第5章

対戦ゲームや自動運転AIの基本アルゴリズム「Qラーニング」

牧野 浩二

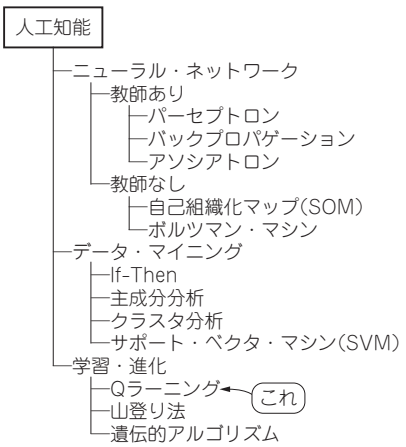

図1　人工知能のアルゴリズムあれこれ

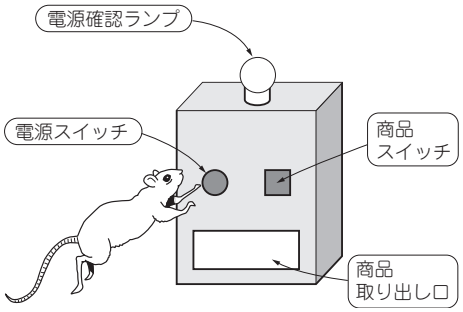

図2 「Qラーニング」は報酬がたくさん得られるように学習する
電源スイッチと商品スイッチが付いた自動販売機があるかごの中に1匹のネズミがいる

進化アルゴリズム「Qラーニング」

人工知能アルゴリズムの中で学習しながら進化するアルゴリズムとしては，次の3つがあります．

- 地味にパラメータを少しずつ変えながら答えに近づく「山登り法」
- 2台のパラメータを交配させて生物のように進化させる「遺伝的アルゴリズム」
- 良かった行動だけに報酬が与えられ「良かった行動」が選ばれるようパラメータを書き換える「Qラーニング」

Qラーニングは，ある決められた行動をしたときに報酬が得られるようになっています．行動するごとに「Q値と呼ばれる行動に対する評価」が得られるようになっていて，たくさん報酬が得られるように，Q値をどんどん更新する方法です．

▶**利点**：望ましい動作だけを与えていればよくて教師データを必要としない
▶**欠点**：望ましい動作をうまく決める必要がある

Qラーニングを応用すると，ロボットの走行/歩行動作の獲得，My検索エンジンの最適化，メールの仕分け，パケットの仕分け，エレベータ制御の最適化などが可能になります．

このQラーニングを応用した「ディープQネットワーク」が注目を集めています．対戦ゲームでは，人間がかなわないような得点を出すことができるようになりました．また，トヨタやNTT，Preferred Networksによる「ぶつからない車」にも使われています(1)．最新の人工知能のアルゴリズムを理解するためにも元となるQラーニングの仕組みを知っておくことが重要です．

Qラーニングのお約束の数式

いきなりQラーニングの数式を書きます．Qラーニングはこの式に従って計算が進んでいくので，説明のために必要となります．この後簡単な例を用いて説明していきます．

$$Q(s,a) \leftarrow (1-\alpha)\,Q(s,a) + \alpha\,(R(s,a) + \gamma \max Q(s',a')) \quad \cdots\cdots(1)$$

ただし，s：今回の状態，a：今回の状態での行動，s'：次の状態，a'：次の状態での行動，α：定数（学習率），γ：定数（割引率），R：報酬，Q：行動選択に必要な値とします．