

審査の結果の要旨

氏 名 丁 弘 毅

本論文は、機械学習分野における点過程時系列データの解析のための統計的モデリングについて、ベイズ統計的なアプローチを試みている。点過程時系列データとは、繰り返し起こるイベントについて、観測された時刻を記録したデータであり、病気の症状、市場における売買、地震など医学、経済学、気象学などさまざまな分野において観測される。このような点過程時系列データは、イベントが起こった時刻を正確に同定できているケース（再発イベントデータ）と、重複のない時間区間の中でイベントが起こったカウントのみがデータとして得られるケース（パネルカウントデータ）に大別される。後者においては、例えば、ウイルス感染による嘔吐が診療と診療の間に起こった回数のようなデータとなる。このような追加提示系列データのモデリングが可能となれば、将来のイベントの予測が可能となり、イベントの背景にあるメカニズムの解明にも繋がるため有用である。前者の再発イベントデータにおいては、各サンプルの再発イベントデータを確率過程と見なし、データの従う確率的インテンシティ関数について、その平均インテンシティ関数をガウシアンプロセスにより生成し推定する方法など、さまざまな情報解析技術がこれまでに研究されていた。ベイズ的なアプローチは、データ解析においてさまざまな事前情報やモデルの構造を推論に付加することができる柔軟な枠組みであり、この再発イベントデータについても、変分ベイズ法を用いた方法は、高い精度を示し、付加的な離散化の処理なども不要であることから広く用いられている。一方、パネルカウントデータについて、平均インテンシティ関数の推測論としては、最尤原理に基づく研究のみがなされ、ベイズ的なアプローチは取られていなかった。本論文では、まず、複数のサンプルに対して得られたパネルカウントデータに対して平均インテンシティ関数を推定するベイズモデルの枠組みを与え、その推定方法として変分ベイズに基づく手続きを導出している。更に、新たなパラメータを付与することで、平均インテンシティ関数ではなく各サンプルにおける確率的インテンシティ関数を推定する方法論を構築した。この方法は、例えば、あるウイルス感染に対する複数の処置の比較などに用いることができる。更に、複数のサンプルの確率的インテンシティ関数が、小数の確率的インテンシティ関数の混合になっている場合について、再発イベントデータからこの混合インテンシティ関数を推定する方法を開発している。このとき、混合するインテンシティ関数の数が重要となる。本論文では、スティックブレイキングプロセスを用

いることで、混合数の推定を一つのパラメータ値の決定問題に置き換え、再発イベントデータの生成された背景となる複雑なメカニズムの同定問題に対する解析手法の開発にも成功している。

本論文は五つの章からなる。第一章では、機械学習分野における学習理論について、特にベイズ統計的なアプローチを説明し、点過程時系列データを実際の例を示しながら再発イベントデータ、パネルカウントデータについて述べている。第二章では、点過程時系列データ解析へのアプローチとして、確率過程、特にガウシアンプロセス、スティックブレッキングプロセスについて導入し、再発イベントデータ、パネルカウントデータにおける確率的インテンシティ関数について説明している。第三章では、パネルカウントデータにおけるベイズ的モデリングのフレームワークを定義し、その推測論について論じている。また、平均インテンシティ関数を各サンプルの確率的インテンシティ関数に拡張する方法論について説明している。性能の評価としては、シミュレーションデータ、および膀胱癌の再発に関するデータを用いて、開発した方法論の精度を検証している。第四章では、混合インテンシティ関数の推定手法についてその理論的枠組みを示し、シミュレーションデータ、および web データ、引用データを用いて性能の高さを示すことに成功している。

本論文は、機械学習分野における点過程時系列データ解析の基盤の一つとなるベイズモデルをパネルカウントデータに対して初めて構築し、平均インテンシティ関数から各サンプルのインテンシティ関数への拡張を行い、再発イベントデータの背景にある混合インテンシティ関数のメカニズムに迫るための拡張まで行ったものであり、今後の点過程時系列モデリングの研究への指針を示すという点で大きな貢献をなすものであり評価に値する。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。