

学業・研究の目的、内容およびその実績(別紙添付も可)

卒業論文のテーマとして「表形式データの自己教師あり学習モデルを対象とした効果的なデータ拡張法」を研究している。本研究は、人力データに基づいて、そのデータが属するクラスを予測する分類問題において、表形式データに特化した自己教師あり学習手法を提案する。具体的には、データを一部破損させた後の置換方法を改良する。研究目的としては、これまでは画像データやテキストデータに多く適用されてきた自己教師あり学習を表形式データに対して適用し、予測精度を向上させることである。その予測精度を向上させるために、既存手法では表現できていなかった実データの特徴量の偏りを考慮する手法を提案する。

分類問題とは、人力データに基づいて、そのデータが属するクラスを予測する問題のことである。そのような分類問題の精度を上げるための手法として、データ拡張や事前学習が挙げられる。データ拡張とは、「既存データが少ない時に、人工的にデータ数を増やす手法」のことを指す。既存データをモデルに通すことで増強し、出力された拡張データも学習に用いることで汎化性能を向上させる。事前学習とは、「大量のデータから、有用な特徴を事前に学習する手法」のことを指す。モデルに学習用データを用意することで、基礎的な情報を理解させ、より少ないデータ数や計算量で精度が向上するようにする。

本研究では、事前学習の中でも「ラベルのないデータ自身から擬似ラベルを生成し、有用な特徴を学習する手法」である自己教師あり学習に着目する。自己教師あり学習手法の研究は、表形式データではあまり行われていない。なぜなら、「セル内には量的変数と質的変数から成るさまざまなタイプのデータが混在しており、データの特性を考慮した学習が必要」だからだ。

その中で、表形式データに特化した自己教師あり学習手法の一つとして、SCARFが提案されている。SCARFはデータをランダムに置換、破損することでデータ拡張を行う。SCARFの問題点として、データを特徴量の一樣分布で置換していることが挙げられる。データを一樣分布で置換すると実データの特徴量の偏りを考慮できないため、現実的には起こり得ないデータを生成する確率が高くなると考えられる。そのため、特徴量の置換方法に着目してラベルごとのデータ分布の差異を考慮したデータ拡張手法を提案し、より実データに近いラベルごとの経験分布に基づく置換を行った。具体的には、データが属するクラスにおける特徴量の分布を算出し、復元抽出のようにして置換を行った。結果として、量的変数のデータセットに対して分類精度を向上させたため、効果的な手法と言える。今後の課題としては、質的変数のデータセットに対する分類精度の向上が挙げられる。この研究は、2024年10月末に開催された学会で発表した。

学業・研究の目的、内容およびその実績(別紙添付も可)

修士課程において、当初は量子情報理論、特に未知の量子状態を効率的に推定する方法を探る量子推定理論の学習に取り組んだ。この分野では、測定回数を抑えつつ高精度で量子状態を推定する技術が求められる。量子推定理論に関するレビュー研究を行い、国際会議でポスター発表を行った(研究業績書参照)。しかし、より独自性の高いテーマを追求する中で、量子情報理論から1次元量子系へと研究対象を移行した。このテーマ変更は、新たな課題に挑戦し、学術的な幅を広げるためのものである。

現在は、量子多体系におけるエンタングルメントスワッピングをテーマに研究を進めている。エンタングルメントスワッピングとは、2つのエンタングル状態間で適切な量子測定を行うことで、直接測定していない部分にエンタングルメントを生成する技術である。本研究では、2つのスピン梯子系の間でこの技術を用い、測定による効果が基底状態にどのような影響を与えるかを解析している。スピン梯子系はスピンと呼ばれる2つの自由度をもつ対象が1次元上に並んだもの(スピン鎖)が2本、横方向の相互作用によって結合した系である。このような系では、スピン間の相互作用の強さによって、自明な相と非自明なトポロジカル相が出現することが知られている。本研究では、異なる相の組み合わせで量子測定とポストセレクションを行った場合の挙動を調べることで、新たな知見を得ることを目指している。

研究手法は、主に解析的手法であり、行列積状態(MPS)とボソン化を用いた二本柱で進めている。スピン梯子系の基底状態として現れるトポロジカル相がMPSで表現されることを利用して、エンタングルメントスワッピングがMPSに及ぼす影響を計算している。また、MPSを与える行列の対称性変換則を調べることで、測定後の状態がどの相に対応するかを特定する。一方で、ボソン化を用いた有効場の理論では、例えば相関長の計算を通じて、MPSの計算で得られた結果の整合性を確認する。現時点では、一方のスピン梯子がトポロジカル相、もう一方が自明な相の場合に、量子測定とポストセレクションにより、残りの部分にすでに知られているトポロジカル相が現れることを確認している。

今後は、両方がトポロジカル相の場合や異なる測定を施した場合の挙動を解析し、新たなトポロジカル相の発見につなげたいと考えている。また、量子情報理論と量子多体系という異なる分野の融合を通じて、学術的な進展に寄与することを目指している。本研究は、スピン梯子系というモデルを通じて、量子多体系における測定の影響を深く理解し、トポロジカル相の研究に新たな展望をもたらす可能性を持つ。