

# 審査の結果の要旨

氏名 山根 一航

計算機および計算機科学のめまぐるしい進歩により、コンピュータを使って自動的かつ効率的にデータを有用な知識に変換するための技術である機械学習がその重要性を増している。これまでの機械学習技術の成功の裏には、質の高い多量のデータの存在がある。しかしながら、すべての実応用領域において、十分な量と質を兼ね備えたデータが用意できるとは限らない。このような背景のもと、本論文では、訓練データが限られた情報しかもたない状況下での機械学習について論じている。

本論文は以下のように英文6章からなっている。

第1章は序論であり、機械学習や弱教師付き学習の背景を説明するとともに、本論文で提案する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章では、機械学習の基本的な概念や定義を導入している。具体的には、学習アルゴリズム、仮説関数、訓練データやテストデータ、損失関数や期待損失などの概念を導入し、学習の目的や結果の評価方法が一般的にどのように定義されるかについて述べている。また、本論文の主題である限られた情報からの学習の枠組みに深く関連する、半教師付き学習、弱教師付き学習、転移学習、マルチタスク学習などの話題を紹介している。

第3章では、クラスタ数が未知の場合でも適用可能な最頻値探索クラスタリングや、データ分布の正規性の尺度の推定などに応用を持つ対数密度勾配推定について述べている。データ変数が多次元ベクトル値をとる場合、学習対象である対数密度勾配は多次元ベクトル値関数になる。この問題は各次元を別々の関数とみなして一つ一つ推定することで従来法が適用できる。しかし、これらの次元はすべて同じ原始関数に偏微分演算を施して得られるものであるという共通点を持ち、ある次元の学習結果が別の次元の学習にとって有用な情報をもつことが期待される。このような考え方に基づいて、次元をまたいで情報を共有しながら対数密度勾配を学習するマルチタスク手法を提案している。そして、提案手法が高い精度で対数密度勾配を推定できること、および、提案法に基づく最頻値探索クラスタリングによって高いクラスタリング性能が得られることを計算機実験により示している。

第4章では、互いに似通った複数の主成分分析タスクがある場合に、マルチタスク学習のアプローチを用いてタスク間で情報を共有しながら同時に問題を解く手法について論じている。主成分分析の特殊な性質として、学習対象である射影行列の空間がユークリッド空間とは異なる幾何構造を持っていることが挙げられる。したがって、他の問題で良く行われるように各タスクの学習結果をユークリッド距離に基づいて近づけると、

射影行列の類似性をうまく反映できない恐れがある．また，通常の行列全体の空間は射影行列全体の空間よりも大きな次元を持つため，射影子の行列表現を素朴に扱くと推定や最適化の効率性の観点からも無駄が生じてしまう可能性がある．このような問題に対処すべく，射影行列全体がなす多様体上で直接的に最適化する手法を提案している．そして，人工データや脳・コンピュータ・インターフェイスの実データ実験により，提案法が精度の上でも計算効率的の上でも優れていることを示している．

第5章では，限られた情報からの向上作用モデリングについて論じている．向上作用モデリングの目的は，特定の処置による結果の変化に対する効果を推定することである．向上作用モデリングは，特定の処置をどの個体に対して行えば全体としての処置効果が大きくなるかという点に関心があるため，個体ごとに特徴量で条件づけた場合の処置効果の推定が特に重要である．従来法では，処置と結果を表す2つのラベルの組が付与されたデータの存在を仮定しているが，このようなデータを集めるのが技術的あるいは倫理的・法的な理由で難しい場合がある．そこで，処置とその結果の両方がわかっているデータが一つも与えられない現実的な状況を考え，そのような状況でも学習が可能な新しい手法を提案している．そして，提案法の推定量は解析的かつ効率的に計算でき，理論的にも実験的にも高い精度を持つことを示している．

第6章では，結論と今後の課題について述べている．

以上を要するに，本論文は，訓練データが限られた情報しかもたない状況下での機械学習アルゴリズムを提案するものであり，開発されたアルゴリズムの有用性を理論的に解析するとともに，実証的な有用性を計算機実験により示している．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．

以上1917字