

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

2023年度 理工学研究科

博士学位請求論文 (要旨)

医薬品原薬製造プロセス開発への機械学習の活用

応用化学専攻
森下 敏治

1 問題意識と目的

一般的に新薬の研究・開発は十数年程度の期間と数百億円から数千億円の投資を必要とするが成功率は決して高くない。難病や希少疾患などのいわゆるアンメットメディカルニーズに対する治療薬を患者さんのもとへ届けるためには、低分子医薬だけでなく、抗体医薬をはじめ、核酸医薬や遺伝子治療薬、細胞医薬品のような様々なモダリティの医薬品開発が必要な状況であり、研究開発費は増加を続けている。このように日本だけでなくグローバルにおいても医療費の増加が大きな課題となっており、DX(Digital transformation)による業務効率化がすすめられている。製薬業界を取り巻く経営環境は目まぐるしく変化を続けており、研究・開発の難易度とリスク、開発コストは年々高まっている。医薬品原薬の製造プロセス開発においても、開発スピードの加速化や検討の質の向上がこれまで以上に求められている。限られた開発期間の中で、環境、安全、品質、実験・製造コストなど、さまざまな観点で最適なプロセスの開発を行わなければならないが、研究開発の現場では研究者の勘と経験による試行錯誤で実験条件が決定されることも多い。医薬品原薬の製造プロセス開発に関しても、これまで以上にDXの推進および業務プロセスの大きな改革が求められている状況である。本研究では、医薬品原薬製造プロセス開発への機械学習の活用における課題の把握と対策の実行を行い、製造プロセス開発の効率向上による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、安定生産へとつなげることを目的とする。

CMC(Chemistry, Manufacturing and Control)の原薬部門では、化学合成原薬およびバイオ原薬の製造プロセス開発、工業化研究、生産サイトへの技術移転や承認申請、製品化後のトラブル対応など、医薬品原薬のライフサイクルを通じて技術的な対応を実施している。製剤化研究、臨床試験をスケジュール通りに進めるためにはタイムリーな原薬供給が不可欠であり、高品質かつ効率的、堅牢な原薬製造プロセスを早期に構築することが求められている。そのような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、その結果を分析機器で測定した後、測定結果を解析し、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る。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ないという難しい状況である。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。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、有機過酸化物を対象とした自己促進分解温度(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

重要である。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、分液、濃縮、抽出、晶析、ろ過、乾燥などの単位操作の組み合わせで構成される。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り、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、探索結果の平均値やばらつきで評価を行うこととした。

2 構成及び各章の要約

本論文は1章の序論、2章から4章までの本文と、5章の結論から構成される。

第1章「序論」では、研究背景および研究目的、本論文の構成を述べた。

第2章「自己促進分解温度(SADT)の予測モデルの開発」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飛躍的に予測精度が向上した。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ることを確認した。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る。一方で、SADTの予測精度を向上させるためには、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が、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。

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」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ではうまく機能せず、また、サンプル選択にクラスタの情報を考慮しない。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、BO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わかった。適切なクラスタリングとは、クラスタに属するサンプル数がほぼ一様であり、対象変数に大きく寄与する要因によってクラスタを形成できる場合を指す。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。研究現場では、ベイズ最適化と専門家の知識を組み合わせたい

という要望が多い。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、BOの検索性能をさらに向上させることができる。

第4章「DFTを用いたベイズ最適化探索性能向上」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。BO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、BO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BO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。

第5章「結論」では、本論文の内容に関するまとめを行った。原薬プロセス開発への機械学習の適用を検討したところ様々な場面で活用することができた。計算リソースの問題や心理的なハードルなど、様々な課題が残されているものの、製造プロセス開発の効率向上につなげるために、機械学習のさらなる活用を検討していくべきである。