Information-Based Modeling of Data Learning Algorithms

(データ学習アルゴリズムの情報論的モデル化)

学位論文内容の要旨

近年、情報工学の分野において、様々なデータ(情報)を処理するための道具立てとして学習理論が注目を集めており、計算機技術の進歩に伴い、その有用性はデータを扱う様々な分野に渡って広く認知されている。また、神経回路網の数理モデルに端を発する学習理論の研究は、昨今では、確率・統計・情報理論・統計物理などにおける解析手法を取り入れながら、情報に関する理論的研究分野として急速に発展している。しかしながら、これまで行われてきた学習理論の研究の多くは、個別の学習アルゴリズムに特化した解析に留まっており、それらの根底にある共通の数理的枠組に着目した体系的な研究はそれほど多くなかった。

そこで本研究では、情報論的モデル化という一貫した立場から様々な学習アルゴリズムの解析を行う。この立場は、学習機械を規定するための確率分布(確率モデル)、及びその確率分布を比較するための尺度(分布間情報量)の設定が、情報論的観点から学習アルゴリズムを構成する際の基本要件であるという考え方に根ざしている。すなわち、本研究の目的は、確率分布と分布間情報量を軸とするという立場が、様々な学習アルゴリズムの特性やそれらの関係の明確化、或いは、実用上の制約条件や目的に応じた学習アルゴリズムの設計を行う上で有用であるということを示すものである。

本論文は、6つの章から構成される。第1章では、自乗誤差関数を損失関数とする入出力学習機械の学習アルゴリズム(自乗誤差最小学習)を議論の出発点とし、本研究の動機付けについて述べた。これは、入力に対する出力の条件付確率分布を分散が定数の正規分布として与え、確率分布を比較するための情報量基準としてカルバック-ダイバージェンスを与えることが、自乗誤差最小学習と等価な学習アルゴリズムを導くことに着目したものである。本研究では、この特徴を発展させることによって、様々な学習アルゴリズムの解析を行った。

第2章では、正則化項と類似した付加項を持つ学習アルゴリズムの情報論的モデル化を行った。このモデル化は、入力に対する出力の条件付確率分布として分散が関数(学習機械が持つパラメータを用いて定義される正值関数)の正規分布を用い、確率分布を比較するための情報量基準としてカルバック-ダイバージェンスを用いるものであり、自乗誤差最小学習の一種の拡張となっている。本章では、このようにして導入された学習アルゴリズムの付加項が確率分布の分散と関連していることを明らかにした。また、このことに着目し、付加項付き学習アルゴリズムがある種のロバスト性、すなわち、データに含まれる特殊なノイズを除去する性質を持つことを明らかにした。

第3章では、アンサンブル学習の情報論的モデル化を行った。良く知られたアンサンブル学習には、自乗誤差関数を損失関数とし、複数の学習機械の重み付け平均を新たな学習機械とするものがある。本章では、このアンサンブル学習を特別な場合として含む一般化を行い、そのアルゴリズム構造を特徴付ける以下の知見を得た。

- アンサンブル学習のアルゴリズムは、カルバック-ダイバージェンスに関する三段階の最小

化操作に帰着できる。

- アンサンブル学習機械に対応する確率分布は、それを構成する個々の学習機械に対応する確率分布の指数型混合分布族として与えられる。

さらに、この結果を利用することでアンサンブル学習の有効性に関する以下の知見を得た。

- 学習サンプルを分割してアンサンブル学習機械を構成する個々の学習機械の学習を行うよりも、学習サンプルを分割せずに単一の学習機械の学習を行う方が効果的である。ただし、学習サンプルを分割せずにそれぞれ異なる方法で個々の学習機械の学習を行う場合、アンサンブル学習は効果的である。また、学習済みの複数の学習機械が与えられている場合、アンサンブル学習は効果的である。
- 性能の低い学習機械を用いてアンサンブル学習機械を構成した場合、その重みは負の値を取る。また、この逆も成り立つ。

一方、カルバック - ダイバージェンスに関する三段階の最小化操作によるアンサンブル学習の特徴付けは、 α - ダイバージェンスを用いた場合にも自然に拡張できることを明らかにした。例えば、双対カルバック - ダイバージェンスを用いたアンサンブル学習における確率分布は線型混合分布族に対応することが分かった。

第2, 3章では、主に、確率分布の設定方法に着目し、分布間情報量としてカルバック - ダイバージェンスを用いる学習アルゴリズムの特徴付けを行った。カルバック - ダイバージェンス以外の分布間情報量には、 α - ダイバージェンスやその一般化である f - ダイバージェンスが知られている。しかし、これらの分布間情報量を用いた場合、カルバック - ダイバージェンスを用いた場合とは異なり、実装が容易な学習アルゴリズムを構築することが困難であるという問題点がある。すなわち、学習サンプルの単純な代入操作による学習アルゴリズムの構築が困難となる。こうした背景を受け、第4章では、 f - ダイバージェンスの不定性と呼ばれる特徴を用いることで、正值有限測度への適用を背景とした新しい分布間情報量のクラスを導いた。これは、これまで良く用いられてきた分布間情報量とは異なるクラスの情報量であるが、学習アルゴリズムの構築や実装を容易にする情報量や、アンサンブル学習を陽に定式化して解析するのに適した情報量を含むことを明らかにすることで、適用対象に応じて分布間情報量を拡張することの有効性の一端を示すことができた。例えば、本章で導入される分布間情報量の一つである $(\gamma, 1+\gamma)$ - ダイバージェンスは、学習サンプルの単純な代入操作に基づく簡便な学習アルゴリズムの構成を可能にするという特徴を持つことを示した。また、 $(1, 2)$ - ダイバージェンス ($(\gamma, 1+\gamma)$ - ダイバージェンスの $\gamma=1$ の場合) を用いたアンサンブル学習の構成が可能であり、このアンサンブル学習で用いられる確率モデルは双対カルバック - ダイバージェンスを用いて構成されるアンサンブル学習と同じ線型混合型分布族であることを示した。双対カルバック - ダイバージェンスを用いて構成されるアンサンブル学習は、学習サンプルの単純な代入操作による学習アルゴリズムとはならないため、 $(1, 2)$ - ダイバージェンスを用いる事でこの問題を回避することができる。

また、第5章では、第4章で導入された $(\gamma, 1+\gamma)$ - ダイバージェンスを、通信トラフィックデータの従う裾野分布の分析へ応用した。これは、高速のインターネット回線においてリアルタイムの分析を可能とするものである。本章では、インターネット上の異なるポイントで実測されたデータに対して、上記の分析方法は高精度の結果を与えることを確認した。

第6章は、本論文の結論である。

本研究による成果は、確率分布と情報量基準を軸とした情報論的モデル化という枠組により、様々な学習アルゴリズムの特性の解明や、従来では得られなかった実用的な学習アルゴリズムの構築に利便性を与えるものである。

学位論文審査の要旨

主 査 教 授 大 森 隆 司
副 査 教 授 大 内 東
副 査 教 授 和 田 充 雄

学 位 論 文 題 名

Information-Based Modeling of Data Learning Algorithms

(データ学習アルゴリズムの情報論的モデル化)

近年、情報工学の分野において、様々なデータ（情報）を処理するための道具立てとして学習理論が注目を集めており、その有用性は様々な分野に渡って広く認知されてきた。学習理論の研究は、昨今では、確率・統計・情報理論・統計物理などにおける解析手法を取り入れながら、情報に関する理論的研究分野として急速に発展している。しかし、これまでの学習理論の研究の多くは、個別の学習アルゴリズムに特化した解析に留まり、それらの根底にある共通の数理的枠組についての体系的な研究は多くはない。

この認識のもとで本研究は、情報論的モデル化という立場から様々な学習アルゴリズムの解析を行っている。この基本は、学習を規定する確率分布（確率モデル）、及びその分布を比較する尺度（分布間情報量）の設定が、情報論的観点から学習アルゴリズムを構成する際の基本要件になるという考え方である。すなわち本研究では、確率分布と分布間情報量を軸として、様々な学習アルゴリズムの特性やそれらの関係の明確化、さらには実用上の制約や目的に応じた学習アルゴリズムの設計に有用であることを示している。

この主張に基づき本論文は、6つの章から構成される。第1章は、自乗誤差関数を損失関数とする入出力学習機械の学習アルゴリズム（自乗誤差最小学習）を出発点として、入力に対する出力の条件付確率分布を分散が定数の正規分布として与え、確率分布を比較するための情報量規準としてカルバック・ダイバージェンス (KLD) を与えることが、自乗誤差最小学習と等価な学習アルゴリズムを導くとした。本論文では、この特徴を多様な方面に発展させ、様々な学習アルゴリズムの解析を行ない、新たな知見を得ている。

第2章では、正則化項と類似した付加項を持つ学習アルゴリズムの情報論的モデル化を行っている。このモデル化は、入力に対する出力の条件付確率分布として分散が関数の正規分布を用い、確率分布を比較するための情報量規準としてKLDを用い、自乗誤差最小学習の一種の拡張となっている。本章では、このようにして導入された学習アルゴリズムの付加項が確率分布の分散と関連していることを明らかにした。また、付加項付き学習アルゴリズムがある種のロバスト性、すなわち、データのノイズを除去する性質を持つことを明らかにした。

第3章では、アンサンブル学習の情報論的モデル化を行っている。アンサンブル学習には、自乗誤差関数を損失関数とし、複数の学習機械の重み付け平均を新たな学習機械とするものがある。本章では、このアンサンブル学習を含む一般化を行い、そのアルゴリズムを特徴付ける知見を得ている。

- アンサンブル学習のアルゴリズムは、KLDに関する三段階の最小化操作に帰着できる。
- アンサンブル学習機械の確率分布は、個々の学習機械の確率分布の指数型混合分布族として与えられる。

一方、KLDに関する三段階の最小化操作によるアンサンブル学習の特徴付けは、 α -ダイバージェンスを用いた場合にも自然に拡張でき、例えば双対KLDを用いたアンサンブル学習の確率分布は線型混合分布族に対応することを明らかにした。

KLD以外の分布間情報量である f -ダイバージェンスは、実装が容易な学習アルゴリズムを構築することが困難であるという問題がある。そこで第4章では、 f -ダイバージェンスの不定性を用い、正值有限測度への適用を背景とした新しい分布間情報量のクラスを導いた。これは、学習アルゴリズムの構築や実装を容易にする情報量や、アンサンブル学習を陽に定式化して解析するのに適した情報量を含むことを明らかにし、適用対象に応じて分布間情報量を拡張することの有効性を示した。例えば、 $(1, 2)$ -ダイバージェンスを用いたアンサンブル学習が構成可能であり、この学習の確率モデルは双対KLDを用いたアンサンブル学習と同じ線型混合型分布族であることを示した。

第5章では、第4章で導入された $(\gamma, 1+\gamma)$ -ダイバージェンスを、通信トラフィックデータの従う裾野分布の分析アルゴリズムの合成へ応用した。これは、高速インターネット回線のリアルタイム分析を可能とし、インターネット上の実測データに対して、上記の方法による分布推定モデルは高精度の結果を与えることを確認した。

第6章は、本論文の結論である。確率分布と情報量規準を軸とした情報論的モデル化という枠組により、様々な学習アルゴリズムの特性の解明や、従来では得られなかった実用的な学習アルゴリズムの構築に利便性を与えるとしている。

これを要するに本論文は、データ学習のアルゴリズム群に対して統一的な視点を提供し、問題に応じた学習アルゴリズムを設計する方式についての知見を得たものであり、学習の理論の構築に対して貢献するところが大きい。よって著者は、北海道大学博士（工学）の学位を授与される資格あるものと認める。