ソーシャルメディアにおける COVID-19 に対する集団感情の経時的変化についての分析

Nguyen Cong Khon† 土屋雅稔 † 吉田 光男 ‡
† 豊橋技術科学大学 情報 · 知能工学課程 · 専攻 ‡ 筑波大学 ビジネスサイエンス系
{nguyen,tsuchiya}@is.cs.tut.ac.jp, mitsuo@gssm.otsuka.tsukuba.ac.jp

1 はじめに

COVID-19 の世界的な流行，いわゆる新型コロナ禍は，2022 年 2 月現在も大きな社会的 · 経済的な影響を及ぼしている．新型コロナに対する人間集団の感情を観察することは，各種の災害（新型コロナ禍を含む）が我々の社会に与える社会的 · 経済的な影響を推測するために有用である [1]．特に，ソーシャルメディアを通じた観察は，広範囲の人間集団の動向を迅速に把握することができると期待される．本稿では，広範囲の人間集団によって生成されたテキストを大量に収集することができるソーシャルメディアとして，Twitter[1] に注目する．

Twitter に投稿されたテキスト（以後，ツイートと呼ぶ）に表出された感情を分析する研究は，数多く提案されている．梶原らは，同一のツイートを対象として，投稿者自身による主観的な感情評定と第 3 者による客観的な感情評定を行い，客観的な感情評定は，主観的な感情を過小評価する傾向があることを報告している [2]．つまり，単一のツイートに基づいて第 3 者が投稿者の感情を推測することは困難である．

以上の議論に基づいて，本稿では，以下のような方針により，人間集団の感情の分析を試みる．第 1 に，個人の感情は複雑かつ流動的であるから，単一の投稿者による単一のツイートを対象として感情を分析する代わりに，ある程度の規模の投稿者集団によって投稿されたツイートを対象として感情を分析する．第 2 に，新型コロナ禍は現在なお進行中の現象であり，この現象に対する人間集団の感情もまた日々変化している．そのため，人間集団の感情が経時的にどのように変化しているか，という点に重点をおいて分析する．

1） https://twitter.com/

2 分析手法

2.1 ツイートの収集

この節では，COVID-19 に言及しているツイートを収集する手順について述べる．

2020 年 2 月 28 日に，北海道を対象として緊急事態宣言が発出された．この日以後，COVID-19 に関するツイートをしているユーザの偏りがほぼなくなり，COVID-19 が一般的な話題となったことが報告されている [1]．そのため本稿では，収集対象期間を，2020 年 2 月 29 日から本研究の開始時点 (2021 年 8 月 31 日) までとする．

次に，COVID-19 に言及しているツイートを，以下の手順により収集する．キーワード「コロナ」「COVID-19」「coronavirus」「肺炎」「新コロナ」「デルタ」を 1 つでも含んでいるツイートを，日本語サンプルストリーム[2] から絞り込んで用意した．2020 年 2 月 29 日から 2021 年 8 月 31 日までの日本語サンプルストリームから当該ツイートを抽出したところ，1,518,714 件のツイートが得られた．投稿者自身の感情に注目するため，リツイートは対象外とした．収集されたツイートを，週別に集計した結果を図 1 に示す．図 1 の横軸は，週の開始日であり，縦軸は当該週のツイート件数である．図 1 から，COVID-19 に言及しているツイートが増加している時期と，ツイートが減少している時期を繰り返していることが分かる．これは，感染状況の変化によって関心が高まっている時期と，関心が低下している時期の反映であると考えられる．以後，ツイートが増加している時期を，ピーク期間と言う．図 1 より，ピーク期間は 5 回あったと考えられる．以後の分析では，表 1 に示すピーク期間を用いて分析を行う．

2） Twitter API の GET statuses/sample のエンドポイント [3] に対して，lang=ja で絞り込んだデータ．

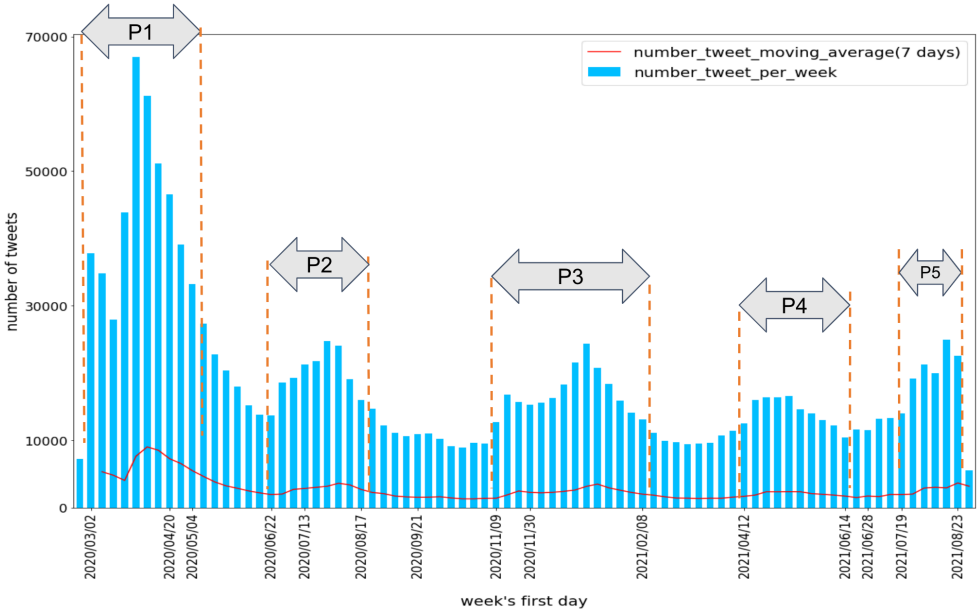

図 1　COVID-19 に言及しているツイートの件数

	開始日	終了日
P_1	2020/3/2	2020/4/20
P_2	2020/6/22	2020/8/17
P_3	2020/11/9	2021/2/8
P_4	2021/4/12	2021/6/14
P_5	2021/7/19	2021/8/23

表 1　ピーク期間

n	$\lvert V_n \rvert$	$\lvert E_n \rvert$
1	381,918	44,968,754
2	42,550	4,985,909
3	11,054	96,2670
4	3,401	194,461
5	933	30,515

表 2　ソーシャルグラフの諸元

2.2　ソーシャルグラフ

本稿では，個人単位の感情ではなく，集団の感情に注目して分析する．集団を取り出すには，何らかの形で個人間の類似性を定義し，個人をクラスタリングする必要がある．鳥海ら [1] は，各個人の発言内容に基づいて類似性を定義するという立場から，リツイートに注目してクラスタリングを行っている．本稿では，各個人の興味に基づいて類似性を定義するという立場をとり，Twitter のソーシャルグラフ上でクラスタリングを行う．

2.1 節で収集した 1,518,714 件のツイートから，明確な意見を表明していないと考えられる極めて短いツイートとして，60 文字未満のツイートを削除すると，878,206 件のツイートが得られた．この 878,206 件のツイートを対象として集計し，表 1 に示す 5 回のピーク期間の内，n 回以上のピーク期間でツイートしている投稿者の集合を V_n とする．集合 V_n に含まれる投稿者相互のフォロワー・フォロイー関係の集合を E_n とする．集合 V_n をノード，集合 E_n を有向エッジとして定義されるソーシャルグラフを G_n とする．得られたソーシャルグラフの諸元を，表 2 に示す．本稿では，集団単位での感情の経時的変化に注目しているため，発言しているピーク期間ができるだけ多いことが望ましいが，V_5 に属する投稿者はかなり少ない．そのため，投稿者数とピーク期間数の両方を考慮して，4 回以上のピーク期間でツイートしている投稿者からなるソーシャルグラフ G_4 を，以後の分析には用いる．

2.3　クラスタリング

本稿では，Louvain 法 [4] を有向グラフに拡張した手法 [5] により，ソーシャルグラフをクラスタリングする．この手法は，式 1 によって定義される目的関数（Modularity）が改善するようにクラスタを結合し，目的関数が改善しなくなるまで繰り返す．

$$Q = \frac{1}{w} \sum_{i,j} \left(A_{ij} - \gamma \frac{d_i^+ d_j^-}{w} \delta(c_i, c_j) \right) \tag{1}$$

ただし，A_{ij} は，ノード i とノード j の辺の重みである．w は，全ての辺の重みの合計であり，$w = 1^T A 1$

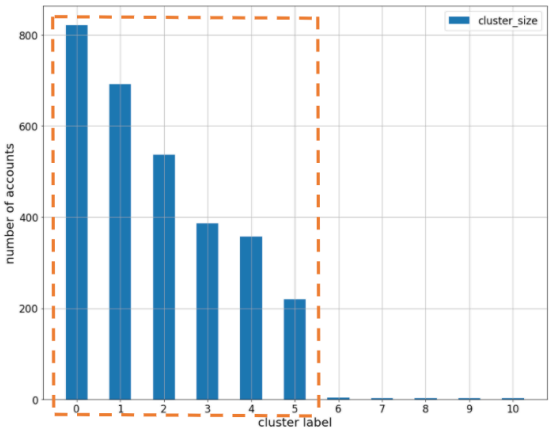

図 2　クラスタリング結果

が成り立つ．d_i^+ は，ノード i と連結している外向きの辺の重み，d_j^- は，ノード j と連結している内向きの辺の重みである．c_i と c_j は，それぞれノード i とノード j が属しているクラスタであり，δ はクロネッカーのデルタである．γ は，ハイパーパラメータである．本稿では $\gamma = 1$ を用いた．

ソーシャルグラフ G_4 を対象として，Scikit-Network ライブラリ [6] を利用してクラスタリングしたところ，11 個のクラスタが得られた．クラスタリング結果を，図 2 に示す．図 2 の横軸はクラスタ番号，縦軸は当該クラスタに含まれるアカウント数である．図 2 より，11 個のクラスタは，クラスタに含まれるアカウント数の観点から，多数のアカウントが含まれる 6 個のクラスタと，極めて少数のアカウントしか含まれない 5 個のクラスタに，明らかに二分することができる．以下の分析では，多数のアカウントが含まれる 6 個のクラスタのみを対象とする．

2.4　投稿意図が記述されていないツイートの削除

収集したツイートには，単なる挨拶のツイートや，感染状況やニュース記事を紹介しているツイートなど，多種多様なツイートが含まれている．本稿では，複数の投稿者からなる集団単位での感情を分析するため，できるだけ投稿者の投稿意図が推測できるツイートに限定して分析する．この節では，限定する方法について述べる．

最初に，ハイパーリンクや改行以外の文字が 100 文字以上含まれていないツイートを，分析対象から除く．この条件に当てはまる 100 文字未満のツイートを予備調査したところ，ツイート単体から投稿者の投稿意図を推測することが困難だったからである．

新型肺炎　情報

福島県
新たに５人の感染が確認されたことについて
福島県が今　会見しています

新たに５人感染確認 福島県会見｜NHK NEWS WEB
https://t.co/D33gw3JdoQ

NHK NEWS WEB 参照　会見終了後配信終了

#新型コロナウィルス
#福島県
#コロナウィルス
#COVID2019

図 3　ニュース記事を紹介するツイートの実例

次に，ニュース記事を紹介するツイートも，分析対象から除く．ニュース記事を紹介するツイートの例を，図 3 に示す．図 3 のツイートのテキストは，ニュースの引用のみからなり，投稿者の投稿意図について記述している部分はない．このようなツイートは，文字数が 100 文字以上であっても，投稿者の投稿意図を推測することは困難である．ニュース記事を紹介しているツイートには投稿者の意見が添えられている場合もあるが，予備調査の結果，投稿者の意見が添えられていない場合の方が多かった．そのため，人手で収集した報道メディア名（「NHK ニュース」「神戸新聞」など）が含まれており，かつ，URL が含まれている，という 2 条件を同時に満たすツイートは，ニュース記事を紹介しているツイートとして分析対象から除く．

さらに，感染状況を紹介するツイートも，分析対象から取り除く．図 4 に，感染状況を紹介するツイートの例を示す．図 4 のツイートのテキストの大部分は，感染状況を説明する地名や数値からなり，投稿者の投稿意図について記述している部分はない．このようなツイートも，文字数が 100 文字以上であっても，投稿者の投稿意図を推測することは困難である．そのため，ツイート中に 20 文字以上の数字が含まれるツイートは，感染状況を紹介するツイートとして分析対象から除く．

最後に，COVID-19 について議論するために重要な単語を含まないツイートも，分析対象から除く．収集されたツイートを単語分割 [7] して得られた単語を，TF-IDF 法によってスコアリングし，上位 1000 語を重要単語とする．ただし，ツイート単体は文書とみなすには極めて短いため，あるアカウントが，あるピーク期間に投稿している全てのツイートを連結し，1 つの文書とみなす．

#武漢肺炎 4/17現在の感染者数と前7日間
（4/10→4/17）増加率
国内9849(+59%
死者 207(+73%

東京都2794(+64%
大阪府1075(+54%
神奈川 708(+62%
千葉県 630(+63%
埼玉県 590(+75%
兵庫県 481(+52%
福岡県 479(+66%

愛知県 382(+21%
北海道 369(+54%
京都府 235(+35%
石川県 160(+74%
岐阜県 136(+40% https://t.co/WX9Ysmn1oq

図 4　コロナ感染状況を紹介するツイートの実例

2.5　感情成分の検出

テキストからの感情成分の検出には，ML-Ask[8] の Python 実装 [9] を用いる．ML-Ask は，2100 語の感情語辞書 [10] のパターンマッチングにより，以下の 10 種類の感情を推定する．

喜，怒，哀，怖，恥，好，嫌，昂，安，驚

ML-Ask Python は，入力されたテキストから検出された感情を全て出力する．例えば,「コロナで外に出れないし昼から鬼滅の刃のアニメを 6 話も観ました。面白いけど疲れました. 雨が降ってきたのでジョギングできないので、もう少ししたら筋トレ始めます！」を入力すると，出力として「厭」「喜」が得られる．あるツイートから m 個の複数の感情が検出された場合，このツイートは m 個の感情を $\frac{1}{m}$ ずつ含むものとして集計する．言い換えると，本稿では，あるツイートの感情の総和は 0 または 1 である．

3　分析結果

分析対象となるツイート全体の感情の割合を集計した結果を，表 3 に示す. 感情が検出されないツイート (表 3 の None 列) は，50.3%と全体の半数に達している．感情が検出されたツイートについては，「嫌」がもっとも多く 19.5%である．また，分析対象となるツイート全体の感情の割合をピーク期間別に集計した結果を，図 5 に示す．図 5 の横軸は，ピーク期間 (表 1) であり，縦軸は，ピーク期間別に表 3 と同様の方法で集計した感情の割合である．図 5 より，感情が検出されるツイートが徐々に増加していることが分かる．また，感情が検出されたツイートについては，一貫して「嫌」がもっとも多い割合を占めている．

感情	スコア	割合
喜 (yorokobi)	502.5	6.5
怒 (ikari)	376.7	4.9
哀 (aware)	291.6	3.8
怖 (kowa)	460.1	5.9
恥 (haji)	31.6	0.4
好 (suki)	136.9	1.8
嫌 (iya)	1508.7	19.5
昂 (takaburi)	109.5	1.4
安 (yasu)	349.8	4.5
驚 (odoroki)	81.5	1.1
None	3892.0	50.3

表 3　全体の感情成分の集計

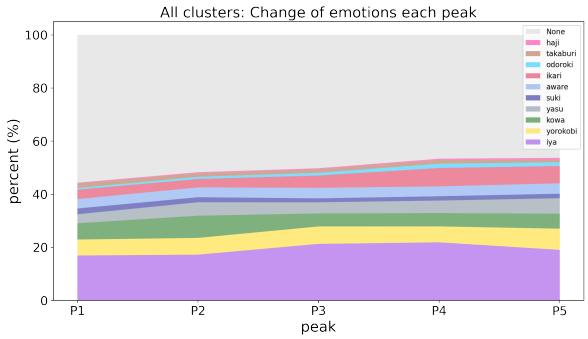

図 5　クラスタ全体の平均のピークによる感情成分の変化

次に，クラスタとピーク期間別に感情の割合を集計した結果を，図 6~11 に示す．図 5 と図 6~11 を比較すると，集団（クラスタ）の感情の割合の時間変化は，それぞれ少しずつ異なることが分かる．特に，特徴的に異なるクラスタとして，クラスタ 2（図 8）とクラスタ 4（図 10）がある．図 8 から，クラスタ 2 においては，全体（図 5）よりも明らかに「怒」感情が多く，特に 3 回目のピーク期間以後に増加していることが分かる．クラスタ 2 によるツイートの内,「怒」感情が検出されたツイートの実例を図 12 に示す．図 12 より，クラスタ 2 は，政権による政策に対して批判的なアカウントからなるクラスタであることが分かる．

図 5 と図 10 を比較すると，クラスタ 4 においては，全体よりも「喜」感情と「安」感情が多いことが分かる．クラスタ 4 によるツイートの内,「喜」感情が検出されたツイートの実例を図 13 に示す．「コロナに負けない」のように COVID-19 に対する不安感を打ち消すツイートや，COVID-19 パンデミック渦中の日常についてのツイートが多い．クラスタ 4 によるツイートの内,「安」感情が検出されたツ

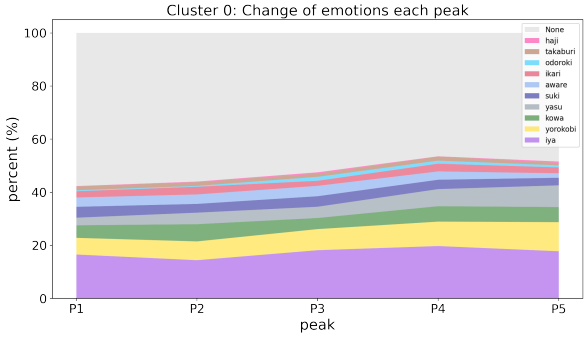

図 6　クラスタ 0 のピークによる感情成分の変化

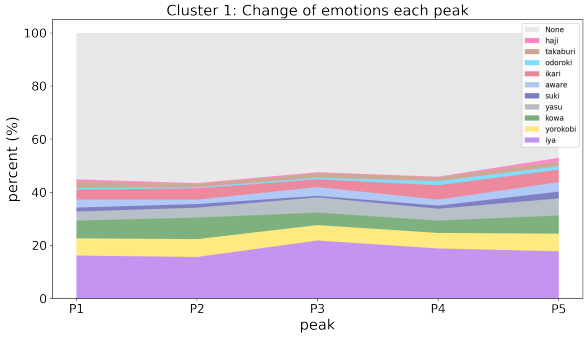

図 7　クラスタ 1 のピークによる感情成分の変化

イートの実例を図 14 に示す．コロナワクチン接種に安心する，感染対策に安心するなどのツイートが多い．以上から，クラスタ 4 は，COVID-19 に対して楽観的なアカウントからなるクラスタであることが分かる．

4　おわりに

本稿では，ソーシャルメディアとして Twitter に注目し，COVID-19 に言及しているツイートを収集して，感情の検出を行い，ソーシャルグラフ上のクラスタリングによって得られた集団の感情の経時的変化について分析した．本稿で用いたソーシャルグラフは，Twitter 上のフォロワー・フォロイー関係に基づいて作成されたグラフであるため，分析に用い

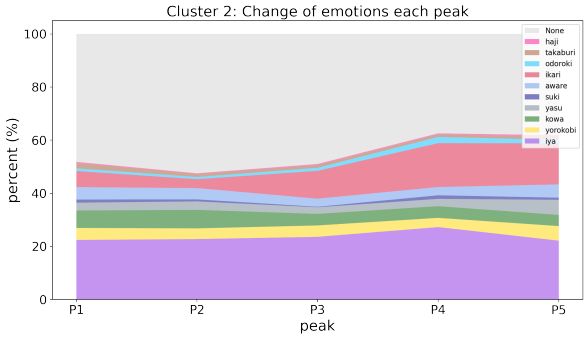

図 8　クラスタ 2 のピークによる感情成分の変化

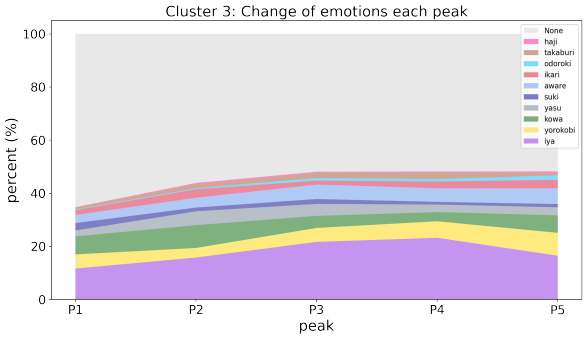

図 9　クラスタ 3 のピークによる感情成分の変化

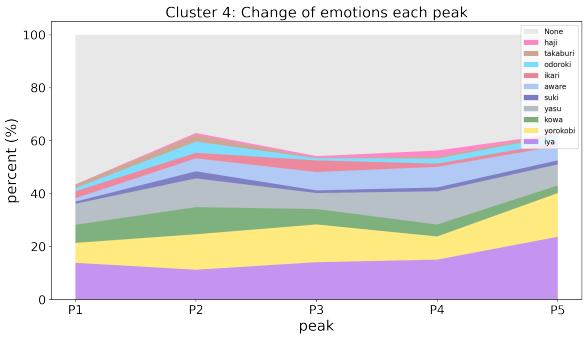

図 10　クラスタ 4 のピークによる感情成分の変化

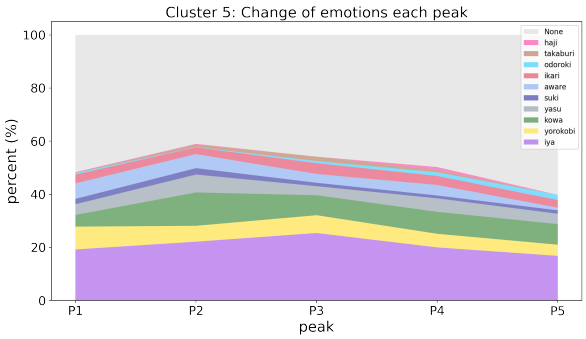

図 11　クラスタ 5 のピークによる感情成分の変化

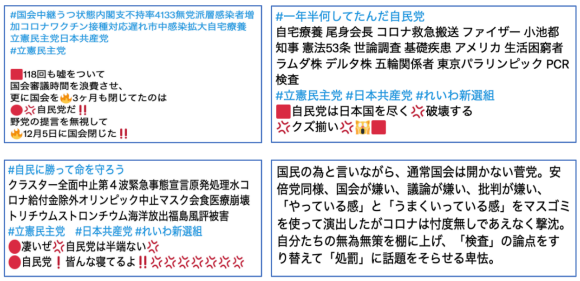

図 12　クラスタ 2 の「怒」の実例

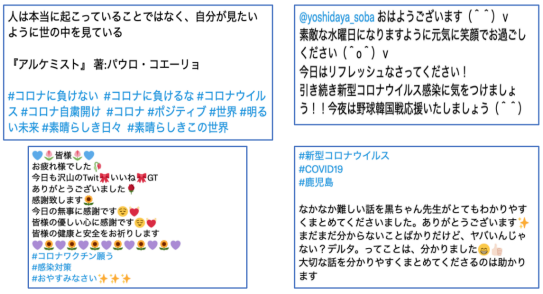

図 13　クラスタ 4 の「喜」の実例

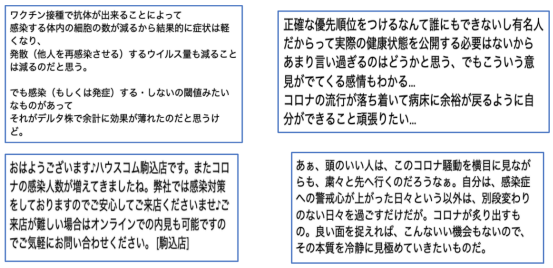

図 14　クラスタ 4 の「安」の実例

た集団は，興味の方向性で類似している集団であると考えられる．このような集団を単位とした分析においても，COVID-19 に対する感情が，集団別に異なっていることが観察された．

今後の課題としては，感情の検出方法の高精度化がある．感情が検出されなかったツイートから，100 件のツイートをランダムサンプリングして調査したところ，感情語が出現しないという観点からは感情的ではないが，実際には疑問や反語の形で様々な意図が表現されているツイートが多く観察された．このようなツイートをどのように分類し，検出するかは今後の課題である．

謝辞

本研究は，JST RISTEX JPMJRX20J3 および JST 未来社会創造事業 JPMJMI20B4 の支援を受けたものである．

参考文献

[1]鳥海不二夫, 榊剛史, 吉田光男. 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, Vol. 35, No. 4, pp. 1–7, 2020.

[2]梶原智之, Chenhui Chu, 武村紀子, 中島悠太, 長原一. 主観感情と客観感情の強度推定のための日本語データセット. 言語処理学会 第 27 回年次大会 発表論文集, pp. 523–527, 2021.

[3]Twitter. Sample realtime tweets.

[4]Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, Vol. 2008, No. 10, p. P10008, 2008.

[5]Nicolas Dugué and Anthony Perez. Directed louvain and maximizing modularity in directed networks. In *Laboratoire d'Informatique Fondamentale d'Orléans*, 2015.

[6]Scikit-Network. Scikit-network clustering.

[7]工藤拓. MeCab: Yet another part-of-speech and morphological analyzer. 2006.

[8]M. Ptaszynski, P. Dybala, R. Rzepka, K. Araki, and F. Masui. Ml-ask: Open source affect analysis software for textual input in japanese. *Journal of Open Research Software*, Vol. 5(1), p. 16, 2008.

[9]Yukino Ikegami. Ml-ask python: Emotion analyzer for japanese text. 2017.

[10]中村明. 『感情表現辞典』. 1993.