

Social Homeostatic Reinforcement Learning (SHRL)

背景

Background.pdfの背景を参照

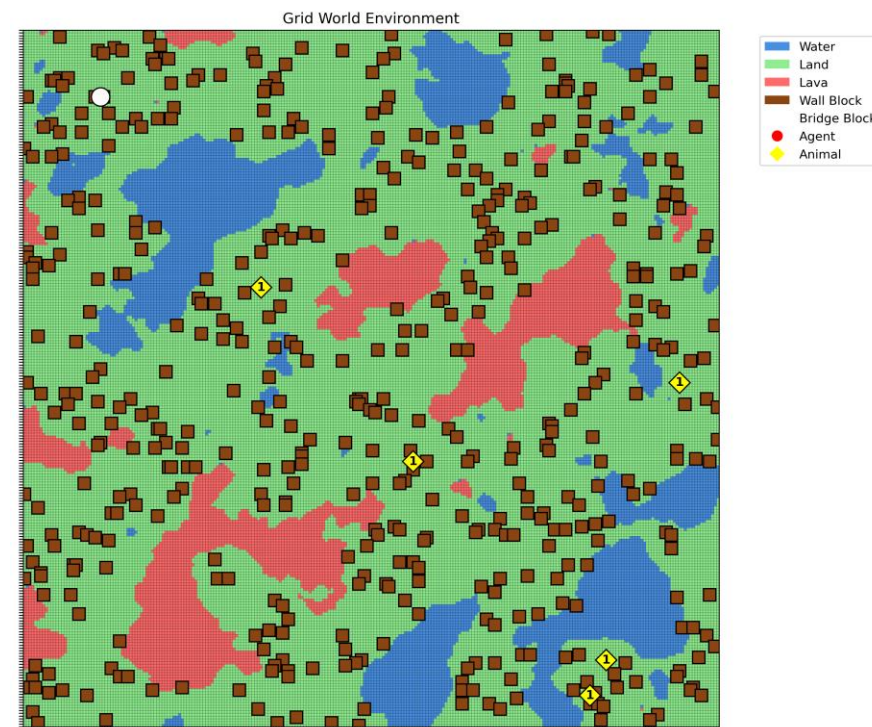
目的

ホメオスタティック強化学習(HRL)にSocial homeostasisによる報酬を追加することによって(SHRL)、エージェント間に社会行動が創発されるかどうか検証

方法

- Social homeostatic reinforcement learning(SHRL)を自作環境にて行った。
- 強化学習モデルはPPOを使用
- 環境の詳細はcode/env_要件.txtを参照
- エージェントは肉を地面に置いたりほかのエージェントに渡したりすることができるので、そのような行動がマルチエージェント強化学習をしなくても創発されるかをみたい。

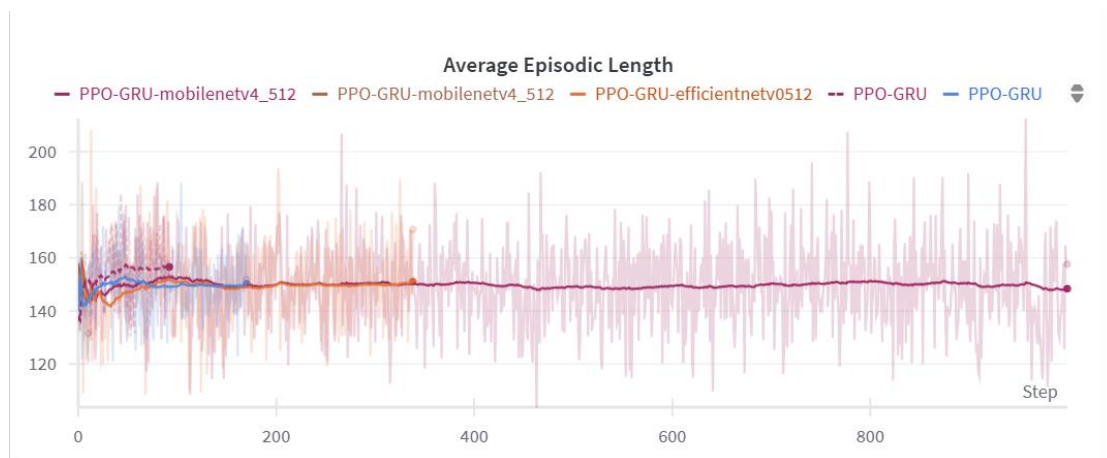
自作環境



結果

結果

- そもそも生存すら怪しい(学習が進まなかった)
- 食料や水の確保もHRLで行っていたが、そもそもHRLでは学習が収束するのに 4×10^9 回ループを回さないといけないらしく、学習を断念



各エピソードでの生存ステップ数
条件を変えてみたが、学習は進まず

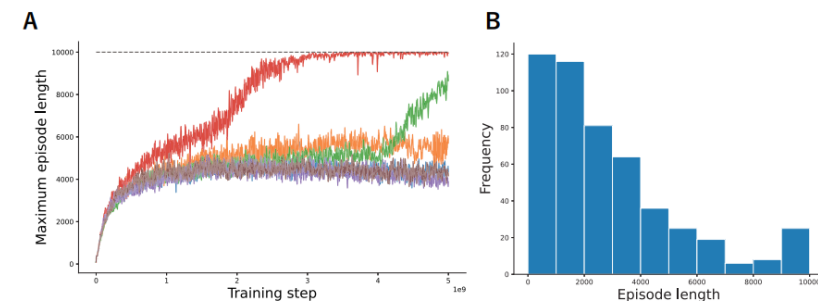


図 4: 異なる 6 回の試行の恒常性エージェントの生存ステップ数の変化. 破線は環境の最大ステップ数 (= 10,000) を表す. それぞれ直近 100 データの平均で平滑化している. A: 100 エピソード間の最大生存ステップ数. B: 最も性能のよいエージェントの 500 回の試行での生存ステップ数の分布.

吉田ら(2024)

https://doi.org/10.11517/pjsai.JSAI2024.0_4S1OS30a01より
収束に 10^9 回ほど学習させなければならない