

(続紙 1)

京都大学	博士 (情報学)	氏名	近原 鷹一
論文題目	Causal Inference for Scientific Discoveries and Fairness-Aware Machine Learning (科学的発見と公平な機械学習を志向した因果推論)		
(論文内容の要旨)			
<p>本論文は、科学の様々な分野におけるデータ駆動型発見や、人工知能の社会的受容のための公平性の担保の実現へ向けた、統計的因果推論と機械学習の立場からの種々のアプローチについての研究結果をまとめたものであり、全6章から構成されている。</p> <p>第1章は序論であり、本論文の目的とその内容を概観している。科学における未知の事象について、観察データをもとに新たな知見を発見するためには、その背後にある生成メカニズムを正しくとらえることが重要である。そのためには、主に2つのタスク、すなわち因果関係の発見と因果効果の推定が重要な課題となる。一方、近年では機械学習をはじめとする人工知能技術の様々な実世界利用が検討されているが、これらが社会に受け入れられ、重要な場面で用いられるようになるためには、公平性などの社会受容を促す性質を担保する必要がある。本論文では、科学的発見を加速し、人工知能の判断の公平性を担保するという2つの目的に対して、因果推論の立場からこれを実現することを目指している。</p> <p>第2章では、本論文の中心的な技術背景として、因果推論の基本的な考え方や手法についてまとめている。まず、グレンジャー因果を中心に時系列データにおける因果関係の定義とその推定方法について紹介したのち、統計的因果推論の一般的枠組みである潜在アウトカムの枠組みにおける、データからの処置効果推定の考え方、さらに、因果関係の構造をより厳密に議論するための構造方程式モデルについて簡潔にまとめている。</p> <p>第3章では、時系列データから因果関係を発見するための教師付き学習アプローチを提案している。2つの時系列の間に因果関係があるか、あるとすればその向きを判定する問題は、因果推論の基本問題のひとつであるが、従来法であるグレンジャー因果に基づく手法は、対象の事前知識を用いて基礎となるモデルを適切に設計する必要があった。本論文では、すでに因果関係の有無・方向が分かっているデータに基づき、因果推論の問題を教師付き学習として新たに定式化し、Maximum Mean Discrepancy (MMD)を用いた分布比較に基づく特徴量を用いることによって、人工データと実データの両方において、従来手法よりも良い判定精度を実現している。</p> <p>第4章では、個別の対象に対する処置効果の違いが、対象のどの特徴によって生じるかを発見するための特徴選択問題を取り扱っている。この問題は、医療・教育・経済など様々な分野における発見や意思決定を助ける重要な課題であるが、従来は処置効果の平均値にのみ着目していたため、重要な特徴を正しく発見することができなかった。本論文では、処置効果の分布を考慮した特徴重要度尺度をMMDによって定義し、処置効果の分布に寄与する特徴を統計的検定によって発見する方法を提案することで、個別の因果効果とその原因を特定することを実現している。人工データと実データを用いた実験では、提案手法が従来手法より高精度で、処置効果に影響を与える</p>			

重要な特徴を選択できることを示している。

第5章では、精度と公平性のバランスを取った予測を行う新たな機械学習法を提案している。近年、人工知能の実利用が進んでいるが、過去のデータにもとづく差別的な判断（たとえば雇用に関する判断）を、予測器が引き継いで判断を行ってしまうという問題が浮上している。この問題に対して、人種などの差別的判断に利用される、いわゆるセンシティブ属性が、予測に影響しないようにする方法が提案されているが、従来手法では、センシティブ属性を予測対象と無関係にするという、やや保守的な戦略によって、必要以上に予測精度を損なうという問題があった。本論文では、たとえば業務に必須の身体的能力などのように、センシティブ属性とは関連するが、予測結果への影響が公平であると広く社会に受け入れられている情報を活用するための新たな公平性規準PIUと、その上界を制約として課しながら最適化する手法を提案することによって、従来手法と比較して、予測精度の低下を回避することに成功している。

第6章は結論であり、本論文で得られた成果を要約している。即ち本論文は、科学的発見を加速し、機械学習の予測の信頼性を向上することを目指して、時系列データからの因果関係の推定精度を高める教師付き学習アプローチ、処置効果に影響を与える対象の特徴を検出し解釈するための特徴選択アプローチ、予測精度と公平性を両立する予測器を学習するアプローチを提案し、その有効性を理論的に、かつ人工データと実データを用いた検証によって示したものである。本論文では最後に、将来の課題・展望として、本論文で提案した個別の問題に対する各々の手法を俯瞰したうえで、統計や機械学習などの分野において著しく発展している様々な技術を取り込むことによって、因果推論の方法論をさらに汎用的に実用的に発展させることを挙げ、本論文を結んでいる。

(続紙 2)

(論文審査の結果の要旨)

本論文は、科学におけるデータ駆動型発見や、人工知能システムの公平性の担保の実現へ向けた、統計的因果推論と機械学習の立場からの種々のアプローチについての研究に取り組んだものであり、得られた主な成果は次の通りである。

1. 時系列データからの因果関係発見のための教師付き学習アプローチを提案した。提案手法では、すでに因果関係の有無や向きが分かっている2つの時系列を訓練データとして、因果関係を判定する問題を教師付き学習の問題として定式化し、ある種の分布間距離を用いた分布比較に基づく特徴量を用いることでこれを実現している。人工データと実データを用いた実験では、従来手法と比較して提案手法がより高い判定精度を実現できることが示された。

2. 個別の対象に対する処置効果の違いが、対象のどの特徴によって生じるかを発見するための特徴選択を、処置効果の分布を考慮した特徴重要度尺度に基づいて行う手法を提案した。提案手法では、平均的な効果のみに着目する従来手法とは異なり、効果の分布に着目し、これに寄与する特徴を統計的検定によって発見することで、個別の処置効果とその原因を特定することを実現している。人工データと実データを用いた実験では、提案手法が従来手法より高精度で処置効果に影響を与える重要な特徴を選択できることが示された。

3. 精度と公平性を両立する予測を行う新たな機械学習法を提案した。従来手法では、公平性に関わるセンシティブ属性を予測対象と無関係にすることで必要以上に予測精度を損なうという問題があったのに対し、提案手法は、センシティブ属性とは関連するが、予測対象に直接な影響はもたない関係を区別した新たな公平性規準を制約として最適化を行うことで、予測精度の不要な低下を回避している。人工データと実データを用いた実験では、従来手法と比較して、提案手法が精度と公平性をともに維持できることが示された。

以上、本論文は、科学的発見を加速し、機械学習の予測の信頼性を向上することを目指して、因果推論と機械学習の手法を融合させた手法を提案したものであり、学術上・実応用上寄与するところが少なくない。よって、本論文は博士(情報学)の学位論文として価値あるものと認める。また、令和4年8月29日に実施した論文内容とそれに関連した口頭試問の結果、合格と認めた。なお、本論文は、京都大学学位規程第14条第2項に該当するものと判断し、公表に際しては、当面の間、当該論文の全文に代えてその内容を要約したものとすることを認める。

要旨公開可能日： 年 月 日以降