

学位論文審査の結果の要旨

報告番号	先端科学技術甲第168号	氏名	神谷 匠
論文題目	Softsatisficing: 確率論的満足化方策		
論文審査委員会	委員(主査) 教授 高橋 達二 委員(副査) 教授 山口 正二 委員(副査) 教授 中山 洋 委員(副査) 教授 柴山 拓郎		

研究の背景

環境から獲得する報酬を最大化する行動系列を試行錯誤によって見つけようとする強化学習は、近年目覚ましい発展を遂げた。しかし、従来の強化学習アルゴリズムは、あくまで最適な行動の発見を目標に据えていることが仇となり、現実世界(きわめて複雑でありうる環境)のタスクに応用しようとするならば、時間的・物理的な制約から有効でない可能性が高い。本研究では、人間や動物が実際に用いている生存戦略にヒントを得て、Simon が提唱した「満足化 satisficing」という原理(戦略)に着目する。満足化の原理とは、人間や動物は一定の基準(満足化基準)を超える行動を探し続けるが、一旦、その基準を超える行動が見つければ満足して探索を止め、基準以上の価値を持つ行動を探し続けることはしない、というものである。しかし、従来の満足化原理の研究においては、具体的な探索の方法やありべき性質についての議論が欠けていた。

これに関し、先行研究で提案された満足化価値関数RS(risk-sensitive satisficing)は、この満足化のアイデアに、リスク態度の自然な切り替えを組み合わせることで、極度に効率的な満足化を実現するものであり、またその振る舞いには数学的な保証も与えられている。しかしながらRSは、各行動の価値(各行動のもたらす報酬の平均)だけでなく、各行動を過去に実施した回数のカウントをも必要とする価値関数である。カウントを必要とすること自体は、強化学習が扱うタスクにおいて最も単純なクラスをなすK本腕バンディット問題に用いられる行動方策としては一般的であるものの、深層強化学習でないと扱えないような広大な状態行動空間においては、他の技術(「疑似カウント」など)と組み合わせることが必要となるため、RSにより効率的な探索を行う満足化行動方策を強化学習のタスク一般へと全面的に適用する際の障害となっていた。

研究の目的

本研究の目的は二つある。上記の行動選択カウントの必要性という問題を解決し、カウントを必要とせずに、強化学習全般の行動選択で用いることのできるアルゴリズムを提案するのが本研究の目的の一つ目(工学的な目的)である。また、カウントを不要とすることにより、動物や人間、人間組織が実際にどのような満足化基準を前提に(あるいはその基準を動的に更新しながら)行動選択を行っているのか、という点に関して定量的な統計分析を可能とすることがもう一つの研究目的(科学的な目的)である。

研究の内容

本研究では、確率論的な方策である Softsatisficing を提案した。Softsatisficing は、既存の満足化価値関数RSに新しい分析を施すことで導出された。RSと同等の振る舞いとパフォーマンスを示すと共に、カウントを必要としないことによりRSに比して飛躍的にその適用範囲の広がったこの手法については、次の論文において正当性のある導出法を示し、またシミュレーションにより有効性の検証を公表している。

Takumi Kamiya, Tatsuji Takahashi,
Softsatisficing: Risk-sensitive softmax action selection,

Biosystems, 2022, 104633. (採録決定, IF 1.97)

<https://doi.org/10.1016/j.biosystems.2022.104633>

以上、本論文において著者が得た結果は、工学的な観点からは、効率的な満足化行動選択の原理の応用範囲を飛躍的に広げ、今後の産業的な応用が期待される。また科学的な観点からは、実際の動物や人間、そして人間組織の行動データに対する定量的なモデリングへと適用可能となったことにより、動物、人間、人間組織の理解に寄与しうるため、神経科学、心理学、経営学や経済学などにインパクトを与える可能性がある。以上のように、本論文の価値は科学的ならびに工学的な観点から十分に評価できる。よって、本論文は博士(情報学)の学位論文として十分な価値を有するものと認められる。