://taku910.github.io/mecab/>
- [6] 金田一春彦 (1950). 「国語動詞の一分類」『言語研究』15, 48-63