

審査の結果の要旨

氏名 坂井 智哉

コンピュータの飛躍的な性能向上やインターネットの爆発的な普及に伴い、大量のデータが日々生み出されており、これらのデータの背後に潜む規則性や構造が抽出できれば、将来の予測や現象の理解などに役立てられる。しかし、膨大な量のデータを人間が解析することは難しいため、ヒトのような知的な情報処理をコンピュータで実現することを目指す機械学習と呼ばれる分野が注目を集めている。

例えば、画像を入力として、その画像に写っている物体を答えとして出力する機械学習アルゴリズムを、教師付き学習とよばれる枠組みで構築する場合、入力画像とその答え（ラベル）の組をたくさん準備し、それらラベル付き画像を用いて画像分類器を学習する。しかし、一般にラベル付きデータの収集にはコストがかかるため、大量にデータを集めることが困難な場合がある。本博士論文の主題である弱教師付き機械学習は、ラベル付きデータよりも収集コストの低い弱教師付きデータを活用して、低コストで精度良く教師付き学習を実現することが目的である。

本論文は以下のように英文6章からなる。

第1章は序論であり、本論文の対象である弱教師付き学習の基礎となる教師付き学習、教師なし学習、半教師付き学習、正例とラベルなしデータからの学習の概要を説明するとともに、提案手法の概要、および、本論文の構成について述べている。

第2章は、従来の教師付き学習、教師なし学習、半教師付き学習、正例とラベルなしデータからの学習の問題を数学的に定式化し、主要なアルゴリズムの技術的な詳細、および、長所・短所を解説している。

第3章では、正例とラベルなしデータのみが与えられる状況で、ラベルなしデータ内の正例と負例の割合に大きく偏りがある場合に、分類器を精度良く学習する方法を提案している。従来法では、分類器を訓練する際の学習規準が真の分類リスクに対してバイアスがあるという問題があった。それに対して本章では、バイアスのない新しい学習規準を導き、その汎化性能を理論的・実験的に評価するとともに、提案手法の有用性を計算機実験により実証している。

第4章では、正例とラベルなしデータのみが与えられる状況で、入出力間の統計的な依存関係を解析する手法を提案している。具体的には、二乗損失相互情報量とよばれる統計的従属性尺度を正例とラベルなしデータのみから推定する手法を開発し、その推定精度を理論的・実験的に評価している。また、提案した二乗損失相互情報量の推定量を用いて、次元削減と独立性検定のアルゴリズムも構築しており、計算機実験によりそれら有用性を実証している。

第5章では、正例・負例・ラベルなしデータを用いる半教師付き学習の新しい手法を提案している。従来の半教師付き学習法では、クラスタ仮定とよばれる、データ分布に対する強い条件が仮定されており、実用上、大きな制約となっていた。それに対して本章では、正例とラベルな

しデータからの学習法を活用することにより，そのような仮定が必要ない新しい半教師付き学習手法を提案している．そして，その汎化性能を理論的・実験的に評価し，提案手法の有用性を計算機実験により実証している．

第6章では，結論と今後の課題について述べている．

以上要するに本論文は，限られた教師情報から精度良く機械学習を行うための新しい枠組みを提案するものであり，開発されたアルゴリズムは有用性が理論的に保証されるという特徴を持つ．また，それらの実際的な有用性も，計算機実験により実証されている．このような研究成果は，複雑理工学，特に複雑系科学の理解のために必要な機械学習分野に対する貢献が大きい．

よって本論文は博士（科学）の学位請求論文として合格と認められる．