

長期 Twitter ユーザにおける投稿内容の経時的分析

廣田 一輝[†] 村山 太一[†] 矢田峻太郎[†] KongmengLiew[†] 若宮 翔子[†]
荒牧 英治[†]

[†] 奈良先端科学技術大学院大学 〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: †{hirota.kazuki.hj0,murayama.taichi.mk1,s-yada,liew.kongmeng,wakamiya,aramaki}@is.naist.jp

あらまし Social media は 2000 年代初頭に生まれ、今では誰もが使用するツールとなった。また、中には 10 年以上の長期にわたって運営されているものもあり、人々の日常が蓄積された貴重な情報源である。しかしながら、実際にそのような Social media から人々の変化に焦点を当てて分析した研究はあまりない。したがって、本研究では Social media の 1 つである Twitter を取り上げ、長期 Twitter ユーザにおける投稿内容の経時的分析を行った。また、本研究では時間と共に変化が見られると考えられるユーザの (1) 投稿内容の類似率, (2) クラスタごとのユーザの移動距離, (3) 単語の意味に関する観点から変化を検討する。そして、これらの 3 つの変化に関して、分類モデルを用いた話題の類似率に関する実験、ユーザの移動距離に関する実験、単語の意味的变化に関する実験を行った結果、人々の変化を示唆する結果がいくつか得られた。それは、長期間経つとユーザの投稿内容は 60% 程度一致しなくなり、70% 程度のユーザがクラスタの遷移をみせ、“コロナ” などをはじめとした単語に意味的变化が見られることである。本稿ではそれらについて報告することに加え、3 つ目の実験により得られた意味的变化が見られた単語について考察する。

キーワード Twitter, データマイニング, クラスタ分析, 機械学習

1 はじめに

Social media が生まれて長期間が経過した。特に、Twitter は日本において長年親しまれてきている social media の 1 つであり、そのユーザ数は今もなお増加傾向にある [3]。したがって、そのような Twitter を長期にわたって使用してきたユーザにはその間の社会的な変化や個人的な変化が大きく反映されているのではないかと考えられ、ユーザ数の増加が見られる現状において、今後も人々の変化に関する知見を得ることは重要な課題となってくる。そこで、本研究では Twitter におけるユーザとその投稿文を用いて長期 Twitter ユーザにおける投稿内容の経時的分析を行うことにした。

人々は時間が経つと言及する内容に変化が見られるものであり、それは話題や単語として現れると考えられる。そこで、次の 3 つの変化に着目して分析を行う。

- (1) 投稿内容の類似率
- (2) クラスタごとのユーザの移動距離
- (3) 単語の意味

まず、人々の変化を考えた時、長期間経つとどれくらいユーザの投稿する内容は類似しなくなるのか疑問点として挙げられる。したがって、1 つ目の変化に関する分析ではユーザの過去と過去、過去と現在におけるそれぞれの投稿内容に関する類似率を求めることによって、変化の定量化を試みることにした。また、この実験ではユーザの過去の投稿文から同一人物による投稿であるかを判定する分類モデルの構築を行い、過去と過去、過去と現在における投稿文の 2 つの推論を行うことによって、一般にユーザの投稿文はどれほど一致しなくなるのかを検

証する。

そして、2 つ目に検討するのがクラスタごとのユーザの移動距離に関する分析である。これは、1 つ目の変化をより発展させた分析として捉えることができ、ユーザの過去と現在における変化をユーザを表現するベクトルの移動距離とクラスタごとの分析を試みるものである。より具体的には、ユーザの過去と現在におけるそれぞれの投稿文集合から 2 つのユーザベクトルを作成し、その遷移を可視化、定量化することによってユーザの変化を分析する。

最後に 3 つ目に検討するのが 10 年経つとユーザの投稿する単語自体にも変化が見られるのではないかとという単語の意味に関する分析である。これは、投稿内容やクラスタの遷移と同様、過去と現在におけるユーザの投稿文から時代ごとの単語の埋め込み表現の獲得し、そのベクトルの差異を求め、意味的变化の見られる単語の検知を行う。

2 関連研究

Twitter における投稿内容の分析としては感情分析 [7], [9] などが主流である一方で、経時的な観点に着目して取り組まれた研究はあまり存在しておらず、この 10 年における Social media に対する分析手法をまとめた論文 [2] や政治に焦点を当てて分析を試みた論文 [10] に限られている。

本研究における分析 2 のクラスタの遷移に関する分析と関連する研究としては、動的にクラスタリングを行った研究が存在している。たとえば、Spiliopoulou らの論文 [5] は十数年前に発表され、通時的なクラスタ分析の先駆けとなった研究として位置付けられる。加えて、情報工学の分野から言語に関する変

表 1: データセットの詳細

デーセット	2009's	2020's
全ユーザ数	99,966	54,277
アクティブ率 [%]	54.4	-
取得期間 (初め)	2009-11-03	2020-01-01
取得期間 (終わり)	2010-03-2	2020-11-04
1 人当たりの投稿数 (平均)	2535.7	670
1 日当たりの投稿数 (平均)	13.0	17.8

化を取り出す研究が存在しており、単語の意味的变化に関する分析手法もわずかながら開発されている [6], [8]. そのなかでも特に hamilton らの論文 [1] が単語の意味的变化に関する手法、現象の両観点からまとめられた論文として数多く引用されている。なお、本研究でもそれに則して Twitter から単語の意味的变化に関する検討を分析 3 で行うことにした。

3 データセット

データセットは我々が過去に取得していた 2009 年の Twitter データ (過去) と、そのデータに含まれるユーザを対象に新たに取得し直した 2020 年の Twitter データ (現在) から構成される。なお、2009 年のデータに含まれていた 99,966 人のアカウントを対象として再びツイートをクロウリングし、現在におけるユーザの最近の 1000 件分の投稿データを取得した。表 1 にデータセットの詳細を示す。

再クロウリングを行なった結果、2020 年現在も Twitter を使用しているユーザの率 (アクティブ率) はおおよそ 5 割程度であるということがわかった。加えて、1 日あたりのツイート数¹の数は、過去から現在にかけて僅かながら上昇していることがわかった。これらもまた、一つの時代間における Twitter の変化であるといえ、10 年経つと半分程度のユーザはアカウントを削除、もしくはアカウントが別の人物に移っているかプライベートにしており、使用し続けているユーザに関しては 1 日あたりに投稿する数が増えているということが判明した。

4 分析 1: 投稿内容に関する変化の分析

4.1 分類モデルを用いた話題の類似率に関する分析

まず、1 つ目の投稿内容の変化を分析するために、分類モデルを用いた話題の類似率に関する分析を行う。この分析では、2 つの投稿文が同一人物によって投稿されたものであるか否かを判定する 2 値分類問題を解くことにより、ユーザの過去と現在における投稿内容の変化を分析する。また、そのユーザの過去の投稿文が学習された分類モデルを用いて過去と現在、過去と過去の投稿文のそれぞれの正解率を比較することにより変化を分析する。本実験における概略図を図 1 に示す。

1: 各ユーザの最新の投稿から上位 1000 件を取得するようにクロールを行ったため、単純に表 1 で示した 1 人あたりのツイート数/日数とはなっていない

4.1.1 実験設定

本実験では分類モデル、訓練データ、テストデータは各ユーザごとに作成した。訓練データは、同一人物による投稿文ペアからなる正解データと他人同士による 2 組の投稿文からなる不正解を合わせた計 200 件のデータである。また、正解・不正解ともに投稿文ペアは各ユーザの過去の投稿文から、また他人の投稿文からランダムに抽出された 100 件の投稿文を用いており、それらの正解、不正解データを用いて分類モデルの訓練を行った。また、テストデータでは、各ユーザの過去と過去、過去と現在の投稿文からなる 2 つのテストデータを用意し、それらの正解率の比較を行うことによって投稿内容の変化を分析する。テストデータは訓練データと同様に、同一人物による過去と過去、過去と現在における投稿文ペアを使用した。なお、過去と過去におけるデータには訓練に使用したデータは含まれない。

この実験を全ユーザに対して行い、全ユーザの正解率の平均から一人当たりの過去と過去、過去と現在における正解率 (同一人物であると判定される割合) の平均を求め、10 年における Twitter ユーザの話題内容に関する変化を分析する。

4.1.2 モデルの構築

決定木アルゴリズムを使用して分類モデルを構築した。実装には scikit-learn [11](version 0.22.1) の DecisionTreeClassifier をデフォルト設定で用いた。これは、比較的シンプルな分類モデルを使用することで問題を単純化させるためである。

モデルに入力される 2 つの投稿文はそれぞれ各次元に各単語の頻度が入る Bag of words (BoW) の形式でベクトルを定義し、それら 2 つのベクトルを連結して分類モデルの入力に使用した。前処理としては連続した 3 つの投稿文をまとめ、MeCab によって分かち書きされた名詞のみの単語を取り出した。ここで、3 つの投稿文をまとめたのは、1 つの投稿文だと文字数が少なくなり、内容的なまとまりがなくなることを防ぐためである。

4.2 実験結果と考察

全ユーザの過去と過去、過去と現在における平均推論精度とその標準偏差を表 2 に示す。過去と過去 (2009 と 2009)、過去と現在 (2009 と 2020) における同一人物の投稿文から同一人物と判定される正解率に対する人数の頻度を 0.01 きざみで取ったヒストグラムを図 2 に示す。

これらの結果から、過去と過去におけるテストデータに対しては 70% 程度の精度で正解していた一方で、過去と現在におけるテストデータに対する精度は 30% 程度と半分よりも低くなっていることがわかる。また、それは、図 2 を見ても明らかであり、過去と過去である 2009 と 2009 の頻度分布は 1 で最頻値を取り、0.8 付近で次に頻度が高くなっているが、過去と現在である 2009 と 2020 の頻度分布は 0 で最頻値を取り、分布がなだらかとなっていることがわかる。これは、明らかに分類モデルとしては同一人物の過去の投稿と投稿が同一人物によるものであると判定できていたのに対し、過去と現在における 2 つの投稿が類似しなくなったことによって、精度が低下したのだと考えられる。したがって、ユーザの変化の定量化を試みた結果、ユーザの過去と現在における投稿内容は 10 年経つと平均

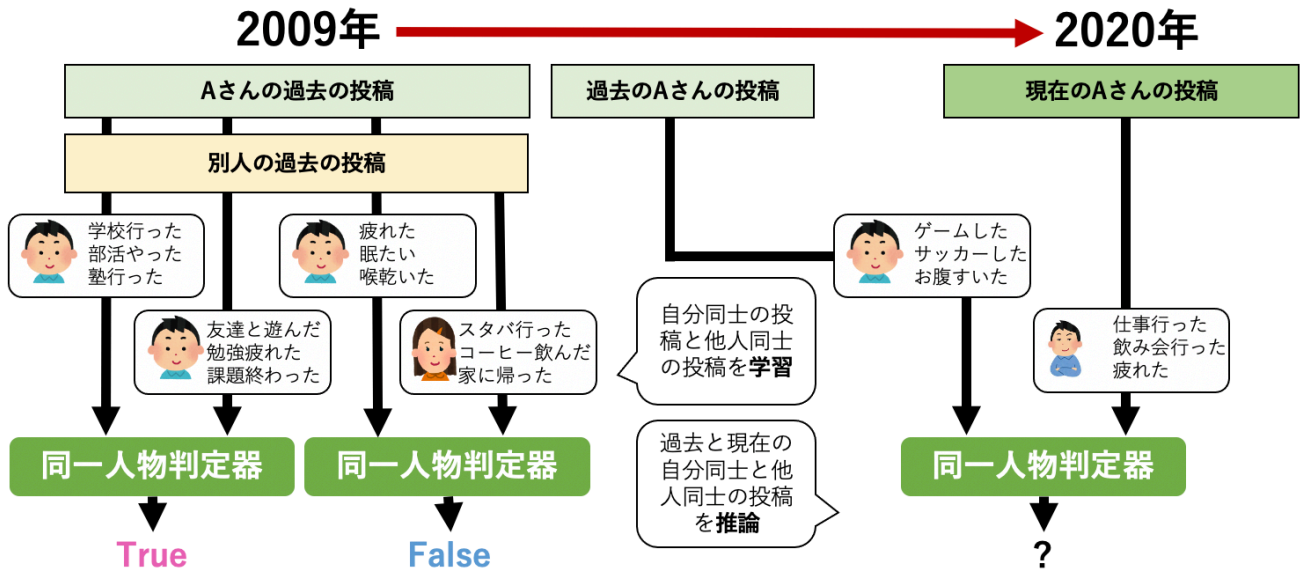


図 1: 投稿内容に関する変化の分析（分析 1）の概略

表 2: 同一人物の投稿文に対する平均推論精度とその標準偏差

精度比較	精度 [%]
過去と過去	74.0 ± 18.9
過去と現在	36.7 ± 30.0

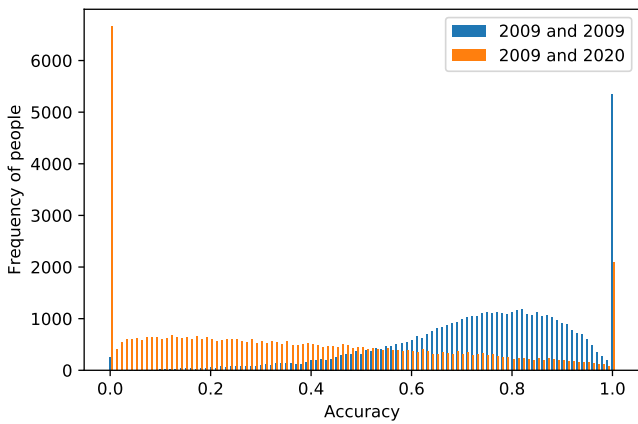


図 2: 過去と過去 (2009 と 2009), 過去と現在 (2009 と 2020) における同一人物の投稿文を用いた同一人物判定の正解率に対する人数の頻度を 0.01 きざみで取ったヒストグラム

値として 63.3%(100-36.7) 程度が類似しなくなり, 過去と現在でおおよそ 37.3%(74.0-36.7) の変化が見られることが判明したと言える。

なお, これほどの変化が見られるということは, 個人が言及する話題にはなんらかの変化が生じているのではないかと考えられる。そこで, 次の分析ではユーザの過去と現在の投稿文から定義される 2 つのユーザベクトルを用いて, 過去と現在にかけてユーザのトピックがどのような遷移を見せているのか, 通時的なクラスタ分析を通して検討した。

5 分析 2: クラスタに関する変化の分析

5.1 ユーザの通時的なクラスタ分析

2 つ目の分析では, ユーザの過去と現在の投稿文から定義される 2 つのユーザベクトルを用いて過去と現在におけるユーザの遷移について分析する。この分析では同じ人物でも過去と現在では別人であると仮定し, 過去と現在の全てのユーザベクトルのクラスタリングを行う。この結果, 過去も現在も関係なく, ユーザの言及するトピックが類似しているユーザベクトルの距離は近くなるはずであり, クラスタごとに変化に違いが見られるかもしれない。つまり, 過去も現在も投稿内容が似ているのであれば同一人物の過去と現在のユーザベクトルの距離は小さくなるはずであり, それが一般的にどれくらいの大きさになるのかを分析することによって, ここでは人々の変化を調査する。なお, クラスタリングを適用するデータは全ユーザにおける過去と現在の全てのユーザベクトルである。

5.2 ユーザベクトルの定義

クラスタリングを適用するユーザベクトルはユーザの過去と現在における投稿文からそれぞれ BoW によって特徴化されたベクトルである。なお, ベクトル化する際, 前処理として絵文字, url, RT つき投稿, 「今日」, 「明日」, 「自分」などといった単語は除いた。これらは, あらゆるクラスタのトピックの上位に来ていた単語であり, トピックとして特に大きな意味合いはないと判断した。

BoW で定義されたユーザベクトルは数万次元のベクトルとなるため, 2 次元に次元削減したユーザベクトルを用いてクラスタリングを行った。なお, 次元削減には, 近年, 比較的性能の高いことが知られている Umap [4] と呼ばれる手法を用いた。また, 次元削減されたユーザベクトルは描画する上で見やすいように z-score による標準化を行なっている。

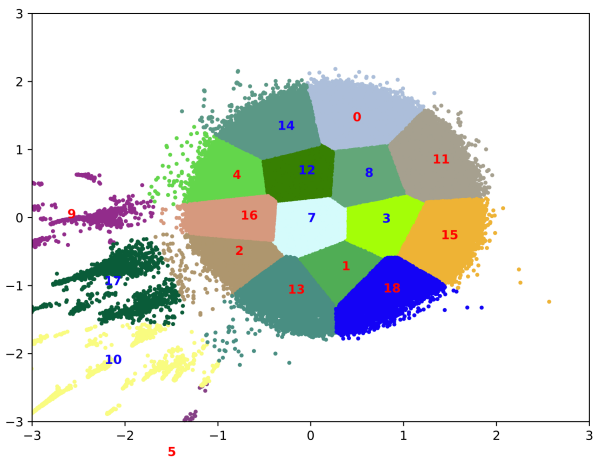


図 3: 過去と現在の全ユーザベクトルに対するクラスタリング結果. 赤字が現在より過去のユーザが多いクラスター, 青字が過去より現在のユーザが多いクラスター

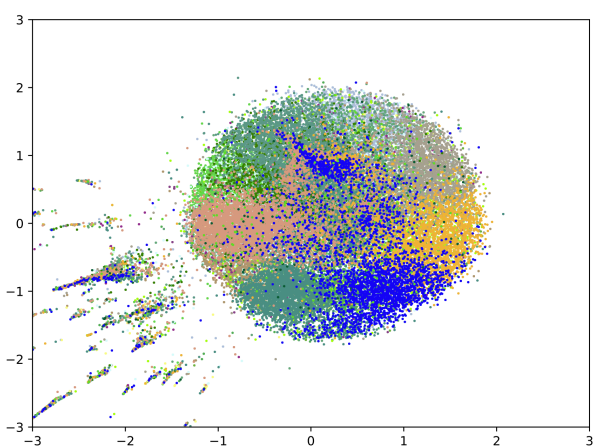


図 4: 過去と現在の全ユーザベクトルに対するクラスタリング結果において, 現在のユーザベクトルに同一人物の過去のクラスターの色を付与した時のクラスタリング結果

5.3 クラスタリング

解釈のしやすさを考慮して, 今回はクラスター数は 20 とした. クラスタリングには, scikit-learn で実装された Kmeans 法の一つである MinibatchKmeans を用いた.

各ユーザの各時代におけるユーザベクトルごとにクラスターラベルを付与し, 各ユーザの過去と現在のユーザベクトルがどのように遷移しているのかを可視化することにより, 人々が言及するトピックの変化を分析した. なお, クラスターごとのトピックを特定するために gensim で実装されたトピックモデルである HdpModel を使用した.

5.4 実験結果

図 3 に過去と現在の全ユーザに対してクラスター数を 20 としたクラスタリングを行った結果を示す. 表 3 に今回注目すべきクラスターごとの特徴と上位 2 つのトピックを示す. 図 4 に現在のユーザベクトルに各ユーザの過去のクラスターの色を付与して可視化した図を示す. また, 図 3 と図 4 は見やすいように描画範囲を -3 から 3 として設定しているが, 描画範囲を超えた領

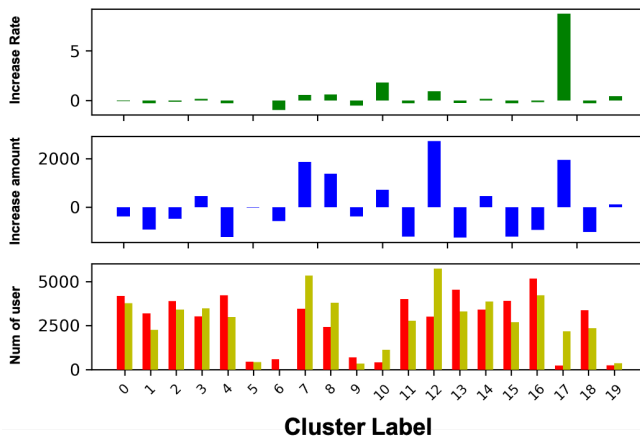


図 5: 各クラスターのユーザの増加率・増加量・時代ごとの所属人数. 赤の棒グラフは 2009 年のユーザ数, 青の棒グラフは 2020 年のユーザ数を示す

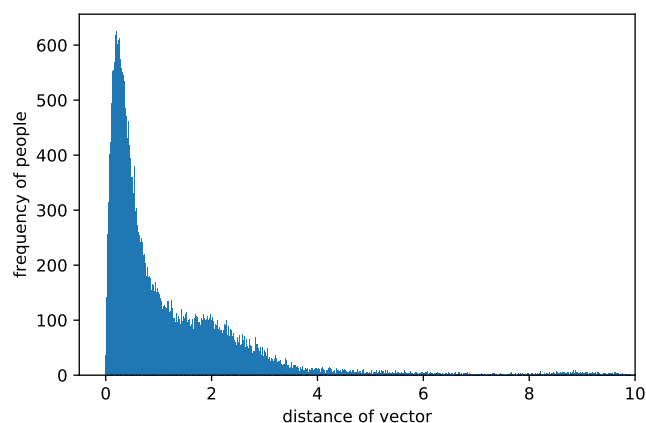


図 6: 0.1 きざみで過去と現在におけるユーザの移動距離に対する人数を取ったヒストグラム. 移動距離が 0.21 のとき最頻値が 601 であった.

域にもいくつか点は存在しており, 20 クラスター全部が図に反映されているわけではない. 図 5 に各クラスターのユーザの増加率・増加量・時代ごとのユーザ数を示す. 図 6 に 0.1 きざみで過去と現在におけるユーザの移動距離に対する人数を取ったヒストグラムを示す. 赤字でラベル付けされたクラスターが過去のユーザが多いクラスター, 青字でラベル付けされたクラスターが現在のユーザが多いクラスターを表している. なお, この過去と現在における人数の差異は, 図 5 の一番下図のクラスターごとの人数差 (棒グラフの差) と対応しており, 過去と現在におけるクラスターの違いを区別するために色分けしている.

5.5 考察

5.4 節の実験結果から時代ごとの異なるトピックを持ったユーザが集まっていること, 大半のユーザはベクトルの移動距離にして少なからず 0.2 程度は移動していること, クラスター毎で変化の仕方に違いが見られることがわかった.

まず初めに, この実験によって時代間におけるトピック的な違いが明らかになったと言える. それは, 図 5 からどのクラスターも過去と現在でそれほどユーザ数に差異は見られないが, 一

表 3: クラスタごとの特徴と上位 2 つのトピック

クラスタ	クラスタの特徴	過去のユーザがもつトピック		現在のユーザがもつトピック	
4	減少量の最も大きい (過去>現在)	トピック 1	$orz*0.00188 + 2010*0.00133 + ちよ*0.0013 + live*0.00127 + つぶやき*0.00127$	トピック 1	$マスク*0.0022 + 2020*0.00218 + めちゃくちゃ*0.00212 + オンライン*0.00211 + 感染*0.00148$
		トピック 2	$orz*0.00233 + live*0.0015 + ちよ*0.00137 + チョコ*0.00103 + つぶやき*0.00101$	トピック 2	$フラン*0.00156 + マスク*0.0014 + amp*0.00137 + 120*0.00118 + 2020*0.00109$
17	増加率が最も高い (過去<現在)	トピック 1	$マイ*0.03809 + 保存*0.03038 + 200*0.02734 + リア*0.02289 + 07*0.02166$	トピック 1	$アイマス*0.02949 + マスター*0.02873 + アイドル*0.0264 + 業務*0.02532 + 02*0.02499$
		トピック 2	$オンライン*0.0042 + kg*0.00377 + 07*0.0037 + 08*0.0035 + 05*0.00291$	トピック 2	$アイドル*0.00648 + マスター*0.00459 + コミック*0.00433 + オンライン*0.00406 + 特典*0.00348$

部, ユーザ数の増加率が極端に高いクラスタやユーザ数の減少量が大きいクラスタが現れていることがわかる. 例えば, 表 3 に示したクラスタ 17 とクラスタ 4 において, それぞれ, 過去に話題となったトピックや現在に話題となっているトピックが多数存在することがわかった. したがって, 各クラスタにはそういった時代ごとの固有のトピックに基づいて似たようなユーザが強く結びついていると考えられる.

次に, 図 4 より, クラスタごとの配置 (色ごとの配置) は大きく変わっていないように見えるが, 実際に他のクラスタに移動した人の割合を求めてみたところ, 約 65% のユーザが異なるクラスタに遷移していることがわかった. これは, あくまでクラスタ数を 20 としたときの他クラスタへの遷移率であるが, 分析 1 で検討した過去と過去, 過去と現在における投稿内容の正解率の変化と同様, やはり多くのユーザは過去から現在にかけて変化しているのだと考えられる.

しかしながら, 今回は時代間のトピック的な変化の違いは見出すことはできたものの, 個人に関する変化傾向は特に見られなかった. この理由としては, 人の変化に関してなんらかの傾向を取り出せることを期待していたが, 個人に関する変化傾向を取り出すには 10 年という期間が少し短すぎた可能性や, 本来, 多様である人々に関して実験に用いたサンプル数が少なかつた可能性がある. また, 今回, クラスタごとにトピックモデルを適用することによって, 政治, 音楽, スポーツといったなんらかの明確なトピックが現れることを期待していたが, どのクラスタでも似たようなトピックが上位にあがってきてしまい, 結果としてクラスタ間の違いを見出すことはできなかった. それは, ユーザが特定のトピックのみを投稿しているわけではなく, 日常的な単語や様々な話題を投稿しているためであると考えられる. したがって, 今後はそういったトピック的な差異が明確に現れるようなベクトルの定義の仕方 (特徴量) を考えなければならない.

加えて, 今回の実験でクラスタ数を 20 としたのに対する妥当性は特になく, クラスタは本来人々の趣味などに基づいて分けられるべきである. そして, 今回は過去と現在に存在しているユーザは別人であり, 全て同一の 2 次元空間に配置されるも

のであるということ暗に仮定して分析を進めてきた. それは, つまり, トピックが 2009 年と 2020 年のみで構成され, 同一人物, また時間における関係性などは考慮していなかったということの意味している. しかしながら, 実際は社会的な影響や日々積み重なる人々の変化に関する制約などがあるべきであり, ユーザベクトルはそれにしたがって定義されるべきである. したがって, 今後の課題としてユーザベクトルの定義の仕方について考えていく必要がある.

最後に, 人々の遷移の傾向を分析するにあたり, 今回はクラスタ間で遷移の多かったものを取り上げて分析してきたが, 条件を絞って分析してみるのも 1 つの手段と考えられる. たとえば, 男性特有のクラスタの遷移や女性特有のクラスタの遷移のように, 属性に基づいた特徴的なクラスタの遷移の分析について今後の課題として検討していきたい.

6 分析 3: 単語の意味に関する変化の分析

6.1 意味的变化の見られる単語の検知

3 つ目の分析では人々の単語の使用状況に関する変化の分析を行う. そのために, 過去と現在で使用されている単語の出現頻度と意味的变化が生じた単語を抽出して分析した.

単語の意味的变化を取り出すにあたり, 全ユーザの過去と現在の投稿文から単語ベクトルの学習を行なった. そして, Diachronic Word Embeddings [1] で提案されている手法を用いて過去と現在における単語ベクトルのそれぞれの空間が揃うような回転を行い単語ベクトルを獲得した.

各単語の差異は, ユーザベクトルの移動距離として, 過去と現在のそれぞれの単語ベクトルを引いたものに L2 ノルムをとることによって求めた. そして, その移動距離を参考にして, 過去と現在で距離の遠いものから順にランク付を行い, 大きく意味的变化が見られる単語からリストアップを行った. また, 過去と現在で意味的变化が大きい単語を分析するため, 出現頻度の差が大きい単語のみを抽出した.

6.2 実験結果と考察

表 4 に過去と現在で単語の出現頻度が大きく変わり意味的変

化が見られた上位 25 の単語を示す。表 4 より、基本的には社会的な影響にしたがって意味的变化が起こったと考えられる単語が多く検知されたことがわかる。なお、これは出現頻度の大きく変わった単語であるため、大半が 2009 年には存在していなかったが新たに価値が付与されたことにより意味的に変化したと考えられる。

例えば、順位が一番高い“メルカリ”は 2009 年にはメルカリ震度階級を意味する単語であったが、現在はオンラインショッピングサービスとしての意味で一般的に使われるようになった。また、昨今話題となっている“コロナ”も大きく意味的变化が見られる単語として検知された。これは、コロナが 2009 年にはビールを意味して使われていたが、現在ではウイルスの意味で一般的に使われるようになったため、意味的变化が見られたのだと考えられる。

さらに、このリストの中でも、“二次元”は明らかに文脈的に変化した単語であると言える。これは移動距離も大きく、かつ、過去と現在で共起頻度の高い語が共通しているため、単純に単語としての印象に変化が生じたと考えられる。したがって、この実験によってユーザーの変化が単語にも強く現れることがわかった。

7 おわりに

本研究では人々の変化を捉えるために、Twitter という大規模な人々に関する情報源を用いて、投稿内容の類似率、クラスターごとのユーザーの移動距離、単語の意味の 3 つの観点から分析を行った。そして、それに関する実験により、10 年経つと (1) 一般的にユーザーの 60% 程度の投稿内容に一致が見られなくなり、(2) クラスタ数を 20 とした時、65% 程度のユーザーがクラスタの遷移をみせ、(3) “コロナ”などをはじめとしたいくつかの単語に意味的变化が見られる、ことなどが判明した。

また、今回の実験を通して、時代背景に基づく様々な変化を取り出すことはできたが、個人に関する経年変化、例えば、あらゆる人における変化傾向を取り出すことはできていない。したがって、今後の課題としては、個人に関する変化をより詳細に分析するために、どうやって社会の影響を切り離して考えられるか、また、2009 年と 2020 年という断片的な変化ではなく、より連続した変化を分析するための方法など、検討していく予定である。

謝 辞

本研究の一部は、JSPS 科研費 JP16H06395, JP16H06399, 19H01118 の支援を受けたものです。

文 献

- [1] Hamilton, William L. et al, “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change”, ACL, pp. 489–1501, 2016.
- [2] Keith Cortis, et al, “Over a Decade of Social Opinion Mining”, arXiv preprint arXiv:2012.03091, 2020.
- [3] statista, “Number of social network users worldwide from 2017 to 2025”, <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- [4] McInnes, et al, “Umap: Uniform manifold approximation and projection for dimension reduction.”, arXiv preprint arXiv:1802.03426, 2018.
- [5] Myra Spiliopoulou, et al, “MONIC - Modeling and monitoring cluster transitions”, DBLP, 2006.
- [6] Liao, Keng-Te, et al, “Explaining Word Embeddings via Disentangled Representation”, ACL, pp. 720–725 2020.
- [7] Kenyon-Dean, et al, “Sentiment Analysis: It’s Complicated!”, ACL, pp. 1886–1895, 2018.
- [8] Carlos Selmo, et al, “Learning language variations in news corpora through differential embeddings”, arXiv preprint arXiv:2011.06949, 2020.
- [9] Hassan SaifHassan Saif, et al, “Semantic Sentiment Analysis of Twitter”, ISWC, 2012.
- [10] Kiran Garimella, et al. “A Long-Term Analysis of Polarization on Twitter”, ICWSM, 2017.
- [11] scikit-learn, “Version 0.22.2.post1”, https://scikit-learn.org/stable/whats_new/v0.22.html#version-0-22-1.

表 4: 出現頻度が大きく変わった単語の中で意味的变化の見える上位 25 の単語

順位	単語	移動距離	現在の共起頻度の高い語	過去の共起頻度の高い語
1	メルカリ	1.479	出品, 販売, フリマアプリ, ヤフオク, 購入, 転売, 発送, 商品, 値段, やつ	震度, 気象庁, 以上, 四川大地震, 確認, あれだけ, 破壊, 綿陽, 荒砥沢ダム, 山体崩壊
2	コントレイル	1.478	菊花賞, 無敗, アリストテレス, サリオス, ヴェルトライゼンデ, 三冠, デアリングタクト, 競馬, 馬券, 達成	きっと, 空気, 予定, 一月, 半ば, 久々, 4 年生, 濃厚, 長谷川, くん
3	帝人	1.476	3401, ファーマ, 訪問看護, 日本経済新聞, 企業, 事業, マスク, 製菓, ソロ, テックス	正臣, くん, 好き, セルティ, 覚醒, ちゃん, 杏里, さま, シズちゃん, デュラ
4	二次元	1.474	三次元, キャラ, 好き, 世界, オタク, アイドル, 美少女, 現実, 情報, エロ	三次元, キャラ, 好き, 規制, 世界, 台詞, アニメ, 体力, 現実, 筋力
4	富岳	1.451	スパコン, 飛沫, 計算, 世界一, PC Watch, 駅名, シミュレーション, マスク, 日本, 理研	出て, 要塞, 新作, 2001 年, 刊行, おなか, いっぱい, 機動力, まごころ, かぼちゃ
5	ロシエツト	1.443	はづき, ちの, 声優, 福山潤	13, 選手, キム・ヨナ, 演技, キムヨナ, 安藤, 真央ちゃん, フィギュア, 浅田真央, 浅田
6	シルシルミシル	1.433	テレビ, 唐突, さっき, グルメ, ロケ, リポーター, AD, くん, 久々, mixi	特集, 天下一品, くん, Google, 番組, 録画, 引き, AD, 天一, ラーメン
7	コミスペ	1.419	レビュー, 日替わり, インタビュー, 漫画, おすすめ, よろしくお願ひします, 作品, 勇者, 桜井, のりお	水戸, E2, 4 階, 22 日, 22, ID, Skype, kurokai, よるほ, in
8	アズレン	1.412	イベント, 10, キャラ, ゲーム, amp, ツイート, 今回, 公式アカウント, 抽選, 生放送	うがい, スプレー, 水溶性, スルホン酸, 炎症, 殺菌, カモミール, ヨード, トローチ, 成分
9	フレボ	1.411	カツ, 礼装, 召喚, 信勝, イベント, FGO, 経験値, 限定, イベ, リリィ	通称, 大好き, 日本語版, 典型, 陸の孤島, 立派, キャンパス, 食堂, 購買, ナチュラル
10	小糸	1.408	ちゃん, 福丸, 雛菜, 樋口, Wo, ノクテル, 円香, 会話, アイドル, 浅倉	座席, 偽装, 強度, 影響, 全日空, 航空, 日航, トヨタ, 旅客機, 問題
11	スティックカム	1.405	配信, ニコ生, オン, 2007 年, Podcast, Live, ログイン, 相互, しょうが, リアルタイム	配信, 終了, 作業, スカイク, チャット, ニコ生, ライブ, 登録, 原稿, pixiv
12	アレクサ	1.404	再生, 反応, 音楽, テレビ, 電気, Amazon, リモコン, Echo, 設定, タイマー	タスク, ナイスク, スタアレ, 大人, ガル, ちゃん, Alexa, ラッド, 美人, あと
13	コロナ	1.401	対策, 影響, 日本, 感染, せい, マスク, 仕事, 今年, ニュース, 時代	ライム, ビール, 好き, ブックス, QT, 太陽, 作家, あと, エントリ, ブログペット
14	ツシマ	1.398	オブ, ゴースト, ゲーム, マルチ, クリア, オンライン, プレイ, PS4, アプデ, 蒙古	大好き, 武器, あと, センス, フリー, 企画, こだな, 新高円寺, どこ, 登録
15	マイクラ	1.398	配信, スマブラ, ゲーム, 参戦, YouTube, 息子, 動画, 肝試し, 最近, ちゃん	マーチ, 日産, 新型, CC, ジュネーブ, デザイン, 披露, コンパクトカー, 日本, QT
16	マイナ	1.395	ポイント, 登録, マイナンバーカード, チャージ, 申請, 還元, 5000 円, 予約, 申し込み, PayPay	ちゃん, ジャンル, 好き, マイナス, 作品, 日本, 映画, ベリカン, ロード, メジャー
17	デカダンス	1.392	アニメ, 2 話, 最終回, ナツメ, 世界, 今期, カブラギ, 展開, 設定, 好き	ミルクチョコレート, ショコラ, DJ, 東京, チョコ, 通りがかり, ゴティバ, コトノハ, チョコレート, 退廃
18	半沢	1.39	大和田, 半沢直樹, ドラマ, 最終回, 録画, ロス, 頭取, 黒崎, 最後, 歌舞伎	FreeTEMPO, くん, 名義, CD, 菅野, ラスト, Free, FreeTempo, 写真, A1
19	ジュイス	1.388	正義, ループ, やり方, アーティファクト, とき, アンディ, アンデラ, ビリー, 過去, わけ	金額, 救世主, Oblige, Noble, sse, 今後, 000, 要望, 受理, 最近
20	カツ	1.386	ノップ, フレボ, 実装, 再臨, 卑弥呼, 礼装, 水着, イベント, ???, FGO	ワイン, リチャード, シュワルツ, HDD, シュバルツ, 社会, ラベル, 問題, ドイツワイン, 黒猫
21	アンナチュラル	1.386	MIU404, ドラマ, アマプラ, 1 話, MIU, 脚本, 中堂, 好き, 石原さとみ, あと	ナチュラル, Vol, 05, 自然体, 番組, 起床, マッチョ, ねま, No, Playing
22	クリスタ	1.384	iPad, 素材, 機能, ペン, PC, 設定, ブラシ, レイヤー, 作業, フォトショ	QT, cristal, filemente, phantzm, haramihitosuji, maxafuku, tagani, iphone, bro, 休み
23	ガシャ	1.382	限定, 無料, SSR, アイドル, 記念, 10, スカウト, SR, フェス, 5 周年	ウィーン, 食玩, ガイアメモリ, メモリ, お嬢様, 樋口, NISSAN, TMS, イメージ, ウィーーン
24	義勇	1.381	鬼滅の刃, 治郎, 好き, しのぶ, キャラ, 富岡, 煉獄, 禰豆, タイプ, あなた	師団, SS, キャラ, オープン, 装甲, S S, 擲弾兵, 連隊, 戦闘, アニメ
25	スブラ	1.379	フェス, ゲーム, 配信, あつ森, チーム, 最近, 久々, Sitch, プレイ, あと	(r y, ちゃん, かー, のりピーちゃん, ポジョレ, TOT, シー, ランド, titter, サービス