

医薬品原薬製造プロセス開発への機械学習の活用

メタデータ	言語: Japanese 出版者: 公開日: 2024-03-27 キーワード (Ja): キーワード (En): 作成者: 森下,敏治 メールアドレス: 所属:
URL	http://hdl.handle.net/10291/0002000345

「博士学位請求論文」審査報告書

審査委員 (主査) 理工学部 専任准教授

氏名 金子 弘昌

(副査) 理工学部 専任教授

氏名 深澤 倫子

(副査) 理工学部 専任教授

氏名 土本 晃久

1 論文提出者 森下 敏治

2 論文題名 医薬品原薬製造プロセス開発への機械学習の活用

(英文題) Application of Machine Learning to the Development of Active
Pharmaceutical Ingredients Manufacturing Processes

3 論文の構成

本論文は、序論（第1章）と、3章からなる本論（第2章～第4章）、そして結論（第5章）から構成されている。

第1章 研究背景および研究目的、本論文の構成

第2章 自己促進分解温度の予測モデルの開発

第3章 クラスタリングを用いたベイズ最適化の初期条件の決定法

第4章 密度汎関数法(DFT)を用いたベイズ最適化探索性能向上

第5章 結論

4 論文の概要

本論文は、検討期間の短縮による上市までのスピードアップや開発コスト削減などを目的として、医薬品原薬製造プロセス開発における機械学習活用についての検討をまとめたものである。

第1章では、研究背景および研究目的、論文の構成が述べられている。高品質かつ効率的、堅牢な原薬製造プロセスを開発するために、研究者は効率的な実験を計画・実施し、その結果を解析し、次の実験が必要かどうか、行う場合はどのような実験条件とすべきかの判

断を繰り返す。これらの作業は非常に属人的であり、手動操作も多く残されている状況であった。また、反応予測、安全リスク、環境影響、品質・物性予測などに各種シミュレーション技術(量子化学計算, 分子動力学(MD: Molecular Dynamics)計算, 流体解析(CFD: Computational Fluid Dynamics), プロセスシミュレーションなど)が活用される場合もあるが、多様なモダリティに技術開発が追いついておらず、十分に活用できているとはいえない状況であった。予測モデルを用いて収率、物性、品質などを予測し、最良となる条件を少ない試行回数で探索することができれば、開発期間を大きく短縮させることができると考え、実験・分析・解析・判断のサイクルを、より迅速に、正確に、効率よく回すために、機械学習を用いた予測モデルの開発とベイズ最適化(BO: Bayesian Optimization)を用いた適応的実験計画法の活用を検討することとした。機械学習予測モデルの開発対象として、有機過酸化物の自己促進分解温度(SADT)予測モデルを構築することとした。熱的な危険性を把握することは化合物の開発や保管、輸送、製造等において非常に重要であり、本モデル開発が医薬品開発の安定生産に貢献できると考え、従来法での課題であった計算負荷の削減と計算精度の向上を試みた。また、BOを用いた適応的実験計画法を有機合成反応条件探索に適用することとした。合成原薬の製造プロセスは、反応にて目的物を得た後、その後の操作で不純物を除去して高い品質の化合物を得る。特に反応は最初に行われる最も重要な工程であり、この工程の完成度次第でその後の工程に要求されるレベルが大きく変化する。その中でも金属触媒を用いた反応は、特異的な挙動を示す場合があり、専門家でも反応挙動を正確に予測できないことが多い。最適条件を探索するためには網羅的にスクリーニングをする必要があるが、大域的な最適解により少ない試行回数で到達することができるBOは、金属触媒反応における最適条件探索と相性が良いと考えた。一方で、BOを合成反応条件探索に適用する上で、様々な課題が残されていた。BOを現場で活用する上で生じる数々の課題の中で、初期条件の決定方法と化合物情報の記述子計算方法に絞り、研究者のドメイン知識や記述子情報の適切な活用により、BOの最適解探索性能を向上することができないか可能性を模索した。性能評価のためには評価指標が必要となるが、BOの性能は初期サンプルの影響を大きく受け、恣意的にサンプルを選択することはできない。そのため、試行回数を十分に増やして、数値実験における探索結果の平均値やばらつきでBOの探索性能評価を行うこととした。

第2章では、有機過酸化物の構造式から自己促進分解温度(SADT)を推算するモデルの構築を試みた。化合物の構造式から記述子を計算し、機械学習予測モデルを適用した結果、比較的精度良くSADTを予測できるモデルを構築することができた。モデル構築時の前処理として分子力学(MM: Molecular Mechanics)計算や密度汎関数(DFT: Density Functional Theory)計算、変数選択に遺伝的アルゴリズム(GA: Genetic Algorithm)を用いたところ、適用しない場合と比べて予測精度が向上し、既存のモデルと比較しても遜色ない精度で予測をできていた。精度が良いモデルを構築するためには十分な前処理と適切な変数選択が重要であり、より厳密に化合物構造の最適化を実施してから記述子の計算を行う方が望ましいものの、計算負荷が非常に重くなるため、予測精度と計算負荷バランスの最適化を考える必要があると述べた。

第3章では、BOの初期条件決定方法について考察を行った。BOで最適解を効率的に探索

するには、ガウス過程回帰(GPR)モデルを構築する際に適切な初期サンプルを用いる必要がある。化合物を用いた実験デザインの場合、化学構造から計算される分子記述子間には常に高い相関が存在し、類似した構造を持つ化合物は化学空間においてクラスタを形成する。そのため、実験条件の情報が最大となる初期試料を得るためには、各クラスタから一様に初期試料を選択することが望ましい。D最適基準のような実験空間の類似性を利用したサンプリング手法は、相関性の高い分子記述子を対象とした場合に想定通りに機能しない場合がある。本研究では、目的変数に大きな影響を与える因子をカバーするクラスタリング情報に基づく初期サンプル選択手法を提案し、B0とのカップリング反応条件の最適化に適用した。その結果、クラスタを適切に形成し、各クラスタから初期サンプルを選択した場合、提案手法はランダムサンプリングやD最適基準に基づくサンプリングよりも少ない実験回数で最適解に到達することを確認した。さらに、1ラウンドあたりの実験回数が少ない場合、提案手法の効果はクラスタ情報を含まない他の手法よりも大きく、探索に必要なラウンド数を削減できることも確認された。クラスタリングは教師なし学習であり、目的変数(実験結果)に関する情報を実験前に知ることはできない。これらの情報を結びつけるためには、付加的な情報(専門家の知識)が必要となる。実験空間構築時に領域知識を活用して適切なクラスタを形成し、初期条件を決定することができれば、反応条件最適化のような化合物の組み合わせの場合、B0の検索性能をさらに向上させることができるとまとめた。

第4章では、分子構造情報からDFT計算で得られた記述子を利用する際に、ベイズ最適化の探索性能を向上させる方法について検討を行った。B0で化合物を取り扱う場合、分子構造情報から計算された記述子を説明変数として利用することができる。様々な組み合わせの基底関数・汎関数でDFT計算された記述子を活用して、B0の探索性能を向上させることができる方法を開発した。今回の検討では直接的アリアル化およびカップリング反応に対して、いくつかの基底関数・汎関数の組み合わせで記述子群を作成し、それらから作成したデータセットを用いて探索性能の検証を実施した。複数の記述子群を平均化して作成したデータセットをB0に用いた場合、単一の記述子群から作成したデータセットを用いる場合よりも探索性能が向上した。複数の記述子群を平均化する方法は計算負荷も小さく、量子科学計算に関する知識があまりない研究者でも簡単に利用可能な方法であるとまとめた。

第5章では、各章の内容を完結に記述し、原薬のプロセス開発の様々な場面で機械学習を活用した予測モデルおよびそれを用いた最適化を活用できると結論付けた。一方で、さらなる活用には実験・分析作業の自動化やデジタル化が必要であり、機械学習を十分に活用できる環境構築と本技術の普及・活用を継続すると述べた。

5 論文の特質

本論文は、原薬の製造プロセス開発に機械学習を活用することを検討したものである。自己促進分解温度(SADT)予測モデル開発では、従来法の課題であった計算時間の短縮や予測精度の改善に取り組み、既存法を上回る性能を持つ予測モデルを開発した。また、その過程において、DFT活用による分子構造の最適化や変数選択の重要性について述べ、予測に影響を与える因子について明らかにした。反応条件最適化問題に対してベイズ最適化の適用を検討し、課題となる初期条件の決定や記述子計算に関して、複数の反応セット、膨大な条件

で数値実験を繰り返し実施し、B0 の最適解探索性能を向上させる方法を見出した。また、その過程において、これまで実施されてこなかったベイズ最適化の探索性能を定量的に評価する方法を新たに確立した。従来法を上回る性能を持った予測モデルを開発したこと、ベイズ最適化の活用において課題となる点を複数解決したこと、これらの点が本論文の特質である。

6 論文の評価

本論文は、原薬の製造プロセス開発に対して機械学習を活用する上で重要となる知見を多く含んでいる。加えて、本論文の研究内容はすでに評価ある国際誌に発表されている。内容の新規性、独創性、将来への発展性、医薬品原薬プロセス開発分野への貢献等を総合的に判断して本論文内容は高く評価できる。

7 論文の判定

本学位請求論文は、理工学研究科において必要な研究指導を受けたうえ提出されたものであり、本学学位規程の手続きに従い、審査委員全員による所定の審査及び最終試験に合格したので、博士（工学）の学位を授与するに値するものと判定する。

以上

主査氏名（自署） 金子 弘昌