審 査 の 結 果 の 要 旨

論 文 提 出 者 氏 名　　松　島　　慎

機械処理の対象となるデータの大規模化とそれを処理する計算機環境は大きく変化してきている。データ処理として機械学習を考えたとき、差し迫って問題となるのはデータが大規模化して全データをメモリに乗せきれない場合である。処理に必要なデータを得るためにその都度ディスクアクセスをすると、膨大な処理時間を要する。この状況を解決するために近年、メモリに乗る大きさにデータをブロックに分割して処理するサポートベクターマシンのアルゴリズムが提案されるようになっている。しかし、速度に関しては改善すべき点が多い。一方、大規模データの多様性も問題になっており、分類器の学習も２クラス分類だけでなく、分類クラスが多数存在する多クラス分類へ拡張する必要がある。メモリに乗りきらない大規模データを対象にする場合、１データ毎に分類器の学習を行うオンラインアルゴリズムの多クラス分類化が重要な課題となっている。

本論文は「A Study on Efficient Algorithms for Machine Learning from Large-scale Data」（大規模データからの機械学習のための効率的なアルゴリズムに関する研究）と題し、６章からなる。

第１章「Introduction」(序論)では、データ大規模化や計算機資源の能力など機械学習を取り巻く環境の変化を概観し、次にバッチ学習、オンライン学習という機械学習の分類を行っている。バッチ学習に関しては大規模データの分散処理が行われるとしても、単一計算機での処理を効率化する問題は依然として重要であることを指摘し、本論文で扱う問題として設定している。次に１データ毎に学習を行うオンライン学習に関しては、多様なデータを効率的に処理するアルゴリズム開発を本論文で扱う問題として設定している。さらに１データ毎の学習では、それ以前に用いたデータは使わず、その１データだけに限定するオンライン形式アルゴリズムを導入している。

第２章「Modeling Batch Learning」(バッチ学習のモデル化)では、バッチ学習に関して、精度とデータ量の関係を述べ、次に経験リスク最小化の枠組みとして定式化している。バッチ学習に対して第１章で導入したオンライン形式アルゴリズムを定式化し、処理対象のデータ量に対する計算量について述べている。

第３章「Modeling Online Learning」(オンライン学習のモデル化)では、オンライン学習に関して、モデル化と学習の効率を定義している。次にFollow-The-Regularized-Leaderアルゴリズムという一般性のある枠組みについて説明し、それをオンライン形式アルゴリズム化する例を示している。さらに、全データのある部分まで使って学習したときの学習結果と理想的な学習結果の差であるリグレット及び分類の失敗回数の上限を明らかにしている。第２章と第３章は、バッチ学習とオンライン学習という分類における教師あり学習に関する従来の知見をオンライン形式アルゴリズムという視点で系統的に整理した内容が記述されている。

第４章「Dual Cached Loops」(デュアル･キャッシュドループ)では、大規模データのストリーム処

理によるサポートベクターマシンのアルゴリズムを提案している。提案されたアルゴリズムは、従来提案されていた全データを複数ブロックに分割して順次メモリに読み込んで学習することを繰り返すアルゴリズムに似ている。しかし、読み込み単位がブロックであるという制限を外しており、ディスクからメモリへの読み込み処理とＣＰＵにおけるサポートベクターマシンの更新処理をスレッド単位で完全に非同期化した点が異なっている。これによって、大規模な実験データを対象にした実験では、既存アルゴリズムに比べて目的関数の同じ値を達成するまでの時間を１桁以上の短縮することに成功している。

第５章「Exact Passive-Aggressive Algorithms in Multiclass Classification」(多クラス分類における厳密なPassive-Aggressiveアルゴリズム)は、本論文のもう一つの貢献であるサポートクラス Passive-Aggressive アルゴリズムを提案している。標準的なオンライン学習手法である従来の Passive- Aggressive アルゴリズムでは多クラス分類は近似解であったが、ここでは入力データが対応するクラスの分類器を更新するとき、他のいくつかのクラスの分類器もこの入力データを誤認識しないように更新する厳密解を導出している。すなわち更新すべきクラスであるサポートクラスを求める閉じた形式のアルゴリズムを明らかにし、学習の効率化と分類精度の向上を実現している。

第６章「Conclusions」(結論)は、本論文のまとめである。

以上を要するに、本論文は機械学習における教師あり学習をバッチ学習、オンライン学習に大別し、その各々において新規性がありかつ従来手法より性能が向上した方法として、大規模データからの学習が可能なストリーム処理によるサポートベクターマシンと、オンライン多クラス分類を高効率かつ高精度で行えるサポートクラスPassive-Aggressiveアルゴリズムを提案し、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。