

科学研究費助成事業 研究成果報告書

平成 26 年 8 月 5 日現在

機関番号：32689

研究種目：挑戦的萌芽研究

研究期間：2011～2013

課題番号：23656370

研究課題名(和文) ソーシャルネットワーキングサービスに投稿された記事に基づく都市・建築空間像の解析

研究課題名(英文) A City and Architecture Analysis based on Social Networking Service

研究代表者

渡辺 仁史 (Watanabe, Hitoshi)

早稲田大学・理工学術院・教授

研究者番号：40063804

交付決定額(研究期間全体)：(直接経費) 2,900,000円、(間接経費) 870,000円

研究成果の概要(和文)：近年のスマートフォンの普及により、SNSがより馴染みのあるものになった。とりわけ、ユーザーがコンテンツを生み出すメディアの中では、地理情報を付与したSNSが徐々に普及してきた。このメディアの特性はユーザーが場所と関連した文章や写真を投稿する傾向にある。これらの文章から自動分析を施して、今までの都市調査のようにある一時的な情報を取得するのではなく、都市の動的な状態を発見出来ると思われる。それを最終目標におき、本研究では、1. 利用者の利用単語頻度から導き出される都市のスケープ、2. 都市で使われている単語から導き出される都市の分類、3. 震災時のSNS利用者の感情変化を主題として研究を行った。

研究成果の概要(英文)：Recently, due to the development of web eco-system and so-called "smart phone," social networking service becomes common. Especially, among the user-generated media, location based social network, social networking service with location data, has been an useful data collection media for researching on city. Traits of this data collection method radically varies from the obsolete research methods in terms of data quantity and quality of data: user's post data describes characteristics, habits, where to go, etc. Following these habits of the Location Based Social Media, we surveyed researches from three viewpoints: 1. semantic analysis of city from quantitative analytics of social network user activity 2. user activity on city from text mining with user activity 3. emotional alteration between disaster.

研究分野：工学

科研費の分科・細目：建築学

キーワード：ソーシャルネットワークサービス SNS ビッグデータ 感性単語 クラスタ分析 時系列分析 地理情報 デジタル・エスノグラフィー

1. 研究開始当初の背景

申請者らは今までに人間行動のあり方を分析し、建築計画に活用してきた。しかしながら、そのモデル化をするためには情報収集から分析に至るまで非常に長い時間がかかっていた。そのような時に、スマートフォンが徐々に普及してきたことで、GPS 情報をマイクロブログに掲載する例が徐々に増えつつあった。これらのユーザーが情報を発信するメディアの特性として、利用者の状態がそのメディアに含まれているという前提がある。そのために、ソーシャルメディアを用いた情報収集は、今までの追跡調査や PT 調査といった手法とは違う、人の特性や状態を含み、なおかつ場所を示すジオタグという情報が記事に含まれている場合がある。これらを数多く取得することが出来るために、今までとは異なる観点から分析という意味以外にも、リアルタイムに都市の状態をセンシングし、それに対する対策をリアルタイムにアクチュエーティングしていくシステムが可能になっていくだろう。以上の観点から、申請者らは、(1) 都市でどのような発言がなされているのか、(2) 都市にどのような人が集まっているのか、(3) 災害時の被災者の感情という視点で分析を行うことにした。

2. 研究の目的

ソーシャルメディアの情報を上記の三点から分析することで、都市・建築に対する利用者の需要を導き出すことを目的に研究活動を行ってきた。以下、具体的な研究活動での目的を記す。

(1) では、都市の特性をソーシャルメディア上で観測される単語をマッピングすることにより、その場所の特徴を示すことで、都市計画やまちづくりに必要な基礎的な資料づくりに対する方法を確立することを目的とした。

(2) では、ソーシャルメディアの利用者の特性を単語の頻度と街の利用頻度という観点で分析することで、その街がどのように変容する可能性があり得るのか検討することを目的とした。

(3) では、東日本大震災前後のソーシャルメディアで観測された形容詞の利用頻度を追跡することで、被災地と都心部での感性単語表現が異なることを示し、今後の災害復興に向けた被災者のリモートセンシング手法の一種としてのソーシャルメディアのあり方を検討することを目的とした。

3. 研究の方法

本報では、マイクロブログサービスである「Twitter」をデータ収集の対象とした。ユーザーが Twitter に投稿する記事にはテキストとしての「ツイート」だけでなく、時刻ユーザー名、ジオタグ(緯度経度情報)などが付帯されたデータとして記録されている。これらのデータセットのことを、ここでは

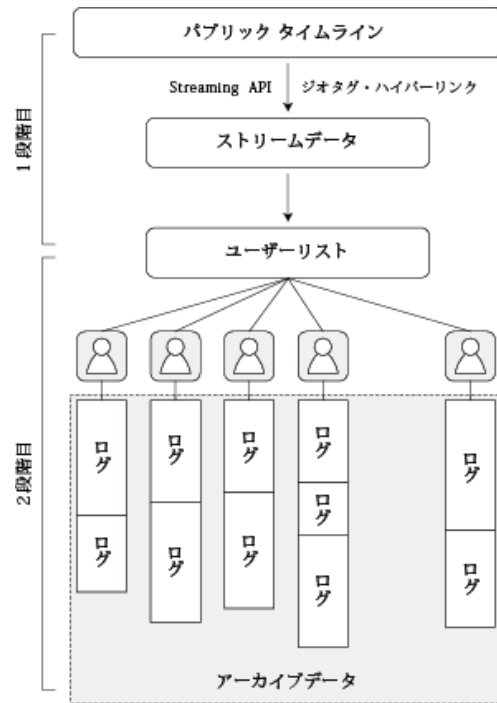


図1 Twitterからの情報取得方法

「ツイートデータ」と呼ぶこととする。

Twitter は第三者がツイートを取得するための API (Application Programming Interface) を公開している。筆者はこの API の仕様に基づいてツイートデータを自動的に取得するプログラムを作成した。ツイートデータの取得過程は 2 段階に分かれており、以下に詳述する。

1 段階目として、記事の収集対象とするユーザーを選定する。公開されている API のひとつである「Streaming API」を常時モニタリングし、その中でもジオタグが付与され、なおかつ、画像投稿サイトへのハイパーリンクが存在する記事のみを選別して記録する。理由は以下の 2 つである。ユーザーによるなんらかの思い入れや感情的な昂ぶりが表現されている可能性があると考えられ、情報の質が高いと推測されるため。そのようなツイートを発するユーザーは、サービスを使いこなしているユーザーであると考えられ、質の高い情報が記事になっていると推測されるため。

また、記録を除外するツイートデータは、以下の 2 つの条件のいずれかにあてはまるものとした。他人の記事を引用するリツイート (RT) 記事は、元の記事がユーザー本人の記事ではないため、記録しない。日本語を母国語とするユーザーを収集するため、取得された記事のジオタグから日本国外と判定されたものは記録しない。

これらの条件を満たした記事を収集する。収集されたデータを「ストリームデータ」と呼ぶこととする。またこれと同時に、投稿ユーザー名をリストアップする。

2 段階目として、前項で作成したユーザー名リストに基づき、リスト中のユーザーの記

事を過去に遡ってこれを収集するプログラムを作成した。これにより得られたツイートデータを「アーカイブデータ」と呼ぶこととする。

アーカイブデータはストリームデータとは異なり、記事中のハイパーリンクやジオタグの有無に関わらず記録する。遡る期間はAPI のシステム上、可能な限り過去の記事まで遡って収集する。アーカイブデータはユーザーごとにデータファイルを分けて保存される。作成したプログラムは定期的なリストを巡回し、各ユーザーの新規記事を取得し、データファイルに追記する。

ここで得られたアーカイブデータとストリームデータを用いて、都市の特性、利用者の特性、災害時の被災者感情の変化を捉えようとした。

3.2 都市に対する発言の分析

都市でどのような活動がなされているのかを把握することが都市計画や建築計画に重要な資料になりうる。そこで、申請者らは都市においてどのような発言がなされているのかを取得するために、単語の集計とクラスタリングから都市の動態を抜き出そうとした。まず、アーカイブデータのうち、ジオタグがあるものだけ発言のみを抜き出した。それらに対して MeCab により形態素解析をし、動詞を取得した。また単語の頻度とその重要性から「食べる」、「待つ」、「撮影する」、「見る」、「買う」、「休む」という6つの動詞に絞り込み、時系列ごとに集計した。また、集計結果を3次元カーネル密度推定によりこれらの単語頻度を可視化した。そこから都市の動態を把握することにした。

また、上記の研究で日常的な単語がその都市の特性を記す可能性が高いという示唆から、日常生活に関連する単語集を作成し、このデータからクラスタ分析を用いて、山手線の駅で発言された内容から駅を意味上で分類した。まず、日常生活に関連する単語を得るために、雑誌や Web などから得られた記事の単語頻度が高い名詞を 4000 語ほど選定した。これらの単語を駅半径 400m 以内でジオタグのあったものを集計し、これらの単語を元に KMeans 法にてクラスタ分析を行った。

3.3 都市の利用者の分析

都市の利用者がどのような興味や趣味を持っているのかを把握した上で、その人たちがどこに潜在的に集まりやすいのかを把握することは不動産開発のみならず、一般的なマーケティング以上に強力な手法になりうる。そこで、申請者らはアーカイブデータの利用者のすべてのデータの名詞の頻度に注目した。この頻度によりその人となりが見えると推定し、その人がどこの駅周辺に集まりやすいのかを集計し、その場所に固有な単語を抜き出した。その後、単語の抽象度を高めて同じカテゴリ

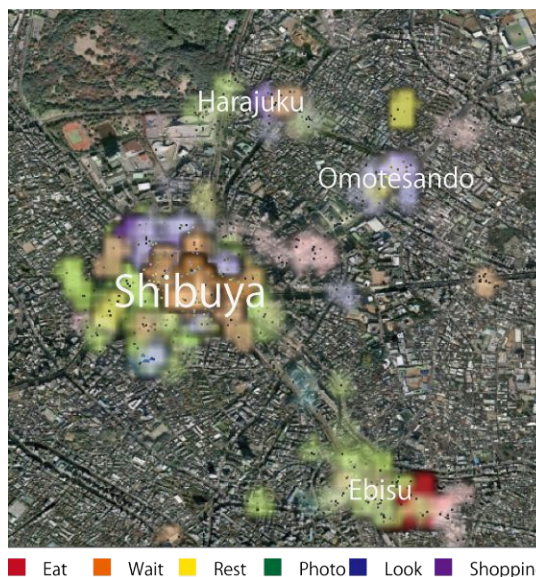


図 2 渋谷周辺での動詞

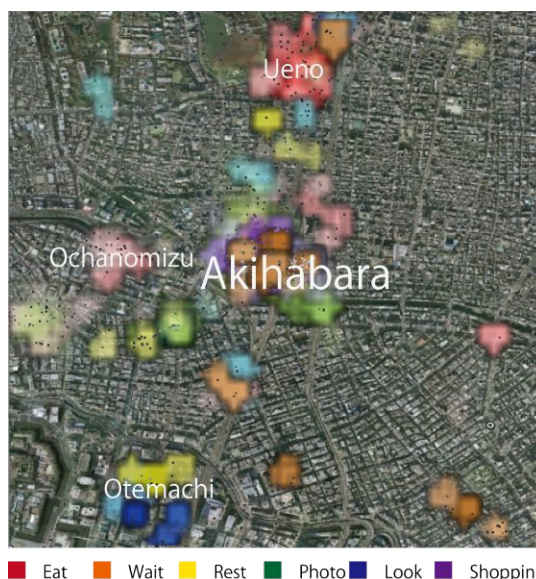


図 3 秋葉原周辺での動詞

の単語頻度を集計した。

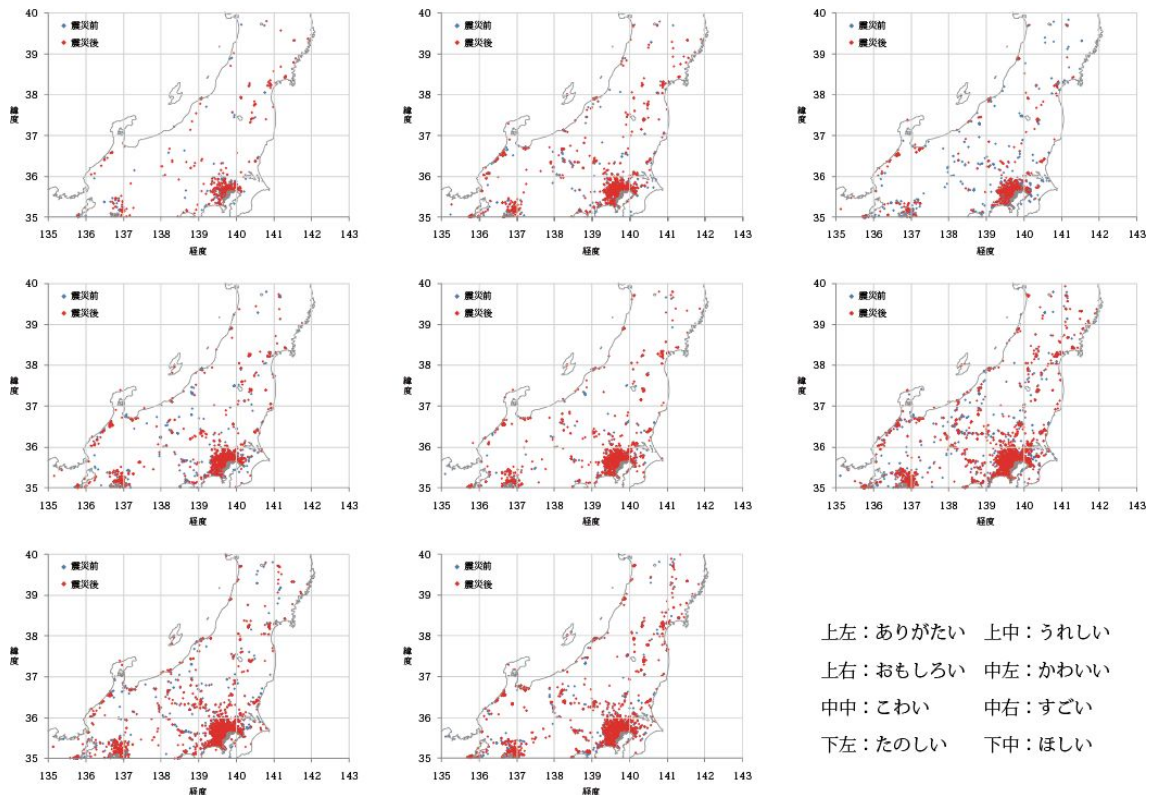
3.4 東日本大震災前後の被災者感情

災害時の被災者状態をうまく把握することは極めて重要だと思われる。そのために、被災者の状態を把握するために、アーカイブデータを用いた。これらのツイート記事を形態素解析により形容詞を抜き出しその頻度と重要性から「ありがたい」、「おもしろい」、「かわいい」、「たのしい」、「うれしい」、「かわいい」、「すごい」、「ほしい」という8つの単語を抜き出した。これらを震災前と震災後で集計を行い、それらの特徴を記した。

4. 研究成果

4.1 都市の発言に対する分析

6つの動詞を用いて渋谷、池袋、新宿、秋葉原、六本木、銀座に対して、可視化した。各々の街では固有の動作が出てきていた。例



上左：ありがたい 上中：うれしい
 上右：おもしろい 中左：かわいい
 中中：こわい 中右：すごい
 下左：たのしい 下中：ほしい

図 5 感性表現単語 出現位置

表 2 感性表現単語ごとの震災前後における共起回数比較

| 順位 | こわい | | | | ほしい | | | | | | | |
|----|------------|------|------------|------|------------|------|------------|------|-----|----|----|-----|
| | 震災前 共起数(回) | | 震災後 共起数(回) | | 震災前 共起数(回) | | 震災後 共起数(回) | | | | | |
| | 被災地周辺 | それ以外 | 被災地周辺 | それ以外 | 被災地周辺 | それ以外 | 被災地周辺 | それ以外 | | | | |
| 1 | 明日 | 2 | 155 | 笑 | 18 | 413 | AppBank | 14 | 346 | 人 | 22 | 768 |
| 2 | 顔 | 2 | 113 | 一 | 11 | 354 | アップル | 14 | 229 | 一 | 19 | 479 |
| 3 | 音 | 2 | 110 | 人 | 8 | 273 | 福袋 | 14 | 204 | 笑 | 16 | 340 |
| 4 | 耳 | 1 | 108 | 地震 | 8 | 171 | アプリ | 7 | 192 | 今日 | 14 | 330 |
| 5 | 人 | 1 | 86 | 耳 | 6 | 171 | iPhone | 6 | 190 | 今 | 13 | 269 |
| 6 | 今日 | 1 | 70 | 今日 | 6 | 161 | ノ | 6 | 141 | 福島 | 13 | 246 |
| 7 | * | 1 | 66 | w | 6 | 139 | o | 5 | 128 | 情報 | 12 | 204 |
| 8 | 前 | 1 | 52 | * | 5 | 137 | 一 | 4 | 116 | 車 | 11 | 193 |
| 9 | 気 | 1 | 51 | 今 | 5 | 125 | 写真 | 4 | 105 | 仙台 | 11 | 189 |
| 10 | 家 | 1 | 49 | 車 | 5 | 106 | 笑 | 4 | 101 | 原発 | 11 | 171 |
| 11 | 一番 | 1 | 47 | 揺れ | 5 | 96 | DK | 3 | 97 | 自分 | 10 | 141 |
| 12 | まじ | 1 | 45 | 前 | 4 | 80 | ハジレット | 3 | 93 | o | 10 | 136 |
| 13 | 外 | 1 | 32 | T | 4 | 77 | デノマーク | 3 | 93 | 時間 | 9 | 135 |
| 14 | 店 | 1 | 30 | 夜 | 4 | 70 | ガングラッシュ | 3 | 88 | T | 8 | 106 |
| 15 | 会社 | 1 | 30 | 余震 | 4 | 66 | リュック | 3 | 82 | 気 | 7 | 94 |
| 16 | y | 1 | 27 | 一 | 4 | 66 | 稲本 | 3 | 73 | 家 | 7 | 93 |
| 17 | 病院 | 1 | 26 | 逆 | 4 | 65 | 河原町 | 3 | 68 | 感じ | 6 | 88 |
| 18 | 普通 | 1 | 26 | 震度 | 4 | 61 | 勤め | 3 | 58 | 石巻 | 6 | 88 |
| 19 | 無事 | 1 | 26 | 最近 | 3 | 56 | 間 | 3 | 57 | 相馬 | 6 | 87 |
| 20 | え | 1 | 25 | 東京 | 3 | 53 | 今日 | 3 | 51 | 明日 | 5 | 83 |

(濃い網掛け：災害に関係する単語 薄い網掛け：顔文字などに使われる記号)

ていた。「ほしい」という単語も震災前では被災地近辺では製品名が多いが、震災後は被災地名が出てきていた。それに対して、被災地以外の場所では同様な傾向は見られないため、被災地との関心の違いが伺えた。

5. 主な発表論文等

〔雑誌論文〕(計 1 件)

Kikuchi K., Kihara T., Enta A., Takayanagi H., Kimura T., Hayashida K. and Watanabe H., Collective Background Extraction for Station Market Area by

表 3 感性表現単語ごとの震災前後における記事数増減比

| | 記事数 [件] | | | | 増減比 [] | |
|-------|---------|-------|-------|-------|---------|-------|
| | 震災前 | | 震災後 | | 震災後/震災前 | |
| | 被災地周辺 | それ以外 | 被災地周辺 | それ以外 | 被災地周辺 | それ以外 |
| ありがたい | 12 | 664 | 61 | 944 | 5.083 | 1.422 |
| うれしい | 45 | 2283 | 123 | 3521 | 2.733 | 1.542 |
| おもしろい | 54 | 3762 | 110 | 3911 | 2.037 | 1.040 |
| かわいい | 48 | 3143 | 195 | 5843 | 4.063 | 1.859 |
| こわい | 28 | 2014 | 130 | 3574 | 4.643 | 1.775 |
| すごい | 134 | 7266 | 284 | 8332 | 2.119 | 1.147 |
| たのしい | 80 | 5121 | 211 | 8081 | 2.638 | 1.578 |
| ほしい | 101 | 4911 | 285 | 6605 | 2.822 | 1.345 |
| 合計 | 502 | 29164 | 1399 | 40811 | 2.787 | 1.399 |

Using Location Based Social Network,
Journal of Civil Engineering and
Architecture, Vol.7, No.3, 282-289、査読有

〔学会発表〕(計 10 件)

菊地 弘祐、木原 己人、遠田 敦、高柳 英明、木村 謙、林田 和人、渡辺 仁史、LBSN を用いた都市利用者の特徴単語と都市の機能との関係抽出、2013年度日本建築学会情報システム技術委員会 第35回情報システム・利用・技術シンポジウム、11、2013.12、東京

林田 和人、遠田 敦、菊地 弘祐、木原 己人、木村 謙、高柳 英明、渡辺 仁史、LBSN を用いた駅周辺地域の類型化とその時間変化、2013年度日本建築学会情報システム技術委員会 第35回情報システム・利用・技術シンポジウム、12、2013.12、東京

遠田 敦、菊地 弘祐、林田 和人、木原 己人、木村 謙、高柳 英明、渡辺 仁史、ソーシャルメディアに投稿された記事に基づく都市・建築空間像の解析～東日本大震災の前後に投稿された記事における感性表現単語と地理情報との関係～、2013年日本建築学会全国大会 学術講演梗概集(情報システム技術) 11014、29-32、2013.08、北海道、招待講演

遠田 敦、菊地 弘祐、林田 和人、木原 己人、木村 謙、高柳 英明、渡辺 仁史、ソーシャルメディアに投稿された記事に基づく都市・建築空間像の解析～建築・都市防災メディアとしての活用方策に関する基礎的検討～、2013年度日本建築学会関東支部、5038、2013.09、愛知

菊地 弘祐、木原 己人、遠田 敦、高柳 英明、木村 謙、林田 和人、渡辺 仁史、ソーシャル・ネットワーキング・サービスを用いた都市・建築空間像の抽出 その1～利用者背景の分析と可視化の基本的手法の解明～、2012年度日本建築学会関東支部、5036、2013.03、東京

菊地 弘祐、木原 己人、遠田 敦、高柳 英明、木村 謙、林田 和人、渡辺 仁史、ソーシャル・ネットワーキング・サービスを用いた都市・建築空間像の抽出 その2～利用者背景分析と駅商圈分析手法の開発～、2012年度日本建築学会関東支部、5037、2013.03、東京

菊地 弘祐、木原 己人、遠田 敦、高柳 英明、木村 謙、林田 和人、渡辺 仁史、SNS 情報に基づく都市・建築空間像の解析、2012年度日本建築学会情報システム委員会 第34回情報システム・利用・技術シンポジウム、72、2012.12、東京

Kikuchi K., Kihara T., Enta A., Takayanagi H., Kimura T., Hayashida K. and Watanabe H., Collective Background Extraction for Station Market Area by

using Location Based Social Network, 2012, The 9th International Symposium on Architectural Interchanges in Asia 2012, G-5-4, 2012.10, Gwangju, South Korea, 査読有

菊地 弘祐、木原 己人、遠田 敦、高柳 英明、木村 謙、林田 和人、渡辺 仁史、都市・建築像解析に向けた単語辞書データベース構築、2012年度日本建築学会全国大会 学術講演梗概集(情報システム技術)、11008、15-16、2012.08、東京

Kikuchi K., Kihara T., Enta A., Takayanagi H., Kimura T., Hayashida K. and Watanabe H., A New Method for Analyzing the Relationship between City and Human Behavior using Geo-Tagging Social Networking Service, International Union of Architecture The 24th World Congress of Architecture Academic Session, 10074, 388-393, 2011.10, Tokyo, Japan 査読有

〔図書〕(計 1 件)

菊地 弘祐、Location Based Social Network による都市分析 都市・建築分野でのビッグデータ解析の試み、2013年度日本建築学会全国大会情報システム技術部門研究集会資料 建築におけるセンシングからビッグデータまでを考える、60-67、2013

〔その他〕

2012年度 日本建築学会関東支部研究発表会 若手優秀研究報告賞 受賞

6. 研究組織

(1) 研究代表者

渡辺 仁史 (WATANABE, Hitoshi)

早稲田大学・理工学術院・教授

研究者番号：40063804

(2) 研究分担者

林田 和人 (HAYASHIDA, Kazuto)

早稲田大学・理工学術院・准教授

研究者番号：10277759

木村 謙 (KIMURA, Takeshi)

早稲田大学・理工学術院・講師

研究者番号：10277824

高柳 英明 (TAKAYANAGI, Hideaki)

滋賀県立大学・環境科学部・准教授

研究者番号：70344968

遠田 敦 (ENTA, Atsushi)

東京理科大学・理工学部・助教

研究者番号：90468851