

京都大学	博士 (工学)	氏名	李 佳睿
論文題目	Incorporating Domain Experts' Knowledge into Machine Learning for Enhancing Reliability to Human Users (領域専門家の知識活用によるユーザへの親和性を重視した機械学習)		
(論文内容の要旨)			
<p>人工知能における機械学習 (ML) は、工学分野の実践で広く使用されている。機械学習モデルの有効性評価の基準としては、第一には入力に対する出力の高精度性が求められる。さらにモデル学習のためのデータ量の爆発的な増加に伴い、学習に要する計算時間も重要な基準となる。しかしこれらの評価基準はベースラインにすぎず、機械学習モデルの出力を人間が有効に活用できるための基準としてより重要になるのは、機械学習モデルの学習した内容が人間にとって親和性の高いものであることである。例えば、AI ロボットが機械学習モデルに基づいて判断を下す場合に、そのユーザとなる人間にとっては、その判断結果の正しさや学習時間が短いことに加えて、なぜそのように判断したのかについての理由や根拠を示せること (reliability) が求められる。すなわち機械学習モデルに求められる説明可能性 (explainability) の基準であり、これを充足することで、人間にとって透明性、解釈容易性、可読性の高い信頼できる機械学習モデルになる。本論文では説明可能性を有する機械学習として二種類の領域専門家の知識を導入することを提案している。一つは当該領域における判別問題の結果に関して専門家が経験的に獲得している知識であり、もう一つは当該領域において収集されるデータの生成過程として背後に存在している潜在変数間の因果関係に係る知識である。本論文は、これら二種類の領域専門家の知識を活用してユーザへの親和性を重視する機械学習モデルについて提案を行うもので、以下の全 6 章から構成される。</p> <p>第 1 章は緒言であり、機械学習モデルの有効性評価の基準として、高精度と高効率だけでなく、ユーザが有効に活用できるための基準として、機械学習モデルの学習した内容が人間にとって親和性の高いものでなければならないことについて述べている。そのための基準として、機械学習モデルが判断する理由や根拠を示せること、すなわち説明可能性の基準についてまとめている。さらに説明可能性を有する機械学習モデルの実現のためには、領域専門家が判別に係る先験的知識、あるいは潜在変数間の因果関係に係る知識の二種類の知識が必要になることについてまとめている。</p> <p>第 2 章では、説明可能な機械学習モデルを作成するために、領域に固有な二種類の専門知識を組み込む方法について示している。ここでは、当該領域で経験的に知られている判別に係る先験的知識が、機械学習を効率的に進めることができ人間にとっての可読性を高めるために有効であることを述べている。さらに領域専門家の推論ロジックを表現した因果関係知識は、観測できるデータの背後にあるデータの生成過程を表すものであることから、機械学習の進行を人間にとって親和性の高いものにすることができるとしている。さらに因果関係知識は、既存の学習済みモデルを出発点として新しく与えた学習対象のデータを学習させる転移学習にも有効で、教師信号となるラベル付きデータが少数しか得られていない異なる系における学習の性能向上に寄与できるとしている。</p> <p>つづく第 3 章から第 5 章では、領域専門家の知識を活用した機械学習モデルとして、改良型 Fuzzy c-Means 法 (CSCD-FCM)、構造方程式モデリングに基づく説明可能な機械学習モデル (SEM-EML)、リレーショナル転移学習アルゴリズム (RF-TL) の提案を行い、その有効性を数値実験の結果とともに示している。</p> <p>まず第 3 章で提案する改良型 Fuzzy c-Means 法 (CSCD-FCM) では、領域専門家が有する分類結果の事前分布に係る先験的知識を導入することで、人間にとって親和性の高いクラスタリング結果が得られることを示している。すなわち、当該領域で判別を要する分類数 (クラスタ数) と分類クラ</p>			

京都大学	博士（工学）	氏名	李 佳睿
------	--------	----	------

ス毎のデータ数（クラスタサイズ）に係る先験的知識が与えられたもとのファジィ集合としてのクラスタ生成を最適化する方法を提案しており、この手法を閉塞性睡眠時無呼吸症候群の患者データに対して適用して得られた学習モデルによる数値実験結果を示している。またこの学習モデルを用いた新規の患者に対する診断結果の説明に際して、典型的な病態との類似性から説得力を有する説明を行うことができることを述べている。

つぎに第 4 章の構造方程式モデリングに基づく説明可能な機械学習モデル（SEM-EML）では、まず集積されたデータの生成過程を説明するための因果関係を同定するべく、潜在変数や観測変数の間の相関関係を探し出し構造方程式として構築する。この構造方程式モデリングでは、諸変数の間に任意の因果関係を設定できるという柔軟性をもつが、変数間の因果関係を決定論的な関数関係として捉えるために、本研究では遺伝的アルゴリズム（Genetic Algorithm）を導入することで、潜在変数間の因果関係の構造を規定するモデルの学習を行う。この進化型学習により、強い相関関係に基づく変数間の関係に則って複数個の因果関係モデル候補が得られる。しかし相関性があれば必ず因果性があることにはならないことから、いずれのモデル候補が真の因果関係モデルを表すものであるかは特定することができない。この問題に対して本研究では当該領域の専門家による介入（intervention）を行う手法を提案している。すなわち、専門家が潜在変数のいずれかに対して介入を行った時に、その介入により特定の変数（目的変数）がどれほど変化するかをみることで、原因が結果に実際に影響を与えているかどうかを判定する。提案手法を前章で扱った閉塞性睡眠時無呼吸症候群の患者データに対して適用し、構造方程式では閉塞性睡眠時無呼吸症候群と相関の高い 13 項目の質問項目と 5 つの潜在要因との関係を特定した。そして最終的に無呼吸低呼吸指数（AHI）を目的変数として説明づける因果関係モデルを導出し、本モデルが他の機械学習手法による結果と比較して、因果関係による説明可能性と高精度の両面で優位な学習モデルであることを実証している。

第 5 章のリレーショナル転移学習アルゴリズム（RF-TL）では、前章の手法により可能となる因果関係モデルによってある系のデータ生成過程を捉えることができた場合に、新しいデータで予測したい問題設定と既存の学習済みモデルの問題設定が大きく違わなければ、学習済みの因果関係モデルの転移活用が可能となる点に着目し、そのための介入を領域専門家が行う手法を提案している。ソース領域（転移元）からターゲット領域（転移先）に写像可能な関係の吟味のために、まずソース領域での反実仮想説明として原因の潜在要因が変化するとしたときに他の変数間の相関がどのように変化するかを見極め、この関係をターゲット領域での因果関係モデルに写像して構成する。これにより転移先の少量のデータからでも高精度な学習モデルを構築することができる。この手法を新型コロナウイルス感染症患者のデータに対して適用し、高齢者患者についての因果関係モデルから若年者患者における因果関係モデルへの転移学習として例証し、患者の呈する兆候から ICU での治療を要するか否かを判定する診断問題に対して、他の診断手法と比較して最も高い精度で診断できることと、出力結果の解釈容易性の向上にも寄与できることを実証している。

最後に第 6 章では結論として、本論文の内容をまとめたあと、今後の展望について述べている。

氏名	李 佳睿
----	------

(論文審査の結果の要旨)

本論文では説明可能性を有する機械学習として二種類の領域専門家の知識を導入することを提案した。一つは当該領域において専門家が有する経験則に関する知識であり、もう一つは当該領域において明らかにされている因果関係の知識である。本論文は、二種類の領域専門家の知識を活用した機械学習モデルについてまとめており、得られた成果は以下の通りである。

機械学習モデルの有効性評価の基準として、高精度と高効率だけでなく、ユーザが有効に活用できるための基準として、機械学習モデルの学習した内容が人間にとって親和性の高いものでなければならないことを明らかにした。そのための基準として、機械学習モデルが判断する理由や根拠を示せる説明可能性の基準について着目し、これを充足する機械学習モデルの実現のためには、領域専門家が判別に係る先験的知識と潜在変数間の因果関係に係る知識の二種類の知識が必要になることを明らかにした。

改良型 Fuzzy c-Means 法 (CSCD-FCM) では、領域専門家が有する分類結果の事前分布に係る先験的知識を導入し、人間にとって親和性の高いクラスタリング結果が得られることを示した。この手法を閉塞性睡眠時無呼吸症候群の患者データに適用し学習モデルによる数値実験結果を示した。

構造方程式モデリングに基づく説明可能な機械学習モデル (SEM-EML) では、まず集積されたデータの生成過程を説明するための因果関係を同定するべく、潜在変数や観測変数間の相関関係を探し出し構造方程式として構築し、これに遺伝的アルゴリズム (Genetic Algorithm) を導入することで、潜在変数間の因果関係の構造を規定したモデル学習手法の提案を行なった。さらに当該領域の専門家による介入を行うことで、原因が結果に実際に影響を与えているかどうかを判定する手法を提案した。提案手法を前章で扱った閉塞性睡眠時無呼吸症候群の患者データに対して適用し、無呼吸低呼吸指数 (AHI) を目的変数として説明づける因果関係モデルを導出し、説明可能性と高精度の両面で優位な学習モデルが構築されることを実証した。

リレーショナル転移学習アルゴリズム (RF-TL) では、領域専門家の因果関係に係る知識を用いた介入が、教師信号となるラベル付きデータが少数しか得られていない異なる系における転移学習の性能向上に寄与できることを示し、因果関係モデルによって捉えられるデータ生成過程の情報が、異なる系での転移学習・メタ学習の手がかりになり得ることを提案した。この手法を新型コロナウイルス感染症患者のデータに対して適用し、診断の精度と説明容易性の両面から他手法に比べて優れた結果が得られることを実証した。

以上要するに、本論文は人間にとって親和性の高い機械学習モデルの構築について有用な知見を与えたものであり、その成果は学術上、実際上寄与するところが少なくない。よって、本論文は博士 (工学) の学位論文として価値あるものと認める。

また、令和3年12月13日、論文内容とそれに関連した事項について試問を行って、申請者が博士後期課程学位取得基準を満たしていることを確認し、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、論文内の一部に関して論文掲載が完了するまで、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。