

審査の結果の要旨

氏名 福島 真太郎

時系列データからその変化を検知する技術は、データサイエンス上重要な技術である。例えば、センサーデータ時系列の変化を検知することによりシステムの障害を早期検知する可能性がある。また、行動データの変化を検知することにより市場の動きを把握できる。このような変化検知技術において、従来では、変化の時点を同定することが主な目的とされていた。しかしながら、変化の起こり方に関する様態を詳しく解析することは知識発見上重要なテーマであるが未発達であった。例えば、変化が起こる時空間パターンや、変化の持続性や、変化の原因を解析することは、データの動きをより深く解析する上で重要な問題であるが、従来の研究ではそこまで踏み込んで変化を解析する研究が発達していなかった。本学位論文は、上記の問題を解決するために、「階層的变化検知」という概念を導入し、階層的に変化をより深く読み解くための新しい解析手法を与えた。階層的变化検知としては1)「メタ変化検知」、2)「変化持続時間検知」、3)「潜在変数モデルの階層的变化検知」という3種類の問題を設定した。1)は変化の起こり方の変化といったメタ変化の階層を扱い、2)は変化の起こり方が持続的か否かの階層を扱い、3)は変化の原因についての階層を扱っている。本論文ではこれらの問題に対して統一的な情報理論的アプローチを展開している。特に、1)、3)の問題に対してはMinimum Description Length(MDL)原理に基づいて、記述長という観点から変化を定量化し、階層ごとの変化量の分解や階層間の変化量の統合を実現した。2)の問題に対してはオンライン学習の立場から持続的变化と非持続的变化の判定をする方法論を与えた。このようなアプローチから、上記3つの問題に対して、具体的なアルゴリズムを構築し、理論的考察と実験的検証を行うことにより、変化検知の新しい体系を築いた。

本論文は「An Information-Theoretic Approach to Hierarchical Change Detection」(階層的变化検知への情報理論的アプローチ)と題し、6章からなる。

第1章「Introduction」(序)では、階層的变化検知の問題を提起し、1)「メタ変化検知」、2)「変化持続時間検知」、3)「潜在変数モデルの階層的变化検知」という3つの問題を設定し、その基本的課題と貢献のポイントを示している。

第2章「Preliminaries」(準備)では、本論文を貫く情報論的アプローチの要である、MDL原理とオンライン学習の基本概念をまとめている。

第3章「Detecting Metachanges in Data Streams from a Viewpoint of MDL」(MDLの観点からのデータストリームにおけるメタ変化検知)では、変化が起こる様態の変化

として「メタ変化検知」を扱っている。メタ変化には、変化の発生間隔が変化する「時間方向のメタ変化」と変化の大きさが変化する「状態方向のメタ変化」があるとしている。いずれも変化の起こり方のパターンを学習して、これに基づいて変化の度合いを記述長の観点から定量化し、これがある基準値を超えるとそれぞれの変化を検知する手法を提案した。また、MDL原理の立場から時間方向と状態方向のメタ変化を統合して全体の変化を検知する手法を与えた。上記手法について人工データを用いて評価したところ、変化検知の早さと誤警報率のAUC(Area under curve)を評価基準として、既存の変化点検知手法よりも有意に優れていること示した。また、人間行動の加速度データや工場での製造過程データを用いて、既存手法では見逃された変化を検知できることを実証した。

第4章「Online Robust and Adaptive Learning from Data Streams」(データストリームからのオンラインロバスト適応学習)では、「変化持続時間検知」を扱っている。外れ値を含み、変化を伴う時系列データからの学習問題を考えるとき、外れ値にロバストであろうとすると変化を見逃すことになり、外れ値に過敏になると変化を検知できるが、学習がロバストでなくなるといった、学習のロバスト性と適応性にトレードオフの関係がある。このトレードオフを勘案しながらどこまで過敏に外れ値を検知するかといった「変化持続時間検知問題」をオンライン学習の枠組みを用いて定量的に評価した。そこでは、期待累積予測損失が最小になるように、ロバスト性と適応性の最適なバランスを実現するようなオンラインアルゴリズムを提案している。その収束性を理論的に評価するとともに人工データ及び核磁気共鳴の実データを用いて有効性を実証している。

第5章「Hierarchical Change Detection for Latent Variable Models」(潜在変数モデルに対する階層的变化検知)では、潜在変数モデルでデータをモデル化する場合に変化はA)パラメータレベル、B)潜在変数レベル、C)構造(潜在変数の数)レベルのうちどこから生じているかを同定する問題を扱っている。例えば、ネットワークデータを対象にして確率ブロックモデルのような潜在変数モデルを当てはめる際に、変化はブロック内のパラメータの変化によるのか、ブロックの構造の変化によるのかを同定することは変化を説明する上で重要である。ここでは、潜在変数モデルに対するMDL変化統計量と分解型正規化最尤符号長といった既存の手法を組み合わせたアルゴリズムを開発することにより、変化の検知のみならず、その要因を分析することを可能にした。本アルゴリズムを人工データ及び人口移動データ、Enronデータといった実データを用いて有効性を検証し、変化説明可能性の新しい方法論を提供した。

第6章「Conclusion」(結論)では全体を総括し、将来の展望を与えている。

以上を要するに、本論文は、変化検知といったデータサイエンス上重要な問題に対して、階層的变化検知という新しい概念を提示して、統一的な情報理論的アプローチを体系的に提示しており、数理情報学の発展に大きく寄与している。

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。