

審査の結果の要旨

氏 名 渡邊 千紘

機械学習および統計的データ解析において、異なる事物間の関係性を用いた知識発見は重要な問題である。事物間の関係性を記述する関係データはこれまで多くの研究の解析対象となり、関係データ解析のための手法が数多く提案されてきた。しかし、この種の解析手法においてモデルや推定結果の信頼性を定量的に評価する方法論はまだまだ発展途上であり、特に数理的な研究が望まれている状況である。本論文では、関係データからパターンを抽出する統計解析に関して、柔軟なモデル評価の方法論を打ち立てることを目標とし、特に共クラスタ構造の検定および深層学習による構造の抽出方法の提案を行っている。各提案手法には、理論解析による正当性の保証、もしくは数値実験による有効性の検証が与えられている。

本論文は「Statistical Analysis for Pattern Extraction from Relational Data Matrix」（関係データ行列のパターン抽出に関する統計解析）と題し6章からなる。論文は大きく分けて三つのパートに分かれており、それぞれ「潜在ブロックモデルの共クラスタ数の検定」「関係データ行列の共クラスタ構造に関する選択的推論」「深層学習による行列並べ替え」に主眼を置いている。

第1章「Introduction」では、関係データ解析に関する研究背景のまとめと、本論文の構成および主結果の概要を与えている。

第2章「Preliminaries」では、本論文で扱う関係データ解析の問題設定および必要な数学的道具の準備を与えている。

第3章「Statistical test on the number of biclusters in a latent block model」では、潜在ブロックモデルを用いた共クラスタリングにおけるブロック数に関する検定手法を提案している。共クラスタリングのブロック数検定はこれまで確率的ブロックモデルに対しては提案されていたが、行と列が異なるオブジェクトに対応する潜在ブロックモデルに対しては検定手法が提案されていなかった。本章では、残差行列の特異値を用いた検定統計量を考察し、それが行列サイズ無限大の極限でTracy-Widom分布に分布収束することを示した。これは、ランダム行列理論の結果および非局在化と呼ばれる性質を用いて示されている。漸近分布は残差が劣指数的な分布であれば十分であり、正規分布に従うという仮定はいらない。さらに、この結果は、行列サイズの増加とともにク

クラスタ数が増大する場合にも拡張できることが示されている。実験的にも提案手法が理論解析通りに振る舞い、漸近分布を用いた検定が実用的にも使えることを検証している。

第4章「Statistical test on the estimated bicluster structure of a relational data matrix」では、共クラスタリングにおけるクラスタ構造の検定手法はクラスタリングに用いたデータを再利用していることから、バイアスを生じさせているという問題を扱い、この問題を選択的推論の枠組みで解決する手法を提案している。そのため、残差の分布はガウス分布であると仮定し、クラスタリングアルゴリズムとして二乗残差を最小にするものを考えることで、クラスタリング結果で条件付けた検定統計量が切断カイ分布に従うことを導出している。この理論結果の正当性を確認すべく、数値実験を通じた検証によって実際にバイアスが除去されていることが数値的にも確認されている。

第5章「Matrix reordering method for capturing flexible structural patterns in a relational data matrix」では、与えられた関係データの行と列を並べ替えて意味のある構造を抽出するための深層学習を用いた手法を提案している。この手法では、深層ニューラルネットワークを用いて行および列を一次元特徴量にエンコードし、その一次元特徴量を昇順または降順に並べ替えることで行列の並べ替えを実現する。また、エンコードされた特徴量から深層ニューラルネットワークを用いて行列値をデコードすることで行列の雑音除去も実行できる。本手法は行列の潜在的な構造のタイプによらず利用することができ、共クラスタ構造だけでなく他の様々な構造に対して利用できるといった利点がある。本手法によって実際に構造を抽出できることをいくつかの人工データを用いて実証し、また実データへ適用することにより実データ解析における有用性も検証している。

第6章「Conclusion」では、本論文のまとめおよび各種拡張に関して議論し、今後の課題を述べている。

以上を要するに、本論文では関係データ行列からパターンを抽出する統計解析手法に関して、そのモデルを評価する検定手法および行列並べ替えの新しい手法を提案した。これらは、クラスタ数が固定されていない共クラスタリング、および特徴量の抽出形式が固定されていない行列の並べ替えを可能にし、関係データ解析の可能性を押し広げるものである。また、提案手法は理論解析だけでなく、十分な数値実験によってその有効性が検証されている。共クラスタリングという応用上有用な問題への理論保証付きの検定手法の提供は当該分野において重要な貢献であり、行列並べ替えに対する深層学習を用いた新しい手法の提案は関連研究に対するインパクトも大きい。関係データという機械学習および統計解析における重要な問題において、本論文は理論と実用の両面から意義のある貢献を果たしており、数理情報学の発展に寄与している。

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。