

事例：深層学習による物体認識

出典: A. Krizhevsky et al., ImageNet classification with deep convolutional neural networks, 2012.



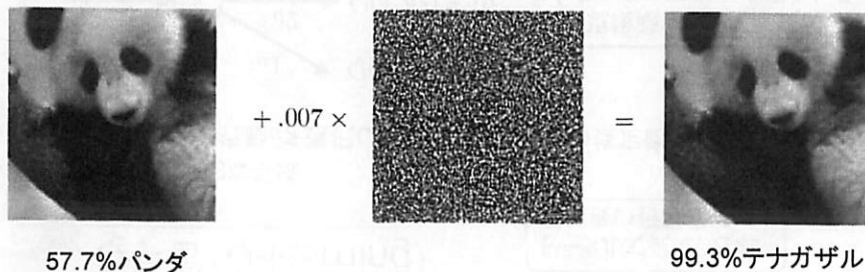
- 1000種類のクラス
- 各クラス1000枚の画像から学習

14層, ニューロン
65万, パラメータ
6000万

17

深層学習の特徴と(現時点での)限界

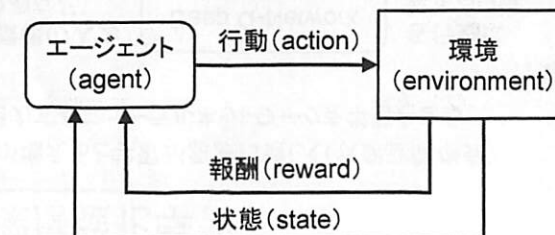
- 大量のデータから, 自動的に学習をさせられる(ビッグデータの活用)
- 汎化 (generalization) 能力の高さ
 - 学習で得た知識や経験を一般化 ⇒ 未知のデータにも対応できる
- 過学習は抑制したい
- 結局は, 多数のパラメータ(重み)の自動調整をしている
- AIはだませる?



19
図出典 Goodfellow et al., Explaining and Harnessing Adversarial Examples

強化学習 (Reinforcement learning)

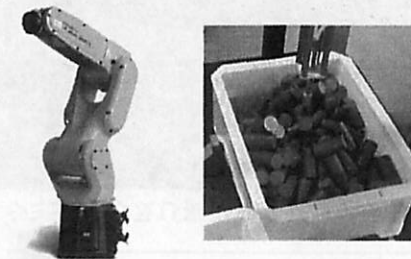
- 試行錯誤をしながらエージェントが最適な行動を見つけ出す機械学習
- エージェントの行動により, 環境の状態が変化し, 報酬を受けとる。
- 総報酬 (収益, return) が最も良くなる方策 (行動戦略, policy) を学習する。
- 試してみないと, どれが良い行動か分からない (探索と利用のトレードオフ)
- ある行動をとった直後の報酬 (即時報酬) が低くとも, うまく行動していけば, その後高い報酬をもらえるかも (遅延報酬)



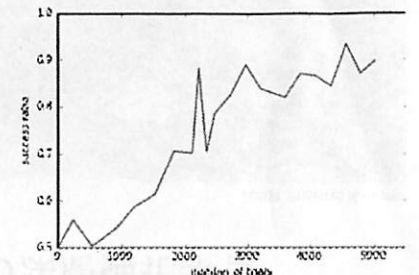
20

事例：深層学習でバラ積みロボットの学習

- Preferred Networks (2016~)
 - 深層学習を様々な分野に展開しての研究, 開発
 - FUNUC, トヨタ, パナソニックなどの企業との協業
- 乱雑に置かれたワークから物体を取り出す動きを学習 (国際ロボット展2015)
- 事前知識無しの状態から約10時間で成功率90%に



FANUC ロボットアーム
LR Mate 200iD



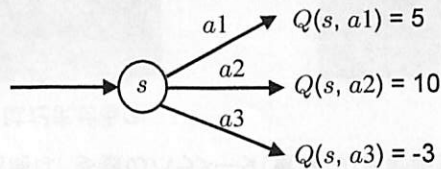
学習に伴う取得成功率の向上

図出典: Preferred Research

Q学習 (Q-learning)

トータルで見ても、
どれだけ良い行動か？

- 強化学習の代表的な手法
- $Q(s, a)$: 状態 s で行動 a を選択したときの期待収益の推定値 (Q値, 行動価値)



(現段階では) 行動 a_2 が
最も良い行動と推定される

- 行動 a の結果の状態 s' と、受け取った報酬 r により、Q値を更新していく

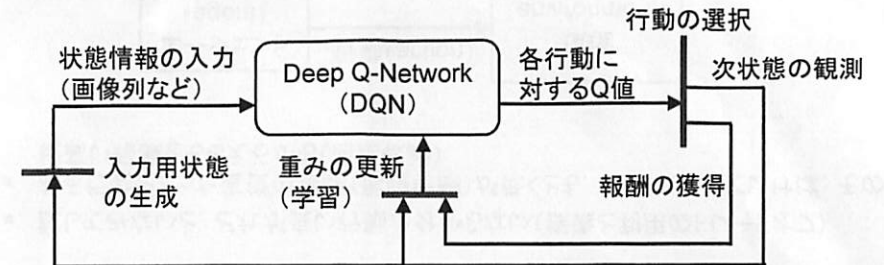
$$Q(s, a) \leftarrow Q(s, a) + \alpha \left(r + \max_{a' \in A(s')} Q(s', a') - Q(s, a) \right)$$

- 試行錯誤を重ねることで、Q値の推定がより正しく \Rightarrow 最適な行動が分かる
- 報酬をどのように設定するか？

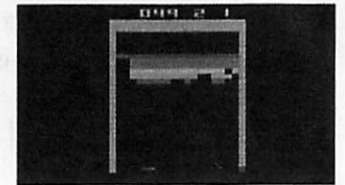
21

深層強化学習

- 状態数が増えると学習が急激に難しく(状態空間爆発)
- Q値を出力するニューラルネットワークを学習させる



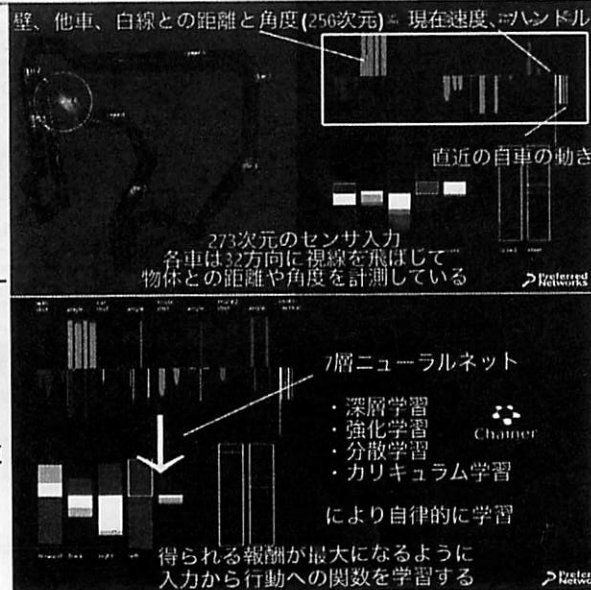
- 一部のゲームで、画面情報のみから人間以上のスコアを獲得できることが示された (DeepMind, Nature誌, 2015)



事例:レーシングカーの深層強化学習

図出典: Preferred Networks

- Preferred Networks (2015)
- ロボットカーが、ぶつからな動作を0から学習
- 周囲の状況をセンサで認識し、ニューラルネットワークに入力
- 道に沿って速い速度で進めばプラスの報酬を
- 壁や他の車への衝突、逃走にはマイナスの報酬



事例:レーシングカーの深層強化学習

- シミュレータ及び実機で有効性が示された

図出典: Preferred Networks



図出典: Car Watch

- CES2016での、深層強化学習を使った自動運転車のデモ (トヨタ自動車, Preferred Networks)

