



琉球大学学術リポジトリ

University of the Ryukyus Repository

Title	A Study on data integration based on its structure and characters(Review_審査要旨)
Author(s)	岡崎, 威生
Citation	
Issue Date	2014-09-10
URL	http://hdl.handle.net/20.500.12000/29679
Rights	

平成26年 8月 4日

琉球大学大学院
理工学研究科長 殿

論文審査委員

主査 氏名 和田 知久
副査 氏名 玉城 史朗
副査 氏名 名嘉村 盛和
副査 氏名 黒田 登美雄



学位（博士）論文審査及び学力確認終了報告書

学位（博士）の申請に対し、学位論文の審査及び学力確認を終了したので、下記のとおり報告します。

記

申請者	氏名 岡崎 威生		
現住所			
成績評価	学位論文 <input checked="" type="checkbox"/> 合格 <input type="checkbox"/> 不合格	学力確認 <input checked="" type="checkbox"/> 合格 <input type="checkbox"/> 不合格	
論文題目	A study on data integration based on its structure and characters (構造と特徴によるデータ再構成に関する研究)		
審査要旨（2000字以内）			
<p>データサイエンスは、数学、確率モデル、統計、パターン認識、機械学習、視覚化、データベース等の知識概念を統合した学問領域である。データ計算の進化と、スマートデバイス、Web、モバイル、ソーシャルメディア等の発達による大量データ生成が一般化を加速してきた。そのためデータマイニング技術がデータサイエンスの注目される潮流となっているが、データ分析の視点から見ると、仮説検証型から仮説発見型への変化だと見ることができる。言い換えると、小標本問題から大標本あるいは複合データ問題への転換となる。しかし、大量データは高信頼性と同義ではない。大量データに対峙し、</p>			

(裏面へ続く)

データマイニングにより仮説発見が試みられた時、その仮説に対する客観性や妥当性は検証されなければならず、検証においては適切な変数選択やデータ変換が重要となってくる。つまりデータ再構成となる。本研究では仮説検証での統計的解析精度向上のためのデータ再構成手続きとして、(1)データ併合、(2)外れ値検出、(3)データ標準化の3項目を提案している。

データ併合においては、正規母集団における母平均と母分散に対して、予備検定を適用したTest and Estimate型推定量を提案し、その統計的特性として確率密度関数、期待値、Mean Square Errorを導出し、不偏性を証明した。そして、従来の推定量評価基準とは異なるL1測度を用いたMean Absolute Errorを提案した。また、ノンパラメトリック統計量の中央値に対しても、同様のTE型推定量を提案し、統計的特性を導出した。予備検定を適用する際に必要となる有意水準を決定するために、事前情報を伴う相対リスク基準とミニマックスリグレット基準をとりあげ、MAEとMSEの比較と合せて最適有意水準について考察した。ミニマックスリグレット基準は保守的性質を持ち、MAEはMSEに比較して小さい最適有意水準を与えることを示した。

潜在構造モデルにおけるデータ併合問題として、因子分析モデルを取り上げ、観測変数補完を目的としたパラメータ推定方法を提案した。欠測値を含む因子分析の状況を考慮し、完全情報最尤推定値を得るためのEMアルゴリズムを適用した。非識別条件と欠測構造の視点から補完可能性について考察し、ベンチマークデータを利用した提案法の数値評価を行ない、有効性について確認した。

外れ値検出においては、非対称関係に対応できる2次元配置技術を提案した。対象データを対称部と歪部に分解しそれぞれを2次元配置した。対称部においては、対象間距離と配置の相関係数を目的関数として、最適化手法を提案し、従来最適化法のPowell法とBFGS法に対する頑健性での優位を数値実験で示した。歪部に対しては、歪みの方向と大きさの2軸で表現し、数値実験により外れ値検出効果を検証した。

データ標準化においては、ゲノム解析の基盤データであるDNAマイクロアレイデータに対して、その信頼性確立を目的とした標準化法を提案した。実験環境におけるバイアス発生要因とその修正法を体系化し、特定データの補正に依存しないよう各バイアス発生の有無を含めたモデル化を行った。モデル化の結果得られた候補モデルの中から、オーバーフィッティングに配慮してBICを適用したモデル選択法を提案した。酵母、大腸菌、ヒトのマイクロアレイデータに適用し、高い正規化結果が得られた。

データ獲得技術の更なる発展が予想され、得られたデータの解析戦略が重要となる中、本研究成果は現代データサイエンスあるいはデータ解析に対して高い信頼性を与えることが期待される。

したがって、本研究成果は工学的に有用であり、提出された学位論文は博士の学位論文に相当するものと判断し、学位論文の審査を合格とする。また、論文発表会における発表ならびに質疑応答において、申請者は専門分野および関連分野の十分な知識ならびに十分な研究能力を有していることが確認できたので最終試験を合格とする。さらに学力確認のための外国語筆記試験において優秀な成績を修めたので学力試験を合格とする。