

## 研究 成 果 報 告 書

(国立情報学研究所の民間助成研究成果概要データベース・登録原稿)

研究テーマ (和文) AB		統計力学的アプローチによる、大規模スパースモデリングの理論と技術の強化			
研究テーマ (欧文) AZ		Statistical mechanical approach to massive sparse modeling techniques			
研究氏 代 表 名 者	カナ CC	姓)オブチ	名)トモユキ	研究期間 B	2017 ~ 2019 年
	漢字 CB	小淵	智之	報告年度 YR	2019 年
	ローマ字 CZ	Obuchi	Tomoyuki	研究機関名	東京工業大学
研究代表者 CD 所属機関・職名		東京工業大学・助教			
<p>概要 EA (600 字～800 字程度にまとめてください。)</p> <p>スパースモデリング (SpM) とは、対象となるデータを少数の変数・自由度で表現することにより、膨大な高次元データの処理を可能にするモデリング技術である。本研究では、SpM の実行において問題になる計算量的困難を統計力学的手法によって克服し、SpM の理論的技術的基盤を強化することを目指したものである。</p> <p>申請時に目標に掲げたのは次の3つである：</p> <ol style="list-style-type: none"> <li>1. 統計力学におけるモンテカルロ法を駆使することで、効率的に変数選択を行うアルゴリズムを構築する。</li> <li>2. ベイズ推定・統計力学の考え方を文脈に応じて選択適用することにより、各変数に選択される確率分布を設定し、それを利用して各変数の重要度が評価する。それを実際に変数選択に応用する。</li> <li>3. モデルの解釈・選択を行う上で、汎化誤差の評価を行う一般的枠組みである交差検定法が重要であるが、一般に計算量的な負荷がたかい。そこでデータが大規模であることを逆用して摂動論の技術を用いて交差検定法を低計算量で実行する方法論を開発する。</li> </ol> <p>実際の研究において、この3つすべてに対応する研究成果を上げることが出来た。次ページに載せた発表文献の1つ目が1.、2つ目が3. 3つめが2. に対応する。これ以外にも2つ原著論文を査読あり雑誌に発表することが出来たが、紙面制約上次ページでは省略する。発表論文はいずれも統計力学/機械学習分野で知られた著名な雑誌に掲載されている。開発したアルゴリズムは広くユーザーに使われるよう Github などを利用して公開している。</p>					
キーワード FA	スパースモデリング	統計力学	機械学習	多自由度推定	

(以下は記入しないでください。)

助成財団コード TA					研究課題番号 AA								
研究機関番号 AC					シート番号								

発表文献（この研究を発表した雑誌・図書について記入してください。）										
雑誌	論文標題 <sup>GB</sup>	Statistical mechanical analysis of sparse linear regression as a variable selection problem								
	著者名 <sup>GA</sup>	Tomoyuki Obuchi et al.	雑誌名 <sup>GC</sup>	Journal of Statistical Mechanics						
	ページ <sup>GF</sup>	1~41	発行年 <sup>GE</sup>	2	0	1	8	巻号 <sup>GD</sup>	103401	
雑誌	論文標題 <sup>GB</sup>	Accelerating Cross-Validation in Multinomial Logistic Regression with L1-Regularization								
	著者名 <sup>GA</sup>	Tomoyuki Obuchi and Yoshiyuki Kabashima	雑誌名 <sup>GC</sup>	Journal of Machine Learning Research						
	ページ <sup>GF</sup>	1~30	発行年 <sup>GE</sup>	2	0	1	8	巻号 <sup>GD</sup>	19 (52)	
雑誌	論文標題 <sup>GB</sup>	Semi-Analytic Resampling in Lasso								
	著者名 <sup>GA</sup>	Tomoyuki Obuchi and Yoshiyuki Kabashima	雑誌名 <sup>GC</sup>	Journal of Machine Learning Research						
	ページ <sup>GF</sup>	1~33	発行年 <sup>GE</sup>	2	0	1	9	巻号 <sup>GD</sup>	20 (70)	
図書	著者名 <sup>HA</sup>									
	書名 <sup>HC</sup>									
	出版者 <sup>HB</sup>		発行年 <sup>HD</sup>					総ページ <sup>HE</sup>		
図書	著者名 <sup>HA</sup>									
	書名 <sup>HC</sup>									
	出版者 <sup>HB</sup>		発行年 <sup>HD</sup>					総ページ <sup>HE</sup>		

#### 欧文概要<sup>EZ</sup>

Sparse Modeling (SpM) is a modelling technique by representing target data with a small number of variables and degrees of freedom. This technique enables the processing of vast amounts of high-dimensional data using the sparsity of the representation, but some techniques in SpM tend to suffer from computational complexity. In this study, We overcome the computational difficulties by using statistical mechanical methods. By this, we aim at strengthening the theoretical and technical basis of SpM.

We have suggested the following three goals in the beginning of this project:

1. Invention of an algorithm for efficient variable selection by using Monte Carlo method in statistical mechanics.
2. Invention of a framework computing the importance of each variable in the model by using the Bayesian statistics and statistical mechanics. Application of it to actual variable selection algorithm.
3. Cross-validation is an important method for selecting a model in a model family. However, it tends to suffer from the computational complexity problem. We invent methods to overcome this computational difficulty by using some perturbative techniques.

In the actual research, we have achieved the results corresponding to all these three. The above publications are corresponding to all these three points. Additional two publications were obtained in this project but they are not shown above due to the restriction of the space. All of the published papers have been published in well-known journals in the field of statistical mechanics/machine learning. The algorithms we developed are available on Github and other sites for wider use.