

多面的な評価指標を用いた報酬の自己生成

-センサ入力に対する危険性を考慮した慣れの実現による危険認識の向上-

坂本 悠真 (Yuma Sakamoto)

February 2022

- 報酬の自己生成においてどの入力に対しても慣れるまでの時間が同じであるという問題点に着目
 - 我々はセンサ入力より多面的な評価を行い環境に合わせた報酬の自己生成をエージェントに行わせる研究を行っている．慣れの値の変化量を定める慣れ速度が定数であるため，危険な入力に対して安全な入力と同様に容易に慣れてしまい短時間で危険でないという評価を出すという問題点があった．
- 入力の大きさによって異なる危険性を考慮することで危険の認知向上を実現
 - 本研究では入力の大きさによって異なる危険性を考慮して慣れる速度の値が変化する手法を提案する．危険な入力であるほど慣れる速度の値を小さくすることで，危険な入力に対して長時間危険であると評価を示すようになる．
- 経路学習実験により，センサ入力の大きさに対する危険認識を検証
 - 災害後の環境を想定したシミュレーション実験で行った．危険な入力に対して慣れる速度の値を小さくすることで，安定して危険を回避する行動を学習することを確認した．



Fig.1 慣れる速度の値の変化のイメージ