

平成 29 年 6 月 9 日現在

機関番号：12608

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26870185

研究課題名(和文)スピングラス理論に基づいた学習・推定・逆問題の総合的研究

研究課題名(英文)A comprehensive study of learning, inference, and inverse problems based on the spin-glass theory

研究代表者

小淵 智之(Obuchi, Tomoyuki)

東京工業大学・情報理工学院・助教

研究者番号：40588448

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究は逆問題の理論的諸課題に、スピングラス理論に基づいて取り組んだものである。逆問題とは正解の確率分布を限られた観測から当てる問題であり、正解に原理的にどこまで近づけるかという理論限界・アルゴリズム開発・現実のデータ解析への応用、など、様々な側面からの研究がある。研究は2つのフェイズに分かれた。前半では理論限界を求めるための解析に注力した。後半では、アルゴリズム開発とニューロン・自然画像などの実データを解析する研究を行った。結果、近年データ解析によく使われる手法である最大エントロピー法やスパースモデリングの理論限界が明らかとなり、他の手法と比較してどのような利点・欠点があるかが明確となった。

研究成果の概要(英文)：The purpose of this study is to clarify general and theoretical aspects of statistical learning theory and inverse problems by using the spin-glass theory from statistical physics. The objective of the discipline is to infer a correct probability distribution from a limited number of observations. Hence common important theoretical questions are as follows: Clarifying the achievable theoretic limit given the limited observations; designing algorithms to achieve the limit; applying the algorithms to real-world dataset. Our actual research processes were roughly categorized into two processes: One is to clarify the theoretical limit by utilizing the spin-glass theory; the other is to invent algorithms and apply them to neurons' firing data and natural image processing. Our study has clarified that two commonly used frameworks, the maximum entropy principle and sparse modelling, have their own theoretical limits and have contrasting pros and cons.

研究分野：統計物理学

キーワード：逆問題・多自由度推定 統計的学習理論・機械学習 統計物理学 最大エントロピー法 スパースモデリング

### 1. 研究開始当初の背景

(1) 機械学習の発展により、学習理論や逆問題の理論的・実際の課題が、多く蓄積されるようになってきている。本研究開始当初では、その理論的課題のうち、特に最大エントロピー法と呼ばれる、分野にまたがって使用されていた枠組みの理論的理解が急務であったため、この課題を設定した。

(2) 機械学習はその理論・技術の発展が速く、最大エントロピー法に代わる枠組み・技術の候補もいくつもあった。そのため学習理論・逆問題の系統的研究方法を追求することが重要であったためこの課題を設定した。

### 2. 研究の目的

(1) 最大エントロピー法の枠組みの理論的理解を得ること、及びこの枠組みが他の方法論と比較してどのような利点・欠点があるのかを明らかにすること。

(2) 最大エントロピー法ではうまくいかない例における代替枠組みを見出し、その理論的理解を得ること。

(3) 以上の理解を元に実用的なアルゴリズムを開発し、生命系のデータ解析に応用すること。

### 3. 研究の方法

申請者はこれまで統計物理学の理論研究に注力してきた。以下の研究方法には、この経験が強く生かされている：

(1) 最大エントロピー法及び代替枠組みの理論解析のため、統計物理学で使われる平均場近似を用いて、その理論性能に関する理論的理解・描像を提出。

(2) 平均場近似による解析のバックアップ及び実用的アルゴリズムとして、モンテカルロ法を用いたアルゴリズムを設計。

(3) 生命科学で研究を行っている共同研究者と協力し、ニューロンのデータ解析を上記の手法を用いて行う。

### 4. 研究成果

(1) 最大エントロピー分布とそれ以外の分布の網羅的比較 (発表論文、 ):

統計力学的手法を用いて、拘束条件 (= 観測結果) を満たす確率分布のアンサンブルを考え、それを解析した。最大エントロピー分布とそれ以外の分布で、どちらが推定においてより好ましい性質を持つかを解析的に議論した。その結果、最大エントロピー分布が、他の分布よりも優れているとする根拠は見つけられなかった。これは解析の都合導入した“ランダムな”拘束条件下では、最大エントロピー分布に利点は無く、逆に言えばある特別な拘束条件の元でのみ、最大エントロピー法は力を発揮することを意味する。

また、この解析に対応したモンテカルロ法を開発し、確率分布そのもののアンサンブルを取り扱うアルゴリズムを開発した。これにより理論解析のバックアップを行った。この

アルゴリズムはデータ解析に用いることも出来る。

(2) スパースモデリングの理論解析 (発表論文、 ):

最大エントロピー法が、与えられた条件以外にはなるべく事前知識を入れない方法論とすれば、スパースモデリングは、適切な基底においてデータがスパースになる、という事前知識を強く利用することで、高度な情報処理を可能にする方法論である。スパースモデリングを用いたとき、情報理論的にどこまで元のシグナルを復元できるか、あるいは情報を圧縮できるか、というのを理論的に明らかにすることは重要である。

統計力学的手法を用いて、これの情報理論限界を導出した。またよく使われる代表的なアルゴリズム・緩和法では、これに到達できないことを数値計算により明らかにした。

(3) スパース線形回帰における緩和法に頼らないアルゴリズムの開発 (発表論文、 ):

上で述べたスパースモデリングにおける情報理論限界を到達するためのアルゴリズムをモンテカルロ法を用いて開発した。実際に数値計算を行い、きちんとデータのスケールリングを行うことで、たしかにその情報理論限界に、多項式時間かつ十分実用に与える時間内で、到達することを明らかにした。

(4) スパース線形回帰における近似的交差検証法の開発 (発表論文、 ):

スパースモデリングでは、解にスパース性を導入するため、正則化項と呼ばれるある種の人為的な項を問題に導入する。この人為的な項をどのようにハンドリングするかが、しばしば情報処理において重要と成る。多くの場合は、その項の係数をモデルの予測性能や一致性などに基づいて最適化することで、これを達成する。

モデルの予測性能は、交差検証法と呼ぶ方法で測るのが一般的である。交差検証法はデータの生成プロセスに強い仮定を置かないため、非常に汎用的で大変広い文脈で使うことが出来る。しかしながら、その実行には計算コストが多くかかる傾向があり、多くの応用においてそれが律速と成っている。

本研究では、この交差検証法を近似的に行う半解析的な公式を導出し、これにより追加の計算量をほとんどかけることなく交差検証法を実行可能にし、そしてそれが上手く働くことをシミュレーションデータと実データの双方で示した。この方法は、多くの問題で利用・実装され始めている。

(5) ボルツマン学習における自然画像の特徴表現 (発表論文、 ):

ニューラルネットワークという生物の神経細胞の動作原理を模して提案された、一群

の機械学習モデルがある。ボルツマンマシンはその初期のモデルの1つであり、最大エントロピー法と、極めて密接に関連している。

このややオールドファッションなモデルを用いて、自然画像を学習した際にどのような特徴が学習結果に現れるかを調べた。結果、学習したモデルの相互作用には、ユニバーサルな側面と、個別的な側面とが分離して現れることがわかった。ユニバーサルな側面は、相互作用が短距離的になること及び速やかに減衰する強磁性相互作用が現れることである。個別的側面として、ピクセル間の最近接相互作用が画像セットに応じて大きく異なること(負の値も取れ、フラストレーションがあらわれること)がわかった。

これらの知見をさらにスパースモデリングや臨界現象の知見と比較し、学習したボルツマンマシンが(自然画像やスパースモデリングによって得られたモデルがそうであるにも関わらず)臨界現象を表さないことから、自然画像には高次の相互作用に臨界性が繰り込まれていることが明らかとなった。

(6) ニューロンのスパイクデータにおける最大エントロピー分布の利点(発表論文):

(1) で開発した確率分布のアンサンブルを生成するモンテカルロ法を用いて、タイガーサラマンダー網膜細胞から取られたニューロンのスパイクデータを解析した。(1) では解析の都合上、観測量がランダムに生成されると仮定されたが、実データでは状況が大きく異なるため、最大エントロピー分布が他の分布と比較して有利になる可能性がある。これを吟味した。

結果、最大エントロピー分布が、特にシステムサイズが大きくなるにつれ、他の分布の大多数より優れていること、また最大エントロピー分布が負ける際には、3体以上の相関が強いておきであることを定量的に示した。

(7) 統計力学的手法そのものの拡張(発表論文、):

統計力学をデータ解析に使うだけでなく、統計力学的手法そのものを拡張することも重要である。

発表論文、では、疎結合スピングラスの理論を生態系のモデル解析に適用した。このモデルでは、一体化した分布をどのようにパラメトライゼーションするかが非自明であり、そのハンドリングに新しい理論的工夫が必要となる。ここでは、ある種の摂動展開とガウス近似を用いて、あるパラメータ領域で厳密とかんがえられる解を導いた。それによって、種分布に極めて非自明な多峰分布が自然に現れる場合があることを導き、それと実際に観測されている種分布との関係性を論じた。また、ガウス近似が破綻する領域でも正しく一体化分布をハンドリングできる方法についても示唆を得た。

発表論文、では、量子スピン系の波動関数

のオーバーラップを一般的な状況で近似計算手法を導いた。これは無限次元極限で厳密に成ると考えられ、ある種の新しい平均場近似と捉えることが出来る。これによりそのオーバーラップに現れる動的特異性を初めて定量的に捉え包括的に議論した。実験における観測可能性や量子情報・エンジニアリングにおける応用可能性についても論じた。

## 5. 主な発表論文等

[雑誌論文](計12件)

Tomoyuki Obuchi, Simona Cocco, Rémi Monasson, **Learning probabilities from random observables in high dimensions: the maximum entropy distribution and others**, 査読有, J. Stat. Phys. 161(3) (2015) 598-632

Tomoyuki Obuchi and Rémi Monasson, **Learning probability distributions from smooth observables and the maximum entropy principle: some remarks**, 査読有, J. Phys.: Conf. Ser. 638 (2015) 012018(1-23)

Yoshinori Nakanishi-Ohno, Tomoyuki Obuchi, Masato Okada, Yoshiyuki Kabashima, **Sparse approximation based on a random overcomplete basis**, 査読有, J. Stat. Mech. (2016) 063302 (1-30)

Tomoyuki Obuchi and Yoshiyuki Kabashima, **Sampling approach to sparse approximation problem: determining degrees of freedom by simulated annealing**, 査読有, DOI: 10.1109/EUSIPCO.2016.7760448

Tomoyuki Obuchi and Yoshiyuki Kabashima, **Sparse approximation problem: how rapid simulated annealing succeeds and fails**, 査読有, J. Phys.: Conf. Ser. 699 (2016) 012017(1-12)

Tomoyuki Obuchi, Yoshiyuki Kabashima, and Kei Tokita, **Relative species abundance of replicator dynamics with sparse interactions**, 査読有, J. Stat. Mech. (2016) 113502 (1-28)

Tomoyuki Obuchi, Hirokazu Koma, and Muneki Yasuda, **Boltzmann-machine learning of prior distributions of binarized natural images**, 査読有, J. Phys. Soc. Jpn. (2016) 85 114803 (1-17)

Tomoyuki Obuchi, Yoshiyuki Kabashima, and Kei Tokita, **Multiple peaks of species abundance distributions induced by sparse interactions**, 査読有, Phys. Rev. E 94

(2016) 022312 (1-5)

Tomoyuki Obuchi and Yoshiyuki Kabashima, **Cross validation in LASSO and its acceleration**, 査読有, J. Stat. Mech. (2016) 053304 (1-36)

Yoshiyuki Kabashima, Tomoyuki Obuchi, and Makoto Uemura, **Approximate cross-validation formula for Bayesian linear regression**, 査読有, Proceedings of 54th Annual Allerton Conference

Ulisse Ferrari, Tomoyuki Obuchi, Thierry Mora, **Random versus maximum entropy models of neural population activity**, 査読有, Phys. Rev. E 95 (2017) 042321 (1-5)

Tomoyuki Obuchi, Sei Suzuki, Kazutaka Takahashi, **Complex semiclassical analysis of the Loschmidt amplitude and dynamical quantum phase transitions**, 査読有, Phys. Rev. B 95 (2017) 174305 (1-11)

〔学会発表〕(計 18 件)  
以下主要な国際会議での発表のみ記述する。

Tomoyuki Obuchi  
Workshop on Statistical physics, Learning, Inference and Networks (L'Ecole de Physique des Houches, France, 2017/2/26-3/3, Poster)

Tomoyuki Obuchi  
Entanglement and non-equilibrium physics of pure and disordered systems (ICTP, Trieste, Italy, 2016/7/25-27, Poster)

Tomoyuki Obuchi  
Statphys26 (Centre de Congrès de Lyon, Lyon, France, 2016/7/18-22, Poster)

Tomoyuki Obuchi  
Statistical physics methods in biology and computer science (Ecole Normale Supérieure, Paris, France, 2016/7/11-16, Oral)

Tomoyuki Obuchi, Remi Monasson, and Simona Cocco  
APS march meeting (Baltimore Convention Center, Baltimore, USA, 2016/3/14-18, Oral)

Tomoyuki Obuchi  
Physics Informed Machine Learning (Inn at Loretto, 211 Old Santa Fe Trail, Santa Fe, USA, 2016/1/19-22), Poster)

Tomoyuki Obuchi

New Horizons of Quantum and Classical Information 2015 (NHQCI2015) (Tokyo Tech, Tokyo, 2015/8/3-5), Oral, Invited)

Tomoyuki Obuchi, Remi Monasson, and Simona Cocco  
Statphys-Kolkata VIII (S.N.Bose Laboratory, Kolkata, 2014/11/30-12/6, Oral, Invited)

Tomoyuki Obuchi, Remi Monasson, and Simona Cocco  
Collective Dynamics in Information Systems (KITPC, Beijing, 2014/10/8-31, Oral)

〔図書〕(計 0 件)

〔産業財産権〕  
無し

出願状況 (計 0 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
出願年月日 :  
国内外の別 :

取得状況 (計 0 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
取得年月日 :  
国内外の別 :

〔その他〕  
ホームページ等

## 6. 研究組織

### (1) 研究代表者

小淵 智之 (OBUCHI Tomoyuki)  
東京工業大学・情報理工学院・助教  
研究者番号 : 40588448

### (2) 研究分担者

( )

研究者番号 :

### (3) 連携研究者

( )

研究者番号：

(4)研究協力者 ( )