強化学習の知見を書く

- 遅延報酬の問題：行動の結果がどうだったかを評価するには十分時間が経過してないとわからない。
    - 対策：TD法やモンテカルロ法で効率的に報酬を与えられるようなアルゴリズムを選択。
- 報酬のスパース性：迷路みたいに報酬が最後までもらえないとき報酬が希薄になること。
    - 対策:報酬形成で中間的な目標を設定する。
- ε貪欲法のεの値は0.05~0.1がよく使われる。