論 文 内 容 の 要 旨

博士論文題目 ハイスループット反応実験データにおける高精度な収率予測モデルの構築

Constructing highly accurate yield prediction models for high-throughput experiment reaction data

氏　名　　佐藤　彰准

（論文内容の要旨）

コンピュータを利用した有機合成経路設計システム構築のためには、反応条件を考慮した反応経路を提案する必要がある。そのためには化学反応における収率を予測するモデルが必要である。本学位論文では、均質な実験条件のもとで計測されたハイスループット反応実験データを対象として、高精度な収率予測モデルの構築法を提案している。

本学位論文は、第1章から第4章で構成されており、第1章ではコンピュター支援による反応設計に関する既往の研究をレビューし、溶媒や触媒などの反応条件を考慮した収率予測モデルの必要性、および研究スコープを明確にしている。また、既往の収率予測モデルが汎用的に使用することができない理由など、手法としての限界を議論する形で本論文の背景と目的を述べる。

第2章では構造式を化合物表現とする収率予測モデルにおける「原子」の表現方法を検討する。公開データである C-N クロスカップリング反応における反応実験データを利用した検証から、原子環境の類似性を考慮した Mol2Vec を「原子」の表現として使用することの優位性を示した。ここでは、計算化学により算出した特徴量に基づく既往の予測モデルよりも提案モデルの精度が高いことを確認している。

第3章では第2章の予測モデルでは不十分だった予測精度の改善に加えて、反応に使用する試薬数の変化に対応可能なモデル構造を提案する。予測精度の改善方法として事前学習とアンサンブル手法を適用し、Transformer エンコーダーを message passing neural network (MPNN) と組み合わせる。2つのハイスループット反応実験データに対して内挿性・外挿性において提案モデルが先行研究よりも優れていることが示された。さらに、アブレーション検証から、事前学習による予測精度の改善と MPNN と Transformer エンコーダーを組み合わせることの有効性を示した。

第4章では、本研究で提案したモデルの限界と今後の展望について述べる。特に、多様な文献から取得する化学反応データは、反応温度・反応時間・試薬量などの反応条件が統一されておらず、実験誤差も不明である。このような反応データを統一した方法でモデル構築に利用するために、予測値の不確実性を含め収率予測するモデルを展望として述べている。

氏　名	佐藤　彰准

（論文審査結果の要旨）

一般的に、コンピュータを利用した有機合成経路設計では、反応物と生成物の関係をモデル化し、生成物から妥当な反応物をモデルが提案する。ここでは、反応収率に大きな影響を与える反応条件は考慮されていない。実用的な有機合成反応設計システム構築のためには、反応条件を考慮する収率予測モデルが必要である。佐藤彰准氏は、モデルの解釈方法を含めた、反応条件を考慮した収率予測モデルの構築法に関する研究を提出した。

本論文では、均質な実験条件のもとで計測されたハイスループット反応実験データを対象として、高精度な収率予測モデル構築方法の開発と、異なる化学反応に適用可能なモデル構造を提案している。反応条件としては、反応に用いられる溶媒、触媒を含めた化合物を対象としている。本論文の主要な結果は以下の通りである。

１．構造式を化合物表現とする収率予測モデルにおける適切な「原子」の表現方法の提案。「原子」の表現は深層学習モデルの入力として用いられる。C-N クロスカップリング反応実験データを用いた検証から、佐藤彰准氏は原子環境の類似性を考慮した Mol2Vec を「原子」の表現とすることの優位性を精度と解釈性の観点から示した。特に、注意機構（attention mechanism）により、収率予測に重要な部分構造を構造式上に示す「モデルの解釈」を行なった結果、Mol2Vec を「原子」の表現とした場合に、先行研究で重要と示されていた官能基を重要な部分構造と正しく捉えていた。加えて、これまで標準的な方法であった「計算化学から算出した電子状態に基づく特徴量と機械学習モデルの組み合わせ」よりも提案モデルの精度が同等であることを確認しており、モデル構築手法として新規性と有用性がある。

２．異なる化学反応に適用可能な柔軟な収率予測モデル構造の提案。実際の化学反応では、試薬数や基質の数、触媒の有無などの因子は反応毎に異なる。佐藤彰准氏は、入力する化合物数が可変となる場合にも対応可能なモデル構造を提案した。具体的には、message passing neural network（MPNN）の出力を Transformer エンコーダーの入力として利用し、反応における化合物の役割をエンコーダー入力時に付与する。佐藤彰准氏は二つのハイスループット反応実験データに対して収率予測精度の検証を行い、提案手法が先行研究のモデルよりも優れていることを示した。さらに、佐藤彰准氏は、アブレーション検証から、事前学習による予測精度の改善と MPNN と Transformer エンコーダーを組み合わせることの有効性を示した。

以上、本論文ではハイスループット反応実験データを対象として、高精度な収率予測モデル構築方法などについてまとめられている。学術的にも大きな意義があり、審査委員一同は本論文が博士（工学）の学位論文として価値のあるものと認めた。