

1 研究目的、研究方法など

本研究計画調書は「小区分」の審査区分で審査されます。記述に当たっては、「科学研究費助成事業における審査及び評価に関する規程」（公募要領 19 頁参照）を参考にすること。

本研究の目的と方法などについて、4 頁以内で記述すること。

冒頭にその概要を簡潔にまとめて記述し、本文には、(1) 本研究の学術的背景、研究課題の核心をなす学術的「問い」、(2) 本研究の目的および学術的独自性と創造性、(3) 本研究の着想に至った経緯や、関連する国内外の研究動向と本研究の位置づけ (4) 本研究で何をどのように、どこまで明らかにしようとするのか、(5) 本研究の目的を達成するための準備状況、について具体的かつ明確に記述すること。

(概要)

本研究では、工場センサー、IoT デバイス、Web 上のオンライン活動等、様々なドメインから生成される多種多様な時系列ビッグデータに対して、汎用的な埋め込み表現を獲得し複数のタスクに対応できる新たなモデルおよびアルゴリズムの開発を目的とする。具体的には以下の課題に取り組む：①大規模時系列データの状態や時系列同士の関係を解釈性の高い形に可視化するような埋め込み表現を学習し、②その埋め込み表現を用いることで分類・回帰・異常値検出のタスクに対し同時に適用できる汎用アルゴリズムの開発を行う。③さらに、刻々と生成され続けるデータに適応するために、埋め込み表現を更新可能にすることを目指す。

(本文)

(1) 本研究の学術的背景、研究課題の核心をなす学術的「問い」

[学術的背景]

近年、IT 技術の急速な進展により、様々な産業、社会活動においてデータ量が飛躍的に増大している。製造業における高度 IoT センシングデバイス導入によるスマート工場化、Web や SNS 上で大量に生成される巨大ソーシャルネットワークにおける社会活動データ、医療現場における患者の生体データ、等数多くの分野で様々なデータが生み出されている。いずれの分野においても、観測されたデータの順序とイベントが記載された「時系列」が重要な意味を持ち、これらを解析し予測や分類、異常値検知などの応用タスクを行うことでビッグデータ活用が実現している。

ビッグデータ活用における時系列解析の重要な課題の 1 つは、様々なドメイン・分野から生成されるデータに適用可能な汎用的なモデル・アルゴリズムを構築することである。時系列データは様々なドメインから大量に生成されているが、各ドメインごとに最適なモデルの構築を行わずに汎用的なモデルを用いて解析することは、モデル構築コストの削減につながる。例えば、工場の生産ラインで得られるデータとインフルエンザ罹患患者数のデータは同じ「時系列」という形式で表現できるものだが、季節性の存在やノイズの存在などによって全く異なるデータとなる。理想的には、データの種類の区別なく、依存関係を持つ時系列の関係やパターンの種類を明らかで、分類や予測などの様々なタスクに容易に適用可能とするモデルが存在すると便利である。しかし、実際には、1 ドメイン 1 タスクごとに最適なモデル設計を行うのが一般的である。

[学術的「問い」]

本研究の核心をなす学術的「問い」は、「様々な時系列タスクに適用可能な、非定常な時系列データの汎用的な埋め込み表現構築は可能であるか？」である。実社会の多くの時系列は、様々な時系列パターンで構成され時間とともに傾向が変化する非定常性を有し、時系列同士の依存関係も複雑である。加えて、それらはデータに応じて、サンプリング頻度や期間、時系列の性質も大きく異なる。このような実社会に存在する各々の特性を持った時系列を、汎用的な表現方法に埋め込むことは、パターン抽出や下流タスクである予測や分類タスクの適用を容易にすることもからも重要である。

本研究の達成において必要な具体的な要件は、①入力される時系列のデータ特性に依存することなく、時系列に含まれるパターンや相互作用を解釈性の高い形で示すこと、②時系列予測、分類、異常値検知など時系列に関する下流タスクに容易に適用可能であること、③実社会の時系列

【1 研究目的、研究方法など（つづき）】

データのように刻々と変化していく状況に適応し、常に最適な結果を得られること、の3点である。特定のドメインのデータやモデルにおいて、これらの問題を個別に扱う既存理論は存在するが、これらの要件を同時に達成するような、時系列解析を容易にする埋め込み表現の開発には、依然として多くの課題が残されている。

(2) 本研究の目的および学術的独自性と創造性

[本研究の目的]

上記のような学術的背景を踏まえ、本研究では、**時系列データの解釈・意思決定・将来予測などの様々な下流タスクに適用可能な、データ特性に依存しない時系列データの埋め込み表現モデル、および、動的な埋め込み表現モデル更新**の技術を確立する。より具体的には、以下の項目に沿って研究を行う。

- 研究項目①：**データ特性に依存しない汎用的な時系列データの埋め込み表現モデル**
特定の点もしくは特定期間の時系列データを埋め込み表現に変換することで、時系列の中に存在するパターンや時系列間の関係の可視化を実現する。
- 研究項目②：**埋め込み表現に基づく下流タスクの汎用学習アルゴリズム**
下流タスクにおいても有効な埋め込み表現を獲得するために、埋め込み表現の構造化と下流タスクにおいても有効な事前学習を提案する。
- 研究項目③：**動的に更新可能な埋め込み表現モデルへの拡張**
研究項目①・研究項目②で開発した基盤技術をベースとし、逐次的にデータが送られてくるストリームデータへ対応するために、学習済み埋め込み表現モデルを動的に更新可能な形に拡張することを目指す。

[学術的独自性と創造性]

最近のAI関連の技術では、画像や自然言語処理、ロボット等の分野を中心に深層学習を用いた研究が活発に行われ、Webサービス、金融、医療・ヘルスケア等、急速に応用範囲が拡大している。時系列解析の分野では予測・分類精度の高さから深層学習も多く活用されるようになった一方で、他分野と比較すると、状態空間モデルなどの統計的手法が依然として中心の技術となっている。これは、深層学習技術では達成が困難であった学習結果の説明性や刻々と変化する環境の中で柔軟に適応するリアルタイム学習などが、時系列解析において重要視されていることが要因である。

そこで本研究は、性能の高い深層学習で説明性を実現するために、**時系列解析と、自然言語処理などの分野で発展してきた埋め込み表現学習の融合を試みる独自性の強いアプローチ**を検討する。データ特性に依らない時系列データの埋め込み表現を構築することで、時系列パターンやその遷移過程、傾向を容易に理解できる。また、構築した埋め込み表現モデルをストリームデータの時系列解析の手法と組み合わせることで、難易度の高いオンライン学習の実現を試みる。本研究は、一般的なアプローチで多くの時系列データに適用できる可能性が高いことから、ドメインに応じてモデル設計する必要が無くなり**人的・時間的コストの削減につながるなど大きな社会的インパクトも期待**できる。

(3) 本研究の着想に至った経緯、関連する国内外の研究動向と本研究の位置付け

[本研究の着想に至った経緯]

これまで、申請者は時系列データ解析に関する研究に従事し、主に深層学習を用い予測性能の向上を達成しつつ解釈性の高さも両立させるモデルを模索してきた。位置情報の有効性を明らかにする感染症予測モデルの構築 (PLoS ONE'21 [5]) やフェイクニュースの拡散の傾向を明らかにする時系列モデルの構築 (FakeModel@PLoS ONE'21 [4]) など、特定ドメインにおいて有効な時系列モデルの構築に取り組んだ。一方で、ドメインに応じたモデル構築の労力の高さも理解し、ドメインに依存しない時系列モデルを構築することを意識することとなった。そして、異なる傾

【1 研究目的、研究方法など（つづき）】

向を持つ複数の時系列データを1つの深層学習モデルで学習する時系列モデル (Single@ECML-PKDD'21 [3]) や拡散反応方程式に基づいて多変量時系列間の相互作用を抽出する時系列モデル (FluxCube@CIKM'22 [1]) など、深層学習ベースでありながらも汎用性と解釈性を両立するようなモデルを構築してきた。本研究では、これまで開発した要素技術を深化させ、時系列データの汎用的な埋め込み表現のためのモデル、および、アルゴリズムを開発する。

申請者は複数件の共同研究を通して、解釈性が高く更新可能な時系列解析手法の構築に需要を感じている。時系列ビッグデータを用いた要因分析によって設備故障時期の予測などを行う際に、高い予測性能だけでなく根拠となった特徴をヒトが見て理解できることが重要となる。また、コロナ禍などの外部状況によって生じうる、時系列ビッグデータを構成するパターンの変化に応じてモデルを更新していくことも重要となる。これらの**社会的な需要の高さと申請者のこれまでの研究経歴を踏まえて本研究の着想を行い**、実社会の様々な時系列データに適用できる革新的な技術を社会に向けて創出することを目標とする。

[国内外の研究動向と本研究の位置付け]

[本研究の目的] における研究項目①と研究項目②は時系列データの深層学習技術と埋め込み表現学習の技術の延長線上に存在する。また、研究項目③は時系列深層学習モデルに、Stream Algorithm、もしくはContinuous Learningの枠組みの適用を検討する挑戦的なものである。関連するトピックの研究動向を以下に示す。

- 時系列データの深層学習技術：再帰的構造を持つ Recurrent Neural Network (RNN) をきっかけに、深層学習に基づく時系列解析は発展してきた。近年では、多センサ間の依存関係を捉える Graph Neural Network (GNN) や、長期依存性・長期予測を達成する Transformer ベースのモデルの開発が盛んである。深層学習ベースのモデルは非構造データ含む多様なデータを柔軟に扱い下流タスクの性能を向上させた一方で、**モデルの解釈性が低いことが課題**となっている (Salinas et al. Journal of Forecasting'20, de Bézenac et al. NeurIPS'20)。
- 埋め込み表現学習：与えられたデータを n 次元の連続ベクトル空間の1点に埋め込む手法で、主に自然言語処理で発展してきた手法である。単語をベクトル表現とし単語同士の意味の距離を測定する Word2Vec という画期的な手法 (Mikolov et al. NAACL'13) が提案され、以後多くの埋め込み表現学習手法 (Pennington et al. EMNLP'14, Bojanowski et al. TACL'17) が提案されてきた。最近、時系列データの埋め込み表現学習も検討され始めている (Franceschi et al. NeurIPS'19, Zerveas et al. KDD'21, Shao et al. KDD'22) が、発展途上の分野であることから、**適応できるタスクの少なさや解釈性の点を重視していないといった課題**が存在する。
- Stream Algorithm/Continuous Learning：新しい入力に対し、構築済みモデルのパラメータを更新し、最適な形に収束することを目指す手法は、大量に生成され続ける時系列データ処理において有用である。状態空間モデルのような代表的なものだけでなく、MDLを用いた最適なモデル更新・選択アルゴリズムの提案 (Gong et al. VLDB'17, Kawabata et al. CIKM'19) など統計的時系列モデルの更新手法である Stream Algorithm が多く提案されてきた。一方で、深層学習ベースのモデル更新は、更新によって過去の学習を忘却する破滅的忘却が課題であったが、最近では Continuous Learning という枠組みでパラメータ更新を最適化する試み (Kirkpatrick et al. PNAS'17) が行われている。本研究では、未だに明らかになっていない、**時系列ベースの深層学習モデルにおける最適なモデル更新方法**について挑戦的に取り組む。

(4) 本研究で何をどのように、どこまで明らかにしようとするのか

[研究項目①：データ特性に依存しない汎用的な時系列データの埋め込み表現モデル]

申請者は、非定常時系列データの汎用的な埋め込み表現の構築において、時系列パターンの理解などの解釈を容易にするため、**①入力長に依存しない特性、および、②時系列の構成要素に分解可能な特性**をもつモデルを検討する。①の特性は、時系列の任意の期間を同じ連続ベクトル空間内の点に埋め込むことを目指したもので、入力長に関わらず似たパターンは近いベクトル空間

【1 研究目的、研究方法など（つづき）】

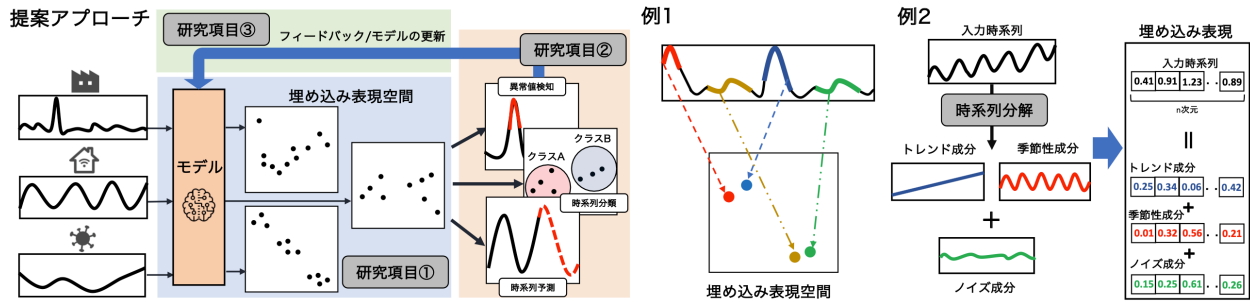


図 1: 左：提案アプローチの概要図，右：提案時系列埋め込み表現の例

に埋め込まれることが期待できる（図1の例1）。また、②の特性について、時系列データはトレンド成分、季節成分、ノイズによって構成されることから、各成分を同一空間の一部に埋め込むことを目指すものである。同時に、各成分の埋め込みベクトルを足し合わせると元の時系列の埋め込みベクトルと近似するように学習することで、時系列データを多角的に捉え、パターン間の類似度や関係性などの構造的な理解が可能となる（図1の例2）。上記の特性を達成するため、**表現能力の高い Transformer ベースのモデルと、時系列の構成要素を明らかにする時系列分解の2つの手法の拡張・融合**を検討し、パターン同士の構造的な理解を可能にするモデルを作成する。そして、共通パターンや似た時系列同士の埋め込み表現が近くなるよう、**距離学習ベースのモデル学習方法**を開発する。

[研究項目②：埋め込み表現に基づく下流タスクの汎用学習アルゴリズム]

研究項目①によって得られる埋め込み表現は、時系列データ自身の特徴の要約であることから、様々な下流タスクに容易に適用できる。しかしながら、構築された表現が予測・分類・異常値検知などの下流タスクに最適であるとは限らない。抽出された埋め込み表現を様々な下流タスクで性能高く利用するためにも、表現学習の段階で下流タスクにおける有用なパターンの情報を同時に抽出し埋め込むことが必要である。そこで申請者は、**下流タスクにおいて有用なパターンを推定・抽出するための、埋め込み表現学習の事前学習手法、もしくは、最適化関数を開発**することで、汎用的な学習アルゴリズムを実現する。

[研究項目③：動的に更新可能な埋め込み表現モデルへの拡張]

最後に、逐次的にデータが送られてくるストリームデータへの対応も見据え、研究項目①と研究項目②で構築した埋め込み表現モデルを動的に更新するためのアルゴリズムを開発する。外部要因や時間経過によって生じる時系列データ中のパターン構造の変化は、下流タスクの結果に影響を与えるため、新しく送られてくるデータに応じてモデルを動的に更新する必要がある。例えば、新たな時系列データの各成分のうち、季節性成分は変化せずトレンド成分のみが変化した場合、トレンド成分の埋め込みに用いられるモデルパラメータのみ更新して徐々にモデルを変化させていくべきである。モデルの更新を最小限に抑えることで、データの変化も容易に解釈できる。上記のようなモデル更新を実現するために、**Stream Algorithm や Continuous Learning の枠組みを発展させる**ことで取り組む。

申請者は上記の研究項目に取り組むことで、最終的な目標である**様々な時系列タスクに適用可能、かつ、動的に更新可能な時系列データの埋め込み表現モデル**の完成を目指す。

(5) 本研究の目的を達成するための準備状況

申請者はこれまでの研究を通じて解釈性の高い時系列解析の知見を得ている。(FluxCube@CIKM'22 [1])では拡散反応方程式を深層学習を用いて拡張することで、時系列データ間の関係性の発見と高い予測性能を実現した。(Single@ECML-PKDD'21 [3])では時系列予測において有用な特徴量を発見する手法を提案した。また、埋め込み表現技術についても、自然言語処理に関する研究(Mitigaiton@W-NUT'21)で扱っており、これらの技術を深化させ本研究に取り組む予定である。

2 応募者の研究遂行能力及び研究環境

応募者の研究計画の実行可能性を示すため、(1)これまでの研究活動、(2)研究環境（研究遂行に必要な研究施設・設備・研究資料等を含む）について2頁以内で記述すること。

「(1)これまでの研究活動」の記述には、研究活動を中断していた期間がある場合にはその説明などを含めてもよい。

(1) これまでの研究活動

申請者は、時系列を中心としたビッグデータを用いることで自然現象や社会現象の時間発展を解析し、予測・モデリングするための先駆的な技術の開発や実社会問題への適用に取り組んでいる。これまで開発した技術は、データマイニング分野のトップ国際会議である CIKM, ECML-PKDD や、有名オープンアクセスジャーナルの PLoS ONE, Scientific Reports, JMIR, そして実社会問題への適用として他分野の有名国際会議である LREC (自然言語処理) や IC2S2 (計算社会科学) など幅広く継続的に採択され、国際的に高く評価されている。最近では、時系列データのモデリング技術 (FluxCube@CIKM'22 [1]) を開発した。これは拡散反応方程式を深層学習によって表現力を向上させたもので、予測精度の向上や時系列データの隠れた相互関係の発見を達成している (図2)。申請者は、時系列データの基礎技術研究のみでなく、ソーシャルメディア投稿の時系列を用いたフェイクニュース検出・モデリング [4,6,11] や感染症流行の将来予測 [3,5,8] など、様々な社会問題に対しても適切な時系列技術を開発することで積極的に取り組んでいる。本研究においても、これまでの時系列に関する基礎研究や実社会応用の経験を活かし、基礎研究の開発・実用化に取り組む。

申請者の研究業績は様々な形で評価されている。第10回 Web とデータベースに関するフォーラム企業賞、第13回データ工学と情報マネジメントに関するフォーラム学生プレゼンテーション賞などの受賞。さらに、2020年度においてはデータベース学会で優秀な研究発表を行った若手会員に贈呈されるコンピュータサイエンス領域奨励賞を受賞した。受賞以外においても、国際ワークショップの1つである“Social Computing Workshop 2021 -Social Computing in the post-pandemic era-”での講演や2022年セキュリティ・キャンプ全国大会での講師を務めるなど、研究以外の社会活動も積極的に行っている。

コミュニケーション能力を生かした他研究者との連携に関しても研究計画の実行可能性を高めている。これまで、申請者は多くの研究者と共同研究を行っており、彼らと積極的に交流・議論を深めることによって多くの論文文化を行ってきた。Yahoo! Japan との共同研究によって [5,8,9] を、小林亮太准教授との共同研究で [4] の成果を実現している。最近では、千葉大地教授のCRESTへの参加や、若手研究者同士による勉強会や共同研究も行っている。このような他研究者との交流・協力により、本研究もより高いレベルで遂行することが期待できる。

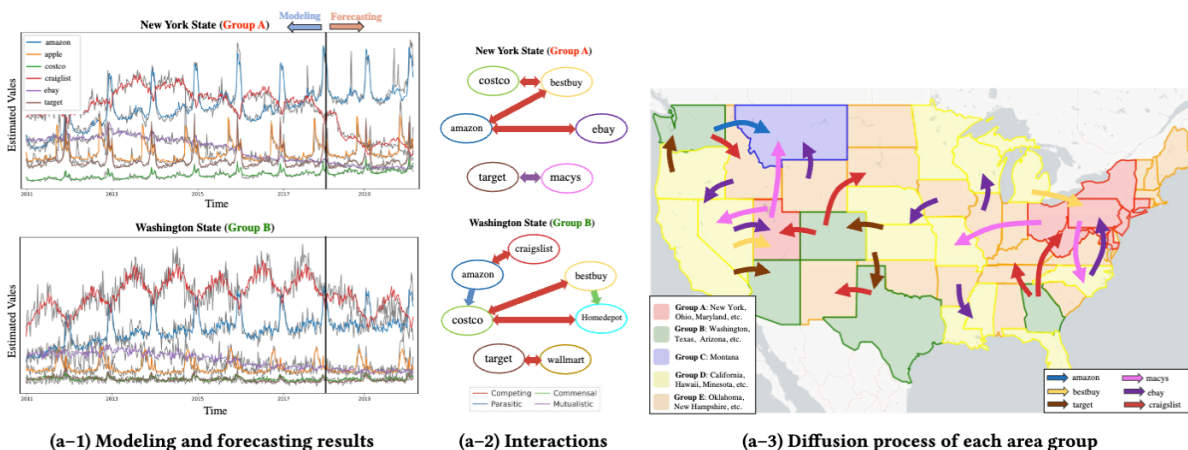


図2: Web データから抽出したキーワード同士の関係の発見例 (FluxCube@CIKM'22 [1])

【2 応募者の研究遂行能力及び研究環境（つづき）】

[関連する主な論文・国際会議]

1. Taichi Murayama, Yasuko Matsubara, Yasushi Sakurai, “Mining Reaction and Diffusion Dynamics in Social Activities”, In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM), (2022).
2. Taichi Murayama, Shohei Hisada, Makoto Uehara, Shoko Wakamiya, Eiji Aramaki, “Annotation-Scheme Reconstruction for “Fake News” and Japanese Fake News Dataset”, In Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC), (2022).
3. Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, “Single Model for Influenza Forecasting of Multiple Countries by Multi-task Learning”, The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), (2021).
4. Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, Ryota Kobayashi, “Modeling the Spread of Fake News on Twitter”, PLoS ONE, 16(4): e250419, (2021).
5. Taichi Murayama, Nobuyuki Shimizu, Sumio Fujita, Shoko Wakamiya, Eiji Aramaki, “Predicting Regional Influenza Epidemics with Uncertainty Estimation using Commuting Data in Japan”, PLoS ONE, 16(4): e250417, (2021).
6. Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, Ryota Kobayashi, “Modeling the Spread of Fake News on Twitter”, 7th International Conference on Computational Social Science (IC2S2), (2021).
7. Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, “Comparable Analysis of News Diffusion between Mainstream and Alternative Media in Twitter”, 7th International Conference on Computational Social Science (IC2S2), (2021).
8. Taichi Murayama, Nobuyuki Shimizu, Sumio Fujita, Shoko Wakamiya, Eiji Aramaki, “Robust two-stage influenza prediction model considering regular and irregular trends”, PLoS ONE, 15(5): e0233126, (2020).
9. Shohei Hisada, Taichi Murayama, Kota Tsubouchi, Sumio Fujita, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki, “Surveillance of early stage COVID-19 clusters using search query logs and mobile device-based location information”, Scientific Reports, 10(1): 18680, (2020).
10. Kazuya Taira, Taichi Murayama, Sumio Fujita, Mikiko Ito, Kei Kamide, Eiji Aramaki, “Comparing medical term usage patterns of professionals and search engine and community question answering service users in Japan: Log analysis”, Journal of Medical Internet Research (JMIR), 22(4): e13369, (2020).
11. Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, “Fake News Detection using Temporal Features Extracted via Point Process”, In Proceedings of CySoc 2020 International Workshop on Cyber Social Threats (ICWSM Workshop), (2020).

(2) 研究環境

申請者の所属研究室(大阪大学産業科学研究所)は時系列モデル構築やリアルタイム予測などの時系列解析を専門としており、本研究を行うのに適した技術や資源に恵まれている。また、所属研究室では産業科学 AI センターが発足、それに伴い人工知能技術を介した他分野との共同研究が複数件実施された。共同研究ではスマート工場や車載センサなどの時系列データを活用した産業貢献を向けた取り組みを積極的に行っており、申請者は様々な実データを扱う機会や研究者・経営者らと議論する場があり、データ解析の需要や要望について理解を深めることができる。

3 人権の保護及び法令等の遵守への対応（公募要領 4 頁参照）

本研究を遂行するに当たって、相手方の同意・協力を必要とする研究、個人情報の取扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究など指針・法令等（国際共同研究を行う国・地域の指針・法令等を含む）に基づく手続が必要な研究が含まれている場合、講じる対策と措置を、1 頁以内で記述すること。

個人情報を伴うアンケート調査・インタビュー調査・行動調査（個人履歴・映像を含む）、提供を受けた試料の使用、ヒト遺伝子解析研究、遺伝子組換え実験、動物実験など、研究機関内外の倫理委員会等における承認手続が必要となる調査・研究・実験などが対象となります。

該当しない場合には、その旨記述すること。

本研究では、ウェブ上で公開されているパブリックデータを使用する。この際、個人が特定されることのないよう匿名化を行い、利用規約を遵守する。プライバシー情報の取り扱いについては、我が国が定める個人情報保護法を遵守し、研究機関における情報セキュリティ関連の委員会等の監督の下、厳重に管理するものとする。