

第1回 強化学習勉強会

10/22 19:00～@リクルート

慶應義塾大学大学院 理工学研究科 田中一樹

慶應義塾大学 理工学部4年 池田春之介

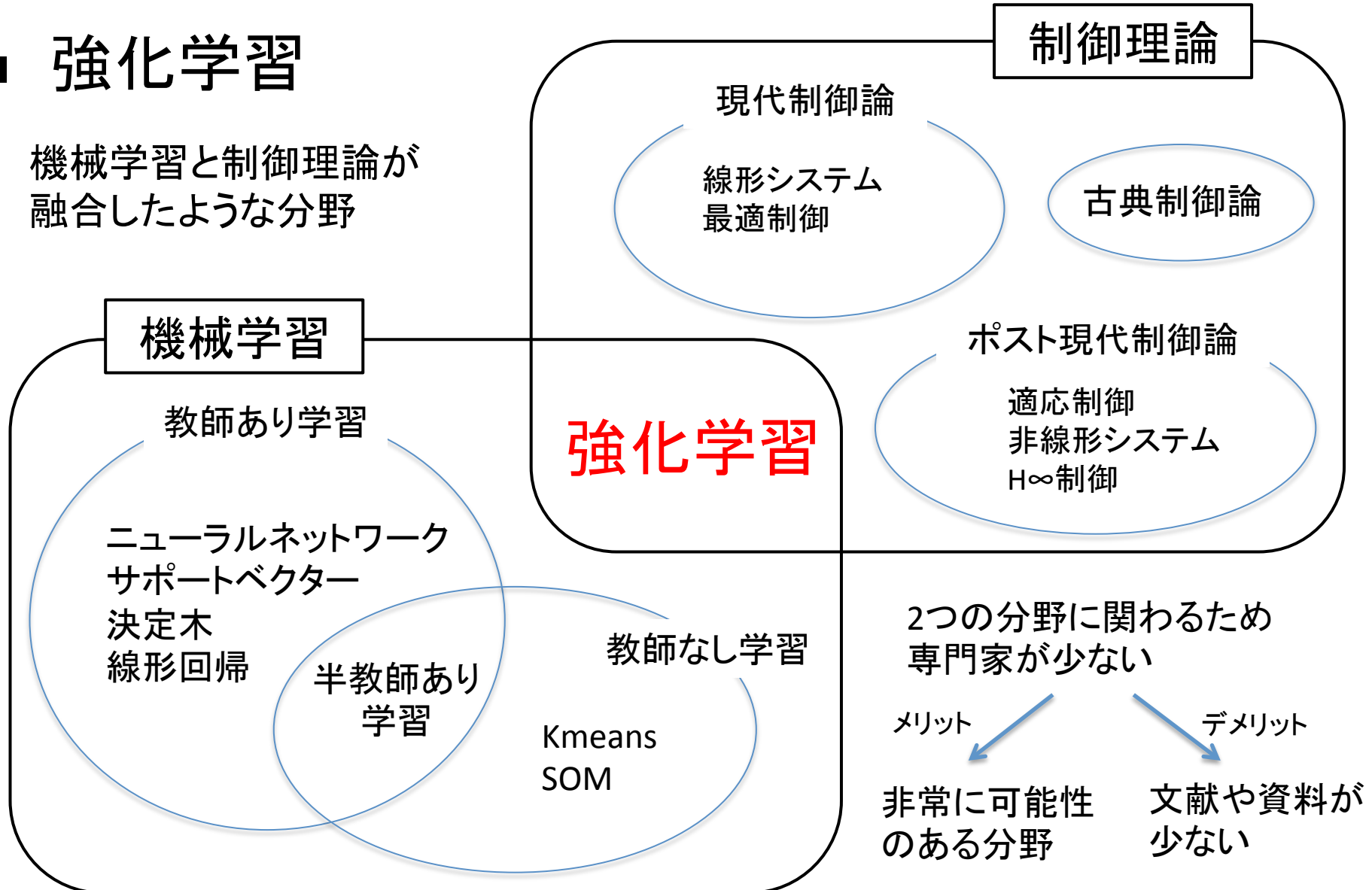
目次

1. 概要
2. 先端研究
3. 参考図書等
4. 自分の研究について

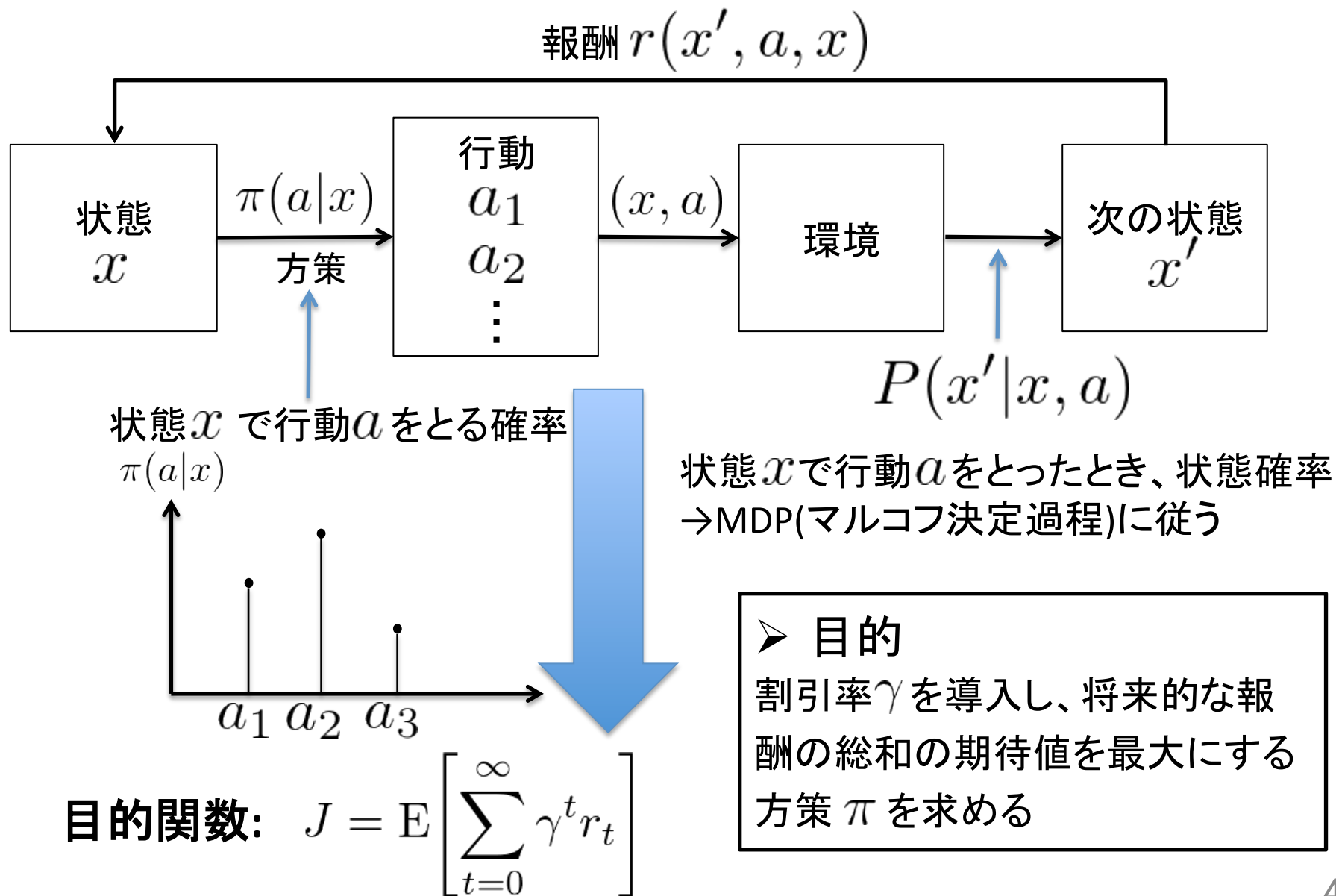
強化学習の概要

■ 強化学習

機械学習と制御理論が
融合したような分野



強化学習の概要(1)

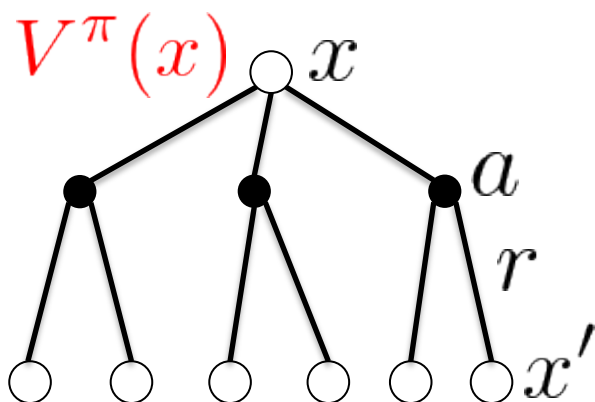


強化学習の概要(2)

➤ 価値関数

状態価値関数

$$V^{\pi}(x) = E^{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r_t \middle| x_0 = x \right]$$



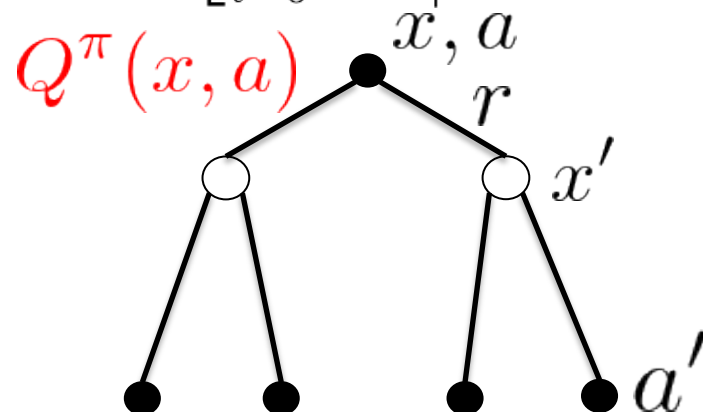
最適方策

$\pi^*(x) = \arg \max_a Q^{\pi}(x, a)$ を求めたいがすべてのパスを計算するのは大変

➡ 最適方策 $\pi^*(x)$ を近似する ➡

行動価値関数

$$Q^{\pi}(x, a) = E^{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r_t \middle| x_0 = x, a_0 = a \right]$$



モンテカルロ法、TD法、Q学習、
アクター・クリティック法 etc...

先端研究(1)

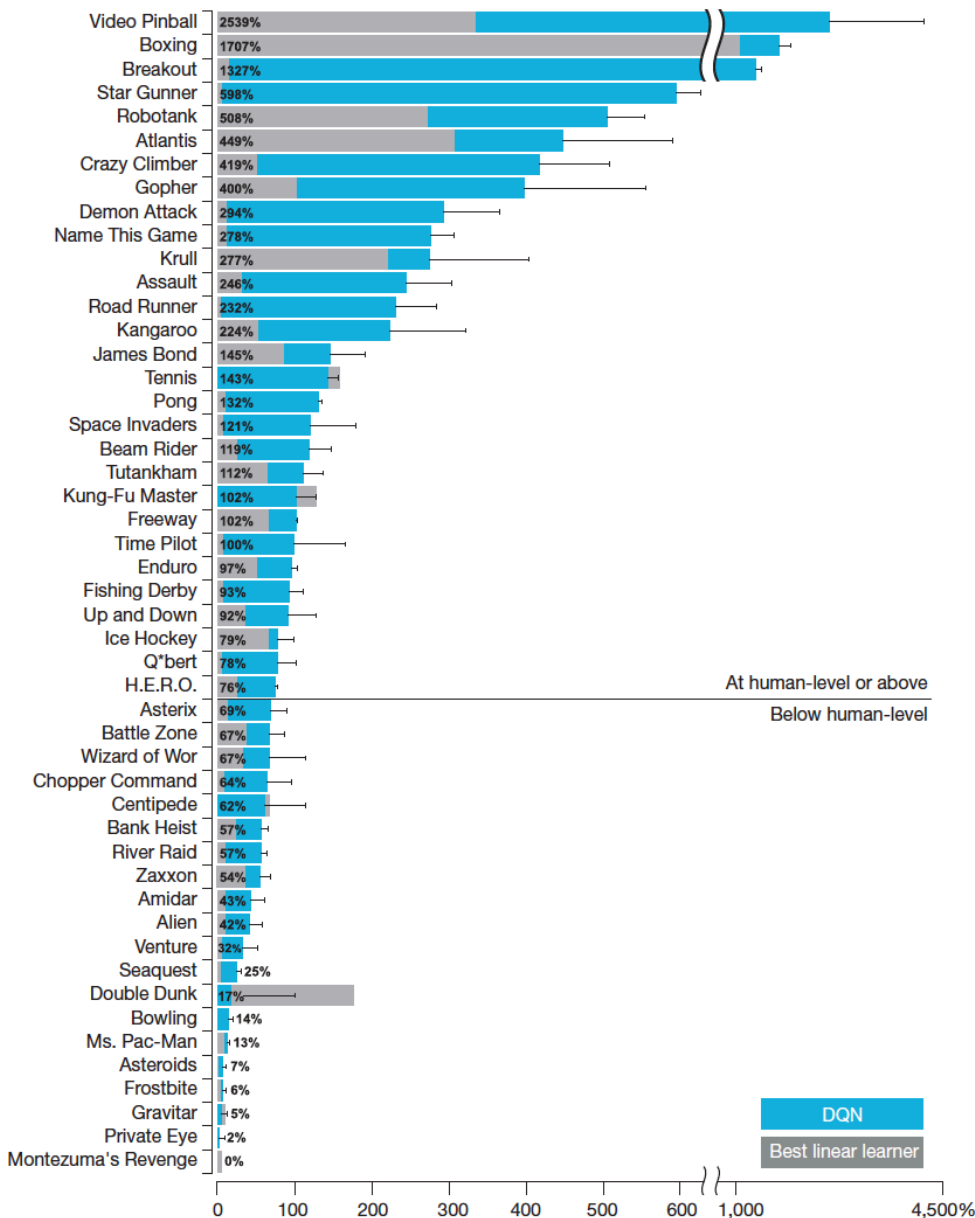
■ DQN(Deep Q-Network)

- 強化学習と深層学習を組み合わせた人工知能
- Q関数をDeep learningで近似
- Googleの子会社Deep mindが開発
- 論文

V. Mnih *et al.*, "Playing atari with deep reinforcement learning"

V. Mnih *et al.*, "Human-level control through deep reinforcement learning"

先端研究(1)



人工知能
「DQN」



Atari 2600のゲーム49種類を
プレイさせた




49ゲーム中43ゲームで、機械
学習手法を上回り、29のゲーム
ではプロのゲーマーと同等以上
のパフォーマンスを見せる

DQN
Best linear learner

先端研究(2)

■ 分散深層強化学習

- ロボット制御に深層強化学習を組み込む
- ロボット同士が協調して衝突を防ぐ

 Preferred Networks 分散深層強化学習デモ

Reinforcement

×

Distributed

×


Deep Learning

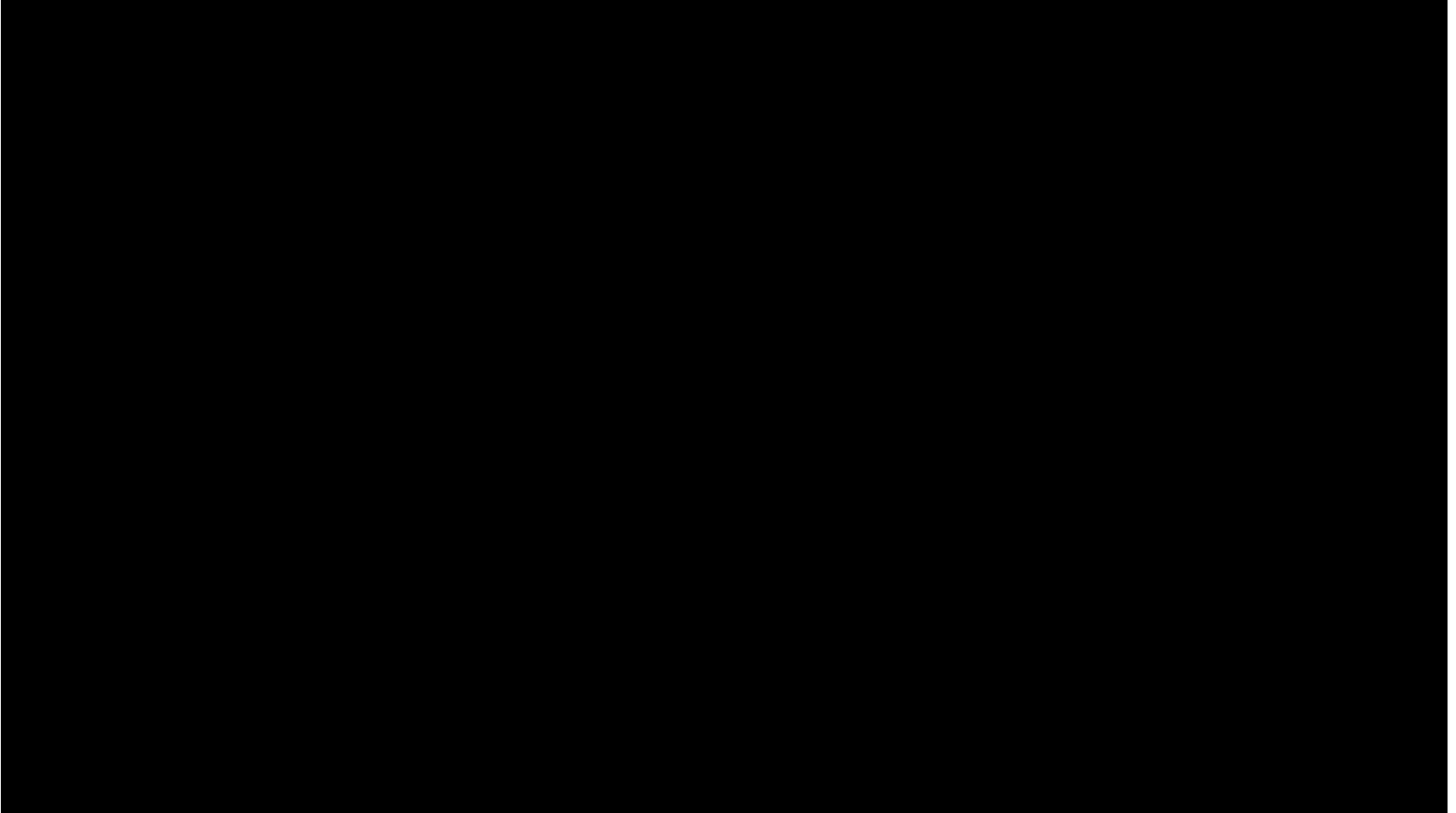
先端研究(3)

■ 逆強化学習(Inverse Reinforcement Learning)

- 行動系列を教師データとして、ある環境における理想的な行動規則を学習(報酬関数を学習)する技術
- 応用例
 - ・ヘリコプター制御:
 - ➡ 人間のエキスパートの技を学習させるため、報酬関数を未知として、人間のエキスパートの行動を教師データとし報酬関数を推定する
 - ・自然言語処理:
 - ➡ 正解データから得られる最適訓練行動列をもとにして、逆強化学習によって報酬関数を設計する

先端研究(3)

ヘリコプター制御の映像 



先端研究(3)

■ 階層的強化学習

- 複雑なタスクを階層的に分解し、各部分に局所的な方策を学習してから統合することによって大域的な方策を学習

- 応用例

音声対話システム:

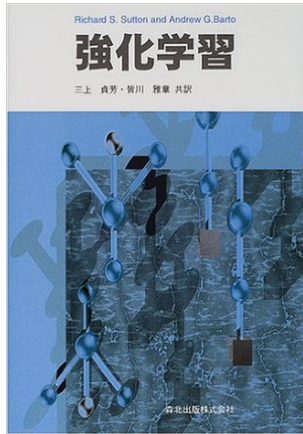
➡ ユーザーの発話内容を観測し要求を明らかにして対応するシステムにおいて、ノイズによって発話内容が正しく観測できない場合でも、従来のシステムに比べはるかに良い応答

論文:

Yasuhide Kishimoto *et al.*, "Spoken Dialogue Manager Using Partially Observable Markov Decision Processes with Hierarchical Reinforcement Learning "

Steve Young *et al.*, "Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning"

参考図書



Sutton and Barto (1998),
Reinforcement Learning

- 定番の一冊
- わかりにくい？

原文(HTML)



<http://webdocs.cs.ualberta.ca/~sutton/book/ebook/the-book.html>



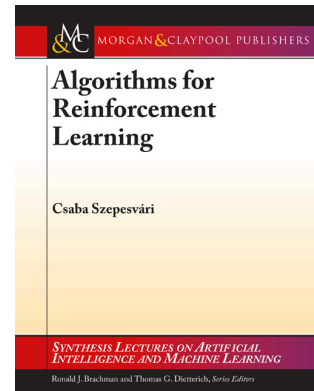
八谷 大岳, 杉山 将,
強くなるロボティック・ゲーム
プレイヤーの作り方

- 絶版？Kindle版のみ？
- サンプルコードが載っていてわかりやすい



Wiering and van Otterlo (2012),
Reinforcement Learning

- 応用例が豊富



Szepesvári (2010),
*Algorithms for
Reinforcement Learning*

- 100ページ未満(英語)

ツール・ライブラリ

- RL Community

http://rl-community.org/wiki/Main_Page

強化学習の競技会のために作成された、エージェント・環境・実験プログラムの多言語インタフェース RL-Glue、そのインタフェースに沿って実装されたさまざまなアルゴリズム・環境のライブラリ RL-Library が入手できる。

- Policy Gradient Toolbox

<http://www.robot-learning.de/Research/PolicyGradientToolbox>

方策勾配 (Policy Gradient) 法を利用した強化学習の MATLAB 用ツールボックス

- Reinforcement Learning Toolbox

<http://www.igi.tugraz.at/ril-toolbox/general/overview.html>

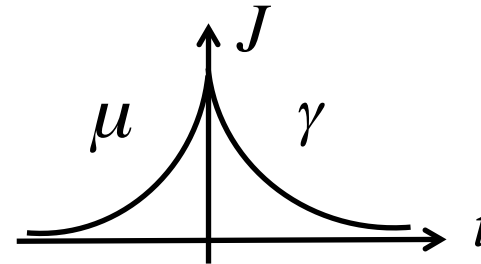
C++ で記述された強化学習のツールボックス

自分の研究について

強化学習のアルゴリズムのひとつであるRLS-TD(0) (TD誤差の逐次最小二乗法) を改良

過去の情報に関する忘却率 μ と未来の情報に関する割引率 γ の両方が含まれるより性能の良い評価関数の提案

評価関数: $J = h(\mu, \gamma)$



得られた新たな更新則をブラックジャックのようなアプリケーションに組み込んで更なる収束速度と汎化能力の向上を目指す