

審査の結果の要旨

論文提出者氏名 大岩 秀 和

機械学習において、データとその正しい分類結果すなわち正解の組が教師データとして与えられたとき、それを利用して分類器を学習することは主要な課題である。典型的な分類器は高次元のデータと重みベクトルの内積の正負によって分類結果を決める。分類器の学習において、1組のデータと正解が処理される毎に重みベクトルの値を更新する方法をオンライン学習と呼ぶ。次元が高いデータの分類を高速に行う方法として重みベクトルにおけるゼロの成分を増やして簡素な分類器にすることが有力である。そのために、オンライン学習においては分類器の予測誤りを表す損失と、重みベクトルの簡素化を実現するL1正則化項を合わせた評価関数を最適化する手法が有力であるが、重要な特徴に対応する成分までゼロにしてかえって性能が悪化する現象が問題である。この問題への対策はオンライン学習の性能向上における課題である。重要な特徴をゼロにならないように保護する正則化項の拡張することが有力な方策である。さらに人間が分類問題を解く際に使う認知バイアスを取り込むオンライン学習の拡張の有力な手法である。本論文ではこれらの問題の解決を目標としている。

本論文は「Subgradient-based Online and Stochastic Learning with Biases」(劣勾配ベースのバイアス考慮型オンライン学習および確率的学習)と題し、6章からなる。

第1章「Introduction」(序論)では、本論文のテーマである分類器のオンライン学習と確率的学習における基本手法である劣勾配ベースの逐次更新型学習を導入する。この手法を利用して展開される提案手法の概要を説明している。

第2章「Online Learning」(オンライン学習)では、本論文の提案の枠組みであるオンライン学習に関して、目標をリグレットの最小化とし、これを劣勾配ベースのオンライン学習に適用している。次にオンライン学習の種々のアルゴリズムを説明し、学習における失敗回数の上界を抑えることを目的とした学習アルゴリズムの枠組みを紹介している。

第3章「Stochastic Learning」(確率的学習)では、劣勾配ベースの確率的学習に関して損失期待値の最小化および経験的リスク最小化を導入し、この最適化を基礎にする学習アルゴリズムを説明している。

第4章「Feature-aware Regularization for Sparse Online Learning」(疎なオンライン学習のための特徴考慮型正則化)は、本論文における第一の提案である。すなわちL1

正則化は分類のための重みベクトルにおける成分選択能力が高いがゆえに重要な特徴に対応する成分まで削除し分類性能がかえって劣化してしまう。この問題に対処するための正則化項の拡張方法について述べている。これは、オンライン学習の1回の更新毎の重みベクトルの各成分の劣勾配の q ノルムを正則化項に乗ずることによって実現する。この拡張を高い性能を示す既存のオンライン学習手法に適用して実験評価し、十分に疎でありかつ分類性能が良い重みベクトルを得られることを示している。さらに、提案手法におけるリグレットの上界を理論的に求めている。

第5章「Online and Stochastic Learning with a Human Cognitive Bias」(人間の認知バイアスによるオンライン確率的学習)は、本論文における第二の提案である。すなわち、人間は過去に正解したデータが学習の進行によって不正解になることを好まないという認知的なバイアスがある。これを学習アルゴリズムに組み込む方法として、分類器における重みベクトルの成分がある更新で削減されなければその値を増加するという手法を提案している。提案手法のリグレットの理論的上界を求めるとともに、実験的にも評価し種々の分類タスクで高い性能が得られることを示している。

第6章「Conclusion」(結論)は、本論文のまとめである。

以上を要するに、本論文は分類器のオンライン学習で求める重みベクトルにおいて無駄な特徴量を削減すると同時に重要な特徴を削減せずに残して高い分類性能を得る学習アルゴリズムを提案して、その有効性を示すことによって、数理情報学分野の技術発展に寄与した。

よって本論文は博士(情報理工学)の学位請求論文として合格と認められる。