

平成 26 年 6 月 18 日現在

機関番号：12301

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500344

研究課題名(和文) 時空間上の大規模履歴の自己組織化による状態遷移確率モデルの構築

研究課題名(英文) Development of probabilistic transition models using self-organizing of a large-scale historical data on space-time dimension

研究代表者

関 庸一 (Seki, Yoichi)

群馬大学・理工学研究科・教授

研究者番号：90196949

交付決定額(研究期間全体)：(直接経費) 4,000,000円、(間接経費) 1,200,000円

研究成果の概要(和文)：大量の時空間履歴データを分析しようとする際の困難は、その履歴の多様性にある。本研究では、履歴群のもつ位相的構造や部分系列の構造を利用することで、履歴の遷移を確率モデルとして抽出する方法論を与えた。具体的には、マップ範囲をデータの位相構造に適合させて変形させる可変自己組織化法(Flexible SOM [FSOM])を開発し、そのマップ上で遷移として履歴を捉える方法などを与えた。得られた方法論は、流通や金融などでのサービス利用履歴や、レセプトなどの医療サービス履歴、都市空間中の移動履歴などの実データで検証し、その有効性を明らかにした。

研究成果の概要(英文)：When we analyze a large-scale historical data on space-time dimension, there is the difficulty with the variety of the history. In this study, we proposed methodologies to develop probabilistic models of the transitions in histories, using their topological structure and the structure of the subsequence of history. We develop flexible self-organization map (FSOM), which is a SOM whose map range is automatically adjusted by the topological structure of history. Transitions in histories are grasped by the structure derived from the FSOM and so on. We verified our methodologies with the real world data, for example, service use history in merchandising or the finance business, health care service histories such as the receipt, and movement histories in city space.

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：データマイニング 自己組織化 SOM 多項ロジットモデル 医療保険サービス ID付POS 移動履歴
共起行列

1. 研究開始当初の背景

近年、情報収集技術の進展とともに、蓄積される情報が時空間上で稠密なものとなって来ている。つまり、より多くの対象から長期間多面的、かつ、悉皆に収集できるようになってきている。代表的な例としては、流通や金融などでのサービス利用履歴などの時系列上の特徴量の変化履歴や、都市での交通手段利用など 空間中の移動履歴がある。

これらのデータは、その中に多面的な知識を含むと考えられ、適切なモデルと分析方法が提供されれば、各種の意思決定支援の可能性は大きく広がる。しかし、これらの履歴データ全体を通じて、個体の履歴を総括的に追跡して分析した研究はない。

たとえば、サービス利用履歴データに関しては、マーケティングや医療管理分野で既に多くの分析がなされているが、時間軸に関し定常的なモデルが多く、変化を扱う購買ブランドのスイッチモデルなどもあるが、データの特定の一面を抽出して分析するものに留まる。また、空間中の移動履歴のデータについては、データ収集方法の研究段階で、収集されたデータの分析方法はこれからの課題となっている。

これらを分析する際の困難さは、まず、履歴の多様性にある。データが他の業務などに用いるために収集されたトランザクションであり、分析のために計画して選別されたものではないため、背後に多様で異なった生成過程がある場合が多く、全データを一括したモデル化に無理があったからである。これに対し、本研究では、以下の研究成果を基礎として、履歴データを自己組織化することにより、自然な状態分割を実現し、状態遷移確率モデルを構築することを可能とする。

申請者は、上述のようなデータに対し、平成 19~21 年度科学研究費：「大規模履歴データからの因果確率モデル構築のための時間発展類型の抽出」および、それに至る一連の研究において、SOM (Self-Organizing Maps, 自己組織化マップ) を用い、個体集団の細分を作成するアプローチによる事例の蓄積を行って来ている。SOM は、平面上に格子状に配置されたノードごとに、代表類型(参照ベクトル)を求め、個体を最も類似した類型へ層別する方法である。計算量がサンプル数 n に比例のオーダーであるため、多量サンプルデータにも適用可能である点が利点となる。

2. 研究の目的

本研究では、多様性に満ちた大量の時空間履歴データを自己組織化等により、履歴群のもつ位相的構造や部分系列の構造を利用することで、履歴の遷移を確率モデルとして抽出する方法論を与えることを目的とした。得られた方法論は、以下のような実データで検証し、その有効性を明らかとした。

(1) 時系列上の特徴量の変化履歴

(ア) 流通や金融などでのサービス利用履歴(ID 付き POS / クレジットカードの私企業データ)

(イ) 医療の治療・サービス履歴(レセプトの健康保険/行政データ)

(2) 空間中の移動履歴

(ア) 都市での交通手段利用(市民の地点間移動調査結果)

これらのデータは、その中に多面的な知識を含むと考えられ、適切なモデルと分析方法が提供することで、各種の意思決定支援の可能性は大きく広がる。

3. 研究の方法

本研究では、次の 2 点を実施した。

[1] SOM を拡張して、マップ範囲をデータの位相構造に適合させて変形させる可変自己組織化法 (FSOM) を確立する。FSOM の基本アイデアはこれまでの研究で与えてきたが、従来算法は計算量が膨大で収束条件が不明という問題があった。そこで、実用算法の開発をめざした。

[2] 開発した算法などを基礎解析として、これを各種の大規模履歴データに適用することで、データの特性にあった状態遷移確率モデルの方法論を提案し、その有効性の実証を実データで検証した。特徴量の変化履歴データ(ア),(イ) では履歴状態についての、空間中の移動履歴データ(ウ) では移動部分系列についてのモデル化を行い、この上での遷移を確率モデルとして定式化し、その推定法を与えた。

4. 研究成果

[1] 可変自己組織化マップの実用算法

SOM は計算量がサンプル数 n に比例のオーダーであるため、多量サンプルデータにも適用可能である点が利点となる。しかし、従来法では、事前に指定された形状の空間に、強制的にサンプルを配分するため、必ずしも隣接したノードが類似した参照ベクトルを持つとは限らず、構成されたマップ上での隣接関係を利用して、新たな分析を行おうとする問題が生じていた。そこで、マップ範囲をデータの位相構造に適合させて変形させる可変自己組織化法(Flexible SOM [FSOM])を開発し論文として発表した。人工・実データを使った評価実験により、提案手法が従来の SOM よりも参照ベクトルの連続性と適合性に関する指標を向上させることを示した。これから実証研究をさらに進める中で改良が必要とは思われるが、FSOM の算法を確立することができた。

また、二つの自己組織化マップの間での対応をとる方法について検討を進め、共起行列の二つの相をそれぞれマップとするアルゴリズムを開発した。これを、モニタと商品の間の共起行列に対して適用した事例を、平成 25 年度データ解析コンペティションで提供されたモニタの購買履歴データに対し作成

し発表した。この方法は、疎な共起行列に対してその構造を発見するために有効と考えられる。今後、さらに研究を進める計画である。

[2] 各種データベースの整備とその解析

各種時空間履歴データについてデータベースの整備を行い、それらのデータを説明できるモデルを検討し、その解析を行った。時系列履歴からの状態遷移については、まず、クレジットカードの利用履歴データに対し、FSOMの適用事例を作成した。また、多変量の場合の傾向変化検出や単変量についての異常検出について手法の検討を行った。さらに、応用事例として、大規模購買履歴データから、価格分布の時間変化要因に注目し、いくつかのモデルを検討し発表した。また、ネットワークの時間発展について応用事例を検討した。

一方、状態遷移確率モデルとして多項ロジットモデルに注目し、地理情報システムを利用して準備した空間移動履歴での遷移について、その因果を説明しようとしたものである。関連して、大規模データによるlogitモデルの可能性について招待論文を発表し、応用事例を作成した。

以下に利用したデータ種別ごとに研究をまとめる。

(1) クレジットカード利用履歴データに対しFSOMによる推定を与え、顧客のクレジットカードの利用状態を類型化した(図1)。FSOMでは標準のSOMに比べてマップ連続性の高いマップを構成できたため、マップの位相関係を利用してマップ上での時間的遷移関係を分析する際に有利であることが明らかとなった(図2)。

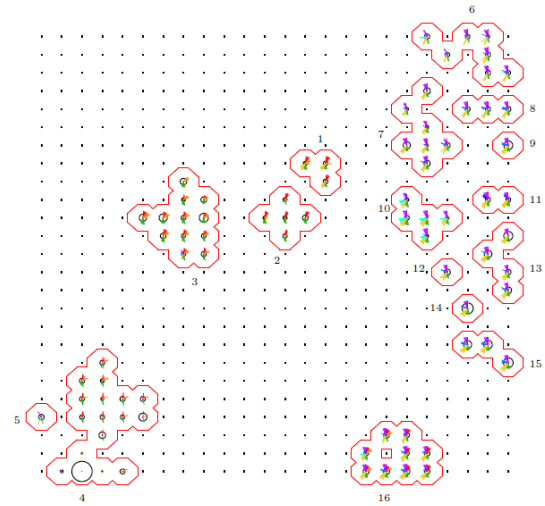
(2) 自治医科大との共同研究において全国規模の健康保険レセプトデータのデータベース化に協力し、このうち群馬県でのデータに対し分析を行った。具体的には患者の位置情報と受診履歴を利用して、地域医療計画における病床数計画などに有益な基礎情報を提供できると考えられる次の研究を行った。(1-1) 患者の受療選択の構造を明らかにするため、患者の医療機関選択モデルを提案した。地理情報システムを利用して、患者と選択候補病院との距離を求め、病院までの距離とデモグラフィックな要因との交互作用が患者の病院選択に与える影響を、多項ロジットモデルで分析する方法を与え、論文発表した。(1-2) 受療について二項分布を仮定し、地理的加重を用いた一般化線形モデルを用いて、デモグラフィックな補助変数を含む受療率モデルを推定する方法を与えた。

(3) 地方都市での移動行動データとして、JST採択課題「地域力による脱温暖化と未来の街-桐生の構築」と共同で、桐生市民からの無作為抽出アンケートを実施し2963名

の移動行動の記録を得た。これなどに基づき、以下の分析を行った。まず、桐生市民の利用交通手段と移動先の選択について多項ロジットモデルによるデータ解析を行い学会および論文発表をした。また、付随して、市民の店舗選択について、商店街側からデータ解析を行い発表した。さらに、付帯データを利用し、ロコミネットワークの時間発展についてグラフ理論を利用した検討を行い論文発表をした。

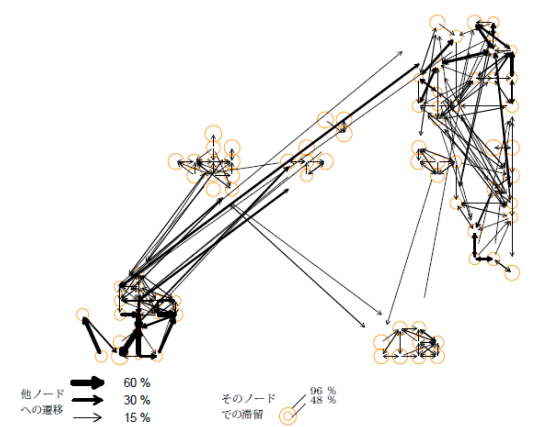
(4) 数年間に渡る全国の電子基準点データに対し、時系列データとしての地盤変動パターンの分析を進め、傾向変化検出法について検討を進めた。また、時系列データからのクラスタリングによる異常検出法を検討した結果を、論文発表した。

(5) 価格分布の検討事例として、平成23年度データ解析コンペティションで提供されたドラッグストアID-POSデータから価格弾力性のモデルを検討した。また、電子部品の長期の販売履歴から価格販売量時系列変化、パーコレーションモデルを利用したモデルを検討した。



格子のうち、丸で囲まれた部分がFSOMにより利用している格子を示す。

図1 抽出されたカード利用類型



矢印の太さは、各ノードの所属人数のうち、他のノードへ遷移した人数の割合を示す。上の凡例では代表的割合での太さを示す。円は、各ノードの所属人数のうち、そのノードに滞留した人数の割合を示す。

図2 カード利用類型間の顧客遷移

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 9 件)

- ① 金井康弘, 阿部圭司, 関庸一, 価格-販売量推移曲線のモデル化と活用法 —民生エレクトロニクス関連製品への適用, Transactions of the Operations Research Society of Japan, 査読有, Vol. 57, 掲載決定 (2014).
- ② Gao Wa, Yoichi Seki, Kenji Amagai, Takayuki Takarada, Analysis for Activation of Old Shopping Streets based on the Surveys of Citizens and Storekeepers, International Journal of Marketing Studies, 査読有, Vol. 6, No. 2, pp.92-101 (2014).
- ③ 仁藤慎也, 藍原雅一, 関庸一, 地域医療データバンクレセプトデータを活用した病院選択行動のロジット分析, 医療情報学, 査読有, Vol.33, No.5, pp. 243-251 (2014).
- ④ Yoichi Seki, Discrete Choice Model in the Big Data Era—Creating Abundant Datasets about Choice Situations—, Journal of Japan Industrial Management Association, 査読無(Invited Paper), Vol. 64, No. 2E, pp. 343-347 (2013).
- ⑤ Shin Ando, Theerasak Thanomphongphan, Yoichi Seki and Einoshin Suzuki, Ensemble anomaly detection from multi-resolution trajectory features, 査読有, Data Mining and Knowledge Discovery, 1 August (DOI 10.1007/s10618-013-0334-x) (2013).
- ⑥ Gao wa, Yoichi seki, Takayuki takarada, Relationship between Word-of-Mouth Network and Behavioral Tendency — Preliminary Research for Local Downtown Activation, International Journal of Marketing Studies, 査読有, Vol. 5, No. 5, pp. 1-11 (2013).
- ⑦ 関庸一, 亀倉大和, 一般化線形モデルによる価格弾力性推定に基づく粗利最大化, 日本経営工学会誌, 査読有, Vol. 63, No. 3, pp. 161-172 (2012).
- ⑧ Yoichi Seki, Luis C. Manrique, Kenji Amagai, Takayuki Takarada, Evaluation of Micro EV's Spreading to Local Community by Multinomial Logit Model, Industrial Engineering & Management Systems, 査読有, Vol. 11, No. 2, pp.147-153 (2012).
- ⑨ 多賀谷侑史, 安藤晋, 関庸一, サンプルの所属度に応じた可変自己組織化マップ, 情報処理学会論文誌. 数理モデル化と応用, 査読有, Vol. 5, No. 3, pp. 1-13 (2012).

[学会発表] (計 6 件)

- ① Yasuhiro Kanai, Keiji Abe, Yoichi Seki, An analysis of the fluctuation in electronic component prices using a price percolation model, The 14th Asia Pacific Industrial Engineering and Management Systems Conference, Dec 4-6, 2013, Cebu, Philippines.
- ② 折本拓真, 関庸一, 店舗業態評価と来店行動との多項ロジット分析に基づく商店街活性化策, 日本経営工学会平成 24 年度秋季研究大会予稿集, pp.94-95, 大阪工業大学, 2012 年 11 月 17 日.
- ③ 仁藤慎也, 藍原雅一, 関庸一, 地域医療データバンクのレセプトデータを活用した病院選択行動のロジット分析, 第 32 回医療情報学連合大会, 朱鷺メッセ新潟コンベンションセンター, 2012 年 11 月 16 日.
- ④ 小林貴子, 藍原雅一, 関庸一, 地理的加重一般化モデルを用いた疾病地図の推定, 2012 年度統計関連学会連合大会, 北海道大学高等教育推進機構, 2012 年 9 月 12 日.
- ⑤ Yoichi Seki, Luis C. Manrique, Kenji Amagai, Takayuki Takarada, Evaluation of Micro EV's Spreading to Local Community by Multinomial Logit Model, The 12th Asia Pacific Industrial Engineering and Management Systems Conference, October 14-16, 2011, Beijing, China.
- ⑥ 関庸一, 川連峻幸, 時系列データに対する自己相関補正累積 MT 法, 日本経営工学会平成 23 年度春季研究大会予稿集, pp.118-119, 愛知学院大学日進キャンパス, 2011 年 5 月 29 日.

[図書] (計 0 件)

[産業財産権]

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

[その他]

ホームページ等なし

6. 研究組織

(1) 研究代表者

関庸一 (SEKI YOICHI)
群馬大学・理工学研究院・教授
研究者番号: 90196949

(2) 研究分担者

なし

(3) 連携研究者

なし