多安定な価値構造を獲得する強化学習モデルと その学習ダイナミクス

理工学研究科数理情報学専攻
T12M006 濱口 大樹
指導教員 中野 浩

概 要

機械学習の一種である強化学習には、学習設計上の 2 つの問題が存在している。1 つは、状態空間に対する適切な設定方針が存在しないという問題であり、他方は、探索と知識利用のトレードオフと呼ばれる問題である。本研究では、強化学習に「動的な関数近似器」と「活性度による心的飽和効果」の導入によって、これらの問題を解決できることを示す。更に、これらの提案手法によって、強化学習を適用できる問題領域が拡張され、この新たな問題領域での学習過程を「学習ダイナミクス」として再解釈することを試みる。

強化学習には、未知の問題に対して適切に状態空間を分割し表現するための設計指針が存在しない。とくに実環境のような連続な状態空間では、有限個のパラメータで状態空間をどのように表現するかが強化学習のパフォーマンスに大きな影響を与える。この問題を解決するために、本研究では実状態空間表現のための「動的な関数近似器」を提案する。本研究で提案する関数近似器は、それぞれが独立に機能する局所的な関数近似モジュールにより構成される。必要なモジュールをボトムアップ的に生成・配置しつつ、不要となったモジュールを統合・削除することで、記憶容量と計算量を抑制可能な関数近似器である。この関数近似器を簡単な関数近似問題に適用しその有効性を示す。また、提案する関数近似器では、学習過程で頻繁に増減する多数の近似モジュール間で最近傍探索を行う必要がある。ここでは、このような動的な状況下でも有効な最近傍探索アルゴリズムを提案し、既存の手法との比較実験も行っている。

強化学習におけるもう一方の問題は、探索と知識利用のトレードオフ問題である。これは、学習過程でより良い解を探索するため、未知領域への新たな無作為探索と、これまでに獲得された知識とをどのようにバランス良く利用すべきかという問題である。この問題に対して、「活性度による心的飽和効果」に注目した新しい強化学習モデル手法を提案する。提案モデルは、Q–Learning の状態行動対に行動活性度を導入したものである。同じ状態行動対が高頻度で選択されて、行動活性度が高まると、「心的飽和（飽き）」の状態となる。それにより、その状態行動は一時的に選択されなくなり、無知的な行動を選択する。ここでは、いくつかの迷路タスクに対して心的飽和効果を適用し、トレードオフ問題、およびに局所解への落ち込みなどに対する有効性を示す。

強化学習の価値構造は、複数の局所解が遍在するような特徴をもつことがある。このような場合、強化学習は局所解に陥る可能性がある。行動活性度の導入は、それら局所解からの脱出を可能とし、それら局所解の遷移は連鎖的学習経路を自律的に構成する。これは結果的にサブゴールのような振舞いをみせる。

また、初期状態とゴールが明確に定義できないような問題に対して、行動活性度は学習に対する新たな視点を提供する。このような問題設定では、心的飽和効果が学習過程を停留させないため、学習収束後も状態遷移のダイナミクスが生じる。これは従来のゴール指向的学習パラダイムを拡張し、「永続的な学習過程」として捉えることが可能である。本研究では、迷路タスクを簡単化した Toy モデルを用いて、この永続的な学習ダイナミクスを解析する。このような学習ダイナミクスはヒトの学習過程とも類似し、学習に対してより動的な観点の提供が期待できる。