実践研究論文

ユーザ属性を考慮した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実践研究†

坂本 一磨*¹・中村 健二*²・山本 雄平*³・田中 成典*⁴・中村 竜也*¹

CGM（Consumer Generated Media）の普及に伴い、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。これらのデータは、社会動向、経済動向に加えて大規模災害などの多種多様な社会事象を計測するソーシャルセンシング技術の向上に用いられている。著者らは、ユーザの習慣行動を用いて、その習慣行動の違いから実世界における事象を抽出する新たなソーシャルセンシング手法を提案し、有用性を証明した。その技術を適用した社会事象との関係性を実践的に評価するためには、ユーザ属性や異なるデータを用いた汎用性を明らかにする必要がある。そこで、本研究では、ユーザの習慣行動を属性ごとに分析し、ユーザ属性ごとの平時と異なる行動から非習慣行動を抽出する。実証実験では、実世界における社会事象をユーザ属性ごとに詳細な粒度で見出すことが可能かを検証する。

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、ユーザ属性

1. はじめに

近年、CGM（Consumer Generated Media）の普及に伴い、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ている。これらのデータはユーザの性別や職業といった属性推定や情報推薦に関する研究に活用されている。また、社会動向、事件、大規模災害などの多種多様な社会における変化（以下、社会事象）を検知するソーシャルセンシングの技術にも用いられている。ソーシャルセンシングでは、ローカルイベントやゲリラ豪雨をユーザの投稿からリアルタイムかつ速やかに検出できる。ソーシャルセンシングに関する研究は、大きく3つに分けることができる。

1つ目は、検索エンジンの検索履歴を用いた手法である。既存研究 [1, 2] では、検索履歴からイベントの検出や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、解析対象とする検索履歴の入手が難しい課題がある。

2つ目は、ブログを対象とした手法である。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特徴的な単語を用いて事象を予測している。この研究では、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。

3つ目は、データが入手し易くリアルタイム性の高い、マイクロブログを用いる手法である。具体的には、マイクロブログのユーザをソーシャルセンサ [4] として捉え、実世界の事象を解析し、イベントの検出や予測を行っている。マイクロブログを用いる方法では、特徴として、即時性と拡散性に優れていることからソーシャルセンシングの適用が多く行われている。

マイクロブログを対象とした既存の取組みとして、地震や台風といった災害を検知する研究 [5–7]、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 [10–12]、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 [14–16]、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、事象を検知する。しかし、事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ることなどの課題が発生する。そこで、著者らは、既存研究の課題に対応するため、ユーザの生活習慣に応じた行動（以下、習慣行動）を用いて、キーワードに依存しない新たなソーシャルセンシング手法を提案 [19] した。

本研究では、提案手法 [19] を応用して実践的な分析を行う。本手法では、「ユーザ属性ごとに平時と異なる行動（以下、非習慣行動）につながる社会事象は異なる」と、「同様の社会事象であってもユーザ属性に応じて得られる情報粒度が異なる」という仮説を設定し、このユーザ行動の違いを用いて、社会事象の抽出を試み、有用性を証明する。これにより、キーワードの出現数や文脈のみに依存せず、多種多様な事象を見出すことが可能となる。

2. 研究の概要

2.1 研究の位置づけ

本実践研究では、インターネット上に蓄積されているビッ

† Practical Study Regarding Social Sensing Technologies for Extracting Unordinary Phenomena Considering User Attributes with Focus on Different Behavior from Normal Time
Kazuma SAKAMOTO, Kenji NAKAMURA, Yuhei YAMAMOTO, Shigenori TANAKA, and Tatsuya NAKAMURA

*1 関西大学大学院総合情報学研究科
Graduate School of Informatics, Kansai University

*2 大阪経済大学情報社会学部
Faculty of Information Technology and Social Sciences,
Osaka University of Economics

*3 大阪工業大学情報科学部
Faculty of Information Science and Technology,
Osaka Institute of Technology

*4 関西大学総合情報学部
Faculty of Informatics, Kansai University

図1 既存研究と本研究の位置づけ

クデータを対象としたソーシャルセンシング手法を適用する。ソーシャルセンシングに関する既存研究 [1-3,5-18] と本研究との位置づけを図1に示す。従来のソーシャルセンシングでは、社会事象に関するキーワードや属性全体の情報や投稿内容を対象としている。そのため、指定したキーワードのみの事象や性別、年代といったユーザのパーソナル属性を考慮できずに投稿内容のみで抽出を行っている。これにより、解析者のバイアスがかかることや雑多なデータが多くなる課題がある。著者が提案した既存手法 [19] では、ユーザの習慣行動に着目し、平常時の習慣行動の違いから社会事象を抽出する手法を提案した。その手法では、まず、ユーザの習慣行動を「起床・就寝」、「在宅」、「出勤」と「帰宅」の4行動に分類し、一年間の行動を平常時習慣行動の基準とした。次に、各月の習慣行動を基準と比較して、閾値を超える場合に非習慣行動とした。そして、各非習慣行動の時間帯の投稿を Blei らによって提案された潜在的ディリクレ配分法 [20] (LDA: Latent Dirichlet Allocation) を用いて解析し、トピック (話題) を検出した。その結果、キーワードで検出する手法よりも詳細なトピックを取得できることが明らかとなり、提案手法の有用性を証明した。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。しかし、トピックが全国的な社会事象のみで、ユーザの性別や職業などの属性を考慮した詳細な分析ができない課

題があった。

そこで、本研究では、著者らがこれまで行ってきたユーザの属性や習慣行動を推定する研究 [21-23] を応用し、ユーザ属性ごとの習慣行動に着目して、非習慣行動を抽出する手法 [19] を適用する。ユーザの属性や習慣行動を推定する研究では、職業推定に関する手法 [21] とユーザの習慣に基づく行動推定手法 [22] を応用し、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手法が提案されている。そして、その属性を考慮した行動推定手法 [23] において、属性推定や習慣行動の予測に対する課題に有用であることを証明している。その解析方法を用いて、既存研究 [19] を深く追求することを目的とする。方法としては、基準の習慣行動と各属性の各月の習慣行動を比較することで、各属性の非習慣行動を分析し、既存手法 [19] よりも詳細なトピックを抽出することを目指す。これにより、同じオリンピックの話題であっても性別 (男性や女性)、職業 (社会人、主婦など) と地域 (北海道・東北地方、関東地方など) によって注目されるトピックが異なると考えられる。

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、提案手法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: ソーシャルセンサの特性として、ユーザの属性ごとに習慣行動が相違し、非習慣行動の時間帯が異なること

多くの研究 [1-3, 5-18] では、事象ごとに特定のキーワードを事前に指定するため、多種多様な事象を広範囲に把握することが困難であった。これに対し、既存研究 [19] では、複数のユーザの習慣行動から非習慣行動を取り出すことにより、キーワードに依存せず、実世界における事象を抽出する手法を提案した。これにより、キーワードの出現数や文脈に頼らないソーシャルセンシング手法 [19] を開発し、その有用性を証明した。しかし、ユーザの属性を考慮せずに、社会事象を取得したため、注目度の大きいイベントや災害などの社会事象に関する話題に集中した。これは、ユーザ群の属性ごとに、社会事象を抽出することにより、改善できると考えられる。例えば、職業のユーザ属性に着目すると社会人は、朝外出し、夜に帰宅するといった一般的な社会人の特性と、夜に外出して朝に帰宅するなどの夜勤の行動特性がある。学生は、朝から授業があり、夕方はアルバイトを行っていることなどが予測できる。これにより、各属性によって、ニュースや情報収集する時間帯が異なり、社会事象に対して反応する時間帯が相違すると考えられる。本研究では、属性が既知のユーザから習慣行動を算出し、各属性ごとの時間帯で非習慣行動となるかを把握し、本仮説を証明する。

2.2.2 研究課題2: ユーザ属性ごとに反応や興味を持つ社会事象が異なり、同じ社会事象でも抽出内容が異なること

本研究では、ユーザ属性ごとにトピック分析を行うことにより、属性によって反応する社会事象が変化することや同じ社会

図2 処理フロー

事象でも収集できる情報が詳細になることを実証実験から証明する。既存研究 [19] では、起床・就寝時には、エンタメ、帰宅時には政治・経済といった傾向があることを確認した。そのため、習慣行動と収集したトピックの関係を分析することで、各属性の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・経済など）が変化すると考えられる。また、ユーザー属性ごとに習慣行動を分析することで、トピックのキーワードが詳細に把握でき、社会事象が的確に抽出できることや取得できる内容自体が変化すると考えられる。

3. ユーザー属性ごとの習慣行動の相違を用いたソーシャルセンシング技術

3.1 処理フロー

ソーシャルセンシングとは、マイクログ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して捉え、実世界の事象を観測する方法である。ソーシャルセンサは、情報の拡散性と速度が速く、不特定多数のユーザーが情報発信しているため抽出可能な事象の範囲が広いことが特徴である。また、解析データの取得が容易といった利便性がある。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。

本手法の処理フローを図2に示す。本手法は、既存研究 [19] を参考に「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。

3.2 ユーザーの習慣行動の定量化機能

本機能では、ユーザー属性ごとの非習慣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、ユーザー属性ごと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。本研究では、1年間の習慣行動を平時の習慣行動とし、比較対象として、1ヶ月ごとの習慣行動を用いた。平時の習慣行動を1年間とした理由は、1年間の習慣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。さ

表1 行動辞書に登録した用語の例

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
外出	出勤、通勤、通学、行ってきます
帰宅	帰宅、帰る、退勤、退社、下校

らに、社会事象は、日付単位や週単位で変化する可能性があるが、ユーザー属性ごとの習慣行動と非習慣行動に着目することで、社会事象の内容に変化が生じるかを目的とした。なお、本論文では、各属性の習慣行動の違いを分析し、非習慣行動を用いて、既存手法 [19] よりも詳細なトピックを抽出することを目的とするため、月単位で比較する。その比較すべき理由は、1週間ごとにすると、何気ない突発的な行動の差が非習慣行動として分類される可能性があり、1週間だけではユーザー属性ごとの習慣行動を取り出せないと考えた。加えて、ユーザー属性に応じて、平時の生活習慣と特定期間の生活習慣のタイムスパン（年、月、週、日）は変化し、取得できる社会事象も変化する可能性がある。しかし、本研究では、ユーザー属性に応じた平時の生活習慣と特定期間の生活習慣のタイムスパンの関係性が明らかになっていないため今後の議論とする。

3.2.1 行動抽出処理

本処理では、ユーザーの投稿履歴から、既存研究 [19] の行動抽出処理で用いた日本語語彙大系 [24] を参考に行動辞書を作成し、行動情報を抽出する。そして、習慣行動に関する単語の時間ごとの出現回数を示す生活習慣ベクトルを作成する。まず、生活習慣ベクトル作成のため、習慣行動の選定を行う。習慣行動として、既存研究 [21–23] を参考に「起床・就寝」、「在宅」、「外出」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 [21–23] を参照されたい。

3.2.2 生活習慣ベクトル作成処理

本処理では、ユーザーの習慣行動を素性とした4次元（習慣行動）×7次元（曜日）×24次元（時間帯）の672次元で構成される生活習慣ベクトルを作成する。平時のユーザーの習慣行動を表す生活習慣ベクトルを平時習慣ベクトルとし、特定の期間の習慣行動を表す生活習慣ベクトルを特定習慣ベクトルとした。平時習慣ベクトルは、年間の投稿から作成した生活習慣ベクトルとし、特定習慣ベクトルは、月ごとの投稿から作成した生活習慣ベクトルとして表した。

ユーザー属性 $attribute_j$ の年間の行動 beh_x における各曜日の時間帯 h の生活習慣ベクトル $YearPost(attribute_j, beh_x)$ を式 (1) に示す。

$$\begin{aligned} & \text{YearPost}(\text{attribute}_j \text{beh}_x) \\ &= \{YPost_{\text{attribute}_1 \text{beh}_x(0)}, YPost_{\text{attribute}_1 \text{beh}_x(1)}, \dots, \\ & \quad YPost_{\text{attribute}_j \text{beh}_x(h)}\} \end{aligned} \quad (1)$$

式(1)において、 h は、7 次元 (曜日) $\times$ 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 時を示す。 j は各属性を表しているため、性別、職業と地域であり、最大は 12 である。 x は、各習慣行動を表しているため、「起床・就寝」、「在宅」、「外出」と「帰宅」であり、最大は 4 である。 $YPost_{\text{beh}_x(h)}$ 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 秒までの間に習慣行動 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 秒までを示す。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した。

$$\begin{aligned} & \text{YearPost}(\text{attribute}_j \text{beh}_x)' = \\ & \frac{\text{YearPost}(\text{attribute}_j \text{beh}_x) - \text{MinYearPost}(\text{attribute}_j \text{beh}_x)}{\text{MaxYearPost}(\text{attribute}_j \text{beh}_x) - \text{MinYearPost}(\text{attribute}_j \text{beh}_x)} \end{aligned} \quad (2)$$

式(2)において、 $\text{MaxYearPost}(\text{attribute}_j \text{beh}_x)$ は、日曜日から土曜日までの最大値の投稿数であり、 $\text{MinYearPost}(\text{attribute}_j \text{beh}_x)$ は、日曜日から土曜日までの最小値の投稿数のことを指す。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。各ユーザ属性 attribute_j の各月の行動 beh_x における各曜日の時間帯 h の生活習慣ベクトル $\text{MonthPost}(\text{attribute}_j \text{beh}_x)$ を式(3)に示す。

$$\begin{aligned} & \text{MonthPost}(\text{attribute}_j \text{beh}_x) \\ &= \{MPost_{1, \text{attribute}_1 \text{beh}_x(0)}, MPost_{1, \text{attribute}_1 \text{beh}_x(1)}, \dots, \\ & \quad MPost_{m, \text{attribute}_j \text{beh}_x(h)}\} \end{aligned} \quad (3)$$

式(3)において、 m は月を表し、 $MPost_{1, \text{attribute}_j \text{beh}_x(0)}$ の場合、1 月の日曜日の 0 時 00 分 00 秒から 0 時 59 分 59 秒までの間にユーザ属性 attribute_j が習慣行動 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習慣行動の抽出機能

本機能では、式(2)で示した平時習慣ベクトルと式(3)で示したユーザ属性の特定習慣ベクトルとを比較し、差分が閾値を超える時間帯を抽出する。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非習慣行動を平時と異なる事象とする。

3.3.1 非習慣行動抽出処理

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とを比較し、ユーザ属性ごとの時間帯ごとに差分を抽出する。ユーザ属性ごとの attribute_j の beh_x に関する任意の時間帯 h の平時習慣ベクトル $YPost_{\text{attribute}_j \text{beh}_x(h)}$ と特定習慣ベクトル $MPost_{m, \text{attribute}_j \text{beh}_x(h)}$ における差分 S は、

式(4)にて算出する。

$$\begin{aligned} & S(m, h) \\ &= \frac{\sum_{k=h-\alpha}^{h+\alpha} |YPost_{\text{attribute}_j \text{beh}_x(k)} - MPost_{m, \text{attribute}_j \text{beh}_x(k)}|}{2\alpha + 1} \end{aligned} \quad (4)$$

式(4)において、 α は求める時間 h の前後の時間数を示す。本研究では、 α 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非習慣行動として定義する。非習慣行動の判断式を式(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、非習慣行動と判断する時の閾値には、全て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)により判定することで非習慣行動を抽出する。

3.3.2 トピック抽出処理

本処理では、既存研究[19]と同様に Blei らによって提案された潜在的ディリクレ配分法[20]を用いて、トピックを抽出する。潜在的ディリクレ配分法[20]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一つである。3.3.1 項の非習慣時から算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[25]を用いて、トピックを分析する。

4. 評価実験

4.1 実験内容

本実験では、第 2 章で設定した研究課題 1 の「ソーシャルセンサの特性として、ユーザの属性ごとに習慣行動が相違し、非習慣行動の時間帯が異なること」と研究課題 2 の「ユーザ属性ごとに反応や興味を持つ社会事象が異なり、同じ社会事象でも抽出内容が異なること」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。

実験では、研究課題を実証するため、平時習慣ベクトルとユーザ属性ごとの特定習慣ベクトルを基に非習慣行動の時間帯を算出する。そして、ユーザの属性ごとに各行動の非習慣行動の時間数が異なることを示し、研究課題 1 を証明する。また、非習慣行動の時間帯の期間の全投稿のトピック、ユーザ属性を指定した非習慣行動の時間帯に行動した習慣行動のトピックを比較し、研究課題 2 を立証する。そして、ユーザ属性ごとに異なる社会事象を取得可能であることを検証し、提案手法の有効性を証明する。

表2 実験データ

	社会人		学生		主婦	フリーター	
	男性	女性	男性	女性	女性	男性	女性
10代	0	0	198	102	0	3	11
20代	52	7	0	101	68	91	42
30代	85	0	0	0	109	64	25
40代以上	92	0	0	0	85	27	7
北海道・東北	15	10	13	6	34	14	14
関東	126	34	83	58	175	93	38
中部	39	14	37	16	60	26	13
近畿	43	6	32	1	69	24	16
中国・四国	19	1	14	4	24	11	2
九州・沖縄県	15	1	12	8	30	17	4

4.2 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 [26] 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 [27] を解析して、2014 年 1 月～12 月に投稿しているユーザを収集し、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。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。ただし、習慣行動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。

STEP 4: STEP 3 で収集したユーザのプロフィールを確認し、ユーザの属性を手で付与する。

以上の手順にて収集した実験データの詳細は、ユーザ数が 1,514 ユーザ、投稿件数が 3,331,423 件である。各属性のユーザ数の構成を表 2 に示す。

本実験では、非習慣行動算出に関わるパラメータ α とトピック抽出で用いる潜在的ディリクレ配分法 [20] でトピック数を設定する。

(1) パラメータ α

パラメータ α は、非習慣行動抽出処理において、平時習慣ベクトルと特定習慣ベクトルの差分を算出する時に用いる値である。式 (4) において、 h は求める対象の時間帯、パラメータ α は h を算出する際に考慮する前後の時間を示す。 α の値が大きくなると、個々の時間の情報を読み取ることが難しくなる。

本実験では、既存研究 [19] を参考に、1 日のサイクルは 0 時

図3 2月と7月の Coherence の結果

図4 2月と7月の Perplexity の結果

～6 時、6 時～12 時、12 時～18 時と 18 時～24 時の 6 時間ごととし、 α の値を 3 に設定した。

(2) トピック数

最適なトピック数を算出するため、評価指標として Coherence (トピックの品質) [28] と Perplexity (予測性能) [29] を用いた。Coherence と Perplexity のグラフを図 3 と図 4 に示す。縦軸は、各 Coherence と Perplexity の値を示し、横軸はトピック数である。図 3 より、Coherence がトピック数 100 から 160 を境に横ばいになっていることがわかる。図 4 の Perplexity のグラフは 7 月に関しては、下がり続けているが、既存研究 [30] においても最適なトピック数は 90 から 140 とされている。そのため、本研究では、潜在的ディリクレ配分法の手法 [20] で用いるトピック数は、100 トピックとし、トピック

表3 2月における非習慣行動の抽出時間数

分類	行動内容	非習慣行動 合計時間数	非習慣行動 重複無し 時間数	非習慣行動重複あり時間数／分類					
				社会人	主婦	フリーター	学生	男性	女性
社会人	起床・就寝	76	7		49	40	—	60	50
	在宅	91	3		56	55	—	74	69
	外出・帰宅	159	18		97	38	33	115	87
主婦	起床・就寝	74	2	60		41	—	55	67
	在宅	69	3	56		43	—	48	54
	外出・帰宅	153	6	97		48	41	103	112
フリーター	起床・就寝	79	16	40	41		—	44	47
	在宅	85	20	55	43		—	52	50
	外出・帰宅	80	6	38	48		—	36	52
学生	起床・就寝	—	—	—	—	—		—	—
	在宅	—	—	—	—	—		—	—
	外出・帰宅	77	26	33	41	—		44	34
男性	起床・就寝	73	0	60	55	44	—		60
	在宅	80	1	74	48	52	—		62
	外出・帰宅	149	1	115	103	36	44		89
女性	起床・就寝	80	4	50	67	47	—	60	
	在宅	79	1	69	54	50	—	62	
	外出・帰宅	141	3	87	112	52	34	89	

推定の反復回数は50回とした。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トピックとして、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。例えば、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トピックとした。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。

4.3 実験1：非習慣行動の抽出実験

4.3.1 実験概要

本実験では、2.2.1項の研究課題1を証明するため、2014年2月と7月の投稿内容を対象に、ユーザ属性ごとの非習慣行動を算出し、習慣行動が相違していることを示して、研究課題1を証明する。ユーザの属性は、性別（男性、女性）、職業（社会人、学生、主婦、フリーター）と地域（北海道・東北地方、関東地方、中部地方、近畿地方、中国・四国地方、九州地方・沖縄県）を対象とする。

4.3.2 実験手順

本実験では、次の手順に従って分析を実施する。

STEP 1: 平時のユーザの習慣行動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2月と7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2月と7月の特定習慣ベクトルを比較し、非習

慣行動が発生した時間帯を抽出する。

以上の手順で抽出した各属性の非習慣行動の時間を表3から表6に示す。表3と表4は、各月の習慣行動の非習慣行動の時間数を属性ごとにクロス集計し、表5と表6は地域ごとにクロス集計した値であり、各属性の非習慣行動の時間数を定量的に示した値である。ここでは、一つの属性に着目し、同時間中に他の属性が非習慣行動を行っていない時間帯のことを「重複無し」、一方、同時間中に他の属性が非習慣行動を行った時間がある場合を「重複あり」として扱う。また、非習慣行動の抽出時間の上位と下位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5に示す。

4.3.3 実験結果

表3から表6、図5より、次のことがわかった。

・各属性の非習慣行動の時間数が異なることがわかった。

表3と表4より、非習慣行動重複無しの時間が2月は平均7.3時間、7月は平均5.6時間あり、各属性で非習慣行動の時間が異なった。その詳細を確認すると、学生の外出・帰宅の行動の非習慣行動の時間帯が最も多いことがわかった。これは、非習慣行動の時間帯を外出と帰宅の行動で一緒にした要因も推察できるが、中学生と高校生、大学生では授業やアルバイトなどの関係によって、外出や帰宅の時間帯が変化することが原因と考えられる。また、フリーターの起床・就寝の行動に関しても朝勤や夜勤のユーザが同等に存在するからであると推測できる。

表5と表6より、非習慣行動の時間数の重複無しの時間数は、関東地方で多くなっており、他の地方では全体的に少なくなっていることがわかった。

表 4 7 月における非習慣行動の抽出時間数

分類	行動内容	非習慣行動 合計時間数	非習慣行動 重複無し 時間数	非習慣行動重複あり時間数／分類					
				社会人	主婦	フリーター	学生	男性	女性
社会人	起床・就寝	69	4		46	24	44	55	56
	在宅	71	3		52	38	35	52	57
	外出・帰宅	153	8		76	84	86	90	94
主婦	起床・就寝	74	2	46		32	35	41	65
	在宅	72	0	52		38	38	44	62
	外出・帰宅	146	6	101		60	94	78	98
フリーター	起床・就寝	79	11	24	32		39	28	37
	在宅	78	5	38	38		48	42	34
	外出・帰宅	149	10	84	60		80	81	83
学生	起床・就寝	70	4	44	35	39		40	40
	在宅	78	6	35	38	48		39	42
	外出・帰宅	166	27	86	89	80		76	85
男性	起床・就寝	68	0	55	41	28	40		58
	在宅	68	4	52	44	42	39		47
	外出・帰宅	135	3	104	78	81	76		81
女性	起床・就寝	85	1	56	65	37	40	58	
	在宅	80	6	57	62	34	42	47	
	外出・帰宅	141	2	101	98	83	85	81	

表 5 2 月における地域属性ごとの非習慣行動の抽出時間数

分類	行動内容	非習慣行動 合計時間数	非習慣行動 重複無し 時間数	非習慣行動重複あり時間数／分類					
				北海道・ 東北	関東	中部	近畿	中国・四国	九州・ 沖縄県
北海道・ 東北地方	起床・就寝	—	—		—	—	—	—	—
	在宅	—	—		—	—	—	—	—
	外出・帰宅	155	27		95	51	128	81	77
関東地方	起床・就寝	80	12	—		—	55	55	63
	在宅	81	13	—		—	55	63	56
	外出・帰宅	152	14	95		51	113	54	46
中部地方	起床・就寝	—	—	—	—		—	—	—
	在宅	—	—	—	—		—	—	—
	外出・帰宅	79	5	51	51		59	—	—
近畿地方	起床・就寝	68	0	—	55	—		58	62
	在宅	80	4	—	55	—		64	66
	外出・帰宅	158	2	99	113	59		74	57
中国・ 四国地方	起床・就寝	75	8	—	55	—	58		62
	在宅	77	0	—	63	—	64		63
	外出・帰宅	81	0	55	54	—	74		55
九州地方・ 沖縄県	起床・就寝	89	13	—	63	—	62	62	
	在宅	76	1	—	55	—	66	63	
	外出・帰宅	77	11	45	46	—	57	55	

・同様の習慣行動でも属性によって、非習慣行動の時間帯が相違することがわかった。

図 5 より、同じ習慣行動においても属性が異なると、抽出できる非習慣行動の時間帯が相違することがわかった。例えば、起床・就寝の投稿では、男性で抽出されていない金曜日の 7 時から 10 時は、社会人では非習慣行動として算出されている。

他にも同様の時間帯も存在するものの各属性によって、非習慣行動の時間が明確に相違することがわかった。

以上より、属性ごとに非習慣行動が異なることを証明でき、非習慣行動の時間帯が異なるため、取得できる投稿内容も変化する。これにより、研究課題 1 を証明できたと考えられる。

表6 7月における地域属性ごとの非習慣行動の抽出時間数

分類	行動内容	非習慣行動 合計時間数	非習慣行動 重複無し 時間数	非習慣行動重複あり時間数／分類					
				北海道・ 東北	関東	中部	近畿	中国・四国	九州・ 沖縄県
北海道・ 東北地方	起床・就寝	75	3		36	45	55	51	58
	在宅	76	9		45	53	51	55	51
	外出・帰宅	143	12		76	105	102	98	88
関東地方	起床・就寝	76	16	36		48	35	34	48
	在宅	68	3	45		47	51	50	56
	外出・帰宅	142	16	76		86	87	87	95
中部地方	起床・就寝	82	2	45	48		50	51	57
	在宅	71	3	53	47		49	53	53
	外出・帰宅	150	8	105	86		103	106	91
近畿地方	起床・就寝	73	0	55	35	50		65	57
	在宅	73	4	51	51	49		57	57
	外出・帰宅	142	1	102	87	103		122	97
中国・ 四国地方	起床・就寝	71	0	51	34	51	65		54
	在宅	72	0	55	50	53	57		53
	外出・帰宅	143	3	98	87	106	122		106
九州地方・ 沖縄県	起床・就寝	85	0	58	48	57	57	54	
	在宅	73	3	51	56	53	57	53	
	外出・帰宅	145	12	88	95	91	97	106	

4.4 実験2：ユーザの属性ごとの非習慣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、2.2.2項で上げた研究課題2を証明するため、非習慣時と算出された期間の投稿をトピック分類した結果に基づき検討する。本実験では、実験1で抽出した非習慣行動の時間帯を基に2月と7月を対象に属性ごとに解析する。

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月における非習慣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習慣行動ごとに区分する。

STEP 2: ユーザ属性ごとの2月と7月の全投稿を収集し、それらの投稿を月ごとに選別する。

STEP 3: STEP 1とSTEP 2で整理した投稿を既存手法のトピックと提案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て該当トピックをノイズとして扱う。

STEP 5: STEP 4で取得した各月の既存手法と提案手法のトピック内の単語が5単語以上一致する場合は、同一トピックとして取り扱う。

以上の処理手順でトピックを分類する。

4.4.3 実験結果

既存研究 [19] の非習慣行動の時間帯のみの投稿内容から抽

出した社会事象に関するトピックの結果を表7に示す。また、既存研究 [19] と提案手法で、取得できたトピック数を性別と職業は表8、地域ごとは表9に示す。表8と表9は、既存手法 [19] で、取得したトピックに対し、性別と職業ごと、地域ごとの非習慣行動の時間帯を考慮して、取り出したトピックの一致、不一致数を定量的に示している。不一致数は、表7で示したトピック抽出内容と一致していない内容のトピック数を指す。これにより、既存手法と比較して新たなトピックを抽出できていることがわかる。表10と表11は、提案手法において、不一致の代表的なトピックをまとめた一覧である。なお、下線のトピックは、既存手法と同様のトピックを示す。表8と表9より、各属性で重複しない多くのトピックを抽出できたことがわかった。表8から表9より、次に示すことが明らかになった。

・既存手法と提案手法ともに異なるトピックが抽出できることがわかった。

表8と表9ともに多くのトピックで、トピックが一致しないことがわかった。これは、各属性のトピックの内容が異なり、同様の内容でも興味を示す内容が違うため、トピック内の単語が一致しないからだと考えられる。トピックを詳細に把握するため、提案手法の不一致の代表的なトピックをまとめた表10と表11より、次に示すことが明らかとなった。

・提案手法ではより多くのトピックを抽出していることがわかった。

既存手法で抽出されている社会事象（ソチオリンピックや大雪被害、サッカー、台風）に関するトピックを、提案手法においても抽出できていることがわかった。提案手法で新たに出てきているトピックを確認すると、ユーザの詳細がわかるような

図5 上位と下位の習慣行動解析結果

普段の生活パターンの内容が確認できた。

具体的には、主婦は子供を心配しているなど、子供との生活を示す内容が多く、男性は外出と帰宅時に大雪に困っているなどの個人的なことに関連した社会事象が把握できた。

- 行動ごとに各属性で抽出できるトピックが異なることがわかった。

代表的なトピックを確認すると、ソチオリンピックや大雪、

地震などに関するトピックを取り出せることがわかった。ソチオリンピックに関するトピックで社会人や主婦は、外出と帰宅時の行動で算出できているのに対し、フリーターは起床・就寝時に抽出できていることがわかった。これは、朝外出して、夜帰宅するといった規則的な生活を送る社会人は、深夜に行われているオリンピックをリアルタイムに視聴することが困難であるためだと考えられる。そのため、外出や帰宅時に試合結果や

表 7 既存手法 [19] の社会事象に関する代表的なトピックの抽出結果

月	既存手法	
	行動	解釈トピック (キーワード)
2月	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) ・安倍政権が調査 (安倍, 経済, 政府, 調査)
	在宅	・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) ・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)
	外出・帰宅	・浅田真央の活躍 (選手, オリンピック, 感動, 埼玉, 彼女, 五, ソチ) ・舩添氏が都知事選に関わった (都知事, 東京, 舩添) ・原発で問題発生 (原発, NHK, 問題, 職員) ・横浜と千葉で地震発生 (横浜, 発生, 千葉, 地震)
7月	起床・就寝	・野々村議員号泣 (反対, 野々村, 説明, 号泣, 選挙, 行為) ・原発に関するニュース (原発, ニュース, 経済, 研究, 避難)
	在宅	・安倍政権が集団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) ・宮崎で台風により大雨 (台風, 影響, 大雨, 発生, 宮崎) ・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) ・安倍政権が地方支援を検討 (安倍, 政権, 支援, 地方, メンバー)
	外出・帰宅	・平和記念ドラマ放送 (ドラマ, 平和, 70, ぶり)

表 8 既存手法 [19] とユーザ属性ごとの抽出トピック数

月	既存手法			提案手法					
	行動	抽出トピック数		抽出トピック数					
				社会人	主婦	フリーター	学生	男性	女性
2月	起床・就寝	抽出数	48	45	47	48	—	45	53
		不一致数	43	44	45	46	—	44	52
	在宅	抽出数	55	59	48	55	—	52	20
		不一致数	50	56	46	54	—	49	19
	外出・帰宅	抽出数	100	102	97	52	46	100	33
		不一致数	67	93	88	48	40	88	24
7月	起床・就寝	抽出数	42	43	45	48	46	45	53
		不一致数	37	41	43	46	42	42	50
	在宅	抽出数	53	50	45	43	54	46	54
		不一致数	49	47	44	41	52	42	53
	外出・帰宅	抽出数	103	99	101	94	108	99	88
		不一致数	91	95	97	90	104	96	84

表 9 既存手法 [19] と地域属性ごとの抽出トピック数

月	既存手法			提案手法					
	行動	抽出トピック数		抽出トピック数					
				北海道・東北	関東	中部	近畿	中国・四国	九州・沖縄県
2月	起床・就寝	抽出数	48	—	52	—	45	49	56
		不一致数	42	—	50	—	44	47	55
	在宅	抽出数	55	—	53	—	53	52	51
		不一致数	50	—	51	—	50	50	50
	外出・帰宅	抽出数	100	97	99	49	108	53	50
		不一致数	72	89	93	45	99	51	45
7月	起床・就寝	抽出数	42	47	52	48	55	43	53
		不一致数	38	44	50	45	52	40	52
	在宅	抽出数	53	47	46	46	50	47	51
		不一致数	49	45	44	44	47	45	49
	外出・帰宅	抽出数	103	86	95	103	88	92	103
		不一致数	89	82	87	97	83	88	100

状況を確認するユーザが多く、主婦も相手の生活リズムに合わせる傾向があることから、このような結果になったと推察できる。性別、地域や月ごとの特徴的な内容を次に示す。

- ・男性と女性では、男性は雪や地震といったトピックが抽出

できていることがわかり、一方、女性では、子供や献立といった家庭内の出来事に関するトピックが確認できる。これは、男性は仕事に影響を与える天候などの情報を確認している傾向があると考えられる。また、女性は主婦のユー

表 10 ユーザ属性ごとの代表的なトピック抽出結果

月	行動	提案手法					
		代表的なトピック内容					
		社会人	主婦	フリーター	学生	男性	女性
2月	起床・就寝	食事の献立 (今日, 食事, 献立, 味)	恵方巻を旦那と一緒に食べるために 買い物 (旦那, 一緒, 巻き, お腹, 帰り, 恵方, スーパー, 通り)	オリンピック応援に関する内容 (ソチ, オリンピック, 修造) 雪の影響により早めの帰宅をした男性ユーザ (雪, 帰宅, 関東, 僕, 早め) 毎日昼頃にアルバイトを行うユーザ (今日, 毎日, 昼, バイト)	—	東京にて大雪に見舞われるニュースと家族の予定 (雪, ニュース, 東京, 幼稚園, 家族, 遊び, 旅, 温泉, 延期, 明後日) 母親の献立に関する投稿 (食事, 献立, ママ, 具)	実家での生活を楽しむ女性 (献立, 幸せ, ママ, 実家, パパ, TV)
	在宅	関東で行われるライブ (昨日, 夜, たくさん, ライブ, 埼玉, 購入, 準備, 栃木, チケット, ベスト)	子供と車周辺の除雪作業を行う様子 (車, 子, 運転, 除雪) バレンタインを家族で過ごす様子 (チョコ, 家族, 実家, バレンタイン, ダンナ)	関東地方での大雪 (雪, 大変, 東京, 関東) インフルエンザで病院に行くユーザ (病院, 大事, インフルエンザ, 年齢)	—	雪かきを行う様子 (自分, 家, 雪かき, 異常, 被害) スーパーへ買い物に向かう親子 (車, 先, 親, 薬, スーパー)	ユーザの子供の部活に関する内容 (娘, 子供, 大変, 会, 吹奏楽, 本, トランペット)
	外出・帰宅	オリンピック出場の上村愛子の活躍 (雪, 上村, 愛子, 準備, 夜中) インフルエンザを危惧するユーザ (ホント, 病院, インフルエンザ, 白湯)	大雪被害に関する報道 (日本, 東京, 大雪, 心配, 電話, NHK, 交通) 北上温泉を楽しむユーザ (北上, 風呂, 新幹線, エステ)	自宅でオリンピック応援 (ごはん, オリンピック, 感情, 酒, スキー, 部屋, パソコン) 久々にライブを楽しむユーザ (ライブ, 期間, 久しぶり, 大人, 最高)	アルバイト出勤前の学生 (明日, バイト, 出勤, 朝)	浅田真央の応援 (ちゃん, 真央, 感動, 浅田) 占い結果を確認するユーザ (姜, 仕事, 運, 金, 健康, 運勢)	ユーザの子供の誕生日 (日, 今日, 子供, 誕生)
7月	起床・就寝	エアコン掃除に関する話題 (設定, エアコン, 掃除, 対策)	夏休みのため, 家事が減った主婦 (夏休み, 幸せ, 弁当, 高校)	動画サイトでサッカー観戦 (YouTube, 2525, サッカー, ツアー, 頻度, 機能, パソコン)	部活動の野球の試合について (市, 試合, 今朝, 野球)	花火大会に関するニュース (ニュース, 花火, 本日, 大会)	子供の学校の怖い話 (息子, 君, 学校, 花子, 女子)
	在宅	ユーザの息子と花火大会 (子供, 息子, 花火, 大会, 影響)	梅雨明けでエアコンを起動し, 快適に過ごす様子 (エアコン, 梅雨, 明け, 快適)	ネットを楽しむユーザ (わたし, sumo, YouTube, ブログ)	アルバイトを考慮する学生 (店, 今年, 失敗, バイト, 久々, 近所, 経験, 夕方)	地震により被災した地域 (時間, 分, 震度, 地震, 県, 大変)	花火大会に関する内容 (花火, 大会, 電車, 帰宅, 準備, 予想)
	外出・帰宅	松坂桃李出演のドラマ (くん, 楽しみ, ドラマ, 今度, 桃李, 撮影, 共演)	サッカーの試合観戦 (最後, 試合, サッカー, 初日, 上達) 東京に台風が迫る状況 (台風, 東京, 空, 準備)	勉強ではチョコレートが良い (勉強, 楽, チョコ, 脳)	食事を1人で済ます学生 (ごはん, ひとり, 食事, 状況, コンビニ)	渋谷の大雨, 雷に遭うユーザの内容 (周り, 雷, 渋谷, 大雨) 花火を楽しむにユーザ (楽しみ, 花火, 福, 天気)	子供の熱中症を考慮する様子 (娘, 雨, 子供, 症, 熱中, 塩, 頻度)

ザも多いことから, 家庭状況を気にかけている様子が多く抽出されている傾向がある。

- ・地域属性を考慮した場合の代表的なトピックを確認すると, 職業属性を考慮した場合と同様に, ソチオリンピックや大雪, 地震に関するトピックを見出せることがわかった。

- ・2月は, 大雪に関するトピックも多く, 近畿地方で, 在宅時の行動として, 雪によって外出に影響したと推考できる内容が取り出せることがわかった。また, 九州地方・沖縄県では, 起床・就寝時の行動で, ユーザの実家で積雪したことに関するトピックを確認できた。また, 在宅時の行

表 11 地域属性ごとの代表的なトピック抽出結果

月	行動	提案手法					
		代表的なトピック内容					
		北海道・東北	関東	中部	近畿	中国・四国	九州・沖縄県
2月	起床・就寝	—	東京都知事選挙 (岡田, 職員, 選挙, 外添, 会長, 団体, 議員) 節分を楽しむユーザ (恵方, 電話, 巻き, 豆)	—	浅田真央に関する話題 (真央, ソチ, オリンピック, 五輪, 浅田, アスリート) スーパーでカレーを購入するユーザ (カレー, スーパー, 屋, 残り, 味噌汁)	家族で生活する様子 (みんな, 一緒, 旦那, たくさん, 幸せ, 息子, 電話)	昨夜にかけて実家が積雪 (昨日, 夜, 実家, ビックリ, 自然, 積雪, 深夜, センチ) ラーメンファンによる投稿 (気持ち, 腹, ファン, ラーメン, 追加, ごちそう)
	在宅	—	ソチオリンピックに関する話題 (雪, 2014, 今年, 生, ソチ, 修造, 松岡) 夕方にご飯を考慮するユーザ (味噌汁, 夕方, 生姜, 焼き, サラダ, 節約)	—	雪によって出勤に影響 (雪, 明日, 影響, 出勤, 状況) インフルエンザによる影響 (影響, 出勤, インフル, 発生, 状況, クラス)	大学受験におけるリスニングの話題 (リスニング, どちら, 記憶, 大学)	大雪により休み (休み, 大雪, 今秋, 雪かき, 予報) インフルエンザによる影響を受けるユーザ (昨日, 朝, 目, インフルエンザ, 最悪)
	外出・帰宅	—	夕飯を考慮するユーザ (ごはん, スーパー, 子供, 鍋, すき, 満足)	インフルエンザを危惧するユーザ (元気, 子供, 病院, インフルエンザ)	ソチオリンピックに関する話題 (ソチ, 気温, 修造, 松岡, 高梨) NHKの織田信長に関する番組を視聴するユーザ (織田, 信長, 放送, NHK, 詳細)	羽生選手を応援 (選手, LOVE, 羽生, ファン, インタビュー, スケート, ラブ) ビジネス投資を考慮するユーザ (円, ビジネス, お金, 億, 生活, カード, 投資, 券)	フィギュアスケート選手の話 (選手, 羽生, 高橋) 停電が発生した話題 (大雪, 予想, 停電, 苦手)
7月	起床・就寝	録画していた番組を見終えたユーザ (終了, 番組, 録画, 昼, アイス, 実家, シーン)	主婦が家で過ごす様子 (ニュース, ご飯, 生活, 旦那, 梅雨明け)	大阪にて梅雨明け (大阪, エアコン, 梅雨, 今朝, 明け, 快適)	地元のスーパーで買い物するユーザ (外, 買い物, スーパー, 地元, 駄馬, 自宅)	ユーザの子供の昼ご飯に関する内容 (娘, ご飯, お昼, 昼, 実家, 我が家)	録画作品に熱中するユーザ (帰宅, 録画, 作品, 熱中, 酒)
	在宅	地震の備えについて (自身, 震度, 水, 実家, 速報, 食)	地震の報道 (地震, 放送, 震度, 防災, 速報)	大阪での花火大会に関する話題 (花火, 大阪, 大会, 最終)	次の週にゴジラが放映 (ゴジラ, 夏休み, ネット, 動画, 来週)	昼間に地震発生 (地震, 速報, 昼, 震度, 緊急)	ユーザの子供が大阪から帰宅 (明日, 息子, 大阪, 帰宅)
	外出・帰宅	占いに関する話題 (運, 勤, 健康, 仕事, 恋愛, 運勢, 勉強)	仕事前にニュースを確認するユーザ (今日, 仕事, ニュース, 朝, 事件, 疲れ, 詳細)	バイトを終えたユーザ (お金, 終了, バイト, 体調)	子供の夏休みの予定 (子供, 今回, 夏休み, 花火, 元気, テレビ, 大会)	ユーザの娘が花火大会に行く (今日, 娘, 元気, 花火, 大会, レベル, 光)	仕事が終わって帰宅した子持ちユーザ (息子, 会社, 帰宅, エアコン)

動では、大雪の影響を受けて、休みにするといったトピックも抽出されていることがわかった。これは、当時、近畿地方や九州地方・沖縄県といった地方にて、大雪に見舞われ、積雪などにより外出等に影響を与えたからだと考えられる。

- ・7月は、地震に関するトピックが多いことがわかった。在宅時の行動で、地震に関するトピックが出現している傾向があることから、当時、この時間帯に地震が発生したと推測できる。また、主に外出と帰宅時の行動で、花火大会に

関するトピックが多く出現していることがわかった。これは、花火大会が夜に開催されるため、帰宅の時間帯で多く投稿される傾向があることによると考えられる。

- ・地方に関連したトピックを抽出していることがわかった。

7月における中部地方にて、起床・就寝時の行動で、大阪にて梅雨明けが発表されたトピックが取得でき、在宅時の行動で、大阪で開催される花火大会に関するトピックが抽出されていることがわかった。これは、中部地方と近畿地方が近い気候であることから、大阪で梅雨明けした情報にユーザが反応し

た可能性がある。また、花火大会のトピックを抽出できた理由として、大阪にて天神祭りや淀川花火大会といった全国的に有名な花火大会を示すことからこのような結果になったと考えられる。

以上の結果からユーザ属性を考慮し、非習慣行動を算出することで、詳細なトピックを見出すことができる。これにより、研究課題2を証明できた。

5. おわりに

本研究では、ユーザの習慣行動を用いて、その習慣行動の違いから、実世界における事象を抽出するソーシャルセンシング手法を適用した。そして、社会事象との関係性を実践的に評価するために、ユーザ属性ごとにトピックを分類する手法の汎用性を明らかにした。これにより、同一トピックに関してもトピックの抽出内容が変化し、既存研究 [19] よりも社会事象を詳細に把握できた。よって、研究課題1と研究課題2を解消することができた。

本研究を通じて、次に示す課題が明らかになった。

課題1: ユーザ属性に応じて、平時習慣ベクトルと特定習慣ベクトルのタイムスパン(年, 月, 週, 日)ごとの変化を考慮できていないこと。

課題2: 投稿内容を習慣行動のみで絞ったため、トピック内に判断が難しいキーワードがあったこと。

今後は、上記の2つの課題を対応する方策を検討し、社会事象のセンシング技術の実用化に向けて、季節や気象などの他の指標の関係を明らかにする予定である。

謝辞

本研究の一部は、2018年度関西大学学術研究員研究費によって行った。ここに記し、感謝の意を表する。

参 考 文 献

- [1] Q. Zhao, T. Liu, S. Bhowmick, and W. Ma: "Event Detection from Evolution of Click-Through Data," *Proc. of the 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 484-493, 2006.
- [2] J. Ginsberg, M. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant.: "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, Vol.457, pp. 1012-1014, 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, 2013.
- [4] 榊剛史, 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, 2012.
- [5] T. Sakaki, M. Okazaki, and Y. Matsuo: "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," *Proc. of the 19th Int. Conf. on World Wide Web*, pp. 851-860, 2010.
- [6] 榊剛史, 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, 2012.
- [7] A. Dingli, L. Mercieca, R. Spina, and M. Galea: "Event Detection Using Social Sensors," *Proc. of the 2nd Int. Conf. on Information and Communication Technologies for Disaster Management*, pp. 35-41, 2015.
- [8] S. Zhao, L. Zhong, J. Wickramasuriya, and V. Vasudevan: "Human as Real-Time Sensors of Social and Physical Events: A Case Study of Twitter and Sports Games," *Rice University Technical Report*, No.TR0620, 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, 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, 2014.
- [11] T. Georgiou, A. Abbadi, X. Yan, and J. George: "Mining Complaints for Traffic-Jam Estimation: A Social Sensor Application," *Proc. of the 2015 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, pp. 330-335, 2015.
- [12] M. Congosto, D. Fuentes, and L. Sanchez: "Microbloggers as Sensors for Public Transport Breakdowns," *Proc. of the IEEE Internet Computing*, Vol.19, No.6, pp. 18-25, 2015.
- [13] S. Asur and B. Huberman: "Predicting the Future with Social Media," *Proc. of the 2010 IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology*, Vol.1, pp. 492-499, 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, 2012.
- [15] J. Bollen, H. Mao, and X. Zeng: "Twitter Mood Predicts the Stock Market," *J. of Computational Science*, Vol.2, No.1, pp. 1-8, 2011.
- [16] J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, and A. Jaimes: "Correlating Financial Time Series with Micro-Blogging Activity," *Proc. of the 5th ACM Int. Conf. on Web search and Data Mining*, pp. 513-522, 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報告, Vol.2011-SLP-86, No.1, pp. 1-8, 2011.
- [18] V. Lampos, T. Bie, and N. Cristianini: "Flu Detector - Tracking Epidemics on Twitter," *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases 2010*, Vol. 6323, 2010.
- [19] 坂本一磨, 中村健二, 山本雄平, 田中成典: "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," 情報処理学会論文誌, Vol.59, No.10, pp.1-14, 2018.
- [20] D. Blei, A. Ng, and M. Jordan: "Latent Dirichlet Allocation," *The J. of Machine Learning Research*, Vol.3, pp. 993-1022, 2003.
- [21] 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, 2013.
- [22] 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, 2013.
- [23] 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, 2016.
- [24] 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.
- [25] Radim Řehůřek, gensim: <https://radimrehurek.com/gensim/> [accessed Apr. 15, 2019]
- [26] Twitter, Twitter Developers: <https://dev.twitter.com/> [accessed Apr. 15, 2019]
- [27] ropross.net, Twilog: <http://twilog.org/> [accessed Dec. 14, 2018]
- [28] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. McCallum: "Optimizing semantic coherence in topic models," *Proc. of the Conf. on Empirical Methods in Natural Language Processing*, pp. 262-272, 2011.
- [29] Radim Řehůřek, gensim: <https://radimrehurek.com/gensim/models/ldamodel.html> [accessed Apr. 15, 2019]
- [30] Murzintcev Nikita, Select number of topics for LDA model: <https://cran.r-project.org/web/packages/ldatuning/vignettes/>

topics.html [accessed Apr. 15, 2019]

(2019 年 1 月 18 日 受付)
(2019 年 6 月 12 日 採録)

[問い合わせ先]

〒569-1095 大阪府高槻市霊仙寺町 2-1-1

関西大学大学院総合情報学研究科

坂本 一磨

TEL: 072-690-2151

E-mail: k245159@kansai-u.ac.jp

—— 著 者 紹 介 ——

さかもと かずま
坂本 一磨 [学生会員]

1991 年生。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。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。Web マイニング、自然言語処理、システム設計等の研究開発に従事。

なかむら けんじ
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 年度科学技術分野の文部科学大臣表彰科学技術賞「科学技術振興部門」受賞。

やまもと ゆうへい
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）。2015 年関西大学先端科学技術推進機構特別任命助教、2018 年同特別任命准教授、2019 年から大阪工業大学情報科学部特任講師。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。

たなか しげのり
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及び関西大学先端科学技術推進機構社会空間情報科学研究センター長。

2016 年度科学技術分野の文部科学大臣表彰科学技術賞「科学技術振興部門」受賞。

なかむら たつや
中村 竜也 [非会員]

1994 年生。2018 年関西大学総合情報学部総合情報学科卒業。現在、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在学中。ビッグデータ解析に関連する研究に従事。

Practical Study Regarding Social Sensing Technologies for Extracting Unordinary Phenomena Considering User Attributes with Focus on Different Behavior from Normal Time

by

Kazuma SAKAMOTO, Kenji NAKAMURA, Yuhei YAMAMOTO, Shigenori TANAKA, and Tatsuya NAKAMURA

Abstract:

With the spread of CGM (Consumer Generated Media), a huge amount of digital data has been accumulated on the Internet. These data are utilized for improving social sensing technologies to measure not only social and economic trends but also various kinds of phenomena such as large-scale disasters. Using habitual behavior of users, authors proposed a new social sensing method for extracting phenomena in the actual world from the difference in the habitual behavior, and proved its usefulness. To practically evaluate the versatility of the application examples of the technology, it is necessary to clarify whether the data with different user attributes are also applicable. In this study, the habitual behavior of the users is analyzed attribute by attribute, and abnormal behavior is extracted based on different behavior from the normal time for each user attribute. Demonstration experiments were conducted to verify whether it is possible to find out social trends in the actual world or social phenomena for each user attribute on a detailed granularity.

Keywords: microblog, social sensing, social phenomena, habitual behavior, user attributes

Contact Address: **Kazuma SAKAMOTO**

Graduate School of Informatics, Kansai University

2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

TEL: +81-72-690-2151

E-mail: k245159@kansai-u.ac.jp