

氏名 川島 貴大

学位(専攻分野) 博士(統計科学)

学位記番号 総研大甲第 2492 号

学位授与の日付 2024 年 3 月 22 日

学位授与の要件 複合科学研究科 統計科学専攻
学位規則第6条第1項該当

学位論文題目 Probabilistic Models Characterized by a Kernel Matrix and
Their Learning Methods

論文審査委員 主査 鎌谷 研吾
統計科学コース 教授
持橋 大地
統計科学コース 准教授
日野 英逸
統計科学コース 教授
武石 直也
東京大学 大学院工学系研究科 講師

博士論文の要旨

氏 名：川島 貴大

論文題目：Probabilistic Models Characterized by a Kernel Matrix and Their Learning Methods

Positive definite kernels play a significant role in modern machine learning. Kernel methods opened up possibilities for analyzing complex data that may be governed by nonlinear structure because of the rich representational power and nice theoretical properties. Intuitively a positive definite kernel realizes nonlinear data processing on the input space in which data points are located by determining a metric on a possibly infinite-dimensional space. Related to the kernel methods, probabilistic models characterized by a positive definite kernel have also attracted attention. Gaussian processes (GPs) and determinantal point processes (DPPs) representatively fall into these models. These models offer choices of methodologies for dealing with complex data as well as the kernel methods. This dissertation addresses the following two topics about probabilistic models with a positive definite kernel:

- i. proposing a GP-based generative model for multivariate time-series data via a physics approach and developing an efficient inference method for the model
- ii. developing a simple and fast learning method for DPPs

For (i), we propose a nonlinear and probabilistic generative model of Koopman mode decomposition (KMD) based on the framework of unsupervised GPs. Differential equations appear in many fields of science, including materials science, geophysics, epidemiology, and social informatics. In these fields, multivariate time-series data governed by an unknown differential equation is sometimes obtained, and we may want to know about the underlying dynamics. One of the factors that make the estimation problem of the underlying dynamics difficult is nonlinearity. Through KMD, nonlinear dynamics on a finite-dimensional space is lifted into an infinite-dimensional space in which the dynamics behaves linearly. That leads concrete algorithms to find the modes characterizing the dynamics, such as dynamic mode decomposition (DMD). While DMD and other related algorithms have been successful in many fields, resulting values yielded by the algorithms are sometimes not very easy-to-interpret. On the other hand, our model makes it possible to estimate the physical quantities associated with the Koopman modes and the (low-dimensional) latent variables simultaneously by taking an approach of generative modeling. Our model is the first to give a way to estimate KMD latent variables, and we show the usefulness through some numerical

experiments with both of synthetic and real-world datasets. Moreover, we develop a scheme that reduces the computational complexity to learn our model for scalability.

For (ii), we develop a fast, stable, and simple learning rule for DPPs on the basis of MM (minorization-maximization) algorithms, which increases the objective values monotonically. DPPs are powerful probabilistic models that generate random subsets with diverse items from a ground set. For example, let us consider a recommender system on an e-commerce site in which a variety of home appliances are handled. Then, the purchasing histories can be regarded as samples of a DPP on the finite ground set, which consists of all the products handled by the site. Now, we may want to assume that “rarely do consumers buy more than one refrigerator at a time,” or more conceptually: “a random subset tends not to have similar items simultaneously.” Such a concept is called negative dependence, and DPPs take it into account. Since the similarities between items are parameterized as a kernel matrix in DPPs on a finite set, the fitting problem of DPPs becomes a problem of estimating a positive definite matrix. Although some existing studies have addressed the problem, there is room for improving the stability and speed of convergence. In this work, we show that the learning problem of DPPs can be resulted in iterative solving of a continuous algebraic Riccati equation (CARE), which is a solvable class of quadratic matrix equations. The monotonicity of our algorithm follows the property of MM algorithms. We also develop an acceleration technique for our algorithm by introducing a step size parameter whose value can be determined adaptively in each iteration. We numerically compare our algorithm and existing methods with synthetic and real-world data in experiments. Our algorithm outperforms existing methods in convergence speed for most of the datasets, and we additionally discuss what contributes the efficiency of our algorithm.

This dissertation is organized as follows. In Chapter 1, we give an introduction motivating us to study probabilistic models parameterized by a kernel matrix. In Chapter 2, we present technical preliminaries related to kernel methods, GPs, and DPPs. In Chapter 3, we develop GPKMD based on Bayesian DMD. We show that Bayesian DMD can be extended to a GP-based probabilistic generative model naturally. We also propose a computational scheme to improve the time complexity of learning GPKMD. Experimental results find that GPKMD can capture important dynamics from observed data. In Chapter 4, we study an efficient and simple rule to learn full-rank DPPs. We prove that maximum likelihood estimation of a DPP can be reduced to iterative solving of some matrix quadratic equation by using MM algorithm. We also develop an accelerated version of the algorithm which is no longer monotone increasing but possibly converges faster. Numerical results on both synthetic and real-world datasets show our algorithm outperforms existing methods. Finally, we give concluding remarks in Chapter 5.

博士論文審査結果

Name in Full
氏名 川島 貴大

Title
論文題目 Probabilistic Models Characterized by a Kernel Matrix and Their Learning Methods

2024年2月1日午前10時から約2時間にわたり川島貴大氏の博士論文審査委員会を開催した。出願者による約1時間の公開発表による概要説明と質疑応答、さらに約1時間の審査委員のみによる審査を行った結果、審査委員会は本論文が学位の授与に値すると判断した。

[論文の概要]

提出された論文は5章111ページからなり、英語で書かれている。正定値カーネルを用いたカーネル法はその豊かな表現力と理論的特性により、機械学習において重要な役割を果たしている。カーネル法に関連する確率モデルとして、ガウス過程および行列式点過程があり、多様なデータとその生成プロセスをモデル化する手法として広く利用されている。本論文は、多変量時系列データのためのガウス過程ベースの生成モデルとその推論方法、及び、行列式点過程のための高速な学習方法を論じるものである。

第1章では、はじめに機械学習を実問題に適用する際に現れる複雑なデータを例示し、それらの取り扱いにおいてデータ同士の距離あるいは類似度の概念が重要であることと、カーネル法を用いた類似性の表現の重要性と柔軟性を説明している。その上で、カーネル法に関連するモデルとしてガウス過程と行列式点過程を挙げ、それらの有効性を説明することで、本論文で扱う問題の重要性と動機を述べた上で、本論文の貢献と研究全体の概要を示している。

第2章は線形回帰にベイズ推論の視点を導入する形で、リッジ回帰を通して正定値カーネルを用いた予測モデルを導入している。正定値カーネルと再生核ヒルベルトの性質を簡潔に紹介し、ガウス過程回帰モデルと行列式点過程モデルの定式化と基本的な性質を記述している。

第3章では、多次元時系列データを複数のモードに分解することで系のダイナミクスを調べる方法として近年注目を集めているKoopmanモード分解を説明し、既存手法をより柔軟に拡張したガウス過程Koopmanモード分解を定式化している。Koopmanモード分解は、状態空間モデルのように、時間発展する潜在変数から観測変数がデータとして得られるという仮定をしている。Koopmanモード分解は生成モデルとして定式化されたものではなかったが、これを教師なしガウス過程モデル帰着することで、潜在変数の直接的な推論が可能になる。このベイズ的なKoopmanモード分解のモデリングにより、欠損データへの対応や、実データに現れるスパイク状の特異な振る舞いを捉えることができることを実験的に確認している。

行列式点過程は, 有限集合からその部分集合が生起する確率をモデル化したものであり, 集合要素同士のカーネル関数で定まるグラム行列の行列式でその生起確率が表される. 実際に観測された要素の共起データから行列式点過程を特徴づけるグラム行列を推定するためには, 計算量削減のため低ランク性や特殊な構造を仮定することが多い一方で, 小中規模の問題に対しては余分な制約を与えずにグラム行列を推定することが望ましい. そこで第4章では行列式点過程の学習方法として, 従来の EM アルゴリズムや不動点アルゴリズムと比べて安定性が高く, 高速かつ実装が容易なアルゴリズムを MM アルゴリズムに基づき開発し, 理論的な性質の解析及び実データによる評価を行っている. さらに, パラメタの更新ステップを動的かつ自動的に定める加速アルゴリズムを開発し, 提案手法が高速かつ安定的に収束することを示している.

第5章では, 本論文の寄与であるガウス過程 Koopman モード分解手法と, MM アルゴリズムに基づく行列式点過程学習アルゴリズムの利点と限界を具体的に述べている. ガウス過程 Koopman モード分解手法の問題点として, モードを表現する固有関数がカーネル関数値として隠れてしまっていることを指摘しており, 固有関数から情報を得る方法を課題としている. 行列式点過程学習アルゴリズムに関しては, 提案手法を支持する理論的解析の必要性と, 大規模データへの応用に不可欠な低ランク近似法の開発といった展望を述べている.

[論文の評価]

データ間の関係性を表現するツールとしてカーネル関数は盛んに研究されており, ガウス過程回帰の応用は非常に多岐にわたる. 高次元時系列データの解析手法として動的モード分解を始めとする Koopman モード分解は近年注目を集めているが, そのベイズ的な定式化の研究は少なく, 川島氏が提案するガウス過程 Koopman モード分解は, Koopman モード分解あるいはより一般に作用素論的データ解析にベイズ推論の視点を成功裏に導入した例といえる. また, 行列式点過程は物理学に由来し, 古くから知られている確率モデルであり, 近年では推薦システムや情報検索への応用を通して機械学習分野でも注目を集めているが, 近似によらない効率的な学習アルゴリズムは僅かしか知られていなかった. 川島氏が開発したアルゴリズムは安定かつ高速で, 実装が容易な方法であり, そのシンプルさから, 行列式点過程の学習アルゴリズムのさらなる開発の嚆矢となりうるものである. 以上の理由により, 本論文は統計科学の博士論文として十分な意義を持つと判断される.

なお, 第3章の内容は, 査読付き国際論文誌 *Neural Computation*(vol 35(1), 82-103, 2023)に, 第4章の内容は, 査読付き国際論文誌 *Transactions on Machine Learning Research* (<https://openreview.net/forum?id=65AzNvY73Q>, 2023)に掲載されている.