

未来社会創造事業 探索加速型
「超スマート社会の実現」領域
終了報告書(探索研究)

令和2年度
終了報告書

[研究開発代表者名:石川 冬樹]

[情報・システム研究機構 国立情報学研究所 准教授]

[研究開発課題名:機械学習を用いたシステムの高品質化・実用化を加速する
"Engineerable AI"技術の開発]

実施期間 : 平成30年11月15日～令和2年12月31日
(令和2年度より研究開発課題名「高信頼な機械学習応用システムによる価値創造」を
再編し継続)

§ 1. 研究実施体制

(1)「NII」グループ(情報システム研究・機構 国立情報学研究所)

1. 研究開発代表者:石川 冬樹 (情報システム研究・機構 国立情報学研究所 アーキテクチャ科学研究系、准教授)

2. 研究項目

- ・自動車業界およびソフトウェア開発業界における課題調査
- ・深層学習自動デバッグツールの試作・評価

(2)「東工大」グループ(東京工業大学)

1. 主たる共同研究者:鈴木 賢治 (東京工業大学 科学技術創成研究院 未来産業技術研究所、特任教授)

2. 研究項目

- ・少数データ学習可能な深層学習モデルの検討
- ・少数データモデルの診断支援 AI への応用

(3)「阪大」グループ(大阪大学)

1. 主たる共同研究者:松原 崇 (大阪大学 基礎工学研究科、准教授)

2. 研究項目

- ・システム論と機械学習の融合による制御可能な意思決定システムの構築
- ・ベイズ的深層学習を用いた小規模データ解析技術の開発

(4)「九大」グループ(九州大学)

1. 主たる共同研究者:馬 雷(九州大学 システム情報科学研究院、准教授)

2. 研究項目

- ・機械学習モデルデバッグにおける分布外検知の活用可能性検討
- ・深層学習自動デバッグツールの試作・評価
- ・自動運転におけるアーキテクチャと信頼性に関する検討

(5)「早大」グループ(早稲田大学)

1. 主たる共同研究者:鷺崎 弘宜 (早稲田大学 基幹理工学部情報理工学科、教授)

2. 研究項目

- ・高信頼な機械学習応用システムの設計上の頻出の課題およびその実践的な解決策を抽象化した機械学習工学アーキテクチャ・デザインパターン集
- ・機械学習工学アーキテクチャ・デザインパターン集の認知や活用実態およびプロセスモデルへの組み入れを見据えたモデル化

§ 2. 研究実施の概要

機械学習技術を用いた AI（以後単に AI）の実用化が盛んに追求されている。機械学習の利点は、データセットを与え訓練することにより、認識や予測など明確な仕様定義ができないファジーな機能を実現できることである。特に深層学習技術の発展により、画像認識をはじめとして、従来は実現が困難であった機能が高精度で実現できるようになり、幅広い領域での産業応用が進もうとしている。

一方で、現状の AI に対しては、その品質、特に安全性や信頼性に対して大きな懸念がある。その原因は、大量データから機能全体を大まかに導くというアプローチにある。大量データへの依存性、および機能の制御困難性から、細やかなニーズに対応するように AI を構築したり修正したりすることができない。AI を用いた製品・サービスをニーズに応じ仕立て上げるための工学（Engineering）が確立されていないといえる。

本研究開発では新たに“Engineerable AI (eAI)” というビジョンを提唱し、AI システムの開発・品質保証・運用を安定化・効率化することで、その高度化および実用化を大きく加速させる。上述のように、大量データへの依存性および制御困難性を現状の AI の根本的な課題ととらえ、これらに対応する技術に取り組む。

2020 年度の探索研究においては、2019 年度まで同様な課題に取り組んできた 3 つの探索研究の合併・再編を進めつつ、事業統括会議からのフィードバックを踏まえつつ、実務者との協働により課題の明確化と技術の実証に取り組んだ。

第一に、医療分野の実務者との議論を通し、専門医不足や医療の質のばらつきという社会問題に対し、従来技術では大量データが得られない非典型的な病変を高信頼に検出するような AI の構築はできないという課題に着目した。この課題に対応する POC を設定するとともに、技術の初期実証に取り組んだ。この初期実証では、比較的少数データで学習可能な深層学習技術を検討し、比較的少数データにより高信頼な医療診断 AI が構築できる可能性を示した。

第二に、交通分野の実務者との議論を通し、自動運転において、安全性向上のために AI を改善しようとしても、変更により AI 全体の動作を意図した形で修正できないために、安全性を効率的・安定的に高め続け保証していくことができないという課題に着目した。この課題に対応する POC を設定するとともに、技術の初期実証に取り組んだ。この初期実証では、プログラム自動修正技術を深層学習モデルに適合させることにより、他の認識対象に対する性能劣化をおさえつつ、特定の認識対象に対する認識性能を向上できることを示した。

以上の取り組みに基づき、本格研究の成果が社会にもたらす効果を最大化するよう POC および研究開発の方向性を定めるとともに、そのフィジビリティを示すことができた。