

論文内容の要旨

博士論文題目 コミュニケーション実現のための階層型モジュール強化学習

氏 名 杉本 徳和

ヒトは複雑な環境をシンボル表現する事で高度な推論や認知的活動を可能にしており、そして見まね学習や協調作業など他者との相互作用がシンボル生成の源であるとされている。連続で高次元な環境にシンボルを割り当てる事や他者が用いているシンボルの意図を推定する事は一般に困難な問題であるが、本研究ではエージェントが過去の経験により獲得している行動即がそれらを解決すると仮定する。その時、エージェントの制御アルゴリズムが生成されるシンボルの性能を大きく左右する。そこで本研究では他の手法よりも柔軟な構造を持つ階層型強化学習手法を提案し、その枠組みを用いて他者のシンボルを理解したり新たなシンボルを生み出す枠組みを提案する。

本稿ではまず、**MOSAIC** モデルに基づいたモジュール構造を持つ 2 つの強化学習アルゴリズムについて説明する。1 つ目は複数個の状態予測モデルと報酬予測モデル、強化学習コントローラを持ち、環境の変化に応じて 3 者を適応的に切り替えながら制御を行うアルゴリズムである。2 つ目は **MOSAIC** モデルに基づいて環境を抽象化する階層型モジュール強化学習である。このアルゴリズムは状態予測誤差に基づいて環境を抽象化する下位層と、抽象化された状態空間で強化学習を行う上位層から構成される。上位層は各抽象化状態に対して選択すべきサブゴールを下位層に指示し、下位層がサブゴールを目標とし環境への出力を行う構造になっている。

後半では階層型モジュール強化学習を用いたコミュニケーションのモデル化について説明する。我々がコミュニケーションを行う際、観測された他者の軌道からその裏に隠れた高次の情報を推定している事は想像に難くない。その推定は一般に不良設定問題となるが、提案するモデルでは他者エージェントも **MOSAIC** モデルを基礎とし、似たような階層構造を持っていると仮定することで他者がどのモジュールを用いているかが推定できる。他者が使用しているモジュール系列をまねる見まね学習と他者のモジュール系列に応じて自身の行動選択を行う協調作業の枠組みを定式化する。

(論文審査結果の要旨)

強化学習は報酬信号をもとに試行錯誤的に学習を行う理論である。環境に対する事前知識がない場合でも自律的に制御則が獲得できることから、ロボットの制御アルゴリズムとしての期待が大きい。しかし学習に多くの試行数を必要とし、また環境が動的に変化する場合はそのつど学習をやり直す必要があるため実用性の面で未だ発展途上と言える。その様な欠点を克服するには階層構造の導入が有効な手段である。階層型強化学習の特徴は上位層が抽象化された状態空間で働く事により探索行動の高効率化が行える点にある。しかし課題ごとに設計者が上位層の設定を行う事は現実的ではなく、自動的に上位層を獲得できかつ適応範囲が広い手法が求められる。また観測された他者の運動軌道からなんらかの知識を得る見まねも学習速度を改善する有効な手段である。しかし一般に物理パラメータや内部構造には個体差があるため、単純に運動軌道を追従するだけでは有効な知識は得られない。

本論文が提案する階層型強化学習は複数の状態予測モデルを用意し、それぞれの予測性に基づいて環境の状態を分節化する手法である。各状態予測モデルはダイナミクスの入力空間に対して担当範囲を持っており、その距離に応じて選択確率の事前分布が計算される。そして観測された状態変化との誤差に基づいた尤度が計算される。それら事前分布と尤度の積より事後分布が求められ、その分布が各分節の活性度合いとなる。ダイナミクスの入出力空間の両方において分節化を行う事により、環境のダイナミクスが非定常に変化する課題にも適応可能という利点がある。通常の階層型強化学習は上位層の初期設定によって性能が制限されてしまうが、提案手法は下位の制御則を最適化する枠組みを示している点も評価できる。また本論文は見まねにより学習速度を改善する枠組みも定式化している。抽象化された空間で働く上位層の情報を推定し真似る事により物理パラメータなどの個体差を解消できるが、運動軌道のみからそのような高次の情報を推定する事は一般に困難である。提案手法は過去の経験からエージェントが獲得しているモジュール構造を用いる事でその推定を可能に

している．先行研究では上位層の状態のみ推定を行っているが，提案手法は行動も推定できている点が評価できる．

本論文は自律性と学習速度という相反する要求を同時に解決できるという点が大きく評価できる．またその柔軟な内部構造が可能にする他者の内部状態の推定手法は学習速度を改善するだけでなく，協調作業といったコミュニケーション課題を解決する事にも役立っている．今後はロボットの社会への進出が活発になると予想され，自律性・速い学習速度・社会性が強く求められる．そのような状況において本論文の成果は広い範囲で応用が可能である．このように本研究は情報工学としての貢献のみならず，社会科学や人間科学にも貢献するものであり，その学術的意義は大きい．よって博士（工学）の学位論文として価値があるものと認める．