

ルールベース型エージェントフレームワークにおける 学習エージェント設計支援機構の拡張

日比野 雅人†

打矢 隆弘‡

内匠 逸‡

†名古屋工業大学 工学部 情報工学科

‡名古屋工業大学 大学院 工学研究科

〒466-8555 愛知県名古屋市昭和区御器所町

〒466-8555 愛知県名古屋市昭和区御器所町

1 はじめに

学習エージェントとは、環境から得られる利益を大きくするために、過去の行動を学習して次の行動を決定するエージェントである。エージェントは、マルチエージェント環境下ではエージェント間で行動を阻害しないことが望ましい。これを実現する学習方法として Nash-Q 学習がある。しかし、エージェントの運用・開発に必要なフレームワークに、Nash-Q 学習を支援するものは存在しないため、利用するには開発者が一から実装しなければならない。そこで、本研究ではエージェントフレームワーク DASH[1] における Nash-Q 学習エージェントの開発支援機構を提案する。

2 DASH

DASH フレームワークにおけるエージェントは推論機構(図 1)と呼ばれる、if-then 型のルールを用いて行動を決定する仕組みを持つ。推論機構は推論エンジン、ワーキングメモリ、ルール集合から構成される。推論エンジンがワーキングメモリを参照し、マッチするルールを検索・実行することで動作する。

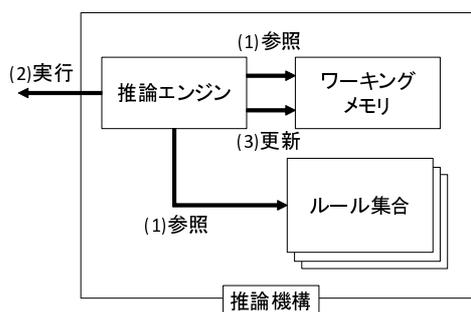


図 1: DASH エージェントの推論機構

3 先行研究

先行研究 [2] では、自動学習機能を DASH に付加した。自動学習機能は以下の機構によって実現されている。

学習データベース

ワーキングメモリとルールの組に付加されたルール優先度*を保存するデータベース。

行動選択エンジン

ルール優先度を基に次の行動を選択するエンジン。従来の DASH はファーストマッチで行動を選択していたが、ルール優先度を参照し行動選択することで学習性を実現する。

学習エンジン

実行結果から優先度を更新するエンジン。優先度を更新する手法として Q-Learning(以下 QL) と Profit Sharing(以下 PS) が利用可能である。

この機構を利用することで、フレームワーク側からルールの優先度を設定・更新することが可能となり、学習エージェントを開発する際にフレームワークから支援を受け取ることが可能になった。本研究では、先行研究で DASH に実装された機構を、Nash-Q 学習を用いることができるよう拡張する。

続いて、QL と PS の特徴を述べる。

Q-Learning

マルコフ決定過程†において最適性がある。よってシングルエージェント環境にて有用性がある。

Profit Sharing

最適解への収束は保障されないが、マルチエージェント環境に適用可能。

QL・PS は自分の利益を最大化する学習方法である。しかし、他のエージェントの行動を考慮することができず、マルチエージェント環境下では他のエージェントと互いに行動を阻害しあう可能性がある。

Expansion of Learning Agent Design Support Mechanism for Rule-based Agent Framework

†Masato HIBINO ‡Takahiro UCHIYA ‡Ichi TAKUMI

†School of Engineering, Nagoya Institute of Technology

‡Graduate School of Engineering, Nagoya Institute of Technology

*そのルールを実行する価値を示すパラメータ。

†現状態と選択した行動により次状態が決定するモデル。

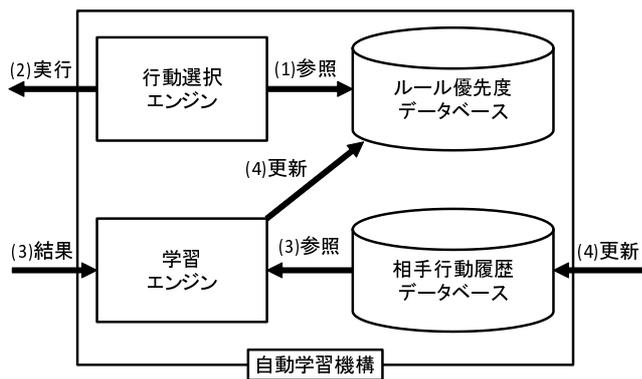


図 2: 提案した自動学習機能

4 Nash-Q 学習

Nash-Q 学習は、マルチエージェント環境においても、他エージェントとのナッシュ均衡*を学習することで最適性をもたせるために考案された学習方法である [3]。相手のエージェントの行動を蓄積していき、現状からのナッシュ均衡をこれまでの履歴から予想し、そこから逆算して相手の行動を予測することで、マルチエージェント環境をマルコフ決定過程に近づけることを目指す。これにより、各エージェントがお互いの行動に干渉しないような行動の学習をすることができる。学習は以下の流れで行う。

(1) 優先度を基に行動を選択・実行

優先度を参照し、優先度が高い行動を優先して実行する。

(2) 相手の行動と、自身が得られる報酬を観測

環境状態と相手の行動を対応付けて保存する。

(3) 相手の行動履歴から次の相手の行動を予測

蓄積したデータから相手の傾向を学習し、次の行動を予測する。

(4) 予測した行動に対するナッシュ均衡を導出

他のエージェントと干渉しあわないためにナッシュ均衡を利用する。

(5) 優先度を変更し、(1)へ

ナッシュ均衡になると思われる行動の優先度を高く更新する。

5 提案機構

本研究で提案する Nash-Q 学習エージェントの開発支援機構について述べる。

* 自他共に行動を変える動機のない行動の組み合わせ

5.1 概要

本研究では、他エージェントの行動履歴を管理する機構と、エージェントの持つ各ルールに対して Nash-Q 学習を用いて自動的に優先度を付加する機構を導入する。これにより、Nash-Q 学習を用いたルール優先度の自動決定を実現する (図 2)。拡張した自動学習機構を導入したシステム全体の流れは以下のようになる。

1. 推論エンジンが現在のワーキングメモリを参照
2. 検索されたルールを行動選択エンジンへ送信
3. 行動選択エンジンが優先度を参照し、選択
4. 選択したルールを推論エンジンに送信
5. ルールを実行
6. ワーキングメモリを更新
7. 環境から得られた報酬と相手の行動履歴から、実行したルールの優先度を更新
8. 相手の行動を取得して (1) へ戻る

5.2 提案機構の利点

提案機構を導入することによって、Nash-Q 学習エージェントを開発・運用する際に、フレームワークからの支援を受け取ることができる。これによって、開発者がエージェントに Nash-Q 学習を実装する負担を減らすことができる。また、Nash-Q 学習を利用することで先行研究では不可能だった、相手に干渉しない行動を学習することが可能となる。

6 まとめ

マルチエージェント環境下で有効である、Nash-Q 学習エージェントの開発支援機構を DASH に付加することを提案した。今後は、提案した機構の実装を進め、有用性の検証のために評価実験を行う。

参考文献

- [1] “DASH ユーザマニュアル”, <http://uchiya.web.nitech.ac.jp/idea/html/>
- [2] 板津呂 翔, 打矢 隆弘, 内匠 逸, 木下 哲男, “知的マルチエージェントシステムにおける強化学習エージェント設計支援機構”, JAWS2012 講演論文集, 2012.
- [3] Junling Hu, “Nash Q-Learning for General-Sum Stochastic Games”, Journal of Machine Learning Research 4, pp.1039-1069, 2003.