

令和 6 年 4 月 9 日現在

機関番号：17102

研究種目：挑戦的研究（萌芽）

研究期間：2020～2023

課題番号：20K20884

研究課題名（和文）行列式確率場と機械学習

研究課題名（英文）Determinantal point fields and machine learning

研究代表者

白井 朋之（Shirai, Tomoyuki）

九州大学・マス・フォア・インダストリ研究所・教授

研究者番号：70302932

交付決定額（研究期間全体）：（直接経費） 4,400,000円

研究成果の概要（和文）：行列式確率場の機械学習の理解を深めるために以下のような研究を実施した。完全2部グラフ上のサイクルの個数を固定したときの連結成分の数え上げの問題、立方格子複体上のラプラシアン固有値・固有ベクトルの決定とその応用として全域非輪体の重み付きの数え上げと指数母関数の族の満たす偏微分方程式系の導出、行列式確率場を正定値行列と歪対称行列を含めたL-行列を用いてモデル化した場合のモデルの表現力の比較、パーシステント図間の距離および点配置の最近接間距離のヒストグラムに関する距離を用いた摂動格子によるギニブル確率場の近似。これらの研究はこれまでに国際的な学術誌に掲載されるとともに、現在投稿中である。

研究成果の学術的意義や社会的意義

行列式確率場は本来純粋数学的興味から研究が始まった研究対象であるが、機械学習への応用が提案されて以来、応用分野でも関心をもたれてきた。本研究は行列式確率場を具体的な問題に応用するという観点から実施し、特にハッシング、セルラーネットワーク、機械学習などの具体例に現れる行列式確率場に関わる数学的問題をとりあげた。これらの研究成果は、数学としての確率論および組合せ論の一分野に貢献するとともに、実世界の問題において応用される可能性も期待される。

研究成果の概要（英文）：To deepen understanding of determinantal probability in machine learning, the following research was conducted: enumeration of connected components when the number of cycles on a complete bipartite graph is fixed, determination of eigenvalues/eigenvectors of Laplacians on cubical complexes and their application in counting weighted spanning acycles and the system of partial differential equations satisfied by a family of exponential generating functions, comparison of representability of statistical models when modeling determinantal probability using L-matrices encompassing positive definite matrices and skew-symmetric matrices, approximation of Ginibre random point fields via perturbed lattices using distances based on persistence diagram and histograms of nearest-neighbors in point configurations. These studies have been published in international academic journals and are currently under submission.

研究分野：確率論

キーワード：行列式確率場 立方格子複体 パーシステントホモロジー 全域木・連結成分

## 1. 研究開始当初の背景

行列式確率場は物理のフェルミオン系やランダム行列の固有値を記述する重要な確率場として、1990年後半から確率論の分野では中心的な研究課題の一つであり続けている。2012年Kulesza-Taskerらによって行列式確率場が機械学習の確率モデルとして応用されて以来、工学や情報の分野でも応用的視点から行列式確率場が注目を集めている。

## 2. 研究の目的

物理・数学のみの観点からは見えてこなかった行列式確率場の問題について、機械学習やアルゴリズム論の視点から実践的・理論的研究を推進するとともに、深層学習(ディープラーニング)などでブラックボックス化されている部分を数学的に解析することを長期的目標と見据え、関連する基礎的な研究を推進することを目指す。また、行列式確率場を含む確率場の理論と機械学習で培われてきた理論と技術の融合により、双方の分野における新たな研究方向の開拓と具体的問題の提供に寄与することも目標とする。双方向的な研究の推進は、数理学における確率場の研究とともに、工学における機械学習ひいては人工知能の研究において大きな意義をもつと考える。本研究では、具体的かつ意味のある確率モデルに対して、上述のブラックボックスの中身を数学的に解析することを長期的な目標とし、種々の確率場に関する機械学習の実践と理論的研究を並行して行うことにより、その長期的目標の基礎を確立することを目指す。本研究を通して、これまでとは異なる視点から、既存の確率モデル(特に確率場)の数学的考察を行うとともに、新たな数学的問題の発掘、確率モデルの構成などに寄与することも目標の一つとする。

## 3. 研究の方法

本研究では、研究目的に述べた解析を進めるために、確率モデルとして行列式確率場を取り上げる。行列式確率場の概念は、1975年頃マッキが荷電粒子などを記述するフェルミオン多粒子系の統計的性質を記述するために考察したのが始まりで、2000年頃申請者は高橋陽一郎(当時京大数理研)との共同研究で、その概念を一般の抽象空間上の確率場として拡張して基礎理論を整備し、その適用範囲を大いに広げた。本研究では、連続行列式確率場の例として、特に種々の計算が容易なギニブル確率場を取り上げ、独立なポアソン確率場で計算されてきた種々の観測量の比較のために用いる。また、離散空間上の行列式確率場は、離散集合のべき集合の上の確率測度であり、非負定値行列をパラメータとする確率分布族として確率モデルを設定する。特に、離散空間としてグラフの辺集合やもっと一般に複体内の単体の集合ととり、その上で定義される行列式確率場の性質を調べる。

## 4. 研究成果

- (1) 生成モデルは科学・工学の現実の問題において重要な役割を果たし、特徴量(確率変数)の条件付き独立性を基礎にしたグラフィカルモデルがその文脈でしばしば現れる。生成モデルの学習においては、考えている現象における事前知識の導入が計算コストの削減において不可欠となる。グラフィカルモデルを構築する際には、すべてのデータ構造が

明らかになる必要があるが、一般的な問題ではそのような状況はまれで、部分的な構造のみが既知という状況で問題を考える必要がある。これまで、機械学習においてこの視点を入れた研究はそれなりに存在するが、生成モデルにおいてはあまり研究されていなかった。分担者河原は、生成モデルに対して部分的な事前知識の情報をいかにして導入するかについて考察し、分布間のヒルベルトシュミット独立性規準 (HSIC) による正則化の方法を提案した。

- (2)  $d$ 次元整数格子をベースとする立方格子複体上のアップラプリアンの固有値を完全に決定して、各次元の固有値および固有関数の間にある種の入れ子構造があることを示した。単体複体およびセル複体に拡張された行列木定理を用いることにより、立方格子複体における全域非輪体のねじれ群の位数の二乗の重みを付けた個数の標識を与えて、その極限を考えることにより、立方格子複体のねじれ重み付き全域非輪体エントロピーが、対数的マラー測度の線形結合として表されることを示した。また、全域非輪体全体上に上記の重み付きによる確率測度を定義すると行列式確率場になり、その観点からも新たな研究の方向性があることを報告した。この結果は、平岡との共同研究で出版された。
- (3) 完全2部グラフ上のサイクルが複数ある場合の連結成分の数え上げの問題はカッコーハッシングの関連で応用場も重要で、特にユニサイクルと呼ばれる場合には考察されていた。藪奥・蓮井らとの共同研究で、ベッチ数を固定した場合の連結成分の総数が、ベッチ数をパラメータとする漸化式を満たし、その指数型母関数を用いる方法で偏微分方程式系を導き、その方程式を解くことによりその係数の漸近挙動を求めた論文を投稿中である。
- (4) Ghosh・三好との共同研究では、三角格子の各点を独立にガウス分布に従って摂動することによって得られる摂動格子で、ギニブル確率場を近似することを考えた。各ガウス分布の分散のパラメータを0から無限大に動かしていくと三角格子からポアソン点過程までを補間することができるが、途中適当な分散の値でギニブル確率場に近づくことを、点配置の間の2種類の距離を用いて測り、ほぼ同じ分散の値においてギニブル場に最も近づくことを観察した。それをワイヤレスネットワークのSINRの問題に応用した論文が出版された。摂動格子は数値実験にも適しており、これを適当に修正することにより、離散空間上の行列式確率場をサンプルするための一つの方法とすることは次の問題として考えうる。
- (5) 離散空間および連続空間の行列式確率場は、その確率場を特徴つけるパラメータとして、縮小かつ正定値行列(正定値カーネル)が主に使われていた。Kulesza-Taskerらによって始められた行列式確率場を基礎とした機械学習では、その行列を種々の手法で学習するが、そのパラメータの空間をもう少し広くいわゆる $P$ 行列(もしくは $P_0$ 行列)とよばれる、すべての主小行列式が正(もしくは非負)である行列の族に広げるのは自然であり、Gartrell et al. ではその点について歪対称行列を含む形で考察していた。行列式確率場を正定値行列と歪対称行列をそれぞれ $L$ -行列としてモデル化して、その階数などモデルのパラメータの自由度を同じにした際に、確率測度の表現力を数値実験のレベルではあるが比較した。

## 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 1件/うちオープンアクセス 2件）

1. 著者名 Hiraoka Yasuaki, Shirai Tomoyuki	4. 巻 0
2. 論文標題 Torsion-weighted spanning acycle entropy in cubical lattices and Mahler measures	5. 発行年 2024年
3. 雑誌名 Journal of Applied and Computational Topology	6. 最初と最後の頁 1-31
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s41468-024-00163-y	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 Makoto Katori and Tomoyuki Shirai	4. 巻 98
2. 論文標題 Local universality of determinantal point processes on Riemannian manifolds	5. 発行年 2022年
3. 雑誌名 Proceedings of Japan Academy	6. 最初と最後の頁 95-100
掲載論文のDOI（デジタルオブジェクト識別子） 10.3792/pjaa.98.018	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 Subhroshekhar Ghosh, Naoto Miyoshi and Tomoyuki Shirai	4. 巻 -
2. 論文標題 Disordered complex networks : energy optimal lattices and persistent homology	5. 発行年 2022年
3. 雑誌名 IEEE Transactions on Information Theory	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TIT.2022.3163604	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 該当する

1. 著者名 河原吉伸	4. 巻 IJCAI-20
2. 論文標題 Knowledge-Based Regularization in Generative Modeling	5. 発行年 2021年
3. 雑誌名 Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence	6. 最初と最後の頁 2390-2396
掲載論文のDOI（デジタルオブジェクト識別子） 10.24963/ijcai.2020/331	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2件（うち招待講演 2件 / うち国際学会 1件）

1. 発表者名 白井朋之
2. 発表標題 行列式点過程の数理とその応用 --- ランダム行列から機械学習へ
3. 学会等名 AIE講演会（東北大）（招待講演）
4. 発表年 2022年

1. 発表者名 Tomoyuki Shirai
2. 発表標題 Persistent homology and its applications
3. 学会等名 Blockchain Kaigi 2022 (BCK2022)（招待講演）（国際学会）
4. 発表年 2022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
研究分担者	河原 吉伸  (Kawahara Yoshinobu)  (00514796)	大阪大学・大学院情報科学研究科・教授    (14401)	

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		
シンガポール	National University of Singapore		