

氏名(国籍)	つう 竹	つうい 翠(中国)
学位の種類	博士(工学)	
学位記番号	博甲第3964号	
学位授与年月日	平成18年3月24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	A Study on Example-Based Outlier Detection in Databases (例示に基づくデータベース中の外れ値検出に関する研究)	
主査	筑波大学教授	理学博士 北川博之
副査	筑波大学教授	理学博士 大保信夫
副査	筑波大学教授	工学博士 西原清一
副査	筑波大学助教授	博士(工学) 石川佳治
副査	筑波大学助教授	博士(工学) 福井和広

論文の内容の要旨

大規模データベースに対するデータ分析の一つとして、データ中の外れ値(例外値)の検出がある。外れ値は、通常のオブジェクトと異なる振る舞いを示すデータオブジェクトであり、クレジットカードやネットワーク利用における不正行為の検出、データからの興味ある知見の発見等、様々な利用が考えられる。これまでも、外れ値検出手法がいくつか提案されてきたが、データの性質に関する詳細な事前知識を必要としたり、手法を適用するためのパラメタ設定に試行錯誤が必要等の問題があった。

本論文は、これらの問題に対する一つのアプローチとして、利用者が提示する例示外れ値に基づく外れ値検出手法を提案し、実験評価により有効性の検証を行ったものである。

本論文では、まず、主に低次元のデータ集合を対象とする OBE (Outlier By Example) 手法を提案している。OBE では、オブジェクトが外れ値である程度を、当該オブジェクトから一定距離内の近傍にいくつのオブジェクトが存在するかを相対比較する MDEF (Multi-granularity Deviation Factor) を用いて測定し、各オブジェクトを特徴ベクトルに変換する。次に、例示外れ値オブジェクトを正例とする SVM (Support Vector Machine) を利用した学習手法により、例示外れ値に類似した外れ値の検出を行う。具体的手法としては、検出すべき外れ値の割合を手がかりとする QBE-Fraction 法と、利用者からのフィードバックを用いる OBE-Relevance Feedback 法の2つの手法を提案している。さらに、実データおよび人工データを用いた実験により、提案手法の有効性を検証している。

高次元データ集合では、距離や密度によるオブジェクトの特徴付けが困難となることが知られている。したがって、OBE を高次元データ集合に直接適用することは難しい。そこで、本論文では、例示外れ値オブジェクトが外れ値としての強い特性を示す低次元の部分空間を、遺伝的アルゴリズムを用いて探索し、発見された部分空間における外れ値を検出する手法である OBE/H 手法を提案している。OBE/H に関しては、利用者が与える外れ値が全て信頼できる場合を対象とする OBE/H-B 法と、与えられた例示外れ値にノイズが含まれる可能性を考慮した OBE/H-R 法を検討している。さらに、実データおよび人工データを用いた実験により、提案手法の有効性を検証している。

審査の結果の要旨

セキュアな情報利用への関心が高まる中、外れ値検出はますます重要性が増している研究課題である。本研究は、利用者が提示する例示外れ値を用いた外れ値検出という視点から、低次元データ集合と高次元データ集合の両者を対象とする独自性のある各種手法を提案し、実験評価によりその有効性を検証した点が高く評価できる。よって、本研究はデータ利用の高度化に関わる情報工学上の貢献が大きいと認められる。今後、より本格的な実用場面への適用や、手法のさらなる一般化等、残された研究課題への取組が期待される。

よって、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。