

ソーシャルネットワークにおける 共通の友人に着目した実世界イベント分類手法

佐藤 翔野^{1,a)} 米澤 拓郎^{1,b)} 河野 慎^{1,c)} 中澤 仁^{1,2,d)} 川崎 仁嗣^{3,e)}
太田 賢^{3,f)} 稲村 浩^{3,g)} 徳田 英幸^{1,2,h)}

受付日 2015年12月20日, 採録日 2016年7月5日

概要: 近年, ソーシャルネットワーク上の情報を使い実世界のイベントを検出・分類する研究がさかに行われている. 既存の研究ではソーシャルネットワーク上に投稿された発言のテキスト解析が主であり, 発言にイベントに関する内容が含まれていないと解析が困難であるという問題が存在する. 本研究では, Twitter 上の位置情報付き発言を用いて, ある空間に存在する Twitter ユーザの興味を発言のテキスト解析を行うことなく抽出し, それに基づいた実世界イベントの検出と分類を試みる. 提案手法では, 空間内のユーザ群の「共通の友人」を分析し, その友人の属性情報を Wikipedia を参照して解析することにより, イベントの属性を抽出する. 提案手法により自動抽出されたイベント分類と, 27 名の被験者によるイベント分類とを比較した結果, 趣味的の高いイベントに関して高い類似性を持ったイベント分類が可能であることが分かった.

キーワード: ソーシャルネットワーク, 位置情報サービス, 実世界イベント検知

Classifying Urban Events by Analyzing Common Friends in Location-based Social Network

SHOYA SATO^{1,a)} TAKURO YONEZAWA^{1,b)} MAKOTO KAWANO^{1,c)} JIN NAKAZAWA^{1,2,d)}
HITOSHI KAWASAKI^{3,e)} KEN OOTA^{3,f)} HIROSHI INAMURA^{3,g)} HIDEYUKI TOKUDA^{1,2,h)}

Received: December 20, 2015, Accepted: July 5, 2016

Abstract: Recently, many researchers focus to detect and classify urban events by analyzing information on social network. Previous work mainly use text analysis of users' posts on social networks for detecting urban events; however, this approach has a limitation that users' posts must mention event information. We develop a new method to detect and classify urban events by extracting users' interests from location-based social network information without using text analysis. Our method analyses common friends in users who exist in the area of on-going events, and extract common friends' attributes from related Wikipedia information. We designed and implemented the proposed method, and carried out experiment for evaluating our method. Our experimental result shows that our method can classify events, where participants have similar interests, with high similarity by compared with ground truth created by questionnaire.

Keywords: Social networks, Location based services, Real world detection

¹ 慶應義塾大学大学院政策メディア研究科
Graduate School of Media and Governance, Keio University,
Fujisawa, Kanagawa 252-0082, Japan
² 慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University,
Fujisawa, Kanagawa 252-0082, Japan
³ 株式会社 NTT ドコモ
NTT DoCoMo, Inc., Yokosuka, Kanagawa 239-8536, Japan
a) shoya@ht.sfc.keio.ac.jp
b) takuro@ht.sfc.keio.ac.jp
c) makora@ht.sfc.keio.ac.jp

1. はじめに

スマートフォンやソーシャルネットワーキングサービス (SNS) の普及により, 多くのユーザが GPS の搭載された

d) jin@ht.sfc.keio.ac.jp
e) satoshi.kawasaki@nttdocomo.com
f) ootaken@nttdocomo.com
g) inamura@nttdocomo.com
h) hxt@ht.sfc.keio.ac.jp

デバイスを持有し、様々な時間や場所で情報へのアクセスおよび発信が可能となった。SNS はインターネットを介して個人間のコミュニケーションを促す社会的要素を含んだメディアであり、日本における 2014 年 9 月時点での主な SNS の利用率は、57.1%^{*1}と過半数を超えている。このような技術の発達および普及により、多くのユーザが GPS を用いた位置情報と紐付いた実世界の情報をリアルタイムに SNS 上に投稿することが可能になっている。本研究では、このような SNS 上の投稿に位置情報が付加されたもののことを位置情報付き発言と呼ぶ。位置情報付き発言は、その場所に起因する実世界上の出来事（実世界イベント、以下イベントと呼ぶ）について投稿されることが多い。そのため、位置情報付き発言を収集し、解析することにより情報空間から実世界のイベントを検出・分類する研究がさかんに行われている [1], [2], [8]。SNS からリアルタイムに実世界イベントを検出・分類可能とすることで、イベント推薦や交通ナビゲーションシステムなどへの応用が期待されている。

SNS 上の位置情報付き発言を使用してイベント検出・分類するための手段としては、主にテキスト解析が用いられる場合が多い。しかし、位置情報付き発言には必ずしもそのイベントに関連するテキストが含まれているとは限らない。このような発言のことを本研究ではノイズと呼ぶ。発言内容がノイズのみである場合、テキスト解析ではイベントの検出は不可能になってしまう。既存研究では話題トピック検出やバースト検出を試みている [7], [9] が、ノイズは考慮されていないため、検出可能なイベントは位置情報付き発言が多数収集可能な大規模イベントに限られている。またユーザの位置情報付き発言の空間的・時間的履歴からイベント検出を試みる研究 [10], [12], [13] も存在するが、これらの手法はイベントの検出こそ可能であるが、具体的なイベントの分類などは位置情報だけでは難しい。これらの問題を解決するため、本研究では、テキスト解析および位置情報の履歴解析のアプローチではなく、空間内のユーザ群の「共通の友人」に着目したイベント検出・分類手法を提案する。具体的には、Twitter を用いて位置情報付き発言を収集し、特定の空間内に存在しているユーザ群の共通のフォロイ（友人）のランキングを作成する。ランキングに登場する共通のフォロイの多くは、多くのフォロワを有するアカウントであることが想定されるため、そのアカウントの属性に関連する Wikipedia から抽出し、解析した結果をイベントの属性として用いる。本研究では提案した手法に基づきイベントの分類を可能とするインタラクティブな Web ツールの設計・実装を行った。提案手法により自動抽出されたイベント分類と、27 名の被験者によ

るイベント分類とを比較した結果、趣味性の高いイベントに関して高い類似性を持ったイベント分類が可能であることが分かった。

本稿の構成は以下のとおりである。まず 2 章で位置情報付き発言を用いた実世界イベント検出・分類についての問題意識を説明し、関連研究を紹介した後、本研究の目的を述べる。3 章でイベント参加者の共通の友人を解析することで、イベントの属性を推定する手法について述べる。4 章では、本手法を用いた解析ツールの設計と実装について述べ、5 章でそのツールを用いた解析結果と人によるタグクラウドと Wikipedia によるタグクラウドを比較し、考察を述べる。6 章で今後の課題について述べ、最後に 7 章で本研究のまとめと今後の展望を述べる。

2. 位置情報付き発言を用いた実世界イベントの検出と分類

本章では、まず問題意識を述べた後、本研究の関連研究について述べる。その後、本研究の目的について述べる。

2.1 問題意識

Twitter 情報を活用した実世界イベントの検出・分類手法が多く提案されている [1]。都市で発生するイベントは、コンサートなどあらかじめ計画されたイベントに加え、事故やサッカー W 杯後の渋谷のスクランブル交差点でのお祭り騒ぎなど突発的なイベントが含まれる。前者のイベントは Web 検索などによっても発見・分類することは可能であるが、後者のイベントなども含めたうえでリアルタイムに都市イベントの検出・分類を行うことで、イベント推薦や近隣住民・店舗への喚起、行政の都市マネジメントなどのアプリケーションが構築可能となる。これまで提案されている実世界イベント検出・分類手法は、大きく分けて (1) 位置情報付き発言内容のテキスト解析 [2], [7], (2) 位置情報の空間的・時間的解析 [10], [11], の 2 種類に分類される。両者の手法に共通している大前提として、実世界イベントが発生している空間で位置情報付き発言が存在することがあげられるが、それに加え、それぞれの手法には追加の前提や、それにとまなう利点・欠点が存在する。

位置情報付き発言内容のテキスト解析手法では、発言内容を解析し、実世界イベントの名称や規模の推測が行われる。同手法ではイベントの検知および分類が可能となっているが、追加の前提として、発言内容にイベントに関する内容が含まれる必要がある。しかし、位置情報付き発言には必ずしもそのイベントに関連するテキストが発言されているとは限らない。たとえば音楽のイベントに参加しているユーザであっても「このあとなに食べよう」のように、イベントに関するテキストを発言していない場合も多く存在する。このようなイベントに関係のない発言を本研究ではノイズと呼ぶ。ノイズを含んだ発言をそのまま利用して

^{*1} 総務省 http://www.soumu.go.jp/iicp/chousakenkyu/data/research/survey/telecom/2014/h25mediariyou_2summary.pdf

表 1 実世界イベント検出・分類の2種類の手法

Table 1 Comparison of two methods for urban event detection and classification.

	位置情報付き発言の テキスト解析	位置情報の空間的・ 時間的解析
前提	・位置情報付き発言 が存在する ・発言内容がイベン トに言及している	・位置情報付き発言 が存在する
イベント検 出	○	○
イベント分 類	○	×

しまうと判定結果にもノイズが多くなってしまいます。また、発言がノイズしか含まれていない場合、イベント検知・分類は不可能となる。

一方、位置情報の空間的・時間的解析では発言内容は考慮せず、ある空間内の発言数の時間的遷移などを用いてイベント検出を行っている。この手法では発言内容は用いないため、前述したノイズの影響を受けることなく、イベントの検出が可能となる。一方で、位置情報の遷移だけでは、具体的なイベントの名称・属性などの分類は難しい。よって、同手法ではイベントの検知は可能であるが、実際にそのイベントがどのようなものであるか推測するためには、結局発言内容の分析を行う必要がある。

これら2つの手法の特徴をまとめたものを、表1に示す。イベントの推薦などのアプリケーションを構築するためには、イベントの検知だけでなく、イベントの分類が重要となる。位置情報の空間的・時間的解析では必要となる前提は1つでよいが、イベントの具体的な分類は難しい。一方、テキスト解析の手法では、発言内容がイベントに言及しているという前提が必要であるが、Hirutaらによると、ある空間に関係のある発言は位置情報付き発言全体の55.93%にすぎないという報告[6]もあり、分類可能なイベントは大幅に限られることが想定される。よって、イベントの検知・分類を少ない前提条件で行うためには、既存手法とは異なるアプローチが必要である。

2.2 関連研究

2.2.1 位置情報付き発言内容のテキスト解析

位置情報付き発言のテキスト解析をする既存研究として、Ishikawaら[7]は、ある地域のある時間において発生するホットトピックの検出のため、TwitterとWikipediaを使って作成した意味的辞書と位置情報付き発言を用いてバースト検出を行い、ホットトピックを検出している。また、金子ら[9]は位置情報付き発言のうち、画像を添付して投稿されている発言に注目し、特定の地域での投稿をリアルタイムに監視し前日と比べ出現頻度が上昇したキーワー

ドを含む位置情報付き発言の画像の解析を行うことにより、イベントの特徴的な画像を抽出している。つまりこれらの研究はイベント検出・分類に成功しているといえるが、ノイズとなる位置情報付き発言を考慮していない。そのため、イベントの判定結果もノイズが多くなってしまう。また発言がノイズだけであった場合、分析をすることすら不可能になる。一方、Marcusら[15]はあるイベントが進行している際に、そのイベントに関するホットトピックがどう変遷しているかについて、ツイート内容をもとに分析・可視化するシステムを構築している。たとえば、あるサッカーの試合に関して、football, soccer, premierleagueなどのキーワードをもとに検索されたツイート内容を分析し、ホットトピックの解析を行っている。同様にNicholsら[16]は、あるイベントのサマリをその内容や盛り上がりをもとにツイート内容を分析することで自動生成する手法を提案している。これらの手法は解析対象のイベントがあらかじめ決定され、特定のキーワードをもとに十分な量のツイートが検索可能であれば有効に動作するが、本研究が対象としているようにあらかじめ何のイベントが行われているか分からない場合には、利用が難しい。

2.2.2 位置情報の空間的・時間的解析

位置情報の空間的・時間的解析をする既存研究として、Leeら[10]は、位置情報付き発言の増加率によるイベント検出の手法を提案している。特定の空間内の通常時の位置情報付き発言の数を推測し、通常時よりも位置情報付き発言が多く投稿されたときに、イベントが発生したとしている。また藤坂ら[13]はあるイベントへ参加していると推測されるユーザたちの移動履歴を分析することで、そのイベントの影響範囲を推定している。さらに藤坂らはイベントの詳細を知るためにテキスト解析を行っているが、実際ノイズを多く含んでおりイベントに関するキーワードをうまく抽出できていない。このように、位置情報の遷移だけではイベントの検出こそは可能だが、イベントの属性などの分類は難しい。また、詳細を知るために発言内容の解析を行うとノイズの影響を受けてしまう。

2.2.3 その他のアプローチ

河野ら[14]はイベント参加者の多様性を意味する「大衆性」を定義し、位置情報付き発言を行っているイベント参加者がフォローしているユーザを解析することにより「大衆性」に基づきイベントの分類する手法を提案している。彼らはテキストを解析せずにイベントを分類しているが、分類指標としてイベントが万人受するものかという「大衆性」のみとなっており、詳細なイベントの属性までは解析できていない。

2.3 目的

以上のことより、イベントの検知・分類を少ない前提条件で行うためには、既存手法とは異なるアプローチが必要

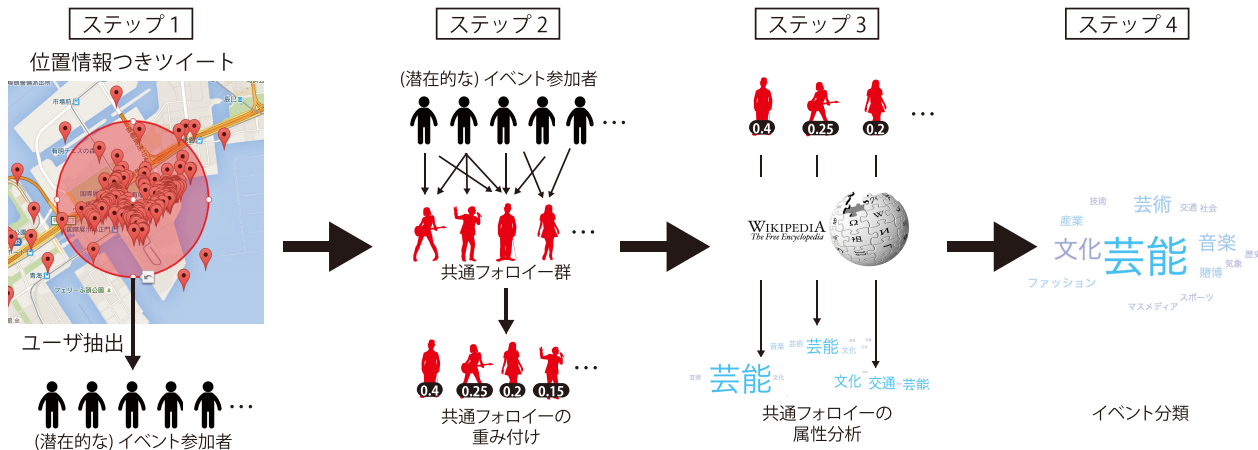


図 1 提案手法の概要

Fig. 1 Overview of proposed method.

である。したがって本研究の目的は、位置情報付き発言を用いて少ない前提条件でイベントの検出・分類する手法を実現することである。

3. 共通の友人分析によるイベント判別手法

この章では、共通の友人からイベントを分類する手法について述べる。まずは本手法の概要を述べたあと、共通の友人の抽出方法、イベントの属性推定手法について述べる。以下は、Twitter を対象とした実験・議論を行うため、共通の友人を共通のフォロイと呼ぶ。

3.1 概要

本研究の目的を達成するために、本稿では空間内に存在するユーザ群の共通のフォロイを抽出・解析し、そのユーザ群の興味を推定し、イベントの属性として用いる手法を提案する。イベントに参加しているユーザは、そのイベントに関する話題に興味を持っており、それは Twitter 上のフォロイに反映されている可能性が高いと考えられる。たとえば事前調査として、「入間基地航空祭」が行われている場所で Twitter の位置情報付き発言を行っていたユーザの共通のフォロイを抽出したところ、「防衛省 航空自衛隊」や「防衛省 海上自衛隊」、「陸上自衛隊」のように軍事に関するアカウントが多くフォローされているという結果が得られた。すなわち、「入間基地航空祭」イベントに参加しているユーザは、常日頃から軍事に興味を持っており、軍事に関する著名なアカウントをフォローしていると考えられる。また複数のアイドルのコンサートが行われている場所での Twitter の位置情報付き発言を行っていたユーザの共通のフォロイを抽出したところ、同様にそのアイドルに関するアカウントが多くフォローされていた。この情報を用い、本研究ではある空間内のユーザ群が持つ共通のフォロイがどのような属性なのかを解析することにより、イベントに参加しているユーザ群の興味として用い、それをその

場所のイベントの属性情報として利用可能とする手法を構築する。本手法を用いることで位置情報付き発言自体のテキスト解析を行う必要がないため、位置情報付き発言が存在する、という仮定だけでイベントの分類が可能となる。

また、ある特定の空間における属性情報を定点観測すれば、その属性の変化からその空間におけるユーザ群の興味がどう変化するか観測できるため、その変化からイベントの発生を検知することも可能となると考えられる。すなわち、本研究の最終的な目的は、イベント発見と分類を同時に行う手法の構築である。本稿では、まずその最初の段階として、イベントの分類が提案手法によりどの程度正確に行えるか、実際のイベントを対象とした評価を行う。本研究では上述したように、イベントに参加しているユーザの興味は Twitter 上のフォロイに反映されている、という仮説に基づくが、イベントによってはそのイベントに対応した Twitter アカウントが存在しない場合も想定される。また、イベント参加者の Twitter 利用率が低い場合や、特定のユーザ層のみに偏っている場合も考えられる。よって仮説の成立はイベントの種別や参加者の性質に依存することもあると考えられる。よって本稿では、上述した仮説に基づいた手法の有効性を議論するために、イベント分類手法を構築し、複数のイベントを対象とした分類実験を行う。

本提案手法の概要を、図 1 に示す。提案手法は、4つのステップから構成される。まず、(1) 解析対象となる空間に存在する Twitter ユーザを、対象空間・時間内に観測された位置情報付き発言より抽出する。次に、(2) 抽出されたユーザが共通してフォローしている、共通フォロイを抽出し、その数より重み付けを行う（共通フォロイランキング）。次に、(3) 抽出された共通フォロイの属性を、その共通フォロイを説明する Wikipedia 記事より分析し、分析された Wikipedia カテゴリの集合を特徴ベクトルとして計算する。最後に、(4) 各共通フォロイの特徴ベクトルを集計し、対象空間で発生しているイベントの特徴ベクトルと



図 2 フォロイランキンクの例：孫正義や NHK 広報局などは多くのユーザからともフォローされているため、ランキンク上位に入っている場合が多い

Fig. 2 Examples of followees ranking: Accounts such as Son-Masayoshi and NHK are followed by most of users, so they often appears on ranking.

して集計する。すなわち、本研究における分類とは、特徴ベクトルの次元数を分類軸として扱い、各分類軸に重みを与えることによって表現されたものである。以下に、(2)、(3)、(4) の各ステップについて詳しく説明を行う。

3.2 イベントに特徴的な共通フォロイランキンク生成方法

共通のフォロイとは、イベント参加者で構成されるソーシャルグラフにおいて、2人以上からフォローされているユーザのことである。予備実験として、8つのイベントでフォローされている数の多い順に共通のフォロイランキンクを作成した(図 2)。すると、孫正義(@masason)や地震速報(@earthquake.jp)のようなアカウントが多く見受けられた。これらのアカウントは、日本でトップクラスのフォロワー数を持つユーザである(孫正義のフォロワー数：2,595,630人、地震速報のフォロワー数：2,173,396人)。そのため、相対的にイベントに参加しているユーザが多くフォローしていた可能性があり、イベント特有のフォロイではない可能性がある。このようなユーザはイベントによってノイズとして扱われるユーザであり、イベント参加者に特徴的なフォロイか、そうでないかを判別する必要がある。そこで、本手法ではイベント特有のフォロイであるかを判別するために、各共通のフォロイに対して FeatureScore を割り振る。FeatureScore の算出は、以下の 3つの手順で行う。

手順 1

特定の空間・時間内で位置情報付き発言をしたユーザのフォロイのうち、2人以上のユーザからフォローされているフォロイの集合を $F = \{f_0, f_1, \dots, f_n\}$ とする。

手順 2

集合 F に含まれるそれぞれのフォロイの EventScore と BasicScore を計算する。

集合 F の任意の要素 f_x がイベント参加者からフォローされている数を $EventFollow(f_x)$ とすると、 $EventScore(f_x)$ は以下のように定義される。

$$EventScore(f_x) = \frac{EventFollow(f_x)}{\sum_{i=0}^n EventFollow(f_i)}$$

また、集合 F の任意の要素 f_x が Twitter でフォローされている数を $BasicFollow(f_x)$ とすると、 $BasicScore(f_x)$ は以下のように定義される。

$$BasicScore(f_x) = \frac{BasicFollow(f_x)}{\sum_{i=0}^n BasicFollow(f_i)}$$

手順 3

集合 F に含まれるそれぞれのフォロイの FeatureScore を計算する。FeatureScore(f_x) は以下の式で求められる。

$$FeatureScore(f_x) = EventScore(f_x) - BasicScore(f_x)$$

FeatureScore の値が高いほど、その共通のフォロイはイベントに特徴的なフォロイの可能性があり、FeatureScore の値が低いほど、その共通のフォロイはイベントに関係のないフォロイの可能性が高いと判断できる。本研究では、この FeatureScore を各共通のフォロイに割り振り、降順にソートを行い、上位 10 人を使用する。

3.3 イベントの属性推定手法

本手法を実現するために、ある空間におけるユーザ群の共通のフォロイがどのような属性であるのかを抽出する必要がある。まず考えられる方法として、Twitter アカウントに設定されたプロフィールを利用することが想定される。しかし、アカウントのプロフィールを記述しているユーザは限られているため、属性情報の抽出には不十分である。たとえば、ミュージシャンのきゃりーぱみゅぱみゅの Twitter^{*2} のプロフィールでは「無敵」としか書いておらず(2015 年 2 月時点)、属性情報としては利用不可能と考えられる。そこで本研究では、「共通のフォロイとして上位に位置付けられるアカウントは著名なアカウントであり、著名なアカウントであればインターネット上の情報を用いて分類することが可能である」という仮定のもと、属性抽出の手法を構築する。具体的には、そのアカウントに紐付けられる Wikipedia^{*3} のカテゴリ階層を利用してアカウントの分類を行う。事前調査では 8つの異なる実世界イベントに対し、イベント参加者内での共通のフォロイのフォロワー数でランキンクを作成し上位 10 人を確認したと

^{*2} <https://twitter.com/pamyurin>

^{*3} <http://ja.wikipedia.org>

ころ、すべてのユーザが Wikipedia に関連付けられる著名なユーザであった。よって、本研究では、共通のフォロイランキング上位 10 人の Wikipedia のカテゴリ属性に基づく分類を試みる。

3.3.1 Wikipedia を用いたフォロイの属性推定

Wikipedia は誰もが無料で自由に編集に参加できるインターネット百科事典である。Wikipedia に存在するすべての語にはその語に属するカテゴリが存在し、また各カテゴリは 1 つ以上の上位カテゴリに属するという階層構造になっており、このようなカテゴリの階層構造を利用した研究は多く存在している [17], [18]。本研究では、Wikipedia に存在する「百のカテゴリ^{*4}」(表 2)の中から Wikipedia の編集者のためのカテゴリである「総記」「プロジェクト関連文章」「一覧」「ウィキポータル」「画像」「索引」「ウィキペディアのテンプレート」「ヘルプ」と百のカテゴリのなかに「北海道」「関東」など、日本をさらに詳細に分類するカテゴリが存在するため「日本」を削除した、計 91 個のカテゴリを階層構造の第 1 層のカテゴリとする。カテゴリ数は 91 個ではあるが、この階層構造の第 1 層のカテゴリを便宜上「百のカテゴリ」と呼ぶ。その第 1 層のカテゴリから 1 ステップでたどることのできるカテゴリを第 2 層、さらに第 2 層から 1 ステップでたどることのできるカテゴリを第 3 層とし、それ以降の階層も同様にする。そのため、共通のフォロイに関連する Wikipedia 上のページに割り振られている各カテゴリから、幅優先で第 1 層カテゴリのいずれかに到達するまで再帰的に探索を行う。第 1 層カテゴリを発見した場合、発見された百のカテゴリに対してポイント(カテゴリポイント)を 1 Point 割り振る。また、同じ距離に第 1 層カテゴリを複数発見した場合は、発見した各カテゴリに対し 1 Point を割り振る。最終的にすべて割り振り終わったときに、1 Point も割り振られていなかったカテゴリに関しては 0 Point を割り振る。そうすることで、ユー

ザを 91 次元(百のカテゴリの数)の特徴ベクトルで分類することができる。また、ユーザに割り振られうるカテゴリポイントの総数は、Wikipedia の各ページに登録されているタグに依存するため、すべてのユーザを同じスケールで表すために、ユーザの特徴ベクトルの正規化を行う。この正規化後の特徴ベクトルを、その共通のフォロイの属性として利用する。このカテゴリ分類手法を、共通のフォロイとして抽出されたアカウントの上位 10 件に対して適応して行う。

3.3.2 イベントの属性抽出

共通のフォロイとして抽出されたアカウント上位 10 件から、イベントの属性を推測する。共通のフォロイとして抽出された 10 件の集合を $C = \{c_0, c_1, \dots, c_9\}$ とする。また、集合 C の任意の要素 c_x の FeatureScore を $FeatureScore(c_x)$ とし、集合 C の任意の要素 c_x の Wikipedia から抽出された特徴ベクトルを $\vec{c}_x$ とすると、イベントの特徴ベクトル $\vec{e}$ は以下の式で表される。

$$\vec{e} = \sum_{i=0}^9 \frac{FeatureScore(c_i)}{\sum_{j=0}^9 FeatureScore(c_j)} \vec{c}_i$$

4. 設計と実装

本研究では本手法を適用し、インタラクティブに実世界イベントを解析可能な Web アプリケーションの設計・実装を行った。同アプリケーションは以下の機能を有する。

- 位置情報付き発言を収集し、地図上に発言を表示する機能
- 地図上の位置情報付き発言からイベントを登録する機能
- 発見されたイベントの解析を行う機能

システム構成図を図 3 に示す。以下、それぞれの機能の設計・実装の詳細について述べる。

表 2 百のカテゴリ一覧

Table 2 List of one hundred categories.

社会, 政治, 経済, 産業, 交通, 教育, 歴史, 福祉, 医療
健康, 環境, 市民活動, 平和, 軍事, 大学, 芸術, 文化
言語, 宗教, 遊び, 趣味, 伝統芸能, 文学, 音楽, 美術
映画, 演劇, アニメ, 漫画, イラストレーション
スポーツ, ゲーム, 賭博, ファッション, 食文化, 建築
マスメディア, 芸能, 大陸, アジア, アフリカ, オセアニア
北アメリカ, 南アメリカ, ヨーロッパ, 北海道, 東北, 関東
中部, 近畿, 中国, 四国, 九州, 沖縄, 自然, 宇宙, 元素
気象, 災害, 海洋, 生物, 植物, 動物, 鉱物, 学問, 哲学
論理学, 言語学, 心理学, 文学, 宗教学, 政治学, 経営学
法学, 経済学, 社会学, 教育学, 数学, 物理学, 化学, 生物学
人類学, 地球科学, 医学, 薬学, 歯学, 農学, 工学, 技術
コンピュータ, ネットワーク, エレクトロニクス
バイオテクノロジー

^{*4} <https://ja.wikipedia.org/wiki/Wikipedia:カテゴリ>

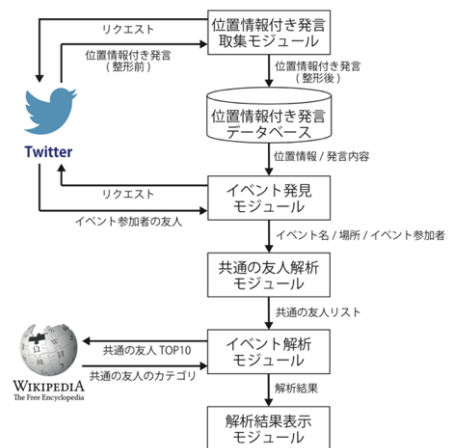


図 3 システム構成図

Fig. 3 System components of the proposed system.

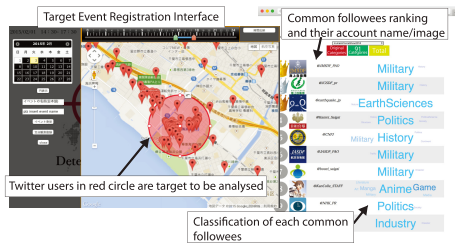


図 4 イベント登録ツール

Fig. 4 Event registration tool.

4.1 位置情報付き発言の収集・表示機能

本研究では位置情報付き発言を収集するために、Twitter の Streaming API^{*5}を用いた。Twitter Streaming API は Twitter 上で投稿される発言をリアルタイムで取得する仕組みである。本研究では対象を日本全国としたため、緯度経度が日本国内を含むように設定をして位置情報付き発言を収集し、データベースに保存を行う。Web アプリケーションでは保存された位置情報付き発言を、任意の日・時間帯を指定して Google Map^{*6}上に表示を行う。

4.2 イベント登録機能

本稿では、イベントが検出された後の分類を主眼をおく。将来的には同手法でのイベントの自動検出を目指す。本稿ではまずは分類を目的とする。そのため、本解析ツールにおいては自動ではなく、利用ユーザに手でイベントを発見・登録を行ってもらう機能を実装した。Google Maps 上にマッピングされた位置情報付き発言から、イベント発見・解析対象としての登録を直感的に行えるインタフェースを実装した。ユーザは以下の手順で、イベントを発見・登録することができる。

- (1) Google Maps を操作し、任意の場所を表示。
- (2) Google Maps 上のピンをクリックすることで位置情報付き発言を表示。
- (3) イベントを発見したら、円で位置情報付き発言を囲む。
- (4) イベントに名前をつけて保存する。

この機能画面を図 4 に示す。イベント登録の際、イベント名と円の範囲内の位置情報付き発言を行ったユーザの ID 情報とテキスト情報がデータベースに保存される。

4.3 イベント解析機能

イベント登録機能により登録されたイベントを上記した手法により、解析を行う。この機能を図 5 に示す。まず図 5 のセレクトボックスよりイベントを選択すると、その後選択されたイベントの共通の友人のランキング 10 位までが表示される。表示方法は「RAW」「百のカテゴリ」「All」の 3 段階用意した。「RAW」では Wikipedia から取得され、第 1 層カテゴリに対応づける前のカテゴリを表示

^{*5} <http://dev.twitter.com>

^{*6} <https://maps.google.co.jp>



図 5 イベント解析結果表示 (百のカテゴリ)

Fig. 5 Result of event analysis in terms of each twitter account with one hundred categories.



図 6 イベント解析結果表示 (All)

Fig. 6 Result of event analysis (All).

する。さらに「百のカテゴリ」では「RAW」カテゴリをそれぞれ第 1 層カテゴリに対応づけ正規化された後の特徴ベクトルをタグクラウドで表示する。また「All」ではランキング 1 位から 10 位の特徴ベクトルから、イベント全体の特徴ベクトルを計算しタグクラウドで表現する (図 6)。タグクラウドに関しては、フォントの大きさによって、そのユーザやイベントの属性の強さが表されている。

5. 評価

本章では、本提案手法によるイベント分類がどの程度可能かどうかを検証するため、実験し、評価を行う。

5.1 評価実験内容

対象となるイベントは、日本イベント産業振興協会^{*7}によって定義された 16 区分のイベントのうち、ツイート収集期間内に得られた、位置情報付きツイートが存在する 13

^{*7} <http://www.jace.or.jp/>

区分のイベント 16 種類を解析した。イベントの発見に利用した位置情報付き発言は、2011 年 11 月 1 日から 2013 年 6 月 14 日までと 2013 年 11 月 1 日以降の期間に投稿されたものである。なお、期間に間があいてしまった原因は TwitterAPI の仕様変更にともない、システムがこれに対応しておらず、停止してしまっていたためである。解析ツールにより、イベントに参加していると考えられるユーザを抽出し、イベントの特徴ベクトルを算出し、分類を行う。本評価では、提案手法で得られた分類結果を、(1) 27 人の被験者によるイベント分類との比較、(2) イベント自体の Wikipedia 記事から抽出した分類との比較、の 2 種類の比較を行う。(1) の比較を行う理由は、提案手法による分類が一般ユーザによる主観的な分類とどの程度近く行えるか、という点が重要であるからである。また、(2) の比較を行う理由は、Wikipedia の記事は方針として客観的な観点での編集がなされており、提案手法および (1) の主観的な分類と比較することにより、提案手法の位置づけがより明確になると考えたからである。また分類結果を比較するためには、同じ分類軸を用いて分類を行う必要があり、本研究は Wikipedia の 91 次元のカテゴリを分類軸として扱っているうえで、Wikipedia の記事との比較が重要であると考えた。

以上のように、本評価では発見された 16 種類のイベントを一般ユーザに 91 種類からなるタグ付けを行ってもらい、その集計結果を人による分類結果として利用した。また 16 種類のイベント中、13 種類のイベントでは Wikipedia に同名のイベントのページが存在しており、そのページのカテゴリから得られる特徴ベクトルは Wikipedia による分類結果として利用した。これらの 2 つの分類結果と本提案手法による分類結果を比較し、評価を行う。なお、これら分類結果は値が 0 の特徴ベクトルを除き、それぞれの特徴ベクトルの重み付けがなされたタグクラウドとして表現する。以下に、詳しく説明を行う。

5.2 Ground truth の準備

5.2.1 人によるタグクラウドの作成

人によるタグクラウドは、一般ユーザによってタグ付けされたデータであるとする。一般ユーザには図 7 のツールを使用し、16 種類のイベントすべてにタグ付けをしてもらった。一般ユーザには、発見された 16 種類のイベントの名称を伝え、本手法によりタグとして割り振られうる Wikipedia の百のカテゴリの中から、そのイベントにふさわしいタグを選択してもらいタグ付けを行ってもらった。一般ユーザは 1 つのイベントに対し、複数のタグを登録することができる。また、イベントをよく知らない場合は検索エンジンでイベントについて調べることを許可した。この人によるタグクラウドの作成には 27 人のユーザに参加してもらった。タグ付けには 18–25 歳の学生もしくは社

図 7 タグ付けアンケート画面

Fig. 7 Questionnaire form for tagging urban event.

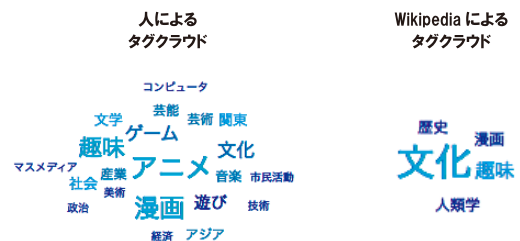


図 8 人によるタグクラウドと Wikipedia によるタグクラウド

Fig. 8 Comparison of questionnaire-based tag cloud and wikipedia-based tag cloud.

会人の男女に行ってもらった。この 27 人分のタグ付けのデータから、91 次元の特徴ベクトルを生成し、これを人によるタグクラウドとする。

5.2.2 Wikipedia によるタグクラウドの作成

Wikipedia のページに付与されているカテゴリはユーザの手で付けられたカテゴリであり、一種のタグ付けであるといえる。本手法により発見された 16 種類のイベント中 13 種類のイベントに関して、同名のイベントが Wikipedia のページとして存在しており、それらのページにもカテゴリが付与されている。このカテゴリをそれぞれ第 1 層カテゴリに対応付け、91 次元の特徴ベクトルを生成し、これを Wikipedia によるタグクラウドとする。

5.2.3 人によるタグクラウドと Wikipedia によるタグクラウドの違い

人によるタグクラウドと Wikipedia によるタグクラウドの違いの例として、図 8 にコミックマーケットを対象とした、人によるタグクラウドと Wikipedia によるタグクラウドを示す。人によるタグクラウドでは、アニメ、漫画、趣味が出現しているのに対し、Wikipedia によるタグクラウドでは、文化が出現し、そもそもアニメが存在していない。

イベント名 (区分)	対象データ	提案手法による タグクラウド	人による タグクラウド	提案手法と人による タグクラウドの コサイン類似度	Wiki による タグクラウド	提案手法と Wiki による タグクラウドの コサイン類似度	人と Wiki による タグクラウドの コサイン類似度
全国高校サッカー選手権大会決勝 (スポーツ・競技)	2015/1/12 8:30 - 19:00 ユーザ数: 237	ヨーロッパ マスメディア スポーツ 健康 南アメリカ 芸術 文化	外編 アジア 音楽 マスメディア 市民活動 大学 北海道 中部 自然 東北	0.982	スポーツ 音楽 人類学 文化 技術 マスメディア	0.542	0.541
ももクロライブ (音楽などの 芸術鑑賞)	2013/4/13 11:00 - 21:30 ユーザ数: 75	趣味 政治 歴史 災害 音楽 芸術 教育 産業 文化	音楽 演劇 アジア 遊び 政治 経済 趣味 芸術 音楽 文化 マスメディア ファッション	0.872			
コミックマーケット (フリマ・ 物産展)	2014/8/15 12:00 - 15:00 ユーザ数: 215	文化 イラストレーション マスメディア 漫画 音楽 政治 経済 芸術 大学 教育 人間学 アニメ コンピュータ 教育 ファッション	コンピュータ 遊び 漫画 文化 アニメ 民間 経済 政治 マスメディア 市民活動 芸術	0.834	歴史 漫画 文化 趣味 人類学	0.599	0.516
FujiRock Festival (音楽などの 芸術鑑賞)	2014/7/25 15:00 - 17:00 ユーザ数: 86	政治 芸術 産業 音楽 マスメディア コンピュータ	コンピュータ マスメディア 音楽 自然 芸術 遊び ファッション 芸術 音楽 文化 芸術 民間 芸術 市民活動 芸術 エレクトロニクス スポーツ	0.811	音楽 芸術	0.879	0.673
入間基地 航空祭 (パフォーマ ンス)	2014/11/3 9:00 - 15:00 ユーザ数: 238	コンピュータ ゲーム マスメディア 経済学 音楽 政治 経済 軍事 大学 政治 芸術 自然 アニメ 産業 オセアニア	伝統芸術 政治学 アジア 政治 経済 大学 音楽 マスメディア 市民活動 社会 文化 スポーツ コンピュータ	0.732			
東京アニメ アワード (コンクール)	2014/3/23 9:00 - 15:30 ユーザ数: 383	オセアニア 漫画 音楽 マスメディア 政治 産業 ゲーム アニメ 趣味 芸術 教育 経済学 文学	東北 社会 経済 遊び アニメ 民間 経済 文化 産業 コンピュータ マスメディア 芸術	0.715	アニメ	0.630	0.896
東京モーターショー (展示会・ 見本市)	2013/11/22 9:30 - 22:00 ユーザ数: 74	北アメリカ 政治 経済 産業 芸術 文化 社会 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	ネットワーク コンピュータ エレクトロニクス 文化 政治 産業 遊び 技術 趣味 経済 政治 芸術 民間 芸術 市民活動 芸術 政治 産業 マスメディア 芸術	0.642	文化 交通 経済	0.093	0.225
さよなら 原発集会 (行列・ パレード)	2012/7/16 11:30 - 14:30 ユーザ数: 96	コンピュータ 産業 ファッション 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	経済学 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.596	物理学 工学 市民活動	0.322	0.344
メディア 芸術祭 (美術などの 芸術鑑賞)	2014/2/11 10:00 - 18:30 ユーザ数: 59	スポーツ コンピュータ 文化 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	アジア エレクトロニクス 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.517	ゲーム 美術 芸術 文化 漫画	0.551	0.842
さっぽろ 雪まつり (自然鑑賞)	2014/2/9 9:00 - 20:30 ユーザ数: 32	マスメディア 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	音楽 社会 経済 遊び 北海道 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.356	文化 歴史 芸術 北海道 人類学 社会	0.399	0.673
博多どんたく (伝統的に 続くお祭り)	2014/5/4 8:30 - 19:00 ユーザ数: 100	趣味 文化 大学 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	マスメディア 伝統芸術 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.346	人類学 文化 歴史	0.263	0.549
外苑前 花火大会 (自然鑑賞以 外の鑑賞)	2014/8/16 19:30 - 20:30 ユーザ数: 117	政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	伝統芸術 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.253	自然 芸術 マスメディア 遊び 文化	0.642	0.404
箱根駅伝 (スポーツ ・競技)	2015/1/3 12:00 - 18:00 ユーザ数: 43	人類学 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	伝統芸術 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.223	技術 歴史 スポーツ マスメディア 大学 産業 文化 芸術	0.481	0.548
神田祭 (伝統的に 続くお祭り)	2013/5/12 8:30 - 19:00 ユーザ数: 44	政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	人類学 政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.186	宗教	0.000	0.218
目黒の さんま祭り (地域の祭 り・行事)	2014/9/7 8:30 - 19:00 ユーザ数: 23	政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.125			
弘前 さくら祭り (自然鑑賞)	2014/4/27 9:00 - 20:30 ユーザ数: 46	政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	政治 経済 大学 政治 産業 民間 経済 技術 民間 教育 コンピュータ スポーツ 芸術	0.036	環境 技術 植物	0.226	0.147

図 9 解析対象イベントと解析結果

Fig. 9 Target events for evaluation and their analysis results.

他のイベントに関しても、出現するタグの数自体基本的に少なく、またイベントの記事が存在しないものもあった(図9参照)。レコメンデーションなどのアプリケーションを考えた際、イベントの分類は人の感覚に沿ったものであったほうが望ましいと考えられるが、人によるタグクラウドと Wikipedia によるタグクラウドを比較した場合、両者には乖離が存在している(定量的な比較は、5.4.2 項で説明する)。そもそも Wikipedia の記事自体、人の手によって書かれたものではあるが、Wikipedia の編集方針^{*8}である「中立性の担保」によって、実際の人の感覚とは乖離した分類になっていることが原因の1つとして考えられる。

本研究ではできるだけ人の感覚に近いイベント分類を目指すため、人によるイベント分類と提案手法によるイベント分類の類似性を重要視する。一方で、「中立性」という観点で行われたイベント分類との比較を行うことも意義があると考えられるため、Wikipedia によるイベント分類に関しても提案手法によるイベント分類との類似性を評価対象とする。

5.3 比較方法

本研究におけるタグクラウドは、ともに 91 次元の特徴ベクトルで表されている。このような n 次元ベクトルどうしがどれくらい類似しているかを評価する指標として、コサイン類似度を用いる。コサイン類似度は n 次元ベクトルの向きの類似性を表す値であり、ベクトル $\vec{x}$ とベクトル $\vec{y}$ のなす角 θ の余弦 $\cos \theta$ である。ベクトル $\vec{x}$ とベクトル $\vec{y}$ のコサイン類似度 $\cos(\vec{x}, \vec{y})$ は以下の式で求められる。

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} = \frac{\sum_{i=1}^{|v|} x_i y_i}{\sqrt{\sum_{i=1}^{|v|} x_i^2} \sqrt{\sum_{i=1}^{|v|} y_i^2}} \quad (1)$$

本手法で得られる特徴ベクトルの値はすべて正であることから、値は 0 から 1 の間の値をとるが、2つのベクトル $\vec{x}, \vec{y}$ の向きが等しいほど 1 に近づき、類似していることがいえる。

5.4 実験結果・考察

それぞれのイベントごとの、本提案手法による分類に基づくタグクラウド、および人による分類に基づくタグクラウド、Wikipedia 記事の分類に基づくタグクラウド、本提案手法と各タグクラウドによるコサイン類似度を図9に示す。また、図9中に、各イベントにおいてデータ分析対象となった Twitter ユーザ数についても記載する。

5.4.1 提案手法と人によるタグクラウドの比較

本提案手法から得られたイベントの特徴ベクトルと、人



図10 全国高校サッカー選手権大会と箱根駅伝

Fig. 10 High school soccer championship tournament and Hakone road relay.

によるタグクラウドの特徴ベクトルをもとにコサイン類似度を算出した。この結果、全国高校サッカー選手権大会、ももクロライブ、コミックマーケット、FujiRockFestival、入間基地航空祭、東京アニメアワードなど、強い類似性を持った分類が行われた。これらのイベントはタグクラウド上からも、特に強く示されている上位のカテゴリが一致しており、類似性が見て取れる。一方で、弘前さくら祭り、目黒のさんま祭り、神田祭、箱根駅伝、さっぽろ雪まつり、外苑前花火大会などではコサイン類似度から類似度が低いと示された。

これらの類似性の高い6つのイベントの共通点は、趣味性が高いイベントであるといえる。コミックマーケットには漫画やアニメが趣味なユーザが、全国高校サッカー選手権大会にはサッカーが好きなユーザが、入間基地航空祭には自衛隊が好きなユーザが参加するだろうと容易に推察できる。しかし、神田祭や外苑前花火大会、弘前さくら祭りなどには、特定の興味を持つユーザではなく老若男女多様なユーザが参加するため、ユーザ群の興味を簡単に導き出すことはできない。実際に、図10のように共通のフォロイとして抽出されたアカウントの上位10件のアカウントを見てみると、類似性の高い全国高校サッカー選手権大会では、サッカーキング (@SoccerKingJP) やゲキスカ (@gekisaka)、日本サッカー協会 (@JFA) などサッカーに関連するアカウントが出ているのが見て取れる。しかし類似性の低い箱根駅伝では、箱根駅伝番組公式 (@hakone-ntv) は出ているものの、これはテレビ番組のアカウントであり、属性としてスポーツが弱く、また他のアカウントは駅伝とは関連していないようなアカウントが抽出されている。これらのイベントは日本イベント産業振興協会の区分で同じ「スポーツ・競技」のイベントであるが、イベントに参加する人の趣味性の強さによって結果が変わってくる。つまり本提案手法は、イベント参加者の共通のフォロイを用いることで、イベントを分類する手法であるため、箱根駅伝のような多様な人々が参加するイベントでは共通のフォロイ

^{*8} <https://ja.wikipedia.org/wiki/Wikipedia:五本の柱>

にも多様なユーザが出現し、イベントを正しく分類することはできなかった。しかし、これはイベント参加者の興味が分散しているといえることから、河野ら [14] の大衆性の高いイベントということは判別できる可能性がある。

5.4.2 提案手法と Wikipedia によるタグクラウドの比較

本手法から得られたイベントの特徴ベクトルと、イベントと同名の Wikipedia ページから得られた特徴ベクトルをもとにコサイン類似度を算出した。この結果、高い類似度を示すものは FujiRockFestival のみであった。5.2.3 項でも述べたように、Wikipedia では、「中立性」という観点でイベント分類がなされていると考えられ、実際に人の感覚とは乖離した分類になっている可能性がある。実際に人によるタグクラウドと、Wikipedia によるタグクラウドをコサイン類似度を用いて比較してみると、13 種類のイベント中 2 種類のイベントしか高い類似性を示していなかった。これにより、Wikipedia によるタグクラウドは人の感覚とは乖離した分類になっているものが多く存在するといえる。このため提案手法と人によるタグクラウドの比較において類似性が高いとされた 6 つのイベント中の 5 つのイベントが Wikipedia によるタグクラウドとの比較では、類似性が低くなったことより、本提案手法は「中立性」という観点より、一般ユーザが考えるイベント像に近い分類を行うことができるといえる。

5.4.3 考察

本提案手法によるタグクラウドと人によるタグクラウドの比較結果および Wikipedia によるタグクラウドの比較結果より、本提案手法は、趣味性の高いイベントに関して高い類似性を持ったイベント分類が可能で、また、Wikipedia によるタグクラウドよりも人によるタグクラウドのほうがコサイン類似度が高いイベントが多く、一般ユーザが考えるイベント像に近い分類を行うことができるといえる。

以下、分類が正しく行えたかどうか判断が難しい個別のイベントに関して、考察を行う。

- さっぽろ雪まつり・博多どんたく

さっぽろ雪まつりを対象とした本提案手法での分類結果では、音楽や工学というキーワードが特徴として現れており、これが人による分類と類似性を低下させる原因であった。また博多どんたくにおいても、元来お祭りに関係のないと推測される芸能や音楽、ファッションなどのキーワードが特徴として現れている。

この原因を探るため、実際どういったアカウントが共通フォロイで現れたか調査したところ、さっぽろ雪まつりでは仮想アイドルの「初音ミク」に関するアカウントが多数、また博多どんたくでは大分のローカルアイドル「Niimo」に関するアカウントが多数上位に現れていた。そこで実際に両イベント当日でどういった内容の催し物が行われたか調査したところ、さっぽろ雪まつりではその中のサブイベントとして「初音ミク

雪祭り」という初音ミクにちなんだイベントが、博多どんたくでは「Niimo ライブ」が行われていたことが分かった。この事実より、提案手法による分類結果は、一般的な両イベントの印象としてではなく、実際現地で行われているイベントに基づいた分類が行われている可能性が高いといえる。一方で、本提案手法はその場所に存在しかつ発言を行った Twitter ユーザの存在に依存しているため、特定のユーザ層の影響を強く受けてしまうという問題点も考えられる。たとえば博多どんたくでは多くの観光客が存在し、様々な催しものが行われているが、特にその中でも特定のアイドルのライブ時に多くの発言が行われたり、そのアイドルのファンの Twitter 使用率が高い場合に、そのユーザ層の影響を強く受けてしまった可能性もある。このように、実際の現地のイベントに基づいて正しく分類が行えているか、それとも現実空間のみの参加ユーザ層と現実・仮想 (Twitter) 両空間とも参加・発言しているユーザ層との乖離による偏ったな分類なのかの判断手法が今後重要となるということが分かる。また、両イベントのようにあるイベントの中で多種多様なサブイベントが行われるイベントでは、特定のイベントのみ Twitter 使用率が高いユーザ層が存在するということも考えられるため、イベント全体として現実空間と仮想 (Twitter) 空間との参加者が乖離し、イベント分類にマイナスの影響を与える可能性も高いと考えられる。その点を考慮した場合、人による分類と比較し高い類似性を有したサッカー選手権大会やももクロライブ、コミックマーケットは比較的均質な性質・嗜好を持ったユーザが現実空間および仮想空間に存在しているとも考えられる。

- メディア芸術祭・東京モーターショー

本提案手法と人によるタグクラウドの比較では、コサイン類似度が 0.642 (東京モーターショー)、0.517 (メディア芸術祭) と、低くはないが、高い相関があるともいえない結果となった。その原因として、これらのイベントに共通することは、共通のフォロイ上位 10 人にイベントに関係するユーザが存在しているが、少数であるため、イベントの特徴ベクトルとして算出した結果、強く現れなかったことにある。また、他の原因として考えられるのは、人による一般的な印象は、東京モーターショーやメディア芸術祭は趣味性の高いイベントとだと考えられがちであるが、実際は様々な参加者を集める大衆的なイベントであるという可能性が高い。実際、河野ら [14] によるイベントの大衆性に関する研究において、東京モーターショーやメディア芸術祭は比較的大衆性が高いと位置づけられている。これら不確定な結果のイベントに関しては、今後イベント主催者への調査などを通じて、実態に関して明らか



図 11 さよなら原発集会分析結果（百のカテゴリ）

Fig. 11 Analysis result of Good-bye nukes event (one hundred categories).

にし、議論を行う必要がある。

● さよなら原発集会

図 11 にさよなら原発集会の共通のフォロイトップ 10 を示す。多くのユーザは反原発活動に関連するユーザであるが、イベントの属性をうまく推定することができなかった。これは、抽出された共通のフォロイには反原発活動以外にも著名なユーザが多く、Wikipedia による解析でより有名な活動のカテゴリに関連付けられたと考えられる。たとえば、坂本龍一氏 (@skmt09) は反原発活動家ではあるが、さらに有名なのは音楽家としての活動である。実際、図 11 の坂本龍一氏の分析結果には音楽カテゴリが一番強く推定されている。

6. 議論・今後の課題

本研究では分析対象となる Twitter ユーザが共通フォロイ上位に現れ、Wikipedia に記事が存在するという仮定・分析に基づき、そのユーザの属性分析を Wikipedia カテゴリを利用して行った。一方、これまでも Twitter ユーザのプロフィール情報、ツイート頻度、ツイート内容などを、LDA (Latent Dirichlet Allocation) 分析などを利用した属性分析が提案されている [3], [4], [19], [20]。これらの分析は単語レベルでのトピック推定として有用であり、ユーザの嗜好など（たとえば文献 [3] ではユーザの支持政党などを推定している）の分析は可能である。一方で、高次元の抽象化されたカテゴリを抽出するには適しておらず、またユーザのツイート数・内容に大きく影響を受けるため、客観的な観点の分類が難しいという問題が存在している。イベント推薦などのアプリケーションを想定した際には、集

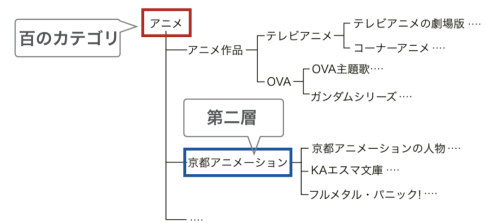


図 12 Wikipedia におけるアニメカテゴリの階層構造

Fig. 12 Category tree of animation category in Wikipedia.

合知としてカテゴリ化された指標に基づいて分類が可能な本研究のアプローチの利点は大きいと考えられる。一方、たとえば反原発集会が活発であった時期は、坂本龍一氏は反原発関係のツイートを多く行ったりしているため、LDA 分析などとの組合せの有効性も検証していく必要がある。Matthew らはユーザの興味の推定を、ツイート内容を本研究と同様 Wikipedia のカテゴリ分析を行って導出している [5]。これらの手法との組合せも検討していく。

本稿においては、イベントの分類軸を Wikipedia の第 1 層である百のカテゴリを利用した。一方、より細かい分類軸でイベントを分類したい場合、Wikipedia のカテゴリ層のレベルを動的に調整することで可能となることも考えられる。たとえば、図 12 は、アニメカテゴリにおける階層構造である。第 2 層のレベルを分類軸とすることで、より細かい分類情報をユーザに提示しつつ、第 1 層レベルの分類軸で他イベントとの類似性を比較することで、より気の利いたイベント発見・推薦アプリケーションが構築可能となると考えられる。こういったカテゴリの階層構造を活用した分析は既存研究のイベント解析では難しいと考えられ、本研究の利点といえる。

また、今回イベント参加者を一定の円の範囲に存在する位置情報付き発言を行っているユーザとしたが、その結果、イベントに参加していないユーザも含まれている可能性があり、その結果分類の精度が下がっている可能性も考えられる。そこで、Hiruta ら [6] や Wang ら [21], [22] のように、機械学習やテキスト解析を行うことで、イベントに参加しているユーザの信頼性を向上させることが期待できる。また、どの程度のユーザ数を分析対象とすれば、どの程度分類精度が得られるかも検証する必要がある。そのほかに、本提案手法では、イベント参加者に多様性があり、興味が分散しているイベントは正確に分類することができないが、このようなイベントは、河野ら [14] の大衆性の高いイベントと判別できる可能性があるため、分析を重ねていきたい。加えて、メディア芸術祭や東京モーターショーなどのように、共通のフォロイの FeatureScore の精度が向上すれば、分類できるイベントの幅が広がる可能性があるため、改善を行っていきたい。

本研究では 16 種類のイベントを対象として分析を行ったが、分析の規模としてはまだまだ小さいと考えられるた

め、今後解析対象数を増やすとともに、ツールの公開^{*9}を行うことで様々なイベントの解析を行っていく。最後に、本稿ではイベントの分類のみを行ったが、本研究の目的にはイベント検知も含まれている。今後、タグクラウドに含まれるカテゴリの変化をとらえることで、ある空間におけるイベントの検出が可能となると考えられる。

7. まとめ

本研究では、位置情報付き発言を用いて、「位置情報付き発言が実世界イベントが発生している場所に存在する」という前提条件のみでイベントの検出・分類を可能とするため、空間内に存在するユーザ群の「共通の友人」を抽出・解析し、そのユーザ群の興味を推定し、イベントの分類に用いる手法を提案した。また提案した手法に基づきイベントの分類を可能とするインタラクティブな Web ツールの設計・実装を行った。解析には、コサイン類似度を用い、27 人の一般ユーザによるイベントのタグクラウドと、イベントと同名の Wikipedia のページのタグクラウドを本提案手法によるタグクラウドと比較した。その結果、本提案手法によるタグクラウドと人によるタグクラウドの比較結果および Wikipedia によるタグクラウドの比較結果より、本提案手法は、趣味的の高いイベントに関して高い類似性を持ったイベント分類が可能で、また、Wikipedia によるタグクラウドよりも人によるタグクラウドのほうがコサイン類似度が高いイベントが多く、一般ユーザが考えるイベント像に近い分類を行うことができるということが分かった。イベント推薦システムなどに本提案手法を取り入れると、ソーシャルメディアなどからユーザの興味のタグクラウドを生成し、本提案手法によるタグクラウドと比較するだけで、そのユーザの興味にあったイベントを簡単に推薦することができ、今後本手法を適用したイベント推薦システムの設計と実装が期待される。

謝辞 本研究の一部は、独立行政法人情報通信研究機構にご支援いただいた。

参考文献

- [1] Steiger, E., de Albuquerque, J.P. and Zipf, A.: An advanced systematic literature review on spatiotemporal analyses of twitter data, *Transactions in GIS*, Vol.19, No.6, pp.809-834 (2015).
- [2] Weng, J. and Lee, B.-S.: Event detection in twitter, *International AAAI Conference on Web and Social Media* (2011).
- [3] Pennacchiotti, M. and Ana-Maria, P.: A Machine Learning Approach to Twitter User Classification, *ICWSM 11*, pp.281-288 (2011).
- [4] Pennacchiotti, M. and Siva, G.: Investigating topic models for social media user recommendation, *Proc. 20th International Conference Companion on World Wide Web*, pp.101-102, ACM (2011).
- [5] Michelson, M. and Macskassy, S.A.: Discovering users' topics of interest on twitter: A first look, *Proc. 4th Workshop on Analytics for Noisy Unstructured Text Data*, ACM (2010).
- [6] Hiruta, S., Yonezawa, T., Jurmu, M. and Tokuda, H.: Detection, classification and visualization of place-triggered geotagged tweets, *Proc. 2012 ACM Conference on Ubiquitous Computing*, pp.956-963 (2012).
- [7] Ishikawa, S., Arakawa, Y., Tagashira, S. and Fukuda, A.: Hot topic detection in local areas using Twitter and Wikipedia, *ARCS Workshops (ARCS 2012)*, pp.1-5, IEEE (2012).
- [8] Lanagan, J. and Smeaton, A.F.: Using twitter to detect and tag important events in live sports, *Artificial Intelligence*, pp.542-545 (2011).
- [9] 金子昂夢, 柳井啓司: Twitter 上の位置情報付き画像を利用したリアルタイムイベント画像検出, 人工知能学会全国大会 (第 28 回) 論文集, 3M4-5 (2014).
- [10] Lee, R. and Sumiya, K.: Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection, *The 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pp.1-10, ACM (2010).
- [11] Boettcher, A., Lee, D. and Eventradar: A real-time local event detection scheme using twitter stream, *2012 IEEE International Conference on Green Computing and Communications (GreenCom)*, pp.358-367, IEEE (2012).
- [12] Thelwall, M., Buckley, K. and Paltoglou, G.: Sentiment in Twitter events, *Journal of the American Society for Information Science and Technology*, Vol.62, No.2, pp.406-418 (2011).
- [13] 藤坂達也, 李 龍, 角谷和俊: 実空間マイクロブログ分析による地域イベントの影響範囲推定, *DEIM Forum* (2010).
- [14] 河野 慎, 米澤拓郎, 中澤 仁, 川崎仁嗣, 太田 賢, 稲村 浩, 徳田英幸: ソーシャルネットワークにおけるフォロー集合分析に基づく実世界イベント分類手法, 情報処理学会論文誌, Vol.56, No.1, pp.72-82 (2015).
- [15] Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S. and Miller, R.C.: Twitinfo: Aggregating and visualizing microblogs for event exploration, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.227-236 (2011).
- [16] Nichols, J., Mahmud, J. and Drews, C.: Summarizing sporting events using twitter, *Proc. 2012 ACM International Conference on Intelligent User Interfaces*, pp.189-198 (2012).
- [17] 吉岡真治: Wikipedia のカテゴリ階層関係の分類を用いた日本語 Wikipedia オントロジーの分析, 人工知能学会全国大会 (第 28 回) 論文集, 2J3-4 (2014).
- [18] 柴木優美, 永田昌明, 山本和英: カテゴリ名と記事名の意味属性分類に基づく Wikipedia からの上位下位関係オントロジーの構築, 自然言語処理, Vol.19, No.4, pp.229-279 (2012).
- [19] 柿 剛史, 松尾 豊: ソーシャルブックマークとしての Twitter リスト機能の応用, 人工知能学会全国大会 (第 24 回) 論文集, 3B3-2 (2010).
- [20] 小坂龍一, 青野雅樹: 機械学習を用いた Tweet の多カテゴリ分類, *DEIM Forum* (2012).
- [21] Wang, D., Abdelzaher, T., Ahmadi, H., Pasternack, J., Roth, D., Gupta, M., Han, J., Fatemeh, O., Le, H. and Aggarwal, C.C.: On Bayesian interpretation of factfinding in information networks, *Proc. 14th International Conference on Information Fusion (FUSION)*, pp.1-8,

^{*9} <http://homin.ht.sfc.keio.ac.jp/index2.html>

IEEE (2011).

- [22] Wang, D., Kaplan, L., Le, H. and Abdelzaher, T.: On truth discovery in social sensing: A maximum likelihood estimation approach, *Proc. 11th International Conference on Information Processing in Sensor Networks*, pp.233-244, ACM (2012).



佐藤 翔野

2014 年青山学院大学理工学部情報テクノロジー学科卒業。2016 年慶應義塾大学大学院政策・メディア研究科前期博士課程修了。同年トヨタ自動車(株)入社。ユビキタスコンピューティングシステム、ソーシャルネット

ワークに関する研究に従事。



米澤 拓郎 (正会員)

慶應義塾大学大学院政策・メディア研究科特任講師。博士(政策・メディア)。主に、ユビキタスコンピューティングシステム、ヒューマンコンピュータインタラクション、アーバンセンシング等の研究に従事。ACM 会員。



河野 慎

2014 年慶應義塾大学環境情報学部卒業。2016 年東京大学大学院修士課程修了。現在、慶應義塾大学大学院政策・メディア研究科後期博士課程在学。主に、機械学習の研究に従事。



中澤 仁 (正会員)

慶應義塾大学環境情報学部准教授。博士(政策・メディア)。ミドルウェア、システムソフトウェア、ユビキタスコンピューティング等の研究に従事。日本ソフトウェア科学会, IEEE 各会員。



川崎 仁嗣

株式会社 NTT ドコモ先進技術研究所勤務。2008 年筑波大学大学院システム情報工学研究科博士前期課程修了。同年(株)NTT ドコモ入社。モバイルコンピューティング, 端末セキュリティ, 分散システムに関する研究に

従事。



太田 賢 (正会員)

1998 年静岡大学大学院博士課程修了。博士(工学)。1999 年 NTT 移動通信網(株)入社。現在, NTT ドコモ先進技術研究所勤務。モバイルコンピューティング, 端末セキュリティ, 分散シ

ステムに関する研究に従事。訳書『コンピュータネットワーク第5版』等。電子情報通信学会会員。



稲村 浩 (正会員)

1990 年慶應義塾大学大学院理工学研究科修士課程修了。同年日本電信電話(株)入社。1998 年より NTT ドコモ。2016 年より公立はこだて未来大学教授。博士(工学)。モバイルネットワーク, スマートデバイスのシステムソフトウェアに関する研究開発に従事。電子情報通信学

会, ACM, IEEE 各会員。本会業績賞。



徳田 英幸 (正会員)

1975 年慶應義塾大学工学部卒業。同大学大学院工学研究科修士。ウォータールー大学計算機科学科博士(Ph.D. in Computer Science)。米国カーネギーメロン大学計算機科学科研究准教授を経て, 1990 年慶應義塾大学環境情報

学部に勤務。慶應義塾常任理事を経て, 現在, 慶應義塾大学環境情報学部教授。専門は, ユビキタスコンピューティングシステム, OS, Cyber-Physical Systems 等。日本ソフトウェア学会フェロー。情報処理学会副会長, 日本学術会議会員, 情報通信審議会委員等を務める。