

07

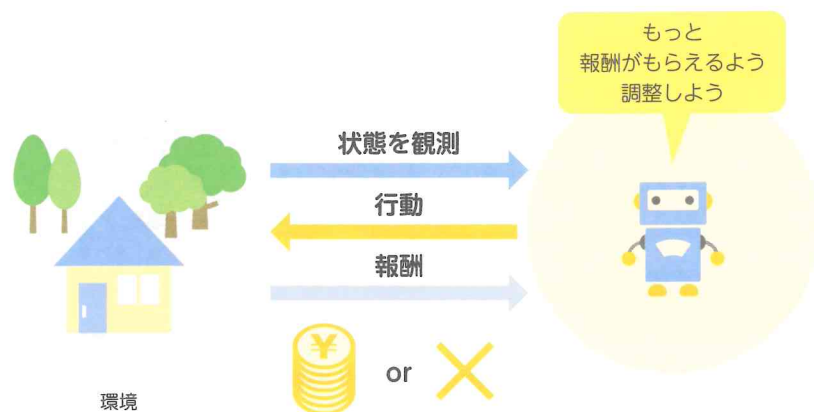
強化学習のしくみ

強化学習とは、与えられた環境とのやり取りから、自らの成果を最大化するように何度も試行錯誤をくり返し、最適な挙動をするように学習することです。教師あり学習とも教師なし学習とも異なる問題設定を持っています。

○ 強化学習とは

強化学習とは、赤ちゃんが一人で立ち上がれるようになるのと同じように、正解を与えなくても試行錯誤をくり返しながらか最適な行動をするように学習する方法のことです。教師あり学習には明示的な正解がありましたが、強化学習にはありません。かわりに、その行動がどれだけよかったのかを報酬として与え、その報酬が高くなるような行動をするように仕向けるのです。教師なし学習も正解はありませんが、強化学習とは性質が全く異なります。前者はデータそのものの特徴を学習しますが、後者は最適な行動を学習するからです。

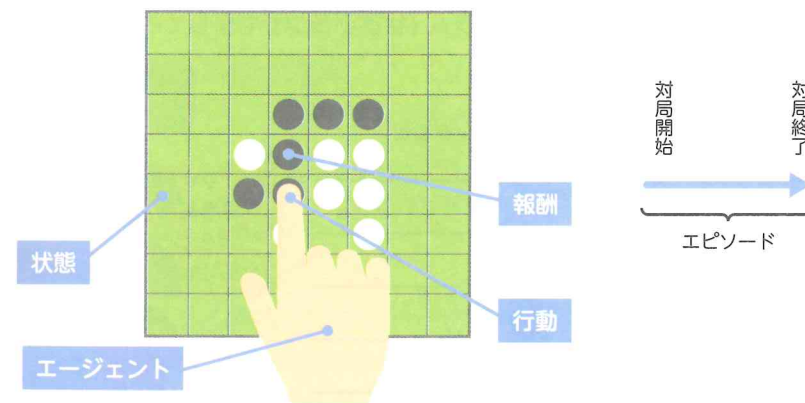
■ 強化学習のメカニズム



強化学習の定義は前ページで述べたとおりです。強化学習でよく使われる用語をオセロを例に紹介しますので、ぜひ頭にいられておいてください。

■ 強化学習で使われる用語

状態 (State)	石の配置・石をおけるマス場所などが「状態」にあたります。
行動 (Action)	石をどこに置くかが「行動」にあたります。
エージェント (Agent)	オセロのプレイヤーのような行動主体を「エージェント」と呼びます。
報酬 (Reward)	行動を起こした結果得られる価値を「報酬」といいます。オセロの場合、石を置いたときに、相手の石をひっくり返した枚数がこれにあたります。
方策 (Policy)	「どの状態のときにどの行動を取るか」という、状態と行動の組み合わせのことを「方策」といいます。
収益 (Return)	「将来まで考えたときに、どれだけの報酬を得られるか」を表します。
Q値 (Q-Value)	「ある状態において、その行動がどれだけよいのか」という行動の価値を表します。目先の報酬だけでなく、将来得られる報酬も考慮します。
V値 (V-Value)	「その状態がどれだけよいのか」という状態の価値を表します。目先の報酬だけでなく、将来得られる報酬も考慮します。
エピソード (Episode)	オセロの1対局のように、行動し始めてから行動ができなくなるまでの一連のまとまりをエピソードといいます。



まとめ

▶ 強化学習はより多くの報酬獲得を目指す