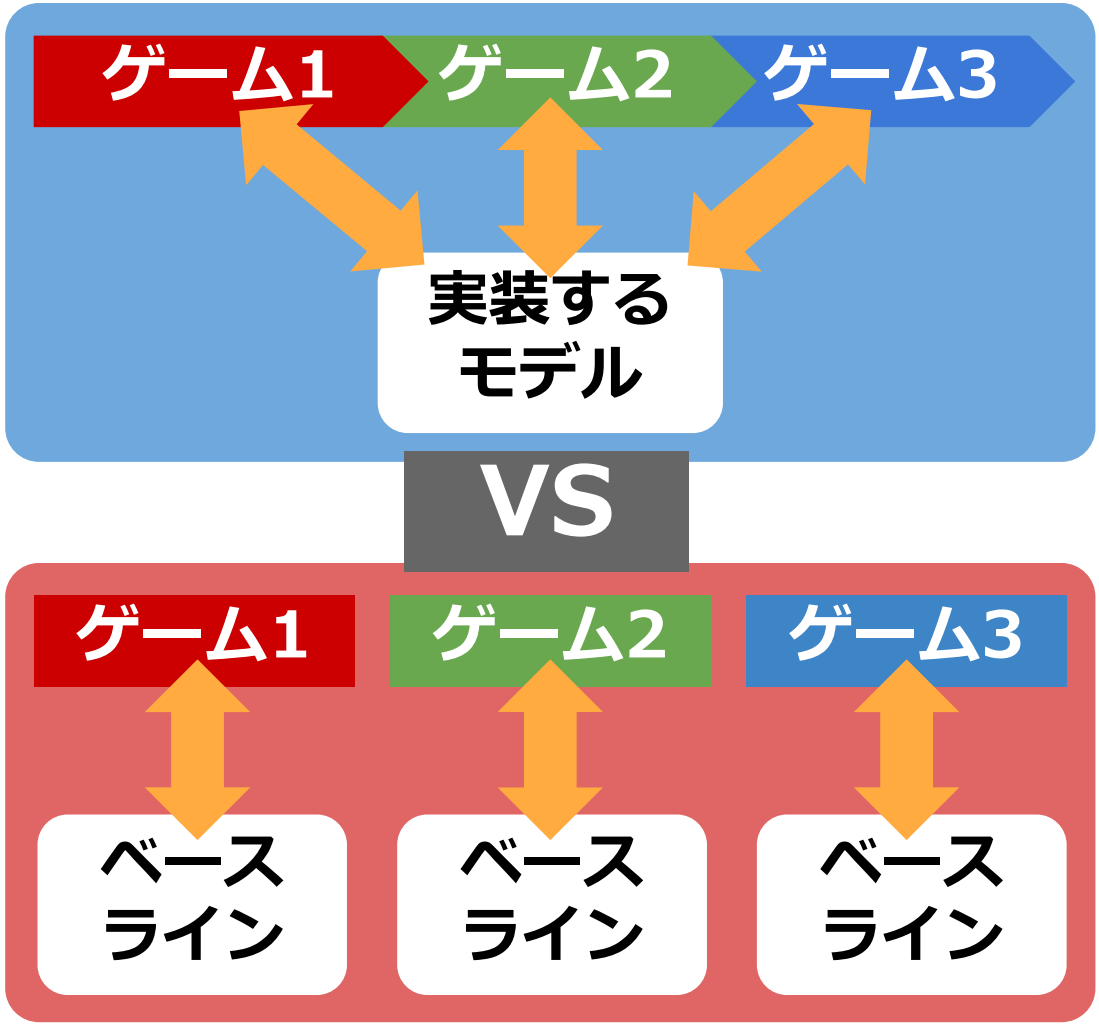
複数のゲームにおける continual learning

慶應大学環境情報学部　野口裕貴

概要

- 複数のゲームを
  逐次的に学習
- 知識の再利用に注目
- 各ゲームを
  独立的に学習した
  ベースラインと比較



知識の再利用

汎用人工知能には過去に獲得した知識を再利用する能力が必要である

It is my conviction that no scheme for learning, or for pattern-recognition, can have very general utility unless there are provisions for recursive, or at least hierarchical, use of previous results.

- Marvin Minsky,
Steps Toward Artificial Intelligence (1960)

知識の再利用｜機能要件

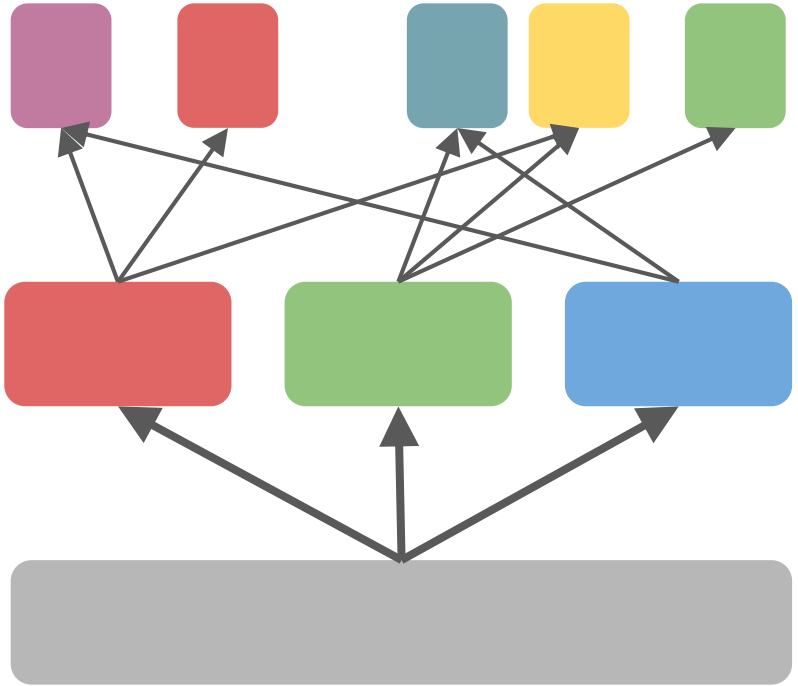
過去に獲得した知識を
保持する

過去に獲得した知識を
新しいタスクに利用する

「continual learning」
「lifelong learning」

理想的な知識構造

タスクに
特化した知識



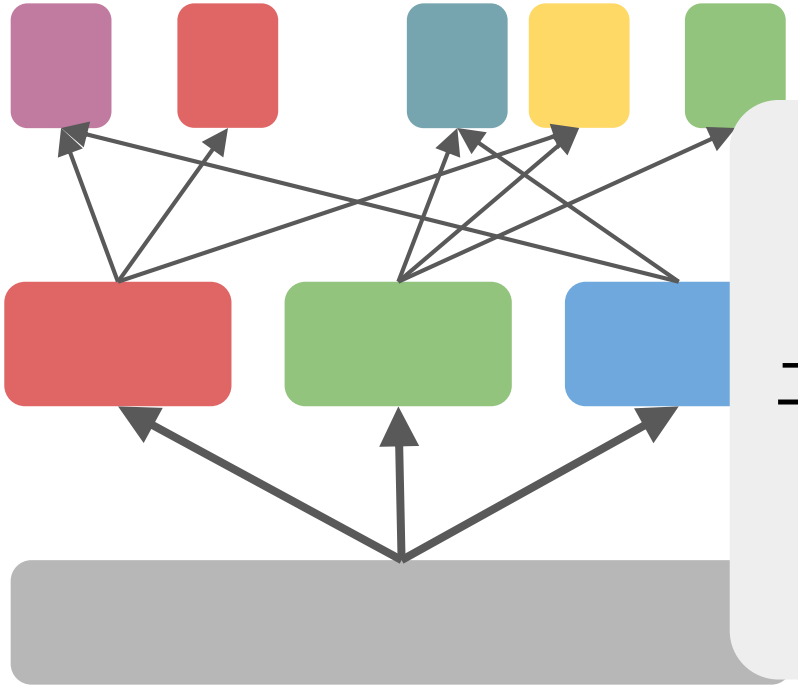
特定の状況に
利用される

汎用的な知識

よく（再）利用
される知識

理想的な知識構造

タスクに
特化した知識



特定の状況に
利用される

汎用的な知識

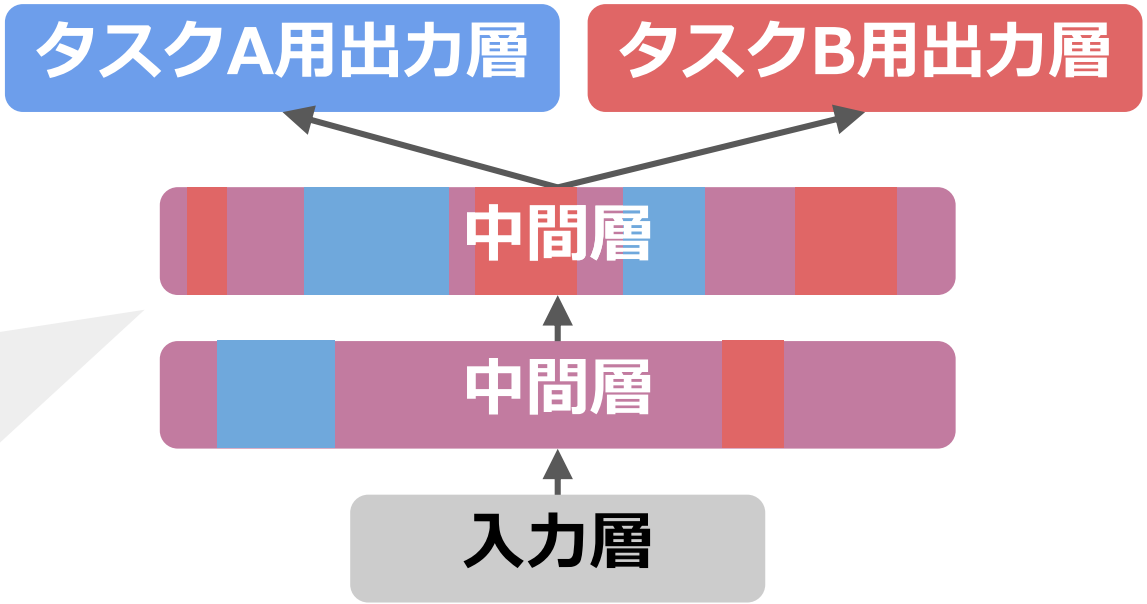
この表現を獲得
するために、
ニューラルネット
の枠組みで
考えてみる

ニューラルネットの性質

典型的なニューラルネットが複数のタスクを学習できるか？

理想

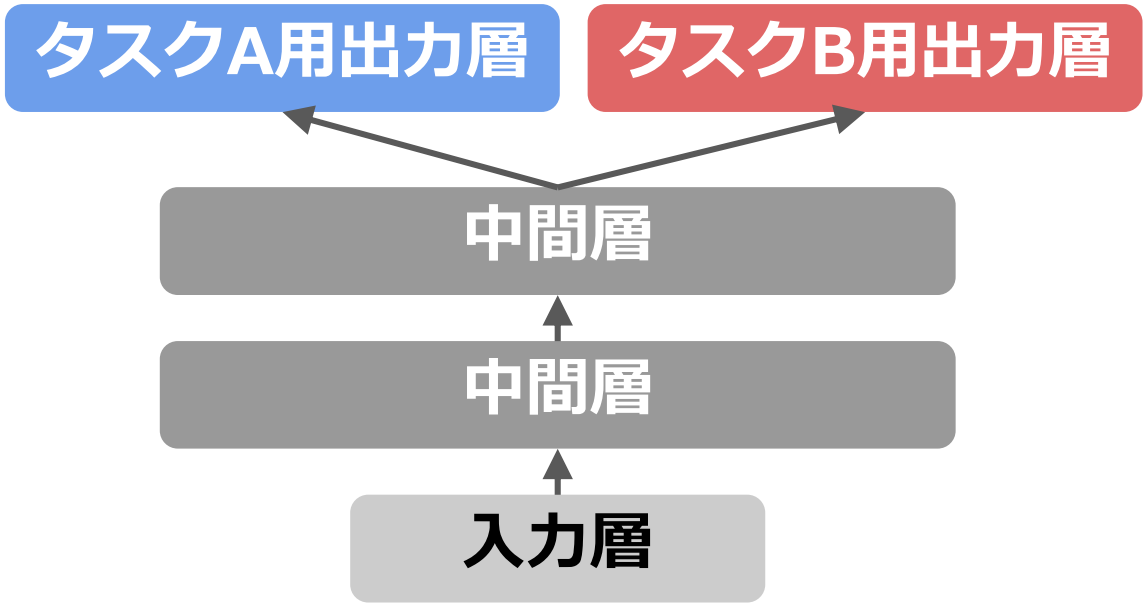
Aに特化、
Bに特化、
さらに**共有**
される知識が
学習される



ニューラルネットの性質

重みに初期値を与える

現実

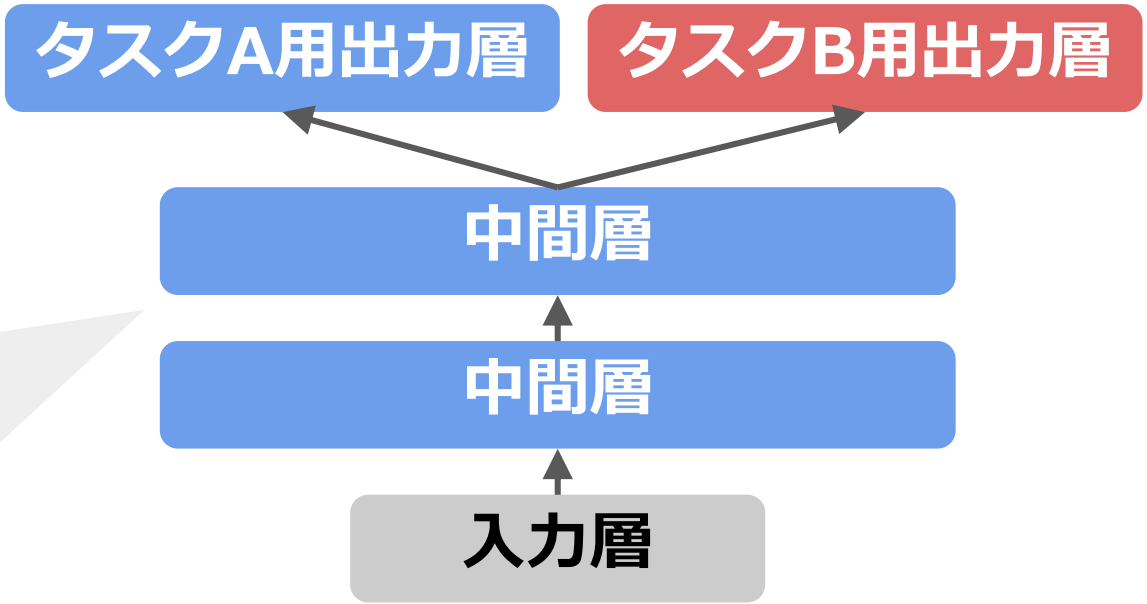


ニューラルネットの性質

まずタスクAを学習

現実

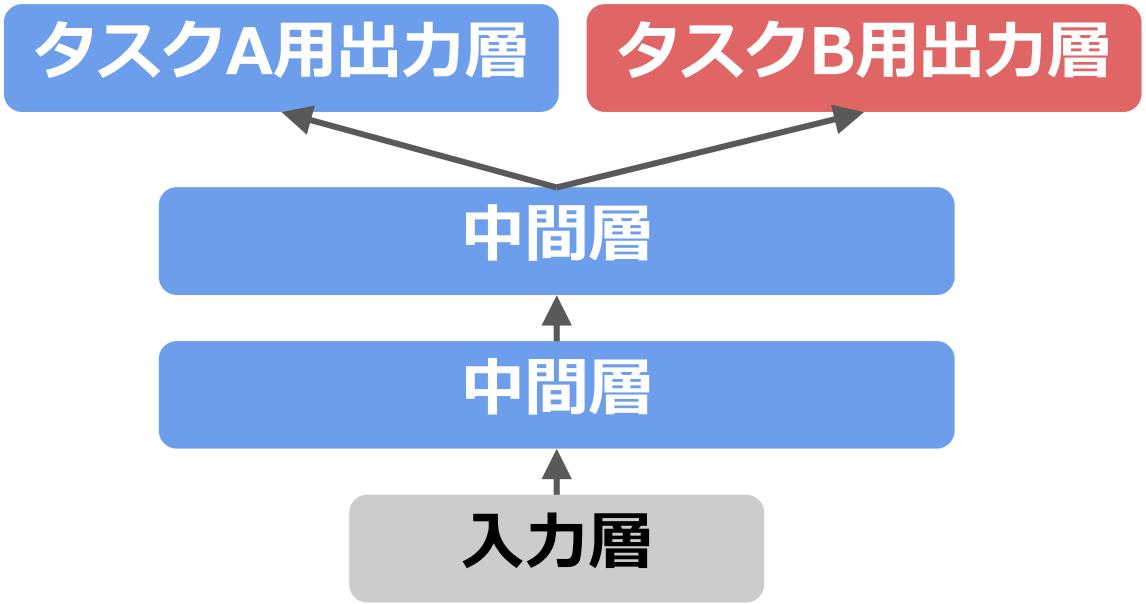
Aに特化した
知識が学習
される



ニューラルネットの性質

次にタスク**B**を学習させれば両方できるようになる．．．？

現実

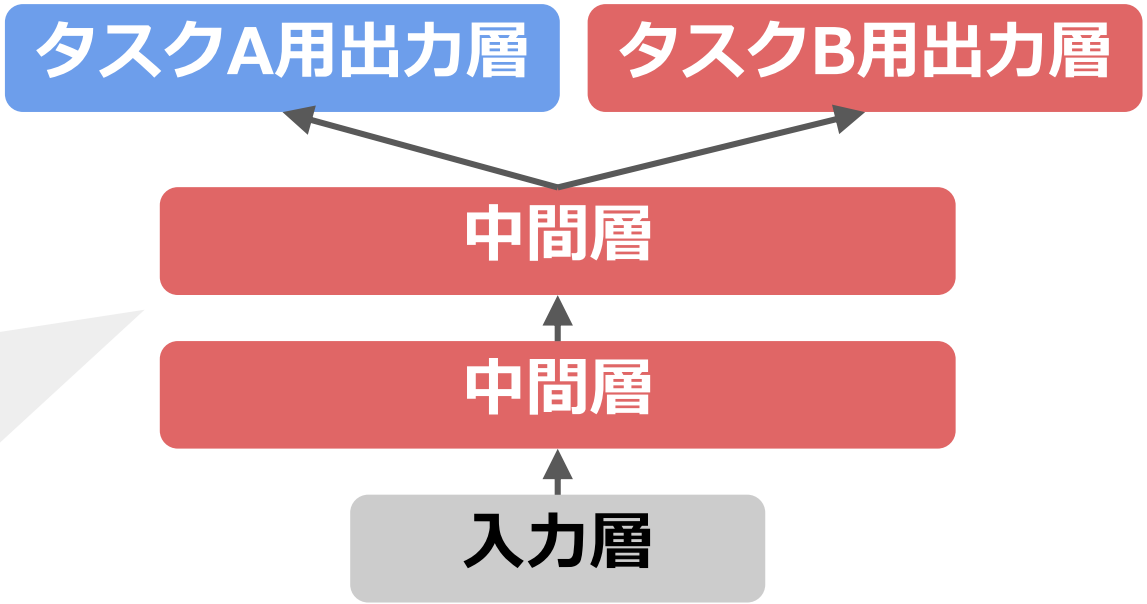


ニューラルネットの性質

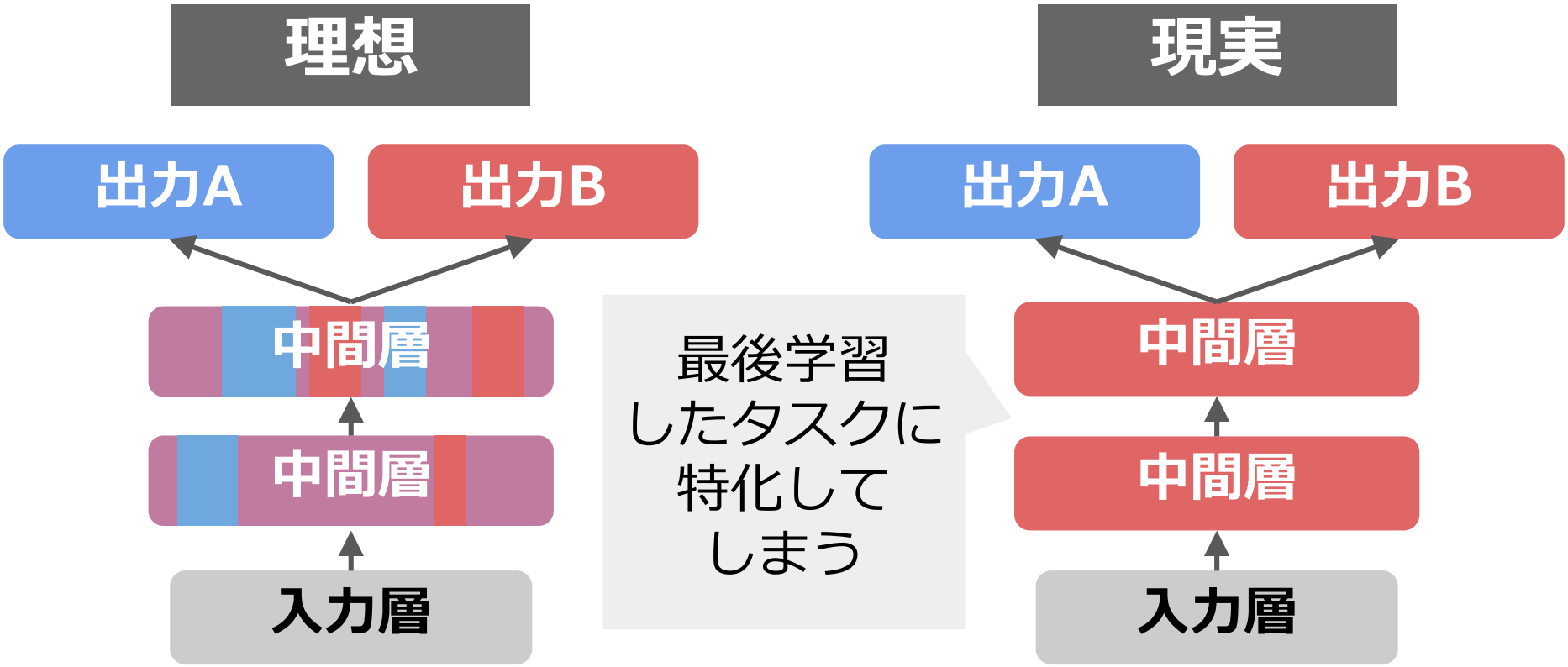
そんなに簡単じゃない．．．

現実

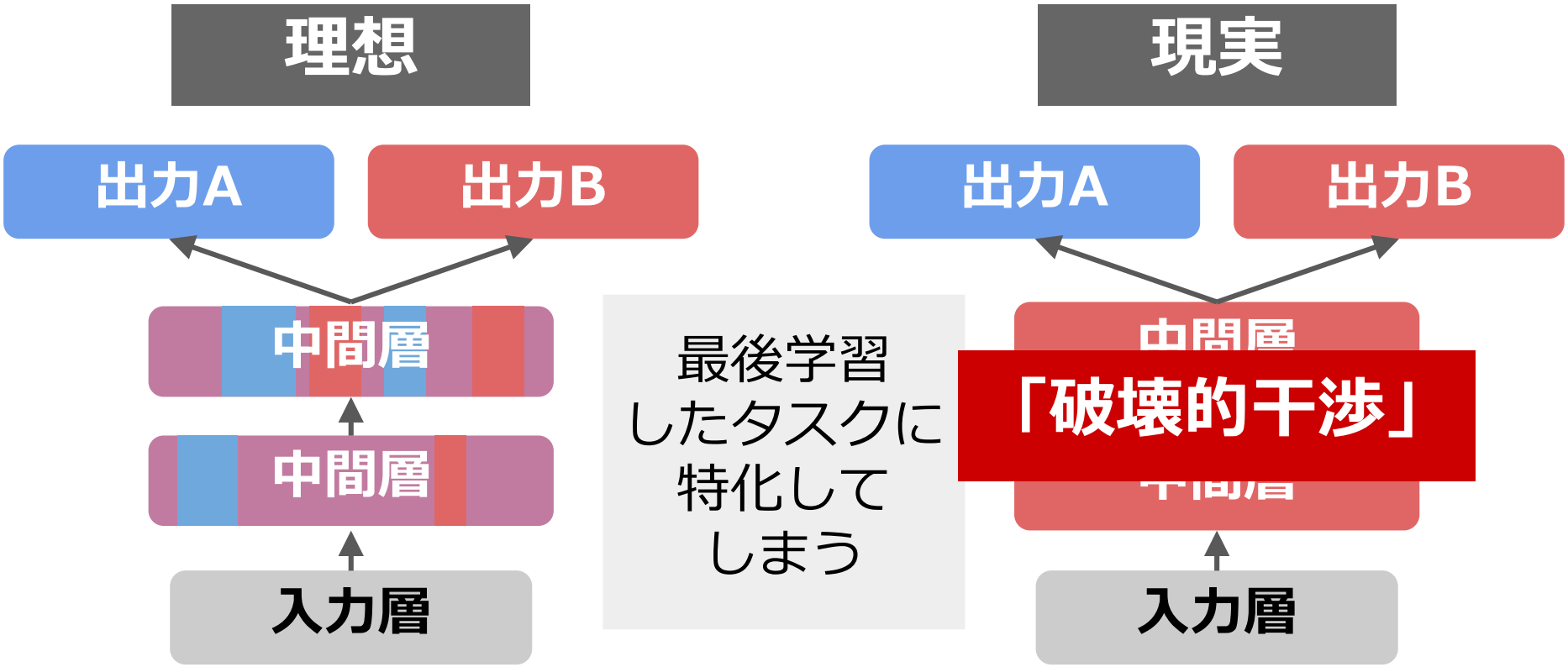
Aに特化した
知識が
「上書き」
され、Bに特化し
てしまう



ニューラルネットの性質



ニューラルネットの性質



破壊的干渉（catastrophic interference）

ニューラルネットの大きな問題

ディープやシャロー、教師あり、教師なしや強化学習
構わず影響される

人間の場合

人間は大事な知識や記憶をある程度忘れない

新たな情報を既存の知識と結びつけることができ、
　それにより学習が速くなる（転移学習）

人間と汎用人工知能

①「**破壊的干渉**」を避ける
過去に獲得した知識を
保持する

②「**転移学習**」を行う
過去に獲得した知識を
新しいタスクに利用する

	①	②
人間	○	○
汎用人工知能	○	○
流行りのNN	×	△

今ハッカソンの目標

Lisを利用し、仮想環境で複数のゲームを遊ばせる
（強化学習）

知識の再利用を行う特殊なモデルを利用する

それとベースラインの学習の速さを比較する

比較と観察を通し、continual learningの有効性を示す

モデル

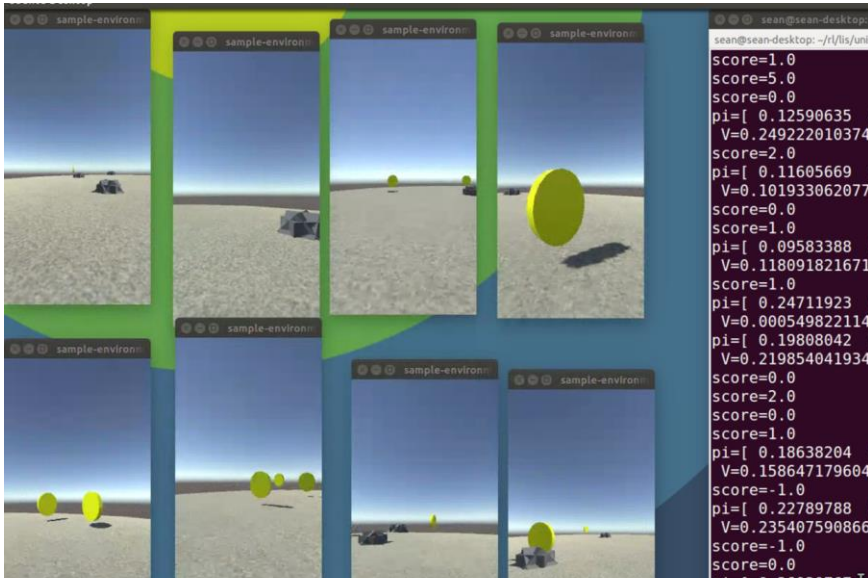
強化学習のモデル：A3C[1]

continual learningを行う手法：Progressive Neural Networks[2]

A3C（Asynchronous Advantage Actor Critic）

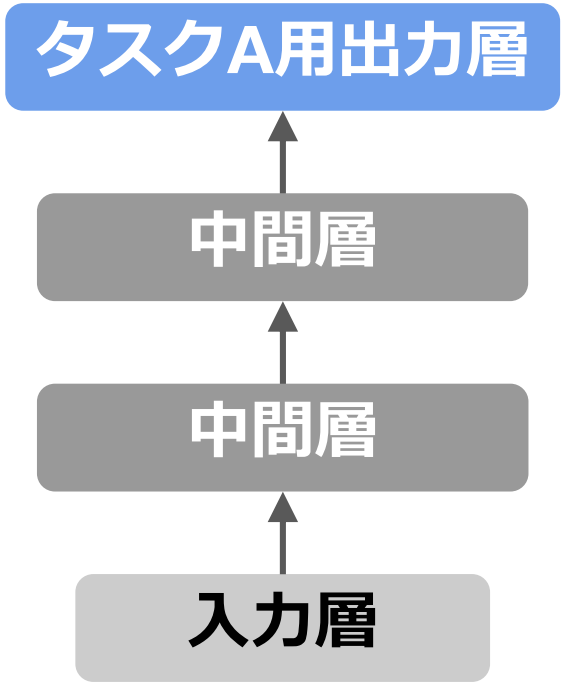
Actor-Critic法で、複数のエージェントを並列に動かし学習

- experience replayが必要なくなる
- DQNより良い結果が出ている
- 今回は８個のエージェントを並列に回した



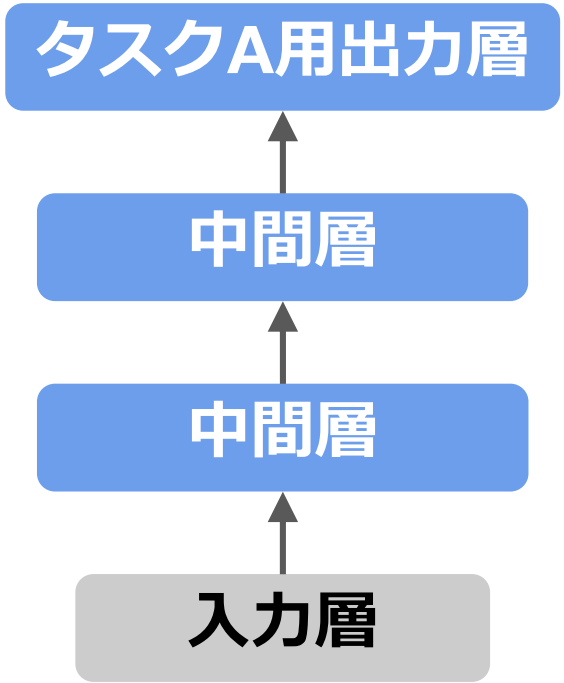
Progressive Neural Networks（PNN）

1．ニューラルネットを用意



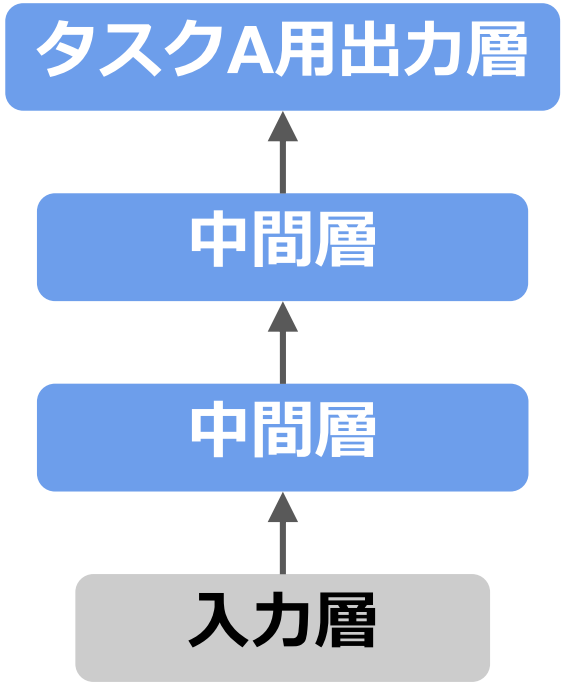
Progressive Neural Networks

２．タスク**A**を学習



Progressive Neural Networks

ここからどうやって破壊的干渉を避けるのか？

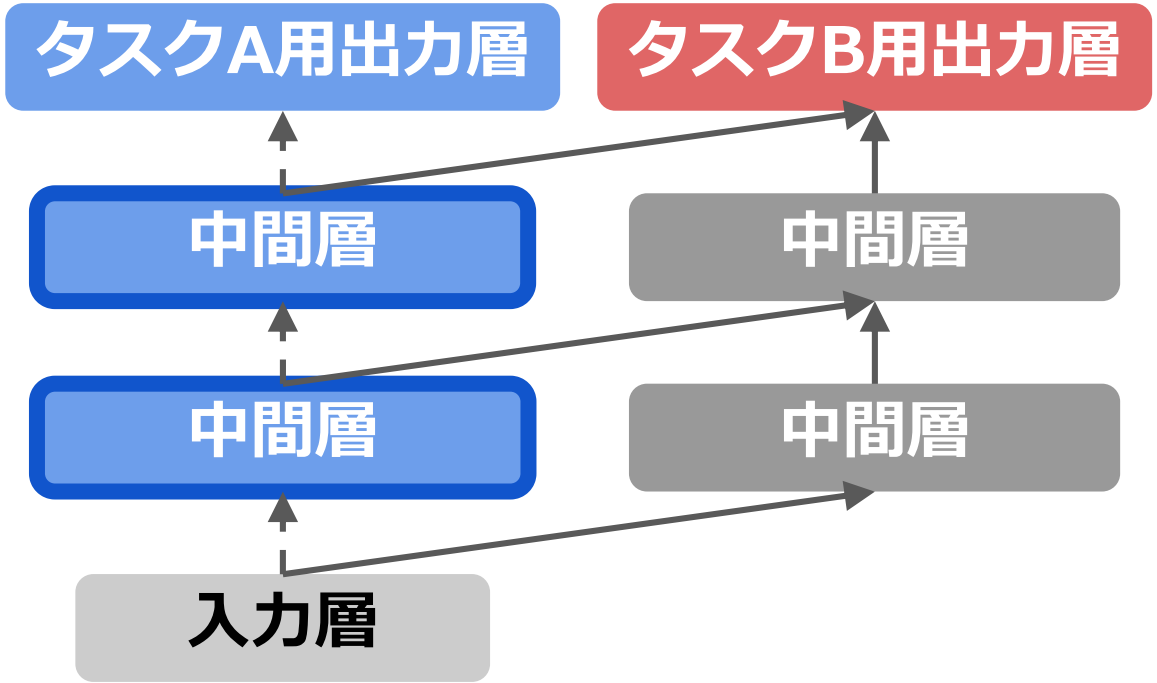


Progressive Neural Networks

4. 重み（パラメータ）を固定

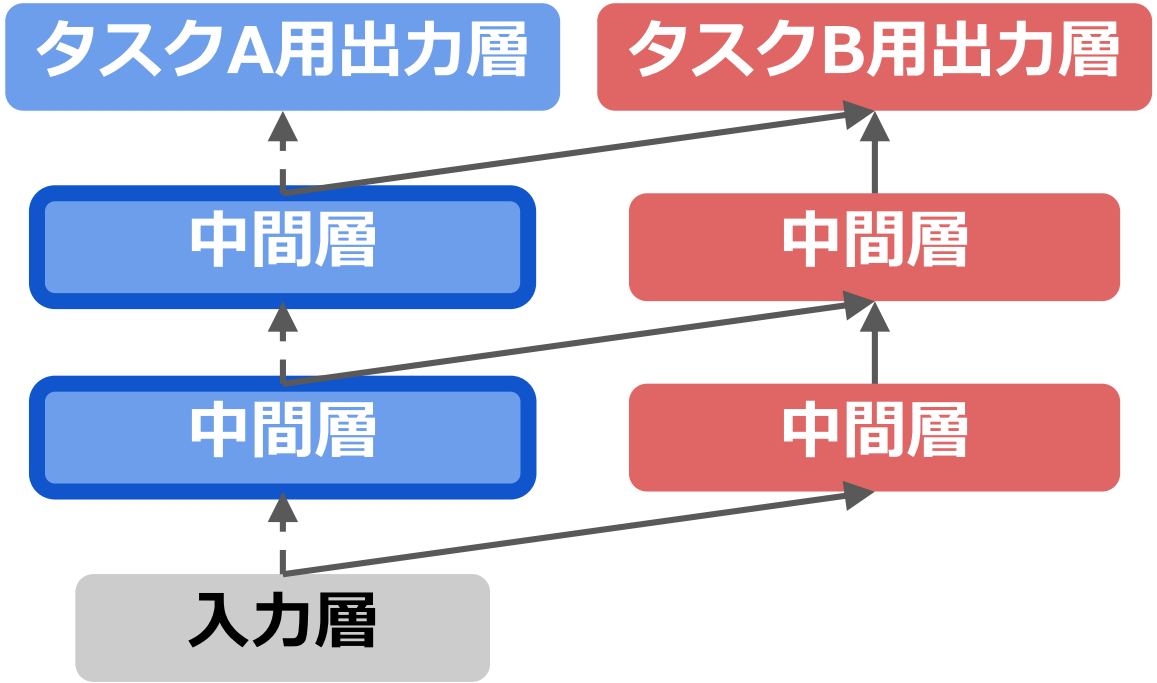
Progressive Neural Networks

3．新しい「カラム」を追加



Progressive Neural Networks

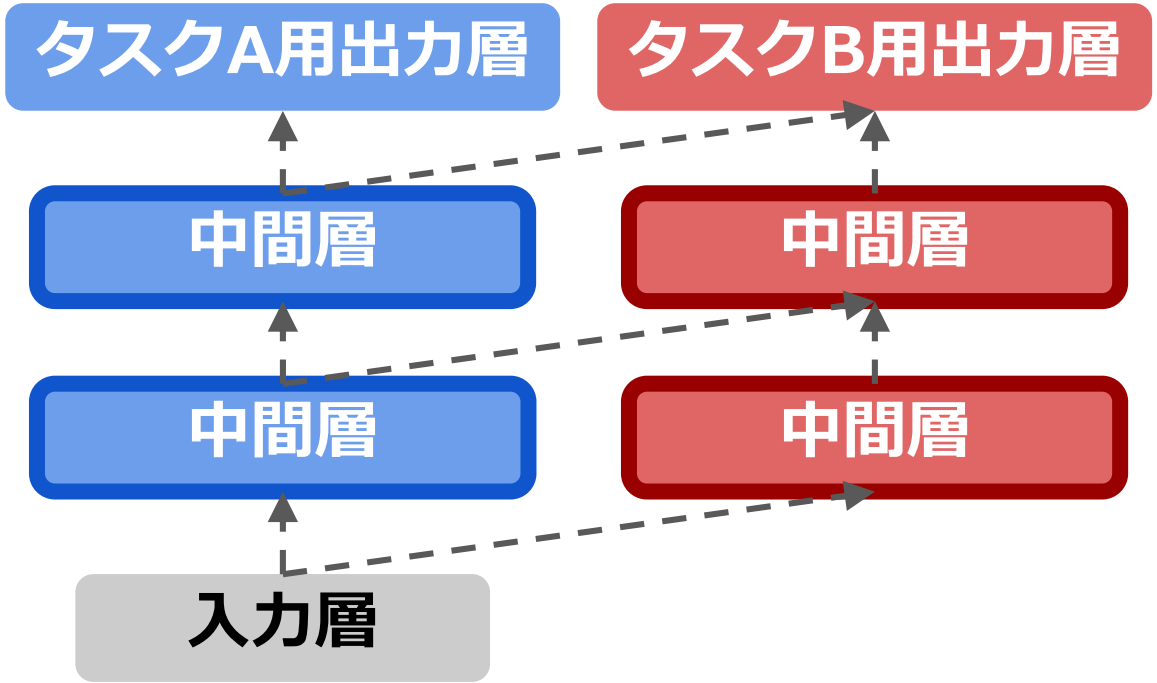
４．カラム２でタスクBを学習



カラム１の重みは固定されているから、タスクBの学習に影響されない

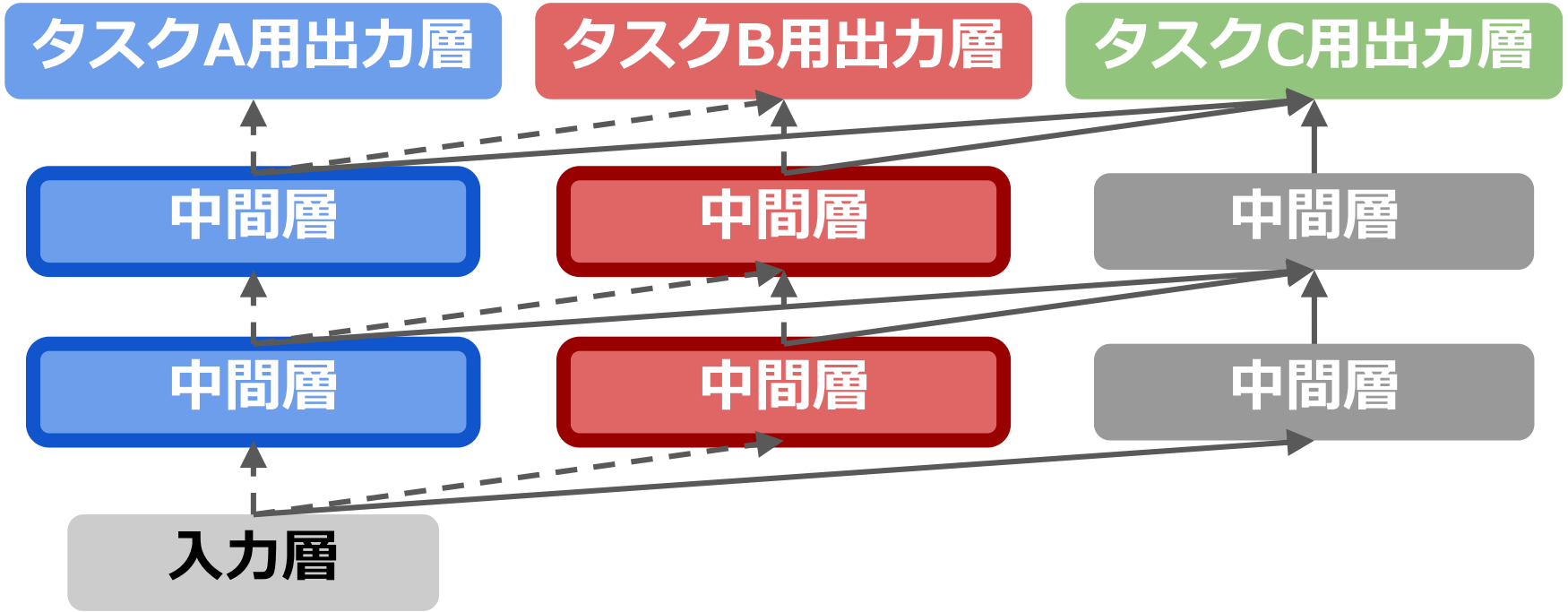
Progressive Neural Networks

5．カラム２のパラメータも固定



Progressive Neural Networks

6．タスクが増えるほどカラムを追加していけばいい



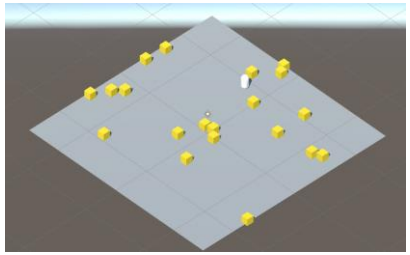
Progressive Neural Networks

- 過去に学習した知識が失われない
  （破壊的干渉を完璧に避ける）
- 過去に学習した知識を新しいタスクに使える
  （転移学習）

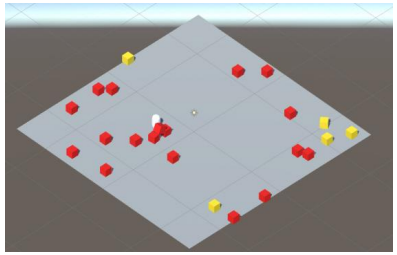
- タスクが増えるほどパラメータが増える
- カラムを追加するタイミングが決まっている

学習させるゲーム

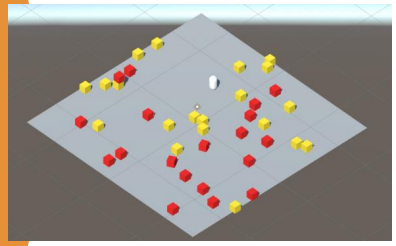
① **報酬**のみのゲーム



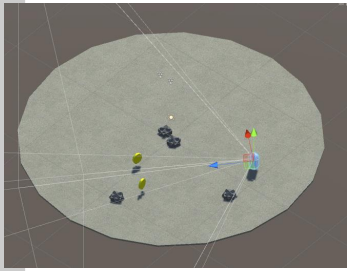
② ほとんど**罰**のゲーム



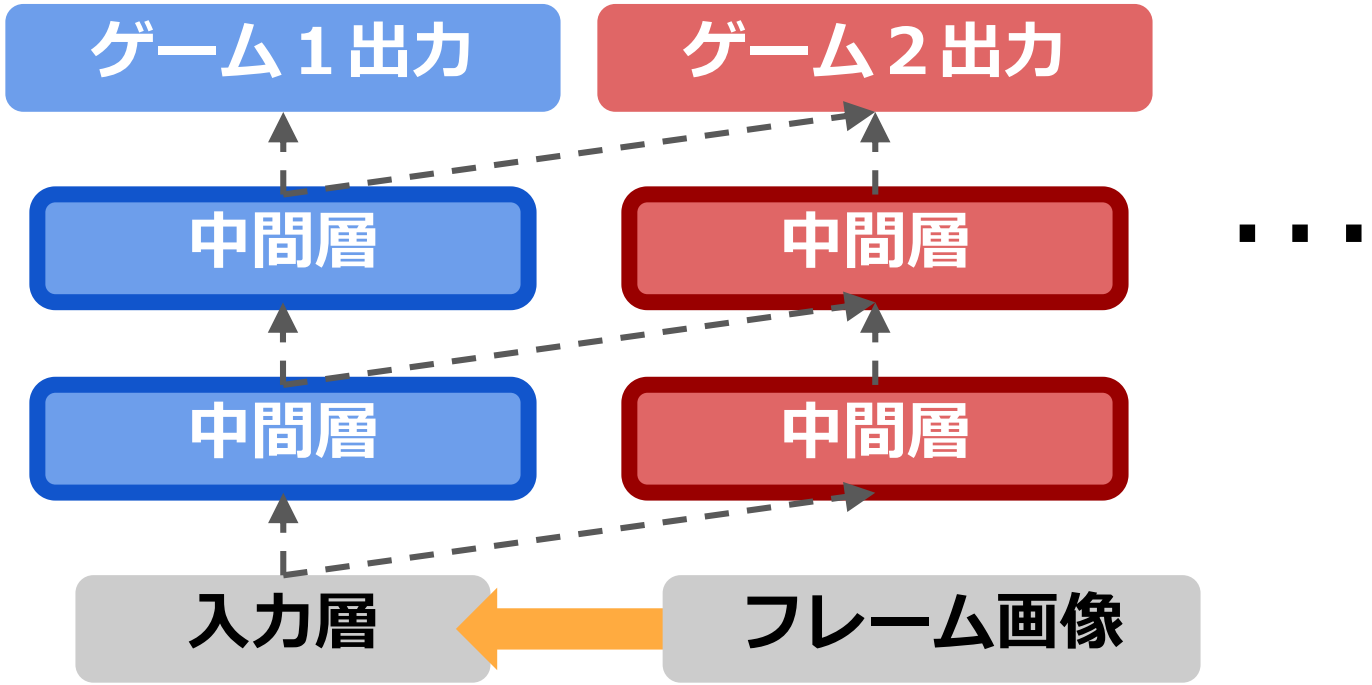
③ **報酬**と**罰**のゲーム



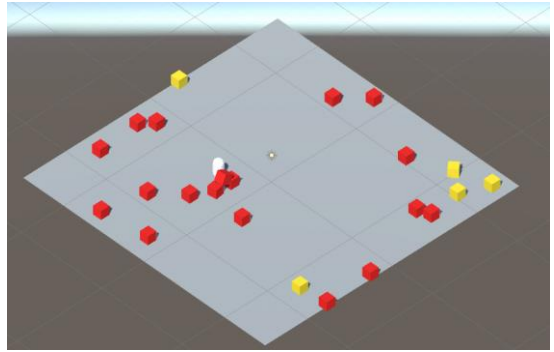
④ アイテムの外見が違うゲーム

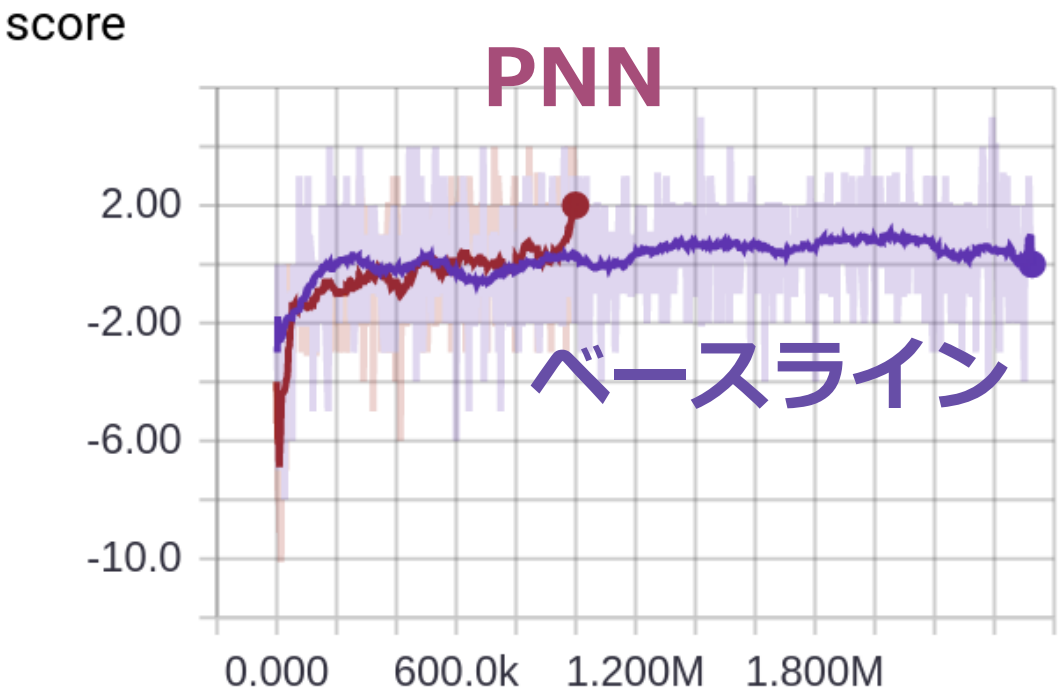


今回の場合

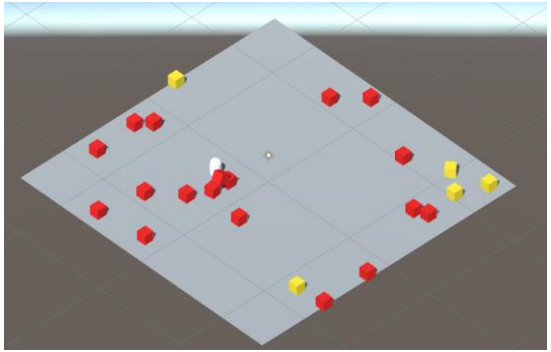


結果｜ゲーム②｜学習

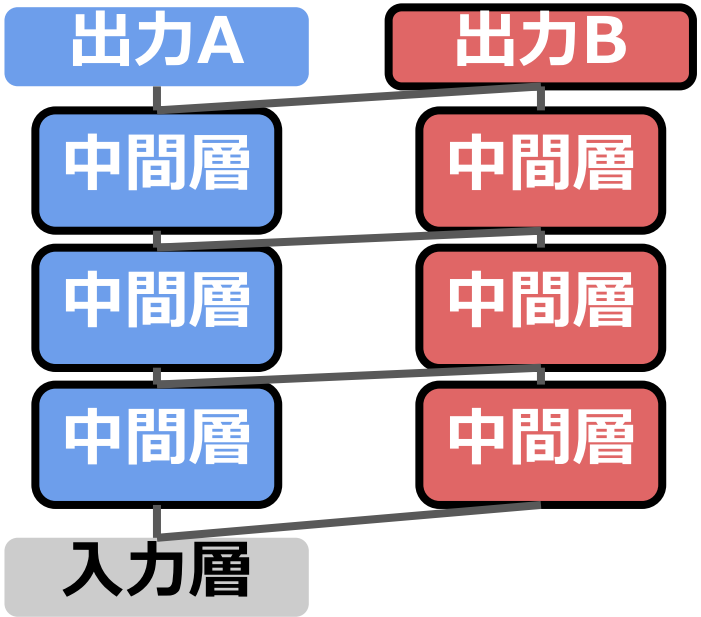




結果｜ゲーム②｜転移

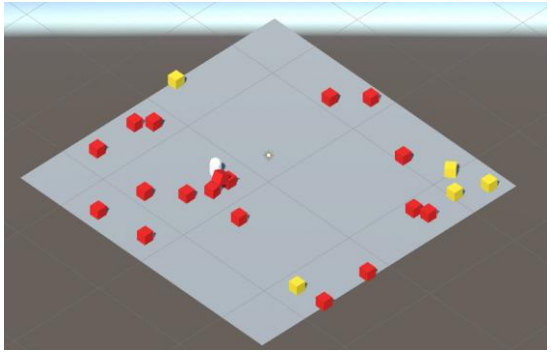


偏微分で、各層の活性が出力に与える
影響力を計算



0.27	**0.75**
0.46	0.54
0.73	0.27

結果｜ゲーム②｜転移



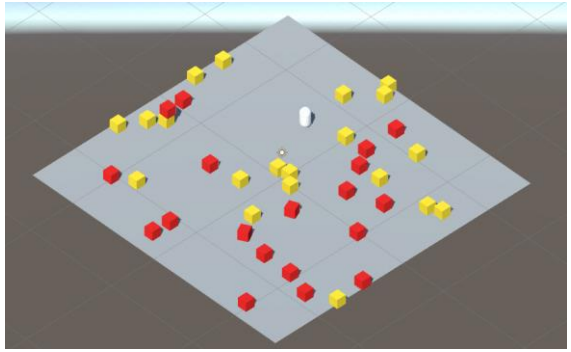
偏微分で、各層の活性が出力に与える
影響力を計算

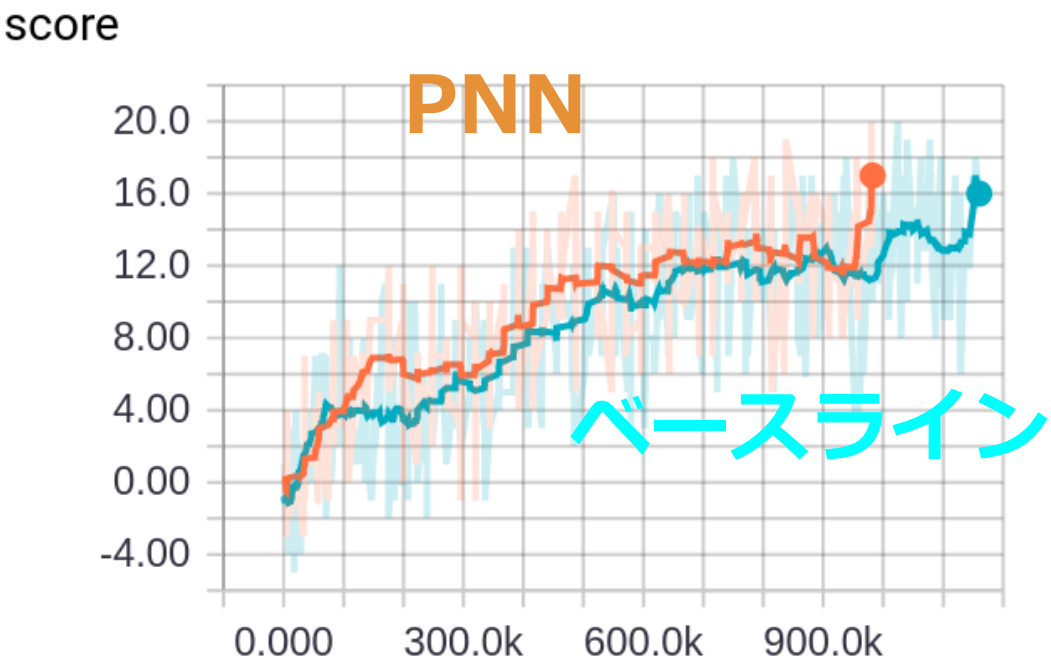
高次な処理はほとんど
新しく学習した

低レベルの特徴は
ゲーム①で
獲得したものを再利用

0.27	**0.75**
0.46	0.54
0.73	0.27

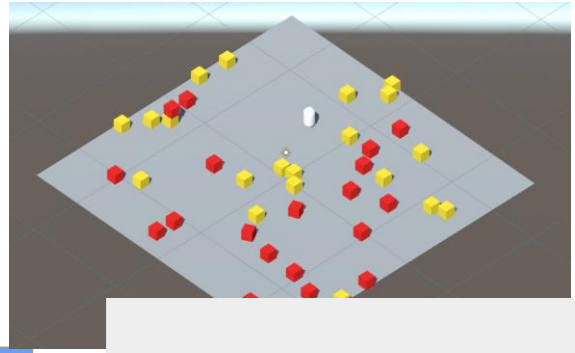
結果｜ゲーム③｜学習




score
PNN
ベースライン
20.0
16.0
12.0
8.00
4.00
0.00
-4.00
0.000
300.0k
600.0k
900.0k

結果｜ゲーム③｜転移

ゲーム②と似たパターン

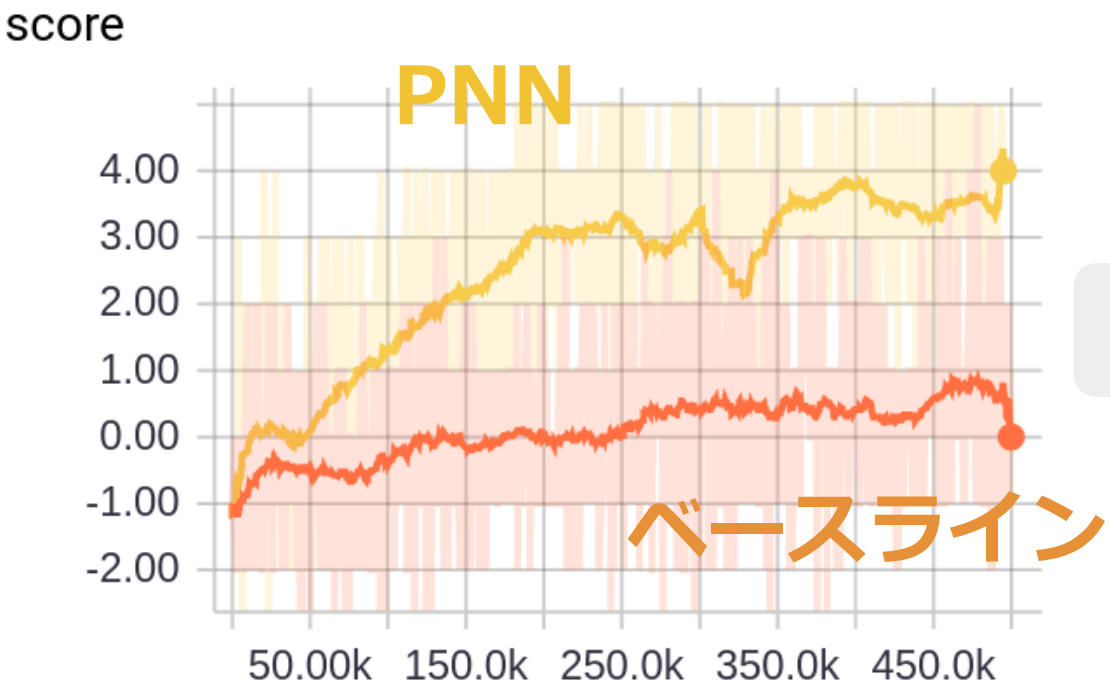


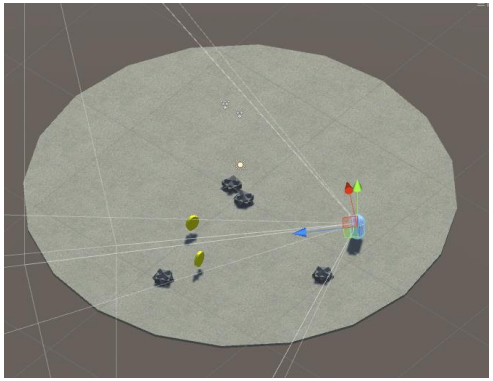
0.14	0.26	**0.61**
0.48	0.23	0.29
0.68	0.16	0.16

高次な特徴は
新しいのを
利用

低レベルの
特徴は
ゲーム①のを再
利用

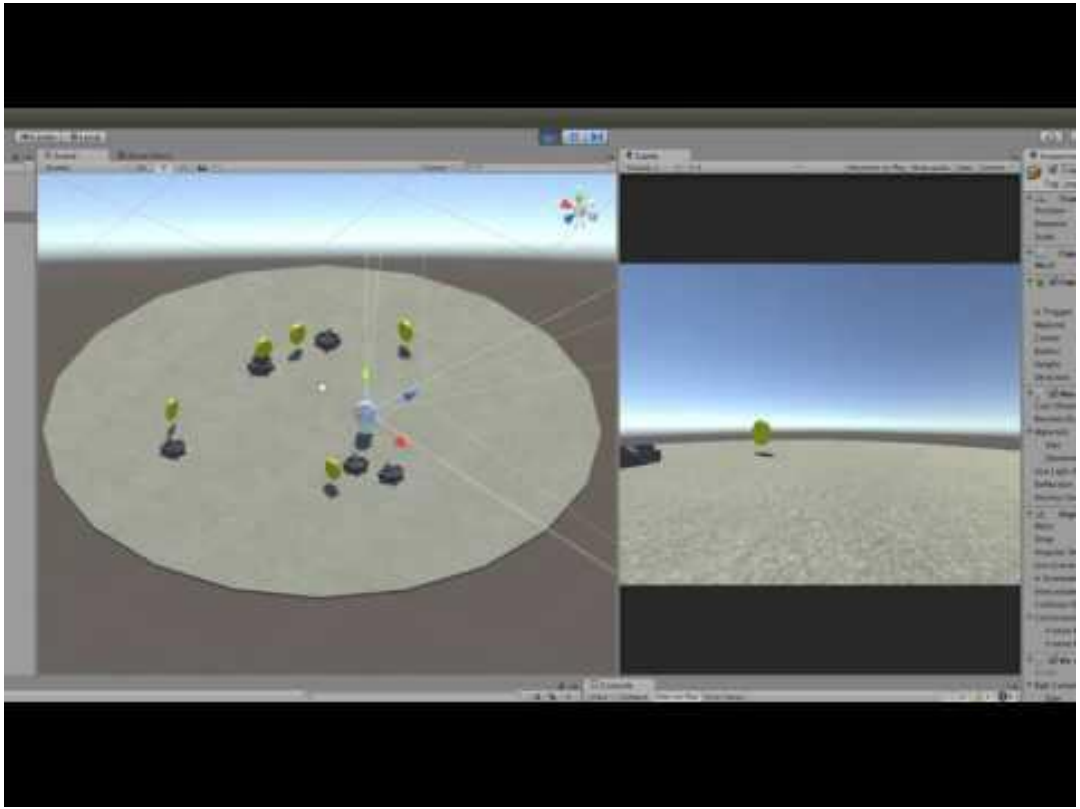
結果｜ゲーム④｜学習



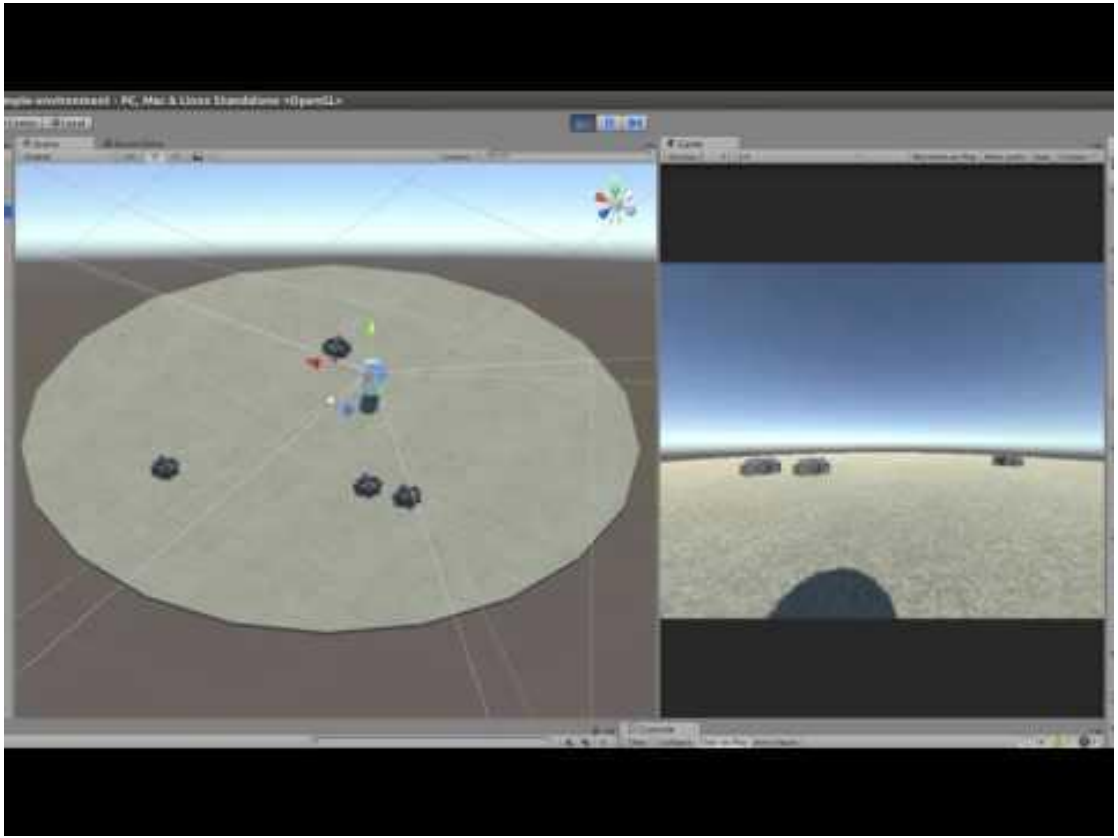


ここで大きな差が出た

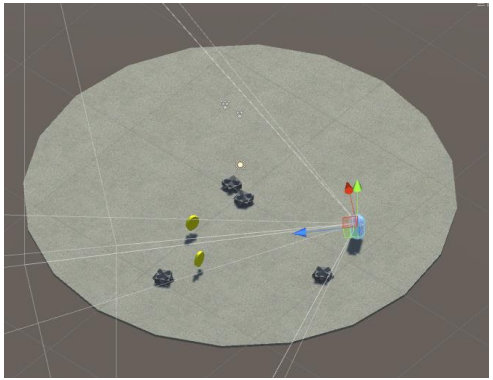
結果｜ゲーム④｜動画｜ベースライン



結果｜ゲーム④｜動画｜PNN



結果｜ゲーム④｜転移



③の「寄る」と「避ける」が転移？

ここの特徴は相変わらずよく利用される

0.16	0.12	**0.38**	**0.33**
0.45	0.16	0.22	0.17
0.80	0.08	0.08	0.04

なぜ④は大きな差を出したのか？

③→④の変化：

見た目

PNNの方は既に基本な形を認識することができたから、
もっと複雑なものの学習が速かった

報酬・罰アイテムの数（③20/20　→　④5/5）

報酬信号がよりまばらになった結果、ゼロから学習する
モデルにはそれほど難しくなったのかもしれない

まとめ

汎用人工知能へ向けて、
　複数のゲームを攻略できるモデルを実装した

学習されたモデルの「中身」を観察した

一つだけのゲームを学習したモデルよりも、
　過去の経験から獲得した知識を再利用するモデルの方
　が学習が速いと示した

今後の展望

自動にカラム・ニューロンを追加・削除

再帰的な重みの追加

外部記憶の追加

教師なし学習を追加（Predictive Codingなど）

今後の展望

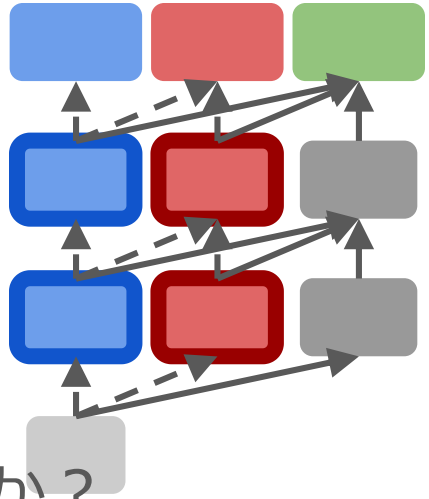
自動にカラム・ニューロンを追加・削除

再帰的な重みの追加

外部記憶の追加

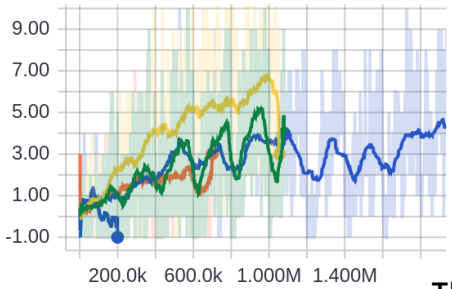
教師なし学習を追加（Predictive Codingなど）

今後の展望｜自動追加・削除



PNNの問題の一つ：
タスクが増えるほどパラメータの数が急増する

あまり使われていないニューロンを削除できるか？
→　性能を保ち、効率化する



現在実験中…

0.16	0.12	**0.38**	**0.33**
0.45	0.16	0.22	0.17
0.80	0.08	0.08	0.04

もっと後の展望

色々学習させてAGIに近づく？？

関連リンク

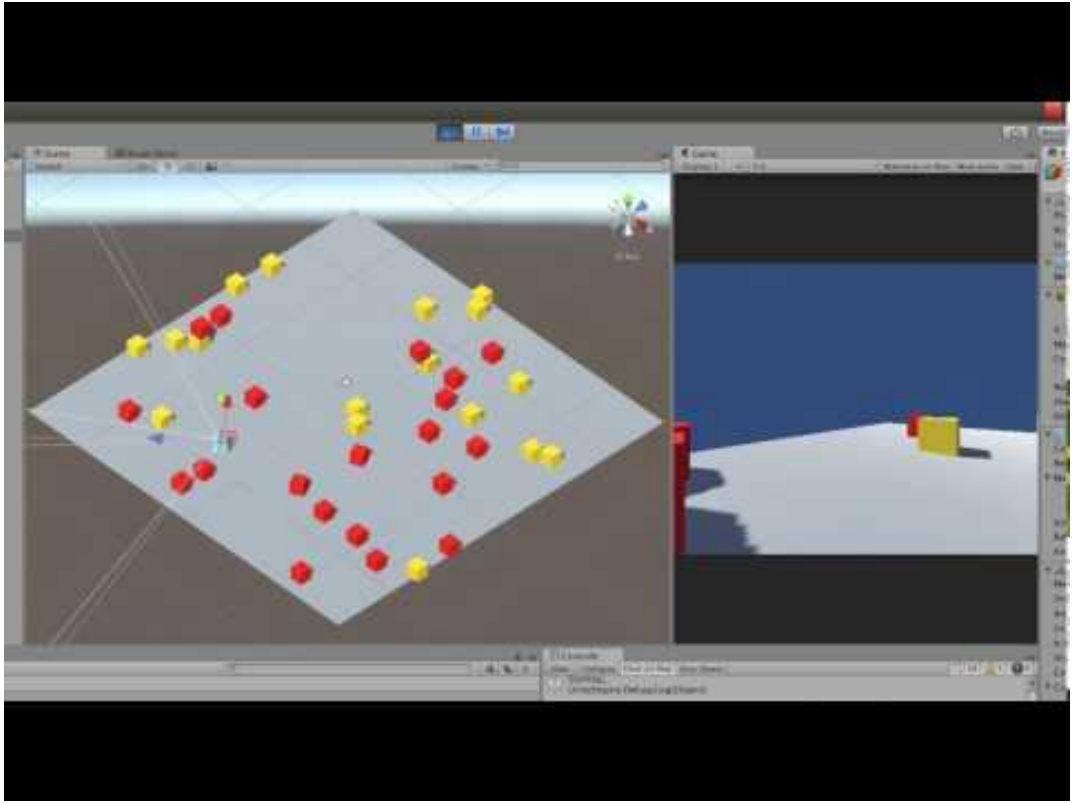
[1] A3C

https://arxiv.org/abs/1602.01783

[2] Progressive Neural Networks

https://arxiv.org/abs/1606.04671

結果｜ゲーム③｜動画｜PNN



より精密な観察を試した

中身をよりよく知るため、
③のモデルをいじった

環境の変化がどう内部・
出力が変わるかを観察した
が、目ではあまりわから
なかった...