

審査の結果の要旨

論文題目 : Nonparametric Regression for Complex Data

題目和訳 : 複雑データのためのノンパラメトリック回帰

氏名 : 今泉 允聡

論文の内容

本論文は、複雑データをノンパラメトリック統計を用いて分析する様々な方法を提案している。複雑データは医学分析・財務データ分析などの様々な応用分野で使用されているが、複雑データを効率的に分析するための方法論には、依然として未解決問題が多く存在している。本論文では、複雑データの中で、テンソルデータ・関数データ・構造データの分析に焦点を当て、ノンパラメトリック統計を応用した分析手法を提案している。ノンパラメトリック統計はモデルの特定化の誤りに起因するバイアスを減らすことができるため、観測値が数多く確保できる状況下では、非常に有効なアプローチである。しかし、複雑データは非常に複雑な構造を持つため、従来のノンパラメトリック統計の手法をそのまま適用すると、推定量は望ましくない挙動を示すことが知られている。本論文は、分析対象のデータが持つ特徴（単純構造・滑かさ・経済的構造）に着目し、その特徴を組み入れた推定量を提唱し、提唱した推定量が優れた理論的特性を持つことを証明している。さらに、数値実験により、提唱した推定量が実験においても優れた特性を示すことを確認している。

第1章は、ノンパラメトリック回帰について概観している。第2章は、高次元テンソルのノンパラメトリック回帰モデルの新しい推定方法を提唱している。従来のノンパラメトリック回帰は、高次元テンソル（＝高次元データ）を想定していないため、高次元テンソルに対しては適用不可能である。第2章は、高次元テンソルのノンパラメトリック関数が、低次元データのノンパラメトリック関数のテンソルとして表せると仮定し、高次元テンソルのノンパラメトリック関数を単純な局所関数に分解する。この分解は、既存のノンパラメトリック回帰における、加法的ノンパラメトリックモデルに相当する。そして、局所関数を事前分布にガウス過程を用いたベイズ推定により推定する。第2章で提唱された回帰手法は、特定の条件下では、単純なノンパラメトリック・アプローチを大幅に上回る収束率を持つことが示された。数値実験の結果は、提唱された手法が、複雑なネットワークの要約統計量を高精度で予測できることを示唆している。

第3章は、テンソルの滑らかさを組み込んだ新しいテンソル分解法を提唱している。現実の世界のテンソルデータは、しばしば、隣接する要素が類似している、もしくは隣接する要素が滑らかに変化するという特徴を持っている。本章では、この滑らかな性質を組み込んだ平滑化タッカー分解 (STD) を提案している。一般的に、滑らかな関数は少数の基底関数によって近似することが可能である。STDは、この原則をテンソルに応用し、滑らかなテンソルデータを、少数の係数テンソルに変換することにより、少数の基底関数によってモデルする。STDの目的は凸問題として定式化されるため、その解決のために、alternating direction method of multipliersに基づくアルゴリズムが導出された。テンソルの平滑さの仮定が満たされれば、STDは従来の方法よりも小さい誤差に達することが示された。さらに、数値計算によりSTDの理論的性質とその性能が検証された。

第4章は、予測変数 X と応答変数 Y の両方がランダム関数である回帰モデルを分析している。ここでは、時点 s における応答変数 Y の条件付き期待値 $E[Y|X](s)$ が予測変数 X の線形関数によって与えられる関数線形モデルを考える。推定の対象は、条件付き期待値演算子の積分カーネル $b(s, t)$ である。本章では、関数主成分分析 (PCA) に基づいた推定量を提唱し、単一打ち切りに基づく推定量と、二重打ち切りに基づく推定量の二つの推定量の収束速度を導出している。正則条件の下で、単一打ち切りに基づく推定量の収束速度は積分カーネル $b(s, t)$ の平滑さと共分散演算子の固有値の減衰率によって特徴付けられ、その速度はミニマックス最適であることが示された。さらに、二重打ち切りに基づく推定量が最適速度を達成する条件も導出された。シミュレーションの結果は、本章が提唱する推定量が既存の推定量よりも優れた性能を持つことを示唆している。

第5章は、状態空間が大きい離散マルコフ選択モデルを推定するための新しい近似手法を提案している。離散マルコフ選択モデルは、経済主体の動学的意思決定をデータから分析することが可能なモデルであり、多くの実証研究で利用されている。しかしながら、離散マルコフ選択モデルを実際に推定する際には、膨大な計算負荷が存在することが知られている。計算コストは、モデルの状態変数の数に対して指数関数的な速度で増加し、この現象は「次元の呪い」と呼ばれている。本章は、大きな状態空間を含む離散マルコフ選択モデルを低い計算コストで解くことができる statistical least square temporal difference method (SLSTD) を提案している。SLSTDは、関数近似と確率的近似の2つの近似技法を組み合わせることにより、離散マルコフ選択モデルが使用するベルマン方程式を高速に解くことを可能にして、計算負荷を大幅に軽減している。シミュレーションでは、SLSTDは他の既存の方法よりも高速かつ正確に実行され、計算時間を短縮することが示された。さらに、SLSTDによるパラメーターの推定量は、一致性と漸近正規性を持つことも証明されている。

論文の評価

本論文は、複雑データ・高次元データに対する新しい分析手法の提案を主要な貢献としている。その際には、データの単純構造（2章）、滑らかさ（3章）、経済的構造（5章）などに着目して、データの特徴を利用した新しい推定法を提案し、一定の条件の下で、提案された手法が既存の手法を上回る性能を示すことを理論的ならびに数値計算で示している。複雑データ・高次元データの統計的分析は、近年活発な研究が行われている分野である。これらのデータに、既存の手法を単純に適用した場合には、得られた推定量はバイアスや分散の面で望ましくない特性を持つことが知られているため、データの特徴を生かした推定法の発見・分析は大いに必要とされている。以下では、各章の主貢献を議論する。

第2章は、テンソルの非線形関数を推定する新しい手法を提案している。提案された手法は、高次元テンソルの非線形関数が、低次元データの非線形関数のテンソルとして表されるならば、その表現に基づいた推定量は優れた理論的特性を持つことを示している。本章で仮定されている構造は、ノンパラメトリックにおける加法性に相当するものである。ここで仮定されている構造が、実際のデータでどの程度（近似的にも）満たされるかは未知であるが、一定の理論的貢献は果たしている。

第3章は、実際のテンソルデータの多くは滑らかな値（近接する要素の値が類似している）をとることに着目し、滑らかなテンソルを少数の基底関数によって近似する手法を提唱している。滑らかな関数の基底関数による近似は、ノンパラメトリック統計学において論文の蓄積がなされてきた分野であるが、テンソルデータへの応用はYokota et al. (2015, Signal Processing) など少数に限られ、Yokota et al. (2015)も理論的性質の分析は行っていない。本章は、テンソルデータを基底関数で近似した推定量の理論的性質を明らかにすることで、テンソル分析の理解を促進する貢献をしている。

第4章が分析しているランダム関数回帰モデルは、近年、数多くの論文が発表されてきた。しかしながら、積分カーネル $b(s, t)$ の推定量の最適収束速度を導出した論文はこれまで存在しなかった。本論文はこの重要な空白を埋め、関数データ分析の理解を促進する貢献をする。さらに、最適収束速度を実際に達成できる推定量の提唱により、実用上も意義のある貢献をしている。

第5章は、離散マルコフ選択モデルの推定における計算量の削減方法を提案している。離散マルコフ選択モデルは、経済主体の行動原理を計量経済モデルの中に明示的に組み入れることにより、政策変化の評価を経済主体の行動の変化を踏まえて分析することを可能にする。そのため、離散マルコフ選択モデルは、政策分析のための有用なツールとして、幅広く応用されており、その計算コストの削減は、計量経済学の実践上の重要な問題である。第5章は、この問題の解決法を提案すること

で、大きな貢献を果たしている。さらに、第5章で提案された手法は、分野を問わず適用できるため、労働経済学・開発経済学・産業組織論といった経済学の広範な分野に貢献することが可能である。

論文審査の結論

以上の評価では、今泉氏の提出論文に対して一致して高い評価が与えられた。提出された論文の第2章は国際的学術雑誌（Journal of Machine Learning Research W & CP）にすでに掲載され、第3章はレフェリー付きの国際学会（NIPS TensorLearn Workshop, 2016）ですでに発表されている。また、第4章・第5章も国際的学術雑誌に投稿される予定である。このように、本論文の水準は、本研究科が要求する学位論文としては十分であり、審査委員会は全員一致で本論文を博士（経済学）の学位授与に値するものであると判断した。

2017年2月
審査委員
下津克己
久保川達也
大森裕浩
加藤賢悟
入江薫