1R-01

Fuzzy-ARTを用いたStepwise Unified HRLの提案

黒須　宏道†　　真部　雄介†
†千葉工業大学大学院　情報科学研究科　情報科学専攻

1　はじめに

強化学習とは，Agentと呼ばれる学習主体が観測可能な環境との相互作用によって適切な行動系列を学習する機械学習手法の1つである．Agentは，設計者が環境に設定した報酬を最大にするような行動を学習する．そのため，試行錯誤的に望ましい行動系列を学習できるが，長期的な戦略を必要とするタスクを解決することが困難であることが知られている．

上記の問題点を解決するアプローチの1つとして，Hierarchical Reinforcement Learning(HRL)[1]がある．HRLとは，事前に用意したサブゴール群から現環境の最適なサブゴールを学習するMeta Controllerと，それらサブゴールまでの行動系列を学習するControllerの2つから構成される階層構造を持った強化学習フレームワークである．サブゴールとは，タスクを細分化する目標のことであり，これによってプランを考慮したかのような行動をAgentに取らせることが可能となる．また，強化学習で解くタスクの長さが一定を超えると，HRLが必要となることがAl-Shedivatらの研究[2]で明らかになっている．そのため，近年ではHRLに関する研究が増加している．

HRLの問題点として，手動でサブゴールを用意する必要がある点が挙げられる．この問題点を解決しようとした研究として，Unified Hierarchical Reinforcement Learning(UHRL)[3]がある．これは，Agentが行動により得た経験をクラスタリングしたクラスタをサブゴールとして用いることができることを証明している．経験とは，行動，環境状態，報酬，次のステップ時間の環境状態の総称である．UHRLでは，経験を獲得させる手法として以下の2つを用いている．1つ目は，ランダム行動を一定回数行う方法で，ランダム行動で問題解決に充分な経験を得られる短いタスクで用いていた．2つ目は，環境画像を用いて離散した初期サブゴールを生成し，その学習過程を用いる方法で，本来HRLが対象としている長いタスクで用いていた．

UHRLの経験獲得手法2の問題点として，強化学習による学習行動を経験獲得に用いているため，安定して新しい経験を得ることができない点が挙げられる．そのため，問題解決に必要なサブゴールを獲得する前に局所解に陥る可能性がある．

そこで本研究では，この問題を解決可能な新たな強化学習フレームワークである，Stepwise Unified Hierarchical Reinforcement Learningを提案する．提案手法は，新しい経験を獲得しやすいランダム行動を段階的に複数回行うことで，徐々に適切なサブゴールを獲得していく手法である．

A Proposal for Stepwise Unified HRL with Fuzzy-ART
†Hiromichi KUROSU　†Yusuke MANABE
†Graduate School of Information and Computer Science, Department of Information and Computer Science, Chiba Institute of Technology

2　提案手法

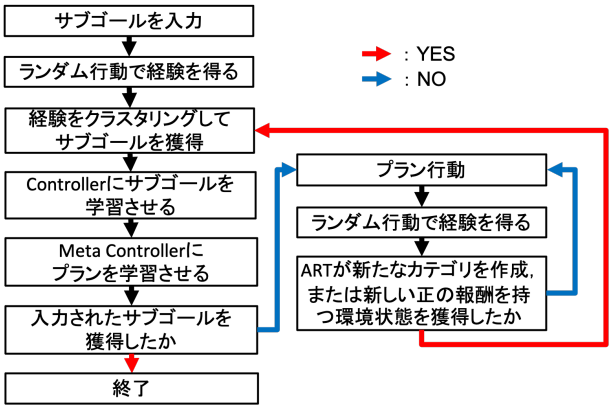

図1　提案手法の流れ図

図1に提案手法の流れ図を示す．

本研究では，初めのランダム行動で獲得したサブゴールを用いたプランの行動後，新たなランダム行動をさせる．このサブゴール獲得処理の繰り返しは，設計者が入力したサブゴールを獲得した際に終了する．

また，この仕組みを実現するためのクラスタリングアルゴリズムとして，適応共鳴理論(Adaptive Resonance Theory:ART)を用いる．ARTとは，ニューラルネットワークの一種であり，入力と記憶の類似度を警戒パラメータという閾値で判別し，どのカテゴリにも属さなかった場合は新たなカテゴリを作る特徴を持つ．この特徴によって，報酬による差異が無い場所でも正確に新しい経験を得た判別が可能だと考える．本論文では，ARTの教師なし学習の1つであるFuzzy-ART[4]を用いる．

プランとは，サブゴールの適用順序のことであり，獲得されたサブゴール群で構成される．本研究のプランには，Meta Controllerが学習によって獲得するプランと，サブゴール群から総当りで作られるプランの2つ存在する．総当たりプランを例を挙げて説明すると，サブゴール群が2つの地点（位置Aと位置B）だった場合，(A), (B), (A→B), (B→A)の4通り生成される．Agentは，Meta Controllerのプランに基づく行動を行った後，総当りプランに基づく行動を実行し，新たな経験獲得のためのランダム行動を行う．上記のAgentによる経験獲得処理は，ランダム行動で得た経験により，ARTが新たなカテゴリを作成する，もしくは新たな正の報酬を持つ環境状態を得るまで実行される．その間，総当たりプランが逐一変更される．

提案手法のメリットとして，Meta Controllerが学習したプランをベースにランダム行動をするため先行研究の利点を無くさずに新しい経験を獲得しやすい点，サブゴール獲得処理の終了条件としてサブゴールを設定する代わりに最終的なゴールを設定できる点が挙げられる．

3　実験・評価方法

実験環境として，UHRLで用いられていたグリッドワールド環境を用いて，複数の部屋を持つ環境(図2)を作成した．行動主体の初期位置は左上で，上下左右への移動が可能であり，ゴールに辿り着くことで正の報酬が与えられ問題解決となる．

本実験では，サブゴール獲得処理を終了するサブゴールにゴールを用いる．

この環境で，ARTの新たなカテゴリ作成時をサブゴール獲得処理の判断基準に用いることができるか確認することで，本フレームワークの評価を行う．

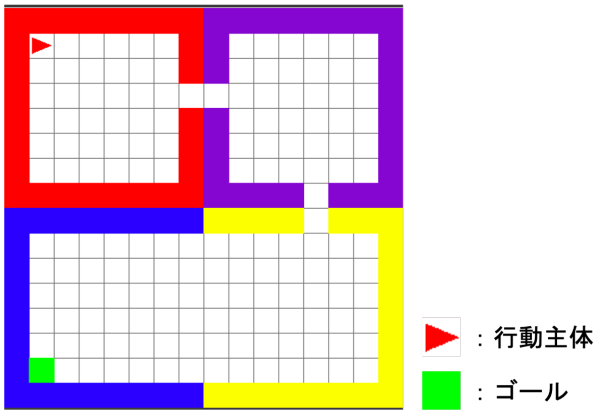

図2　複数の部屋を持つ環境

4　実験結果

1段階目のサブゴール獲得処理で得たサブゴール群を図3に示す．図のデータは，Fuzzy-ARTの学習に用いた経験内の環境状態(座標)であり，同じ色のデータは同じクラスタであることを表している． 1段階目では，1つ目と2つ目の部屋までの経験を獲得し，部屋ごとにクラスタが作成されていることがわかる．

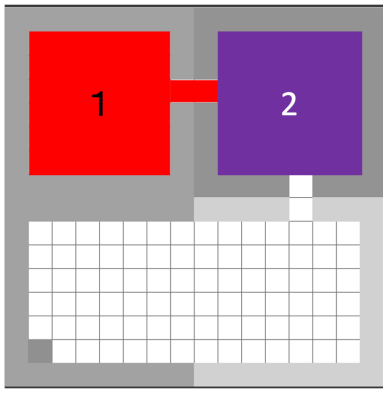

図3　1段階目のサブゴール群

1段階目のサブゴール群を用いて得た2段階目のサブゴール群を図4に示す． 1段階目で得たサブゴールを利用し，1段階目では得られなかった新しい経験を獲得できていることがわかる．

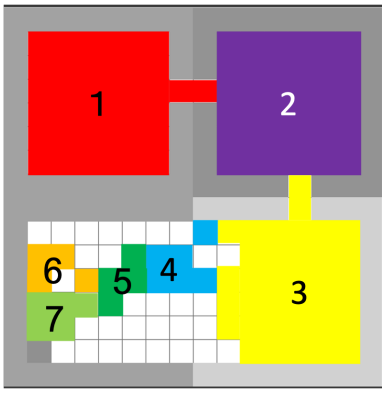

図4　2段階目のサブゴール群

2段階目のサブゴール群を用いて得た3段階目のサブゴール群を図5に示す．全ての経験を獲得し，それぞれクラスタによって分割されていることがわかる．入力したサブゴールであるゴールを獲得したため，ここでサブゴール獲得処理が終了した．

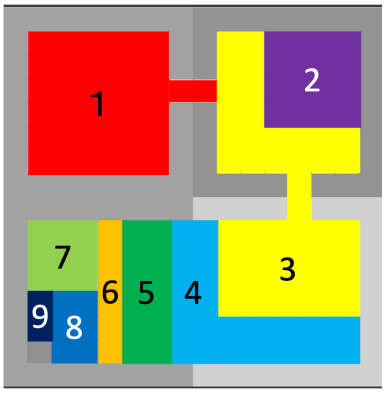

図5　3段階目のサブゴール群

実験結果より，ARTの新たなカテゴリ作成時を報酬の差異が無い場所でのサブゴール獲得処理の判断基準に用いることができることがわかった．

5　おわりに

本研究では，UHRLのサブゴール獲得処理をFuzzy-ARTを用いて段階的にした手法を提案した．実験の結果，ARTの新たなカテゴリ作成時を報酬の差異が無い場所でのサブゴール獲得処理の判断基準に用いることができることがわかった．

今後は提案手法を他の環境を用いて実験を行う．

参考文献

[1] T.D. Kulkarni, et al. “Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation.” *Advances in neural information processing systems 29.* 2016. pp. 3675-3683.

[2] M. Al-Shedivat, et al. “On the Complexity of Exploration in Goal-Driven Navigation.” arXiv preprint arXiv:1811.06889. 2018.

[3] J. Rafati, et al. “Unsupervised Methods For Subgoal Discovery During Intrinsic Motivation in Model-Free Hierarchical Reinforcement Learning.” *33rd AAAI Conference on Ar- tificial Intelligence (AAAI-19). Workshop on Knowledge Extraction From Games. Honolulu, HI, USA.* 2019.

[4] G.A. Carpenter, et al. “Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system.” *Neural networks 4.6* 1991. pp. 759-771.