

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  
年次報告書(探索研究)

H30 年度 研究開発年次報告書
---------------------

平成 30 年度採択研究開発代表者

[研究開発代表者名：潮 俊光]

[大阪大学大学院基礎工学研究科・教授]

[研究開発課題名：形式手法を用いたデータ駆動階層型管理システムの設計]

実施期間：平成 30 年 11 月 15 日～平成 31 年 3 月 31 日

## §1. 研究開発実施体制

### (1) 潮グループ(大阪大学)

- ① 研究開発代表者: 潮 俊光 (大阪大学基礎工学研究科、教授)
- ② 研究項目
  - ・シングルエージェントの管理エージェントの DMD
  - ・ユースケースの探索
  - ・データ駆動階層型管理システム実装における問題点の検討

### (2) 松原グループ(神戸大学)

- ① 主たる共同研究者: 松原 崇 (神戸大学システム情報学研究科、助教)
- ② 研究項目
  - ・分散強化学習の協調動作に向けた意思決定のモデル化

## §2. 研究開発実施の概要

時相論理で書かれた仕様を満たす方策を制御対象のデータから学習する階層型強化学習法を開発した。時相論理を満たす振る舞いのみを受理する一般化ラビンオートマトンの受理条件の構造に対応させて3階層から構成される。一般化ラビンオートマトンでは、複数の受理条件があり、いずれの受理条件も有限回しか訪問できない状態集合と無限回訪問する状態集合の対で記述されている。この受理条件の特徴を利用し、最上位層では採用する受理条件の学習、中間層では訪問の順番の学習、そして最下位層では各時刻でのアクション選択の学習を行う。この3階層からなる学習によって学習の完全な自動化が達成できた。また、ネットワークによって制御対象と制御器との間にデータ転送遅延がある場合、この遅延は確率的に変動するので、ネットワークを利用した強化学習は失敗することがあった。この問題を解決するために、転送中のデータも状態と見なした拡張状態を定義し、それをを用いた強化学習法を提案した。ネットワークでの遅延に変動があっても学習できることをシミュレーションによって確認した。

階層的なシステムにおいては、エージェントは上位エージェント、同階層内のエージェント、そしてシステムの管理下でない人間エージェントとも協調動作をする必要がある。そこで、深層学習を用いて強化学習に模倣学習や逆強化学習を組み合わせることにより、人間エージェントの意思決定や目的をデータ駆動的に学習してモデル化し、他のエージェントと統一的に扱える土台を開発する。本研究項目では、強化学習と人間の模倣学習を組み合わせることで、性能と人間らしい振る舞いを両立するような学習法を提案した。離散行動空間のゲーム **Gopher** 及び自動車運転シミュレータ **Torcs** において、提案手法を検証し、強化学習よりも感性評価において人間らしいと判定される動き、そして人間エージェントの単なる模倣よりも高い性能を得ることができた。