

強化学習レポート

学籍番号:35714121

氏名:福富隆大

総作業時間:3時間

自己評価:S

演習課題のまとめ

作業時間

1時間

課題の概要

演習課題①②の結果のまとめと考察

課題の結果

演習課題1...10/10点

実装方法...障害物の座標を定義し、シートの範囲外かチェックしているときに一緒にチェックする。

考察...学習の際に障害物の座標に行くという挙動を許さないようにすることで、障害物をよけるように学習させることができた。

演習課題2...2/10点

リンゴを獲得した配置

黒...障害物、赤...リンゴ、青...自分の操作するtoio

				赤		
	黒	黒	黒			
	青					

	黒	黒	黒		青	
赤						

実装方法...障害物に当たって進めなくなったら数回ランダムに動くようにする

考察...今回はランダムに動かしているだけなのでリンゴをとれるかは運頼りになっている。同じ障害物に複数回当たっていたので一回当たった場所を記録しておいて、その座標に行こうとしたらランダムな別の座標に行かせるようにすると解決できるのではないかと思う。

強化学習と移動ロボットによる身の回りの課題解決

作業時間

2時間

課題の概要

身の回りの課題を強化学習と移動ロボットを使って解決する方法と定式化を用いたその説明

課題の結果

1.身の回りの課題

洗濯物を自分の部屋に持っていくのがめんどくさい

2.解決システムの概要

複数人が同居する家庭内における畳んでコンテナに入った状態の洗濯物の自動配達システム

以下の機能を備える

- 家庭内のマップに基づく自律移動移動
- 動的な障害物の回避
- コンテナの色で誰の洗濯物かを識別

3.強化学習の定式化

S:ロボットの位置、配送状況、未配送タスク

A:タスクの開始/終了、移動

R:コンテナを正しい配送先に届けると大きな正の報酬、間違った配送先や移動時間の増加には負の報酬

4.学習プロセス

強化学習の目的は、観測される状態S(例:「未配送コンテナ:赤2、青1、現在地:畳み場所」)に対し、最も効率的な行動A(例:「赤のコンテナを搭載し、花子の部屋へ移動」)を選択するポリシー $\pi(a|s)$ を学習すること。

5.まとめ

MDPとして定式化することで、ロボットは試行錯誤を通じて、環境変化に対応できる汎用的かつ最適な行動ポリシーを自律的に獲得できる。

しかし、実際に作成するにはコンテナの積み下ろしや充電、途中で落ちてしまった場合など多くのことを考える必要がある