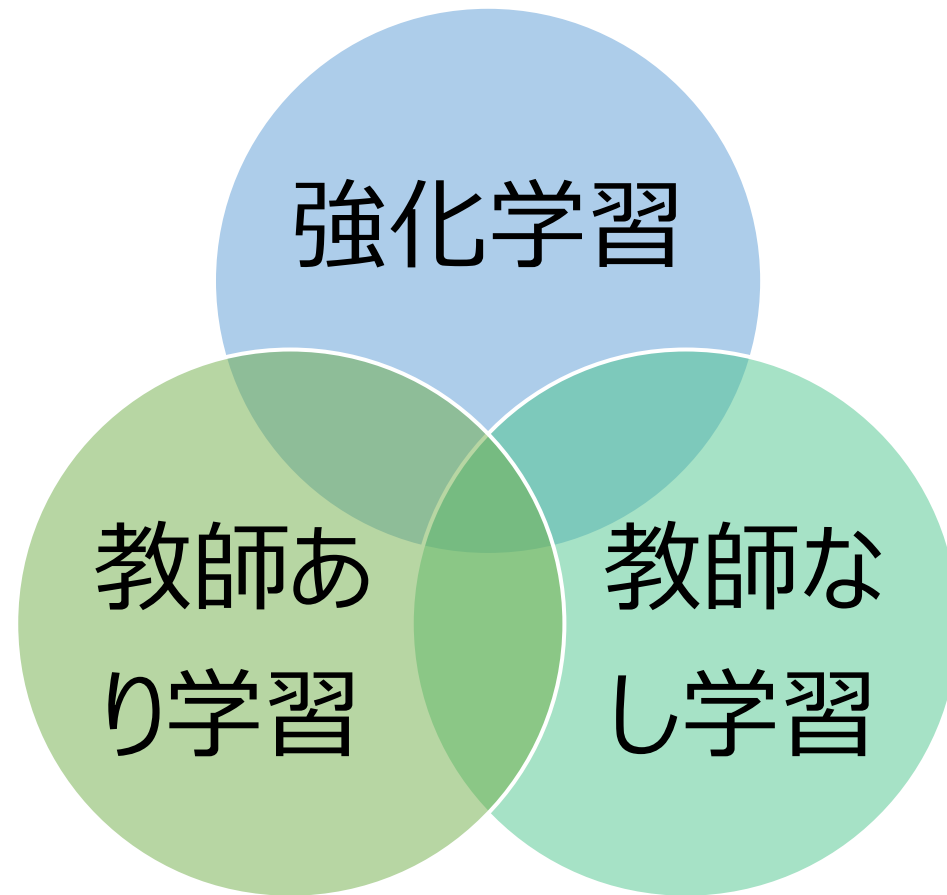


自然言語処理 — 強化学習入門 —

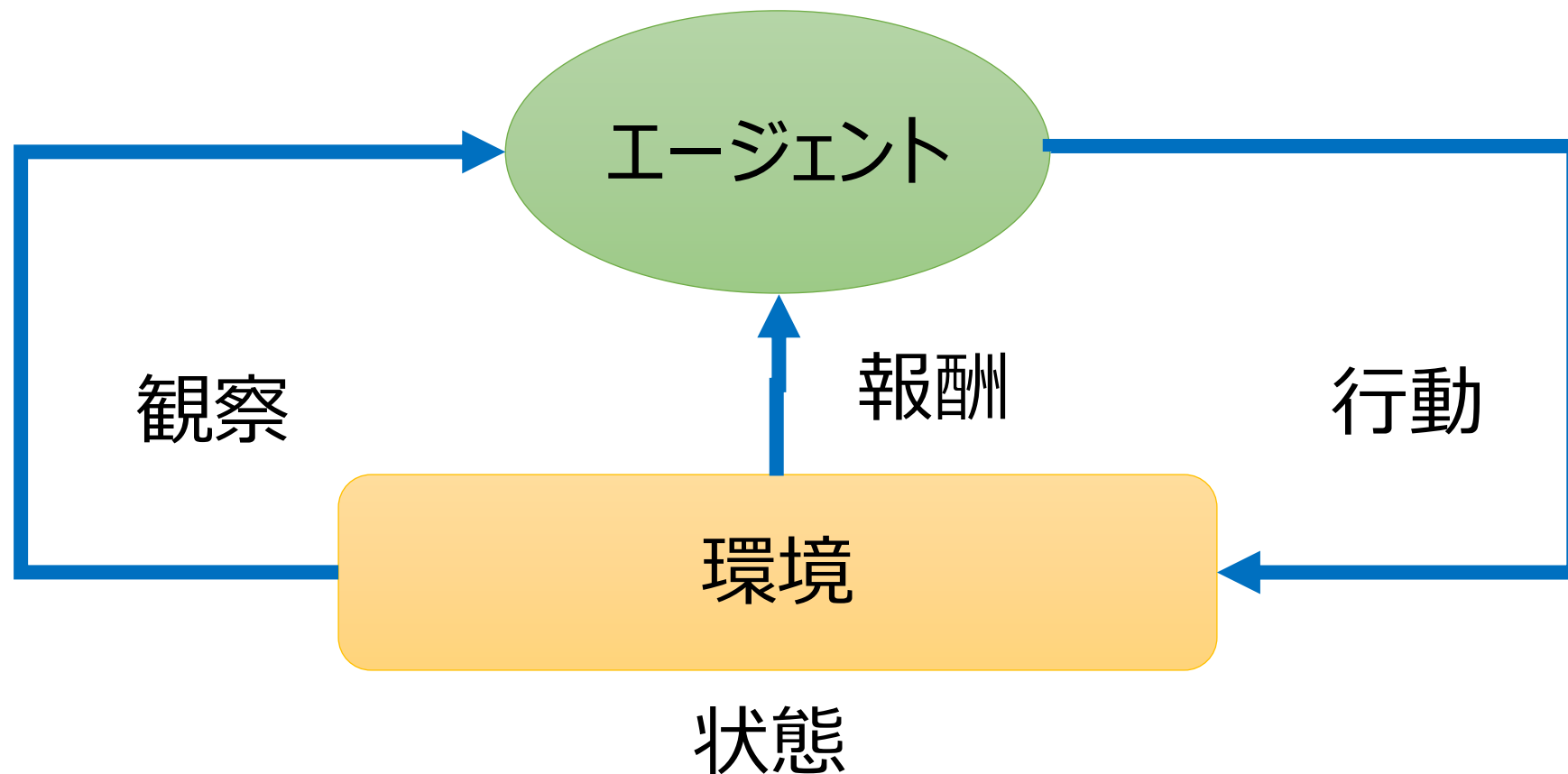
<https://satoyoshiharu.github.io/nlp/>

機械学習

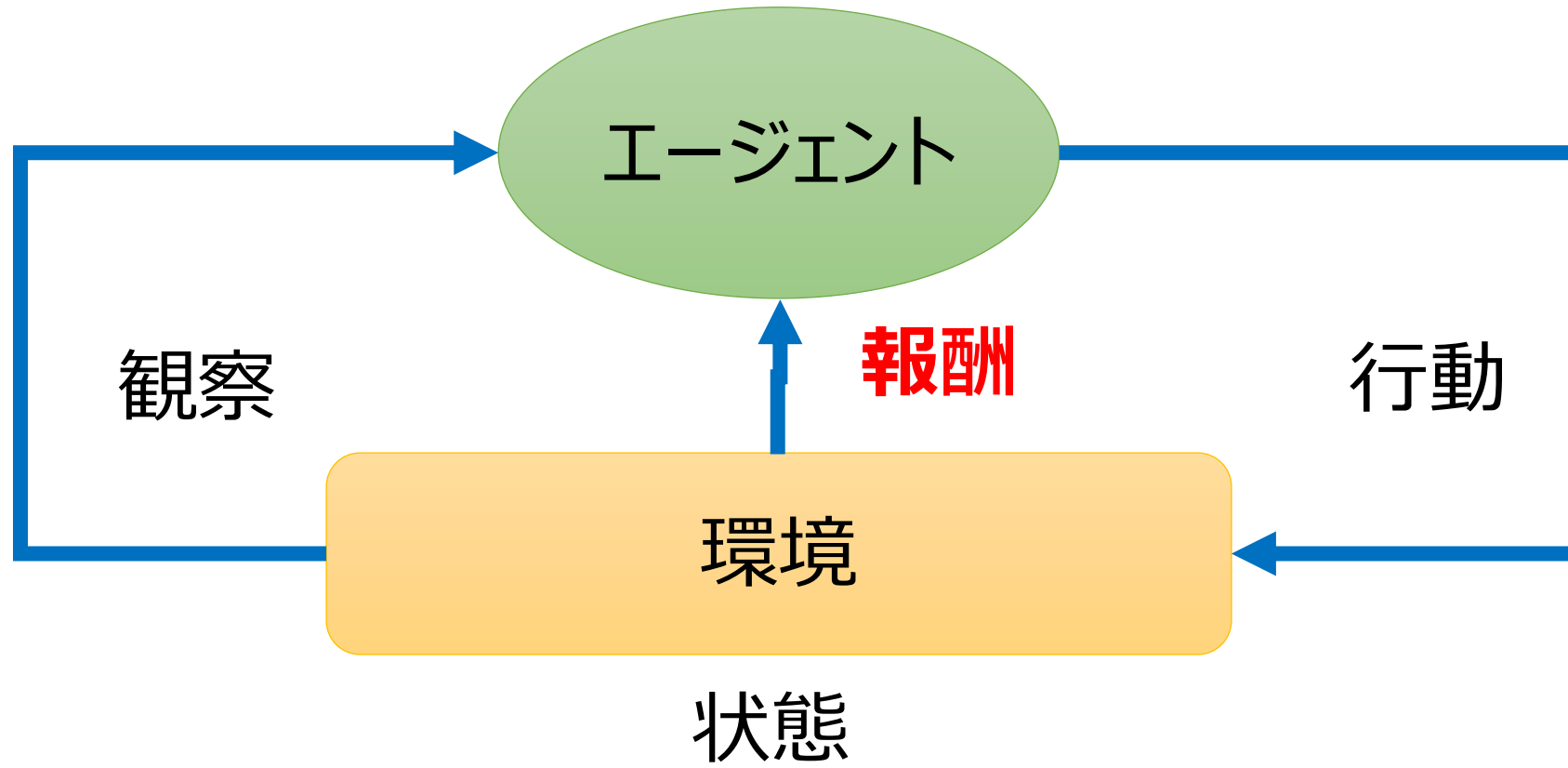


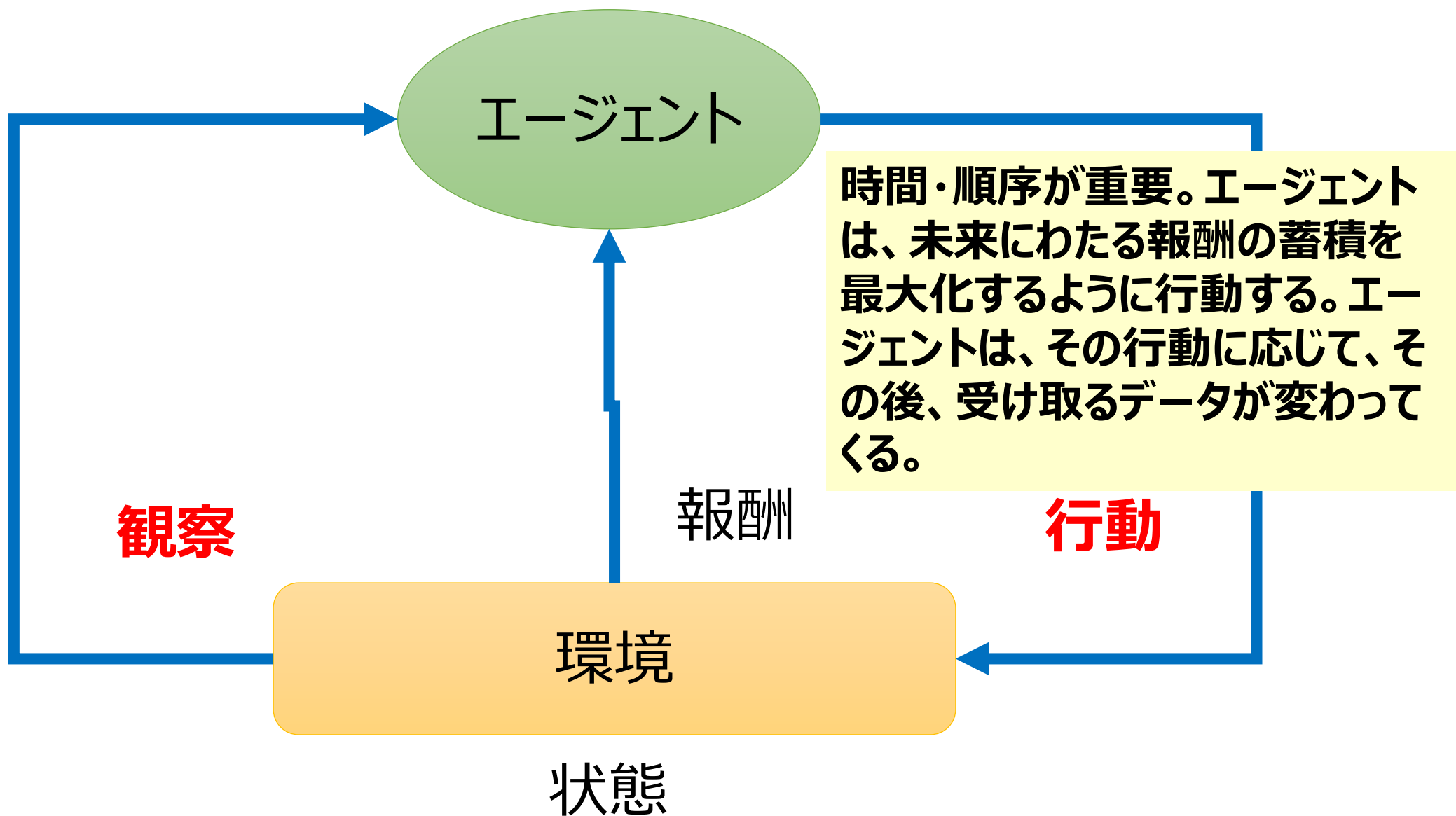
強化学習の基本的な考え方

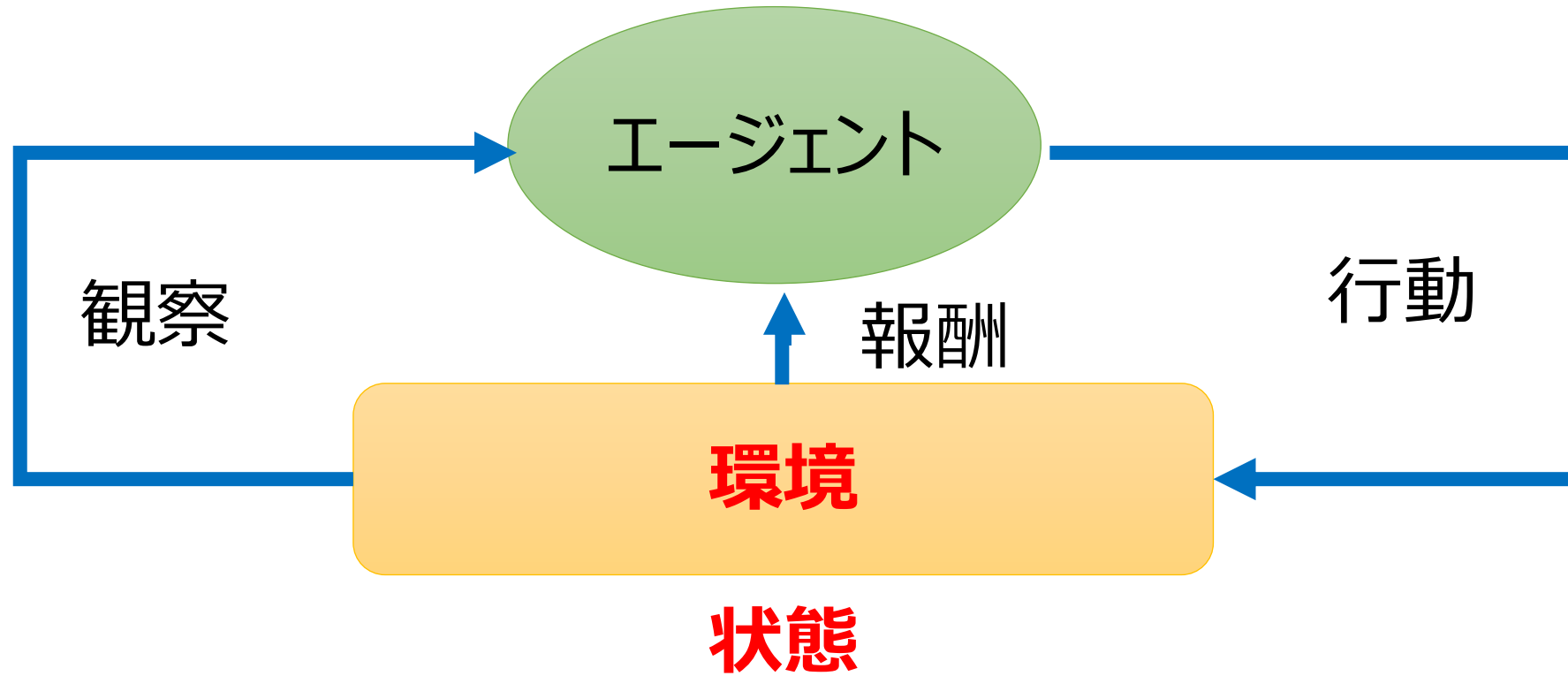
エージェントが環境に働きかけて学習する



教師なし、報酬信号（ある時刻でのスカラー値）だけがある。即時の報酬でなく、未来にわたる報酬の蓄積を気にする。それは後でわかる。例：棒が倒れたかそうでないか、ゲームに勝ったか負けたか。







- 状態が、履歴から独立して、直前の状態だけで決まる場合、マルコフ的という ($P[S_{t+1}|S_t] = P[S_{t+1}|S_1, \dots, S_t]$)。強化学習は、マルコフ状態を前提としたマルコフ決定理論として定式化されている。
- 多くのエージェントは環境の知識(モデル)を持たない。ゲームへの応用ではゲームの規則さえ知らずにいいパフォーマンスを出す。膨大なプレイデータで見様見真似している。

強化学習の構成要素

方策

- エージェントの行動を決めるポリシー

価値

- ある状態ないし行動がどれだけいいか

モデル

- エージェントが環境を表現する

エージェントのタイプ

状態の価値の漸化式的な関係を利用する

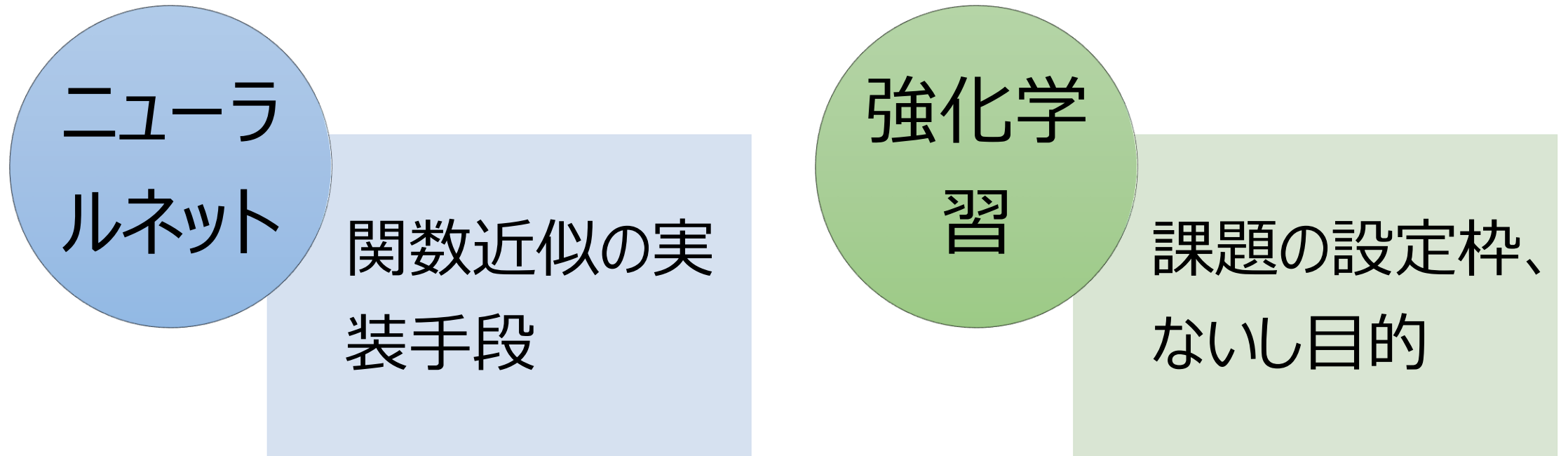
- 代表的なアルゴリズム：Deep Q-Learning
- 応用例：Googleのレコメンド

方策を直接最適化する

- 代表的なアルゴリズム：ポリシー・グラディエント法
- 応用例：AlphaGo

Actor Critic（価値とポリシー両方を最適化する）

強化学習はニューラルネットを利用



強化学習全般に関する参考資料

- [深層強化学習アルゴリズムまとめ](#)
- [ゼロから始める深層強化学習](#)
- [作りながら学ぶ強化学習](#)
- 以下、英語
 - [Introduction to Reinforcement Learning with David Silver](#)
 - [A friendly introduction to deep reinforcement learning, Q-networks and policy gradients](#)
 - [Tutorial: Deep Reinforcement Learning](#)
 - [Reinforcement Learning: An Introduction](#)

Deep Q-Learning 参考資料

- [ゼロからDeepまで学ぶ強化学習](#)
- [【入門】Q学習の解説とpythonでの実装 ～シンプルな迷路問題を例に～](#)
- [PyTorch「強化学習（DQN）チュートリアル」 サンプルコード](#)

Policy Gradient 参考資料

- 英語

- [Deep Reinforcement Learning: Pong from Pixels](#)
- [サンプルコード pg-pong.py](#) (現状ROMなんとかエラーが出る)