

令和 5 年 5 月 25 日現在

機関番号：34310

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K11720

研究課題名（和文）メッシュ位置情報ビッグデータの分析法に関する基礎的研究

研究課題名（英文）Fundamental research for analysis of mesh location data

研究代表者

宿久 洋（Yadohisa, Hiroshi）

同志社大学・文化情報学部・教授

研究者番号：50244223

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究では、位置情報データの分析法に関する基礎的な研究を行なった。特に、「平常」な状態とのギャップに着目し、そのギャップを明らかにするための手法の開発を行なった。その中で、位置情報データの解析手法を改良し、地点の予測の安定性を向上させた手法の開発、オンライン学習に基づいた異常検知のための手法の開発を行なった。これらの手法の開発に加えて、実際にGPSデータを用いて得られたメッシュの異常検知を行い、実際の災害時の人流変化を反映した結果を得ることができた。本研究で開発した手法は、メッシュデータの利用だけでなく、地理座標が付与された位置情報データにも適用可能である。

研究成果の学術的意義や社会的意義

位置情報データは、その取得容易性から収集できるデータ自体は大きいもののそれを活用することは未だ困難な点が多いのが現状である。本研究では、異常検知の観点から位置情報データの解析手法の開発を行い、実際の位置情報データを用いて異常を検知できることを確認した。

本研究で提案した手法により、災害や大規模イベント発生時など、「平常」とされる状態とは異なる事態が発生した際にいち早くその状態を検出することが可能になり、二次災害を防ぐことや災害発生時のリスクマネジメントに役立つことが期待される。

研究成果の概要（英文）：In this study, we conducted fundamental research on methodologies for analyzing location data. Specifically, we focused on the disparity between "normal" and "abnormal" conditions and developed a method to clarify the factors contributing to this discrepancy. We devised a technique to improve the stability of location prediction and an approach for online learning-based anomaly detection. Additionally, we performed anomaly detection on a mesh using GPS-acquired people flow data, which accurately reflected variations in people flow during real disasters. The methodology developed in this study is applicable not only to mesh-based data but also to location data represented by geographic coordinates.

研究分野：統計科学

キーワード：位置情報データ 時空間データ解析 オンライン学習 異常検知 メッシュ統計学 人流データ

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年、情報技術の発達に伴って、GPS 等を用いてスマートフォンなどから容易に位置情報データの取得が可能となった。位置情報データとは、そのデータの取得地点に紐づいて得られるデータのことである。多くの位置情報データでは、ある地点で得られた値がその位置情報座標と共に得られている。位置情報データの各対象は空間的な関係性を持っていることが多く、収集される地点同士が関係を有していることが多い。そのため、位置情報データは通常の変量データとは異なりそのままの形で既存の分析手法の適用を行うことは難しい。活用の現状に関する問題として、集計ベースの分析法の活用が多いことが挙げられる。また、位置情報データは高頻度かつ広範囲に取得可能であることから、収集できるデータ自体は大きいものの活用においてその全てのデータを使うことは困難である。このことから、多くの活用例においては、データの取得対象を特定移動体(タクシーなど)に限定したもの、あるいは、計測箇所を道路に限定したもの、一定の観測時間に限定したものなど、対象・場所・期間を限定しての局所的な活用に留まっている。その一方で、近年、各対象が地域メッシュであるような位置情報データの活用が期待されている。ここで得られるようなデータをメッシュ位置情報データと呼ぶ。このようなデータは、大規模なものとして得られており、そのようなデータをメッシュ位置情報ビッグデータと呼ぶ。

メッシュ位置情報ビッグデータを分析する際に次の2つの問題がある。1つ目の問題として、位置情報データのもつ共通の構造とはどのようなものなのかということが挙げられる。2つ目の問題として、位置情報をもとにして特定の時間・場所における対象の状態を推定できるのかということが挙げられる。1つ目の問題について、既存の位置情報データの分析においては、ある特定の地域のみのような局所的なデータが取得され、そのデータによって交通の分析などが行われてきた(e.g. Liu, et al., 2012)。しかし、地域間で何か潜在的な共通している傾向の存在が考えられるため、隣接関係のみではなく、各地点から得られる様々な要素を考慮して分析を行う必要がある。2つ目の問題について、平日朝・夜などの特定時間、駅・空港などの特定場所の情報が得られた際にデータの取得対象の状態を推定することは重要な課題である。そして、その状態に関する情報はその後、推定された結果を用いて将来の予測や異常検知など様々な利活用されることが期待される。そのため、良い精度での状態推定を行う必要がある。

2. 研究の目的

先に述べた問題に対して、本研究では、特に「平常」とのギャップに着目した位置情報データ分析に関する分析手法について基礎的な研究を行う。本研究課題では、次の目的を達成することを目指す。

「平常」な状態とのギャップを発見するための位置情報データ分析のための手法開発
実世界の位置情報データに対して、開発手法を適用することによる問題解決

1つ目の研究目的は、本研究課題の中の核となるものである。

また、ここで考える手法に対しては、先に述べた本研究課題をめぐる問題意識に関して、与えられている位置情報データの仮定として次の2つが考えられる:

(ア)外部情報を持たない位置情報データ、

(イ)外部情報を持つ位置情報データ。

状況(ア)については、先に述べた1つ目の問題に対応するものである。得られている位置情報データのみを用いることにより、空間的な関係性をもつデータの構造を明らかにするためのデータ分析法を開発することを目指す。具体的には、空間的な関係ないし時間的な関係を持つ単変量のデータを対象とした「平常」の推定を目指す。

状況(イ)については、先に述べた2つ目の問題に対応するものである。実際の位置情報データでは、そのデータが取得された背景となる情報や、位置情報データの中で説明変数となりうる情報が含まれていることがある。これに加えて、すでに平常な状態か異常な状態かというラベル情報が与えられた状態もこの状態に含む。このようなデータを用いて、「平常」の推定と「平常」と「異常」の違いを明らかにすることを目指す。

1つ目の研究目的における、状況(ア)、(イ)のいずれの状況に対しても先述の方法により、位置情報データからその場所の平常状態の定義を与えることができる。このことにより、位置情報データに対する異常検知が可能になる。そして、災害をはじめとする異常事態の早期発見が可能となる。

2つ目の研究目的は、本研究課題の中で、社会実装を目指すためのものである。実際に得られたデータを用いて、「平常」とされる状態の定義とその状態からの乖離を明らかにするための分析を行う。このことにより、位置情報データを用いた実世界の「異常」とされる状態の検知を目指す。

これら2つの目的の達成によって、異常事態の早期発見や社会実装に貢献し、位置情報データを活用した実社会における問題解決に寄与することが期待される。

3. 研究の方法

先述の目的を達成するための方法として、本研究課題では次に述べる3つの課題に取り組む。

- 研究1: 得られている位置情報データを用いた未観測地点の推定
- 研究2: オンライン学習による新規観測の予測
- 研究3: メッシュ型位置情報データを用いた異常検知

研究1は、先の研究目的を達成するための基礎的な研究である。その中で、ここで考える位置情報データの仮定としては、状況(ア)および状況(イ)の両方に対応するものである。具体的には、地球統計学で用いられる手法であるクリギングの方法を用いることにより、位置情報データの特性に応じたモデリング手法の開発を行う。その中で、データが得られている各地点のクラスタリング方法や外部の説明変数を用いた予測方法を検討する。

研究2は、先の研究目的を達成するための基礎的な研究である。その中で、ここで考える位置情報データの仮定としては状況(イ)に対応するものである。具体的には、オンライン線形判別分析の改良手法を開発し、位置情報データの分布の非対称性を考慮した異常検知を行う。特に、中央値や分位点を用いることで、平常とされる状態において分布の右側に裾を引くデータに対する異常な状態の判別を行うための手法を検討する。

研究3は、先の研究目的を達成するための研究であり、実用性に重きを置いた研究である。ここでは、特に状況(ア)の状態を考える。具体的には、GPSを用いて取得されたメッシュ型位置情報データを対象として、オンライン学習の方法を用いた異常検知を行う。オンライン学習の考え方を活用し、災害時の異常検知に焦点を当て、人流の増加や減少といった観点から現実の状態を反映した結果を得る。

4. 研究成果

本研究では、メッシュ位置情報データを初めとした地理情報データを対象としたデータ解析法の開発に取り組んでいる。本研究で対象としている位置情報データに対して、古典的な多変量解析法を用いてそのまま分析することは、対象とするデータの特性やその量が問題となり困難であることが多い。そのため、データの特性を踏まえたデータ解析法が必要になる。そこで本研究では、先の研究方法で述べた研究1から研究3に関して、次の結果を得た。

研究1に関しては、位置情報データに関する分析手法である Geographically Weighted Regression (Brunsdon et al., 1996; Fotheringham et al., 2002)の改良を行った手法の開発を行った。これは、データが得られている各地点を、クラスタリングした下でモデリングする方法に関する研究である。特に、各地点に関する情報が、緯度経度などの位置を表す情報のみである場合について研究を行った。この方法では、予測にベイズ的な方法を用いることでデータの少ない地点においても安定した予測を行うことが可能になった。平常時を定義するためには、平常とされる状態を推定する必要があり、この状態の推定にはその地点の予測値を得ることが有用である。

研究2に関しては、オンライン学習の手法の一つである、オンライン線形判別分析(Liu and Song, 2019)の改良を行った手法を開発した。この方法については、中央値を用いる方法とさらに拡張した分位点を用いる方法を提案している。これは、得られる位置情報データに対して仮定することのできる分布が必ずしも左右対象な分布ではないという考えに基づいている。この方法により、平常とされる状態を想定した時に右に裾を引く分布に対しても平常である状態と異常である状態の判別を可能にしている。これにより、平常と異常の状態を判別する際に、より柔軟な分析が可能となった。

研究3に関しては、GPSを用いて取得されたメッシュ型位置情報データを用いて、各対象となっているメッシュの異常検知を行った。このメッシュ型位置情報データは逐次的にデータが得られているというデータ取得の背景から、この背景情報を考慮し、我々はオンライン学習の考え方を活用してメッシュが「平常」な状態との違いを明らかにすることに取り組んだ。その結果、2018年に発生した2つの災害に着目して、災害時の異常検知を行い、人流の増加・減少の観点から、現実の状態を反映した結果を得ることができた。

本研究課題開始時点当初の研究計画では、メッシュとして得られた位置情報データを中心に研究を進める予定であった。しかし、研究を進める中で、データの取得単位は必ずしもメッシュとして得られる訳ではなく、メッシュとして得られる位置情報データについても、通常の位置情報データと同様の課題が見られることを発見した。そのため、地理座標が付与された地点として得られている状況でも利用可能な手法の開発を進めた。各地点での観測がメッシュとして得られる状況は、地理座標が付与された位置情報データの特長ケースとして考えることができるため、本研究の成果を用いることで今後メッシュ位置情報データへの応用を行うことは可能であると考えられる。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Sakamoto Kenta, Okabe Masaaki, Yadohisa Hiroshi	4. 巻 192
2. 論文標題 Generalized canonical correlation analysis for labeled data	5. 発行年 2021年
3. 雑誌名 Procedia Computer Science	6. 最初と最後の頁 517 ~ 525
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.procs.2021.08.053	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Goto Satoshi, Takagishi Mariko, Yadohisa Hiroshi	4. 巻 156
2. 論文標題 Clustering for time-varying relational count data	5. 発行年 2021年
3. 雑誌名 Computational Statistics & Data Analysis	6. 最初と最後の頁 107123 ~ 107123
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2020.107123	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Duong Thi Binh An, Tsuchida Jun, Yadohisa Hiroshi	4. 巻 48
2. 論文標題 K-means generalized maximum entropy estimation for structural equation modeling	5. 発行年 2020年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 103 ~ 115
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-020-00118-4	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Otani Ryo, Yadohisa Hiroshi	4. 巻 176
2. 論文標題 Estimation of causal effect using propensity score and weighted-average method	5. 発行年 2020年
3. 雑誌名 Procedia Computer Science	6. 最初と最後の頁 810 ~ 817
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.procs.2020.09.076	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計7件（うち招待講演 0件 / うち国際学会 6件）

1. 発表者名 Sakai, T., Tsuchida, J. and Yadohisa, H.
2. 発表標題 A cluster Kriging method considering additional information
3. 学会等名 The 11th Conference of The Asian Regional Section of the International Association for Statistical Computing (国際学会)
4. 発表年 2022年

1. 発表者名 Sakamoto, K. and Yadoshisa, H.
2. 発表標題 Embedding method for multiple multivariate data by using class label information
3. 学会等名 The 11th Conference of The Asian Regional Section of the International Association for Statistical Computing (国際学会)
4. 発表年 2022年

1. 発表者名 Kawarai, S. and Yadohisa, H.
2. 発表標題 Online quantile linear discriminant analysis
3. 学会等名 The 11th Conference of The Asian Regional Section of the International Association for Statistical Computing (国際学会)
4. 発表年 2022年

1. 発表者名 Hashiguchi, M. and Yadohisa, H.
2. 発表標題 Wasserstein discriminant analysis for limited labeled data
3. 学会等名 The 11th Conference of The Asian Regional Section of the International Association for Statistical Computing (国際学会)
4. 発表年 2022年

1. 発表者名 高澤一平, 森岡優輝, 後藤智紀, 鈴木雄也, 加藤有祐, 宿久洋
2. 発表標題 位置情報データを用いた災害時のオンライン異常検知について,
3. 学会等名 2020年度人工知能学会全国大会(第34回)
4. 発表年 2020年

1. 発表者名 Matsuoka, K., Tsuchida, J., Okabe, M. and Yadohisa, H.
2. 発表標題 inary Quantile Regression for Vector Responses
3. 学会等名 Joint Statistical Meeting 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Okabe, M., Yadohisa, H.
2. 発表標題 Adaptive Cost-Sensitive Logit Boost for Optimizing Performance Metrics
3. 学会等名 Joint Statistical Meeting 2020, (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------