

ナッシュ Q 学習を用いた災害救助問題におけるマルチエージェントの協調行動の獲得

Acquisition of Multi-Agent Cooperative Behavior in Disaster Relief Problem Using Nash-Q Learning

中村 蒼紫† 謝 孟春† 村田 充利† 森 徹†
 Soshi Nakamura Mengchun Xie Mitsutoshi Murata Toru Mori

1. はじめに

近年、東日本大震災や熊本地震といった地震や津波などの自然災害が増加している。災害による救助の人員不足や人が救助困難な危険区域での災害救助ロボットが注目されており、災害救助ロボットは自律的で効率的に救助することが望まれる。なお、複数のロボットをエージェントとするマルチエージェントシステム(Multi-Agent System, MAS)として扱うことができる。

先行研究では、エージェントが個々に Q 学習を行い、役割分担に応じて報酬を与えることで協調行動がある程度で獲得できた⁽¹⁾。しかし、MAS において個々での学習では、他エージェントの行動により環境が逐次変化するため、Q 学習の収束が保証されない問題がある。そのため、本研究では、ナッシュ Q 学習を用いて MAS の全体的な学習環境を構築する方法を提案する。災害救助問題を対象とし、ナッシュ Q 学習における協調行動の獲得の有効性を検証する。

2. ナッシュ Q 学習による協調行動の獲得

ナッシュ Q 学習は、Q 学習とナッシュ均衡戦略を組み合わせた手法である⁽²⁾。ナッシュ均衡とは、他のエージェントと自身にとって両方とも最良の行動をとった際の行動の組み合わせである。ナッシュ均衡戦略はナッシュ均衡を得るための方策である。このナッシュ均衡を行動価値関数に反映することで、学習によりエージェント同士が互いに最良な行動を選択することが考えられる。

ナッシュ Q 学習の行動価値関数の更新を式(1)～(2)に記す。Q 学習はエージェント自身の報酬の最大値に基づいて更新することに対して、ナッシュ Q 学習は将来のナッシュ均衡の報酬で更新される。

$$Q_i(s, a_0, \dots, a_n) = (1 - \alpha)Q_i(s, a_0, \dots, a_n) + \alpha(r_i + \gamma \text{Nash}Q_i(s')) \quad (1)$$

$$\text{Nash}Q_i(s') = \pi_0(s') \cdots \pi_n(s')Q_i(s') \quad (2)$$

ただし、 i は i 番目のエージェント、 s は現在の状態、 $a_0 \cdots a_n$ は行動、 r は報酬、 α は学習率、 γ は割引率である。 $\text{Nash}Q_i$ はナッシュ均衡をとるような方策を求める関数であり、 $\pi_j(s')$ はエージェント j の状態 s' における方策である。

本研究は、フィールド上に配置されている全ての負傷者をエージェントが効率よく救助する災害救助問題を対象とする。災害救助問題におけるエージェント同士にナッシュ Q 学習を適用させ、協調行動を獲得する。図 1 に災害救助問題における状態の 1 例を示し、その状態における行動の組み合わせの報酬の 1 例を表 1 に示す。この場合、ナッシュ均衡はエージェント i が上、エージェント j が右に移動する行動の組み合わせとなる。ナッシュ均衡での Q 値は、式(1)により更新され、最終的にナッシュ均衡戦略を獲得する。

表 1 行動の組み合わせの報酬の 1 例

$i \setminus j$	上	下	右	左
上	(10,0)	(10,-10)	(10,10)	(10,0)
下	(-10,0)	(-10,-10)	(-10,10)	(-10,0)
右	(0,0)	(0,-10)	(0,10)	(0,0)
左	(-10,0)	(-10,-10)	(-10,10)	(-10,0)

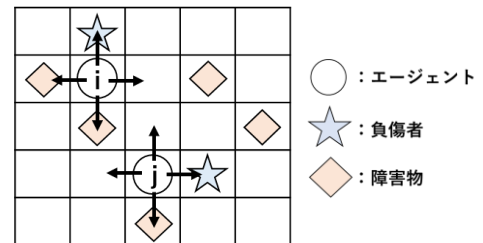


図 1 災害救助問題における行動の 1 例

3. 予備実験

予備実験としてフィールドサイズを 10×10 、エージェントを 2、2000 回の Q 学習の結果を図 2 に示す。横軸はエピソード数、縦軸はタスク完了までのステップ数である。学習率 $\alpha=0.1$ 、割引率 $\gamma=0.9$ である。行動選択として ϵ -greedy 法を用い、 $\epsilon=0.1$ である。

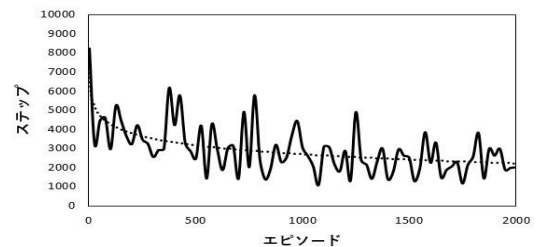


図 2 Q 学習による学習結果

図 2 より、エピソードが進むにつれてタスク達成までのステップ数は減少し、学習ができていくことがわかる。ナッシュ Q 学習による協調行動獲得の結果を大会当日に報告する。

4. まとめ

災害救助問題を対象にナッシュ Q 学習による協調行動の獲得を提案した。今後の課題として、ナッシュ Q 学習と Q 学習が協調行動の獲得に与える影響を比較し、考察する必要がある。

参考文献

- [1] 謝孟春・佐藤翔馬, 役割分担による災害救助エージェントの協調行動の獲得, 第 8 回進化計算学会研究論文集(2015), pp60-65
- [2] J. Hu, Nash Q-Learning for General-Sum Stochastic Games. J. M. L. R4, (2003) pp.1039-1069

†和歌山工業高等専門学校,
 National Institute of Technology, Wakayama College