

令和 6 年 5 月 23 日現在

機関番号：12608

研究種目：研究活動スタート支援

研究期間：2022～2023

課題番号：22K21302

研究課題名（和文）記号論理と深層機械学習の統合による新たな知識を発見するAIに関する研究

研究課題名（英文）Discovering New Knowledge by Combining Symbolic Logic and Deep Learning

研究代表者

Phua Yin Jun (Phua, Yin Jun)

東京工業大学・情報理工学院・助教

研究者番号：20963747

交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：本研究では、深層学習を用いて新たな人間の理解可能な知識を出力する手法について研究を行った。まず、アンサンブルと呼ばれる複数のモデルを同時学習させる手法を用いて、モデルから捉えられる重要な要素の安定性を向上させた。次に、知識抽出から開発されたデータ生成の手法を、生成AIを用いて増分学習の分野での応用を行った。これにより、知識抽出の手法は記号推論だけでなく一般的な深層学習手法でも応用が可能であることを示した。さらに、既存の記号知識に存在するノイズに対する頑健なグラフニューラルネットワークモデルを提案した。

研究成果の学術的意義や社会的意義

学術的意義として、実験データから重要な要素を抽出する基盤技術に貢献することができた。また、記号推論でよく使われる手法として、学習データを生成する手法は深層機械学習にも応用が可能であることを示した。さらに、生成AIの学術的応用を示すこともできた。社会的には、ノイズに対して頑健な手法を提案することで、実世界のデータをそのまま応用することが可能となる技術の開発に貢献した。本研究で開発した手法をさらに展開させることにより、実験データから新たな知識を抽出することができるAI技術へつながることが考えられる。

研究成果の概要（英文）：This study focused on methods for discovering new human-readable knowledge by utilizing deep learning methods. First, we proposed a method that utilizes ensemble, where multiple models are trained, to stabilize the results for extracting features that are considered important by the models. Next, by utilizing a method to generate training data, which has been proposed for symbolic knowledge extraction methods, is applied to using generative AI in the incremental learning field. With this, we show that methods proposed for symbolic knowledge extraction can also be generalized to deep learning methods. Furthermore, we also proposed graph neural network models that are robust to noise that are inherent in graphs.

研究分野：知能情報

キーワード：深層学習 ニューロシンボリック 記号推論 知識発見

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

近年、自然科学の応用において深層機械学習技術を用いてブレークスルーを果たしたが、その本質となるメカニズムの解明に役立っていない。例えば、Deepmind の AlphaFold [Jumper ら 2021] は数十年に渡り解決が難しいと思われていたタンパク質のフォールディング問題において、深層機械学習を用いた手法で、これまでの手法よりも優れた結果を得た。しかし、AlphaFold はタンパク質の構造を高精度で予測することができて、その本質となるタンパク質のフォールディングのメカニズムを解明することに役に立たなかった[Al-Janabi 2022]。

そして、深層機械学習モデルは解釈可能でないため、結果が説明できない問題や、潜んでいるバイアスや未知ケースでの脆弱性の発見および修正が難しい。深層機械学習モデル自体は、大量の数字が並べられ様々な演算を経て結果を出している。各数字がどんな働きをするか説明する手法もあるが、新しいモデルなどに応用する時は都度手法を開発する必要がある。さらに、学習済みのモデルがバイアスを持っている時や、特定のケースだけ弱い時に修正を行うのはとても難しい。一方、記号論理に基づいた機械学習技術は、結果の説明ができ、脆弱性などの修正も可能である。しかし、ノイズに対する頑健性や実世界データの応用は依然として課題となっている。

### 2. 研究の目的

以上の背景から、本研究の目的は、新たな知識を自動的に発見できる人工知能技術の基盤を研究することである。深層機械学習手法を用いて、大量のデータから新たなパターンや因果関係を発見することができる。その発見した新たな知識を記号的に記述することで人間の知識にもなる。申請者はそんな人類の知識に貢献できる手法の開発を目指している。

### 3. 研究の方法

上記の目的を達成するため、以下の方法で研究を実施した。

(1) 実世界データから重要な役割を担う要素を抽出する手法を開発する。具体的には、教師なし学習を用いたオートエンコーダを利用し、縮小した隠れ表現に制限をかけることで重要な要素を学習するよう仕向ける手法を開発した。また、アンサンブルと呼ばれる、複数のモデルを同時に学習させる手法を実装したことで、抽出される要素の安定性を向上させた。本手法の有効性を示すため、生物学の実データである細胞配列データに応用し、その有効性を調べた。

(2) 生成 AI を用いて学習データを生成し、深層学習モデルの破滅的忘却問題に対処する手法を開発した。深層学習モデルが、別のタスクの学習を行うことでその前のタスクで得られた知識がすべて失う現象を破滅的忘却と呼ぶ。記号論理に基づいた記号的手法では、このような破滅的忘却問題が現れない。生成 AI を用いてクラス増分学習の新たな手法を提案した。生成 AI から生成した学習データは、過去のクラスの記憶再生だけでなく、過去の知識を蒸留するためにも利用する手法を確立した。

(3) ノイズに対して頑健なグラフニューラルネットワークモデルを開発した。グラフニューラルネットワークは、記号的なグラフを深層学習で学習するニューロシンボリック手法のひとつとして知られている。グラフに潜在するノード特徴量およびエッジのノイズ、両方に対して同時

に頑健な手法を確立した。ノイズの入ったグラフに対して、一部ノードやエッジを除いて対照的学習を行なった。手法の有効性を実グラフで示した。

#### 4. 研究成果

(1) 提案手法を細胞配列データに応用したことで、より安定した要素を抽出することができることを確認した。それを図1に示す。

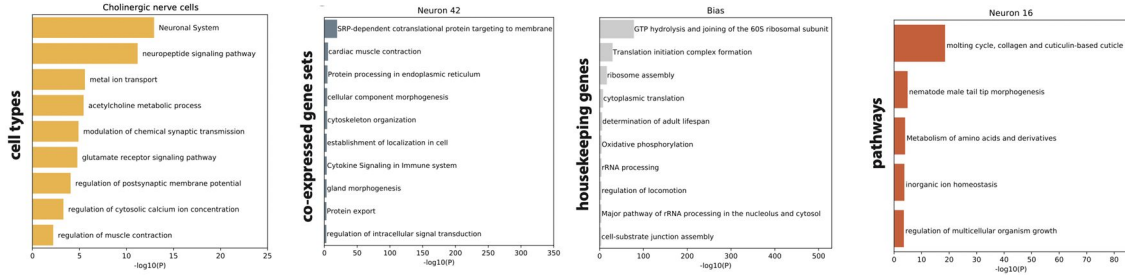


図1 細胞配列データから抽出した重要な遺伝子

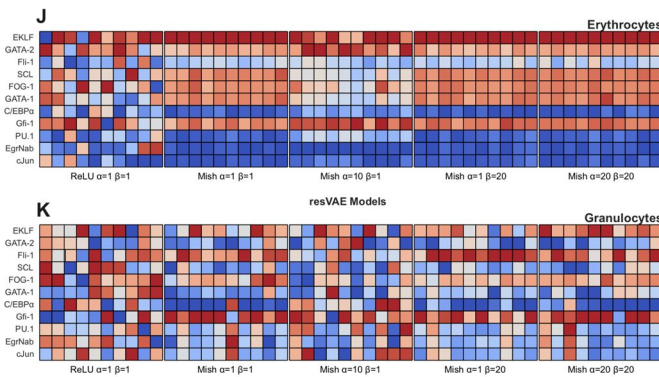


図2 抽出した要素の安定性の向上

図2では、従来手法の抽出したもののKに比べ、提案手法の抽出した重要な要素Jが安定していることがわかる。



図3 抽出した要素がより重要である

図3では、従来手法が抽出した要素(右)に比べ、提案手法が抽出したもの(左)がより有意義であり、直感的なものではないことを示している。

(2) クラス増分学習において生成AIを組み込んだ手法を初めて提案した。その成果が認められ ICCV 2023 Workshop にて最優秀論文賞を受賞した。

(3) ノイズに頑健なグラフニューラルネットワークモデルは従来、ノード特徴量あるいはエッジノイズのどちらかのみ対応ができた。しかし、実グラフでは両方同時に潜在していることが一般的であり、それに対して頑健な手法を提案することができた。これにより、実世界での応用がさらに広がることが期待される。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 1件 / うちオープンアクセス 1件）

1. 著者名 Ten Foo Wei, Yuan Dongsheng, Jabareen Nabil, Phua Yin Jun, Eils Roland, Lukassen Soren, Conrad Christian	4. 巻 11
2. 論文標題 resVAE ensemble: Unsupervised identification of gene sets in multi-modal single-cell sequencing data using deep ensembles	5. 発行年 2023年
3. 雑誌名 Frontiers in Cell and Developmental Biology	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fcell.2023.1091047	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Quentin Jodelet
2. 発表標題 Class-Incremental Learning using Diffusion Model for Distillation and Replay
3. 学会等名 1st Workshop on Visual Continual Learning, ICCV2023（国際学会）
4. 発表年 2023年

1. 発表者名 Tai Hasegawa
2. 発表標題 DEGNN: Dual Experts Graph Neural Network Handling both Edge and Node Feature Noise
3. 学会等名 Pacific-Asia Conference on Knowledge Discovery and Data Mining 2024（国際学会）
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

受賞  
1. Jodelet, Q., Liu, X., Phua, Y. J. and Murata, T.: Best paper award. 1st VCL Workshop in ICCV 2023.

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
--	---------------------------	-----------------------	----

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		
ドイツ	Berlin Institute of Health (BIH)	Charite - Berlin University of Medicine	University of Heidelberg