

## 審査の結果の要旨

氏名 石田 隆

近年、画像理解、音声認識、言語翻訳など様々な実世界の問題に対して、機械学習技術が用いられるようになってきた。しかし現状の機械学習技術では、十分な教師情報が得られない場合に信頼できるシステムを構築することは困難である。本博士論文は、限られた情報からでも信頼性の高いシステムを構築できる機械学習アルゴリズムを提案し、その理論的性質および実験的性能について論じるものである。

1章「Introduction」では、機械学習の背景を概観した後、弱教師付き学習と正則化技術に関する従来研究を紹介し、各手法の利点・欠点について論じている。そして、分類リスクを修正するという本博士論文で提案する機械学習アルゴリズムに共通する基本的な考え方、および、提案する3つのアルゴリズムそれぞれの概要を述べている。

2章「Background and Preliminaries」では、3章以降で提案する機械学習アルゴリズムで必要となる数学的な概念をまとめている。具体的には、データの分布について述べた後、2クラス分類問題に対する分類リスク、ベイズ誤差、学習モデル、損失関数、正則化などの概念を導入している。そして、それらの多クラス分類問題への一般化についても述べている。

3章「Learning from Positive-Confidence Data」では、正と負の2つのクラスへの分類問題を考え、負のデータが与えられない状況でも分類器を学習できる手法について論じている。この問題は、例えばあるユーザが自社の商品を買うかどうかを予測したいときに、自社のユーザのデータ（正データとする）は手に入るが、他社の商品を買ったユーザのデータ（負データとする）は手に入らないという状況に対応している。この問題に対して、正のデータに信頼度の情報が付与されていれば、正信頼度のデータから分類リスクを不偏推定できることを示している。そして、この不偏リスク推定量を最小にする学習結果は、真の解への収束性の意味で最適性を有していることを理論的に証明し、提案法によって比較対象の手法より高い分類性能が得られることを実験的に示している。

4章「Learning from Complementary Labels」では、3クラス以上の多クラス分類問題に対する学習アルゴリズムについて論じている。多クラス分類問題はクラス数が多いため、正しい教師ラベルを付与するのにコストがかかる。一方、正しくない教師ラベルを付与することは容易である。例えば、医用画像処理において、正しい病名を付与するためには医師が時間をかけて診断を行う必要があるが、「がんでない」といった否定ラベルは比較的短時間で付与できる。このような否定の教師情報を補ラベルとよび、補ラベルから多クラスの分類リスクを不偏推定できることを示している。本章の前半では、対称な損失関数に対する不偏リスク推定量を導出し、この不偏リスク推定量を最小にする学習結果が最適な収束率を達成できることを理論的に証明している。また、通常ラベルと補ラベルを組み合わせることによって、分類性能を大幅に

向上させられることを実験的に示している。本章の後半では、不偏リスク推定量の導出において、損失の対称性の制約が完全に除外することを示している。そして、データから推定した分類リスクが負になることを補正する改良法を提案し、深層ニューラルネットワークなどの複雑な学習モデルに対する性能が向上することを計算機実験によって示している。

5章「**Flooding: A Novel Regularizer to Avoid Overfitting**」では、確率的勾配法を用いた学習アルゴリズムにおいて、過適合を防ぐための方策について論じている。通常のカテゴリ分類問題では、クラス間に重なりがあるため、分類器が達成できる最小のカテゴリ分類誤差は一般にはゼロにならない。一方、深層ニューラルネットワークなど多数のパラメータを持つ非常に複雑なモデルでは、与えられた訓練データに対する誤差をゼロにすることは容易である。そこで、訓練誤差がある正の値以上になるように解を制約する、洪水法とよばれる学習アルゴリズムを提案している。洪水法によって、学習過程においてテストデータに対する分類誤差が2回降下する二重降下現象が誘発され、過学習が抑制できることを実験的に示している。

6章「**Conclusions and Future Work**」では、本博士論文の結論と今後の課題について述べている。

以上を要するに、本博士論文は、十分な教師情報が得られない場合でも、信頼の高いシステムを構築できる機械学習アルゴリズムを提案するものであり、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に貢献するものと考えられる。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1834字