

Twitter における市町村関心度の数値化手法

武内奎太^{†1} 新庄雅斗^{†2} 關戸啓人^{†2} 岩崎雅史^{†1}

概要: ソーシャル・ネットワーキング・サービス (SNS) は望みの情報を迅速かつ手軽に入手・共有できることから、世界的に爆発的な広がりを見せている。そのため、企業経営では SNS 上の最適戦略を模索することは必要不可欠といえる。地方自治体の運営においても企業経営と同様、SNS の活用が重要であり、SNS 上の動向に目を向けるべきである。本論文では数ある SNS のうち代表的といえる Twitter に絞り、Twitter 上でどのぐらい注目されているかを市町村ごと明らかにする方法を提案する。提案手法を円滑に実行するために、Google Cloud Platform (GCP) サービスをどのように利用したかについても併せて説明する。さらに、京都府南部の山城地区にある 12 市町村を対象とした分析結果についても報告する。

キーワード: ソーシャル・ネットワーキング・サービス (SNS), Twitter, 市町村, 関心度, Google Cloud Platform (GCP), 京都山城地区

1. はじめに

近年の情報技術の発展は目覚ましく、情報を取得したり発信するための手段の多様化が進み、2023 年現在ではソーシャル・ネットワーキング・サービス (SNS) は多くの人にとって主要な情報源の 1 つと言っても過言ではない。SNS は、登録された利用者間でコミュニケーションできる会員制サービスであり [1], SNS を用いると文章、写真、動画などを投稿・閲覧することができる。SNS の原型と言われている SixDegrees.com が 1997 年にサービスを開始して以来、インターネット通信の高速化やモバイル端末の登場に併せて Facebook, Twitter, Instagram などの SNS が世の中に広く浸透した。日本国内における SNS 利用者は 2022 年時点で 8270 万人に達し、これは日本の総人口約 1.2 億人の約 7 割にあたる数字である [2]。そのため、日本の多くの企業では SNS の活用は非常に重要と考えられ、最近では地方自治体においても例外ではない。

地方自治体では住民の意見や要望を収集して政策に反映させるために、これまで住民アンケートを実施したり意見交換会を開催するなどの取り組みがなされてきた。しかしながら、これらの方法だけでは回答者や参加者の特徴や属性が偏るという欠点があり、特にこれからの社会を担う年齢層の意見や要望を拾い上げづらい点には注意が必要である。このような状況を改善する一手として、住民アンケートや意見交換会に加えて若者の利用率が高い SNS [3] から意見や要望を収集することが有効であると考えられる。一般的に SNS への投稿内容は自治体関係者に依頼されたものではなく自発的なものであるため、自治体に対するより自然な意見や要望が収集されることも併せて期待できる。

SNS への投稿内容から意見や要望を汲み取る以前に、SNS 上ではどの市町村にどの程度の関心が集まっているか

を明確にすることも重要であるため、本論文では市町村ごとに関連投稿数を人口で除した数値を市町村関心度と称し、これを求めるための方法を提案する。市町村関心度の比較によって SNS 上ではどの市町村が注目を集めているかが客観的に明らかとなる。

以降、SNS を Twitter に絞って議論を進める。研究を開始した 2022 年 5 月時点の Twitter では、写真や動画に添えて投稿できるメッセージは、URL などの一部例外を除き全角で 140 文字、半角ならば 280 文字以内の短いものである。一般的に Twitter で投稿されたメッセージはツイートと呼ばれることが多く、本論文でもこれに倣うことにする。Twitter を選択した理由としては、他のユーザーの投稿を共有できるリツイート機能などにより情報の拡散が追いややすく、市町村関心度を求めるという目的に適していると判断したためである。

本論文では、まずツイートデータの収集方法を説明する。どのような技術を組み合わせてこれを実現するかについても併せて説明する。収集されたツイートデータはそのままでは不適切なものも含まれるため、そのようなデータを除く必要があるが、これについては続いて説明する。最後に整理されたツイートデータから求められる Twitter 上での市町村関心度について説明し、実際に京都府南部の山城地区の 12 市町村を対象とした分析結果を報告する。なお、山城地区を構成する 12 市町村は宇治市、城陽市、八幡市、京田辺市、木津川市、久御山町、井手町、宇治田原町、笠置町、和束町、精華町、南山城村である。

本論文の構成は以下の通りである。まず 2 節ではツイートデータを収集するための技術を紹介し、3 節ではツイートデータの収集方法を説明する。続いて 4 節では不適切なツイートデータを除去するための分別法について述べる。5 節では Twitter 上での市町村関心度について説明し、山城

^{†1} 京都府立大学 (連絡先: s822632015@kpu.ac.jp)

^{†2} 大阪成蹊大学

投稿日: 2023 年 7 月 13 日

採択日: 2023 年 9 月 25 日

地区 12 市町村を対象として算出した結果を報告する。最後に 6 節で本論文のまとめを述べる。

2. データ収集に使用した技術

本節では Twitter からツイートを収集するために使用した Google Cloud Platform (GCP) がどのようなものかを説明し、具体的にどのように GCP を用いてツイートを収集したかを述べる。

利用者が非常に多い Twitter を対象とする際にはツイート数は一般的に膨大になるため、そのデータ量の大きさから一般的なデータ処理ツールで扱うことは現実的ではない。いわゆるビッグデータ分析を円滑に扱うには分散処理などの高度な技術や大量のデータを保存できるストレージなど、高価なリソースが必要不可欠となる。しかしながら、研究者自らがこのようなビッグデータを扱えるような環境を用意することは難しい。ツイートデータを収集するために Google が提供するクラウドコンピューティングサービス GCP が有力な選択肢の 1 つである。GCP と類似のサービスはいくつかあるが、そのような中で GCP を選んだ理由については次に述べる。

GCP ではコンピューティングに係わる 100 を超えるリソースをインターネットを介していつでもどこからでも利用でき、それらを組み合わせることで、比較的容易に思い通りのデータ収集が実行できる。GCP リソースの中にはビッグデータ処理に特化したものがあり、BigQuery を用いると膨大なデータの高速処理が可能となる。

GCP で利用できるリソースの保有・管理などはサービス提供者、つまり Google まかせでよく、ユーザの負担増とはならない。そのため、自前でリソースを保有・管理する煩わしさもなく、自前でストレージなどを用意する場合と比較して、計算機環境を整えるためのコストも抑えることができる。GCP の料金体系はサービスを使用した分だけ料金を払う従量課金制であり、一定の使用量に達するまでは無料で使用できる。ちなみに、12 市町村の市町村関心度を分析するため使用した GCP サービスは、無料の範囲内である。

GCP の使用開始の際も、複雑な手順を踏む必要はなく、Google アカウントを所持していれば、標準的なサブスクリプション契約と同じで請求情報などの個人情報をいくつか登録するだけで直ちに使用できる。その他の GCP の使用に関する詳細な情報については、GCP のホームページ[4]を参照されたい。

どのような GCP リソースをどのように組み合わせれば、Twitter からツイートを自動で収集・保存できるかを図 1 に示す。まず、ツイートを収集してそれを保存に適した形式に変換するために Cloud Functions を利用する。Cloud Functions はサーバーやコンテナによる管理なしで、プログ

ラムをクラウド上で実行できるサービスである。実行させたい処理はプログラミング言語の 1 つである Python を用いるとよい。次に、Cloud Functions を用いて収集されたツイートの保存先を BigQuery とする。さらに、Cloud Functions での処理を定期的に自動実行させ新たなツイートを収集するために Pub/Sub, Cloud Scheduler を利用したり、収集されたツイートのバックアップのためにストレージサービスの Cloud Storage も利用する。

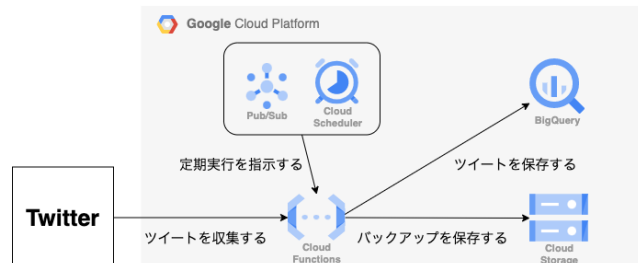


図 1 ツイートデータ収集のイメージ図

3. データ収集方法

本節では GCP サービスを用いて Twitter から京都山城区の 12 市町村に係わると思われるツイートをどのようにして収集したかを説明する。なお、収集されるツイートは係わると思われるものであり、それらのすべてが必ずしも真に係わるものとは限らないことに注意されたい。

ツイートの本文中に「市、町、村を除いた市町村名」、あるいは「先頭に#が付いた市町村名」の少なくともどちらか 1 つが含まれるものを係わると思われるツイートと判定し、2 節で説明した技術を用いて市町村ごとに係わると思われるツイートを収集する。例えば、精華町の場合は「精華」と「#精華町」の少なくともどちらか 1 つが本文中に含まれるツイートを精華町に係わると思われるツイートとして収集する。なお、「精華」が含まれるか否かの判定だけで十分と感ぜられるかもしれないが、「#精華町」が含まれるか否かの判定も必要である。なぜならば、「精華」が含まれるか否かの判定では「精華」、「精華町」が含まれるツイートは収集できるものの、「#精華町」が含まれるツイートは収集できないためである。また、市町村名には漢字が並ぶことからそのままでは中国語で書かれたツイートが含まれてしまうため、「市、町、村を除いた市町村名」が含まれる場合、収集するツイートを日本語で書かれたものに限定する。

これらの処理は、Twitter API (Application Programming Interface) の GET /2/tweets/search/recent を用いて実装できる[5]。GET /2/tweets/search/recent は検索ワード、ユーザーアカウント名、言語や地域などの要素でツイートを検索し、所望のツイートデータを取得する API である。この API を用いて取得できるツイートデータの主な構成は表 1 に示す。

表 1 ツイートデータの主な構成

名前	型	説明
id	文字列	ツイートを識別する一意な識別子
text	文字列	ツイート内容のテキストデータ
created_at	日付 (ISO8601)	ツイートの作成時間

山城地区 12 市町村に係わるとされるツイートの収集期間は 2022 年 6 月 1 日から 2022 年 9 月 8 日までの 100 日間とした。データ収集の頻度については係わりがあると思われるツイートが多い宇治市のみ 1 時間おきに、それ以外の市町村は 1 日おきになるように設定した。

4. ツイートの分別

GCP サービスを用いて収集した市町村名を含むツイートの中には係わりのないツイートが含まれる。よって、正確な市町村関心を数値化するためには、それらのツイートを収集ツイートから取り除く必要がある。本節では、市町村とは係わりのないツイートを取り除く方法について説明する。

例えば、精華町の場合、収集されたツイートの中には「京都精華大学」というワードが本文中に含まれるものが少なくない。京都精華大学関連のツイートが混入してしまうのは本文中に「精華」が含まれるためである。京都精華大学は精華町にあるわけではなく、京都精華大学と精華町は直線距離にして約 35km も離れているため、京都精華大学関連のツイートのほとんどは精華町と係わりがない。

市町村に係わりのないツイートの数は市町村ごとに差はあるが、精華町以外のいずれの市町村でもそのようなツイートが一定数含まれてしまう。収集されたツイートの数は市町村ごとに大きく異なるため、その数が多いか少ないかで不要なツイートを取り除く方法を使い分ける方が得策である。続く 2 つの小節ではこれらの具体的な処理について説明する。以降、簡単のために市町村に係わりのあるツイートを関連ツイート、係わりのないツイートを無関連ツイートと呼ぶことにする。

4.1 収集されたツイートが 10 万件未満の場合

収集されたツイートが 10 万件未満の市町村は宇治市と八幡市を除く 10 市町村であるが、これらについてはツイートの内容を 1 つ 1 つ目視で確認して関連ツイートと無関連ツイートに分別できるレベルである。

作業手順としては、まず BigQuery に保存されたツイートを、1 行に 1 ツイートの内容を収める形式で CSV ファイルに書き出す。次に、その CSV ファイル内に新たに 1 列を追加し、ツイートが関連ツイートならば 1、無関連ツイートならば 0 を記入する。0 が付いた無関連ツイートを除外したものを関連ツイートとして保存する。

市町村との係わりの有無については、ツイートの本文中に含まれる固有名詞が市町村と係わりのあるものか否かで

分別する。市町村との係わりの有無が直ちに分からないような固有名詞については、Google などによる検索結果をもとに判断する。複数人で分別する場合は固有名詞ごとに市町村と係わりがあるものか否かを確認し合い、それらの情報を共有しながら進めるとよい。

このような分別作業は、1 人が 1 日 6 時間で 4000 ツイート程度と考えて差し支えない。まれに 1 つのツイートの分別に 2, 3 分かかるものもあるが、内容的にほとんど同じツイートが圧倒的に多いため、数秒で分別できるものが大半を占める。そのため、1 人が 1 日で 4000 ツイート程度を分別するという設定は容易に達成できるレベルと言える。また、よく知る市町村であれば検索結果をもとに判断する回数が減り、収集されたツイート数が少ない市町村であればツイートに含まれる固有名詞の種類も少なくなるため、1 日に分別できるツイート数はさらに多くなると考えられる。よって、10 万件に近いツイートでも 1 ヶ月あれば 1 人で分別できる分量と考えてよい。データ収集の途中でもツイートデータがある程度収集された段階で分別作業を開始するのがよく、そうすることでデータ収集完了後に分別作業を開始するよりも最終結果を得るまでの日数を短くできる。ちなみに、今回の分別作業はデータ収集と並行して 6 人で少しずつ進めたため、データ収集完了の数日後にはツイートの分別を終えることができた。

4.2 収集されたツイートが 10 万件以上の場合

収集されたツイート数が 10 万件以上となるのは宇治市と八幡市の 2 市である。ツイートのテキストデータに含まれるワードをもとにツイート内容を推測して関連ツイートと無関連ツイートに分別する。具体的な分別の手順については続く 4 つの小節で説明する。

4.2.1 分別用ワードの選定法

最初に行うべきことは、このワードが含まれると関連ツイートと判定する「関連ワード」と、このワードが含まれると無関連ツイートと判定する「無関連ワード」を選定することである。

ワードを選定する際に、予め関連ツイートか無関連ツイートかを分別済みのツイート、つまりテスト用ツイートを複数個用意する必要がある。そのため、収集されたツイートから 4000 ツイートをランダムに抽出し、それらを 1 つ 1 つ目視で関連ツイートと無関連ツイートに分別する。なお、4000 ツイートというのは前述したように 1 人が 1 日で無理なく分別できる分量であり、テスト用ツイートの用意に割くことができる人員や日数に応じて変動させてもよい。テスト用ツイートが用意できたならば、以下に示す 2 つの方法でワードの候補を得る。

1 つ目の方法は、市町村名が含まれるワードに着目してワードの候補を得る方法である。まず、テスト用ツイートのテキストから空白文字を除いた上で、3~10 文字区切りで

テキストを分割する。例えば、ツイートに含まれるテキストが「宇治茶を飲む」の場合、3文字区切りで分割すると「宇治茶」、「を飲む」となる。次に、分割されたワードのうち、「市、町、村を除いた市町村名」が含まれるワードに対して、テスト用ツイートに含まれる関連ツイートすべてと無関連ツイートすべてにおける出現回数からそれぞれの出現割合を算出する。テキスト分割で得られたワード w が関連ツイートで 99 回、無関連ツイートで 1 回現れたならば、関連ツイートにおけるワード w の出現割合は $r_1 = 0.99$ であり、無関連ツイートにおけるワード w の出現割合は $r_0 = 0.01$ である。逆にワード w が関連ツイートで 1 回、無関連ツイートで 99 回現れたならば、それぞれの出現割合は $r_1 = 0.01$, $r_0 = 0.99$ である。これより、 $r_1 > 0.98$ となるワードを関連ワードの候補、 $r_0 > 0.98$ となるワードを無関連ワードの候補とする。

2 つ目は、自然言語処理オープンソースライブラリ GiNZA[6] を用いてテキストデータから固有名詞や時間表現などの固有表現を抽出し、それらの中からワードの候補を得る方法である。まず、GiNZA によってテスト用ツイートから固有表現を抽出する。GiNZA では、固有表現とみなされたワードのほかに固有表現の種類に関する情報も得られるが、これはワードの候補を絞る際には不要であるため用いない。続いて、固有表現として抽出されたワードのうち、複数回抽出されたものに対して、1 つ目の方法と同じように、関連ツイートと無関連ツイートにおける出現割合を算出する。もちろん、 $r_1 > 0.98$ となるワードを関連ワードの候補に、 $r_0 > 0.98$ となるワードを無関連ワードの候補とする。

このような 2 つの方法を用いてワードの候補を得た後に目視で関連ワードと無関連ワードを選定する。目視によるワードの最終選定については、固有名詞が市町村と係わりのあるものか否かを判定するのと同じ要領で行う。

4.2.2 分別用ワードの評価法

関連ワードと無関連ワードの選び方で分別の正確さが変動するため、選定されたワードが適切であるか否かの評価が必要となる。

選定されたワードを評価するには、4.2.1 節で用意したテスト用ツイートを再び用いる。関連ワードと無関連ワードをもとにテスト用ツイートを分別することになるが、その流れについては図 2 に示すとおりである。最初に取り組むべきは関連ワードと無関連ワードが本文中に含まれているか否かの確認である。関連ワードのみが含まれるツイートを関連ツイートに、無関連ワードのみが本文中に含まれるツイートを無関連ツイートに分別する。関連ワードと無関連ワードの両方が含まれるツイートと関連ワードと無関連ワードのどちらも含まれないツイートについてはワードによる分別ができないため、「目視による確認が必要なツイート」とする。

とする。

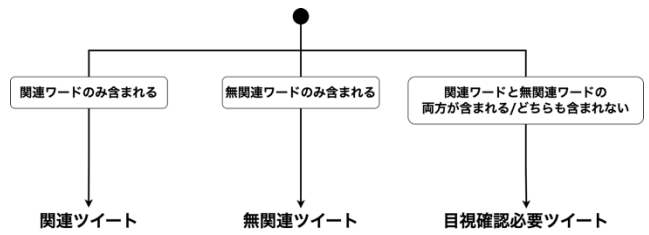


図 2 ワードによる分別の流れ

テスト用ツイートを選定されたワードを用いて分別すると、関連ツイートと分別されたものの中に、目視だけで関連ツイートと無関連ツイートと判定されたものがいくつ含まれるかが分かり、それぞれの個数を N_{11} , N_{10} とすると、関連ツイートの判定精度は $R_1 = N_{11}/(N_{11} + N_{10})$ で求められる。同様に、無関連ツイートと分別されたものの中に目視だけで関連ツイートと判定されたものが N_{01} 個、無関連ツイートと判定されたものが N_{00} 個あるとすると、無関連ツイートの判定精度は $R_0 = N_{00}/(N_{00} + N_{01})$ となる。

山城地区 12 市町村に関する分析では、評価基準として $R_1 > 0.98$, $R_0 > 0.98$ となるように関連ワードと無関連ワードを選んだ。得られた関連ワードと無関連ツイートは付録の表 A.1, A.2 に示すとおりである。

4.2.3 内容ごとにツイートを分類する方法

ツイートが 10 万件以上の場合でもツイート内容を 1 つ 1 つ目視で確認して関連ツイートと無関連ツイートに分別する作業が発生する。複数のツイートの内容が一致する場合、個別に分別するのは明らかに効率が悪い。目視による分別の効率を向上させるために、内容が一致するツイートを 1 つにまとめる方法を提案する。

同一内容のツイートは他のユーザーが投稿したツイートを再度投稿するリツイート機能によるものや BOT によって自動投稿されたものが多い。これらのツイートは、「RT @ユーザーID:」、「https://~」、「@ユーザーID」などの部分を除き、ほぼ同じ文字列で構成される。逆に「RT @ユーザーID:」、「https://~」、「@ユーザーID」の部分はツイート内容を評価する際には重要ではない。よって、ツイート内のテキストに対して以下の手順で処理すると、同一内容のツイートを 1 つにまとめることができる。

1. 「RT @ユーザーID:」、「@ユーザーID」、「https://~」、「空白文字」、「全角/半角数字」をこの順に空文字で置き換える
2. 先頭から 80 文字だけ保存する
3. SHA-256 アルゴリズムを用いて ID を生成する

これらの処理は、BigQuery 上で SQL と呼ばれるクエリ言語を用いて実行できる。ここで、API で取得できる 1 ツイートあたりの文字数が制限されることに注意が必要であ

る。リツイートされたテキストには先頭に「RT@ユーザーID:」が付加されるため、もともとのツイートの内容は API ですべて取得できたとしてもそのリツイートの内容はすべて取得できるとは限らない。文字数が API の取得制限数を超えてしまうと後方から超過分の文字が削除され、その末尾に「...」が付加される。このような場合はもともとのツイートとそのリツイートの内容が同じでも「RT@ユーザーID」の文字数により超過する文字数が異なるため、「...」を単純に空文字に変換するだけでは、「RT@ユーザーID」の文字数分だけ空文字変換後のリツイートの文字列は短くなる。このままでは同一内容のツイートとリツイートが別物と判定される可能性があるため、手順2のようにツイートからリツイートかによらず空文字変換後のテキストの先頭から80文字だけを保存する。同一内容のツイートであれば手順3で生成されたIDが一致するので、同一内容の複数のツイートを1つにまとめることができる。

4.2.4 ワードによるツイートの効率的な分別法

まず、同じIDのツイートを集め、テキストの重複がないように、それらのテキストを結合して一つの長い文字列を作成する。次に、得られた文字列に対して図2の流れで、関連ツイート、無関連ツイート、目視による確認が必要なツイートに分別する。ツイート本文のテキストの結合処理はBigQuery上でSQLを用いると実行できる。宇治市と八幡市に関して、収集ツイートを結合・分別した結果は表2に示すとおりである。ここで、表内の数字はツイート数ではなくIDの個数であることに注意されたい。

関連ツイート、無関連ツイート、目視による確認が必要なツイートをそれぞれ1行に1つのIDの内容を取める形式でCSVファイルに出力する。目視による確認が必要なツイートに関しては、目視による確認で関連ツイートと判定したならば1を、無関連ツイートと判定したならば0を行の最終尾に記入する。この段階で、IDごとに関連ツイートと無関連ツイートのどちらであるかが一目瞭然のCSVファイルが完成する。

表2 分別結果

	関連ツイート	無関連ツイート	目視確認必要 ツイート
宇治市	70,177	12,136	25,365
八幡市	3,932	58,398	24,132

最終的に知りたいのは関連ツイートと無関連ツイートのID数ではなく、関連ツイートと無関連ツイートの数そのものである。そのため、図3のようにIDを通じてCSVファイルの分別結果をツイートに対応付けることで、ツイートごとに関連ツイートと無関連ツイートに分別する。この処理はPythonのプログラムを作成することで容易に実

行できる。収集された全てのツイートデータを関連ツイート、無関連ツイートに分別した後、無関連ツイートを取り除いたものを関連ツイートとして保存する。

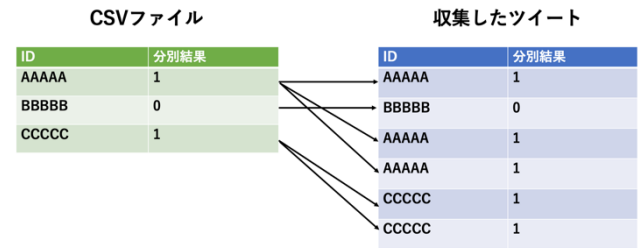


図3 IDごとの分別結果からツイートごとに分別する流れのイメージ

5. 市町村関心度

本節では、関連ツイートに分別された件数から算出される山城地区12市町村に対する市町村関心度について報告する。

Twitter上に市町村名の登場回数が多いほど、その市町村に対する関心が高いと考えるのは妥当である。しかしながら、市町村名の単純な登場回数である関連ツイート数では市町村の規模が全く考慮されていないため、市町村がTwitter上での広報戦略を考える際の直接的な指標にはなり得ない。そこで、1つの関連ツイートが市町村の内外どちらから投稿されたものかは問わず、市町村の1人の住民によるものと見なし、関連ツイート数を人口で除すことで得られる住民1人当たりの平均投稿数を市町村関心度とした。精華町の場合を例に挙げると

$$\frac{\text{精華町の関連ツイート数}}{\text{精華町の人口}}$$

によって精華町のTwitter上での市町村関心度が求まる。市町村の人口についてはWikipediaに掲載の数値を用いた[7]。

表3は、山城地区の12市町村に関して、GCPで収集されたツイート数、関連ツイート数、人口、市町村関心度をまとめたものである。GCPで集められたツイート数は宇治市が最も多く431,448ツイートであり、逆に宇治田原町が最も少なく4,148ツイートであった。関連ツイート数も宇治市が最も多く366,461ツイートであり、逆に井手町が最も少なく1,909ツイートであった。関連ツイート数と人口をもとに算出した市町村関心度は笠置町の6.467が最も高く、逆に最も低い木津川市は0.133であった。笠置町の関連ツイートは笠置キャンプ場絡みのもの、南山城村も関連ツイートは道の駅みなみやましる村絡みのものが支配的であり、宇治市の関連ツイートについては宇治茶や歴史的建造物絡みのものが多く、市町村関心度の高い3市町村に共通するのは観光スポットや特産品が強みという点である。

対して、木津川市は観光スポットや特産品の少ないベッドタウンであるため、市町村関心度が低いと考えられる。

表 3 数値化した結果

対象市町村	ツイート数	関連ツイート数	人口	市町村関心度
精華町	42,425	11,613	35,911	0.323
宇治市	431,448	366,461	177,229	2.068
城陽市	23,367	19,235	73,625	0.261
八幡市	309,065	19,945	69,680	0.286
京田辺市	22,676	21,833	74,331	0.293
木津川市	17,373	10,493	79,010	0.133
久御山町	13,325	12,932	15,014	0.861
井手町	34,973	1,909	7,183	0.266
宇治田原町	4,148	4,134	8,664	0.477
笠置町	19,567	6,823	1,055	6.467
和束町	4,409	4,050	3,313	1.222
南山城村	6,363	6,030	2,311	2.609

6. まとめ

本論文では、GCP サービスを用いて Twitter から山城地区 12 市町村に関連すると考えられるツイートの候補をすべて収集し、その中から本当に関連するものだけに絞る手法を提案した。さらに、Twitter 上での山城地区 12 市町村に対する市町村関心度を算出した。

観光スポットや特産品が市町村関心度に影響を与えるのは明らかとなったが、それら以外の市町村関心度を左右する要因が本論文においてすべて解明できたわけではない。関連ツイートの内容まで分析することで市町村関心度を左右する要因が把握できると思われるが、この点については今後の課題の 1 つである。また、本論文では対象としていない人口 50 万人以上の大都市に対しても提案手法が有効に機能し、本論文と同じように市町村関心度が算出できるかなども検証したい。

謝辞 的確かつ有益なコメントをいただきました査読者に心より御礼申し上げます。

参考文献

[1] “SNS (ソーシャルネットワークワーキングサービス) の仕組み”. https://www.soumu.go.jp/main_sosiki/joho_tsusin/security/basic/service/07.html, (2023年08月31日アクセス).

[2] “2022年度 SNS 利用動向に関する調査”. <https://ict.co.jp/report/20220517-2.html>, (2023年03月29日アクセス).

[3] 「令和3年 情報通信白書」. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/html/nd242120.html>, (2023年03月29日アクセス).

[4] “GCP”. <https://cloud.google.com/?hl=ja>, (2023年05月10日アクセス).

[5] “GET /2/tweets/search/recent”. <https://developer.twitter.com/en/docs/twitter-api/tweets/search/api-reference/get-tweets-search-recent>, (2022年05月20日アクセス).

[6] “GiNZA”. <https://megagonlabs.github.io/ginza/>, (2023年05月26日アクセス).

[7] “Wikipedia 山城地区(南山城地域)”. <https://ja.wikipedia.org/wiki/山城地区>, (2022年10月27日アクセス).

付録 A

表 A.1 関連ワードと無関連ワード (宇治市)

関連ワード	無関連ワード
東宇治, 晴耕雨, 山城, 宇治警察, アルブラ, 宇治しろこ, 宇治手押し, 宇治妙楽, 宇治にまつわる, 宇治の川, プレミアムグラノーラ, パンナコッタ, 宇治行っ, 宇治しのぶ, 宇治ミルク金時, 抹茶ホイップ, 宇治茶 伊藤久右衛門, 和菓子と宇治茶, 宇治ティラミス, 宇治中書島, 宇治ミント抹茶かき氷, 抹茶大福パンケーキ, 宇治市が舞台, 歴史資料館, 宇治探訪, 宇治十帖, 京都 抹茶スイーツ, あじさい, 京都, 久御山, 抹茶グラノーラ, 北宇治中学, 抹茶テリース, 濃い抹茶パフェ, 宇治小倉店, 祇園辻利, 宇治イベント, 抹茶カステラ, 天ヶ瀬, 宇治へ, 宇治 式部, アクトバル, 宇治市文化センター, 宇治市観光大使, 京阪宇治線, 宇治出身, アンサンブルコンテスト編, 宇治の夜景, 宇治のまっちゃん, 夏の宇治, 京都大作戦, 茶壺, 宇治神社, 京の七夕, 煎茶, 木津川, 宇治の老舗, 抹茶ゼリー, 抹茶ドリンク, 小倉町, 宇治市観光協会, 玉露, 宇治市役所, 観光, 京都・宇治, 生八ッ橋, 宇治ミルク, 抹茶の生チョコ, 宇治紫ヶ丘, 宇治近郊, 宇治周辺, #宇治, 京都宇治, 宇治式部, 宇治金時, in 宇治市, いけずな京都旅, in 宇治, 宇治市観光センター, 宇治上神社, 宇治拾遺, トヨタ宇治店, 抹茶カプチーノ, キャラメル抹茶オレ, 濃厚ブラウニー, 宇治へ., JR 宇治駅, 宇治文化, ほうじ茶, 宇治限定, in 宇治市, 宇治へ, 宇治ほうじ, 宇治橋, KBS 京都, 北宇治 0B, 宇治 waoya, 茶畑, 宇治天皇, 抹茶くず餅, 宇治茶レディ, 宇治青年会議, 抹茶共和国, 京	宇治舞, 宇治原 萌, 匿名のメッセージを募集中, 宇治原さん, 宇治原, 宇治原希介, 北宇治高校, 宇治山田博志, 宇治キンちゃん, 北宇治バニー, 宇治野さん, 宇治カズ殿, 北宇治ファイター, 宇治カズ, 北宇治カルテット, 北宇治, 宇治丸, 宇治原史規, 宇治露, 宇治マッサさん, 宇治茶監督, 宇治清高, 宇治達也, 宇治原希, 宇治金時フロスト, 千夜, ヒーヒー!, 宇治鷹, 宇治松さん, 宇治 一茶, 宇治銀時, 宇治紐, みんなからの匿名質問を募集中, 宇治金時さん, 宇治原彩, 宇治郷株, 宇治マッポ, 宇治くん, 宇治特急, 公式吹奏楽コンサート, 北宇治高校吹奏楽部 サマーコンサート, 宇治長次郎, リズと青い鳥, 宇治金時♪, ロザン, 宇治野春, 宇治茶さん, 宇治ちゃん, 宇治海風, 宇治抹茶 TV, 宇治銀時井, 宇治山田博, 宇治工臨, 虹郎宇治, 宇治宮

<p>アニ, 八幡市, 宇治 チョコレート, 宇治園オンラインストア, 紫陽花, ウトロ, 宇治ほうじ茶, 宇治行き, 抹茶プリン, 宇治駐屯地, 萬福寺, 横島, 三室戸, 城陽市, 抹茶ソース, 宇治太陽が丘, 立命館宇治, 宇治でお祭り, ニッキ, お茶っ葉, お茶スイーツ, 抹茶 スイーツ, 宇治市, 宇治徳洲, #宇治, 平等院, かぶせ, 宇治川, 抹茶アイス, 伊藤久右衛門, 天ヶ瀬, 抹茶クリーム, 抹茶パフェ, 宇治公園, 木幡御蔵山, 聖地, 抹茶ジェラート, 宇治抹茶, 抹茶スイーツ, ALCO, JR 奈良線, 宇治駅, 抹茶とうふ, 久世郡, 京阪宇治駅, 宇治徳洲会, 宇治 三室戸寺, 山城総合運動公園 太陽が丘, アクトパル宇治</p>	
---	--

<p>幡, 鳴尾八幡, 八幡堀, 由比ヶ浜結衣, 宮八幡, 八幡たま, 川越八幡, 安久津八幡, 赤羽八幡, 京成佐倉, 八幡社鳥居の龍神様, 八幡橋, 八幡愛, 八幡くん, 八幡イオン, 京成高砂, 大福堂本八幡店, 八幡愛, 渋川八幡, 吉川八幡, 横浜市, れいわ新選組, 諏訪八幡, 都営新宿線, 京成, 白川八幡, 静岡, 市川市, 鎮守八幡宮, 雪ヶ谷八幡, 八幡総合庁舎, 大垣八幡, JR 本八幡駅, 福岡八幡, 青葉区八幡, 牟岐町八幡, 京成本線, 多賀城市, 富岡八幡, 荻野屋 八幡山, 八幡グループ, 山上八幡神, 新潟県新発田市八幡, 滋賀県近江八幡市宮内町, 武田八幡神, 北九州市八幡東, 京成八幡, 八幡神忠隊, 水着八幡, 篠崎八幡, 行田八幡, 鎌形八幡, 八幡撮影所, 八幡和郎, たけべ八幡, 八幡大空襲, 本八幡, 市原市八幡, 北沢八幡, 加賀野八幡, 甲子園, 京成中山, 八幡薙刀, 荒田八幡, 梁川八幡, 南八幡, 僧形八幡, 八幡山, 若宮八幡社, 八幡大菩薩, 川之江八幡, 岐阜, 八幡(やはた)愛, 郡上八幡, 八幡さま, 北澤八幡独演, 大崎八幡, 鶴屋八幡, 八幡の, 静岡八幡店, 三重県, 八幡製鉄, 阿久津八幡, 京成上野, 仙台八幡店, 八幡総合, 百草八幡, 八幡八雲, 八幡さん, 荻野屋八幡山, 高浜八幡神, 八咫八幡, 八幡社旧蹟碑, 雪ノ下陽乃, 皿倉山, 八幡商業, 八幡 和郎, 三嶋八幡, 若宮八幡宮, 八幡茂, 比企谷 八幡, 片田八幡, 地黄八幡の, “八幡”, 仙台市青葉区八幡, 八幡建設, 八幡宇庚田, 宇佐八幡宮, 八幡 PA, 長崎の八幡, 八幡人, 八幡化成, 八幡・伊佐美, 闇之森八幡, 群馬八幡駅, 京成線, 戸越八幡, 滋賀県, 国分八幡, 八幡美咲, 八幡あい, 比企谷小町, 八幡様, 田ノ上八幡, 小野八幡, 八幡餃子, 京成八幡駅, 八幡奈乃花, 北九州八幡, 八幡大樹, JR 総武線・都営新宿線, 富士八幡, 京成小岩, 比企谷八幡, 市原市, 八幡歌</p>
--

表 A.2 関連ワードと無関連ワード (八幡市)

関連ワード	無関連ワード
<p>イズミヤ長岡店, 城陽, 京都の八幡, 飛行神社, 灯りの湯, 京都府, 和束町, 八幡市ユナイテッドフォー, 竹取温泉, 堀口文昭, ラ・ムー八幡, 八幡 JCT, 石清水八幡宮駅, in 八幡市, キングオブキングス八幡店, 久世郡, 京田辺市, 京都京阪バス, 八幡御幸, 八幡市文化, コーナン PRO 八幡, ラムー八幡店, 向日市, 京都府八幡市, 八幡桜まつり, 八幡巻き, 京都八幡高校, 明尊寺, 京田辺, 久御山町役場, 相楽・八幡, 京都府八幡, 八幡(京都, #京都 #八幡, 京都八幡, イズミヤ八幡, 八幡市文化センター, 木津川市, 精華町 #木津川市, #京都#八幡, イズミヤ八幡店, 精華町, 守口手押し, 京都ドッグカフェ, 京都八幡増山事件, アークオアシス八幡, キングオブキングス八幡, 石清水八幡宮, 京都八幡店, 宇治市, in 八幡市, #京都 #八幡, 男山, 八幡市議会議員, スタジアム八幡, コストコ, 京都・八幡, 城陽市</p>	<p>函館, 八幡ツツジ, 八幡神, 下御領八幡, 荒田八幡電停, 比企谷八幡生誕祭, 八幡信仰, 京成津田沼, 若宮八幡, 山上八幡, 発田市八幡, 東京都, 八幡 bot, 八幡舞, 富丘八幡, 八幡ホテル, 八幡営業, 八幡生誕, 滝野川八幡, MLeF 本八幡, 甲子園八幡, 埼玉, 北澤八幡, 桑名八幡郵便局, JR 津田沼駅, 八幡先生, 八幡団地, 高浜八幡, 京成船橋, 八幡みゆき, 到津八幡, 名古屋, 八幡工業, 由比ヶ浜, 八幡家, 八幡奈乃, 八幡丸, JR 本八幡駅, 八幡神社, 京成成田, 蒲生八幡宮, 八幡 entetsu, 八幡の小町, 八幡と, 北九州市八幡東区, 太子堂八幡, 八幡地獄, 八幡太郎, 仙台八幡, 千葉, 石狩市八幡 4 丁目, 八幡インター, 南宮宇佐八幡, 八幡川, 大阪八幡, 雪谷八幡, 中央総武線, 愛知県, 雪ノ下雪乃, 春日部八幡, 函館山, 八幡製鉄政治献金事件, 能登八幡, 八幡氏, 八幡町, 八幡登場, 八幡中央, 香川県観音寺市八幡町, 群馬八幡, 鳩森八幡, 宮城県仙台市青葉区八幡, 寺津八幡, 八幡城, 八幡製鉄所, 蒲田八</p>

	謡祭, 草加八幡店, JR 千葉駅, 八幡 IC, 八幡坂, 银杏岡八幡, 八幡カオ ル, 平山八幡, 大和町八幡, 八幡さ 一ん, 八幡 愛, 八幡下町, 近江八
--	---

	幡, 宇佐八幡, 鶴岡八幡, 飯能市八 幡町, 八幡駅, 比企ヶ谷八幡
--	--

A Quantification Method for Interest Level in Municipality on Twitter

Keita TAKEUCHI^{†1} Masato SHINJO^{†2} Hiroto SEKIDO^{†2}
and Masashi IWASAKI^{†1}

Abstract: Social networking services (SNSs) have been popular worldwide because they allow users to obtain and share desired information quickly and easily. Thus, in recent corporate managements, it is essential to search for optimal strategies on SNS. Similarly to the corporate case, municipalities should utilize SNSs, and pay attention to SNS trends. In this paper, we focus on Twitter, which is a representative SNS among many SNSs, and propose a practical method to identify the interest level on Twitter for each municipality. We also explain how to use Google Cloud Platform (GCP) services to carry out the proposed method smoothly. We describe the analysis results of 12 municipalities in the Yamashiro area in the southern part of Kyoto Prefecture.

Keywords: Social Networking Service (SNS), Twitter, Municipality, Interest level, Google Cloud Platform (GCP), Kyoto Yamashiro area

^{†1} Graduate School of Kyoto Prefectural University (Correspondence Author:s822632015@kpu.ac.jp)
^{†2} Osaka Seikei University