



自律的に創造した遊びを通じて成長する人間型知能の開発

東京大学

特任研究員 田中 一敏

研究目的

人間が遊びを通じて多くのことを学び、育つ。特に、乳幼児は自ら遊びを創造し、創造した遊びを通じて、自分の身体を動かす方法、物体を扱う方法、他人の心を察する方法などを学び、運動能力、認知能力、社会能力を発達させる。このため同様に、自律的に遊びを創造し、創造した遊びを通じて成長する知能のモデルを構築できれば、人間が楽しむ遊びの本質を機械論的な観点から理解できる。加えて、人間と同様様々な課題を解決できる人工知能の実現に貢献できる。さらに、人間の発達を理解するためのモデルを提供できる。これまで、与えられた目標を達成するために副目標を自律的に設定する方法が研究されてきた。しかし、目標自体を自ら設定することや、人間が行う多様な振る舞いは考慮されていない。このため本研究の大目標は、自律的に遊びを創造し、創造した遊びを通じて成長する、人間型の知能を開発することである。

この大目標を達成するために、まずすべきこととして、自身の能力を踏まえながら、適切な難易度の目標を設定し、達成することを繰り返しながら、変動する目標を達成する、汎用的な問題解決能力を自ら鍛える枠組みを構築することである。このため、そのような枠組みの構築が本研究の目的である。

研究の特徴

これまで、与えられた目標を達成するために副目標を設定する方法が研究されてきた。対して本研究は、機械が自ら目標や規則を設定する方法を研究する。

他に、四肢のない、少数の関節しか持たない身体を備え、刺激を求めて遊ぶ知能について研究されてきた。対して本研究は、四肢と多数の関節を備えたより複雑な人間型の身体を有する知能を扱う。本研究が遊びを創造し、成長する知能を開発する。そこで、どのような知能がどう発達したか、どのような遊びが生まれ、どう遊ばれるかが示される。このため、遊びに関する新しい仮説や人間が発達する原理を提示することが期待される。

本研究において設計者は、開発する知能が達成すべき目標を設定しない。このため、開発した知能は、遊びを通じて獲得した範囲のあらゆる問題に対して解答を出せる。ゆえに、汎用人工知能の実現に貢献することが期待される。

関連研究

現在、知能の運動をシミュレーションすることで、多様な運動を獲得する知能について盛んに研究されている。そこでは、平らな床における動物型の身体運動がシミュレーションされる。すなわち、現実世界にある環境の複雑さが欠落している。対して本研究では、複雑な環境において知能が生み出す多様な振る舞いを扱う。それにより、人間型の高度な知能を開発する。

他に、知能が音楽や画像や映像を生成する研究が活発に行われている。そこでは、ある画家の筆致と他の画家の構図を組み合わせる新しい絵画を作成する方法、与えられた顔画像から別の表情を生成する方法などが開発された。対して本研究では、知能によって遊びが生成される。それにより、人類がこれまで体験したことのない遊びが生み出される可能性がある。加えて、既存の遊びを類型化するための知見も期待される。

本研究で取り組んだ経緯

これまでで私は、ロボットと機械学習と統計の技術を用いて、人間が行う運動を再現し、解析し、シミュレーションしてきた。具体的には、人間のスポーツ動作を再現するロボットの開発、卓球を行う人間の運動予測方法の開発、発達に伴う乳児の運動変化の潜在原理の探求、多様な運動を獲得する知能の開発などである。そこで得た技術と知識を活用することで、本研究において知能が遊びを創造し、学習し、経験を蓄積するアルゴリズムを構築し、シミュレーションを行うことに挑んだ。

実験内容

強化学習のマウンテンカー環境を用いた。エージェントの課題は通常通りに目標到達位置が山の上になるのではなく、ランダムに発生する目標に対して、その位置に到達する振る舞いをエージェントが学習する問題として改変した。その中で、エージェントは自ら目標を設定し、達成に取り組むことを繰り返すことで、変動する目標を達成する能力を鍛える。エージェントは、Deep Deterministic Policy Gradient を用いて方策を学習した。提案手法と比較するため、ランダムに目標位置を選択する場合と比較した。このランダムな比較は、現在の最善の手法でもある。

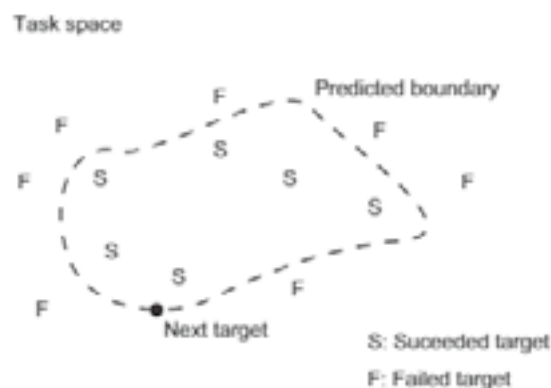


図1：目標選択 [1]

提案手法 (図1)

まず、ありうる目標を多数設定し、それらの目標ごとの取り組み、成功か失敗かの結果を記憶し、目標と成否の対応関係の表を作る。次に、目標と成否との関係を Support Vector Machine を用いて近似し、成功するか失敗するか分からない領域を抽象する。続いて、抽出した領域の中から目標を選択して取り組む。以上を繰り返すことで、変動する目標に対する行動を学習する。学習の過程で適切な難易度の問題に取り組むことが期待できる。

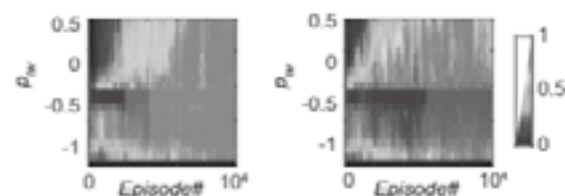


図2：成功率 [1]

実験結果 (図2)

エピソードごとの、目標位置と成功率との関係を示す。図から、多くの目標位置に対して、提案手法の方が早く成功することが分かる。この結果から、提案手法が変動する目標に対する振る舞いを高効率で学習できると分かった。

今後の展望

本研究において、変動する目標に対する振る舞いを、自己の能力を省みて目標を自ら設定して取り組むことで学ぶ方法を構築できた。以上の成果は査読付き国際会議 International Conference on Development and Learning and on Epigenetic Robotics に採択され、発表された。

現在、目標設定のたびに目標ごとに成功するか試す必要があるため、今後、試すことなく結果を予測できるように改良する。他のより複雑なタスクを解るように手法を拡張する。人間型の身体と、人間の暮らす複雑な環境でも、同様に遊びを生成できるように、発展させていく。

参考文献

- [1] Learning Multiple Tasks by Self-Goal-Setting Based on Self-Evaluation, K. Tanaka, I. Karino, H. Kanazawa, R. Niiyama, Y. Kuniyoshi, The 9th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics, 2019.