

忘却学習を取り入れた Dynamic Fuzzy Q-Learning における 協調行動の獲得

日大生産工 ○川野 勝之 日大生産工 山内 ゆかり

1 まえがき

集合とは、要素が集合に属しているか否かを明確に判断するものだが、ファジィ集合とは、ある要素が集合にどれだけ属しているかを 0~1 で表す集合である。例えば、「年齢が若い」というファジィ集合があったとすると 25 歳という年齢は、ファジィ集合に 0.8 属しているなどと表すことができる。このようにファジィ集合は曖昧さを表現する集合を指す。近年ファジィ集合を用いた Q 学習への応用が注目されている。

Er らは、Q 表に状態と行動の組(ファジィルールと呼んでる)を動的に追加できる Dynamic Fuzzy Q-Learning(以下 DQFL に示す)を提案している。1)

細谷らは、DQFL に忘却学習を適用し、不要な状態と行動の組やファジィ集合を削除する方法を提案した。2)

細谷らによって提案された、追跡問題に忘却学習を取り入れた DQFL では、状態を追加したり削除することで無駄なメモリを消費しないが、状態の属性が少ないために学習を早めることができない。hunter の属性としては prey と hunter の距離と角度の 2 つである。hunter の視界に別の hunter がいても考慮がないため hunter 同士が衝突する直前に一方の hunter が止まらなければならないだけでなく無駄なステップを踏むことになる。そこで hunter 同士に協調行動を取り入れることで学習の高速化を提案する。

また、hunter は入力データを得られない時は、ランダム行動するため、hunter の視界に別の hunter を見つければ衝突を防ぐために別の方向を探そう行動をさせる。

本研究では、追跡問題に忘却学習を取り入れた Dynamic Fuzzy Q-Learning(以下 DQFL に示す)を用いる際に、hunter 同士の間に協調行動を取り入れることで学習の高速化に着目した。

2 従来研究

細谷らが提案した追跡問題に忘却学習を取り入れた DQFL とは具体的には、動的に状態と行動の組を追加していく際に TD 誤差を用いてファジィ集合の中心値と幅の調整を行う。さらに忘却を用いることであまり使用されないファジィ集合を削除できるようにする。また、状態の評価値として V 値を使用し、これに忘却を用いることで不要な状態と行動の組の削除を行う方法である。

図 1 に全体のフローチャートを示す。

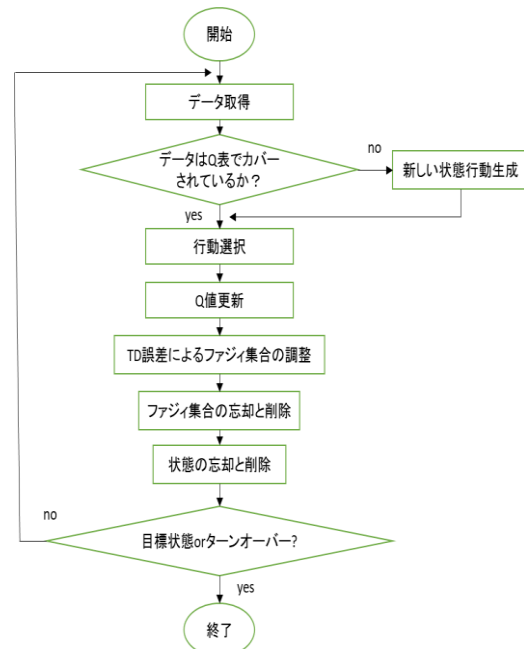


図 1.全体のフローチャート

図 1 では hunter の視界に prey が観測された場合の hunter のデータ観測から状態の忘却と削除までを示している。hunter は目標状態(hunter3 体以上で捕獲)になるか、あるターン数になるまで行動、更新し続ける。提案手法では、Q 表の属性を増やすことなので、この流れは変わらない。

Acquisition of cooperative behavior in Dynamic Fuzzy Q-Learning incorporating forgetting learning

Katuyuki KAWANO and Yukari YAMAUCHI

図 2 に状態と行動の組の生成と削除を繰り返した後の Q 表を示す。

		状態		行動			
No.	V値	距離	角度	上	下	右	左
1	V1	ファジィ集合	ファジィ集合				
2	V2	ファジィ集合	削除				
4	V4	削除	ファジィ集合				
・	・	ファジィ集合	削除				
・	・	ファジィ集合	ファジィ集合				
・	・	削除	ファジィ集合				
n	Vn	ファジィ集合	ファジィ集合				

図 2. 状態と行動の組の生成と削除を繰り返した後の Q 表

入力データを得た時に、ファジィ集合との一致度や V 値を考慮し、行動を ϵ -グリーディ法で選択する。すべての状態で行い統合し行動を決定する。

4 提案手法

本研究では、追跡問題に忘却学習を取り入れた DQFL を用いて、hunter 同士の間に協調行動を行うことを提案する。

他の hunter の状態の情報を属性として追加することで、他の hunter を考慮するため最適な行動を得ることが可能となると考えられる。従来研究では、hunter から prey までの距離と hunter から prey に対しての相対角度を属性としているが、属性の数が少ないために学習スピードが遅いと考えられるため新たに hunter 同士の情報を属性に増やすことを提案する。

5 実験環境

本研究が提案する実験環境としては、縦と横の大きさが 15×15 のトーラス状の実数値空間上に prey を 1 体、hunter を 4 体ランダムに向きを持たせたまま配置する。prey はランダムに行動する。hunter も prey の行動の後にファジィ集合から求めた一致度をもとに行動する。もし、hunter の視界に別の hunter がいるときは衝突を避けるように行動する。また、hunter の目標状態としては、hunter が prey を取り囲んだ状態である。この目標状態は複数の hunter が prey をタッチしている状況(正面方向の半径 $r \cdot 3$ の半円内)にあり、prey と各 hunter の中心を結んだ線分について、隣り合うすべての線分の角度が 180 度以下の時である。

目標状態の例を以下の図 3 に示す。

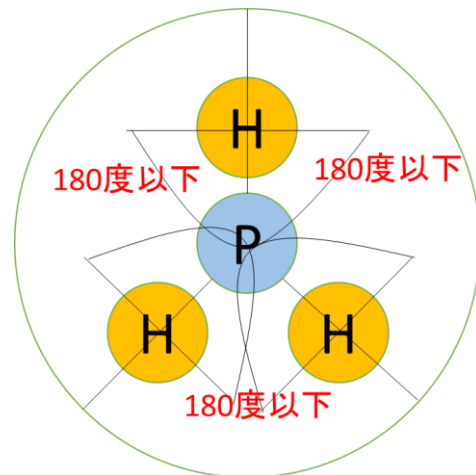


図 3. 目標状態

目標状態の場合は捕獲にかかわっている各 hunter に成功報酬が与えられる。また、目標状態でなくても各ターンにおいて prey にタッチしている状態であれば部分報酬が与えられる。目標状態あるいはターンオーバーになるまで行動を続ける(これを 1 ゲームとする)。この実験を 10000 ゲーム(これを 1 セットという)行い、成功率と 10 セットにおける成功時の平均ターン数を調べる。

6 まとめ

本研究では、追跡問題に忘却学習を取り入れた DQFL を用いて、hunter 同士の間に協調行動を行うことを提案したが、従来研究で問題視した学習速度の精度が改善されたと考えられる。

「参考文献」

- 1) M.J.Er and C.Deng: "Online Turing of Fuzzy Inference Systems Using Dynamic Fuzzy Q-learning. "IEEE Trans. on systems, Man, and Cybernetics, Vol34, Np3, pp.1478-1489, 2004
- 2) 細谷優,馬野元秀, 状態の調整および状態と行動の組の削除機能を持つ動的ファジィ Q-learning, 日本知能情報ファジィ学会誌, Vol26, No5, pp844-854, (2014)