

氏 名	小山田 昌史		
学 位 の 種 類	博 士 (工 学)		
学 位 記 番 号	博 甲 第 8 5 2 4 号		
学位授与年月日	平成 3 0 年 3 月 2 3 日		
学位授与の要件	学位規則第 4 条第 1 項該当		
審 査 研 究 科	システム情報工学研究科		
学位論文題目	Efficient Machine Learning on Relational Data (関係データに対する機械学習の効率化に関する研究)		
主 査	筑波大学 教授	理学博士	北川 博之
副 査	筑波大学 教授	博士 (理学)	加藤 和彦
副 査	筑波大学 教授	博士 (工学)	天笠 俊之
副 査	筑波大学 教授	博士 (工学)	佐久間 淳
副 査	筑波大学 教授	博士 (工学)	佐藤 哲司
	(図書館情報メディア研究科)		

論 文 の 要 旨

本論文は、関係データを対象にした機械学習によるデータ分析処理の効率化に取り組んでいる。関係データに対するデータ分析処理は、多くの場合、(1) ETL (Extract, Transform, and Load) フェーズ、(2) 特徴量作成フェーズ、(3) モデル学習フェーズ、(4) モデル評価フェーズの 4 フェーズからなる。モデル評価フェーズにおいて満足ゆく結果が得られない場合、分析者は前のフェーズに戻り、作成する特徴量やパラメータを変更する。本論文は、そうした試行錯誤を含むデータ分析処理の効率化に向け、特徴量作成の高速化、特徴量作成の不要化、モデル学習・評価の高速化に関わる新たな手法を提案している。

特徴量作成の高速化に関しては、データを事前に集計しておくことで、データ分析に頻出する集計処理を高速化する技術を提案している。関係データからの特徴量の作成においては、データの最大値や標準偏差等の集約値を計算する処理がよく行われる。分析者は集約条件を変えながら何度もデータ集約処理を繰り返すことが多く、その高速化が望まれる。関連する既存技術としては、部分集約法が提案されているが、適切な粒度でグループを事前に作成しておくことが難しいという課題がある。本論文では、分析者が実際に行った集約対象データに応じてグループを動的に作成する APA 木 (Adaptive Partial-Aggregation Tree) を用いる手法を提案し、実験によりその有効性を示している。

特徴量作成の不要化に関しては、関係データとして記述されたエンティティ属性を高精度に推定する機械学習モデルを提案している。関係データのエンティティ属性の欠損値の予測は、顧客の属性情報の推定等、多くの応用を持つ重要なタスクである。従来、予測精度を高めるために、エンティティ

属性の予測の手がかりとなる購買履歴等の情報や機械学習モデルを用いて新たな特徴量を導出し、それを用いた予測を行う等のアプローチが取られてきた。しかし、試行錯誤が必要であったり、生成された特徴量の意味解釈が難しいといった問題があった。本論文では、特徴量を作成することなく購買履歴等の情報を活用する新たなモデルを提案している。提案モデルは、購買履歴等を用いてエンティティをクラスタリングし、各クラスターでエンティティ属性の予測モデルを構築する。これにより、高精度にエンティティ属性値を予測することが可能となることを示している。

モデル学習・評価の高速化に関しては、データ圧縮によるベクトル計算の高速化手法を提案している。ベクトル計算は機械学習処理等において頻発する計算であり、計算時間の削減や省メモリ化は重要な課題である。本論文では、ベクトル計算を疎密両ベクトルについて高速化・省メモリ化するフレームワークである CVS (Compressed Vector Set) を提案し、時間計算量と空間計算量を同時に改善する手法を示している。実験により、CVS は確率勾配法によるリッジ回帰等の実際の機械学習処理において、従来の方式である疎行列表現を上回る高速化・省メモリ化の効果を発揮することを示している。

審 査 の 要 旨

【批評】

本研究では、データ分析の実業務で対象となる関係データを対象にした機械学習によるデータ分析処理の効率化を目的に、(1) 集計処理の高速化、(2) 関係データのエンティティ属性を高精度に推定する機械学習モデル、(3) データ圧縮によるベクトル計算の高速化の3つのテーマに関して新たな手法を提案している。(1)については、分析者が実際に行った集約対象データに応じて動的にグループを作成するAPA木を用いる手法を提案し、データアクセスのためのI/Oが大幅に削減できることを実験により示している。(2)については、購買履歴等を用いたエンティティのクラスター分けとクラスター内のエンティティの属性予測モデルによる予測精度の向上を同時に最適化する新たなモデルを提案し、従来手法に比べて予測精度や予測モデルの解釈容易性が向上することを、実験的に示している。(3)については、ベクトルを次元方向に連長圧縮したままで基本的ベクトル演算を行う手法を提案すると共に、圧縮効率を上げるための多項式時間のヒューリスティックな手法や、量子化による圧縮率向上についても取り組んでいる。このように、実際のデータ分析処理業務を念頭に、複数の新たな手法を提案しその有効性を実験的にも示している点で、本研究は情報工学上の貢献が認められる。今後、実際のデータ分析処理の場面での提案手法の適用や評価が進むことが期待される。

【最終試験の結果】

平成30年2月8日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士(工学)の学位を受けるに十分な資格を有するものと認める。