

# 平時と異なる事象に対するソーシャルセンシング技術に関する研究

坂本 一磨<sup>1,a)</sup> 中村 健二<sup>2</sup> 山本 雄平<sup>3</sup> 田中 成典<sup>4</sup>

受付日 2017年12月25日, 採録日 2018年7月10日

**概要:** CGM (Consumer Generated Media) の普及にともない, インターネット上に膨大なデジタルデータが蓄積されている. このデータを活用し, 社会動向, 経済動向に加え, 大規模災害などの多種多様な現象を計測するソーシャルセンシングが注目されている. ソーシャルセンシングに関する研究には, マイクロブログを用いて特定のキーワードの出現数や文脈を解析して現象を検知する手法がある. しかし, これらの手法では, キーワードを事前に指定する必要があるため網羅的な分析が困難であることや, キーワード選定に解析者のバイアスがかかり分析に偏りがみられることなどの課題が発生する. そこで, 本研究では, ユーザの習慣行動を用いて, 実世界における現象を抽出する新たなソーシャルセンシング手法を提案する. 本手法では, 「平時と異なる行動を起こすユーザ群を特定することで, その異常行動から何らかの大きな社会事象が発生している」という仮説を設定し, ユーザ行動の変化を用いて, 現象を抽出する手法を検討する.

**キーワード:** マイクロブログ, ソーシャルセンシング, 社会事象, 習慣行動

## Research for Social Sensing Technology to Different Phenomena during Normal Times

KAZUMA SAKAMOTO<sup>1,a)</sup> KENJI NAKAMURA<sup>2</sup> YUHEI YAMAMOTO<sup>3</sup> SHIGENORI TANAKA<sup>4</sup>

Received: December 25, 2017, Accepted: July 10, 2018

**Abstract:** The spread of CGM (Consumer Generated Media) has made huge amount of digital data to be accumulated on the Internet. Using this big data, a social sensing to measure various phenomena that are a social tendency, an economic trend, and a huge disaster has been focused on. In the existing research related to the social sensing there are lots of methods that detect various phenomena to analyze the appearance of specific keywords or the context of a passage with microblogs. But, it is difficult for these methods comprehensively to analyze the context because it is necessary in advance to decide on the specific keywords. In addition, we have the subject to discover the appropriate keywords. Then using actual behaviors of people, we propose a method to detect various phenomena in the real world. Therefore, we make an assumption that the large social events would be happened by irregular actions of people if people occur the different events during normal times are specialized. Using actual behaviors and attributes of people, we verify a method to detect various phenomena in the real world.

**Keywords:** microblog, social sensing, social phenomena, habitual behavior

### 1. はじめに

CGM (Consumer Generated Media) の普及にともない, インターネット上には, 多様で膨大なデジタルデータが蓄積され, 様々な分野の研究で活用されている. しかし, こ

<sup>1</sup> 関西大学大学院総合情報学研究所  
Graduate School of Informatics, Kansai University,  
Takatsuki, Osaka 569-1095, Japan

<sup>2</sup> 大阪経済大学情報社会学部  
Faculty of Information Technology and Social Sciences,  
Osaka University of Economics, Osaka 533-8533, Japan

<sup>3</sup> 関西大学先端科学技術推進機構  
Organization for Research and Development of Innovative  
Science and Technology, Kansai University, Suita, Osaka  
564-8680, Japan

<sup>4</sup> 関西大学総合情報学部  
Faculty of Informatics, Kansai University, Takatsuki, Osaka  
569-1095, Japan

a) k245159@kansai-u.ac.jp

これらの研究では、解析対象を特定のユーザやコミュニティに限定していることや、分析対象のトピックを特定の商品やサービス、イベントなどに限定していることから、広範囲な視点からの社会現象の把握ができていない。そこで、社会動向、経済動向に加え、事件、事故、大規模災害などの多種多様な現象（以下、社会事象）を計測するソーシャルセンシングが注目されている。ソーシャルセンシングに関する研究は、検索エンジンを用いて現象を検知する研究 [1], [2] から徐々に増加している。これらの研究では、検索エンジンの検索履歴を用いて、インフルエンザの流行を抽出する研究が行われているが、解析対象の検索履歴の入手が困難であるため、多くの研究は行われていない状況である。また、ブログを対象として、選挙得票と株式市場を予測する研究 [3] では、特定のキーワードが含まれるブログの記事数や相場の上昇と下落時を表す特徴的な単語を用いて現象を予測している。これらの研究では、ブログの特性上、イベントからユーザの投稿までのタイムラグが発生するため、即時性の課題があった。このため、近年では、データが容易に入手でき、即時性と拡散性に優れたマイクロブログを用いてソーシャルセンシングする研究 [4] が注目されている。既存研究 [4] では、マイクロブログのユーザをソーシャルセンサととらえてセンシングする。マイクロブログを対象とした既存研究では、地震や台風といった災害を検知する研究 [5], [6], [7] やスポーツイベントを検知する研究 [8], [9]、鉄道や交通渋滞などの遅延や事故に関する情報を抽出する研究 [10], [11], [12]、映画の興行収入を予測する研究 [13]、経済動向を分析する研究 [14], [15], [16]、インフルエンザの流行を予測する研究 [17], [18] が実施されている。これらのソーシャルセンシングに関する研究では、主に特定のキーワードの出現数や文脈を解析して、現象を検知する手法が利用されている。しかし、現象ごとに特定のキーワードを事前に指定する必要があるため、網羅的な分析が困難であることや、キーワード選定に解析者のバイアスがかかり分析に偏りがみられることなどの課題が発生する。

そこで、本研究では、既存研究の課題に対応するため、ユーザの習慣行動を用いて、実世界における現象を抽出する新たなソーシャルセンシング手法を提案する。本手法では、「平時と異なる行動（以下、異常行動）を起こすユーザ群を特定することで、その異常行動から何らかの大きな社会事象が発生している」という仮説を設定し、このユーザ行動の変化を用いて、社会事象の抽出を試みる。これにより、キーワードの出現数や文脈のみに依存せず、多種多様な現象の抽出が可能になると考える。

本論文の構成は、以下のとおりである。2章では、本研究で証明する仮説とその方策について説明する。3章では、本研究におけるソーシャルセンシング手法に関して説明し、4章では、研究目的と実験計画を整理する。5章では、本研究を総括する。

## 2. 研究の概要

### 2.1 研究目的

著者らは、マイクロブログを用いたユーザの行動を推定する研究 [19], [20], [21] に取り組んできた。行動推定に関する研究 [19] では、ユーザ属性の推定時に投稿時間の情報を考慮することの有効性を証明した。また、属性推定に関する研究 [20] では、マイクロブログの投稿内容および投稿時間に関わる情報を解析することで、習慣行動から属性の推定が可能であることを証明した。さらに、属性を考慮した行動推定手法とユーザの性別、年代と職業といった属性を段階的に推定することの有効性を証明 [21] した。これらの研究により、ユーザ特性を用いたソーシャルセンシングの基盤を構築できた状況である。

本研究では、このソーシャルセンシングの基盤の中で、異常行動を起こすユーザを抽出することで、キーワードに依存せずに、社会事象を抽出する新たな手法の開発に取り組む。なお、異常行動を一定の生活パターンで活動しているユーザにおいて、地震や台風などの災害により変化した急な行動の変化、イベントやビックニュースなどの大衆が注目する事象への反応行動、事件や事故に巻き込まれた際の行動と設定している。また、ツイッターの心理学 [22] では、Tweet をする理由を「自分の感想やニュースに関する事実を知ってほしくて共有したい、伝えたいときに投稿する傾向にある」と説明している。このことから社会事象が起きた際の習慣行動ごとの投稿パターンは、平時の行動での投稿パターンと異なる可能性が高いと考えられる。著者らは、これらの行動の変化に着目して解析することで、異常行動から何らかの大きな社会事象を把握することが可能であるという仮説を設定し、新たなソーシャルセンシング手法を開発する。

### 2.2 研究課題の設定

本研究では、次に示す2つの研究課題を設定し、これらを明らかにすることで、提案のソーシャルセンサの有用性を確認する。

#### 2.2.1 研究課題1：平時と異なる行動を起こすユーザ群を特定してその投稿を解析することで、社会事象を抽出可能であること

現状のソーシャルセンシングに関する研究では、主に特定のキーワードの出現数や文脈を解析して、現象を検知する手法が利用されている。しかし、これらの手法は、現象ごとに特定のキーワードを事前に指定するため、多種多様な現象を広範囲に把握することが困難である。既存研究においても、地震やスポーツイベント、交通ネットワークなどの特定の現象を対象に抽出しているものが多い。

本研究では、複数のユーザの習慣行動から異常行動を抽出することにより、実世界における現象を抽出する手法を

提案する．これにより，キーワードの出現数や文脈のみに頼らないソーシャルセンシング手法を開発し，本仮説を証明する．

### 2.2.2 研究課題 2：ソーシャルセンサの特性としてユーザーの生活習慣を考慮することで，抽出可能な社会事象の粒度やカテゴリが変化すること

提案手法では，ユーザーの生活習慣の変化を用いてキーワードの出現数や，文脈では抽出できない現象の抽出ができると考えられる．また，習慣行動と抽出したトピックの関係を分析することで，行動ごとに関心の高い社会事象のカテゴリ（生活，エンタメ，スポーツ，政治・経済，その他など）が変化すると考えられる．たとえば，睡眠前には，一日の出来事，娯楽情報やエンタメ情報，帰宅時はスポーツや政治・経済に関する情報などである．また，習慣行動ごとに抽出することで，社会事象の起きた時間帯の行動が変化するため，トピックのキーワードが詳細に把握でき，粒度（トピックの内容の詳細度）が変化すると考えられる．

本研究では，ユーザーの習慣行動に着目し，行動の変化を抽出する．そして，異常行動をした期間の投稿を収集してトピック単位に分類する．その後，実世界で発生した特定の現象を抽出し，平時の現象と比較し，本仮説を証明する．

## 3. ユーザーの習慣行動の変化を用いたソーシャルセンシング技術

### 3.1 提案手法

ソーシャルセンシングとは，マイクロブログなどのソーシャルメディア上での利用者をソーシャルセンサととらえ，実世界の現象を観測する方法である．ソーシャルセンサは，抽出可能な現象の範囲が広く，解析データの取得が容易といった特徴がある．本研究では，その中でもリアルタイム性に優れており，投稿される情報量が多い Twitter を採用する．

提案手法の処理フローを図 1 に示す．処理フローに示すとおり，本手法では，「ユーザーの習慣行動の定量化機能」と「平時の習慣行動と異なる行動の検出機能」により構成される．

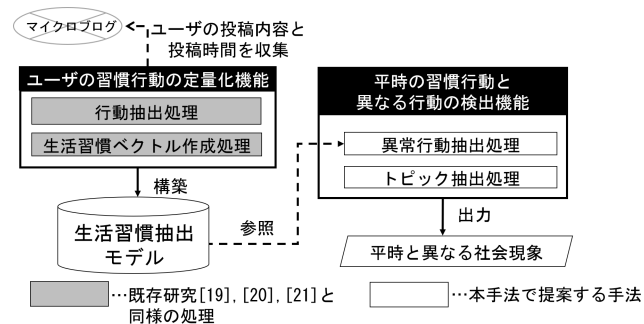


図 1 処理フロー

Fig. 1 Flowchart of process.

### 3.2 ユーザーの習慣行動の定量化機能

本機能では，ユーザーの異常行動を抽出するために用いる生活習慣ベクトルを算出する．生活習慣ベクトルは，ユーザーの投稿傾向を分析し，平常時の習慣行動を定量化して表現したものである．生活習慣ベクトルの算出手順を次に示す．

#### 3.2.1 行動抽出処理

本処理では，習慣行動を分析するために投稿履歴を解析し，各投稿からユーザーの行動情報を抽出する．本研究では，既存研究 [19] の行動抽出処理で用いた NTT コミュニケーション科学基礎研究所が作成した日本語語彙体系 [23] を用いる．習慣行動として，既存研究 [21] を参考に「その他」の項目を「在宅」に絞り，「起床・就寝」，「出勤」，「在宅」と「帰宅」の 4 種類を習慣行動として採用する．なお，生活習慣ベクトルを構成する素性には，あらかじめ構築した行動辞書に登録されている用語を使用する．行動辞書には，日本語語彙大系を参考にして，手作業で行動に関連する用語を習慣行動ごとに選定したものを登録する．行動辞書に登録した用語の例を表 1 に示す．詳細は，既存研究 [19], [20], [21] を参照されたい．

#### 3.2.2 生活習慣ベクトル作成処理

生活習慣ベクトル作成処理では，生活習慣に関する単語の時間ごとの出現回数を示すベクトルを作成する．生活習慣ベクトルはユーザーの習慣行動を素性としたベクトルであり，4次元（習慣行動）×7次元（曜日）×24次元（時間帯）の 672次元で構成する．本研究では，1年間の行動を平時の行動とし，比較は，1か月ごとに行うこととした．理由としては，1年間の行動を平時の行動とすることで，1年間に起きた社会事象を考慮した行動となるため，年ごとに特徴があると考えたからである．また，社会事象は，日付単位や週単位で変化する可能性がある．本研究では，異常行動に着目して社会事象を抽出できるかを目的としたため，月単位で比較する．

年間の行動  $beh_x$  における各曜日の時間帯  $h$  の生活習慣ベクトル  $YearPost(beh_x)$  を式 (1) に示す．

$$YearPost(beh_x) = \{YPost_{beh_1(0)}, YPost_{beh_1(1)}, \dots, YPost_{beh_x(h)}\} \quad (1)$$

式 (1) において， $h$  は，7次元（曜日）×24次元（時間帯）の 1 時間を表す． $h = 0$  の場合，年間の日曜日の 0 時

表 1 行動辞書に登録した用語の例

Table 1 Example of terms on behavior dictionary.

行動	用語
起床・就寝	おはよう，寝る，就寝，おやすみ
在宅	風呂，テレビ，洗濯，支度，掃除
出勤	出勤，通勤，通学，行ってきます
帰宅	帰宅，帰る，退勤，退社，下校

を示す.  $x$  は, 各習慣行動を表しているため, 「起床・就寝」, 「出勤」, 「在宅」と「帰宅」であり, 最大は4である.  $YPost_{beh_x(h)}$  において  $h = 0$  の場合, 年間の日曜日の0時00分00秒から0時59分59秒までの間に生活習慣  $beh_x$  に関連する単語を含む投稿がなされた回数を表す.  $h = 167$  の場合, 土曜日の23時00分00秒から23時59分59秒までを示している. 生活習慣ベクトルを式(1)により作成後, 式(2)により習慣行動ごとに正規化を行った.

$$YearPost(beh_x)' = \frac{YearPost(beh_x) - MinYearPost(beh_x)}{MaxYearPost(beh_x) - MinYearPost(beh_x)} \quad (2)$$

式(2)において,  $MaxYearPost(beh_x)$  は, 日曜日から土曜日までの最大値の投稿数,  $MinYearPost(beh_x)$  は, 日曜日から土曜日までの最低値の投稿数を示す. また, 本研究では, 同様に月ごとでも生活習慣ベクトルを算出する. 各月の行動  $beh_x$  における各曜日の時間帯  $h$  の生活習慣ベクトル  $MonthPost(beh_x)$  を式(3)に示す.

$$MonthPost(beh_x) = \{MPost_{1,beh_x(0)}, MPost_{1,beh_x(1)}, \dots, MPost_{m,beh_x(h)}\} \quad (3)$$

式(3)において,  $m$  は月を表し,  $MPost_{1,beh_x(0)}$  の場合, 1月の日曜日の0時00分00秒から0時59分59秒までの間に生活習慣  $beh_x$  に関連する単語を含む投稿がなされた回数を表す. 式(3)により, 生活習慣ベクトルを作成後, 式(2)と同様に日曜日から土曜日までを算出し, 習慣行動ごとに正規化を行った.

### 3.3 平時の習慣行動と異なる行動の検出機能

本機能では, 式(1)で示した平時のユーザの習慣行動を表す生活習慣ベクトル(以下, 平時習慣ベクトル)と式(3)で示した月ごとの習慣行動を表す生活習慣ベクトル(以下, 特定習慣ベクトル)とを比較し, 時間帯ごとの差分を抽出する. これにより, 平常時と異なる行動を起こすユーザが多い曜日・時間帯を明らかにする. そして, その時間帯に発生している社会事象を平時と異なる現象として抽出する.

#### 3.3.1 異常行動抽出処理

本処理では, 任意の時間帯における平時習慣ベクトルと特定習慣ベクトルの値を比較し, 時間帯ごとに差分を抽出する.  $beh_x$  に関する任意の時間帯  $h$  の平时習慣ベクトル  $YPost_{beh_x(h)}$  と特定習慣ベクトル  $MPost_{m,beh_x(h)}$  における差分  $S$  は, 式(4)で算出する.

$$S(m, h) = \frac{\sum_{k=h-\alpha}^{h+\alpha} |YPost_{beh_x(k)} - MPost_{m,beh_x(k)}|}{2\alpha + 1} \quad (4)$$

式(4)において,  $\alpha$  は求める時間  $h$  の前後の時間数を示

す. 本研究では,  $\alpha$  の値を3時間と設定した. 前後の時間を考慮する理由としては, 短時間の異常投稿による影響を少なくするためである. このことから, 任意の時間  $h$  における差分は, 合計した時間数  $2\alpha + 1$  で除算することで平均値とした. なお, 差分が, マイナス値にならないよう絶対値で計算を行う.

本研究では, 算出した  $S(m, h)$  が一定の閾値を超える場合を異常行動として定義する. 異常行動の判断式を式(5)に示す.

$$S(m, h) \geq \frac{\sum_{k=0}^{167} S(m, k)}{168} \quad (5)$$

式(5)において, 異常行動と判断するときの閾値には, すべての時間帯における平时習慣ベクトルと特定習慣ベクトルの差分の平均値を用いる. 一定の投稿パターンで行われている平常時の習慣行動との差分が大きい箇所は, 何らかの社会事象が発生していると考えられることから, 習慣行動ごとに式(5)により判断することで異常行動を抽出できる.

#### 3.3.2 トピック抽出処理

本処理では, Bleiら[24]によって提案された潜在的ディリクレ配分法(LDA: Latent Dirichlet Allocation)を用いて, トピック(話題)を抽出する. 潜在的ディリクレ配分法とは, 文書に出現する単語に存在するトピックの関係を確率的に表したトピックモデルの1つである. 3.3.1項で抽出した特定の期間と月全体のトピックを比較し, 社会事象を抽出する. 本研究では, Pythonのトピックモデルライブラリであるgensimを用いて, トピックを抽出する. トピックの生成課程を次に示す.

STEP 1: 3.3.1項で抽出した各行動の期間と月全体の投稿に対しMeCabを用いて, 形態素解析を行う.

STEP 2: STEP 1で求めた形態素から名詞のみを用いて, gensimを用いて特徴語の辞書を作成する. そこでは記号やアルファベット一文字などをStopWordと定義し, 除外する.

STEP 3: STEP2で作成した辞書をもとにトピックを抽出する.

## 4. 評価実験

### 4.1 実験内容

#### 4.1.1 実験概要

本実験では, 2章で設定した「異常行動を起こすユーザ群を特定してその投稿を解析することで, 社会事象を抽出可能であること」と「ソーシャルセンサの特性としてユーザの生活習慣を考慮することで, 抽出可能な社会事象の粒度やカテゴリが変化すること」の2つの研究課題に対し, 本提案のソーシャルセンシング技術が有用であることを検証する. 本研究の実験計画を図2に示す. 図2は, 評価実験により検証項目を明確化するため, 図1と対応関係を

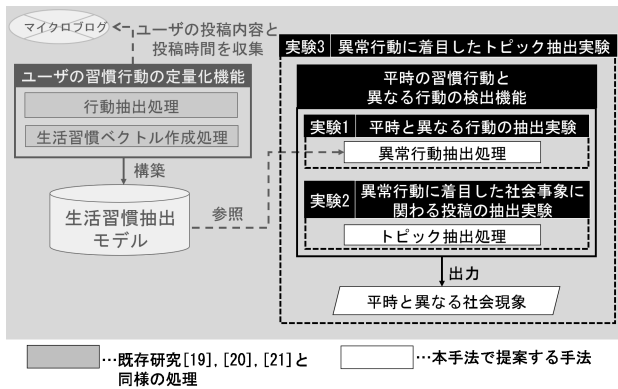


図 2 実験計画  
Fig. 2 Plans of experimentation.

図示したものである。

実験 1 では、研究課題 1 の前提として、平時習慣ベクトルと特定習慣ベクトルをもとに異常行動の時間帯を算出可能か検証する。実験 2 では、Twitter から取得できる社会事象を明らかにし、抽出対象の社会事象を選定する。実験 3 では、対象の期間の全投稿のトピック、キーワードを指定した投稿のトピックと抽出した異常行動の時間帯のトピックとを比較することで、社会事象を取得可能であるかを検証し、本提案手法の有効性を証明する。

#### 4.1.2 実験データ

本実験では、2014 年 1 月～12 月の投稿群を対象とする。実験データの収集方法を次に示す。

STEP 1: TwitterAPI を用いて 2014 年 1 月～12 月に投稿しているユーザおよび投稿内容を取得する。

STEP 2: Twilog [25] を解析して、2014 年 1 月～12 月に投稿しているユーザを抽出し、その投稿内容を取得する。Twilog とは、Twitter に投稿された内容をユーザごとにブログ形式で保存するサービスである。Twilog から取得したユーザが STEP 1 のユーザと重複している場合は、Twilog のデータを優先して採用する。理由としては、TwitterAPI では、投稿内容の取得数制限のため、最大で 3,200 件であるのに対して、Twilog はユーザの全投稿を抽出できるため、投稿数が多いからである。

STEP 3: STEP 1 および STEP 2 で収集した投稿内容の件数が 1,000 件以上のユーザを実験データとする。ただし、ライフスタイルの解析に最低限必要な 1 週間分の投稿内容を取得できないユーザは、実験データから除外する。ユーザにより投稿数に差が出る場合でも習慣行動は取得可能であるため、実験に支障はないものと考えられる。

以上の手順で収集した実験データの詳細は、ユーザ数が 1,440 ユーザ、投稿件数は、1,814,476 件である。なお、投稿件数は TwitterAPI および Twilog の仕様に基づき収集したツイートの件数を示している。

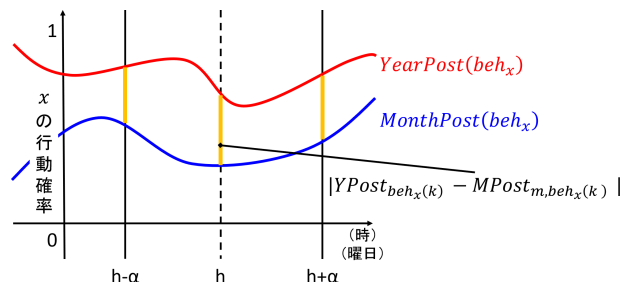


図 3  $h$  と  $\alpha$  の関係性  
Fig. 3 Relationship of  $h$  and  $\alpha$ .

#### 4.1.3 実験条件

本実験では、異常行動算出に関わるパラメータ  $\alpha$  とトピック抽出で用いる LDA の手法でトピック数を設定する。

##### (1) パラメータ $\alpha$

パラメータ  $\alpha$  は、異常行動抽出処理において、平时習慣ベクトルと特定習慣ベクトルの差分を算出するとき用いる値である。式 (4) において、 $h$  は求める対象の時間帯、パラメータ  $\alpha$  は  $h$  を算出する際に考慮する前後の時間を示す。 $\alpha$  の値が大きくなると、個々の時間の情報を読み取りにくくなる。 $h$  と  $\alpha$  の関係を図 3 に示す。図 3 において、赤いグラフが平时習慣ベクトル  $YearPost(beh_x)$ 、青いグラフが月ごとの特定習慣ベクトル  $MonthPost(beh_x)$ 、黄色い箇所が、 $|YPost_{beh_x(k)} - MPost_{m,beh_x(k)}|$  の差分を指す。本実験では、トピック抽出が可能かどうかを検証することを目的とし、1 日のサイクルは 0 時～6 時、6 時～12 時、12 時～18 時と 18 時～24 時の 6 時間ごとにあると考えたため、 $\alpha$  の値を 3 に設定した。

##### (2) トピック数

LDA の手法で用いるトピック数は既存研究 [26] に倣い、100 トピックとした。なお、習慣行動によって投稿件数が増えるため、全体の投稿件数を 100 トピックとした場合、各行動の投稿件数を割合でトピック数を算出し、パラメータとして設定した。たとえば、全体の投稿件数が 10,000 件、各行動の投稿件数が 1,000 件の場合、全体では 100 トピックに対し、各行動では 10 トピックとしている。これにより、全体と各行動の 1 トピックに対する投稿件数の比率が同様となるため、同様の詳細度のトピックを抽出できる。また、反復回数は 50 回とした。

#### 4.2 実験 1 平時と異なる行動の抽出実験

##### 4.2.1 実験概要

本実験では、2.1 節であげた「研究課題 1: 平時と異なる行動を起こすユーザ群を特定してその投稿を解析することで、社会事象を抽出可能であること」を証明するため、平时習慣ベクトルと特定習慣ベクトルをもとに異常行動の時間帯を算出し、その異常行動から社会事象を抽出する。

##### 4.2.2 実験手順

実験では、次の手順に従って分析を実施する。

表 2 各月の異常行動数  
Table 2 Number of irregular actions in each month.

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	平均	標準偏差
起床・就寝	73	74	62	67	66	<b>86</b>	79	73	69	70	75	85	73.3	6.98
在宅	81	84	79	69	<b>86</b>	<u>66</u>	78	81	78	76	<b>86</b>	70	77.8	6.30
出勤	66	<b>81</b>	74	59	73	70	75	73	<u>54</u>	65	74	59	68.6	7.70
帰宅	71	67	75	<b>77</b>	68	71	74	68	67	68	69	<u>62</u>	69.8	3.94
平均	73	<b>77</b>	73	68	73	73	<b>77</b>	74	67	70	76	69	72.5	3.23
標準偏差	5.40	6.58	6.34	6.40	7.79	7.60	2.06	4.66	8.57	4.02	6.20	10.07		

STEP 1: 平時のユーザの生活習慣を明らかにするため、2014年1月~12月の全投稿群を対象に平時習慣ベクトルを作成し、習慣行動ごとに分類する。  
 STEP 2: 2014年の各月で特定習慣ベクトルを作成し、習慣行動ごとに分類する。  
 STEP 3: STEP 1で作成した平時習慣ベクトルとSTEP 2で作成した12カ月分の各特定習慣ベクトルを比較し、異常行動が発生した時間帯を抽出する。  
 STEP 4: 異常行動が発生した時間帯を分析し、その際に発生した社会事象を分析する。

4.2.3 実験結果

各行動の異常行動の抽出時間数を表 2 に示す。各行動で最も多い月は太字、低い月は下線で示す。また、各月の平均の上位を太字の網掛けで示す。表 2 において、異常行動の抽出時間が、最小数と最大数の平時習慣ベクトルと特定習慣ベクトルの比較結果を図 4 に示す。図 4 は縦軸が行動確率、横軸は各曜日の時間を表しており、抽出した異常行動を赤枠で示す。在宅は、6月と5月を図 4 に示す。表 2 と図 4 を分析した結果、明らかになった内容を次に示す。

- 習慣行動から異常行動を抽出可能であることが分かった  
 図 4 の各行動に関して、平時習慣ベクトルと特定習慣ベクトルを確認すると、差分が大きい箇所が異常行動として抽出されていることが分かる。また、前後 3 時間の行動確率を考慮して、異常行動を算出しているため、出勤の水曜日や金曜日などの突発的な異常行動も抽出できていることが分かる。しかし、一見差がみられる箇所が抽出できていないことが分かった。たとえば、在宅の 23 時から 0 時ごろ、出勤の月曜日 9 時~11 時ごろなどが抽出できていないことが分かる。要因として、1 週間を考慮して算出したため、異常行動として抽出できていないことが分かった。これは、異常行動の算出方法を時間帯だけでなく、テキスト内容も考慮することにより正確に算出できると考えられる。
- 生活習慣には一定のリズムがある  
 生活習慣ベクトルを確認すると、習慣行動には一定のパターンがあることが分かった。各行動の平時習慣ベクトルと特定習慣ベクトルを確認すると、各行動を行っている

考えられる時間帯は、行動確率が高くなり、行っていない時間帯は低い傾向があることが分かる。また、そのパターンから外れる箇所が異常行動として抽出できており、異常行動を抽出できていることが分かる。図 4 の起床・就寝では、2 カ月とも同じパターンで、行動を繰り返していることが分かる。出勤についても同じパターンを示していることが分かる。これにより、一般的な生活習慣のパターンを取れていることが分かった。

在宅に関しては、平日と休日あまり差がみられなかった。これは、在宅として取得している投稿は、風呂や掃除など日常的に行う行動であるため、差がみられないと考えられる。また、休日の出勤が高い理由として、今回のユーザは属性を指定していなかったためであると考えられる。大きな要因としては、休日の会社員や自営業のユーザは休みをとっており、学生はアルバイトなどで平日の時間帯と異なる行動パターンになるからであると考えられる。また、一部のユーザが休日にもかかわらず出勤するため、投稿する傾向が高いためであると考えられる。これは、属性も考慮することや出勤の行動辞書の検討などを行うことで、緩和できると考えられる。

図 4 の 5 月の木曜 9 時~10 時の在宅、9 月の木曜日 7 時~10 時の出勤など、行動確率の差が高いにもかかわらず異常行動として抽出できていないことが分かる。これは、3.3.1 項の異常行動抽出処理で示す内容で、異常行動抽出をしている。そのため、各月の各行動で閾値が変化しており、突発的に増加して減少する行動が抽出できていないからである。これは、閾値の算出方法を変えることで対応できると考えられる。

- 異常行動は、各習慣行動に関連して発生する傾向がみられる  
 提案手法では、異常行動の時間帯を月ごとに抽出できることが分かった。表 2 より各行動の中で、最も異常行動の多い月は起床・就寝で 6 月、在宅で 5 月と 11 月、出勤で 2 月、帰宅で 4 月となった。最も低い月は起床・就寝で 3 月、在宅で 6 月、出勤で 9 月、帰宅で 12 月となった。  
 また、2014 年の大きなトピックとして、2 月はソチオリンピック、7 月は集団的自衛権の行使の閣議決定や台風による

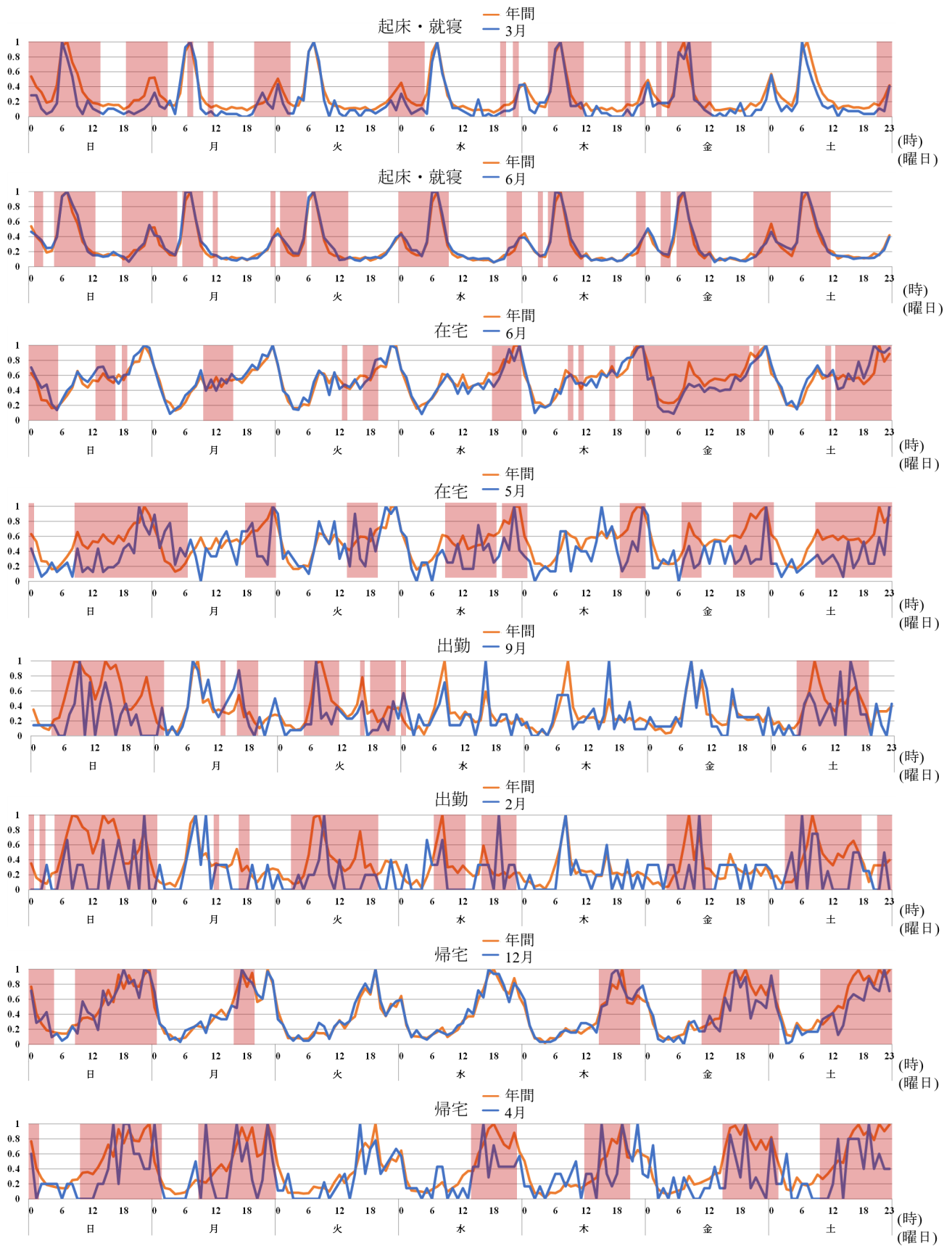


図 4 上位と下位の生活習慣解析結果 (縦軸：行動確率)

Fig. 4 Lifestyle analysis results on top and bottom.

被害など平常時と異なる関心度の高い事象や災害があったことからこのような結果が出たと考えられる。実際に投稿を確認すると「あー心臓バクバク！凄い緊張ですね！！！神様お願い！羽生くんに金！」、「安倍内閣の憲法改正発議、集团的自衛権行使の法制には賛成を投じられたら終わり。ここは非自公+政党助成金目当ての衛星政党でいくしかない。」、「竜巻かと思うほどの風の後に、土砂降りの雨！びっくりするほどすごい！」、などのトピックの状況や感想の投稿が多いことが分かった。

表 2 より、どのユーザも基本的に出勤時間は変わらず、一定のリズムであると考えられる。しかし、2月に最も高い 81 時間という値を示していることが分かる。これは、2月は記録的な大雪に見舞われ公共交通が不安定になったため、異常行動が増えたと考えられる。次いで高い時間を示す 7月 は、台風 8 号の影響により初の特別警報や土砂災害など被害をもたらしたためであると考えられる。

各行動の平均が最も高い在宅に関しては、掃除、テレビの視聴や入浴など日常生活において、必ずしも決まった時間に行われない行動であり、出勤とは逆に一定のリズムが崩れることが多いためであると考えられる。

### 4.3 実験 2 異常行動に着目した社会事象に関わる投稿の抽出実験

#### 4.3.1 実験概要

本実験では、異常行動に着目して社会事象に関わる投稿を適切に抽出可能であることを明らかにするため、任意の月に発生した社会事象の正解データを対象に評価実験する。なお、本実験で対象とする期間は、実験 1 の結果より異常行動の期間が多くみられた 2014 年 2 月と 7 月とする。

#### 4.3.2 実験手順

本実験の手順を次に示す。

STEP 1: 実験対象期間である 2014 年 2 月と 7 月の社会事象を Wikipedia (2014 年の日本) から取得する。

STEP 2: STEP1 で取得した社会事象に関わる投稿を 4.1 節で作成した実験データから抽出し、正解データとして蓄積する。

STEP 3: STEP2 で作成した正解データを対象に、提案手法で各行動に関わる投稿を抽出し、その件数を集計する。

#### 4.3.3 実験結果

実験の結果を表 3 に示す。表 3 より次に示すことが分かった。なお、表 3 では、社会事象に関わる投稿の全体件数および、提案手法の各行動において抽出された投稿件数を指す。なお、全体件数が 0 件の社会事象は除外している。

- 異常行動に着目することで、社会事象に関わる投稿を取得できることが分かった

表 3 を確認すると、社会現象に関わる投稿がないもしくは、きわめて少ない場合を除き、その事象に関わる投稿を

抽出できていることが明らかとなった。また、抽出されている社会事象を確認すると、「ソチオリンピック」や「大雪」などの習慣行動の時間のズレに影響を与える事象や、「STAP 細胞関連論文撤回」や「北朝鮮経済制裁」など、投稿量に影響を与える事象も取得できていることが明らかとなった。

- 習慣行動ごとに社会事象に対する関心が異なることが分かった

表 3 を確認すると、異常行動として抽出された各行動の時間帯の社会事象の投稿量を確認するとソチオリンピックは起床・就寝時に最も多く、大雪では、出勤や帰宅時に多い傾向があることが分かった。これにより、抽出したい社会事象によって行動を変えることにより効率良く社会事象が抽出できると考えられる。

### 4.4 実験 3 異常行動に着目したトピック抽出実験

#### 4.4.1 実験概要

本実験では、2.1 節であげた「ソーシャルセンサの特性としてユーザの生活習慣を考慮することで、抽出可能な社会事象の粒度やカテゴリが変化すること」を証明するため、異常時と抽出された期間の投稿をトピック分類した結果に基づき解析する。表 2 より本実験では、異常行動の時間が最大数の 2 月と 7 月を対象に解析する。

#### 4.4.2 実験手順

実験では、次の手順に従って分析を実施する。

STEP 1: 2 月と 7 月における異常行動とされた期間の投稿を抽出し、それらの投稿を各月の各習慣行動に分類する。

STEP 2: 2 月と 7 月の全投稿を抽出しそれらの投稿を月ごとに分類する。

STEP 3: STEP1 と STEP2 で分類した投稿を提案手法のトピックと既存手法のトピックとして分類する。

STEP 4: 各月の既存手法と提案手法のトピックから、それぞれトピック内の単語を取得し、対応する期間にニュースやイベントがあるかを検索し、該当トピックをノイズに分類する。

STEP 5: 各月の既存手法と提案手法のトピックを構成する単語を比較し、5 個以上の単語が同一であった場合、一致するトピックに分類する。

以上の処理手順でトピックを分類する。以上の処理のフローチャートを図 5 に示す。

#### 4.4.3 実験結果

各月の各行動の一致率を表 4 に示す。本実験では、2014 年の 2 月と 7 月各行動で算出したトピックから実世界で起きた社会事象を抽出した。

- 既存手法と比較して、トピックが異なる傾向がみられた表 4 の既存手法の正常トピック数と提案手法の合計の正常トピック数を比較すると、提案手法の方が多くのトピッ



表 3 社会事象に関する投稿の抽出結果

Table 3 Results of extracting tweet related to social events.

月	社会事象	全体 件数	提案手法の件数			
			起床・就寝	出勤	在宅	帰宅
2 月	ソチオリンピック	505	299	275	210	245
	大雪	504	264	308	246	276
	東京都知事選	308	93	99	92	98
	PS4 発売	78	24	20	19	22
	長崎知事選挙, 岐阜市長選挙	17	7	9	5	5
	大阪都構想	6	3	3	2	2
	架空の国産キノコ販売事業に投資問題	19	9	12	10	10
	佐村河内守氏ゴーストライター問題	22	9	14	8	9
	JR 北海道レール検査改竄問題	4	2	0	0	0
	尖閣諸島中国漁船衝突事件で那覇地方裁判所に提訴	2	1	0	0	1
	元住吉駅下りホームで列車衝突	19	5	2	1	3
	山口県知事選挙	1	1	1	1	1
	TPP 協定交渉	24	15	7	14	16
7 月	集団的自衛権の行使	382	183	152	218	149
	野々村議員政務活動費不正疑惑	162	75	65	84	89
	STAP 細胞関連論文撤回	112	41	30	57	45
	北朝鮮経済制裁	57	35	1	30	25
	沖縄県初の特別警報	194	64	94	127	96
	ベネッセ個人情報流出	194	25	35	109	84
	オーストラリアと経済連携協定連携協定	30	15	15	16	13
	舞鶴高一女子殺害事件無罪確定	9	2	1	6	2
	バブアニューギニア政府開発援助	3	1	2	2	1
	滋賀県知事選挙	20	10	8	14	15
	川内原子力発電所審査合格	9	4	4	1	7
	直木賞	7	1	4	6	2
	嫡出推定	1	0	0	0	0
	生活保護法外国人は適用外	8	8	2	6	7

クを抽出できていることが分かった。その一方で、提案手法は、ノイズに分類されるトピックも多くみられ、トピック抽出後のフィルタリング方策を検討する必要があることが分かった。

表 4 のトピック抽出結果を確認すると、既存手法と提案手法のトピックの非一致数が多いことが分かった。これは、それぞれの手法で、抽出されたトピックが異なることを示している。

● 提案手法では行動ごとに抽出できるトピックが異なる傾向がみられた

習慣行動と抽出トピックとの関係を分析するため、抽出トピックをカテゴリ分類した結果を表 5 に示す。表 5 において、太字下線は、各カテゴリの最大値を示す。本研究

では、生活、エンタメ、スポーツ、政治・経済とその他のカテゴリに分類した。カテゴリ分けに関しては、解析者のバイアスを緩和させるため Yahoo!カテゴリを参考に複数人で行った。

表 5 を確認すると、行動ごとに抽出できるトピックが変化することが分かった。生活のカテゴリは在宅と出勤の行動が高い割合を示している。エンタメのカテゴリは、起床・就寝の行動で高い割合を示している。アニメに関するトピックをエンタメに分類し、深夜のアニメの投稿が目立ったためこのような結果になったと考えられる。政治・経済のカテゴリは、帰宅と出勤の行動で高い割合を示している。このように、習慣行動ごとに抽出されるトピックは異なっている傾向がみられた。

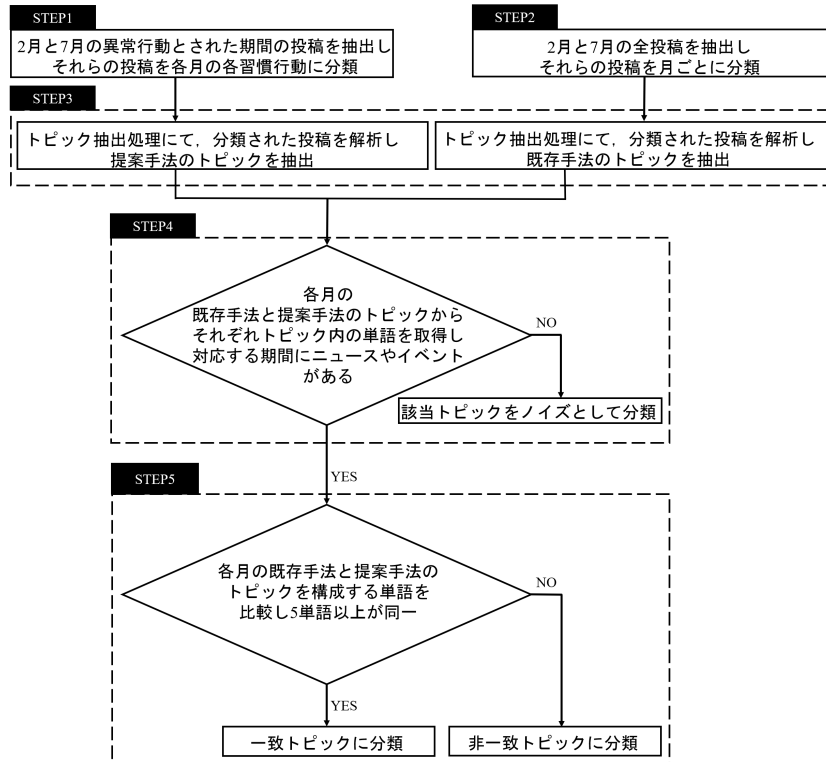


図 5 トピック分類の処理手順

Fig. 5 Processing procedure of topic classification.

表 4 各月の各行動の一致率

Table 4 Rate of concordance of each action in each month.

			既存手法	提案手法				合計
				起床・就寝	在宅	出勤	帰宅	
2月	正常	一致	17(17%)	6(13%)	4(6%)	11(21%)	10(19%)	31(14%)
		非一致	54(54%)	16(33%)	40(63%)	25(48%)	29(56%)	110(51%)
	ノイズ	29(29%)	26(54%)	26(31%)	19(31%)	16(25%)	74(35%)	
7月	正常	一致	11(11%)	6(14%)	6(11%)	6(11%)	3(6%)	21(11%)
		非一致	64(64%)	21(50%)	21(40%)	11(20%)	14(30%)	67(34%)
	ノイズ	25(25%)	15(36%)	26(49%)	37(69%)	30(64%)	108(55%)	

表 5 トピックのカテゴリ

Table 5 Topic category.

	生活	エンタメ	スポーツ	政治・経済	その他	標準偏差
起床・就寝	7(14%)	<b>15(31%)</b>	3(6%)	12(24%)	12(24%)	4.26
在宅	<b>22(33%)</b>	9(14%)	<b>4(6%)</b>	17(26%)	<b>14(21%)</b>	6.24
出勤	18(36%)	2(4%)	3(6%)	18(36%)	9(18%)	6.96
帰宅	12(21%)	11(20%)	3(5%)	<b>19(34%)</b>	11(20%)	5.08
標準偏差	5.72	4.71	0.43	2.69	1.80	

- 既存手法と提案手法とで抽出できるトピックの粒度が異なる傾向がみられた

表 4 において、既存手法と提案手法の合計の一致トピック数を比較すると、提案手法の件数が多くなっていることが分かる。これは、既存手法の 1 つのトピックに対して、

提案手法では、複数のトピックに分割して抽出していることを示している。一致内容の詳細を確認するため、各月において、一致したトピックの対応関係をまとめた結果と網羅性の確認をするため、事前にキーワードを指定し、絞った投稿内容に対して LDA によりトピック抽出した 2 月と

表 6 社会事象に関するトピックの抽出結果①

Table 6 Results of extracting topics related to social events.

月	トピック	指定 キーワード	既存手法	提案手法	
			解釈トピック (キーワード)	行動	解釈トピック (キーワード)
2 月	ソチ オリン ピック	指定なし	オリンピック・登場・話題・羽生	起床・ 就寝	ソチオリンピックを都知事が応援 (都知事, 応援, オリンピック, ソチ)
		オリンピ ック, ソチ, 羽 生, 五輪	羽生選手男子フィギュアで金メダル (オリンピック, ソチ, 羽生, 金メダル, 男子, フィギュア) 浅田選手や高橋選手の活躍 (感動, 真央, 高橋)	在宅  出勤	ソチオリンピック高校生や大学生の活躍 (大学, 高校生, オリンピック, 発見, ソ チ)  浅田真央の活躍 (選手, オリンピック, 感 動, 埼玉, 彼女, 五, ソチ)
	大雪	指定なし	大雪・長期・熊谷, 東京・大阪・雪だるま	在宅	静岡県久しぶりの除雪 (久しぶり, 31, 除雪, 記録, 静岡, 配信, 視聴) 千葉で去年の倍の雪 (心配, 千葉, 大雪, NHK, ぶり, 去年, 倍)
		積雪, 大雪, 関東, 甲信, 仙台	東京, 千葉など各地で大雪 (大雪, 積雪, 埼玉, 新潟, 千葉), (関東, 立ち往生, 東京), (大雪, 大変, 山梨)		
	東京都知 事選	指定なし	田母神・都知事・演説	出勤	舛添氏が都知事選に関わった (都知事, 東京, 舛添)
		東京都知 事選, 舛添	都知事選舛添氏を安倍総理が応援し当選 (都知事, 舛添, 選挙, 安倍, 応援, 当選) 田母神氏が都知事選挙に関わった (都知事, 選挙, 田母神, 東京)		
7 月	集团的 自衛権の 行使	指定なし	自衛・権・マジ・閣議	在宅	安倍政権が集团的自衛権の 行使の閣議決定 (集団, 本日, 安倍, 政権, 閣議, 米国), (自衛, 権, 集団, 的), (集団, 行使, 閣議, 決定)
		集团的 自衛権, 臨時閣議, 閣議決定	安倍政権が集团的自衛権の行使の閣議決定 (自衛, 集団, 行使, 安倍, 首相, ), (自衛, 集団, アメリカ, 日本, 容認, 安倍, 閣議), (自衛, 権, 集団, 閣議, 決定, 行使, 安倍)		
	野々村議 員政務活 動費不正 事件	指定なし	議員, 野々村	起床・ 就寝	野々村議員号泣 (反対, 野々村, 説明, 号泣, 選挙, 行為)
		野々村, 政務活動費 不正, 号泣会見	野々村議員号泣会見 (野々村, 議員, 耳, 手, 号泣), (野々村, 議員, 西宮, 辞職)		
	STAP 細胞 関連論文 撤回	指定なし	—	—	—
		ネイチャー, STAP 細胞, 小保方	小保方氏謝罪 (小保, 方, ニュース, NHK, 謝罪, ) 小保方が早稲田の博士学位 (小保, 博士, 論文, 学位, 早稲田), (方, 小保, 氏, 理研, 博士, 論文, 学位, 早大)		
北朝鮮経 済制裁	指定なし	—	—	—	
	経済制裁, 北 朝鮮	北朝鮮ミサイル発射 (北朝鮮, 政治, 発表, 弾道, ミサイル, 発射) 安倍首相が経済制裁発表 (首相, 安倍, 制裁)			

7月の既存手法と提案手法の結果を表6と表7に示す。なお、PS4発売に関しては、既存手法と提案手法ともにトピックを抽出できなかったため、除外している。表6と表7を確認すると、提案手法では、既存手法と比較して、該当のトピックを詳細に表していることが明らかとなっ

た。各トピックについて、既存手法と提案手法とを比較して、明らかとなった内容について次に示す。

ソチオリンピックのトピックは、ロシアで行われた冬季オリンピックである。既存研究は羽生選手がオリンピックで活躍したことが分かる。しかし、提案手法では、都知事

表 7 社会事象に関するトピックの抽出結果②

Table 7 Results of extracting topics related to social events.

月	トピック	指定 キーワード	既存手法		提案手法	
			解釈トピック (キーワード)	行動	解釈トピック (キーワード)	
7 月	台風	指定なし	雨・台風・傘	在宅	宮崎で台風により大雨 (台風, 影響, 大雨, 発生, 宮崎)	
		特別警報, 大雨, 土石流, 台風 8 号	各地大雨警報で被害 (注意報, 警報, 大雨, 洪水, 浸水, 熊谷), (台風, 警報, 大雨, クラス, 最強, 洪水) 沖縄県で特別警報発表 (台風, 警報, 沖縄, 特別, 避難, 最強, クラス, 発表, 暴風雨)			
	ベネッセ 個人情報 流出	指定なし	—	—	—	
	ワールド カップ	ワールド カップ, FIFA	ワールドカップ・アルゼンチン・過去 ワールドカップドイツとアルゼンチンの決勝 (ワールドカップ, 決勝, ドイツ, アルゼンチン) ワールドカップドイツが優勝 (ワールドカップ, ドイツ, 決勝, 優勝)	在宅	ワールドカップのアルゼンチンの試合が NHK で TV 放送 (放送, NHK, アルゼンチン, ワールドカップ, TV)	

表 8 抽出対象以外の習慣行動のトピック

Table 8 Topics of habitual behavior other than the target of extraction.

月	提案手法	
	行動	解釈トピック (キーワード)
2 月	起床・就寝	安倍政権が調査(安倍, 経済, 政府, 調査)
	出勤	原発で問題発生(原発, NHK, 問題, 職員)
	帰宅	横浜と千葉で地震発生(横浜, 発生, 千葉, 地震)
7 月	起床・就寝	原発に関するニュース(原発, ニュース, 経済, 研究, 避難)
	在宅	安倍政権が地方支援を検討(安倍, 政権, 支援, 地方, メンバー)
	出勤	平和記念ドラマ放送(ドラマ, 平和, 70, ぶり)

がオリンピックに応援を行ったこと、高校生や大学生が活躍したことなど既存手法では抽出できなかったことが抽出できていることが分かる。

大雪のトピックは、都心でも記録的な大雪が降った事象である。既存手法からは東京、大阪や熊谷など都市部でも雪が降った事実が確認できる。しかし、提案手法では、千葉で去年の倍の雪が降ったことが確認できる。実際に確認すると約 2 倍の雪が降っていることを確認することができた。

このように雪が降った事実だけでなく、どれくらい降ったのか詳細を抽出することができた。

都知事選のトピックは、2 月 9 日に執行された東京都知事選挙の事象である。既存手法からは田母神氏が都知事選で演説した事実が確認できる。提案手法では、それに加えて、都知事がソチオリンピックを応援したこと、田母神氏

以外に舛添氏が都知事選に関わった事実が確認できる。

表 7 の台風は、7 月に発生した台風の影響より、全国的に大雨をもたらした事象である。既存手法では台風 8 号に関連するキーワードは台風、雨と傘であったが、提案手法では台風、影響、大雨と宮崎といった被害の大きかった地域も抽出できた。実際に確認すると、宮崎県のえびの市で 1 時間に 77 ミリといった大雨があったことが確認できた。

野々村議員政務活動費不正事件は、野々村議員が政務活動費を不正に受給していた事象である。既存手法では野々村議員のキーワードは抽出できたものの、野々村議員が何をしたのかが判断できる単語は抽出できなかった。提案手法では、野々村議員が号泣した事実が抽出できた。

表 6 と表 7 を確認すると、取得できる内容は多少変化するものの指定した場合と比較し、異なる内容を取得することができた。ソチオリンピックの話題に関する既存手法

は、羽生選手の金メダル獲得や浅田選手や高橋選手の活躍が抽出できたのに対し、提案手法では、都知事が応援したことや高校生や大学生が活躍したことが抽出できた。

大雪では、各地で大雪があり、立ち往生があったことに対し、提案手法において静岡県では、久しぶりの除雪や千葉で去年の倍の雪が降ったなど具体的な規模を示す内容が抽出できたことが分かる。

- 社会事象において一般的に周知の事実は抽出が困難であることが分かった

表 6 と表 7 において、既存手法と提案手法を比較すると、提案手法ではソチオリンピックで羽生選手が金メダル獲得したことや東京都知事選で舛添氏が当選した事実などは、抽出が困難であることが分かった。これは 2.1 節で説明した Twitter の投稿の心理上、社会的ニュースについて自分の感想を知ってほしいや伝えたい、事実を共有したいなどの理由により社会事象の内容を投稿するユーザが多く、その情報が詳細化されるため抽出が困難であると考えられる。そのため、一般的に周知の事実を投稿するユーザが少ないためトピックとして抽出できなかったと考えられる。これは、速報内容のみの投稿の取得や Twitter だけでなく他のマイクロブログも用いることで、抽出できるのではないかと考えられる。

提案手法で抽出した抽出対象以外のトピックを表 8 に示す。表 8 に示すとおり、横浜と千葉の地震や平和記念ドラマの放送など既存手法では抽出できないトピックが検出可能であることが分かった。

## 5. おわりに

本研究では、マイクロブログの投稿からユーザ群の異常行動を抽出することで、キーワードに依存せずに社会事象を抽出する新たな手法を提案した。提案手法において、既存研究の課題である「現象ごとに特定のキーワードを事前に指定する必要があるため網羅的な分析が困難であること」、「キーワード選定に解析者のバイアスがかかり分析に偏りがみられること」の課題に対して、異常行動に着目することで、事前にキーワードを指定せずに社会事象を抽出できた。さらに、同一トピックに関しても既存手法と比較して複数の内容が抽出できており、社会事象を詳細に把握できた。これにより、研究課題 1 と研究課題 2 を解消することができた。なお、本提案手法は、異常行動に着目して社会事象を抽出する手法であり、キーワードを指定する必要がないため、習慣行動を特定する表 1 の行動辞書の語句を他言語へ変換することで、日本語以外へも容易に対応可能である。

本研究を通じて、次に示す課題が明らかになった。

- 課題 1：抽出可能なトピックが全国的な社会事象のみであったこと
- 課題 2：投稿を生活習慣のみで絞ったため、トピック内に

判断が難しいキーワードがあったこと

- 課題 3：提案手法で抽出可能な社会現象は、「対象の事象に関わる投稿が Twitter 上でなされていること」と、「各生活パターンにおいて投稿数のずれが生じる内容であること」の 2 つの条件が満たされる必要があること

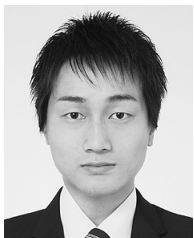
今後は、上述の 3 つの課題に対応する方策を検討しつつ、「平時習慣と特定習慣のタイムスパン（年、月、週、日）の組合せ」と「抽出される社会事象」との関係性を明らかにする予定である。

## 参考文献

- [1] Zhao, Q., Liu, T.Y., Bhowmick, S. and Ma, W.Y.: Event Detection from Evolution of Click-Through Data, *Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.484–493, ACM (2006).
- [2] Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L.: Detecting Influenza Epidemics Using Search Engine Query Data, *Nature*, Vol.457, pp.1012–1014, nature (2009).
- [3] 松尾 豊：ウェブからの実世界の観測と予測，電子情報通信学会論文誌 B, Vol.J96-B, No.12, pp.1309–1315, 電子情報通信学会 (2013).
- [4] 榎 剛史, 松尾 豊：ソーシャルセンサとしての Twitter：ソーシャルセンサは物理センサを凌駕するか？, 人工知能学会誌, Vol.27, No.1, pp.67–74, 人工知能学会 (2012).
- [5] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *Proc. 19th International Conference on World Wide Web*, pp.851–860, ACM (2010).
- [6] 榎 剛史, 松尾 豊, 鳥海不二夫, 篠田孝祐, 栗原 聡, 風間一洋, 野田五十樹：ソーシャルメディアを用いた災害検知及び被災地推定手法の提案, 人工知能学会全国大会論文集, Vol.26, pp.1–4, 人工知能学会 (2012).
- [7] Dingli, A., Mercieca, L., Spina, R. and Galea, M.: Event Detection Using Social Sensors, *Proc. 2nd International Conference on Information and Communication Technologies for Disaster Management*, IEEE (2015).
- [8] Zhao, S., Zhong, L., Wickramasuriya, J. and Vasudevan, V.: Human as Real-Time Sensors of Social and Physical Events: A Case Study of Twitter and Sports Games, Technical Report TR0620-2011, pp.1–9, RICE University and Motorola Labs (2011).
- [9] 富田大志, 道満恵介, 井手一郎, 出口大輔, 村瀬洋：Twitter を用いたスポーツ試合中のイベント検出に関する検討, HCG シンポジウム 2012 論文集, pp.492–498, 電子情報通信学会 (2012).
- [10] 長野伸一：ソーシャルセンサからの情報抽出技術, 東芝レビュー, Vol.69, No.7, pp.19–22, 東芝技術・生産統括部 (2014).
- [11] Georgiou, T., Abbadi, A., Yan, X. and George, J.: Mining Complaints for Traffic-Jam Estimation: A Social Sensor Application, *Proc. 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp.330–335, ACM (2015).
- [12] Congosto, M., Fuentes-Lorenzo, D. and Sanchez, L.: Microbloggers as Sensors for Public Transport Breakdowns, *Proc. IEEE Internet Computing*, Vol.19, No.6, pp.18–25, IEEE (2015).
- [13] Asur, S. and Huberman, B.: Predicting the Future with

Social Media, *Proc. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol.1, pp.492-499, ACM (2010).

- [14] 迫村光秋, 和泉 潔: twitter テキストマイニングによる経済動向分析, 第9回人工知能学会ファイナンスにおける人工知能応用研究会資料, pp.39-41, 人工知能学会 (2012).
- [15] Bollen, J., Mao, H. and Zeng, X.-J.: Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, Vol.2, No.1, pp.1-8, Elsevier (2011).
- [16] Ruiz, J.E., Hristidis, V., Castillo, C., Gionis, A. and Jaimes, A.: Correlating Financial Time Series with Micro-Blogging Activity, *Proc. 5th ACM International Conference on Web Search and Data Mining*, pp.513-522, ACM (2012).
- [17] 荒牧英治, 増川佐知子, 森田瑞樹: Twitter Catches the Flu: 事実性判定を用いたインフルエンザ流行予測, 音声言語情報処理研究会研究報告, Vol.2011-SLP-86, No.1, pp.1-8, 情報処理学会 (2011).
- [18] Lampos, V., Bie, T.D. and Cristianini, N.: Flu Detector - Tracking Epidemics on Twitter, *Proc. ECML-PKDD'10*, pp.599-602 (2010).
- [19] 田中成典, 中村健二, 寺口敏生, 中本聖也, 加藤 諒: マイクロブログから抽出したユーザの習慣に基づく行動推定に関する研究, 情報処理学会論文誌: データベース, Vol.6, No.3, pp.73-89, 情報処理学会 (2013).
- [20] 田中成典, 中村健二, 加藤 諒, 寺口敏生: マイクロブログの投稿時間に着目したユーザの職業推定に関する研究, 情報処理学会論文誌: データベース, Vol.6, No.5, pp.71-84, 情報処理学会 (2013).
- [21] 加藤 諒, 中村健二, 山本雄平, 田中成典, 坂本一磨: マイクロブログにおけるユーザの属性と習慣行動の推定に関する研究, 情報処理学会論文誌, Vol.57, No.5, pp.1421-1435, 情報処理学会 (2016).
- [22] 北村 智, 佐々木裕一, 河井大介: ツイッターの心理学, 誠信書房 (2016).
- [23] 池原 悟, 宮崎正弘, 白井 諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林 良彦: 日本語語彙大系 CD-ROM 版, 岩波書店 (1999).
- [24] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol.3, pp.993-1022, ACM (2003).
- [25] Twilog, available from (<http://twilog.org/>) (accessed 2017-12-20).
- [26] 藤本 拓, 原 隆浩, 西尾章治郎: 時系列の最適平滑化と動的な語彙集合を考慮した時系列文書に対するトピック解析手法, 電子情報通信学会論文誌, Vol.J96-D, No.5, pp.1212-1221, 電子情報通信学会 (2013).



坂本 一磨 (学生会員)

1991年生。2015年関西大学総合情報学部総合情報学科卒業。2018年関西大学大学院総合情報学研究科知識情報学専攻博士課程前期課程修了。修士(情報学)。現在、関西大学大学院総合情報学研究科総合情報学専攻博士課程後期課程在学中。ビックデータ解析, システム設計等の研究開発に従事。



中村 健二 (正会員)

1981年生。2009年関西大学大学院総合情報学研究科総合情報学専攻博士課程後期課程修了。博士(情報学)。現在、大阪経済大学情報社会学部教授。2016年度文部科学大臣表彰科学技術賞「科学技術振興部門」受賞。



山本 雄平 (正会員)

1986年生。2015年関西大学大学院総合情報学研究科総合情報学専攻博士課程後期課程修了。博士(情報学)。現在、関西大学先端科学技術推進機構特別任命准教授。Webマイニング, 自然言語処理, スポーツ情報学に関連する研究に従事。

る研究に従事。



田中 成典 (正会員)

1963年生。1988年関西大学大学院工学研究科土木工学専攻博士課程前期課程修了。博士(工学)。現在、関西大学総合情報学部教授および社会空間情報科学研究センター長。2016年度文部科学大臣表彰科学技術賞「科学技術

振興部門」受賞。