

氏 名 (本籍)	Salman Ahmed SHAIKH (パキスタン)		
学 位 の 種 類	博 士 (工 学)		
学 位 記 番 号	博 甲 第 6870 号		
学 位 授 与 年 月 日	平成26年 3月25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 Distance-based Outlier Detection on Uncertain Data (不確実データに対する距離に基づく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	博士 (工学)	山本 幹雄
副 査	筑波大学 教授	博士 (工学)	福井 和広
副 査	筑波大学 准教授	博士 (工学)	天笠 俊之
副 査	筑波大学 准教授	博士 (工学)	佐久間 淳

論 文 の 要 旨

情報技術の高度化に伴い、様々な実世界のデータが計算機システムで利活用されるようになってきた。それらの実データの中には、ノイズや観測誤差等を含む不確実データが多数存在する。したがって、実データから有用な知識を発見するデータマイニングにおいては、データの不確実性を考慮に入れた分析手法の開発が必要となっている。本論文では、データマイニングにおける重要な課題の一つである外れ値検出を取り上げ、不確実データに対する距離に基づく外れ値検出の手法を提案している。距離に基づく外れ値検出手法は、1998年に Knorr らによって提案されたもので、最も基本的な外れ値検出手法として知られているが、データの不確実性を考慮していない。本研究では、多次元データの各属性値が正規分布で表現される不確実性がある場合を想定し、距離に基づく外れ値検出をこのような不確実データに拡張したアプローチを示している。具体的には、本論文では次の3点について研究を行っている。

1) 近傍オブジェクト数の期待値に基づく外れ値検出手法：Knorr らの手法を不確実データに拡張するため、近傍オブジェクト数の代わりに近傍オブジェクト数の期待値を用いる外れ値の定義を示した。次に、この定義に基づく外れ値検出を効率的に行うための手法として、空間をメッシュ上に分割したセル構造を用いる手法を提案した。セル構造を用いることで、多くのオブジェクトに対する外れ値判定を、計算コストのかかる確率的な計算なしに行うことができる。

2) トップk外れ値検出手法：上記の発展として、近傍オブジェクト数の期待値を尺度として用いることで、外れ値度の高いトップk件のオブジェクトを効率的に検出する手法を提案した。提案手法はPC-List と呼ぶリスト構造を用いることで効率的な外れ値検出が可能である。

3) 不確実データストリームに対する連続的外れ値検出手法：観測データが正規分布の不確実性を有する時系列ストリームデータが多数存在する状況を仮定し、各スナップショットにおける外れ値を効

率良く検出する手法を提案した。提案手法は、上記の手法と同様にセル構造を利用する他、時系列ストリームデータの時間変化があまり急激でない場合は、直前のスナップショットにおける判定結果を利用した差分計算による効率化ができることが特徴である。

上記いずれの手法においても、人工データと実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。また、近似計算を用いた場合の精度や計算時間への影響等についても考察を行っている。

審 査 の 要 旨

【批評】

不確実データに対する外れ値検出は近年研究が盛んになりつつある。本研究は、多次元データの各属性値が正規分布の不確実性を有することを想定し、距離に基づく外れ値検出を拡張したアプローチを取っており、比較的一般性の高い状況を扱いつつも、これまでの他の研究にはない独自性が認められる。また、静的データに対する従来の手法のオーソドックスな拡張手法、トップk検出手法、ストリームに対する連続的外れ値検出手法と、同一のコンセプトのもとに体系的に研究を進捗させている点や、それぞれ実験により有効性を検証している点も評価できる。本研究は不確実データの外れ値検出に関して、複数の視点からの有効な手法を構築したことによる情報工学上の貢献が十分に認められる。今後、大規模実データへの適用、異なる不確実性への対応等、残された課題に対する検討が進められることが期待される。

【最終試験の結果】

平成 26 年 1 月 23 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。