

Twitter における網羅的トレンド解析方式・可視化手法の提案

福島 健四郎[†] 吉田 尚史[‡][†] [‡] 駒澤大学 GMS 学部 〒154-8525 東京都世田谷区駒沢 1-23-1E-mail: [†] kenkensz9@yahoo.co.jp, [‡] naofumi@komazawa-u.ac.jp

あらまし：本論文では、Twitter において、Twitter 社の提供していないより日常用語的な**大局的トレンド**について、単語の他に意味辞書や共起等を用いて収集し、独自の時系列分析・ランク付け・可視化をする新たな方法で、Twitter を通した現実世界の理解を深めることを目的とした。SNS の網羅的トレンドについての先行研究は少ない。従来の変化率だけに頼るのではなく、変化の変化である「**勢いデータ**」を利用し、勢いの大きさを継続する時間の間だけ足し合わせてランク付けをした。更にそれを用いて、「**トレンド早見盤**」による新たな効率的な可視化手法を提案する。また、その可視化から読み取れることについて考察を行う。

キーワード：ツイッター トrend 時系列データ解析 可視化 計算社会科学

1 序論

1-1 背景

ツイッター社の提供するトレンドは、それ自体がメディアやツイートによって言及されるなど注目を集めることが多い。しかし、その提供システムはツイッター上における社会運動の目標とされたり、スパムによる操作がされたり、ツイッター社が操作しているのではないかとされるほど**不確かなものである**。

ツイッター社のトレンド抽出アルゴリズムは増加率に基づいていると言われるが、**それはトレンドの一面しか示していない**。よって、それだけを見ることによって実社会での関心度合いを計ることは難しい。

しかしながらツイッターは、以下のような重視すべき意味[1]を持っていると考える。

- ・人々の欲望や独り言が溢れている。
- ・その場は相互に影響を与え合い、言葉がミームとして行き交うところである。

(パノプティコン＝監視ではなく、あくまで**同意のある相互監視**)

- ・ツイッター自体は意思決定を下さないものの、集合知としても機能する(リツイートや情報共有など)
- ・そのようなビッグデータが刻々と生成されている。

ツイートを集合知・現実の鏡と見て社会を分析する計算社会学等におけるツイッター研究はあるが網羅的、それこそパノプティコン的な全体の監視、**俯瞰的な視点からの情報、すなわちトレンドに関する研究は少ない**。ここで監視者になり得るのはデータを集める側である。また、倫理的問題は傾向を元に個人を特定して悪用した場合に起こり得るとされる。

ツイートは目的もなく保存・公開されているに過ぎないが、ツイッターは大規模であり、どうにか俯瞰的

に全てのデータを用いて捉えよう・活用しようという動きはある。

しかし、ツイッターのビッグデータは意味を持つように構造化されておらず、何かの枠組みで捉えないと意味がない。つまり、目的はない言わばゴミの山であるため、**何らかの捉える枠組みが必要**と言える。

1-2 関連研究

論文[2]では Uchida[3]の手法を用いて、「同じユーザーにリツイートされたツイートは似たような内容であるという前提」で RT に含まれる単語をクラスタリングし tf-idf で抽出し面グラフ(ThemeRiver)で可視化している。この手法では RT を用いるため絶対的なツイート数ではなく強度という指標を用いている。単語単体でなく単語グループで可視化という点は分かりやすいが、2019 年の”箱根駅伝”を含むツイートのみを対象としているため俯瞰できているとは言い難い。

論文[4]では、感情極性辞書と RT 回数・深さを追って時系列的にバーストしたものをトレンドとしている。時系列バーストは短期的移動平均が長期を上回った場合という **FX や株で用いられる手法**を採用している。

広義のトレンドについて、リツイートの深さ・拡散経路を利用する方法もあることが分かる。

論文[5]では、「ソーシャルメディア上の多くのバースト現象は、一部のコミュニティに所属する人々だけが興味を示す状態、すなわちアカウントの偏りが生じていることが多い」としている。具体的には、**2%のアカウントが検察庁法改正案ツイートの 50%を占めていた**という研究である。

本研究ではユーザー名を用いないが、組み合わせると**信憑性のあるトレンドを取れる**と考えられる。

上の3つのような研究はあるが、網羅的かつ従来にない可視化の方法を試しているのは見つからなかった。

上記の関連研究に比べてこの本方式の特徴は、可視化を直感的に、また一部のリツイートではなく全体のツイートを利用し。より大衆的・日常的なトレンドを取得するところにあると言える。

1-3 目的

本論文では、ツイッターにおいてトピックを指定せず、網羅的なトレンドを抽出し独自の大局的な時系列解析・可視化方式を提案することで、SNS に反映される現実世界の傾向をより効率的に網羅的に把握することを目的とする。その情報を元に考察も行う。

なお、このシステムを利用するユーザーとしては、普段からツイッタートレンドに興味を持ち、更にその詳細な情報を求める人を想定している。

網羅的というのは、特定のトピックに限定せずトレンドを抽出し、意味で抽象化した、俯瞰的であることを意味する。

2 提案方式

本研究は3つの段階・提案部分に分けられる。1トレンド候補検出 2トレンドの時系列的ランク付け 3可視化である。

1,トレンド候補検出

ツイートを取得し、まず、大きなトレンド(クラスター)が含まれるかをクラスタリングして調べる。それが含まれない場合に、小さいトレンドを取得する。

小さいトレンドというのは、「単語単体・共起」と国立国語研究所が発行する意味辞書[6]で翻訳・抽象化した「意味単体・共起」を合わせたものである。この辞書は研究所が提供する国語研短単位自動解析用辞書Unidicで形態素解析した後に研究所がgithubで提供されているWLSP2UniDic[7]を用いて意味に抽象化したものである。

そしてこの作業を繰り返し、大小のトレンドの出現回数を逐次記録をする。

2,時系列的ランク付け

それぞれの出現回数推移について、カーネル法(データを高次元の特徴空間上へ写像することによって、線形構造を作りフィッティング)を用いて非線形回帰を行う。その際にカーネル関数(入力データ同士の類似度を計算する関数)として、Laplacian カーネルとRBF カーネルの2つの手法で回帰分析を行い、グラフを作成する。RBFカーネルはRadial basis function kernelの略で、ガウスクーネルとも呼ばれる。Laplacianカーネルはラプラスカーネルとも呼ばれる。

それぞれ式は次のようになる。

・RBFカーネル： $K(x, y) = \exp(-\gamma ||x - y||^2)$

・Laplacianカーネル： $K(x, y) = \exp(-\gamma ||x - y||)$

ここで、 γ はハイパーパラメータを指す。

以後、RBFカーネルをRBF、LaplacianカーネルをLaplacianと省略する。

また、一般に、RBFの方がよく用いられるとされる。

Laplacianはデータが密である場合に用いられる。

次にどれほどトレンドらしいかLaplacianグラフからは単純増加率・RBFグラフからは大局的増加度の2つの指標に基づいたランキングを作る。具体的には、

・単純増加率に基づくトレンドはそのまま出現回数のグラフの内、独自のアルゴリズムで判定した最後の局面における増加率

・大局的増加度に基づくトレンドは、増加幅と増加がどれだけ継続したのかを考慮する。例えば、出現回数が1, 2, 4, 8と増加が継続している場合の階差系列は1, 2, 4となり、大局的増加度は増加部分の階差系列の和、7となる。厳密には、 $\arctan2$ を用いて、 $\arctan2(1, 1) + \arctan2(2, 1) + \arctan2(4, 1) = 8.08 \text{ rad}$ となり階差系列で大きな増加があると抑制(標準化)される。 $\arctan2$ 関数の軌跡は図2-1の通りである。

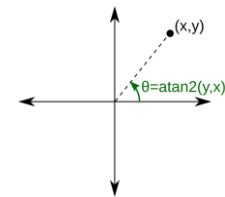


図 2-1：関数 $\arctan2$ の説明[8]

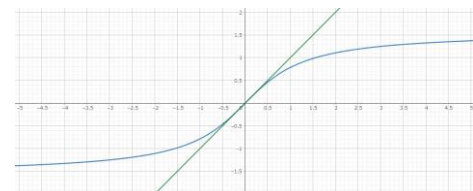


図 2-2：高さ $y=x$ 緑色 と $y=\arctan2(x, 1)$ 青色
 $\arctan2$ は、図 2-1 のように、底辺 x と高さ y から θ を求める関数である。

図 2-2 の緑の線($y=x$)が階差系列における階差とすると、青色の線は $y=\arctan2(x, 1)$ であり、その時点での勢いを指している。勢いは変化が大きいほど $\pi/2$ ラジアンすなわち 1.57 に近づく。つまり、頻度の変化が大きいほど大局的増加度は大きくなる。逆に急な変動も抑えることができる。

3,可視化

前節で作ったランキングについて、上位から増加のペースが落ちているのか増えているのか加速度($\arctan2$ を用いたもので、最大値は $\pi/2$)を縦軸に取り、出現回数を横軸に取って、トレンド早見盤として

可視化を行う。曲線中の点の間隔は 60 分である。他に、「トレンド名」「直近の出現回数」「加速度」「24 時間前との差異」を列とする表を左に、上位トレンドの動きを標準化してクラスタリングした積み上げ棒グラフを下に表示する(時系列的類似クラスター推移)。つまり、複数のグラフの動きを似たグループにまとめて、同じグループは表セルの背景で同じ色で表現する。トレンド早見図はトレンド早見盤と表とクラスター時系列推移を組み合わせた表を言う

図 2-3 は前 2 つの段階で説明した単純増加率を上段に、大局的増加度を下段において、左からトレンドの「単語単体」「意味単体」「5 つを足し合わせた総合」のトレンドを配置したトレンド早見盤である。

テーブル	トレンド早見盤 単語単体 laplacian 回帰による 単純増加率	テーブル	トレンド早見盤 意味単体 laplacian 回帰による 単純増加率	テーブル	トレンド早見盤 総合 laplacian 回帰による 単純増加率
	時系列的類似 クラスター推移		時系列的類似 クラスター推移		時系列的類似 クラスター推移
テーブル	トレンド早見盤 単語単体 RBF 回帰による 大局的増加度	テーブル	トレンド早見盤 意味単体 RBF 回帰による 大局的増加度	テーブル	トレンド早見盤 総合 RBF 回帰による 大局的増加度
	時系列的類似 クラスター推移		時系列的類似 クラスター推移		時系列的類似 クラスター推移

図 2-3：トレンド早見図の配置

ここで実際のトレンド早見図は図 2-4 のようになる。

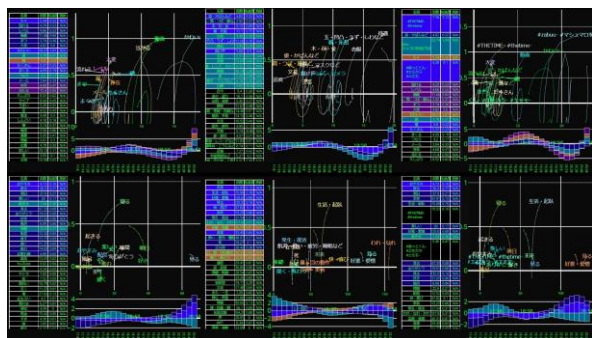


図 2-4：2022/10/21 金曜日 21 時のトレンド早見図

3 実現方式(実装)

3-0 概要

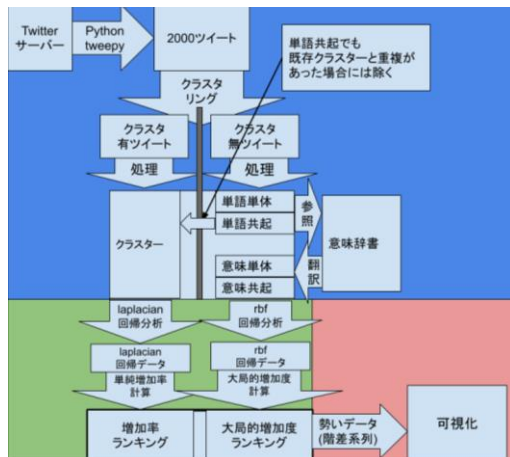


図 3-0-1：システムの全体図

システムは図 3-0-1 で色分けした 3 つに分けられる。

1. トレンド候補検出(青色の領域)
2. 時系列的ランク付け(緑色の領域)
3. 可視化(赤色の領域)

なお、Laplacian は $\alpha=0.01$ 、RBF は $\alpha=0.1$ 、平滑化には `scipy` の `savgol_filter` を用い、区間は長さの 8%、次数は 4 とした。クラスター判定にはダイス類似度(7 割以上で同一)を用いた。単語には、McCab の名詞・形容詞・形容動詞・未知語を取得した。可視化の際に Plotly はラベル重複回避機能が無いため、Github のライブラリ `rectangle-overlap` [9]を用いた。全ての単語を処理する事はできないため、出現回数の平均が 2000 ツイート中で単語は 2,共起は 5 未満のものは除外した。

4 実験

4-0 実験環境

- ・ Windows 10 Core i7-6700K メモリ 16GB
- ・ Anaconda 4.9.2 Python 3.9.0

他に、Chrome において Google Colab も使用した。

実験 1-4 は、2022 年 10 月 23 日 日曜日に収集したデータを用いた。

4-1 実験 1 (トレンド候補検出実験)

4-1-1 目的 5 種類のトレンドの時系列データを集めた集計表が取れることを確認する。

4-1-2 方法 トレンド候補検出プログラムを動作させ、集計表として csv へ出力し、エクセルで開く。

4-1-3 結果

図 4-1-3-1 の A 列はトレンド名、B 列は合計、C 列は平均、D 列以降は観測データである。観測データの上は、日付時刻をシリアル値で表したものである。

	A	B	C	D	E	F	G	H	I
1	trend	count	ave	2022/10/21 00:00:00	2022/10/21 01:00:00	2022/10/21 02:00:00	2022/10/21 03:00:00	2022/10/21 04:00:00	2022/10/21 05:00:00
2	「いーやー」なれ、ふたつ	319019	303.9132	185	166	167	210	165	136
3	「いーやー」なれ、ふたつ	311132	288.0576	182	172	185	145	183	179
4	「いーやー」なれ、ふたつ	280834	145.3879	143	165	135	134	135	139
5	「いーやー」なれ、ふたつ	257027	132.6759	134	122	132	105	115	127
6	「いーやー」なれ、ふたつ	252079	130.18	151	145	129	122	135	141
7	「いーやー」なれ、ふたつ	230034	126.0463	130	147	137	139	142	139
8	「いーやー」なれ、ふたつ	229079	117.839	113	124	132	111	106	114
9	「いーやー」なれ、ふたつ	215686	118.9496	116	114	126	101	103	122
10	「いーやー」なれ、ふたつ	197020	104.7479	101	101	100	108	104	98
11	「いーやー」なれ、ふたつ	189449	102.5972	102	110	120	101	104	105
12	「いーやー」なれ、ふたつ	181430	108.4825	95	92	105	90	80	89

図 4-1-3-1：「(-)<>」に囲われているのが意味単体

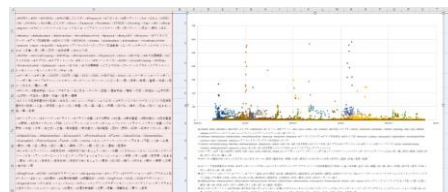


図 4-1-3-2：クラスターの上位 10-19 番目の散布図

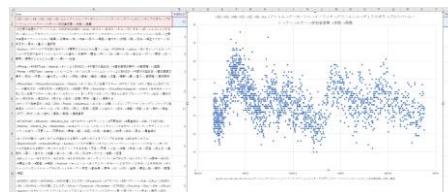


図 4-1-3-3：クラスターの内、最も出現回数が多かった

たもの(スマホゲーム関連)の時系列推移である。
参考として、クラスタリング機能を切った場合には、
クラスターは単語の共起(断片)として現れた。

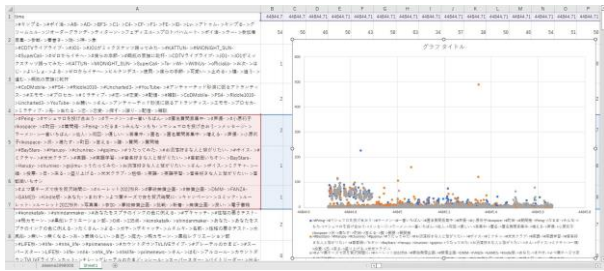


図 4-1-3-4: 参考として、クラスターとの共起 1 つで
同一と見做した場合

4-1-1 考察

全ての種類のトレンドがしっかりと取れていた。
クラスターについては、ここからここまで現れる/
現れないが明確読み取れ、ノイズの削除として機能し
ていると考える。周期的に現れるものもあった。
そして、全体的に観測値の分散は比較的少なかった
ため、信頼できる回帰曲線が得られると言える。
また、実験データには示していないが、分析の経験
から以下のことが言える。

- ・様々なトレンドを含んだトレンドお知らせツイ
ートやスパムによって、誤検知が出ることがあった。
- ・クラスターは長さ 18 に達すると変化しないため、
時々話題の変化についていけないことがある。そのた
め、ハッシュタグは有効活用すべきと言える。

4-2 実験 2 (時系列的ランク付け実験)

4-2-1 目的 時系列解析するには観測データの点を
回帰分析して線にする必要がある。どのような回帰手
法(カーネル)が適しているのかを見る。

4-2-2 方法 Laplacian と RBF の 2 つの方法で回帰
分析し上下を識別し増加率を取る。また、階差系列を
元にして勢いデータを作る。

4-2-3 結果

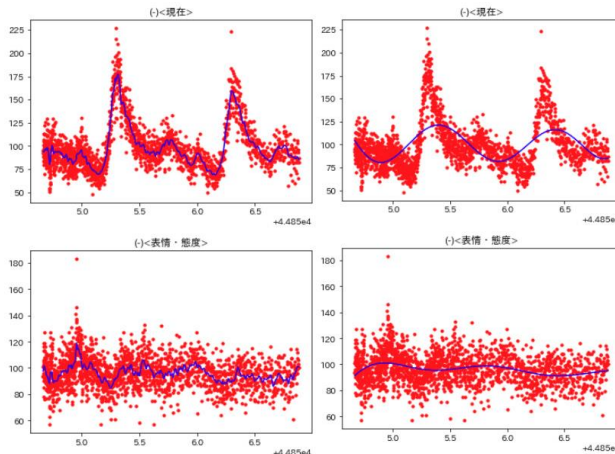


図 4-2-3-1: それぞれのトレンド項目について、

左が Laplacian、右が RBF の場合のグラフである。

横軸は時間、縦軸は出現回数
赤の点は観測された出現回数
青は回帰曲線を示す。

図 4-2-3-2 はそこから上昇と下降を定義し、更に勢
いデータを加えたグラフを示した図である。

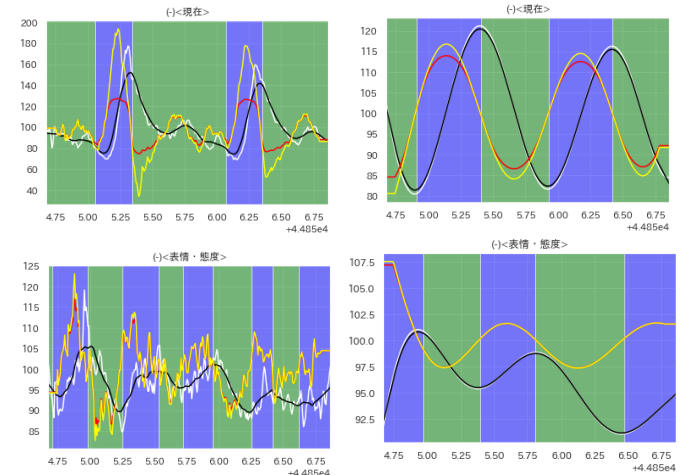


図 4-2-3-2: 左が Laplacian、右が RBF であり、
青い部分は増加、緑は減少と判定された局面を指す。
白グラフは、回帰曲線
黒グラフは、回帰曲線を平滑化したグラフ
黄グラフは、黒グラフを元にした階差系列
赤グラフは、黄グラフを元にした勢いデータを指す。
また、黄・赤のグラフは見やすくするために、20 倍
して 100 を足してある。
そして、図 4-2-3-3 のように増加と減少を数値化し、
単純増加率を計算する際は最後の増減データを用いる。

(-)<現在>	
[値, '16:18', 93.60287825882689]	
[増減, '01:18', 55, '-', '1', -12.31, -0.22]	
[値, '01:18', 82.08163571719048]	
[増減, '08:18', 42, '-', '1', 81.88, 1.95]	
[値, '08:18', 149.29163526305283]	
[増減, '01:38', 104, '-', '1', -46.61, -0.45]	
[値, '01:38', 79.71170491542246]	
[増減, '08:28', 41, '-', '1', 74.97, 1.83]	
[値, '08:28', 139.46838192486877]	
[増減, '20:48', 74, '-', '1', -38.48, -1.92]	
[値, '20:48', 85.79428716368876]	

図 4-2-3-3: 例として意味トレンド「現在」の増減表
他に、図 4-2-3-2 左上の図を見ると、「現在」の勢いデ
ータは arctan2 を用いることによって、階差系列の黄
→赤となり、急な増減が抑制されていることが分かる。

4-2-3 考察

Laplacian 回帰のほうが細かい動きを表している。
RBF は大局的かつ増加と減少を明確にしている。特に、
グラフの最後の部分が上昇か下降かを Laplacian と違
い、RBF は明確に示している。

Laplacian と RBF で増減の領域は、ある程度一致し
ているが、短期的には Laplacian が適していることが
分かる。

4-3 実験3 (ランキング作成実験)

4-3-1 目的 Laplacian と RBF の回帰方法・単純増加率と大局的増加度の測定を組み合わせることで実際にランキング実験を取って、どのようなトレンドが得られるのかを試す。

4-3-2 方法 2022 年 10 月 20 日 16 時 00 分から測定し、53 時間後の 22 日 21 日に取得をする。

4-3-3 結果

まず、Laplacian のランキングを作成した結果は図 4-3-3-1 のようになった。

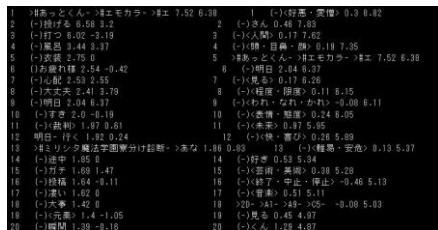


図 4-3-3-1: 左側は Laplacian の増加率を示し右側は Laplacian の大局的増加度を示す。

トレンドの横に並ぶ数字は、10 分あたりの単純増加率と大局的増加度を指している。

図 4-3-3-2 は、図 4-3-3-1 の上位 5 トレンドの、出現回数(上段: 青)と勢いデータ(下段: 赤)を示す。

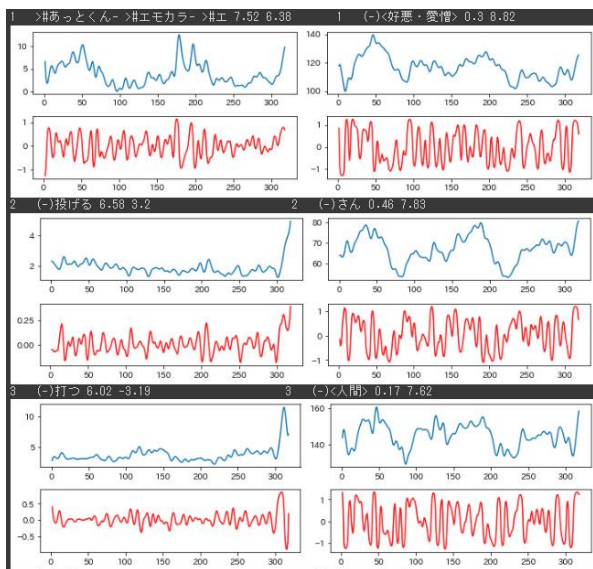


図 4-3-3-2

次に、RBF のランキングを作成した結果を図 4-3-3-3 に示す。

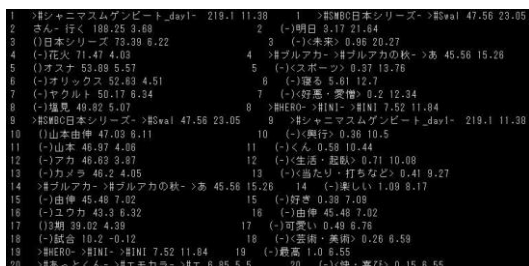


図 4-3-3-3: 左側は RBF の単純増加率を示し、右側

は RBF の大局的増加度を示す。

同様に図 4-3-3-4 は、図 4-3-3-3 の上位 3 トレンドの出現回数(上段青)と勢いデータ(下段: 赤)を示す

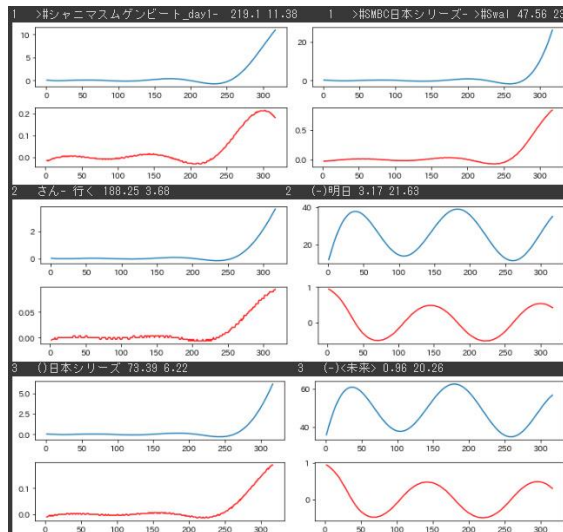


図 4-3-3-4

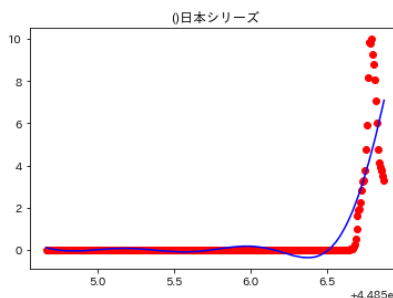


図 4-3-3-5

ここで図 4-3-3-5 に示すように、RBF は、出現回数推移に 0 が続いていると増加の直前がマイナスになってしまうという特徴がある。そのため、増加率との相性は悪い。そこで、増加率の計算で間にマイナスを挟まない場合も調べる必要がある。

マイナスとなる影響を取り除いた RBF 単純増加率によるトレンドは図 4-3-3-6 のようになる。

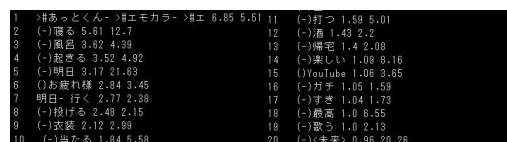


図 4-3-3-6

4-3-4 考察

これは 22 日の 21 時時点のトレンドであるが、RBF の単純増加率を見るとこの日の 18 時 30 分から 22 時 30 分まで行われた野球の単語(日本シリーズ)が多く含まれているが、Laplacian の単純増加率では「投げる」「打つ」の他には含まれていない。詳しく、RBF の単純増加率上位 3 番目「日本シリーズ」を図 4-3-4-1 で見ると、

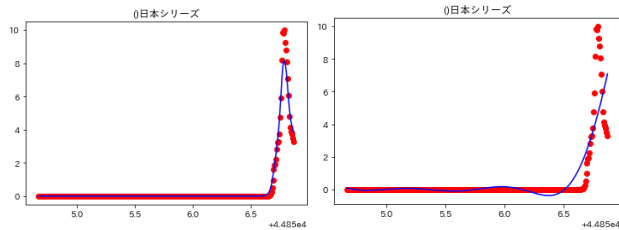


図 4-3-4-1:「投げる」Laplacian(左)と RBF(右)
赤が観測点で青が回帰曲線を示している。

図 4-3-4-1 では、Laplacian では頭打ちになっているが、右の RBF ではまだ増加中となっていた。ここでは、ゆるく回帰をするという RBF の解釈が速さについていけなかったと考えられる。

つまり、Laplacian では単純増加率を取ると、短期的トレンド抽出に有用であると言える。

次に、Laplacian 増加率 1 位 「投げる」を図 4-3-4-2 で見ると、

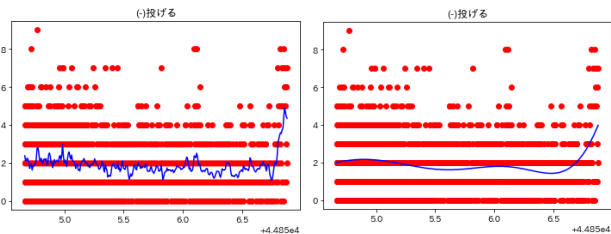


図 4-3-4-2:「投げる」Laplacian(左)と RBF(右)

図 4-3-4-2 の Laplacian ではトレンドと言えるような特徴的なグラフとなっていた。

次に Laplacian 大局的増加度 1 位 「好悪・愛憎」を図 4-3-4-3 で見ると、

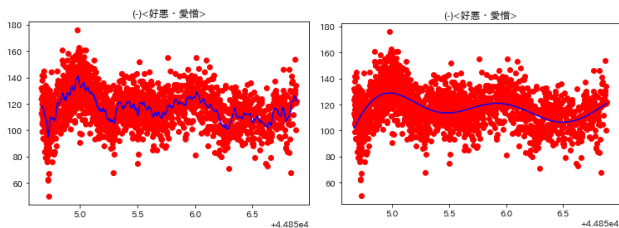


図 4-3-4-3:「好悪・愛憎」Laplacian(左)と RBF(右)

Laplacian でも増加傾向が継続していることが分かる。これは RBF の大局的増加度でも 7 位に入っていた。

また、RBF の大局的増加度 2 番目「明日」、6 位「寝る」を見ると、

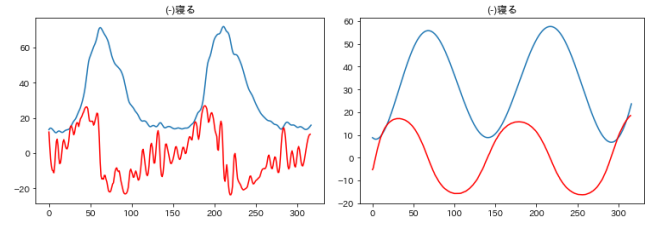
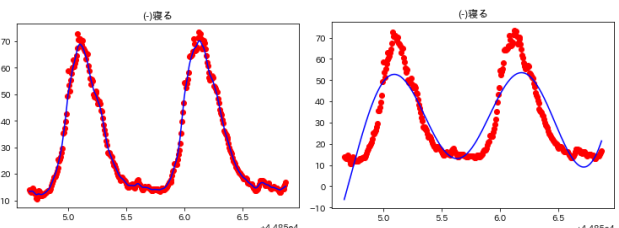


図 4-3-4-5:上から、「寝る」Laplacian(左)と RBF(右)の回帰曲線(青)と勢いデータ(赤)である。

「寝る」は、RBF では増加傾向にあることを既に示していた。RBF はこれまでの推移の解釈を保ちつつ回帰分析を行うため、少しでも上がると解釈された場合には上昇と解釈していると考ええる。

単純増加率は元のデータを利用するが大局的増加度は勢いデータを利用するため、増加がこの時点で大きく捉えられた。

図 4-3-3-4 上部分を見ると、グラフの形は同じでも、図 3-2-2 に示すように出現回数の大きさと変化の大きさの関係性によって(絶対的な)勢いが強いと判断された。よって、出現回数の多いかつ変化の大きい大局的なトレンドが RBF の大局的増加度によって導き出されることが分かる。

それを含めても、Laplacian の増加率と RBF の大局的を用いるのが良いことが分かる。

4-4 実験 4 (可視化実験)

4-4-1 目的 ランク付けしたデータを新たな方法で可視化を行う。

4-4-2 方法 可視化ライブラリの Plotly を用いてランキング順に可視化をする。

4-4-3 結果

まず、マイナスの影響を無視して単語のみを可視化した場合、図 4-4-3-1 のようになった。

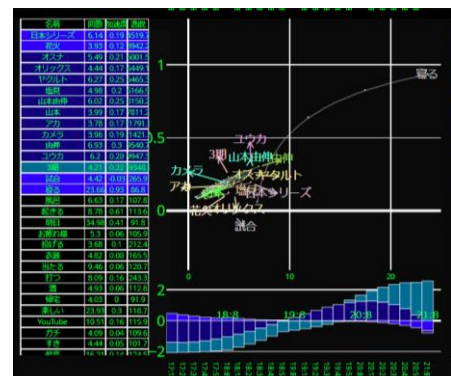


図 4-4-3-1: ここでは並びは、実験 3 と同じく、

Laplacian 単純増加率 Laplacian 大局的増加度
RBF 単純増加率 RBF 大局的増加度
と、なっている。

マイナスの影響を残して単語のみ可視化した場合、図 4-4-3-2 になった。並びは図 4-4-3-1 と同様である。

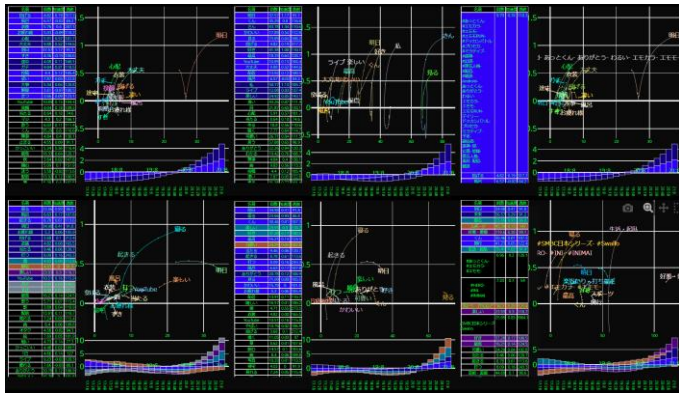


図 4-4-3-2

4-4-4 考察

まず図 4-4-3-1 の左下ブロック (RBF 単純増加率) では、ゆるく回帰して単純増加率を取るという性質上、一気に増加したのがトレンドとして上がりやすく、これを指標として使うことは難しい。図 4-4-3-2 で RBF の単純増加率は RBF の大局的増加度と似ているトレンドが多く、RBF は大局的増加度で使えば良く、単純増加率で使う必要はないと言える。

また、図 4-4-3-1 の右上のブロック (Laplacian の大局的増加度) は厳密に回帰して大局的増加度、つまり増加中を取っているため、単調なグラフになりやすく、可視化の観点から他に Laplacian と RBF の大局的増加度が取れていれば不要と判断した。

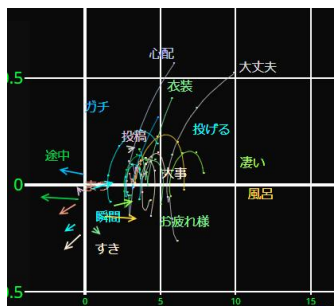


図 4-4-3-3

図 4-4-3-3 は、図 4-4-3-1, 2 の左上 (単純増加率 Laplacian) を拡大した図である。

そのため、図 4-4-3-4 の配置を用いて可視化をする。

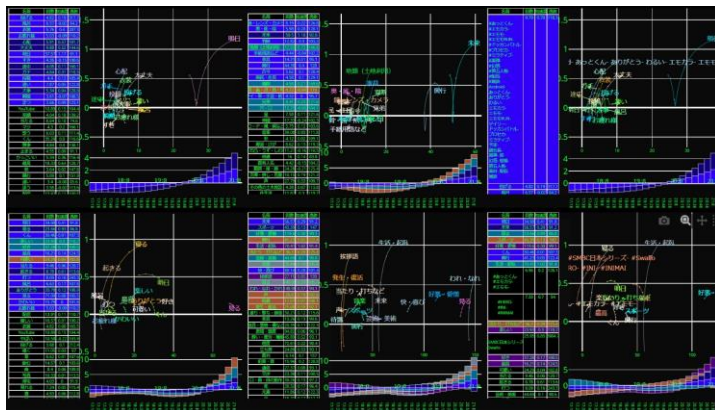


図 4-4-3-4 : トレンド早見図。データは 21 時
どちらも図 2-3 で示したように、左から「単語単体」「意味単体」「5 つのトレンドをまとめた総合」の区分
上段 : Laplacian の単純増加率
下段 : RBF の大局的増加度である。

左を見ると、「明日」「風呂」など、上と下にある単語では、上の方が実際の動きに合っていると思われる。

「明日」は上だと増加しているが、下では増加ペースが鈍っていた。「風呂」は上だと減少に入ったが、下だと微増だった。この可視化手法では上と下で全く同一という訳ではなかった。これはやはり、RBF は幅を取って回帰するために起きると考えられる。

中央下を見ると、順に「未来」「スポーツ」「好悪・愛憎」「興行」「生活・起臥」「当たり・打ちなど」「芸術・美術」「快・喜び」「挨拶語」「見る」が含まれるなど、確かに夜らしいトレンドとなっていた。

逆に中央上を見ると、「鏡・レンズ・カメラ」「奥・底・陰」「地類」「手紙用語」となっていた。一見すると分からないが、何かマイナーな単語に影響している可能性がある。

また、単語を意味に変換する過程で同音異義語ないし複数の意味を持つ単語を取ってしまった可能性もある (できる限り排除はしたもの)。

左下 RBF 大局で現れる「好き」「かわいい」「楽しい」という単語は、意味における「好悪・愛憎」「快・喜び」に現れている。このように、夜になると推しへの愛を語る傾向は毎日確認することができる。

4-5 実験 5 (ツイッタートレンドとの比較実験)

4-5-1 目的 単語を対象に、ツイッタートレンドをどれほど取れているかを検証する

4-5-2 方法 ツレンドと独自トレンドを 10 分ごとに同時に取ってランクをつける

実験日時 : 11 月 16 日 18:30 から翌 15:50

4-5-3 結果

EW	EX	EY	EZ	FA	FB	FC	FD	FE	FF	FG	FH
1:39:23	1:39:23	1:39:23	1:39:23	1:39:23	1:49:23	1:49:23	1:49:23	1:49:23	1:59:23	1:59:23	1:59:23
1682	#JOIAND	#JOIAND	#JOIAND	84	1781	#JOIAND	#JOIAND	90	1683	#JOIAND	#JOIAND
1642	#JOIAND	#JOIAND	#JOIAND	984	1746	#JOIAND	#JOIAND	1146	1629	#JOIAND	#JOIAND
1681	ハッピー	ハッピー	ハッピー	1681	960	#JOIAND	#JOIAND	102	1781	#JOIAND	#JOIAND
426	#JOIAND	#JOIAND	#JOIAND	128	1747	ハッピー	ハッピー	1776		ハッピー	ハッピー
	近況					近況				近況	

図 4-5-3-1 : 「増加率」「ツイッタートレンド」「一致したトレンド」「大局的増加度」の順で取った。

また、上位 50 位以内のトレンドに、独自に抽出した単語・単語共起 (のおよそ 2000 項目) の内いずれかが含まれていた割合は平均すると 22% だった。

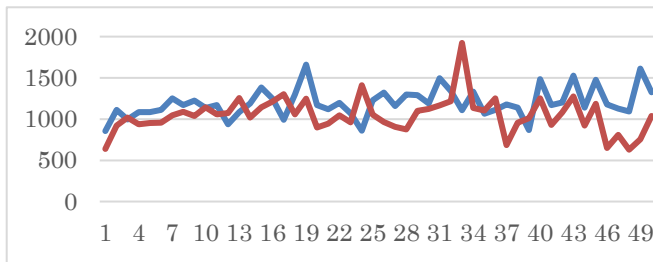


図 4-5-3-2：各ツイッタートレンド順位が平均して独自トレンドで何位に当たるか、青が Laplacian-単純増加率で、赤が RBF-大局的増加度を指している。

4-5-4 考察

図 4-5-3-2 を見ると、Laplacian-単純増加率によるトレンドは、直感と同じく公式下位になるほど、本論文の方式でも低くなる傾向があった。RBF-大局的増加度のトレンドは、10-20 位において高く評価された

すなわち、大局的増加度より単純増加率の方が公式トレンドには近い。

未知語も取得はしているが、ツイッター公式トレンドワードを内包するような単語をそのまま取することは難しかった。逆に言えば日常用語的トレンドを取得すること・可視化が本研究の強みになると言える。

5 結論

5-1 意義

- ・ツイッタートレンドの種類として、意味を用いた抽象化を取り入れた。

- ・大局的トレンドという、単に増加率でない全体を網羅する指標を導入し、より大局的なトレンドを推定した。

- ・勢いデータに基づいた新たな可視化手法を提案し、時系列的増加を視覚的・直感的に捉えられるようになった

5-2 考察

Q. 俯瞰的な視点からの情報、すなわちトレンドに関する研究は少ない。世の中を捉える枠組みを提供することができたか？

単語を更に意味で抽象化し、大局的な増加を取り、トレンド早見盤に短期的・大局的な単語・意味・全ての 6 トレンドごとに 15 グラフを表示して俯瞰的な可視化を実現した。

その上で、ツイッターで取り上げられない単語を取ることができたと言えるだろう。

また、俯瞰的に見ると例えば夜になると推しが好きという文脈で単語が増えるなど、こういった変化は大局的だと言える。もちろん、話題で捉えるべきであって単語や意味で得られるのは一握りだが、それでもツイッタートレンドでない生の視点をトレンド早見盤にできたのは、枠組みを提供したと言えるため有益と言える。

トレンド早見盤に実用性があるのかという点では、トレンド名のラベルが重複する点など見づらいところも多いが、2*3 というブロックを 2*2 にする、ライブラリを変えるなど検討を進めたい。

またトレンドを可視化することで、**ツイッターを通して世界との一体感を感じられる**と考える。

5-3 今後

これまでにない可視化であるため、トレンド早見盤を用いた可視化について、情報を SNS(Twitter や YouTube)で発信するなど、公式と並ぶような存在にしたい。そのために、どこかのサーバー上で常に実行できるようにしたい。

他に取り入れたい機能としては、

- 1, 周期性の考慮。多くの単語には 24 時間の周期性があるため、周期性の有無で単語を分けると特異的な単語を抽出しやすい
- 2, Sentence-BERT を用いて、文脈を考慮した文章類似度に基づいたトレンド。等がある。

本研究は最初に、**ツイート中の複数の単語に共通する意味「潜在意味」**を取ろうとしていたが、上手く取ることができなかった。「共起」も同様だった。トレンドの粒度という観点では単語に落ち着いたが、更に抽象化したトレンドを得ることも目指したい。

参考文献

- [1] 情報社会の〈哲学〉：グーグル・ビッグデータ・人工知能 第二章：ビッグデータの社会哲学的位相
- [2] 小西 敦郎(2019) Twitterトレンドのリツイートによる分析と時系列の可視化
- [3] K.Uchida, F.Toriumi and T.Sakai "Evaluation of Retweet Clustering Method Classification Method Using Retweets on Twitter without Text Data," Proc. WI, pp. 187-194, 2017.
- [4] 須田 剛裕, 小嶋 和徳, 伊藤 慶明, 石亀 昌明, 鳥海 不二夫(2013) 震災時におけるツイッターのトレンドワードと拡散情報を利用したデマ推定の一考察
- [5] 鳥海 不二夫(2021) パースト現象における拡散の定量分析
- [6] 国立国語研究所(2004)『分類語彙表増補改訂版データベース』
- [7] github - 分類語彙表番号-UniDic 語彙素番号対応表 <https://github.com/masayu-a/wlsp2unidic>
- [8] Wikipedia - atan2
- [9] mwklng/rectangle-overlap <https://github.com/mwklng/rectangle-overlap>