

Twitter利用者の 感情・行動表現の時系列的変動

令和2年 2月18日

情報理工学域 メディア情報学プログラム 久野研究室

4年 1710673 吉田 周平

目次

- ・ 研究背景
- ・ 研究目的
- ・ 先行研究
- ・ 準備：コーパス構築
- ・ 実験1：クラスタリング
- ・ 実験2：主成分分析
- ・ まとめと今後の課題
- ・ 参考文献

研究背景

Twitterには多くの人の感情や行動が反映される。

Twitterの時系列変化に関する研究の多くは用途が限定的である。

研究目的

個々の単語のツイートにおける出現数には日や週を単位した周期性があり、人々の行動・感情の周期性と整合するということを確かめる。

先行研究

Diurnal variations of psychometric indicators in Twitter content

(Dzogang ら, 2018)

対象：英国の54都市からTwitterに投稿された約8億件のTweet、70億の単語

期間：2010 – 2014

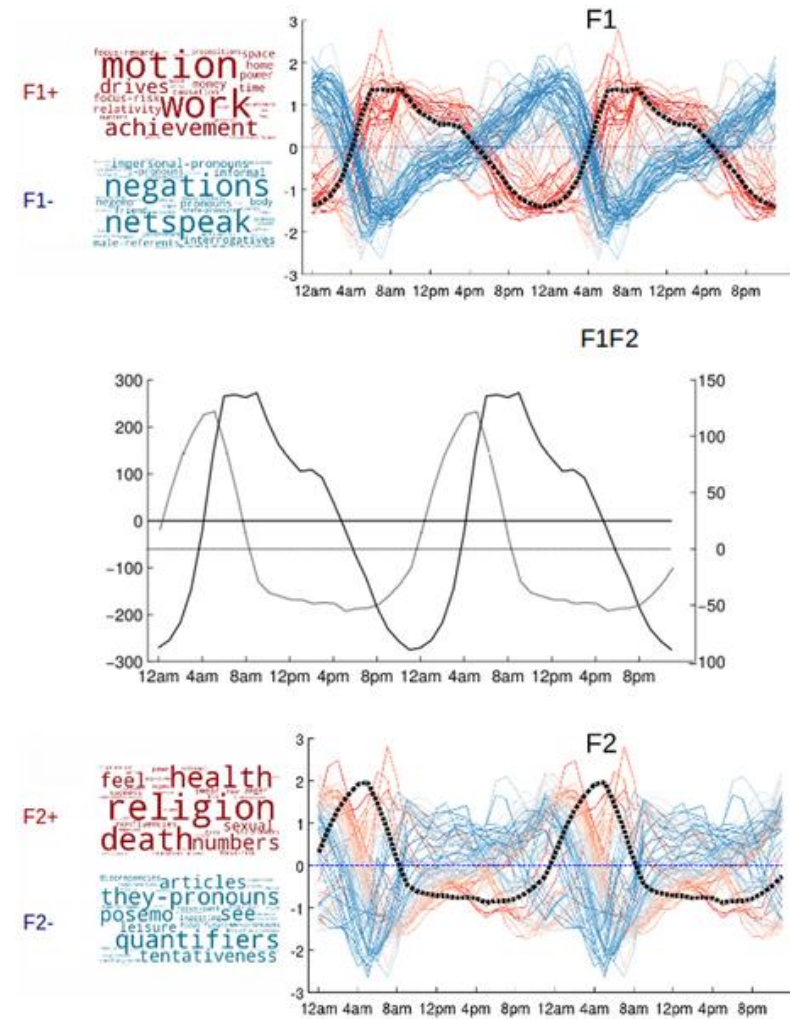
手法：LIWC（語彙を抽象化してカテゴリ化するためのツール）

73の心理測定変数を測定

5:00-6:00 → F1：分析的思考

3:00-4:00 → F2：実存的思考

Fig 3. Leading factors behind the Diurnal Variation Profile of the 73 psychometric categories.



Dzogang F, Lightman S, Cristianini N (2018) Diurnal variations of psychometric indicators in Twitter content. PLOS ONE 13(6): e0197002. <https://doi.org/10.1371/journal.pone.0197002>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0197002>

準備（コーパス構築）：取得データ

使用した属性情報

text

Tweet本文

created_at

Tweetの作成時刻

lang

ユーザの言語設定（「**ja**」のみ取得）

source

Tweetするのに使用したアプリケーション

準備（コーパス構築）：手順

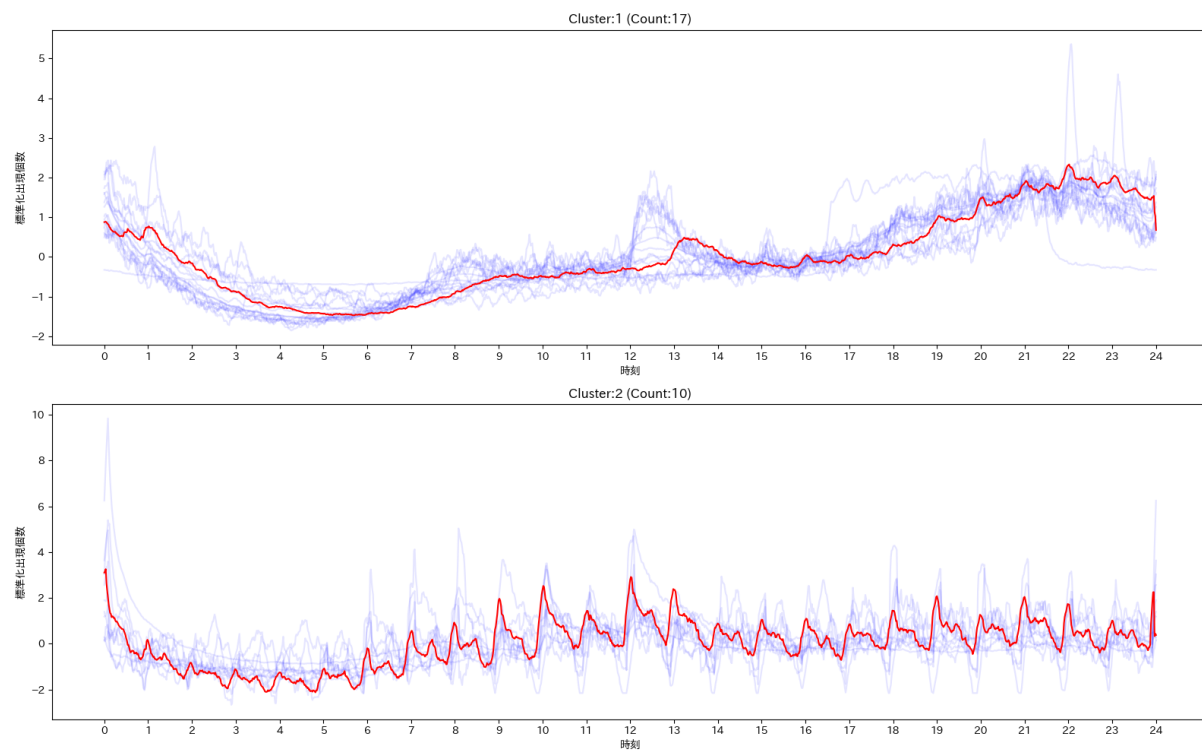
対象：言語設定が日本語であるTweet 約3000万個

期間：2020/12/3 – 2021/2/3

- ①APIで全ツイートの1%をサンプリング
- ②言語設定が日本語のツイートのみを取得
- ③定期投稿と思われるツイートの除外
- ④MeCabでツイート本文を形態素解析
- ⑤形容詞(頻度1万以上の131語)・動詞(頻度5万以上の159語)を抽出
- ⑥出現数を時系列ベクトル化（週単位/日単位）
- ⑦移動平均(前後30分)をとり、標準化

準備（コーパス構築）：定期投稿の除外

ツイートソースでクラスタリングを行うことにより、
1 時間毎の周期性が強いツイートを除外した。



ツイートソース
TweetDeck
Twitter for iPad
グランブルー ファンタジー
Twitter Web App
Twitter for iPhone

ツイートソース
OWNLY Admin
Twitter
WordPress.com
Echoes Act2
Shuttlerock - Bluebird

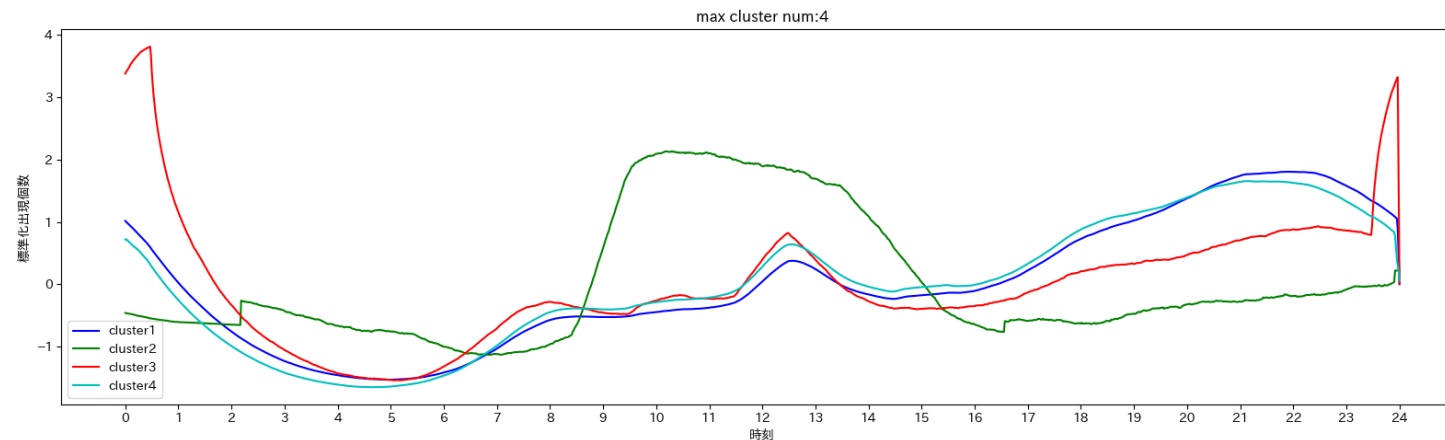
実験 1 : クラスタリング

実験 1 ： 手順

対象：動詞、形容詞それぞれの日変動と週変動

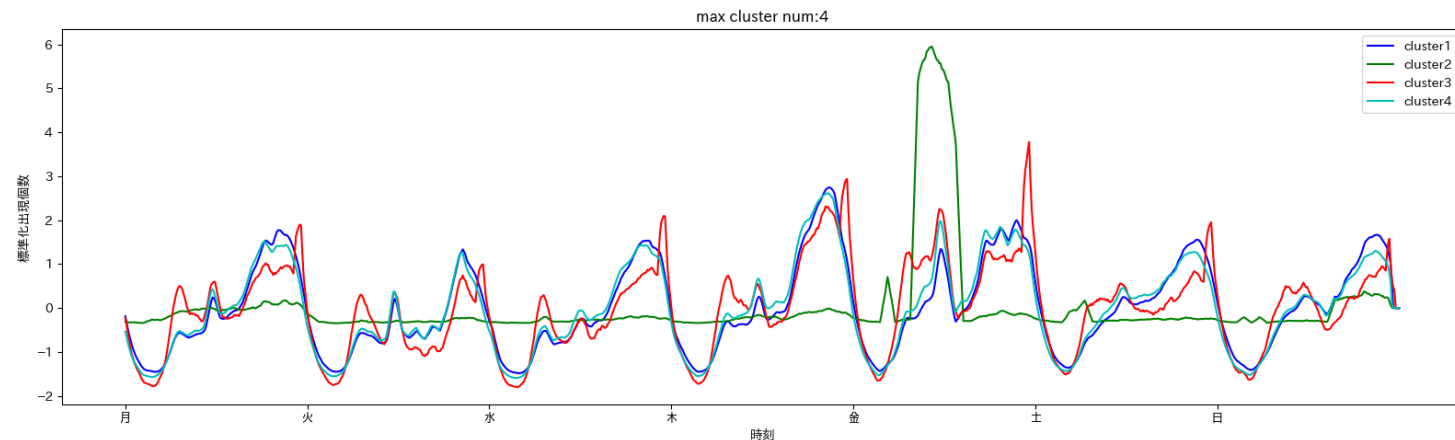
- ①K-shape法を用いて、時系列ベクトルをクラスタリング
- ②エルボー法で最適クラスター数を決定
- ③クラスター中心に近い単語でクラスターの解釈

実験 1 : 結果 動詞(日単位)



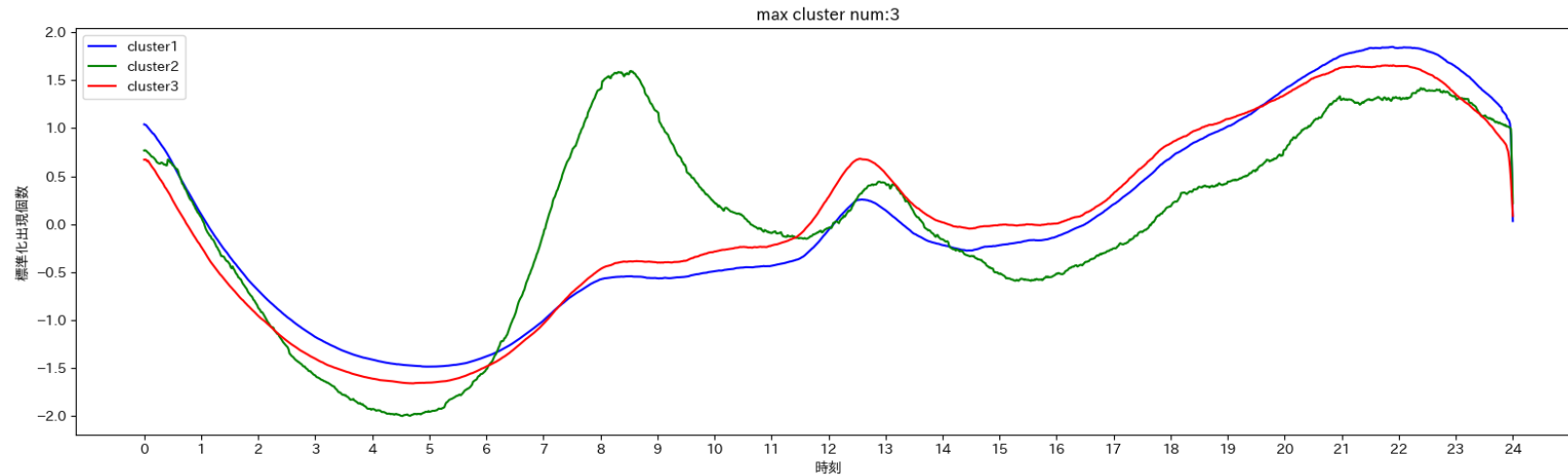
cluster	波形の特徴	単語の特徴	主要な単語
1(青)	夜に比較的多い	瞬間(変化)	分かる/始める/やめる/終わる/疲れる
2(緑)	一定の間多くなる(時刻はまばら)	挨拶/イベント/起床	あける/起きる/答える
3(赤)	日付の境目で急に増える	イベント/就寝	置く/届く/当たる/寝る
4(水)	日中に比較的多い	状態	いる/する/れる/増える/ある

実験 1 : 結果 動詞(週単位)



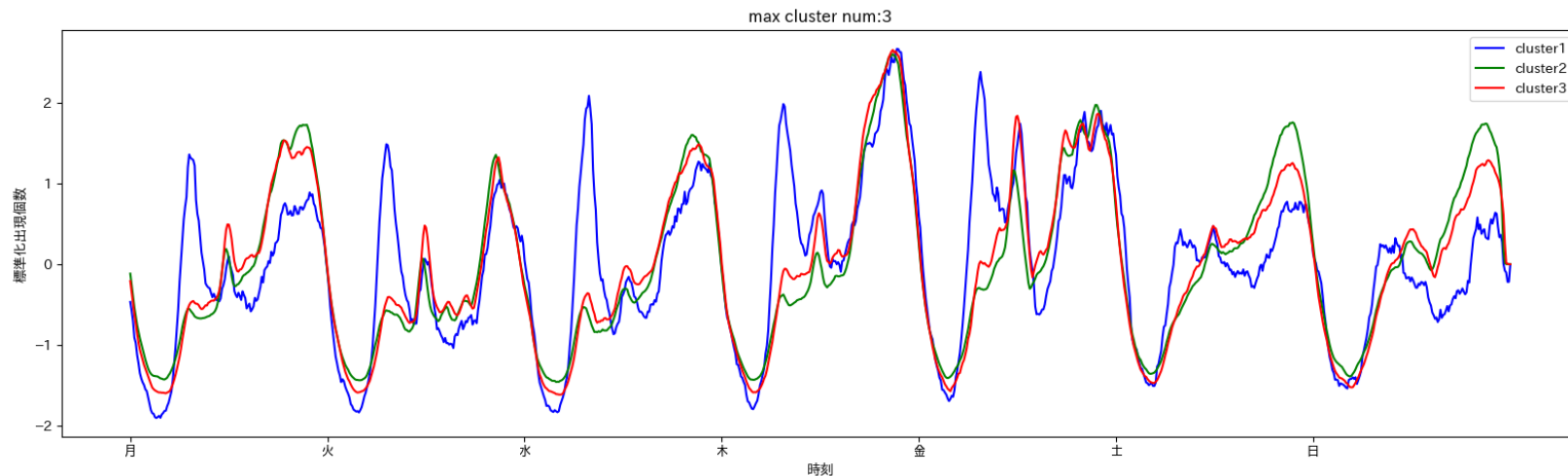
cluster	波形の特徴	単語の特徴	主要な単語
1(青)	夜に比較的多い	瞬間(変化)	分かる/始める/やめる/終わる/疲れる
2(緑)	金曜日のみ非常に多い	新年の挨拶	あける
3(赤)	朝に多い	起床/キャンペーン	頑張る/過ごす/当たる/置く/起きる
4(水)	日中に比較的多い	状態	いる/する/れる/増える/ある

実験 1：結果 形容詞(日単位)



cluster	波形の特徴	単語の特徴	主要な単語
1(青)	夜に比較的多い	主観的	すごい/いい/悲しい/仲良い/よい
2(緑)	朝に非常に多い	温度感覚/眠気	暖かい/ねむい/寒い/さむい/眠い
3(赤)	日中に比較的多い	客観的	高い/大きい/多い/少ない/新しい

実験 1 : 結果 形容詞(週単位)



cluster	波形の特徴	単語の特徴	主要な単語
1(青)	朝に非常に多い,土日は少ない	温度感覚/眠気	暖かい/ねむい/寒い/さむい/眠い
2(緑)	夜に比較的多い	主観的	すごい/いい/悲しい/仲良い/よい
3(赤)	日中に比較的多い	客観的	高い/大きい/多い/少ない/新しい

実験 1 : まとめと考察

- 週の変動から、1日毎に周期があることがわかる。
- 平日と休日で異なる変化をする単語が存在する。
- 行動表現と人々の行動は整合する。
- 朝から夜にかけて状態から瞬間（変化）という動作表現の変化と、客観から主観という感覚表現の変化が現れている。

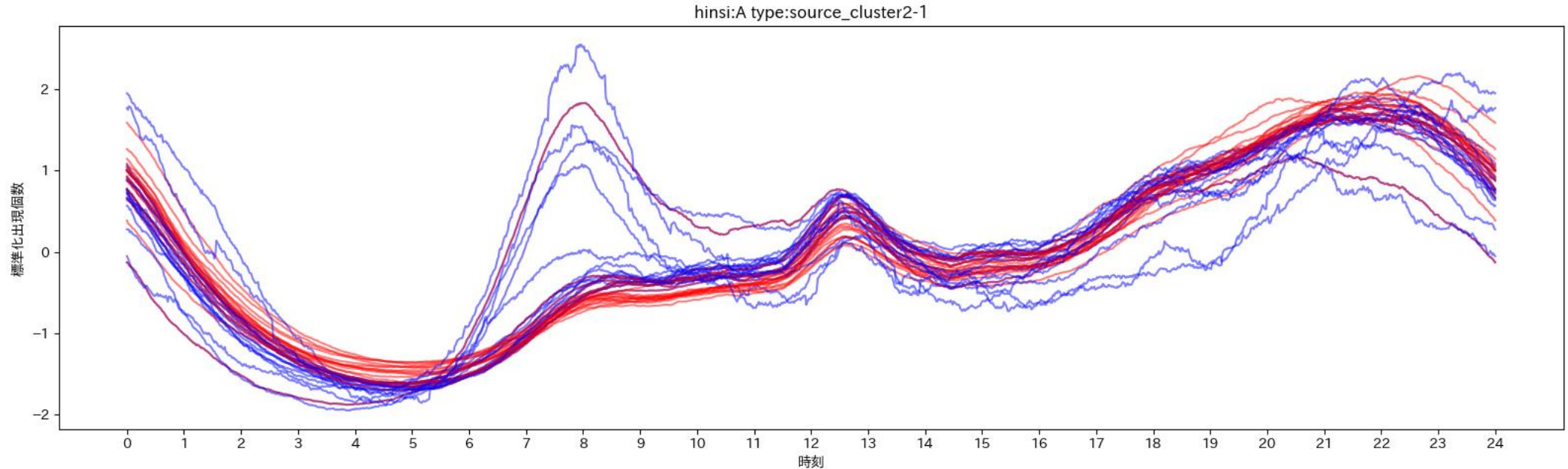
実験 2 : 主成分分析

実験 2 ： 手順

対象：動詞、形容詞それぞれの日変動と週変動

- ①時系列ベクトルを主成分分析
- ②累積寄与率を計算し、主成分選択
- ③主成分得点を計算
- ④主成分の解釈

実験 2 : 主成分得点の高い形容詞(日)



主成分 1 (赤) ; 朝から夜にかけて増える一般的な波形
主成分 2 (青) ; 朝に非常に多くなる波形

まとめ

- ・感情、行動表現の時系列変動に周期性(日/週)が認められた。
 - ・状態を表す動作表現は比較的朝から昼に多い。
 - ・瞬間的な変化を表す動作表現は比較的夜に多い。
 - ・客観的な感覚表現は比較的朝から昼に多い。
 - ・主観的な感覚表現は比較的夜に多い。
- ・行動表現が人々の行動と整合することが認められた。

今後の課題

- ・ 挨拶メッセージの除去など、前処理を工夫する必要がある。
- ・ 単語の種類がかなり絞られてしまった。
さらに長い期間でより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/>
5. Kshape, <https://github.com/tslearn-team/tslearn/blob/a3cf3bf/tslearn/clustering/kshape.py#L21-L291>
6. 金田一春彦 (1950). 「国語動詞の一分類」 『言語研究』 15, 48-63