

令和 5 年 6 月 12 日現在

機関番号：12102  
研究種目：若手研究  
研究期間：2020～2022  
課題番号：20K19747  
研究課題名（和文）Data analysis based on combinatorial structures  
  
研究課題名（英文）Data analysis based on combinatorial structures  
  
研究代表者  
鮎川 矩義（Sukegawa, Noriyoshi）  
  
筑波大学・システム情報系・助教  
  
研究者番号：20757710  
交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：本研究では組合せ構造を活用してデータ解析手法を高性能化する技術を開発した。具体的には、(i)社会科学分野における社会的選択理論の問題、非線形力学分野における(ii)カオス応答写像の実装で生じる周期の問題、(iii)マーク付き点過程データの中央値計算の問題を扱った。各問題に対して、当該分野で従来用いられてきたモデルを、そこに内在する組合せ構造を意識して刷新し、組合せ構造に特化したアルゴリズムや最適化の技術の効果的な応用を可能にした。特に、(iii)の問題に対する計算方法に関する研究では、地震の挙動を予測する問題への応用を行ない、本研究で開発した技術の有用性を示すことができた。

研究成果の学術的意義や社会的意義  
開発したデータ解析手法を適切に応用することで、新たな知見の発掘に寄与することが期待される。本研究で主な対象とした社会科学と非線形力学の分野はもとより、それらの周辺分野や、同様の解析手法への需要がある未開拓の分野への応用が考えられる。たとえば、本研究で対象とした地震の挙動を予測する問題は、非線形力学の問題でもありつつ、自然科学の問題でもあるといえる。本研究によって組合せ構造の活用によるデータ解析手法の高度化にはまだ大きな余地があることが明らかになった。これにより、組合せ構造に関する理論と応用の両面での研究がより活発になることが期待される。

研究成果の概要（英文）：In this research, we developed frameworks to improve the performance of data analysis methods using combinatorial structures. Specifically, we dealt with (i) a ranking problem in social choice theory, (ii) a parameter-tuning problem for a perturbation method used when implementing chaotic maps in nonlinear dynamics, and (iii) median calculation of a finite set of marked point process data. For each problem, the conventional model has been modified with an awareness of the combinatorial structure inherent in it, enabling effective application of algorithms and optimization techniques specialized for combinatorial structure. In particular, in the research on (iii), we applied it to the problem of predicting the behavior of earthquakes to show the usefulness of the technology developed in this research.

研究分野：数理最適化

キーワード：組合せ構造 データ解析 アルゴリズム 数理最適化 カオス理論 社会選択理論 厳密計算 整数論

## 1. 研究開始当初の背景

データサイエンスの応用場面で生じる問題の大規模化・複雑化に耐えうるアルゴリズムの高性能化が進んでいた。特に、グラフ・ネットワーク構造に代表される組合せ構造は頻出の構造であり、組合せ構造に関する最適化やデータ解析のための高性能なアルゴリズムを応用する動きがさまざまな分野で起こっていた。

## 2. 研究の目的

本研究では、組合せ構造に基づくデータサイエンスの手法の応用範囲の拡大を目的とした。モデリングとアルゴリズムのミスマッチに起因する不正確・不安定な解析からの脱却を促すことで、データサイエンスの手法のさらなる高性能化を達成し、より安心・安全な情報化社会の実現につなげることが最終的な目標である。

## 3. 研究の方法

本研究で対象とする具体的な問題をはじめに集めた。基本的な組合せ構造である、グラフ構造、順序構造、階層構造、ネットワーク構造を主なキーワードとして、文献調査、ならびに、専門家の方々との議論によって、

- (1) 非線形時系列解析におけるマーク付き点過程データの解析に現れる問題
- (2) 非線形時系列解析におけるカオス応答写像の実装の問題
- (3) 離散確率分布の推定問題
- (4) 選挙区割とシルバー民主主義の問題
- (5) 社会選択理論における順序付けの問題

を取り上げた。主にこれらの5つの問題に対する研究をほとんど同時進行的に進めた。具体的には、共同研究者との定例打合せ、関連する実データや人工データの解析、モデルとアルゴリズムの開発、結果の解釈とその現実の問題解決への還元可能性についての議論、を繰り返し行なった。特に、既存研究で既に組み立てられた具体的なモデルに対するアルゴリズムの開発を検討するのではなく、データと後工程で用いられるアルゴリズムやソフトウェアとの相性を考慮した、新しいモデルの開発に注力した。

## 4. 研究成果

本研究で得られた研究成果を上述の「3. 研究の方法」にある(1)～(5)のそれぞれに関するものについて順番に報告する。

(1) マーク付き点過程データとは、連続時間上で離散的に発生する事象を記述する時系列データであり、脳科学におけるニューロン発火、自然科学におけるゲリラ豪雨や地震活動など、その応用例は多岐に渡る。マークはひとつひとつの事象の発生の時刻の情報以外の付加情報である。たとえば、地震活動のデータでは、地震のマグニチュードや震源地の緯度経度の情報などが考えられる(図1)。

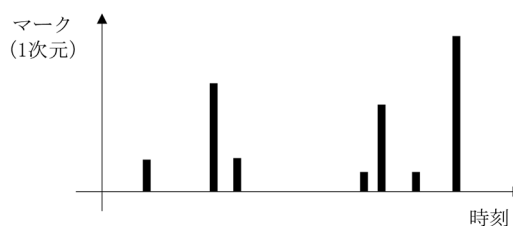


図1 マーク付き点過程データ

マーク付き点過程データの解析では、解析のための距離測度として、Victor と Purpura によって1997年に考案された編集距離(以降、単に、距離)がよく用いられる。マークがない場合の距離計算は動的計画法によって効率的に行なえることが知られていたが、マークがある場合の距離計算については最近まで最善策が見つかっていなかった。これに対し、申請者らの研究グループは、2019年頃、グラフ理論の最小重み完全マッチングを用いたモデルが有効であることを実験的に示した。そこで本研究では、マーク付き点過程データの解析に現れる手法の、グラフ・ネットワーク構造に関するアルゴリズムの応用による、さらなる高性能化を図った。そこで、中央値計算に着目した。ここでの中央値は与えられたデータ点との距離の総和を最小にする点と定義される。従来の中央値計算のための手法は精度に保証のないヒューリスティックスであった。これに対し、本研究では、距離計算の研究での知見を活かし、中央値計算に内在

する組合せ構造、具体的には、2部グラフの構造を指摘し、グラフ理論の言葉で中央値計算を記述した。しかし、距離計算の場合と異なり、同構造に対しては、(申請者の研究グループで検討した限りでは)適用可能なグラフ・ネットワーク理論のアルゴリズムがなかったため、最適化ソルバー(数値最適化の計算を実行する高性能ソフトウェア)の活用を提案した。100程度の点過程データの組(含まれる点の数は合計で1,000程度)の中央値であれば数分程度で厳密な中央値を計算できる。

同手法を地震活動の予測問題に応用した。この文脈での従来手法は、大雑把に言うと、現在の地震発生パターンと類似する過去の地震発生パターンを参照して未来の予測を行なうものであった。これに対し、本研究で確立した中央値計算の方法を用いて、参照するデータ数を増やし予測性能を向上する方法を提案した。図2に、東日本大震災付近での地震活動のデータに対し、提案手法によって予測性能が改善された割合(テストデータにおける手法による予測パターンと現実のパターンとの乖離を距離で表し、距離が削減された割合)を示す。パラメータ設定(A・B・C)を変え、参照するパターン数(括弧内の数字)を変えながら実験を行なった。どの設定でも40%~50%の改善に成功している。

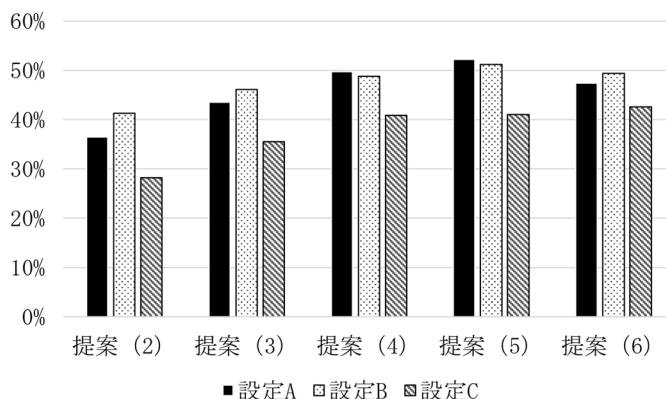


図2 従来手法に対する提案手法の改善率：地震活動の予測

(2) 上述の(1)で述べたマーク付き点過程データはカオス理論の主要なツールであり、同じくカオス理論で主要な役割を果たすツールとして、カオス応答写像と呼ばれるものがあり、現象の理解やアルゴリズムの解析・開発に役立てられている。カオス応答写像の最も基本的なものにベルヌーイシフト写像と呼ばれる写像がある。0以上1未満の実数の入力に対して0以上1未満の実数の無限列を返す写像であり、応用場面では同無限列から自然に定義される周期が長くなるのが望ましい。理論上、カオス応答写像は無理数の入力に対して無限の長さの周期を持つが、通常のコンピュータ上のシミュレーションでは、2進数表現を用いる必要があり、したがって、表現できる入力は有理数であり、無限の周期は得られないという問題があった。本研究では、まず、2進数表現の場合、ビット数の指数オーダーの周期は期待できるが、素朴な実装では線形オーダーの周期しか実現できないことを示した。つぎに、この問題の解決に用いられる摂動法に着目した。摂動法はしばしば指数オーダーに近い周期を実現するものの、パラメータの設定方法の確立やその理論保証が未解決であった。本研究では、良いパラメータというものを形式的に定義し、その性質を整数論の原始根の言葉で言い換え、整数論の技術を応用することでパラメータ設定に対するアルゴリズムを提案した。一方で、所与のビット数に対して良いパラメータが存在する可能性について何らかの理論保証を与えることは、原始根に関する Artin 予想の特殊ケースと密接に関係しており、簡単ではないことを指摘した。

(3) 離散確率分布の推定は統計学の基本的な問題である。いわゆるビッグデータ時代の到来によって、データの度数分布をそのまま母集団分布とみなせる場面も多くなりつつあったが、ユーザの事前知識や直感に符合しない場面も少なくなかった。そこでユーザの事前知識や直感に符合するように、分布を最適化手法によって修正する方法がさまざまな場面で活用されつつあった。本研究では、分布に現れる峰、つまり、データを一列に並べたときの前後の大小関係が表す順序構造を活用した手法を提案し、その有用性を育児関連データの解析で生じる問題に応用し、従来研究やベースラインモデルでは検出しにくい知見が得られることを確認した。また、eコマースの文脈において、データ点が整数の有限列(例:1 3 2や2 2 4)と関連付けられる場合に対し、ある応用場面でのユーザの事前知識や直感が、可能な有限列の集合の上の半順序集合になることを指摘し、そのハッセ図を活用して、消費者の行動パターンを効率良く予測するモデルを確立し、従来モデルに比べて良い性能を持つことを実験的に示した。

(4) 衆議院議員選挙区区割りには、一般に、人口データを用いて一票の格差の観点から公平に決められる。本研究では、区割りに生活圏を反映させるために、生活圏を捉えるためのデータとしてモバイル空間統計による人流の推定値を用いる枠組を提案した。この枠組では、人流を表すネットワークを構築し、その上で、一票の格差の是正と生活圏と行政圏の調和というふたつを目的として区割りを最適化するモデルを考案し、茨城県・栃木県・群馬県を対象としてその有用性を実験的に示した。また、シルバー民主主義の問題を解決するべく、世代を考慮した区割りを扱うために、前述したネットワークの多層化モデルを提案した。各選挙区での有権者の年齢の中央値が高かった上述の3県に対し、提案モデルによって、中央値が高い選挙区と低い選挙区にすみ分けがなされ、若年層の意見が反映しやすい区割りが行なえる可能性を示した。

(5) 複数の審査員が複数の評価対象へのランキングを表明している場面で、審査員からの不満の総和がなるべく小さいランキングを求める問題は、ランキング集約問題と呼ばれ、社会的選択理論における古典的な課題となっている。同問題は、評価対象を頂点、評価対象間の優劣あるいは優劣の度合を有向枝あるいは重み付き有向枝として表現することでネットワーク上の問題として記述できる。同問題においては、公平性の観点から、不満の総和が最小のランキングをひとつ求めるだけでなく、複数あればそれらすべてを求めたいという需要がある。本研究では、同問題の最適解の集合が持つ構造を、問題に付随する上述のネットワークの解析を通して、行ない、既存研究よりも、より詳細な条件の導出に成功した。

## 5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Higashikawa Yuya, Imai Keiko, Shiraga Takeharu, Sukegawa Noriyoshi, Yokosuka Yusuke	4. 巻 36
2. 論文標題 Minimum point-overlap labelling*	5. 発行年 2020年
3. 雑誌名 Optimization Methods and Software	6. 最初と最後の頁 316 ~ 325
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10556788.2020.1833880	査読の有無 無
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 Ando Kazutoshi, Sukegawa Noriyoshi, Takagi Shota	4. 巻 65
2. 論文標題 STRONG CONDORCET CRITERION FOR THE LINEAR ORDERING PROBLEM	5. 発行年 2022年
3. 雑誌名 Journal of the Operations Research Society of Japan	6. 最初と最後の頁 67 ~ 75
掲載論文のDOI (デジタルオブジェクト識別子) 10.15807/jorsj.65.67	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Sukegawa Noriyoshi, Ikeguchi Tohru	4. 巻 165
2. 論文標題 How to perturb Bernoulli shift map	5. 発行年 2022年
3. 雑誌名 Chaos, Solitons & Fractals	6. 最初と最後の頁 112793 ~ 112793
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.chaos.2022.112793	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 Iwanaga Jiro, Nishimura Naoki, Sukegawa Noriyoshi, Takano Yuichi	4. 巻 37
2. 論文標題 Analysis of Information Search around the Time of Childbirth: Estimating Probability Distributions of Search Dates via Mathematical Optimization	5. 発行年 2022年
3. 雑誌名 Transactions of the Japanese Society for Artificial Intelligence	6. 最初と最後の頁 D ~ L74_1-11
掲載論文のDOI (デジタルオブジェクト識別子) 10.1527/tjsai.37-3_D-L74	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計13件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 鈴木祥平、鮎川矩義、池辺淑子
2. 発表標題 点過程のメディアン計算とその地震学への応用
3. 学会等名 日本オペレーションズ・リサーチ学会春季発表会
4. 発表年 2021年

1. 発表者名 丸山 怜、鮎川 矩義、池辺 淑子、立川 智章、山本 啓二、庄司 文由
2. 発表標題 スーパーコンピュータのジョブスケジューリング
3. 学会等名 日本オペレーションズ・リサーチ学会春季発表会
4. 発表年 2021年

1. 発表者名 相原 麟太郎、鮎川 矩義、池辺 淑子、石井 康貴、廣田 正之
2. 発表標題 数理最適化によるテストの配点決定
3. 学会等名 日本オペレーションズ・リサーチ学会春季研究発表会
4. 発表年 2021年

1. 発表者名 鮎川 矩義
2. 発表標題 整凸多面体の組合せ的直径とその周辺
3. 学会等名 離散数学とその応用研究集会（JCCA2021）
4. 発表年 2021年

1. 発表者名 廣田 正之、石井 康貴、池辺 淑子、鮎川 矩義
2. 発表標題 テスト・問題集編成における数理最適化技術の応用
3. 学会等名 日本オペレーションズ・リサーチ学会春季研究発表会
4. 発表年 2022年

1. 発表者名 荻野 奨太、鮎川 矩義、池辺 淑子
2. 発表標題 船の到着時刻の不確実性を考慮したコンテナ事前整列問題
3. 学会等名 「最適化：モデリングとアルゴリズム」政策研究大学院大学研究集会
4. 発表年 2022年

1. 発表者名 荻野 奨太、鮎川 矩義、池辺 淑子
2. 発表標題 船の到着時刻の不確実性を考慮したコンテナ整列問題
3. 学会等名 「組合せ遷移の展開に向けた計算機科学・工学・数学によるアプローチの融合」 学生シンポジウム
4. 発表年 2022年

1. 発表者名 松原千波、鮎川矩義、大澤義明
2. 発表標題 生活圏と世代を意識した選挙区割りの最適化
3. 学会等名 2022年度第36回ARSC研究発表大会
4. 発表年 2022年

1. 発表者名 伊熊大貴、鮎川矩義、高野祐一
2. 発表標題 船舶到着時刻の不確実性を考慮したコンテナ事前整列問題
3. 学会等名 最適化の理論とアルゴリズム：未来を担う若手研究者の集い12023
4. 発表年 2023年

1. 発表者名 柳智也、鮎川矩義、高野祐一
2. 発表標題 データコラボレーション解析を用いた推薦システム
3. 学会等名 最適化の理論とアルゴリズム：未来を担う若手研究者の集い12023
4. 発表年 2023年

1. 発表者名 石井康貴、廣田正之、池辺淑子、鮎川矩義、西田優樹
2. 発表標題 生徒個別最適な問題セットの編成
3. 学会等名 日本オペレーションズ・リサーチ学会2023年春季研究発表会
4. 発表年 2023年

1. 発表者名 松原千波、鮎川矩義、大澤義明
2. 発表標題 生活圏と世代を意識した選挙区割りの最適化
3. 学会等名 日本オペレーションズ・リサーチ学会2023年春季研究発表会
4. 発表年 2023年



1. 発表者名 Noriyoshi Sukegawa, Shohei Suzuki, Yoshiko Ikebe, Yoshito Hirata
2. 発表標題 Optimization approaches in analyzing marked point process data
3. 学会等名 10th International Congress on Industrial and Applied Mathematics (国際学会)
4. 発表年 2023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