

異なる知識を持つ複数エージェントによる単体ロボットの意思決定手法の提案 -学習進度と行動選択確率に基づいた行動価値の統合による学習性能の向上-

酒井 啓伍

February 2024

- 強化学習の環境変化に対して行動価値を上書きする問題に着目
 - 強化学習では環境が変化した際に、変化前の行動価値を用いることで現在の環境に適さない行動を多く取り、行動数が増加するという問題があった。また通常の強化学習では学習空間が一つのため、変化後の環境を学習する際に変化前の過去の行動価値に対して上書きをしてしまう。そのため、一度学習した環境であっても再度訪れた場合には学習を行う必要があり、効率的な探索が行えないという問題があった。
- 過去に学習した環境の行動価値を蓄積して活用するシステムの提案
 - 環境変化を認識し、新たなタスクエージェントの作成と、これまでのタスクエージェントをエージェント群へ格納することで過去の環境の行動価値を蓄積する。また、タスクエージェントの行動価値と過去の環境を学習したエージェント群内の各エージェントの行動価値をタスクエージェントの学習進度とエージェント群内の各エージェントの環境への適合度を表す信用度に基づいて統合することで、知識の活用を実現する。
- 迷路問題を用いて過去の行動価値を活用した効率的な探索による学習性能の向上を検証
 - スタートとゴール、壁が存在する迷路問題を利用し、提案システムにおいて効率的な探索が行われているかの検証実験を行った。実験結果から過去の環境の知識を蓄積して活用する提案システムにより強化学習の問題を改善し学習時における行動数を減少させた。このことから提案システムによりエージェントの学習性能が向上したことを確認した。

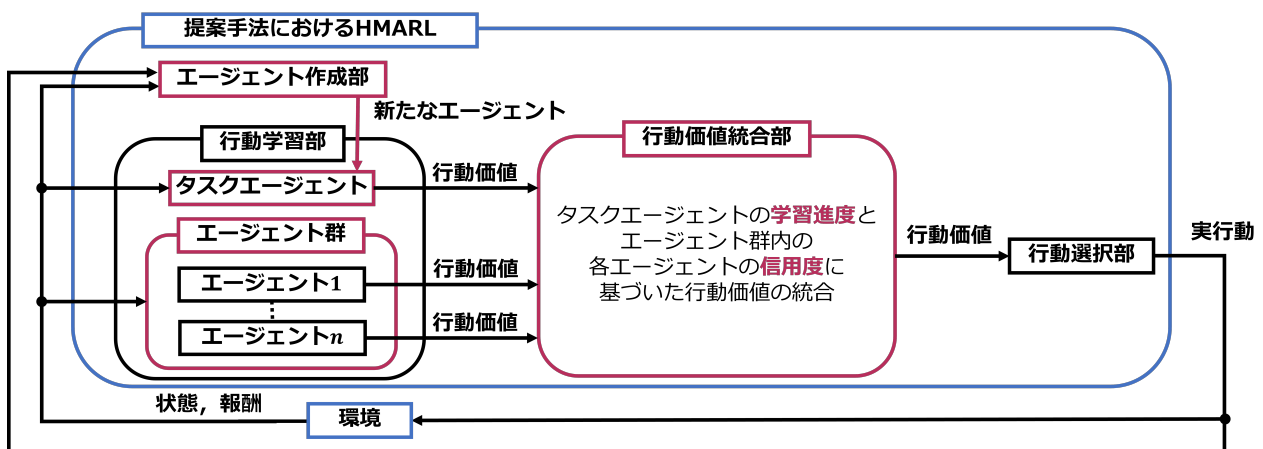


Fig.1 HMARL システムの概要図