

令和 5 年 6 月 5 日現在

機関番号：14301

研究種目：挑戦的研究（萌芽）

研究期間：2021～2022

課題番号：21K18849

研究課題名（和文）文献を解読し第一原理モデルを自動構築する人工知能の開発

研究課題名（英文）Development of Artificial Intelligence that Automatically builds first-principle models from the literature

研究代表者

加納 学（KANO, Manabu）

京都大学・情報学研究科・教授

研究者番号：30263114

交付決定額（研究期間全体）：（直接経費） 5,000,000円

研究成果の概要（和文）：製造プロセスの設計や運転を最適に行うためには、その第一原理モデル（物理モデル）を構築すべきであるが、モデル構築には膨大な労力が必要となる。そこで本研究では、対象プロセスを指定すると、関連する文献を収集し、文献中の変数や数式の意味を理解し、モデル構築に必要な情報を抽出・編集し、第一原理モデルを自動的に構築する人工知能（AI）の実現を究極目標として、変数とその定義の抽出方法、変数や数式同義性判定方法など必要な要素技術とプロトタイプの開発を実施した。

研究成果の学術的意義や社会的意義

様々な能力を持つ人工知能（AI）が開発され、活用されている。しかし、製造業で決定的に重要でありながらその構築が困難な第一原理モデルを自動的に構築してくれるAIを開発しようという試みは本研究が世界初である。自然言語処理や数式処理の新しい方法論を開発するという点で学術的意義が大きいのみならず、第一原理モデル自動構築AIが完成すれば、製造業の生産性を劇的に向上させることができるため、その社会的意義も計り知れない。

研究成果の概要（英文）：To optimize the design and operation of a manufacturing process, a first-principle model (physical model) of the process plays a crucial role. However, building such a model requires an enormous effort. The ultimate goal of this research is to realize an artificial intelligence (AI) that automatically builds a first-principle model by collecting relevant literature, understanding the meaning of variables and formulas in the literature, and extracting and editing information necessary for model construction, given a target process. We developed several elemental technologies, including a method for extracting variables and their definitions and a method for determining the equivalence of variables and formulas, and also we developed a prototype of AI.

研究分野：プロセスシステム工学

キーワード：人工知能 第一原理モデル 物理モデル 自然言語処理 数式処理 化学工学 プロセスシステム工学

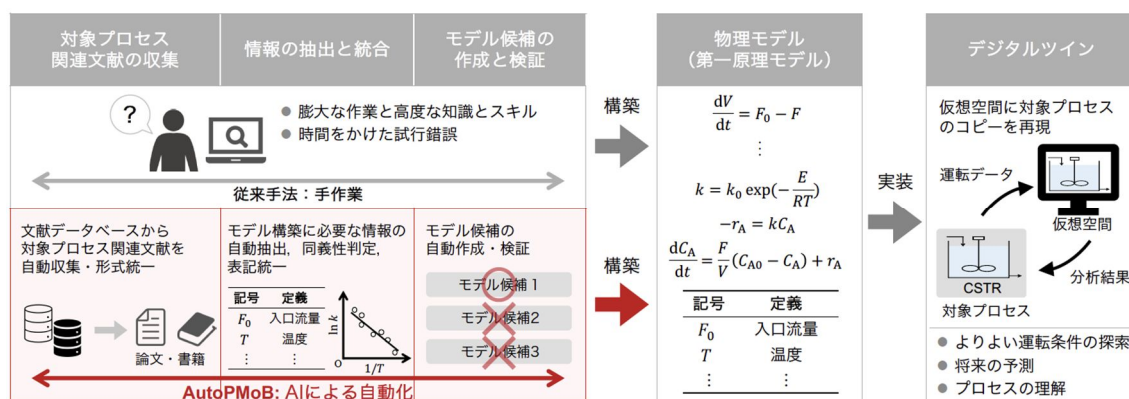
### 1. 研究開始当初の背景

申請者は、プロセスシステム工学分野において、化学、製薬、鉄鋼、半導体など様々な産業プロセスを対象に、オンライン計測が困難な製品特性等を推定する仮想計測技術(ソフトセンサー)や運転支援システムを開発・実用化してきた。これらの研究開発では、対象プロセスの正確なモデルを構築することが成功の鍵を握る。近年、機械学習が人気を博し、統計モデルの構築が極めて活発に行われているが、様々な外乱や異常をも想定し、真に生産現場で活用できるモデル(デジタルツイン)を構築するには、現象論に基づく第一原理モデル(物理モデル)を中核に置くべきとの想いを強くしている。それは、物理モデルと統計モデルを統合するグレイボックスモデルの研究開発を通して、データに依存することの危険性と限界、そして物理モデルを用いることで実現されるシステムの頑健性を目の当たりにしてきたからである。しかし、物理モデルの構築には大変な手間がかかる。その手間を削減し、目的に合致した物理モデルを誰もが構築できるようにしなければ、結局、プロセス産業でデジタルツインは活用されないのではないかと、DXによる生産現場の革新も実現できないのではないかと考え、物理モデル自動構築 A I (AutoPMoB: Automated Physical Model Builder) の開発に着手した。

### 2. 研究の目的

製造業における生産性向上の柱として、デジタルトランスフォーメーション(DX)に大きな期待が寄せられている。深層学習を含む機械学習が様々な分野で革新を起こすなか、データ活用に関心が集中しているが、次世代生産技術の中核を担うデジタルツインの実現に向けて、データ取得が困難な状況でもプロセスの挙動を正確に予測するために、科学法則に基づく第一原理モデル(物理モデル)が必要になる。実際、申請者は数々の社会実装を通して、データに依存することの限界、第一原理モデルに基づくシステムの頑健性を痛感してきた。ただ、現象の理解を前提とする第一原理モデルの構築、特に実プロセスを正確に表現可能なモデルの構築は、専門家が試行錯誤を重ねながら取り組む必要があり、極めて難しく高コストである。さらに、文献情報処理を伴う人工知能(AI)開発において、自然言語とは対照的に、数式を対象とした研究は極めて少なく、実用レベルに達している成果もほとんどない。

本研究では、対象プロセスに関連する文献を収集し、文献中の変数や数式の意味を理解し、モデル構築に必要な情報を抽出・編集し、物理モデルを自動的に構築する人工知能(AI)のプロトタイプを開発し、その実現可能性を世界に先駆けて示すことを目的とする。



### 3. 研究の方法

ほとんどの文献はウェブ上の論文データベースから入手できるため、それらを文献情報として利用する。その上で、研究目的を達成するため、以下に示す要素技術を開発し、それらを統合して第物理モデル自動構築 A I のプロトタイプを開発する。なお、A I の構築と検証に不可欠なコーパス (事前に論文を解析して用意した正解データの集合) も作成する。

A) 変数とその意味、および数式を抽出する。

文献には、変数を表す記号とその意味、変数間の関係を表す数式が記載されている。既存技術では変数の記号を抽出できても変数の意味は抽出できない。そこで、文中の位置関係や、変数の意味が記号に反映されやすい事実 (Pressure を表す変数の記号や添字には P が使われやすい) などを活用して、変数とその意味、および数式を高精度に抽出する方法を開発する。

B) 変数および数式と同義性を判定する。

論文によって変数を表す記号や説明は異なるため、変数の意味を理解し、同一の変数か否かを判定する必要がある。そこで、記号の周辺で用いられている単語や節を比較することで同一の変数であるかを判断する方法を開発する。数式も論文によって表記の仕方が異なる。そこで、マークアップ言語で書かれた数式を木構造で表現し、それを統一した表記ルールによって数式に変換することで、同義性を判定する方法を開発する。

C) 物理モデルを自動で構築する。

無駄のない物理モデルを導出するため、モデルに含めるべき変数を与えたときに必要十分な数式群を判断する方法を開発する。さらに、複数の文献に記載されている数式を関連付け、物理モデルを自動で構築する方法を開発する。このとき、優先順位を付して複数のモデルを提示させる。なお、本研究期間ではモデル構造の決定までを対象とし、実験データや操業データを用いたパラメータフィッティングは含まない。

D) 物理モデル自動構築 A I のプロトタイプを作成する。

最終的に、ユーザがモデルに含めたい変数を指定すれば、キーワード検索で文献を収集し、指定した変数を含む第一原理モデルを出力する物理モデル自動構築 A I ( AutoPMoB: Automated Physical Model Builder ) のプロトタイプを作成する。

#### 4. 研究成果

物理モデルを自動的に構築する人工知能 ( A I ) のプロトタイプを開発し、その実現可能性を世界に先駆けて示すために、必要となる各種要素技術を開発した。

1) 大規模言語モデルである BERT を用いて化学プロセス関連文献から高精度に変数の意味を抽出する方法を開発した。

2) 化学プロセス関連文書に特徴的な変数の使用方法の傾向を明らかにし、それを用いた変数抽出手法を開発した。

3) 関連する約 80 万報の論文を用いて言語モデル ProcessBERT2022 を構築し、文書中に現れる変数の同義性を判定する方法を開発した。

4) 文献から抽出した 2 つの数式群をある記号について数式を解くことで数式群の同義性を判定する方法を開発した。

これらの要素技術開発の成果を結集して、化学工学で重要なプロセスに注目した物理モデル自動構築 A I ( AutoPMoB ) のプロトタイプを作成した。

今後さらに取り組むべき問題が明らかになったため、それらを解決する必要はあるが、第一原理モデル自動構築 A I のプロトタイプ開発という目的を達成し、その実現に向けて大きく前進した。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計35件（うち招待講演 0件 / うち国際学会 12件）

1. 発表者名 加藤祥太
2. 発表標題 化学工学分野の専門用語を理解するBERTモデルの開発
3. 学会等名 NLP若手の会第17回シンポジウム (YANS2022)
4. 発表年 2022年

1. 発表者名 加藤祥太
2. 発表標題 変数消去順序を考慮した数式群同義性判定手法
3. 学会等名 化学工学会第53回秋季大会
4. 発表年 2022年

1. 発表者名 加藤祥太
2. 発表標題 化学工学分野の論文に含まれる命名法に基づく変数記号および定義の解析
3. 学会等名 言語処理学会第29回年次大会 (NLP2023)
4. 発表年 2023年

1. 発表者名 張純朴
2. 発表標題 数学的表現の構造的情報のトークン化手法によるProcessBERTの性能改善
3. 学会等名 言語処理学会第29回年次大会 (NLP2023)
4. 発表年 2023年

1. 発表者名 山本 時志
2. 発表標題 二段階のファインチューニングを行ったBERTによる変数定義抽出
3. 学会等名 言語処理学会第29回年次大会 (NLP2023)
4. 発表年 2023年

1. 発表者名 加藤祥太
2. 発表標題 化学工学関連文書の変数を理解するための言語モデル構築手法
3. 学会等名 化学工学会第88年会
4. 発表年 2023年

1. 発表者名 加藤祥太
2. 発表標題 化学工学関連論文中の命名法に基づく変数定義予測手法
3. 学会等名 情報処理学会 第255回自然言語処理研究会
4. 発表年 2023年

1. 発表者名 Shota Kato
2. 発表標題 Annotation Tool for Variable Extraction from Documents on Manufacturing Processes
3. 学会等名 The 10th Asian Symposium on Process Systems Engineering (PSE Asia 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Shota Kato
2. 発表標題 Variable Definition Extraction from Documents on Chemical Processes Utilizing Semantic Information on Variable Symbols and Definitions
3. 学会等名 The 10th Asian Symposium on Process Systems Engineering (PSE Asia 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Shota Kato
2. 発表標題 Efficient physical model building algorithm using equations extracted from documents
3. 学会等名 33rd European Symposium on Computer-aided Process Engineering (ESCAPE33) (国際学会)
4. 発表年 2023年

1. 発表者名 Shota Kato, Manau Kano
2. 発表標題 Efficient physical model building algorithm using equations extracted from documents
3. 学会等名 33rd European Symposium on Computer-aided Process Engineering (ESCAPE33) (国際学会)
4. 発表年 2023年

1. 発表者名 Masaki Numoto, Shota Kato, Manabu Kano
2. 発表標題 Variable Definition Extraction from Documents on Chemical Processes Utilizing Semantic Information on Variable Symbols and Definitions
3. 学会等名 The 10th Asian Symposium on Process Systems Engineering (PSE Asia 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Shota Kato, Manabu Kano
2. 発表標題 Annotation Tool for Variable Extraction from Documents on Manufacturing Processes
3. 学会等名 The 10th Asian Symposium on Process Systems Engineering (PSE Asia 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Chunpu Zhang, Shota Kato, Manabu Kano
2. 発表標題 Equivalence Judgment of Equation Groups Representing Process Dynamics
3. 学会等名 The 14th International Symposium on Process Systems Engineering (PSE 2021+) (国際学会)
4. 発表年 2022年

1. 発表者名 Shota Kato, Manabu Kano
2. 発表標題 Towards An Automated Physical Model Builder: CSTR Case Study
3. 学会等名 The 14th International Symposium on Process Systems Engineering (PSE 2021+) (国際学会)
4. 発表年 2022年

1. 発表者名 Shota Kato, Kazuki Kanegami, Manabu Kano
2. 発表標題 ProcessBERT: A Pre-trained Language Model for Judging Equivalence of Variable Definitions in Process Models
3. 学会等名 13th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems (DYCOPS 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 加藤 祥太, 加納 学
2. 発表標題 化学工学関連論文での命名法と数式に基づく変数定義予測手法
3. 学会等名 第37回人工知能学会全国大会
4. 発表年 2023年

1. 発表者名 加藤 祥太, 加納 学
2. 発表標題 化学工学関連論文での命名法に基づく変数定義予測手法
3. 学会等名 第255回自然言語処理研究会
4. 発表年 2023年

1. 発表者名 加藤 祥太, 張 純朴, 山本 蒔志, 加納 学
2. 発表標題 化学工学関連文書の変数を理解するための言語モデル構築手法
3. 学会等名 化学工学会第88年会
4. 発表年 2023年

1. 発表者名 山本 蒔志, 加藤 祥太, 加納 学
2. 発表標題 二段階のファインチューニングを行ったBERTによる変数定義抽出
3. 学会等名 言語処理学会第29回年次大会
4. 発表年 2023年



1. 発表者名 張 純朴, 加藤 祥太, 加納 学
2. 発表標題 数学的表現の構造的情報のトークン化手法によるProcessBERT の性能改善
3. 学会等名 言語処理学会第29回年次大会
4. 発表年 2023年

1. 発表者名 加藤 祥太, 加納 学
2. 発表標題 化学工学分野の論文に含まれる命名法に基づく変数記号および定義の解析
3. 学会等名 言語処理学会第29回年次大会
4. 発表年 2023年

1. 発表者名 加藤 祥太, 張 純朴, 加納 学
2. 発表標題 変数消去順序を考慮した数式群同義性判定手法
3. 学会等名 化学工学会第53回秋季大会
4. 発表年 2022年

1. 発表者名 加藤 祥太, 加納 学
2. 発表標題 化学工学分野の専門用語を理解するBERTモデルの開発
3. 学会等名 NLP若手の会第17回シンポジウム
4. 発表年 2022年

1. 発表者名 加藤 祥太
2. 発表標題 Language model for understanding variables in documents on chemical engineering
3. 学会等名 NII MathNL 2nd Seminar
4. 発表年 2023年

1. 発表者名 加藤 祥太
2. 発表標題 Towards Automated Physical Model Builder
3. 学会等名 NII MathNL 1st Seminar
4. 発表年 2022年

1. 発表者名 加藤祥太
2. 発表標題 製造プロセスの物理モデルを文献から自動生成するAIの開発
3. 学会等名 NLP若手の会第16回シンポジウム
4. 発表年 2021年

1. 発表者名 金上和毅
2. 発表標題 複数文献中の変数の同義性判定手法の開発
3. 学会等名 NLP若手の会第16回シンポジウム
4. 発表年 2021年

1. 発表者名 張純朴
2. 発表標題 計算機代数システムを用いた数式群の同義性判定手法
3. 学会等名 NLP若手の会第16回シンポジウム
4. 発表年 2021年

1. 発表者名 加藤祥太
2. 発表標題 物理モデル自動構築に向けて:1)変数アノテーションツールの開発
3. 学会等名 化学工学会第87年会
4. 発表年 2022年

1. 発表者名 加藤祥太
2. 発表標題 物理モデル自動構築に向けて:2)ProcessBERT化学工学のための事前学習言語モデル
3. 学会等名 化学工学会第87年会
4. 発表年 2022年

1. 発表者名 加藤祥太
2. 発表標題 物理モデル自動構築AIの実現に向けたProcessBERTの構築
3. 学会等名 第9回計測自動制御学会制御部門マルチシンポジウム
4. 発表年 2022年

1. 発表者名 Shota Kato
2. 発表標題 ProcessBERT: A Pre-trained Language Model for Judging Equivalence of Variable Definitions in Process Models
3. 学会等名 The 13th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems (DYCOPS) (国際学会)
4. 発表年 2022年

1. 発表者名 Shota Kato
2. 発表標題 Towards an automated physical model building system: CSTR Case Study
3. 学会等名 The 14th International Symposium on Process Systems Engineering (PSE2021+) (国際学会)
4. 発表年 2022年

1. 発表者名 Chunpu Zhang
2. 発表標題 Equivalence judgment of equation groups representing process dynamics
3. 学会等名 The 14th International Symposium on Process Systems Engineering (PSE2021+) (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	加藤 祥太  (KATO Shota)  (60883402)	京都大学・情報学研究科・助教    (14301)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------