

2021年 1月 8日

情報・知能工学専攻	学籍番号	第 143320 号	指導教員	渡辺 一帆
				石田 好輝
氏名	小林真佐大			

論文内容の要旨 (博士)

博士学位論文名	f 分離可能ブレグマン歪み尺度に基づく機械学習アルゴリズムの拡張と統計的性質の解明
---------	---

(要旨 1,200 字程度)

近年、コンピュータの発達によるインターネットの普及やセンサの高集積化による IoT (もののインターネット) の普及により、収集されるデータが大規模化しており、そのデータを有効に活用するための手法が求められている。その手法として、機械学習法が様々な応用で用いられている。機械学習は与えられたデータから、学習を行い未知データの予測や潜在構造の抽出を行う。しかしながら、実際のデータにはヒューマンエラーや機器の故障などの様々な原因により、外れ値と呼ばれる学習に悪影響を与えるデータが含まれる。外れ値の対処のためには、外れ値を除外して学習を行うことが考えられるが、その場合には外れ値とは何かがわかっていなければならない。データを実際に見て、外れ値を除外するのも一つの方法かもしれないが、データ数や次元が大きな場合にはそのようなことは困難である。外れ値に対処するための方法は、古くからロバスト統計の分野で研究されてきた。特に、近年における主流は、外れ値に頑健なダイバージェンスと呼ばれる確率分布間の乖離度を測る指標を用意し、そこからアルゴリズムを構築するといった手順である。このダイバージェンスには、 β 、 γ ダイバージェンスがよく使用されているが、推定量の一致性のために解析的計算が困難となるバイアス補正項を含む。本研究では、外れ値に頑健かつ計算コストが低く簡便な機械学習アルゴリズムの構築を目的とする。具体的には、情報理論の分野で提案された f 分離可能歪み尺度とブレグマンダイバージェンスを組み合わせた f 分離可能ブレグマンダイバージェンス歪み尺度に基づく推定法を考案し、機械学習アルゴリズムへの適用とその統計的な性質を調査した。まず、3章で f 分離可能ブレグマン歪み尺度に基づく推定法を構築した。4章では、外れ値に対する頑健性の評価指標の一つである影響関数を導出し、影響関数の有界性が成り立つための関数 f の条件について議論した。5章では、バイアス補正項なしに推定量の一致性の必要条件である推定方程式の不偏性が成り立つための条件について議論した。6章と7章にて、ディクレ過程平均法と非負値行列分解をそれぞれ f 分離可能ブレグマン歪み尺度に基づくアルゴリズムに一般化し、数値実験によってその性能評価を行った。8章では本論文の総括を行った。

Date of Submission (month day, year) : 1/8/2021

Department of Computer Science and Engineering	Student ID Number D143320	Supervisors Kazuho Watanabe Yoshiteru Ishida
Applicant's name Masahiro Kobayashi		

Abstract (Doctor)

Title of Thesis	Machine Learning Algorithms for f -Separable Bregman Distortion Measures and Analysis of Statistical Properties
-----------------	---

Approx. 800 words

In recent years, with the spread of the Internet by the development of computers and the spread of the IoT (Internet of Things) by the high integration of sensors, the amount of data collected has become large, and methods for effectively utilizing the data are required. For such a purpose, machine learning methods are used in various applications. Machine learning learns from given data, predicts unknown data, and extracts latent structures. However, the actual data includes data called outliers, which adversely affect learning due to various causes such as human error and equipment failure. In order to deal with outliers, it is conceivable to exclude outliers for learning, but in that case, it is necessary to know what the outliers are. It may be one way to actually look at the data and exclude outliers, but it is difficult to do so when the number of data and the dimensions are large. Methods for dealing with outliers have long been studied in the field of robust statistics. In particular, the mainstream in recent years is the procedure of preparing an index for measuring the degree of deviation between probability distributions called divergence, which is robust against outliers, and constructing an algorithm from it. β and γ divergences are often used for this divergence, but they include a bias correction term that makes analytical calculations difficult for the consistency of the estimators. The purpose of this study is to construct a simple machine learning algorithm that is robust against outliers and has low computational cost. Specifically, we devised an estimation method based on the f -separable Bregman divergence distortion measures, which is a combination of the f -separable distortion measures and Bregman divergence proposed in the field of information theory, and introduced it to machine learning algorithms. We investigated the application and its statistical properties. First, in Chapter 3, we construct an estimation method based on f -separable Bregman distortion measures. Second, in Chapters 4 and 5, we discuss robustness against outliers and the condition of unbiased estimation equation as a necessary condition for the consistency of the estimator, respectively. Third, in Chapters 6 and 7, for the Dirichlet process means and non-negative matrix factorization are generalized to algorithms based on f -separable Bregman distortion measures, respectively, and their performance is evaluated by numerical experiments. Finally, Chapter 8 summarizes this thesis.