

# 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それぞれの感情や行動を反映させている。これらのデータは膨大であり、TwitterAPI を用いることによって、これらの情報を時刻やその他のデータとともに入手することができる。また、人間は感情を元に意思や行動を決定している。

そこで、時系列変化について変動する Twitter 利用者の感情・行動表現を観察することで、それぞれの時間帯や曜日における大衆の心理状態やそれに伴う行動の傾向を把握できれば、マーケティング等の分野に役立てることができると考えた。これまでも Twitter を用いて感情や行動の表現の時系列変動を見る研究はあったが、その多くは対象が限定的である。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。

## 2 先行研究

### 2.1 コロナと感情の傾向の変化[1]

この先行研究では、2019 年に確認された新型コロナウイルス感染症が社会的にどのように話題とされ、どのような影響を人々に与えていたかを明らかにするため、2020/1/17-4/30 の日本の Twitter 上でそれに関連した単語を含むツイートを収集し、人々の関心と感情の変化を分析している。

大きく分けて 2 つの手法を用いて分析を行っている。一つはユーザの偏りの評価、もう一つは感情成分の時間的变化である。

どちらの分析からも、3 月の三連休中に存在したと言われる「気の緩み」の存在を示唆する集合現象が観測された。複数の分析から同等の結果が得られたことから、本来定量的には

評価できない「気の緩み」の存在がソーシャルメディアから推定できる可能性があることが示唆された。

より長期間の分析、より精度の高い感情分析手法を今後の課題としている。さらに、トイレットペーパーの買い占めが例に挙げられるインフォでミック対策として、情報拡散メカニズムを明らかにすることの重要性を主張している。

### 2.2 英国での研究例[2]

この先行研究では、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[3]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 予備実験

### 2.1 概要

予備実験では、2020/09/23 の 7:00-8:00 にランダムに取得した 5000 ツイートに対して、動詞、形容詞それぞれの出現頻度上位 30 個を求めた。

## 2.2 手法

TwitterAPI のうち、全ツイートのうちの1%を常にサンプリングできる API[4]を用いて、Python でツイートデータを json ファイルに収集していく。ツイートデータには様々な属性情報がついているが、そのうち、ツイート ID、ツイート時刻、ツイート本文テキスト、ユーザ ID、ツイートのソース(つぶやくのに使ったアプリケーションなど)、返信であれば返信元のツイート ID、ツイートの属性(RT やリプライで有るかどうか)、リツイート数、高評価数、リプライ数、引用された回数の属性情報を取得した。また、言語設定が日本語であるもののみを取得した。

その後、json ファイル中のツイート本文テキストに対して形態素解析器 MeCab[5]で形態素解析[1]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。そして、動詞と形容詞のうち、出現数の多い 30 個を出力した。

## 2.3 結果

表 1 動詞の出現数

動詞	数
する	2310
てる	710
いる	604
なる	532
ある	424
れる	419
思う	272
くる	261
見る	231
やる	216
できる	207
くださる	194
言う	183
みる	154
行く	140
くれる	139
せる	136
頑張る	125
られる	122
当たる	117
出る	114
いく	113
すぎる	105
来る	90
しまう	82
使う	79
く	78
いう	77
おる	75
持つ	73

表 2 形容詞の出現数

形容詞	数
ない	245
いい	245
良い	153
可愛い	78
多い	63
嬉しい	60
楽しい	60
欲しい	56
早い	45
ほしい	44
よい	43
すごい	41
かわいい	38
高い	36
無い	35
強い	33
悪い	31
やすい	28
新しい	28
えらい	27
怖い	24
難しい	23
美味しい	22
面白い	21
涼しい	20
寒い	18
寝い	17
うまい	17
おいしい	16
辛い	15

## 2.4 考察

「する」や「ない」が圧倒的に多い。これらからは感情や行動を具体的に読み取ることは困難である。ゆえに、このような単語はストップワードとして除去することを検討する。しかし、ほかの行動語や感情語の評価をする際に、相対的な指標として役立つ可能性はある。形容詞に関して、どちらかというとポジティブな印象の単語が上位にある。

## 4 まとめ及び今後の展望

TwitterAPI を使用したツイートの収集及び MeCab を用いた形態素解析を行うことにより、使用頻度の高い動詞及び形容詞を調べることができた。

今後は時間帯や曜日ごとにツイートを集積していく。全体を通して出現頻度が上位の単語に対して、時刻や日の変化によってその頻度がどのように変化するかを分析する。

## 参考文献

- [1] 鳥海不二夫、榊剛史、吉田光男(2020)ソーシャルメディアを用いた新型コロナ禍における感情変化の分析、人工知能学会論文誌
- [2] Fabon Dzogang, Stafford Lightman, Nello Cristianini (2018) Diurnal variations of psychometric indicators in Twitter content, *PLOS ONE(Open Access Journal)*
- [3] LIWC, <http://www.liwc.net/>
- [4] 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>
- [5] 工藤拓 (2013) MeCab,  
<http://taku910.github.io/mecab/>