

研究提案の要旨

○要旨

申請者の目指す未来は、IoT 機器から得られる多種大量のデータ (ビッグデータ) を駆使して、様々な対象を自由自在に制御・管理するシステムにより、工場・施設等の自動的かつ高速な最適化が達成される社会である。この社会を実現するためには、図 1 のように、ある対象において目指すべき時間的な推移から、それを達成するために空間に配置された機器等をどのように時間的に推移させればよいかを、即時的に予測することが必須である。例えば製造工場において、製品品質の目標を設定するだけで、その目標を達成するための製造条件や機器の測定値が自動的に提案され、それに基づいて最適な製品が製造されたり、植物工場において、植物の成長の目標を設定するだけで、その目標を達成するための水量や栄養の量等の生育条件や機器の測定値が自動的に提案され、最適な植物が生育されたりする未来を考えている。本研究ではこの未来に向けて、対象の目標からそれを達成するための方法・過程を提案するシステムを開発したい。

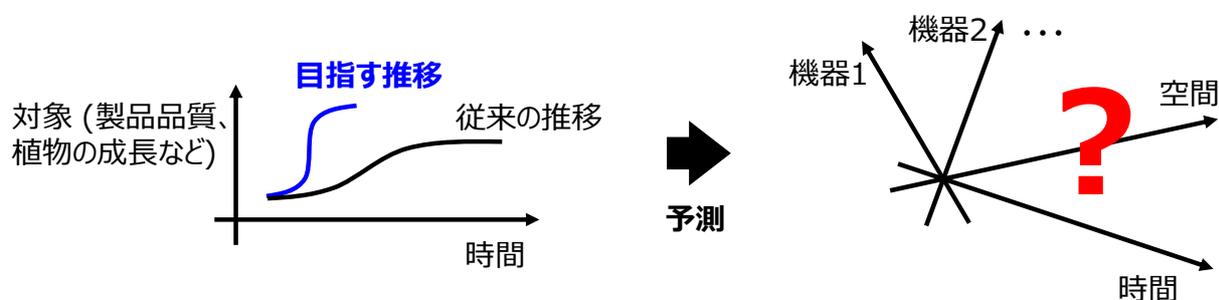


図 1. 本研究構想で対象とする IoT 活用の課題

IoT 環境においては機器等により様々な装置・施設における変数の値が計測され、それらをモニタリングすることが可能である。例えば製造工場において、装置で測定される温度や圧力等のセンサー情報に基づいて、その装置が正常かどうかをモニタリングできる。さらに、そのような IoT 機器から得られる多種大量のデータを用いて、変数間の相関関係をモデル化することで、例えば工場において、製品品質等の測定が困難な変数 Y を温度・圧力等の簡単に測定可能な変数 X から予測できる。また植物工場において、水量や栄養の量等のセンサー情報に基づいて、植物の成長 Y を水量・栄養の量・日射量等 X から予測できる。

工場において IoT 機器により状態をモニタリングしたり予測したりできるが、真に必要なことは、製品品質や植物の成長といった対象 Y が理想的な値になるように、工場全体を最適化する制御方法・運転方法を提案することである。工場における運転最適化の方法を提案することは、目的変数 Y と操作可能な変数 X の間で構築したモデル $Y=f(X)$ に対して、 X を入力して Y を予測する順解析とは逆に、 Y を入力して X を予測する、すなわちモデルを逆解析することに対応する。一般的なモデルの逆解析は、 X の仮想サンプルを計算機で大量に生成し、それらをモデルに入力して Y の値を予測し、予測値が良好な仮想サンプルを選択する、すなわち順解析を網羅的かつ効率的に繰り返す擬似的な逆解析にすぎない。これでは人が事前に想定した X の探索範囲における Y を予測することにすぎず、特に既存の Y の値を超越する最適化を目指すには不十分である。さらに、IoT 環境において製造工場・植物工場等を日々運転し、リアルタイムにデータが取得される状況において真に必要なことは、 Y を理想的に推移させるための X の値を、オンラインで即時的に提案することである。

擬似的な逆解析では、網羅的に仮想サンプルを生成する必要があり計算時間がかかるため、**従来法では即時的な X の提案はまったく不可能**であり、X が予測されるまでの時間がそのまま Y を理想的に推移させる際のロスタイムになってしまう。

申請者は分子の物性・薬理活性および分子構造といった化学のデータを情報科学の手法によって解析し、分子の機能予測や合成経路予測をしようとする学問分野であるケモインフォマティクスにおいて、人工知能 (Artificial Intelligence, AI) モデルに基づいて目的変数 Y (複数可能) の値から説明変数 X の値を直接的に予測する、すなわち AI モデルを直接的に逆解析する手法 (**直接的モデル逆解析法**) を開発した。直接的モデル逆解析法では、従来法と異なり大量の仮想サンプルを生成する必要がなく、**即時的な X の提案が可能**となる。さらに X に上限・下限等の制限を設ける必要もない。**IoT 環境において、Y の理想的な推移を達成するための即時的な X の予測には、直接的モデル逆解析法の原理がカギ**になると考えた。さらに提案法ではデータの可視化も同時に達成できるため、IoT 環境においてはモニタリングと直接的モデル逆解析を同時に達成可能である。

IoT 環境において、最適化したい対象である Y の理想的な時間推移を達成するための、工場全体の制御方法・運転方法を提案するためには、変数 X の時間推移だけでなく工場における X の位置情報を考慮する必要がある。以上の背景を踏まえ本研究では、**図 2** のように IoT 機器から得られる多種大量のデータ (ビッグデータ) を活用することで、**Y (複数の場合あり) の理想的な推移から、高精度かつ即時的に X の時空間変化を提案するシステムを開発することを目的**とする。そして、変数変換や確率分布の多様化によるモデルの柔軟性および予測精度の向上、X や Y における時間および空間を考慮したモデル構築、変数のスパース化および分散モデルによる直接的モデル逆解析の高速化、種々の工場におけるデータへの展開によって目的の達成を目指す。

提案するシステムに Y の目標を入力するだけで、非常に多くの X に対して、目標を達成するための時空間を考慮した推移を自動的に提案でき、インテリジェントな機器等をニーズに合わせて制御することに貢献する。さらに、その X の提案を実施した後の実際の Y の結果に基づいて、さらなる Y の最適化へ自動的に先導できる。これにより、**人が先導して最適化する社会から、IoT 機器およびそこから得られるビッグデータが先導して自動的に最適化する社会への革新**が達成される。

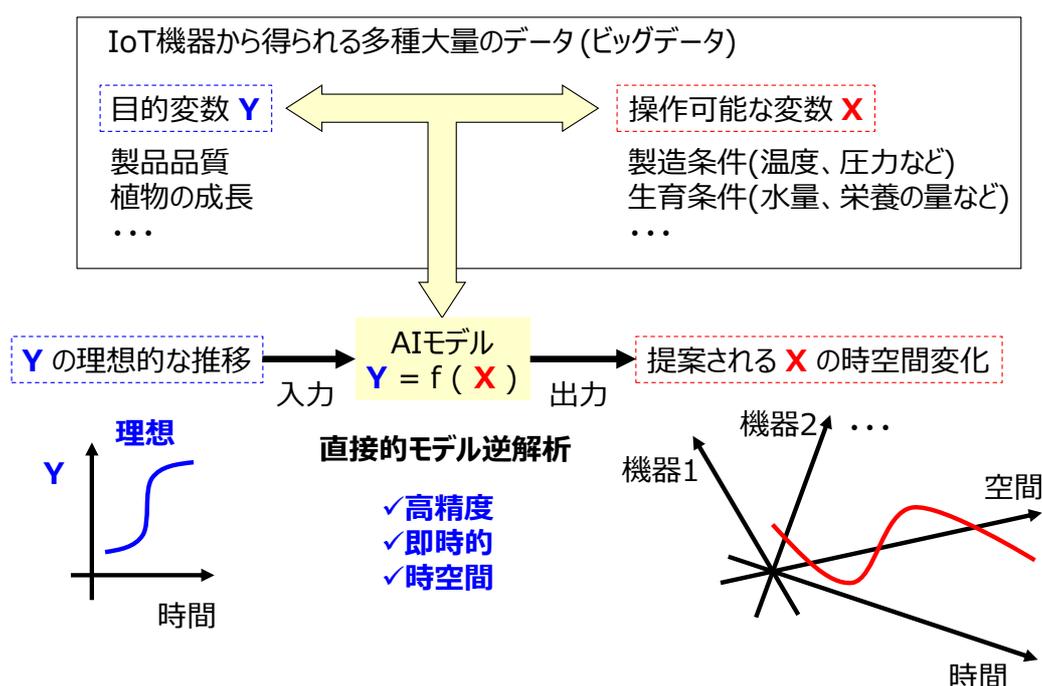


図 2. 本研究構想の目的および概要

研究構想

1. 研究の背景・目的

申請者の目指す未来は、IoT 機器から得られる多種大量のデータ (ビッグデータ) を駆使して、様々な対象を自由自在に制御・管理するシステムにより、工場・施設等の自動的かつ高速な最適化が達成される社会である (図3)。この社会を実現するためには、「研究提案の要旨」の図1のように、ある対象において目指すべき推移があるとき、それを達成するために空間に配置された機器等をどのように時間的に推移させればよいかを、即時的に予測することが必須である。例えば製造工場において、製品品質の目標を設定するだけで、その目標を達成するための製造条件や機器の測定値が自動的に提案され、それに基づいて製造工場が自動制御され最適な製品が製造されたり、植物工場において、植物の成長の目標を設定するだけで、その目標を達成するための水量や栄養の量等の生育条件や機器の測定値が自動的に提案され、それに基づいて植物工場が自動制御され最適な植物が生育されたりする未来を考えている。本研究ではこの未来に向けて、対象の目標からそれを達成するための方法・過程を提案するシステムを開発したい。



図3. 申請者の目指す未来および本研究提案の対象

本研究構想に至った背景・経緯

申請者はこれまでケモインフォマティクス・マテリアルズインフォマティクス・プロセスインフォマティクスの研究を推進してきた。それぞれの詳細を以下に示す。

- ✓ **ケモインフォマティクス:** 分子の物性・薬理活性および分子構造といった化学のデータを情報科学の手法によって解析し、分子の機能予測や合成経路予測をしようとする学問分野。申請者は主に、分子構造 X と物性・薬理活性 Y との間でモデル $Y=f(X)$ を構築し、モデルに基づいて物性・薬理活性の目標値をもつ分子構造を設計する研究を推進 [2-4,6-11,13,16,21,24,27]
- ✓ **マテリアルズインフォマティクス:** 統計解析や機械学習などの手法やデータベースを活用し、新たな材料を探索する学問分野。申請者は主に、材料の実験条件・製造条件 X と材料の物性・特性 Y との間でモデル $Y=f(X)$ を構築し、モデルに基づいて物性・特性の目標値をもつ実験条件・製造条件を設計する研究を推進 [2,3,6-9,11,15,22,23,27,30,40,41]
- ✓ **プロセスインフォマティクス:** IoT 機器を活用しながら産業プラントにおいて計測される温度・圧力等のデータを情報科学の手法によって解析し、製品品質の予測・制御や異常検出等を行う学問分野。申請者は主に、測定が困難なプロセス変数 Y の値を、簡単に測定可能なプロセス変数

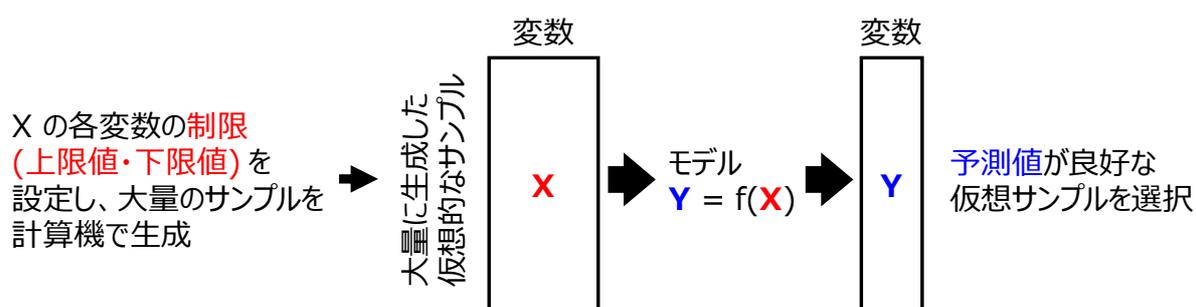
X から予測しながら制御する研究を推進 [1,2,5,7,12,14,17,18-20,25-28,31-39]

以上のように申請者は、X と Y との間のモデル $Y=f(X)$ の構築、およびモデルを用いた Y の値から X の値の設計、すなわち**モデルの逆解析**に関する研究を実施してきた。様々な装置・施設等の情報が繋がるIoT 環境下においては、Y が最適化したい対象、例えば製造工場における製品品質や植物工場における植物の成長であり、X が機器等により測定される変数である。申請者の研究背景を踏まえ、**新たな逆解析技術を開発することで、工場・施設等の時空間を制御して高速かつ自由自在に Y の最適化が可能な次世代IoT 活用基盤システムの構築**を達成できると考えた。

IoT 環境においてはセンサー等により様々な装置・施設における変数の値が計測され、それらをモニタリングすることが可能である。例えば製造工場において、装置で測定される温度や圧力等のセンサー情報に基づいて、その装置が正常かどうかをモニタリングできる。さらに、**IoT 機器から得られる多種大量のデータを用いて、X, Y を含む変数間の関係を解析して相関関係をモデル化**することで、モデルに基づいて X から Y を予測することも可能となる。工場においては、製品品質等の測定が困難な変数 Y を温度・圧力等の簡単に測定可能な変数 X から予測できる。また植物工場において、水量や栄養の量等のセンサー情報に基づいて、工場が正常に運転しているかどうかをモニタリングし、植物の成長 Y を水量・栄養の量・日射量等の変数 X から予測できる。

工場においてIoT 機器により状態をモニタリングしたり予測したりできるが、**真に必要なことは、製品品質や植物の成長といった対象 Y が理想的な値になるように、工場全体を最適化する制御方法・運転方法を提案すること** (「研究提案の要旨」の図1) である。ただし現状では、工場において行われている一般的なプロセス制御 (Comput.Chem.Eng., 128 (2019) 538) は、現在の状態のみを制御するものであり、さらにIoT 環境のように非常に多くの変数が広い空間に配置されている状況にはまったく対応できない。そこでIoT 機器から得られる多種大量のデータ (ビッグデータ) を活用する。

工場における運転最適化の方法を提案することは、目的変数 Y と操作可能な変数 X の間で構築したモデル $Y=f(X)$ に対して、X を入力して Y を予測する順解析とは逆に、Y を入力して X を予測する、すなわち**モデルを逆解析することに対応**する。一般的なモデルの逆解析 (Mol.Inf., 39 (2019) 1900087) は、**図4**のように X の仮想サンプルを計算機で大量に生成し、それらをモデルに入力して Y の値を予測し、予測値が良好な仮想サンプルを選択する、すなわち**順解析を網羅的かつ効率的に繰り返す擬似的な逆解析にすぎない**。実験計画法において近年注目されているベイズ最適化 (P.IEEE, 104 (2016) 148) により、Y の予測値ではなく獲得関数の値を用いることで、データの外挿領域を探索できるようになったが、逆解析の方法としては**ベイズ最適化も擬似的な逆解析**



X を入力して Y を予測し、結果を選択する**擬似的なモデルの逆解析にすぎない**

図4. 従来のモデルの逆解析

である。これでは人が事前に想定した X の探索範囲における Y を予測することにすぎず、特に**既存の Y の値を超越する最適化を目指すには不十分**である。

IoT 環境において製造工場・植物工場等を日々運転し、リアルタイムにデータが取得される状況において**真に必要なことは、 Y を理想的に推移させるために、オンラインで即時的に X の値を提案すること**である。擬似的な逆解析では、網羅的にサンプルを生成する必要があり計算時間がかかるため、**従来法では即時的な X の提案はまったく不可能**であり、 X が予測されるまでの時間がそのまま Y の理想的な推移を達成するまでのロスタイムになってしまう。

申請者はケモインフォマティクスの分野において、Gaussian Mixture Regression (GMR) [16] や Generative Topographic Mapping Regression (GTMR) [3] の手法により構築された生成モデルに基づいて、 Y の値から X の値を直接的に予測する、すなわちモデルを直接的に逆解析する手法 (**直接的モデル逆解析法**) を開発した (図 5)。低分子の有機化合物データにおいて、分子の特徴量を X 、物性を Y として GMR や GTMR で構築した生成モデルを解析することで、 **Y の目標値から直接 X の値を瞬時に予測できる**。直接的モデル逆解析法では、従来法と異なり大量の仮想サンプルを生成する必要がなく、**即時的な X の提案が可能**となる。さらに X の各特徴量に上限・下限を設定する必要もない。**IoT 環境において、 Y の理想的な推移を達成するための即時的な X の予測には、直接的モデル逆解析法の原理がカギ**になると考え、本研究提案に至った。なお GTMR ではモデルの逆解析に加えて、データの可視化も同時に達成できるため、IoT 環境においてはモニタリングと直接的モデル逆解析を同時に達成可能である。

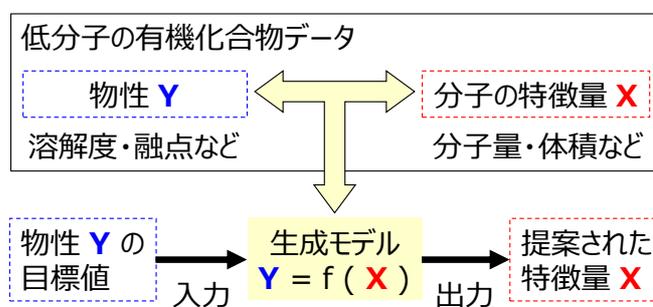


図 5. 直接的モデル逆解析法 [3,16]

本研究構想の目的

IoT 環境において、最適化したい対象である Y の理想的な時間推移を達成するための、工場全体の制御方法・運転方法を提案するためには、変数 X の時間推移だけでなく工場における X の位置情報を考慮する必要がある。以上の背景を踏まえ本研究では、IoT 機器から得られる多種大量のデータ (ビッグデータ) を活用することで、 **Y (複数の場合あり) の理想的な推移から、高精度かつ即時的に X の時空間変化を提案するシステムを開発することを目的**とする (「研究提案の要旨」の図 2)。

「科学技術イノベーションの源泉となる先駆的な成果」としての新技術のシーズ (新しい発見・発明)

- ✓ ビッグデータを用いて高精度かつ即時的にモデル構築とモデルの逆解析を同時に達成する手法
- ✓ IoT 機器から得られる多種大量のデータにおける変数間の関係および最適化する時空間を考慮して、自由自在に対象を制御する技術

「科学技術イノベーション」の観点として、この成果がもたらす科学技術上のインパクトや、科学技術上のインパクトが導く社会のあり方の変化

提案するシステムに Y の目標を入力するだけで、非常に多くの X に対して、目標を達成するための時空間を考慮した推移を自動的に提案できる。さらに、その X の提案に基づいて工場の制御・運転等を実施した後の実際の Y の結果に基づいて、さらなる Y の改善および Y の最適化を自動的に導くことが可能となる。これにより、**人が先導して最適化を実施する社会から、IoT 機器およびそ**

これから得られるビッグデータが先導して自動的に最適化する社会に変革する。これに付随して、エネルギー消費の低減、環境負荷の低減、工場の省人化も達成される。

2. 研究期間内の達成目標 ※100字以内（句読点含む）

最適化したい対象の理想的な値や時間推移を、IoT 機器から得られる多種大量のデータで構築したモデルに入力することで、それを達成する変数の時空間変化を高精度かつ即時的に出力するシステムを開発すること。（96字）

3. 研究計画とその進め方

本研究提案における研究項目は以下の4つである。

研究項目1 モデルの柔軟性および予測精度の向上

現状の直接的モデル逆解析法では、データにおける変数（ X, Y を含む）が複数の正規分布の重ね合わせで表現されることを仮定しなければならず、**モデルの柔軟性が低い** [3,16]。本研究提案におけるIoT 機器から得られる多種大量のデータにおいては、変数間の関係は非常に複雑であり、さらに変数の時空間情報を扱うため、**複数の正規分布の重ね合わせでは変数間を表現不可能**である。

この問題を解決するための一つの方針は、**変数およびその関係の非線形変換**によりモデルの柔軟性を向上させることである。指数変換、対数変換、ロジット変換、交差項の追加等により、 X を非線形変換する。多種多様な X の非線形変換があるため、変数の数が増大することが問題として考えられる。その際は遺伝的アルゴリズムや遺伝的プログラミング等の最適化手法による非線形変換の最適化で対応する。

もう一つの方針は、直接的モデル逆解析法を**正規分布以外も利用できるように再設計**することである。例えば、対数正規分布・ベルヌーイ分布・二項分布・ポアソン分布・指数分布・ t 分布・カイ 2 乗分布・ガンマ分布・ベータ分布・コーシー分布・ラプラス分布・ロジスティック分布・パレート分布等の確率分布で X と Y の同時確率分布を計算した後、ベイズの定理と確率の乗法定理から、 Y が与えられた際の X の事後確率分布を立式する。多くの確率分布があるため、Expectation-Maximization アルゴリズムや上述した最適化手法を駆使してモデル構築を検討する。

さらに、IoT 環境下においては日々新しいデータが測定されるため、そのデータを用いてモデルを効果的に更新することで、モデルの予測精度を向上させる手法を開発する。

研究項目2 X や Y における時間および空間を考慮したモデル構築

IoT 環境下では X, Y の値は時間的に変化し、 X は Y に対して時間遅れを伴って関係している。例えば製造工場における装置において、原料が入ってから製品が出るまでに時間がかかるため、原料の情報が時間遅れを伴って製品品質に影響する。植物工場における植物の育成においては、水量・栄養素の量・日射量等は時間遅れを伴って成長に寄与し、それらの時間遅れは一定ではない。さらに、 Y の理想的な推移から X を逆解析で予測するためには、 X の時間変化だけでなく Y の時間変化も考慮してモデルを構築することが必須である。また植物の成長においては植物の位置を考慮して Y と X を設定する必要がある。以上のことから、ケモインフォマティクスの分野で開発した GMR や GTMR による直接的モデル逆解析法を、**時空間を考慮できるよう抜本的にアップデート**しなければならない。

X の時間遅れに関しては、申請者のプロセスインフォマティクスでの研究である X の時間遅れの最適化 [5,18,19,28,32,33,34,36] の成果を予備的な知見として有効に活用する。さらに本研究では、X だけでなく Y の時間変化を考慮して生成モデルを構築する手法を開発する (図 6)。

X の位置、Y の位置に関しては、位置の (x, y, z) 座標に基づいた近接関係と、変数間の相関関係との間でモデル化する。これにより新たな位置における Y, X についても予測することが可能となる。X, Y の時間的な関係および位置的な関係を考慮した直接的モデル逆解析法を開発することで、時空間を自由自在に制御可能な最適化手法を実現する。

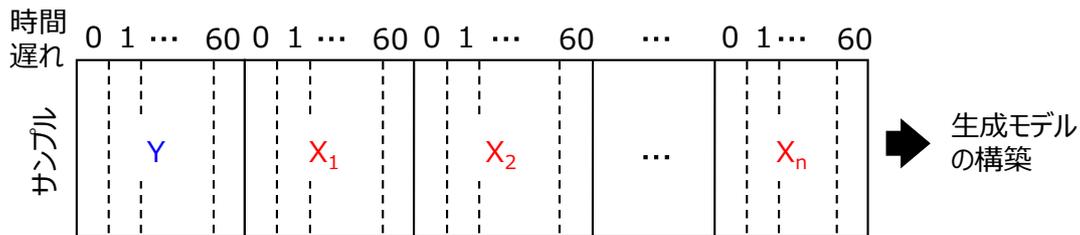


図 6. X, Y の時間変化を考慮した生成モデルの構築

研究項目 3 直接的モデル逆解析の高速化

変数間の関係を複数の正規分布 (研究項目 1 により他の確率分布も考慮可能) の重ね合わせで表現する際、もちろんサンプル数が大きくなるほどモデル構築にかかる時間が大きくなるが、変数の数に対して 2 乗のオーダーで計算時間が増加してしまう。特に IoT 環境下においては多くの変数があり、さらに研究項目 2 において図 6 のように X, Y の時間遅れを考慮すると、非常に多くの変数が存在することになる。直接的モデル逆解析法におけるモデルの構築および逆解析に時間がかかってしまい、達成目標にある即時的な X の時空間変化の提案はまったく不可能である。

本研究では研究項目 2 における時間遅れを考慮するために追加した X の変数の中で、Y と関連のある変数のみを採用することを考える。そのためにモデル構築の際に、正則化項の追加や、変数選択を検討し、変数をスパース化することでモデル構築およびモデルの逆解析を高速化する。

さらに、変数をクラスタリングすることで、関連する変数グループごとに分割し、分割した変数グループごとにモデルを構築する分散モデルを提案する。Y の推移から X の推移を提案する際は、モデルごとに Y の値を入力することで、モデルに対応する X の値が提案され、それを結合することで、X 全体の予測結果が得られる。

研究項目 4 種々の工場において得られるデータへの展開

研究項目 1・2・3 で開発した手法を評価するため、種々の工場において得られるデータを用いて検証を行う。現在も化学工場・植物工場・浄水処理場等でそれぞれセンサー等により実際に計測された時系列データを保持・解析しており、今後も申請者の運営する研究室と共同研究している 12 社の企業から様々な工場等におけるデータを提供いただく予定である。

なお、ビッグデータ解析に関しては「超高速 IoT ビッグデータ解析のための分散アルゴリズム基盤」の研究者である天方大地先生と、時空間データのセンシング技術に関しては「ハイパーモーダル時空間データの超スパース表現」の研究者である田中雄一先生と、欠損値のあるデータに対しては「大規模で不完全なセンサデータに対する高速な最近傍探索」の研究者である松井勇佑先生と、予測精度の

高いモデル構築および即時的なモデルの逆解析について相互触発することにより研究構想を進展させる。

4. 国内外の類似研究との比較、および研究の独創性・新規性

本研究構想は、 Y の理想的な値や時間推移からモデルを逆解析することで、 X の値や時間推移を獲得するものである。ベイズ最適化 (P.IEEE, 104 (2016) 148) をはじめとする現状のモデルの逆解析 (Mol.Inf., 39 (2019) 1900087) では、 X の上限・下限を設定してその範囲内で大量にサンプルを生成し、それらの Y の値を予測した後に良好な予測結果であったサンプルを選択する、すなわちモデルの順解析を効率的かつ網羅的に実施しているにすぎない。また Variational AutoEncoder (arXiv:1312.6114, 2013) や Generative Adversarial Networks (arXiv:1406.2661, 2014) をはじめとする深層生成モデルでも大量にサンプルを生成する必要があり、このような従来の擬似的な逆解析では即時的な X の提案および人の想像を超える最適解の提案はまったく不可能である。本研究の独自性・新規性は Y の値から X の値を直接的に予測することであり、さらに即時的に X の値が得られるといった従来法からの優位性がある。またモデルの逆解析において入力する Y としてその時間推移を扱うことはこれまで皆無であり、さらに X に関しても時間推移を提案する必要があることから、本研究のモデルの逆解析を達成することは挑戦的といえる。

プロセス制御の観点から本研究構想について考える。工場における PID 制御 (J.ProcessContr., 20 (2010) 1220) をはじめとした現状の制御 (Comput.Chem.Eng., 128 (2019) 538) では、 Y の目標値と現在の Y の目標値と実測値との差に基づいて X の値を制御するため、将来の Y の推移に基づいて、今および将来の X を制御することは不可能である。現在・未来だけでなくそれぞれの空間を自在に制御する、まったく新規な本研究提案によって始めて、工場における時空間制御による運転最適化が実現する。ビッグデータを活用したデータ駆動型制御ともいえる独創性があり、現在の制御ではまったく達成不可能であり挑戦的といえる。

5. 研究の将来展望

本研究構想の成果として Y の理想的な時間推移から工場・施設等の変数がどのように時間推移すべきかを即時的に提案できる。本システムは IoT 機器の付加価値になり、IoT 機器の活用が促進すると考える。また、この最適運転の指針に基づいて図 3 の「自動制御(ハードウェア)」を実現する研究が進展することになり、時空間の最適化の実現に向けた新規制御技術等の科学技術の創出に繋がる。

「日本のものづくり」を考えたときに、研究室・実験室でいくら革新的な材料が開発されても、製造工場で生産できないと絵に描いた餅に過ぎない。IoT 機器および本提案による工場最適化の指針、そして自動制御のためのハードウェアの科学技術発展によって、「日本のものづくり」を根本から変革すると確信している。

今回開発する方法論・技術は、一般の方々も利用可能な形にソフトウェア・ツール化し、研究成果および事例と共に研究室のウェブサイト (<https://datachemeng.com/>) や GitHub (<https://github.com/hkaneko1985>) を通じて公開する。さらに学会発表や講演会、そして研究室主催のオンラインサロン (<https://datachemeng.com/onlinesalon/>) で積極的に周知することで、人々の間での利用を促進させる。現状でも 12 社との共同研究をしており、新たな企業を含めた共同研究により本研究のシステムを広く普及させ、さらに実際の工場で活用されることを目指す。