審 査 の 結 果 の 要 旨

氏名　バールフット、ダニエル クレーリー

有限の観測データ集合からそれが従う真の確率分布を推定することは，広汎な統計的認識・学習手法の基盤である．本論文は，極めて複雑な結合分布となり得る実世界データに対して，分布関数の推定・改善と判別条件による分割を再帰的に繰り返すことで，階層的な統計モデルを自動的に構成する手法を提案している．結合条件に起因する組合せ爆発や条件の非独立性に起因する性能低下が少なく，大規模複雑データの処理に適している点が本手法の特徴である．本論文は 11 章からなり，以下の各章では基本概念の導入に始まり，手法の提示を経て，実データによる検証・評価を示した上で，学術的考察を加えている．

第 1 章 “Introduction”では，統計モデル推定理論とアルゴリズム情報理論とを関連付けつつ，本論文の基本概念と問題意識を呈示している．複雑な確率構造の推定について，モデルの推定・改善と判別条件による分割を再帰的に繰り返す基本的な考え方を呈示し，また，この種の問題における重要課題として，特徴選択問題と特徴結合問題を指摘し，その解決のための方針について論じている．

第 2 章 “Model Ensemble Update Method”では，統計モデル集合の逐次更新法の基礎を呈示している．観測データ集合 X とそれを生成する真の確率分布 P(x)に関して，所与の確率分布推定モデル Q(x)の累積分布関数 F(x)による X の確率積分変換 Y=F(X)が，Q(x)=P(x)のとき一様分布となることから，Y の非一様性（乱数欠陥）に基づき F(x)を修正することで Q(x)を P(x)に近づけられる．このとき，Y(PIT 値と呼ぶ)のヒストグラムを用いて F(x)の修正パラメタを得る方法を示すと共に，その場合は修正の効果である符合短縮量が実データやモデルによらず，PIT ヒストグラムの階数のみに依存することを示している．また，上述の方法が複数モデルの集合に対して適用可能なことを指摘している．

第 3 章 “PIT Value Analysis Algorithm”では，本論文の中核的提案である PIT 値解析アルゴリズム(PITA アルゴリズム)を呈示している．入力の各変量ごとにモデル分布と文脈関数を割り当てる．文脈関数とは 2 値の判別条件式であり，その値により入力データ集合を 2 個の部分集合に分割する．分割された各部分集合ごとに 2 章で示したモデル更新法を適用する．事前に用意した多数の文脈関数の中から，毎回，PIT ヒストグラムに基づいて最もモデル改善効果の大きいものを自動選択しつつ，再帰的に上述の分割とモデル更新を繰り返す．この手法の計算量についての検討を示し，特に大規模データへの適用に適していることを論じている．また，本手法が，任意の既存手法が与える推定モデルを初期モデルとしてさらなる改善を得るために使える可能性を指摘している．

第 4 章 “Alternative CDF Transformation Techniques”では，2 章で示した手法以外にモデル集合更新を実現する方法を複数呈示している．各方法は PIT 値の異なる統計に対応しており，各々の場合における修正の効果を表す符合短縮量の評価も与えている．

第 5 章 “Related Work”は，本論文で提案した PITA アルゴリズムを，機械学習や統計学における著名な既存手法と対比し関連性と相違について論じている．特に，AdaBoost をはじめとする boosting 手法，最大エントロピー法，および一連のモデル適合度評価を取り上げている．

第 6 章～10 章では，PITA アルゴリズムを様々な応用問題に適用し，検証と評価を行っている．適用対象は，画像圧縮，2 値パタン識別，英語の形態素モデル，音声モデル，運動認識である．劣化なしの画像圧縮への適用では，PITA アルゴリズムの正当性を確認すると共に，良好な圧縮率を達成した．2 値パタン識別は著名な機械学習用公開ベンチマークデータに適用し，AdaBoost と同等の性能を示した．形態素モデルへの適用では，非数値データに適用する際の問題点を明らかにし，その解決策を呈示した．音声モデルにおいては，本手法による複雑なモデルが，標準的従来手法である Laplacian，Gaussian モデルに比べ格段に良い記述能力を有することを示した．最後に，公開モーションキャプチャデータを用いた運動認識実験においては，PITA

アルゴリズムが，標準的従来手法である HMM と比較して，総合的に優れた認識能力を有することを示した．さらに，HMM に重畳して PITA を用いる方法を示し，それによりさらなる改善が得られることを示した．

第 11 章 “Conclusion”では，全体を総括し学術的考察と今後の展開について論じている．まず，PITA アルゴリズムが，乱数欠陥の探索を基盤とする統計モデル推定手法の一つであるとの位置づけを述べ，乱数欠陥の探索手法を変えることで，同じ基盤に立つ別の方法があり得ることを論じている．次に，PITA アルゴリズムの特長について，高性能の複雑なモデルを過学習に陥りにくく構成可能で大規模データに適していること，他手法で得たモデルの上に重畳可能であることなどを上げてまとめている．最後に，今後の展開について，脳機能モデル化への適用可能性等を含めて論じている．

以上これを要するに，本論文は実世界情報処理における重要課題である大規模複雑データの統計モデル推定に対し，アルゴリズム情報理論と統計モデル推定理論にまたがる独自の考察に基づく新手法を提案し、多様な代表的実データ処理問題に適用・評価してその正当性と有効性を示している．また，任意の既存手法に重畳して性能改善を図る方法を呈示することで有効応用範囲を広げている．

以上の理由から，本論文は知能機械情報学上貢献するところ少なくない．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．