# 審 査 の 結 果 の 要 旨

氏　　名　　孫 暁琳

物質の分子構造の特徴からバンドギャップなどの物性値を機械学習により予測するためには通常大量の学習データが必要である。しかし、個別の研究課題において学習データを用意するためにたくさんの実験を行うことは通常困難である。そこで、実験データに、シミュレーションデータなどの大量の補助データを追加することにより、精度の高い予測モデルを構築する試みが多く行われている。しかし、実験データの値と補助データのデータ値は直接数値値を比較できないことがしばしばある。そこで、本論文においては、機械学習における比較学習の手法を用いて、実験データと補助データのデータ値を直接比較することなしに、物性値を予測する新しい手法を開発した。またこの手法を用いて、分子構造から有機物質の光吸収スペクトルを予測するモデルを構築し、HOMO-LUMO ギャップについてのシミュレーションデータと実験データの統合が実際に予測モデルの性能を向上させることなどを示した。

本論文は４章と補遺により構成されており、第１章は機械学習を用いた物性予測についての導入、第２章ではガウス過程を用いた物性予測モデルの研究、第３章ではニューラルネットワークモデルを用いた物性予測モデルの研究、第４章では本研究の結論、補遺では、予測に使われた特徴のリスト、実験データとシミュレーションデータの値、転移学習と比較学習との精度比較について述べられている。

第１章では、まず比較学習をするためのデータの変換について説明された。この変換では、特徴ベクトルと目的変数値からなるデータを、データ点の全てのペアの組み合わせを考え、各ペアについては、目的変数の相対的な大小関係の情報のみを残すようにした。このようなデータの変換を行うことで、実験データにおける目的変数値と補助データにおける目的変数値が系統的に異なる場合であっても、目的変数値の順位関係を統合的な比較を可能にした。次に、このようにデータ点のペアのデータセットをモデル化するためのガウス過程を導入した。ガウス過程とは、有限時刻点の確率変数の結合分布がガウス分布で与えられる、確率過程の一種である。ガウス過程は関数空間上の確率分布に用いられることが多い。本研究においては、特徴ベクトルから目的変数の代用となるサロゲート値を計算する関数の確率分布としてガウス過程を採用した。これにベイズの定理とカーネルトリックを用いてデータが与えられたもとでのサロゲート値の予測分布を導出し、新たな特徴ベクトルが与えられた時にサロゲート値を計算するアルゴリズムを構築した。次に、この手法を物質の物性予測に用いるために、任意の化学物質を特徴ベクトルに変換する方法について述べられた。化学物質に対し、その化学構造は SMILES という文字列に変換し、化学的な性質については外部ツールを用いて特徴ベクトル化を行った。次にベンチマーク試験に用いるデータセットについて述べられた。１つ目の検証データは無機化合物のバンドギャップについてのデータを用いた。このデータベースでは、各化合物のバンドギャップを精度と計算時間の異なる２つの計算により求めた値が与えられている。そこで計算は速いが精度が劣るシミュレーション方法で計算したバンドギャップのデータを、計算が遅いが精度が高いシ

ミュレーションによるバンドギャップの予測に活用できるか実験を行った結果、４種のデータセット中３種で補助データの追加により性能が向上することが分かった。次に有機分子の光吸収スペクトルのデータセットを用いて実験を行った。このデータセットでは実験データに時間依存密度汎関数法によるシミュレーション結果を追加することで、吸収光の波長の予測性能が向上することが示された。これらによりガウス過程を用いた比較学習により、補助データを活用することの有用性を示すことができた。

ガウス過程を用いた第２章とは異なり、第３章ではニューラルネットワークモデルを用いた比較学習法の開発を行った。この手法では、特徴ベクトルを入力とし、目的変数のサロゲート値を出力するニューラルネットワークを用意する。それからデータ点のペアに対し計算される２つのサロゲート値の大小関係のみに依存する損失関数を定義し、その損失関数を最小化するようにニューラルネットワークのパラメータを学習させる。最初のベンチマーク試験では、PubChemQC データベースにある光吸収波長と HOMO-LUMO ギャップのデータを使用し、HOMO-LUMO ギャップの補助データの活用が、光吸収の波長の予測に効果的であることを示すことができた。もう一つのベンチマーク試験では、第 Xa 因子阻害剤の薬剤効果のデータベースを用いて、異なる研究グループのデータセットを統合することで予測性能を向上できることを示すことができた。

第３章の後半では、ガウス過程に基づく比較学習手法とニューラルネットワークを用いた比較学習手法の比較を行い、計算速度と予測精度の両面でニューラルネットワークが優れていることが示された。ただし、ガウス過程に基づく比較学習手法は、ベイズ最適化を用いた探索的研究に使うことができるため、どちらも実用的な価値をもつことが述べられた。

第４章で本研究で開発した比較学習の手法と得られた結果についてまとめを行い、それらの利点と欠点がまとめられ、今後に残された研究課題がのべられた。

本論文の第２章の一部は共著論文として国際誌に出版されているが、アルゴリズムの実装と計算機実験は学位申請者によって行われており、学位申請者は研究に十分な貢献をしていると判断される。

よって本論文は博士（科学）の学位請求論文として合格と認められる。

以上２３３５字