

Twitter トレンドの地域差

飯 尾 淳*

Regional Differences in Twitter Trends

Iio Jun

In a recent decade, social network services have been popular among all generations. The trend topics provided by those social network vendors are considered as mirrors reflecting social trends. Twitter, one of the major social network services, provides its trend list obtained from a huge number of tweets, and the Twitter trends have regionality. This study was focusing on its regionality, that is, it investigated the differences between the set of trend keywords from each region. The result of research indicates the definite difference in Twitter trends between that for large cities and the others.

キーワード：Twitter trends, Trends API, Regional differences, Levenshtein distance, Social trends

【目次】

1. はじめに
2. Twitterトレンドとは
3. 分析手法
4. 結 果
5. 関連研究
6. おわりに

1. はじめに

Twitter や Facebook といったソーシャル・ネットワーキング・サービス (Social Networking Services, SNS) が登場して 10 年ほど経過した現在, それらのサービスは人々の生活に十分浸透している. それらが爆発的な速度で普及したため, 当初, マスコミはネットに駆逐されてしまいかとさえ目されていたが, 現時点では, マスメディアとネットメディアは相互に依存しあう

* 中央大学国際情報学部教授

状況を確認することができる。たとえば TV 番組では毎日ネットで話題になった出来事を取り上げ、一方の SNS では TV で報じられた話題や TV 番組そのものについて盛り上がるというような図式ができあがっている。

そのような状況において、Twitter は若者に人気のある SNS の代表的なサービスである。140 文字限定¹⁾のマイクロブログ（ミニブログ）として登場した Twitter は 2006 年のサービス開始以来、長い文章を考えなくてよいという簡便性や半匿名で投稿できる気楽さ、リツイート、フォローする、されるというような仕組みが受けて、現在、全世界で 3 億人を超えるユーザーが利用するまでに成長した。日本の利用者数も、このところは月間アクティブユーザー数（Monthly Active Users, MAU）がおおよそ 4,500 万人と、堅調な利用状況を示している。

ひとりで複数のアカウントを保持していたり、企業アカウントのような法人ユーザーがいたり、MAU そのものは個々人と一対一で結び付けられるものでもないが、日本の人口を考慮するとこの 4,50 万 MAU という数字は社会に対する影響を考慮するうえで無視できない数字であると考えられる。一方で、Twitter には「トレンド」と呼ばれる機能が実装されている（図-1）。トレンドを抽出するアルゴリズムは公開されていないが、その時点で多く言及されたキーワー

図-1 Twitter のトレンド機能



注：右側部分。線による強調は筆者。

1) 2019 年の時点では、日本語、韓国語、中国語では 140 文字の制限が残っているが、英文では 280 文字まで拡大されている。

ドが「トレンド」として抽出されているという原理は想像に難くない。したがって、トレンドそのものは、ユーザーのツイートを紹介している程度は世論を反映しているのではないかという仮説が浮かぶ。

ところで、Twitterが提供するトレンドの一覧は地域に依存しているという特性がある。その最も明確な理由を理解するには、たとえば日米での事情を想定すれば分かりやすい。すなわち、日本では日本語のツイートが大多数を占め、一方で米国では英語によるツイートが大多数を占めるという原則があるため、トレンドとして現れるキーワードも、日本では日本語、米国では英語で表現されるはずだという事情である。それだけではなく、社会的なイベントがユーザーの関心の元となっているのであれば、地域によって語られる内容は変わるはずであり、トレンドにも差が出てくるのではないかということも想定される。

本論文では、Twitterが提供しているトレンドデータが地域によってどの程度の差があるのか、それらにはどのような性質の差が見られるのかについて明らかにすることを試みる。本論文では、第2章でTwitterトレンドについてももう少し詳細に報告する。続く第3章では上記の問いに対してどのような手法で分析を行うかについて説明する。さらに第4章ではその結果を報告する。第5章で本研究に関連する研究事例について述べ、最後に第6章で、まとめと今後の展望について報告する。

2. Twitterトレンドとは

Twitterは自身のサイトで、Twitterが提供しているトレンドの決定について以下のように説明している（Twitter 2020）。

トレンドはどのように決定されますか？

トレンドはアルゴリズムによって決定され、初期設定では、フォローしているアカウント、興味関心、位置情報をもとにカスタマイズされています。ここ数日や今日1日で話題になったトピックではなく、今まさに注目されているトピックが選び出されるため、Twitterで盛り上がっている最新の話題をリアルタイムで見つけることができます。

この説明以外にトレンドの決定方法に関する説明はなく、トレンドの一覧を抽出するアルゴリズムは公開されていない。したがって、本論文ではその原理にまでは踏み込まず、「その地域に在住しているユーザーが多く言及した話題を表すキーワード」としてトレンドを扱うものとする。

また、トレンドが地域に紐付けられていることは、トレンドの表示を明示的に地域毎に切り替えられる機能が提供されていることから明らかである。たとえば、図-1では「日本のトレ

ンド」が一覧として表示されているが、トレンド一覧表示の右上に表示されている歯車のマークをクリックすると、表示するトレンドの地域を変更することもできるようになっている（図-2）。歯車のマークをクリックすることにより図-2に示されている画面が表示され、さらにこの画面で「場所を変更」をクリックすると、トレンドを表示させたい地域を好みの地域に変更することができるようになっている。

後述するように、この Twitter のトレンド一覧は、Twitter 開発者（Twitter developer）登録を行うと、アプリケーション・プログラミング・インタフェース（Application Programming Interface, API）を介してプログラムから取得することができるようになる。この API を用いると、世界中で定義されているトレンドに関する地域の一覧も取得可能である。その機能によれば、日本では静岡および堺を除く政令指定都市と、高松、沖縄、東京、および、日本全域の 22 地域が指定可能である。Twitter トレンドの対象となる地域のうち日本に関係する地域名およびその WOEID（Where on Earth ID）の一覧を表-1 に示す。

図-2 トrend地域の変更



表-1 日本に関係する地域および WOEID 一覧

Kitakyushu	1110809	Saitama	1116753	Chiba	1117034
Fukuoka	1117099	Hamamatsu	1117155	Hiroshima	1117227
Kawasaki	1117502	Kobe	1117545	Kumamoto	1117605
Nagoya	1117817	Niigata	1117881	Sagamihara	1118072
Sapporo	1118108	Sendai	1118129	Takamatsu	1118285
Tokyo	1118370	Yokohama	1118550	Okinawa	2345896
Osaka	15015370	Kyoto	15015372	Okayama	90036018
Japan	23424856				

3. 分析手法

本研究の目的は、Twitterが提供するトレンドの一覧に関して地域差がどのように現れているのかを明らかにすることである。そこで、本研究では、以下の手順によりTwitterトレンドの地域差にどのような傾向を見て取ることができるかを明らかにすることを考えた。

1. Twitter API を利用し地域毎のトレンド情報を取得する
2. 得られたトレンド一覧に関して、データ集合としての距離を計測する
3. 距離の違いにどのような傾向が見出されるかを分析する

以下、順にその詳細について説明する。

3.1. トレンドデータの取得

Twitterが提供するWebのインタフェース (<https://twitter.com/> にアクセスするケース) 以外の方法によって各種データを取得するためには、Twitter 開発者登録を行い、いくつかのパスフレーズを取得しなければならない。<https://developer.twitter.com/> にアクセスし、いくつかの情報を入力、審査を経て、開発者登録を行うことができる。開発者登録を行うと、自分のアプリケーションを構築するための情報を登録できる。それらの情報登録が終わると、自身のアプリケーションとTwitterが連携するためのパスフレーズが提示される。パスフレーズは、コンシューマーAPI鍵 (Consumer API key)、コンシューマーAPI秘密鍵 (Consumer API secret key)、アクセストークン (Access token)、アクセストークン秘密鍵 (Access token secret) の4種類²⁾があり、Twitterとアプリケーションが連携する際には、それらすべての情報の提示が求められる。

これらのパスフレーズを送信することで、Twitter APIの機能を利用することができる。図-3はトレンド一覧を取得するためのAPIを説明している画面である。地域は、Yahoo!が提唱する地域ID (Where on Earth ID, WOEID) を用いて指定する。たとえば、23424856は「日本 (Japan)」を示すWOEIDであり、そのIDを引数に与えてTwitterトレンドAPIにアクセスすると、その時点でのその地域におけるトレンドの一覧を取得することができる (図-4)。

なお、トレンドAPIによるデータの取得に関しては制限があり、15分間で75アクセスまでしか許されていない。ただし、トレンド自体は5分おきに更新されるとアナウンスされており、1回のアクセスでトップ50のトレンドを取得することができるので、この制限が実際には問題になることは少ないであろうと考えられる。

2) 歴史的経緯による。

図-3 トレンド一覧の取得

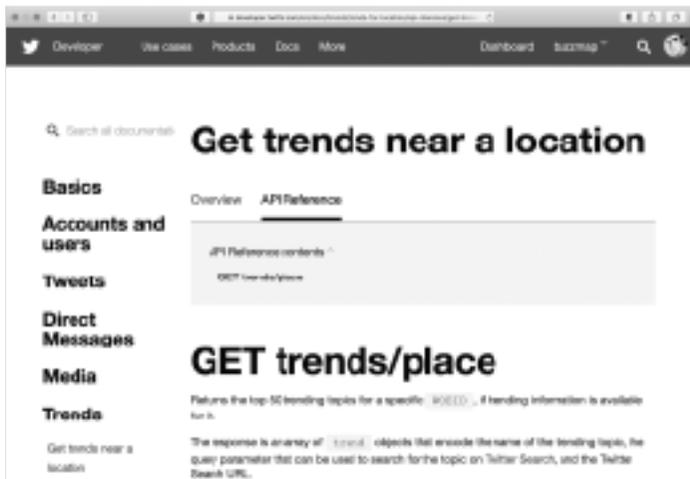


図-4 日本におけるトレンド一覧のプログラムからの取得例



また、トレンドには、一般的なフレーズで表現されるトレンドと、先頭が「#」から始まるトレンドの2種類が存在する。後者はハッシュタグと呼ばれ、フォークソノミー的な方法によりユーザー自身が情報を整理するために、ツイートに恣意的に含められるキーワードがトレンドとして抽出されたものである。トレンドAPIによる一覧の取得においては、「exclude=hashtag」という指定を付け加えることにより、ハッシュタグによるトレンドを除外することができる。

本研究においては、自然発生的なトピックの傾向をトレンドから読み取るという目的との整合性を考慮し、ハッシュタグによるトレンドは除外した。

3.2. 集合間の距離

トレンドの集合が地域によって差があるかどうかを判断する指標として、本研究では編集距離（レーベンシュタイン距離, Levenshtein distance）を応用することとした。集合間の距離を定義する方法はいくつか著名なものが存在するが、今回の場合は、上位のトレンドリストということで、順序も考慮する必要がある。そのため、各トレンドのキーワードを、いわゆる一般的な編集距離でいうところの1文字とし、トレンドの一覧を、一般的な編集距離が取扱う1文と見なすことで、トレンドの一覧A（たとえばある時刻tにおける東京のトレンド）とトレンドの一覧B（同時刻tにおける大阪のトレンド）を比較し、その間の距離を数値で示すことができる考えた。

編集距離とは、AとBの距離を、「追加または削除の処理を何回行くと、AをB（または、BをA）に編集できるか？」で測る³⁾ものである。たとえば、文字列Aとして「そろばん」、文字列Bとして「そばや」を考えたとすると、「そろばん」←（「ろ」を削除）→「そばん」←（「ん」を削除）→「そば」←（「や」を追加）→そばや、となるので、文字列AとBの編集距離は3ということになる。編集距離は、比較する文字列A,Bが同じものとき最小値0をとり、AとBの乖離が大きくなるにつれて編集距離も大きくなるという性質を持つ。

厳密には、文字列 a, b 間の編集距離（lev）は、以下の式で定義される。なお、この式において、添字 i および j は、文字列 a, b 間におけるインデクスである。すなわち、以下の式における $lev_{a,b}(i, j)$ は、文字列 a, b をそれぞれ i 番目および j 番目まで比較した時点での編集距離の値を示す。

$$lev_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1, j) + 1 \\ lev_{a,b}(i, j-1) + 1 \\ lev_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

編集距離には、長さの違いに影響を受けるという問題点がある。前述した「そろばん」と「そばや」の編集距離は3であったが、「東京都渋谷区代々木3丁目」と「東京都渋谷区渋谷3丁目」の編集距離も3である。さらに「千駄ヶ谷」と「谷」の編集距離も3である。これらがすべて同じ編集距離を持つということは、いささか直感に反する。

3) 「置換」の操作も認めて「追加・削除・置換」で測るやり方もある。この場合は、「追加して削除する」の2操作を「置換する」の1操作で置き換えることができるため、編集距離の絶対値は小さくなる。

そこで、長さによる影響を多少なりとも抑えるために、正規の概念を導入し、正規化編集距離 (normalized Levenshtein distance) を考える。正規化編集距離は、通常の手順で求めた編集距離の値を、長い方の文字列の長さで割って求める。正規化編集距離がとり得る値は 0.0 から 1.0 である。比較する文字列が同一であるときに最小値の 0.0 を取り、全く異なるときに最大値の 1.0 をとる。前述の例を正規化すると、「そろばん」と「そばや」の正規化編集距離は 0.75 ($=3/4$)、「千駄ヶ谷」と「谷」の正規化編集距離も 0.75 ($=3/4$)、「東京都渋谷区代々木 3 丁目」と「東京都渋谷区渋谷 3 丁目」の正規化編集距離は 0.25 ($=3/12$) となる。この結果は、単純な編集距離よりは直感的な印象を与える数値となっている。

今回のトレンドリストに関していえば、リストの長さが地域によって大きく変わることはさほどないと考えられるが、若干の差がある可能性も考慮して、正規化編集距離でトレンドリストの地域差を数値指標化することとした。

3.3. 地域差に関する傾向の分析

ここまで述べてきた道具を用いて、地域毎に Twitter トレンドの差異はあるのかに関する調査を実施する。ただし、その範囲は日本に限るものとする。日本と海外のトレンドの差異に関しては、使用する言語が異なるため明らかであり、翻訳も含めて同じ意味的に同じキーワードがどれだけ含まれるかという逆の関連性に関して調査することには若干の意味があると考えられるが、包括的な差分の計測に意味がないことは自明であろう。

同じ日本語によるツイートであっても、地域によってトレンドは異なるのではないかという疑問点に対しては、話題には地域性があるためいくらかの差が出るであろうという仮説が考えられる。たとえば、日本のある地方で事件が発生したとすると、当然、その事件が発生した地域においては、他の地域よりも注目を集める話題になるであろう。また、東日本と西日本という文化的な差異が影響するかもしれない。これらの違いが、Twitter トレンドの違いにどの程度の影響を与えるかが本研究での興味の対象である。

上記の問題に対する仮説を検証するために、実際のデータを集めて分析を行う。トレンドの地域差を求める手順を以下に示す。

1. バッチプログラムを用いて、15 分毎に Twitter トレンド API にアクセスし、地域のトレンドリストを収集する。
2. 収集したトレンドワードのリストに対して、各地域間の正規化編集距離を計算する。その処理により、15 分毎に、各地域間の距離を示した表が作成される。
3. 1. および 2. の手順を 24 時間繰り返す。その結果として作成されたそれぞれの表に関して、時間に関する平均値を計算する。

以上の手順により、地域毎に、他地域との距離の平均値が計算される。それらから、トレンドの地域差を読み取ることが期待できる。本研究においてデータの収集は、2019年10月7日に実施した。

4. 結 果

前章で示した方法により、Twitterトレンドの地域差をどれだけ示すことができたか、また、それらにはどのような特徴を見出すことができたかについて、本節で説明する。

図-5は、地域によるTwitterトレンドを表計算ソフトで並べ、そのスクリーンショットを切り取ったものである。A列には表-1に示した地域名称を、B列以降には、トレンドのリストが並ぶ。順番は、トレンドAPIで取得できた順番⁴⁾である。なお、図-5に示すトレンドのデータは2019年10月6日に取得した。

図-5を見ると、地域差の状況を大まかに確認することができる。すなわち、1行目のChibaから20行目のTakamatsuまでのトレンドは、一連のキーワードがほぼ同じ順序で並んでいる。S列（またはその隣に少しだけ見えているT列）が、一部の地域に関して異なっているだけである。具体的には、Japan, Osaka, Sendaiの3地域に関してS列は「消費税……」となっており、20行目までに示されているその他の地域（Tokyo, Yokohamaを除く地域）のそれが「政治家……」となっている点で異なるのみであり、R列まではすべて一致⁵⁾している。

他方、21行目および22行目に示されているTokyoとYokohamaに関しては、それ以外の地域と大きく異なっており、それら同士も大きく異なっていることが分かる。

図-5 地域によるTwitterトレンドの違いの例

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Chiba	アークAirDro...																		
2	Fukuoka	アークAirDro...																		
3	Hokuriku	アークAirDro...																		
4	Hiroshima	アークAirDro...																		
5	Japan	アークAirDro...	消費税……																	
6	Kanagawa	アークAirDro...	政治家……																	
7	Kanagawa	アークAirDro...	政治家……																	
8	Kobe	アークAirDro...	政治家……																	
9	Kumamoto	アークAirDro...	政治家……																	
10	Kyoto	アークAirDro...	政治家……																	
11	Nagoya	アークAirDro...	政治家……																	
12	Niigata	アークAirDro...	政治家……																	
13	Osagawa	アークAirDro...	政治家……																	
14	Okazaki	アークAirDro...	政治家……																	
15	Osaka	アークAirDro...	消費税……																	
16	Sagamihara	アークAirDro...	政治家……																	
17	Saitama	アークAirDro...	政治家……																	
18	Sagami	アークAirDro...	政治家……																	
19	Sendai	アークAirDro...	消費税……																	
20	Tokushima	アークAirDro...	政治家……																	
21	Tokyo	アークAirDro...	政治家……																	
22	Yokohama	アークAirDro...	政治家……																	

4) APIの説明によればTop 50とあるので、その順番と考えられる。

5) 表ではキーワードの一部しか表示されていないので、図-5に示されている表からでは「すべて一致している」ことは確認できないが、実際のデータは、R列まですべて一致していた。

表-2は、各地域において、他地域と正規化編集距離を計算し、その平均値を示したものである。24時間、15分毎にデータを収集したので、各地域で96(=4*24)件のトレンドデータが収集された。それらを地域毎に総当りで正規化編集距離を求め、求めた96個の値に関して求めた平均値を表-2に並べている。

同じデータの間で定義される距離は0.0であるため、表-2の対角線上には0.0という値が並

表-2 各地域間に関する編集距離の平均値一覧

	Japan	Sapporo	Sendai	Niigata	Saitama	Chiba	Tokyo	Kawasaki	Yokohama	Sagamihara	Hamamatsu
Japan	0.0000	0.0684	0.0443	0.0713	0.0539	0.0703	0.2304	0.0708	0.0352	0.0706	0.0843
Sapporo	—	0.0000	0.0430	0.0475	0.0513	0.0394	0.2630	0.0391	0.0781	0.0419	0.0595
Sendai	—	—	0.0000	0.0566	0.0254	0.0502	0.2384	0.0507	0.0639	0.0526	0.0687
Niigata	—	—	—	0.0000	0.0488	0.0177	0.2657	0.0167	0.0910	0.0088	0.0263
Saitama	—	—	—	—	0.0000	0.0550	0.2482	0.0556	0.0737	0.0575	0.0736
Chiba	—	—	—	—	—	0.0000	0.2638	0.0113	0.0901	0.0120	0.0297
Tokyo	—	—	—	—	—	—	0.0000	0.2649	0.2261	0.2651	0.2769
Kawasaki	—	—	—	—	—	—	—	0.0000	0.0906	0.0110	0.0286
Yokohama	—	—	—	—	—	—	—	—	0.0000	0.0903	0.1046
Sagamihara	—	—	—	—	—	—	—	—	—	0.0000	0.0176
Hamamatsu	—	—	—	—	—	—	—	—	—	—	0.0000

	Nagoya	Kyoto	Osaka	Kobe	Okayama	Hiroshima	Takamatsu	Kitakyushu	Fukuoka	Kumamoto	Okinawa	Average
Japan	0.0519	0.0698	0.0437	0.0748	0.0738	0.0781	0.0670	0.0755	0.0863	0.0644	0.0680	0.07394
Sapporo	0.0259	0.0347	0.0985	0.0438	0.0421	0.0494	0.0318	0.0459	0.0581	0.0352	0.0371	0.05874
Sendai	0.0237	0.0445	0.0754	0.0546	0.0537	0.0602	0.0447	0.0576	0.0694	0.0399	0.0479	0.06025
Niigata	0.0397	0.0487	0.1019	0.0220	0.0174	0.0163	0.0296	0.0137	0.0259	0.0492	0.0115	0.04887
Saitama	0.0286	0.0494	0.0817	0.0595	0.0586	0.0650	0.0504	0.0625	0.0743	0.0448	0.0528	0.06526
Chiba	0.0316	0.0405	0.1022	0.0166	0.0120	0.0196	0.0204	0.0170	0.0291	0.0410	0.0062	0.04646
Tokyo	0.2455	0.2589	0.2643	0.2683	0.2656	0.2692	0.2607	0.2661	0.2766	0.2591	0.2617	0.25898
Kawasaki	0.0321	0.0394	0.1018	0.0156	0.0076	0.0186	0.0193	0.0126	0.0248	0.0416	0.0051	0.04562
Yokohama	0.0717	0.0880	0.0749	0.0945	0.0935	0.0953	0.0866	0.0951	0.1059	0.0857	0.0878	0.09155
Sagamihara	0.0341	0.0430	0.1036	0.0163	0.0117	0.0075	0.0231	0.0080	0.0202	0.0435	0.0059	0.04497
Hamamatsu	0.0501	0.0606	0.1142	0.0322	0.0276	0.0124	0.0407	0.0256	0.0361	0.0611	0.0235	0.05972
Nagoya	0.0000	0.0247	0.0848	0.0361	0.0351	0.0416	0.0250	0.0390	0.0512	0.0194	0.0293	0.04862
Kyoto	—	0.0000	0.1000	0.0450	0.0424	0.0505	0.0322	0.0463	0.0584	0.0341	0.0382	0.05949
Osaka	—	—	0.0000	0.1067	0.1048	0.1096	0.1008	0.1053	0.1184	0.0974	0.0999	0.10428
Kobe	—	—	—	0.0000	0.0079	0.0222	0.0249	0.0212	0.0251	0.0455	0.0104	0.04967
Okayama	—	—	—	—	0.0000	0.0176	0.0223	0.0133	0.0172	0.0445	0.0058	0.04640
Hiroshima	—	—	—	—	—	0.0000	0.0307	0.0156	0.0261	0.0510	0.0134	0.05093
Takamatsu	—	—	—	—	—	—	0.0000	0.0272	0.0393	0.0343	0.0181	0.04901
Kitakyushu	—	—	—	—	—	—	—	0.0000	0.0218	0.0484	0.0108	0.04897
Fukuoka	—	—	—	—	—	—	—	—	0.0000	0.0606	0.0230	0.05941
Kumamoto	—	—	—	—	—	—	—	—	—	0.0000	0.0387	0.05903
Okinawa	—	—	—	—	—	—	—	—	—	—	0.0000	0.04262

ぶ。表-2の2行目、Japanの例で説明すると、1列目のラベルに続き、2列目にJapan-Japan間の距離（の平均値）が記載されたあと、3列目にJapan-Sapporo間の距離（の平均値）、4列目にJapan-Sendai間の距離（の平均値）……というように値が並ぶ。最後の列（一番右の列）は、それらすべての平均値である。すなわち、Japanの行でいえば、SapporoからOkinawaまでの各地域とJapanの間の距離をすべて加算して自らを含まない地域数21で割り平均値を求めたものである。すなわち、その平均値にはJapan-Japanというような自分自身との距離0.0を含んでいない。

なお、表-2は各地域間の距離を示しているため、本来は対角線を軸として線対称となるように対称な値が並ぶ。紙面の都合上、および、見やすさのために、左下半分を表示していないことにも注意されたい。

平均値の欄を見ると、最小値がOkayamaの0.046であり、最大値はTokyoの0.259である。0.1を超えているのはそのTokyoと、Osaka(0.104)の2箇所しかなく、それ以外の地域は他地域との差の平均値はすべて0.1よりも小さい値となっている。なお、その次に大きい値はYokohama(0.092)である。これは、図-5に示した状況と合致する。このように、正規化編集距離を用いてトレンドの集合に関する差を数値化することで、大都市圏とそれ以外という構図が明確になった。

TokyoやOsaka、Yokohamaといった大都市とその他の地域におけるトレンドの傾向が異なる理由として考えられるものは、母集団となるツイートの個数が異なることによる影響である。トレンドキーワードを抽出するアルゴリズムが明らかにされていない以上はあくまで憶測の域を越えることができないが、特定地域のツイート数が大きく異なれば、興味関心が集中する話題の傾向が異なることは十分に考え得る。

一方で、その他の地方都市に関する差分に関しては、特別な差異を見出すことはできなかった。ひとつは、距離の平均値を求めただけでは「特定の差異」を判別できないという手続き上の問題点による。この点については、とくに大きな距離が観測されたようなケースを抽出し、個別のケースを深掘りする必要があったが、今回はそこまで検証することができなかった。また、他の可能性としては、調査したその24時間の範囲では地域差が出るような話題がなかったことも考えられる。今回、データを採取したのは1日24時間だけであったが、より長期のデータを取得して分析する必要もあろう。

以上のように、当初設定した「Twitterで語られる話題には地域性があり、地域によって差が出る」という仮説は、一部については支持されたが完全に支持することはできなかった。不足した点に関しては、今後さらなる研究が望まれるところである。

5. 関連研究

SNS のトレンドに関する研究例は少なくない。我々もこれまで、Twitter トレンドにリストアップされた話題の可視化とその日のトレンドをトピックマップで一目瞭然にするシステムの開発（飯尾 2019, Iio 2019）や、そのシステムを用いた話題の特性に関する分析（金田, 飯尾 2020）、Twitter のつぶやきと感性語の関係の分析（Iio 2019b）などを進めてきた。

Twitter のデータ（ツイート）そのものから Twitter のトレンドを抽出しようという試みもいくつか存在する（Rosa et al. 2014, Dykov & Vorobkalov 2013, Kim et al. 2015, Gaglio et al. 2016）。多くの研究ではツイートを自然言語処理で前処理したうえで、統計的な手法ないしは機械学習の手法を用いてトレンドの抽出を試みている。

Twitter やその他のトレンド情報を用いて、特定のトピックについて分析しようという研究も多い。たとえば、Bian ら（Bian et al. 2016）は、Internet of Things (IoT) の認知状況と消費者の印象を、Google Trends と Twitter の情報を用いて分析した。Bovet ら（Bovet et al. 2016）は、ヒラリー・クリントンとドナルド・トランプが戦った米国大統領選挙の結果を Twitter のトレンドから予想するとの試みに挑戦している。また、Chae ら（Chae & Park 2018）は、企業の社会的責任 (Corporate Social Responsibility, CSR) に関するツイートのトレンドに着目し、Twitter で語られる CSR 関連のトピックと一般的な会話におけるテーマは何か、それらのトピックはどのように相互に関連しているか、さらに、そのようなトピックは時間の経過によってどのように変化したかといった状況を仔細に分析した。これら以外にも、Twitter のトレンドを用いて何らかのテーマ、トピックの構造や認知状況を分析する研究は数多く行われている。

本研究のように、時間的なトレンドの変遷のみならず地理的な分布に関して着目した研究もいくつか存在する（Albishry et al. 2018, Okamoto & Yanai 2019）。ただし、本研究のようなトレンドの地域差に着目したものはなく、その点では、本研究の新規性が認められよう。なお、Okamoto らの研究例のように、位置情報付きのツイートに着目した研究例は他にも存在するが、そもそも位置情報付きのツイートは 1% 程度と極めて少ないことが明らかになっており、現在のところ位置情報を付加する機能は無効とされている（Twitter Support 2019）ため、注意が必要である。

Twitter のトレンドを利用したサービスも、多数、提案されている。日本において日本のトレンドを対象に展開されているサービスには、ついつぶるトレンド | Twitter で【いま】盛り上がりつつある話題は？（<https://tr.twipple.jp/>）、Twittrend（ついつトレンド）- 今、話題になっている各地域の Twitter トレンド（<https://twittrend.jp/>）、トレンド墓地（<https://app.afonomics.com/TrendCemetery/>）などがある。

なお、本研究ではデータ処理に Python と Ruby を用いたが、統計的な分析をする際には R を

用いると効率的である。TwitterのデータをRで処理するためのパッケージが提案されている (Gentry et al. 2016) ので、今後の研究においてはそのようなツールを用いることも検討すべきであろう。

6. おわりに

本研究では、Twitterのトレンドに関する地域差に注目し、同じ日本国内における話題の違いについて地域性があるか否かを検証した。地域差の指標としては正規化編集距離を用い、トレンド一覧の差を定量化し数値で比較できるようにした。

Twitterで指定できる (日本全域を含む) 日本国内22の地域を対象に、TwitterのAPIを利用してデータを収集、分析するプログラムを作成し、15分間隔で24時間、トレンドデータを収集した。さらにそのプログラムにより、各地域、計96個のデータからそれぞれの距離を計算、平均値を求めてその特性を比較した。その結果、東京 (Tokyo) や大阪 (Osaka) といった大都市圏が、他の地域とは異なるという特徴が明らかになった。

他の地域に関しては、残念ながら、特徴的な差異を観測することはできなかった。この点については、より詳細なデータ分析や、さらに長期間、観測を続けることで、特徴的なケースを抽出して分析することが必要である。これらについては、今後の研究課題として残されている。

参考文献

- 飯尾 (2019). Twitterトレンドの分析—SNSでは何の話題で盛り上がるのか?—, 2019年度社会情報学会 (SSI) 学会大会, 232-235, 東京市ヶ谷.
- 金田, 飯尾 (2020). Twitterトレンドの可視化による傾向の分析, 第44回サイバーワールド研究会, CW 2019-09, 東京市ヶ谷.
- Albishry, N., Crick, T., Fagade, T., & Tryfonas, T. (2018, September). Popularity and Geospatial Spread of Trends on Twitter: A Middle Eastern Case Study. In *International Conference on Computational Collective Intelligence* (pp. 167-177). Springer, Cham.
- Bian, J., Yoshigoe, K., Hicks, A., et al. (2016). Mining Twitter to assess the public perception of the "Internet of Things". *PloS one*, 11(7).
- Bovet, A., Morone, F., & Makse, H. A. (2016). Predicting election trends with twitter: Hillary clinton versus donald trump. *arXiv preprint arXiv:1610.01587*.
- Chae, B. K., & Park, E. O. (2018). Corporate social responsibility (CSR): A survey of topics and trends using Twitter data and topic modeling. *Sustainability*, 10(7), 2231.
- Dykov, M. A., & Vorobkalov, P. N. (2013, May). Twitter Trends Detection by Identifying Grammatical Relations. In *the Twenty-Sixth International FLAIRS Conference*.
- Gaglio, S., Re, G. L., & Morana, M. (2016). A framework for real-time Twitter data analysis. *Computer Communications*, 73, 236-242.
- Gentry, J., Gentry, M. J., SQLite, S., & Artistic, R. L. (2016). Package 'twitter'. *R package version*, 1(9).

- Iio, J. (2019a). TWtrends — A Visualization System on Topic Maps Extracted from Twitter Trends, *IADIS International Journal on WWW/Internet*, 17(2), 104-118.
- Iio, J. (2019b). Kawaii in Tweets: What Emotions Does the Word Describe in Social Media? *In Proceedings of the 8th Workshop on Web Services and Social Media (WSSM2019)*, L. Barolli et al. (eds.), *Advances in Networked-based Information Systems, The 22nd International Conference on Network-Based Information Systems (NBIS-2019)*, 715-721, Oita, Japan.
- Kim, D., Kim, D., Hwang, E. et al. (2015). TwitterTrends: a spatio-temporal trend detection and related keywords recommendation scheme. *Multimedia Systems* 21, 73-86. <https://doi.org/10.1007/s00530-013-0342-0>
- Okamoto, K., & Yanai, K. (2019, September). Analyzing regional food trends with geo-tagged twitter food photos. *In 2019 International Conference on Content-Based Multimedia Indexing (CBMI)* (pp. 1-4). IEEE.
- Rosa, H., Carvalho, J. P., & Batista, F. (2014). Detecting a tweet's topic within a large number of Portuguese Twitter trends. *In 3rd Symposium on Languages, Applications and Technologies*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- Twitter (2020). Twitter のトレンドについてのよくある質問, ヘルプセンター, 検索とトレンド, <https://help.twitter.com/ja/using-twitter/twitter-trending-faqs>, 2020 年 3 月 18 日確認.
- Twitter Support [TwitterSupport] (2019, June 19). Most people don't tag their precise location in Tweets, so we're removing this ability to simplify your Tweeting experience. You'll still be able to tag your precise location in Tweets through our updated camera. It's helpful when sharing on-the-ground moments. [TWEET], <https://twitter.com/TwitterSupport/status/1141039841993355264>, 2020 年 3 月 18 日確認.