WBA
Future Leaders
Whole Brain Architecture
Future Leaders

Brain-inspired A3C

Implementing cortical-hippocampal system

Takuma Seno, Shoya Matsumori, Toshiki Kikuchi, Yuki Takimoto

課題結果

1-8b以外は個別に単独で学習

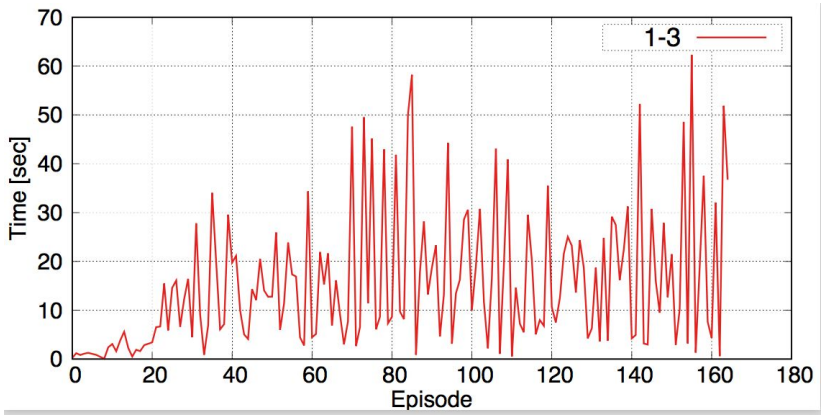
課題番号	成功エピソード数	失敗エピソード数	合計エピソード数 (成功+失敗)	学習状態
1-1	69 (100%)	0 (0%)	69	成功
1-2	69 (100%)	0 (0%)	69	成功
1-3 緑停止	49 (34%)	97 (66%)	146	学習時間不足
1-4 赤停止	53 (32%)	114 (68%)	167	学習時間不足
1-5 青停止	102 (53%)	90 (47%)	192	成功
1-6	65 (43%)	86 (57%)	151	学習時間不足
1-8a 色なし	206(52%)	189(48%)	395	成功
1-8b 色なし	81(49%)	86(51%)	167	学習時間不足
2-1				ロスのみ
2-2				学習中…
3-1	40(29%)	99(71%)	139	学習時間不足
3-2	36(31%)	82(69%)	118	学習時間不足

ログ

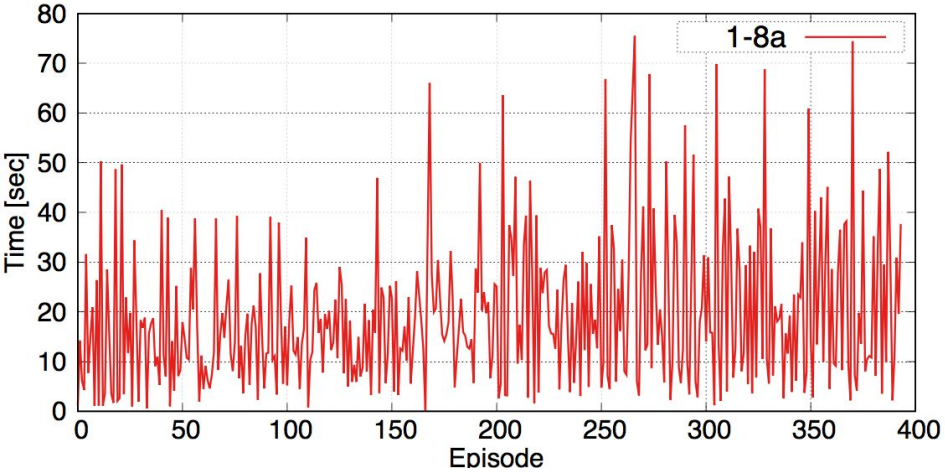
すみません、ログの仕様を理解していなくて消失しました...

CPU時間じゃなくてUnity時間で見て欲しい...

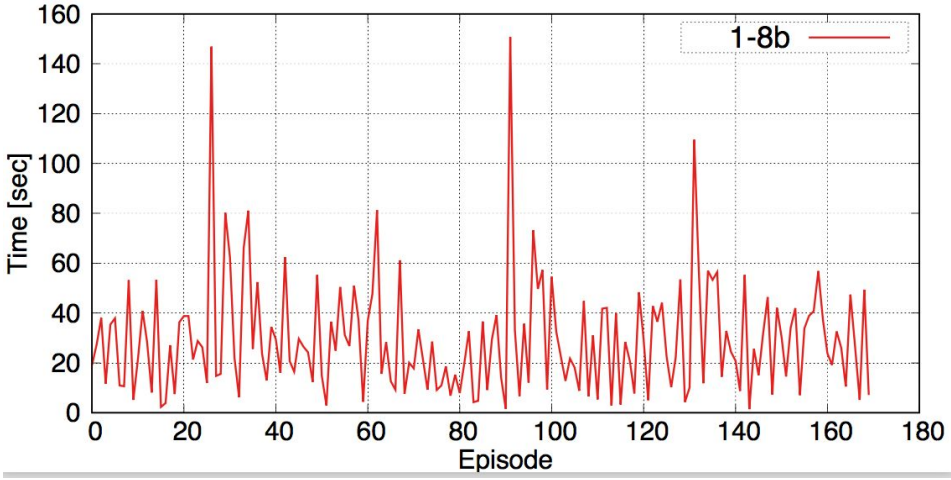
1-3



1-8a



1-8b

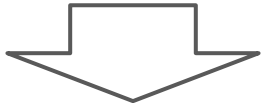


神経科学的妥当性

		✔			✔
海馬内活動	リプレイ	✔	脳領域構造	CA1	✔
	プリプレイ			CA2	
	場所細胞	✔		CA3	✔
	グリッド細胞	✔		歯状回	
	頭部方向細胞	✔		嗅内皮質	✔
	シータ位相歳差			海馬支脚	
	スパース表現			Perirhinal Cortex	
	パターン補完			Postrhinal Cortex	
	細胞新生		その他	コネクトームの導入	✔
行動機能	自律的フェーズ変化			BiCAMON可視化	
	エピソード記憶	✔		その他	✔
	場所の再認	✔			
	記憶転送				
	ナビゲーション/空間認知	✔			
	Path integration				

海馬を世界最先端の工学モデルに組み込む!

Brain-inspired A3C



世界最強の脳型アーキテクチャ

提案

Brain-inspired A3C

海馬を世界最先端の工学モデルに組み込む!

神経科学的:

ヒューリスティクスを排除し
空間認知 + エピソード記憶の機能をニューラルネットで学習

工学的:

最新のモデルを力の限り採用!
新規性: A3C + NEC + 空間認知情報

AGENDA

Brain-Inspired A3C

1. イントロ
2. なぜA3Cなのか
3. アーキテクチャ設計
4. 実装-実験
5. 考察
6. まとめ

(復習) DQNで良いのか (1/3)

タスクが公開された時、我々は考えた

Experience Replayは諸刃の剣

局所解に嵌まることを避けるため、
蓄積したエピソード(s, a, r, s')からランダムサンプリングして学習

ランダムサンプリングする過程で時系列は失われる！

⇩

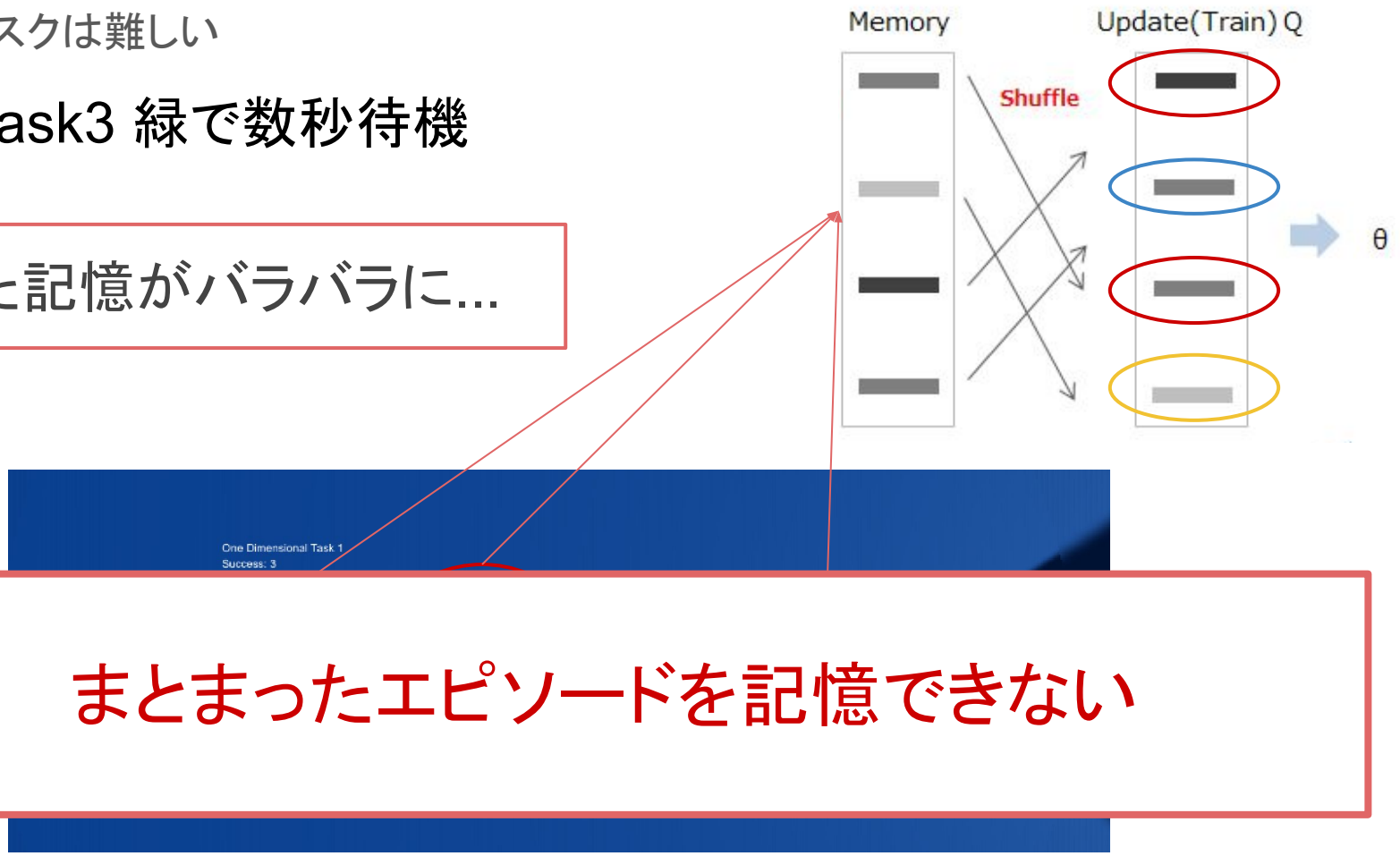
時系列が効いてくるタスクは難しい…

(復習) DQNで良いのか (2/3)

こんなタスクは難しい

例: Task3 緑で数秒待機

待った記憶がバラバラに...



まとまったエピソードを記憶できない

(復習) DQNで良いのか (3/3)

Experience Replayの神経科学的妥当性

- 睡眠中にCA1領域でメモリーリプレイを行なって記憶の定着を行う
  - 時系列的に順方向でのメモリーリプレイ

→ DQNのランダムなサンプリングは神経科学的に妥当ではない

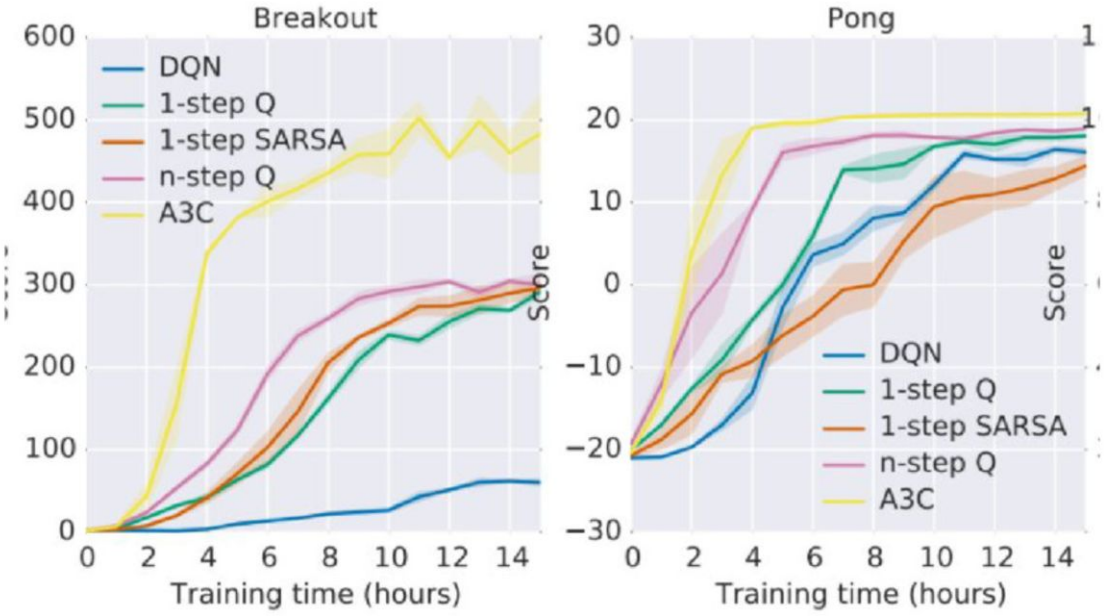
では、時系列性をもたせるか ...?

ER: ランダム性が重要
→ DQN+ERでの実装には限界がある

RLの進化

- A3Cの躍進(2016年)
  「高速」「シンプル」「ロバスト」で
  既存モデルを圧倒 [Minh+ 16]

DQN（GPU使用）よりもA3C（CPU使用）が
短い実行時間で学習できた



A3Cの利点

A

- Asynchronous
  (複数のエージェントで非同期MasterNetを更新)
  → マルチで高速かつロバスト
  → ERの代わりに**RNN使用可(時系列につよい)**

3

- Advantage

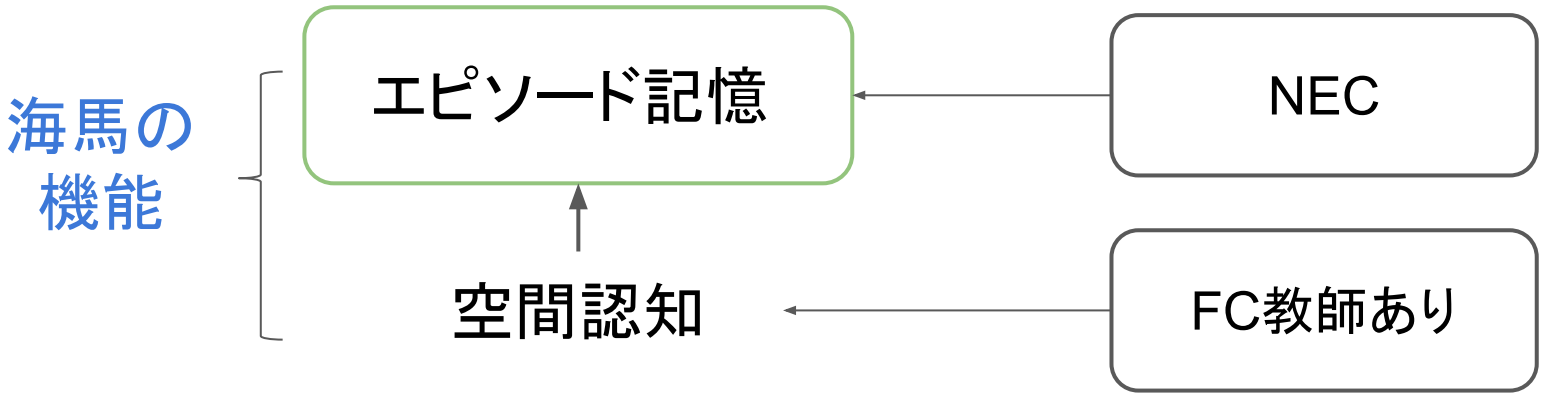
Brain-InspiredなA3Cモデルを作る

- Actor-Critic (πとVを別々に学習)
  π(s): 方策関数
  V(s): 価値関数
  線条体神経科学的妥当性 [Barto 95]
  → 行動の連続値を扱うことができる
  (今回は使用しない)

アーキテクチャ設計

提案アーキテクチャ

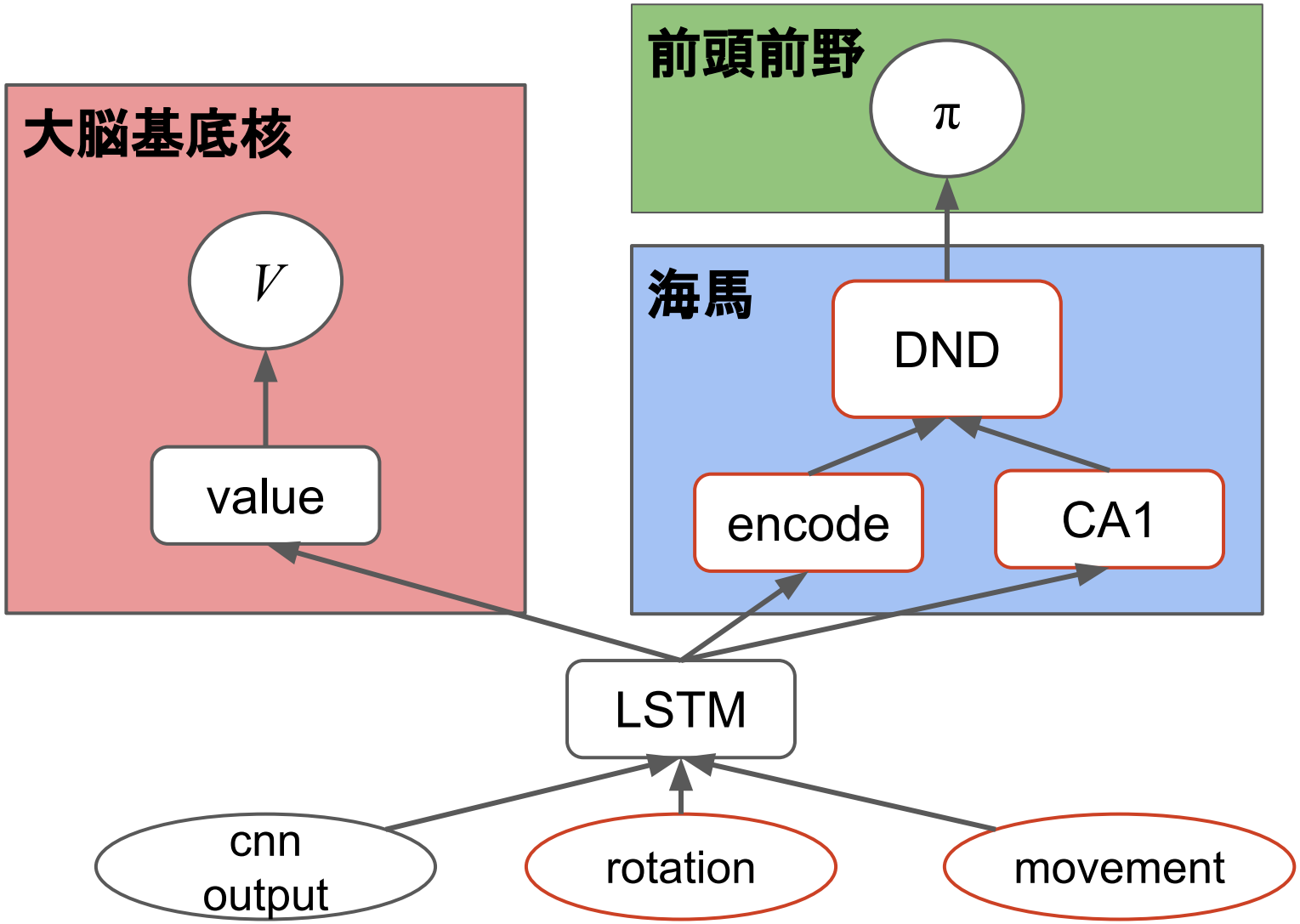
Brain-inspired A3C



エピソード記憶を「価値を伴う記憶」[Wimmer+ 12] として実装
エピソード記憶の鍵として空間情報を付加

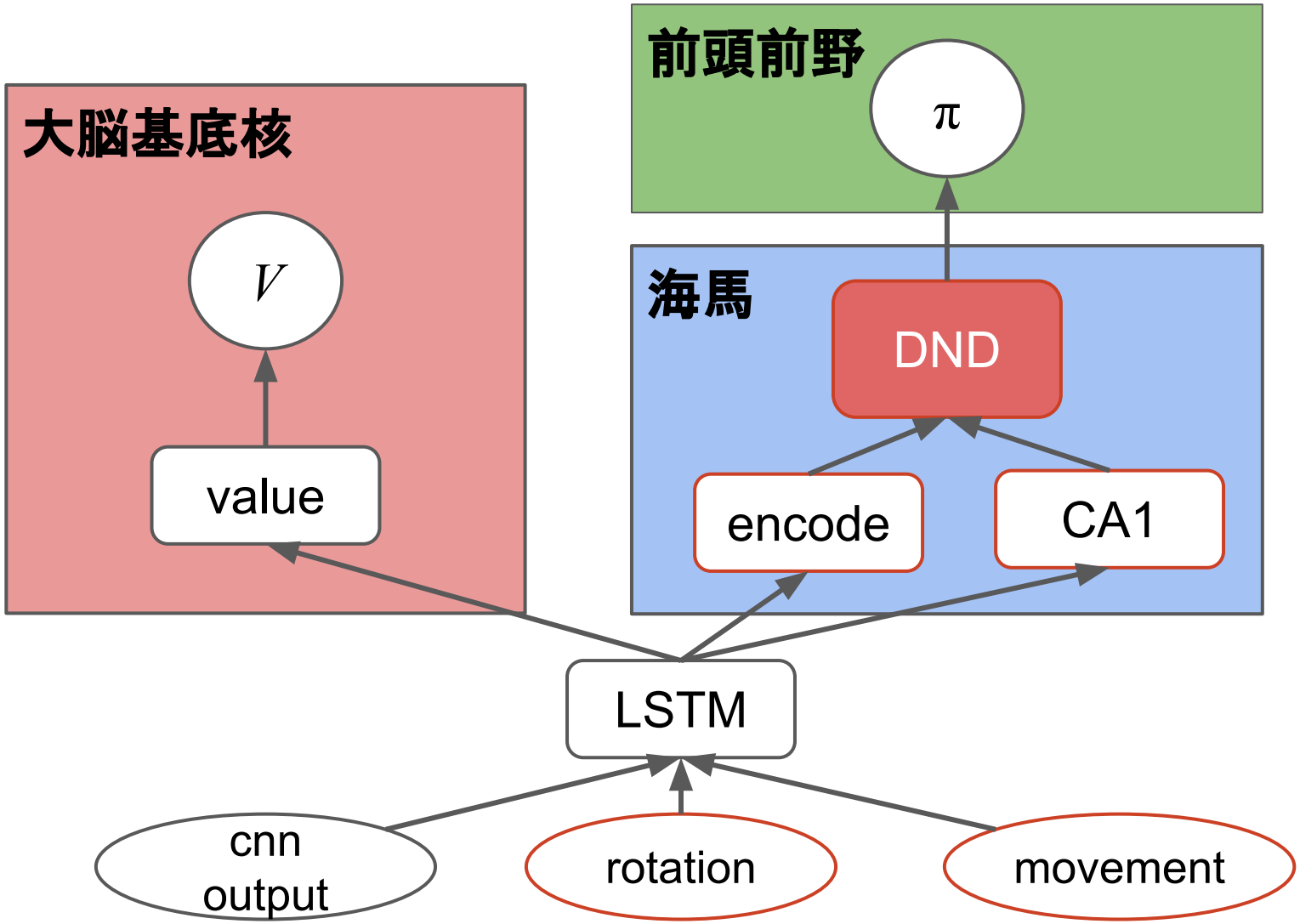
新規性

アーキテクチャ


前頭前野
π
大脳基底核
V
value
海馬
DND
encode
CA1
LSTM
cnn
output
rotation
movement

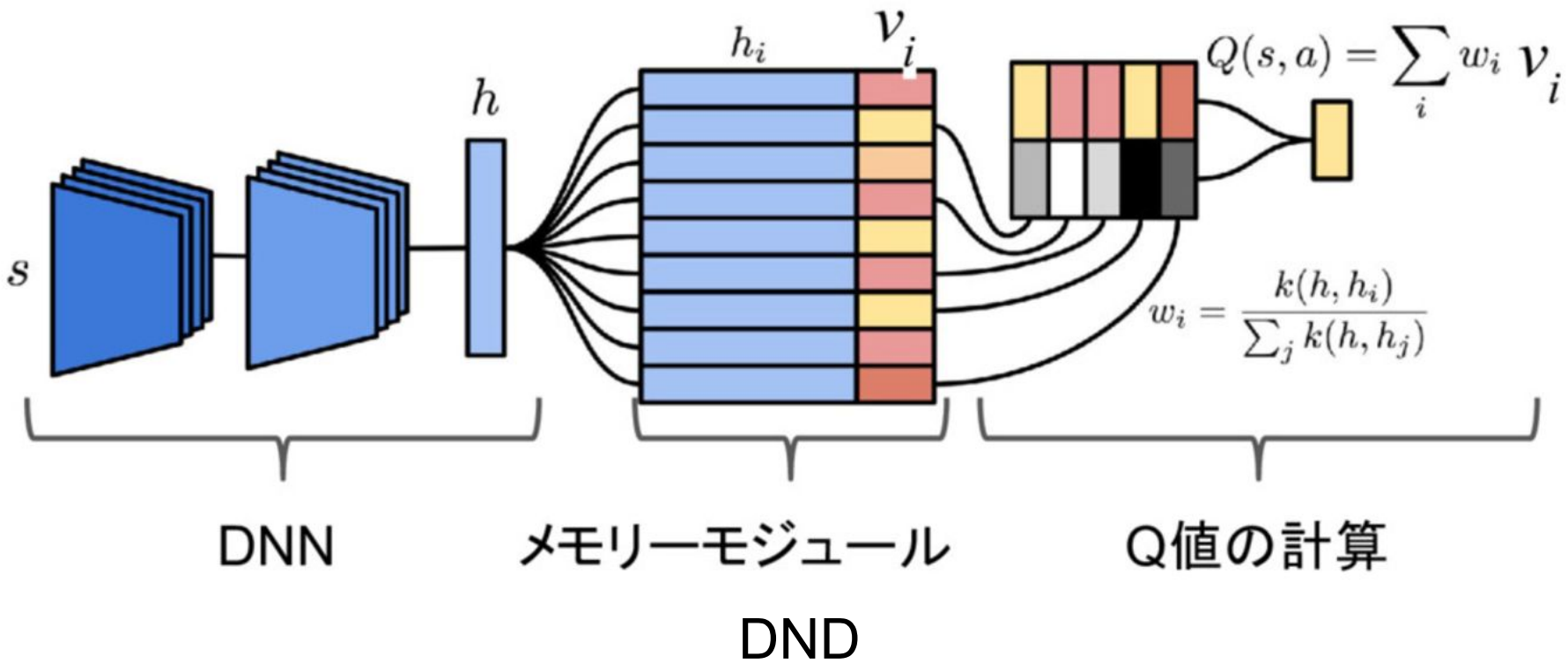
新規性

アーキテクチャ


前頭前野
π
大脳基底核
V
value
海馬
DND
encode
CA1
LSTM
cnn
output
rotation
movement

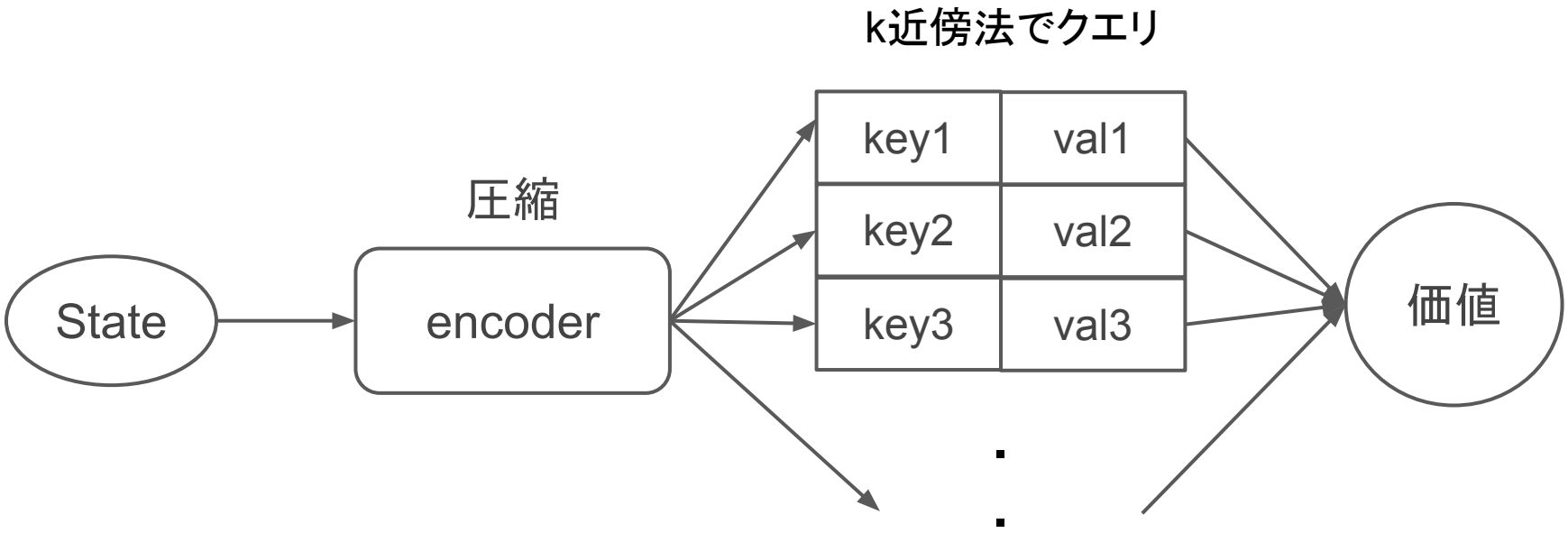
Neural Episodic Control (NEC)

DNNに**メモリーモジュール(DND)**を組み込むことで少ない学習で行動選択を行えるようにした深層強化学習手法



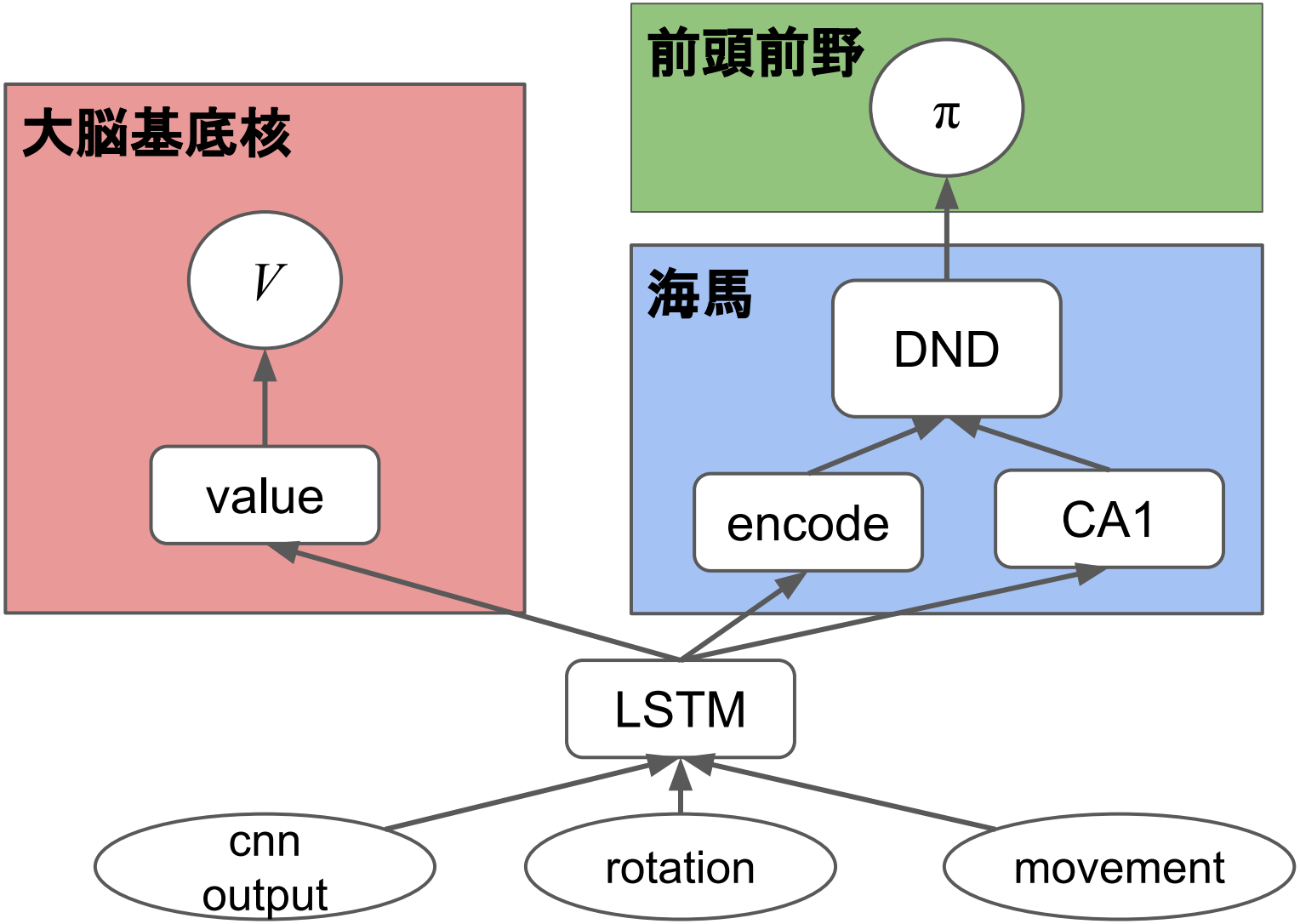
Differentiable Neural Dictionary (DND)

Key-Valueのテーブルに実際に経験した状態に対する価値を
エピソード記憶として保存する

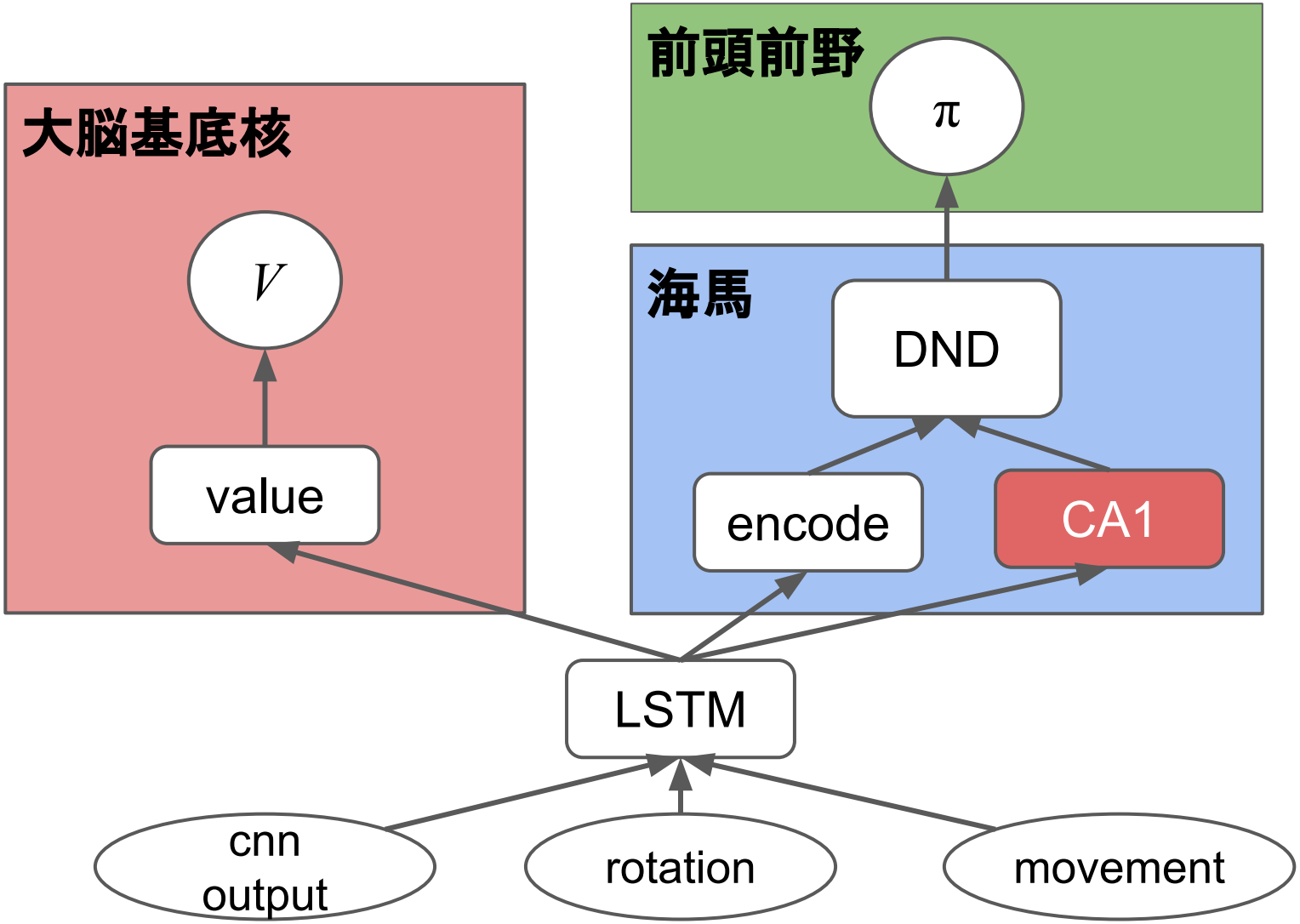


DNNだけよりも少ない学習で推定可能！！

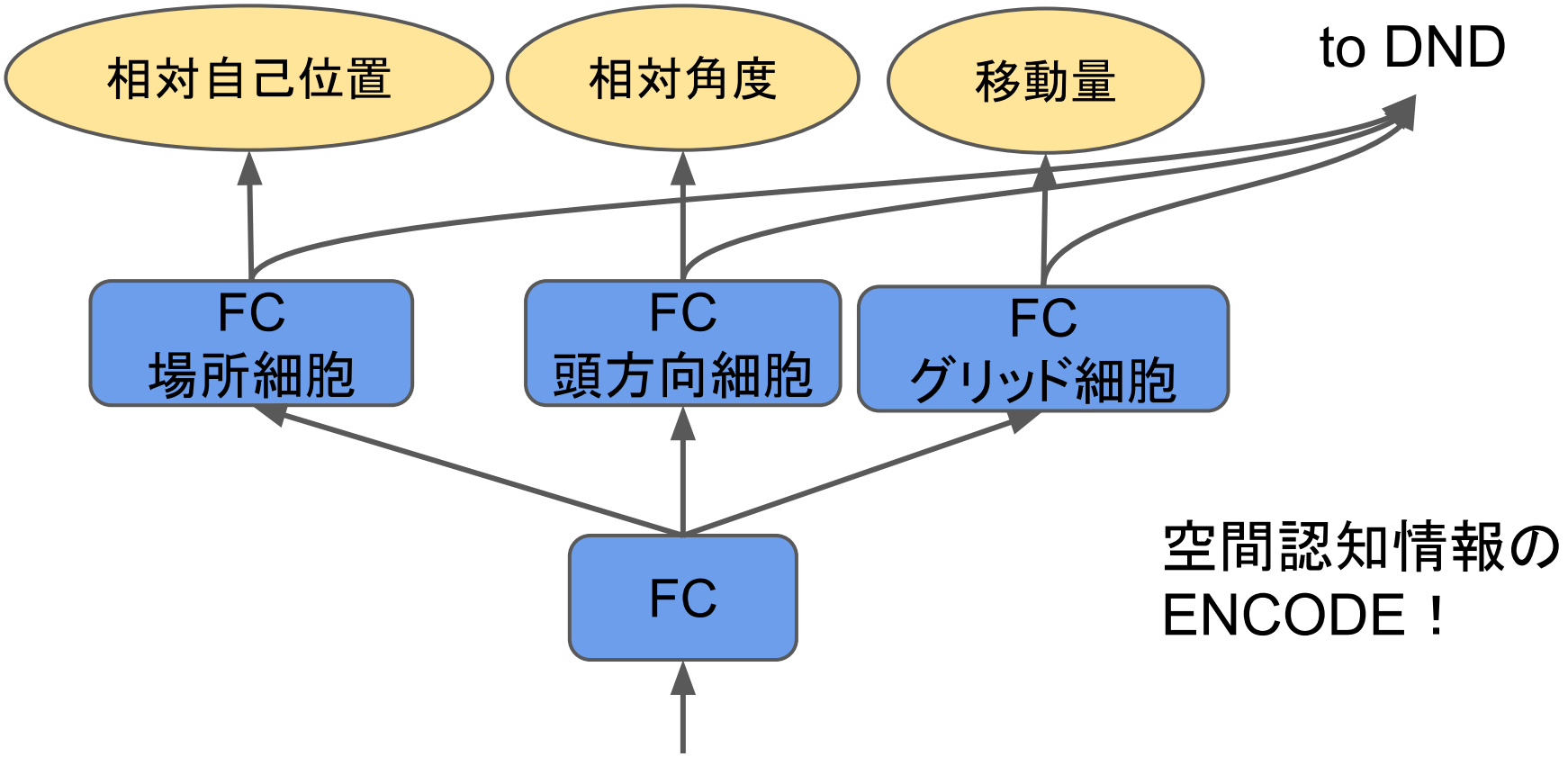
アーキテクチャ


前頭前野
π
大脳基底核
V
value
海馬
DND
encode
CA1
LSTM
cnn
output
rotation
movement

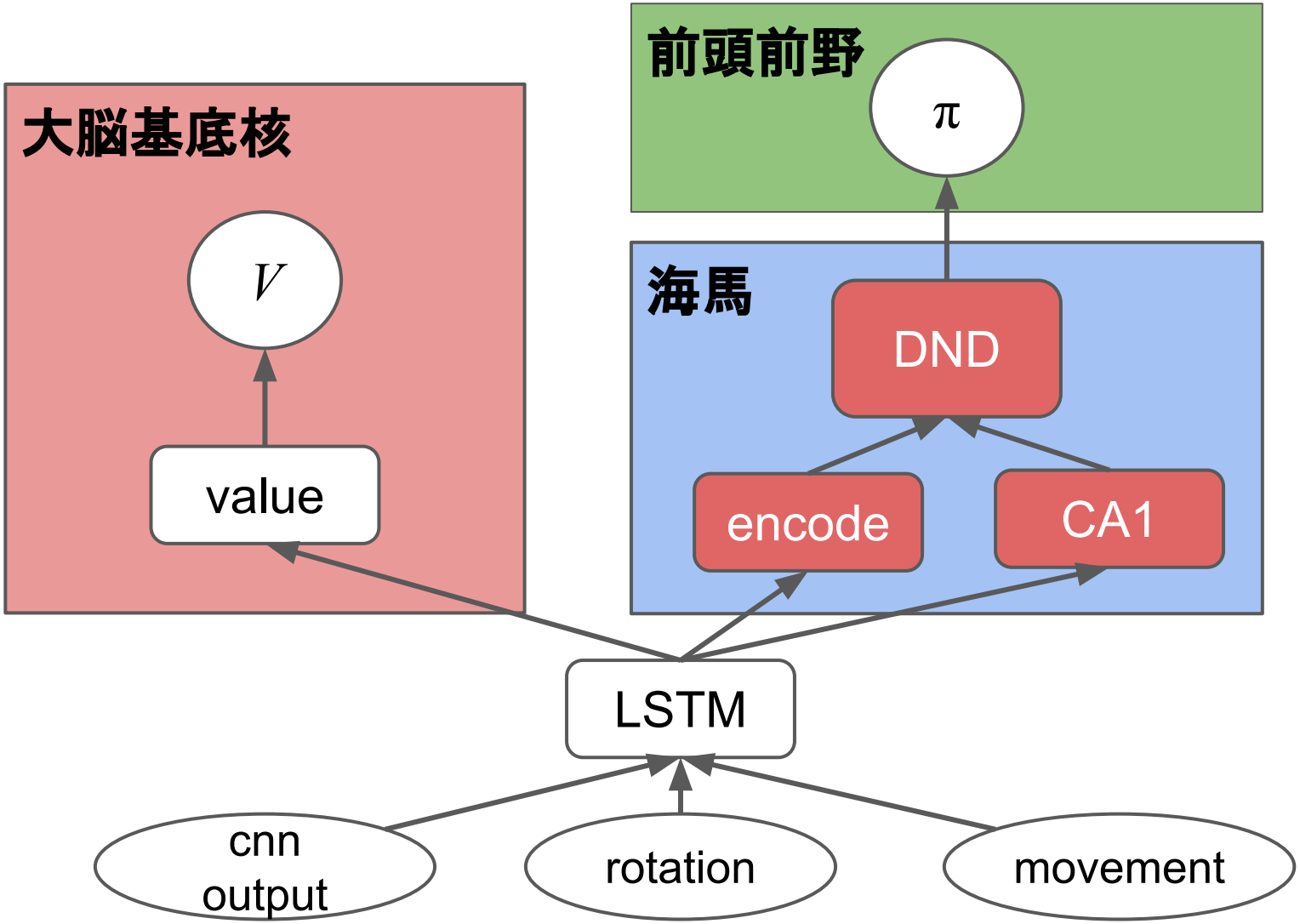
アーキテクチャ


前頭前野
π
大脳基底核
V
value
海馬
DND
encode
CA1
LSTM
cnn
output
rotation
movement

CA1の詳細


相対自己位置
相対角度
移動量
to DND
FC
場所細胞
FC
頭方向細胞
FC
グリッド細胞
FC
空間認知情報の
ENCODE！

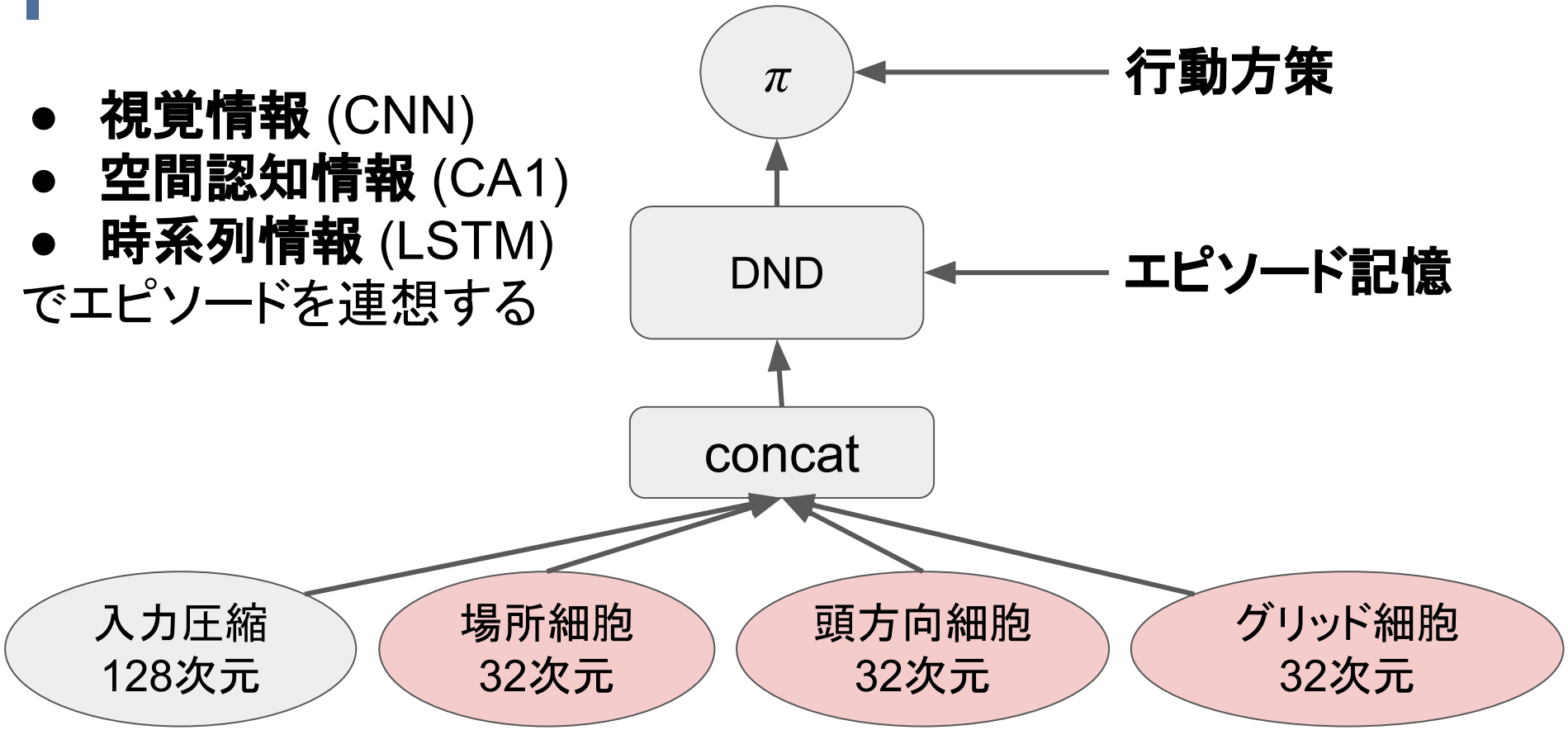
アーキテクチャ


前頭前野
π
大脳基底核
V
value
海馬
DND
encode
CA1
LSTM
cnn
output
rotation
movement

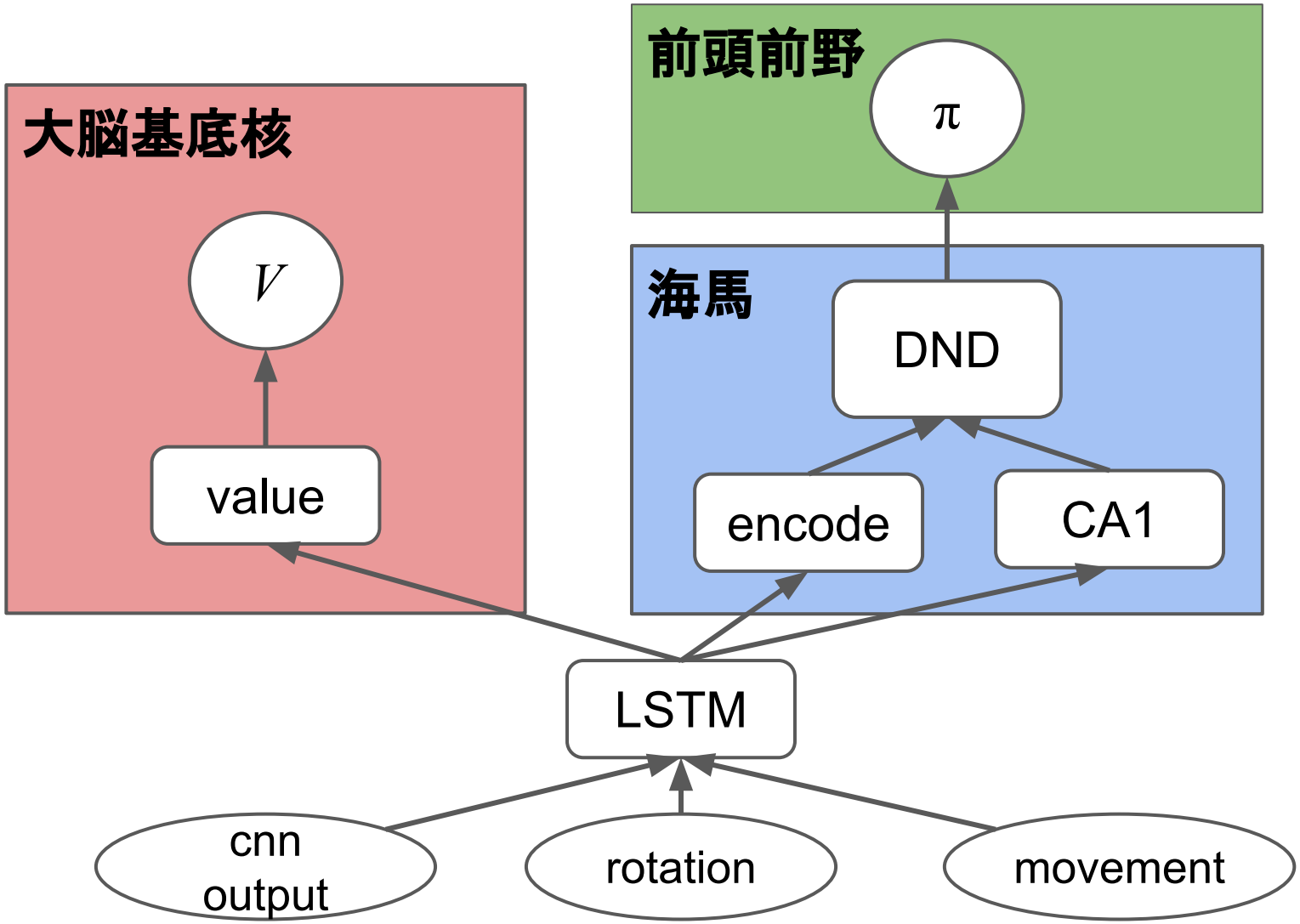
DNDのキー (エピソード記憶の連想)

- **視覚情報** (CNN)
- **空間認知情報** (CA1)
- **時系列情報** (LSTM)

でエピソードを連想する



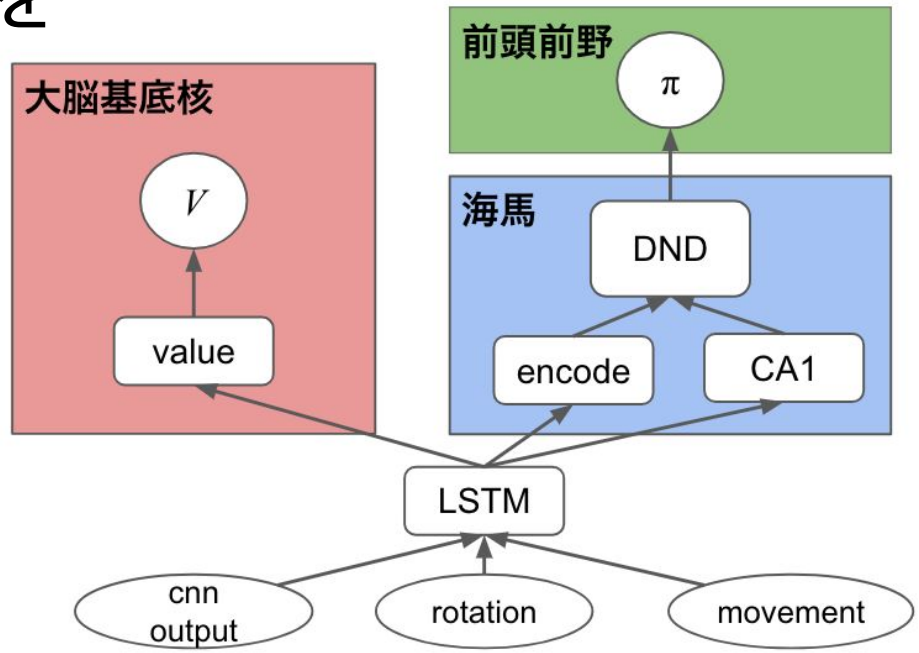
認知情報に基づいて連想できる！


前頭前野
π
大脳基底核
V
value
海馬
DND
encode
CA1
LSTM
cnn
output
rotation
movement

アーキテクチャまとめ

- A3Cによるマルチエージェント学習
- 空間認知 + 視覚 + 時系列情報でエピソードを連想

海馬のエピソード記憶の連想を
工学的、神経科学的に再現



3D空間でのタスクの
性能向上が期待！

実装 - 実験

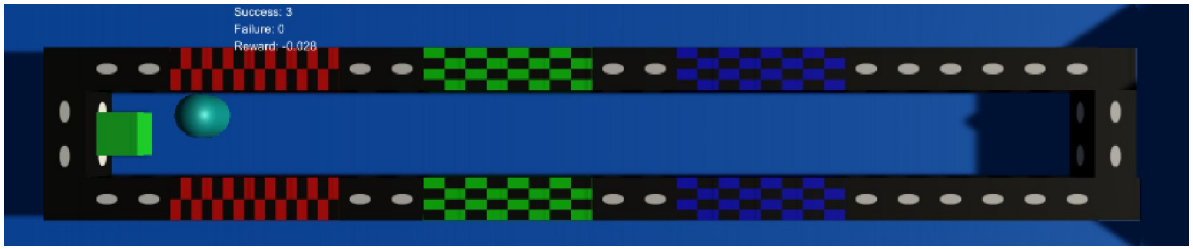
実験

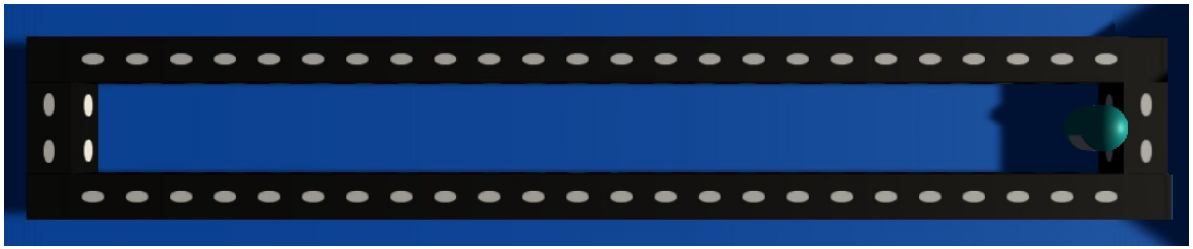
- 提案アーキテクチャを使用
  - エージェント数: 4
  - CA1教師データ: 相対自己位置、相対角度、移動量
    - Depth, Rotation, Movementからヒューリスティックに作製
- 可視化
  - Tensorboadによるロスをモニタリング
  - 価値推定V(s)のモニタリング
  - CA1各細胞の発火状態を保存 (可視化できず)

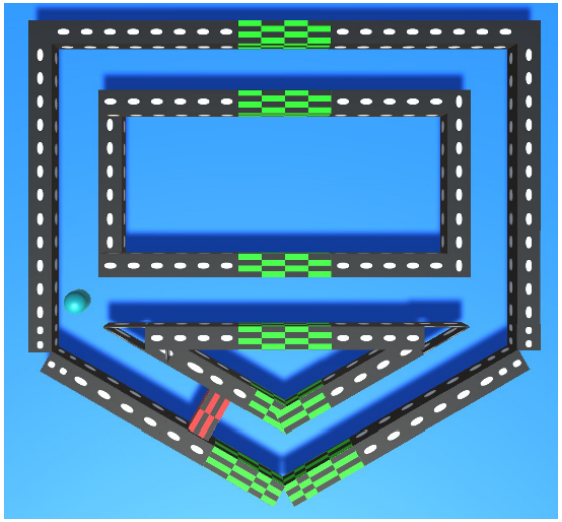
性能評価 (スコア)

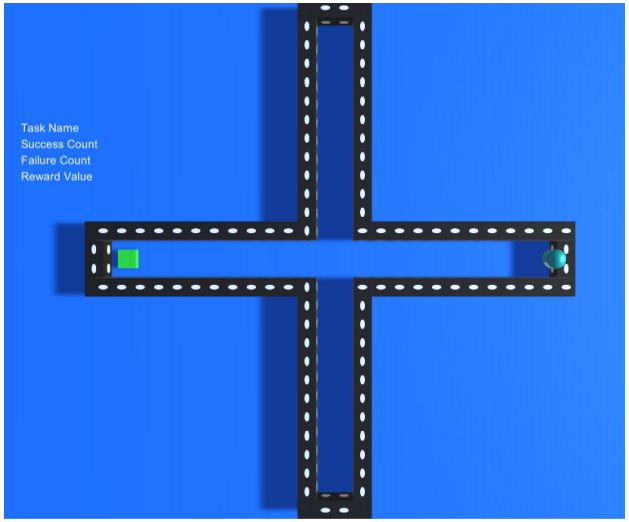
既定課題:

1. 1次元迷路課題
    - 色あり 1~7
    - 色なし 8a, 8b
2. Arrow迷路課題
3. 十字迷路課題









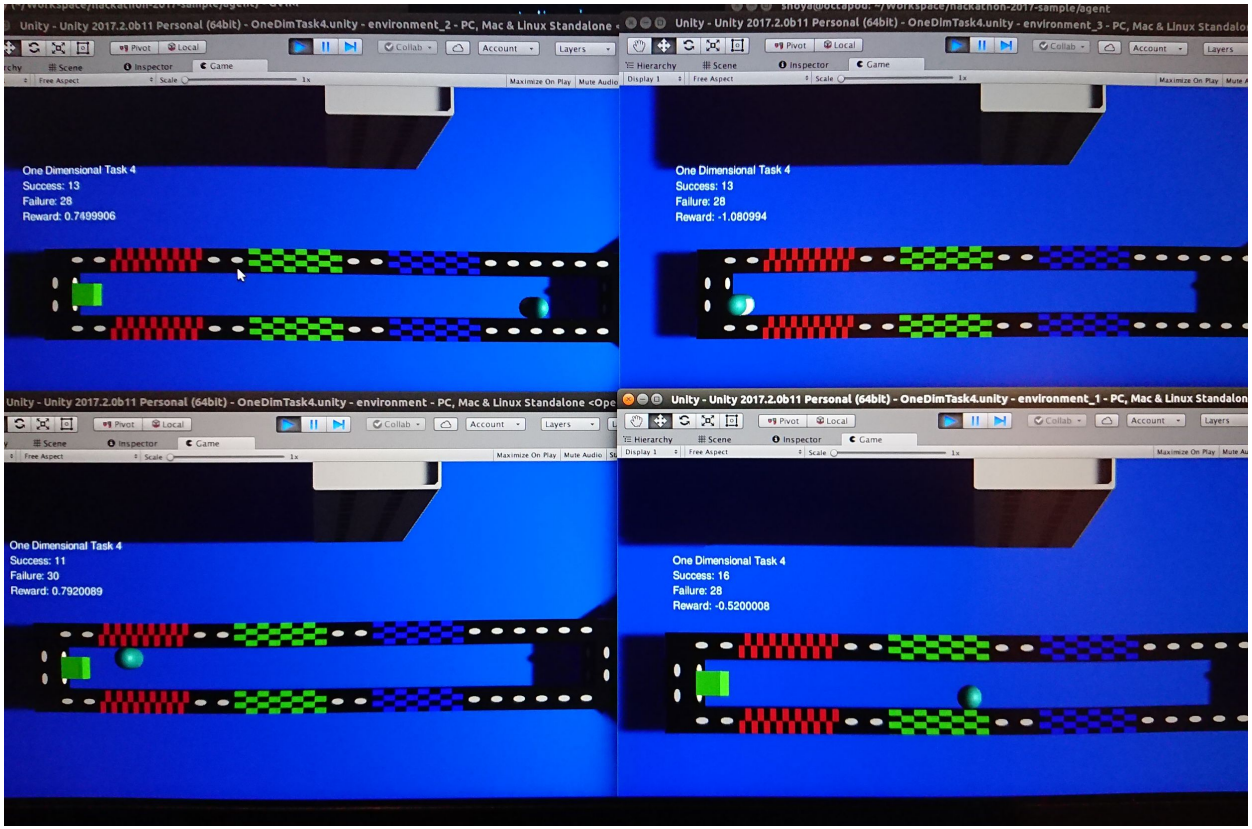
課題結果

1-8b以外は個別に単独で学習

課題番号	成功エピソード数	失敗エピソード数	合計エピソード数 (成功+失敗)	学習状態
1-1	69 (100%)	0 (0%)	69	成功
1-2	69 (100%)	0 (0%)	69	成功
1-3 緑停止	49 (34%)	97 (66%)	146	学習時間不足
1-4 赤停止	53 (32%)	114 (68%)	167	学習時間不足
1-5 青停止	102 (53%)	90 (47%)	192	成功
1-6	65 (43%)	86 (57%)	151	学習時間不足
1-8a 色なし	206(52%)	189(48%)	395	成功
1-8b 色なし	81(49%)	86(51%)	167	学習時間不足
2-1				ロスのみ
2-2				学習中…
3-1	40(29%)	99(71%)	139	学習時間不足
3-2	36(31%)	82(69%)	118	学習時間不足

1-4 学習途中のスクリーンショット

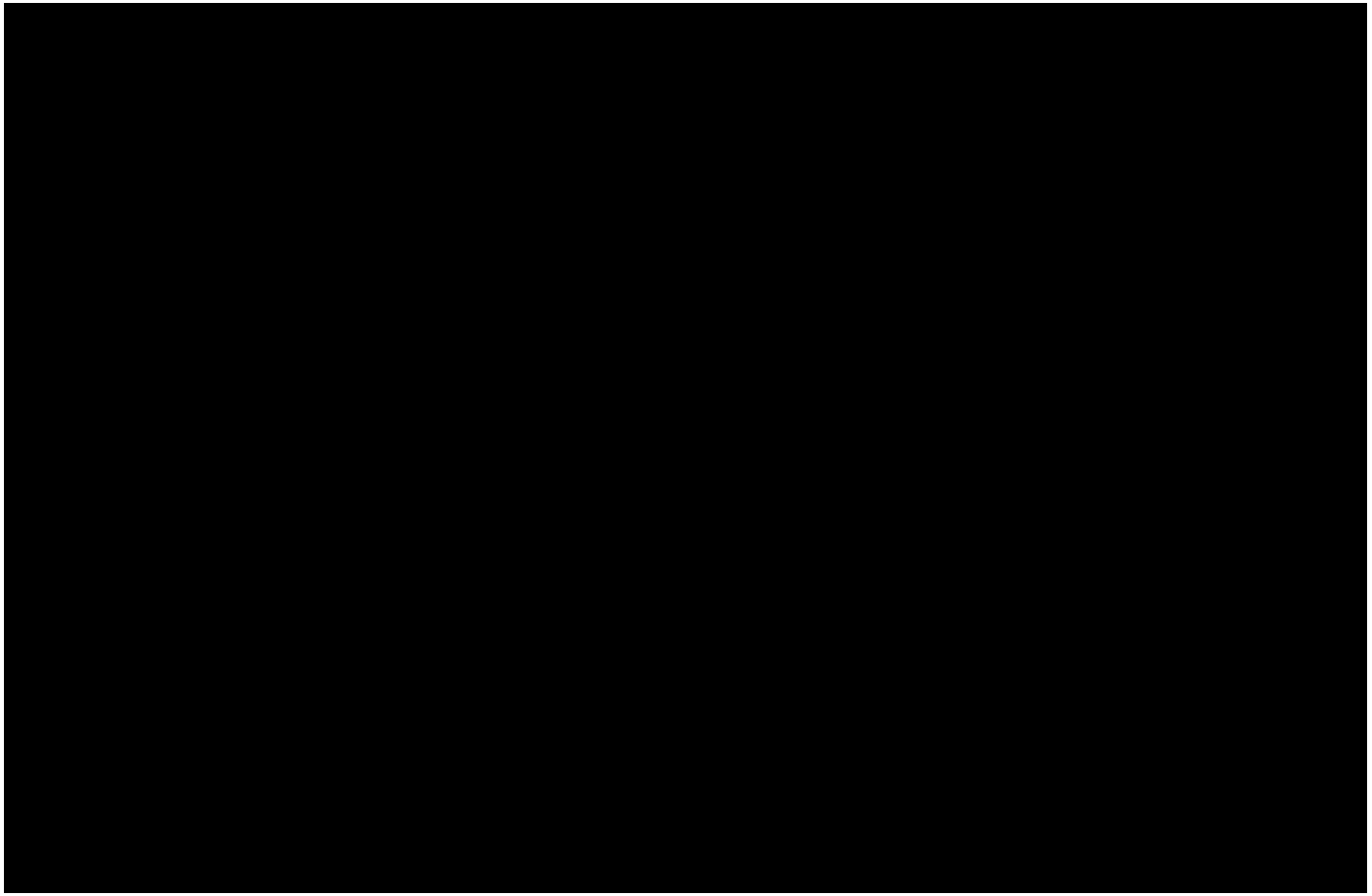
- 赤で停止することでエサが出るタスク



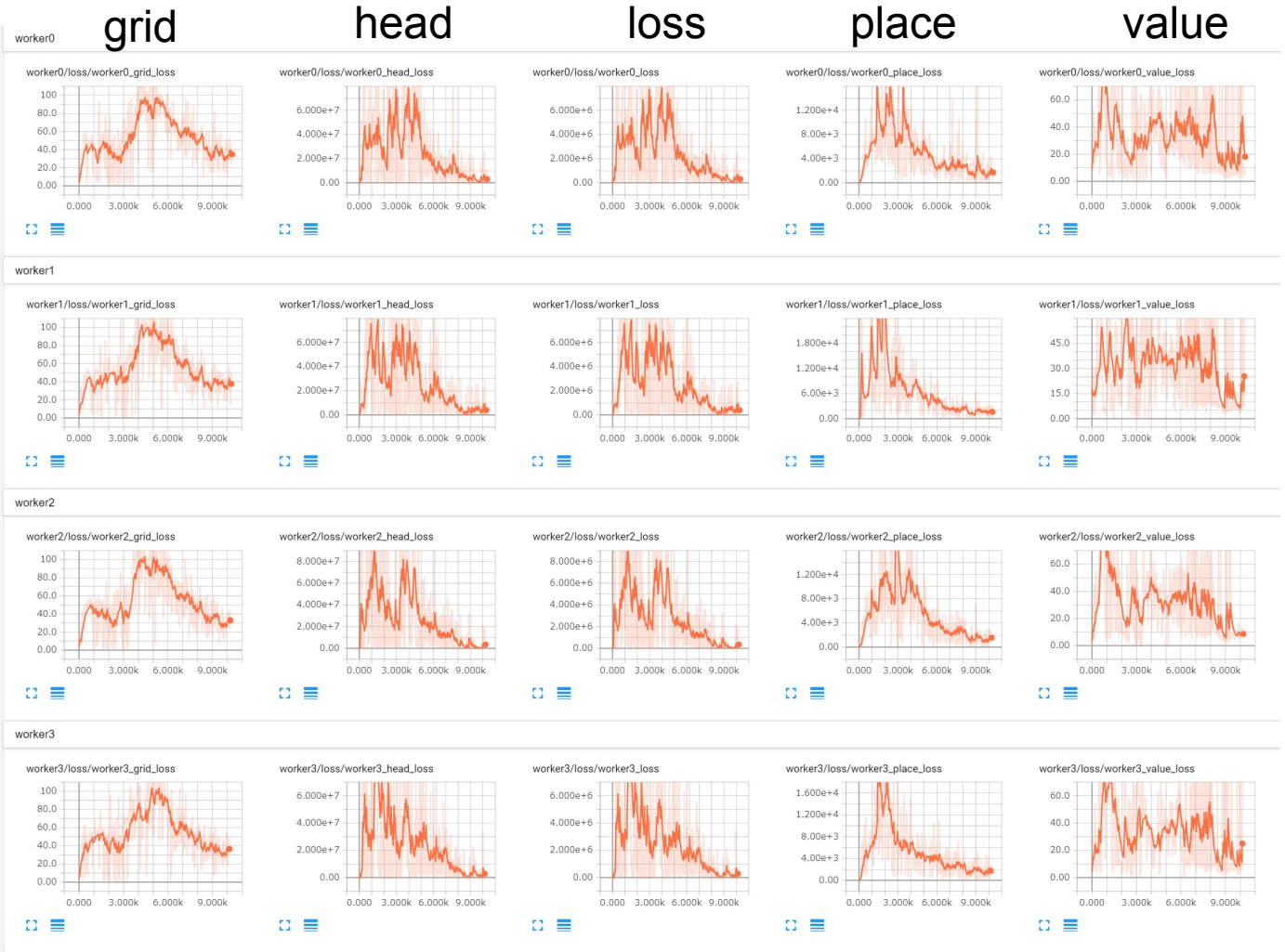
- 4エージェントで学習 → Unityを4つ起動
- 多くの場合で緑の箱を出現させることができた

1-5 終盤の学習の様子

- 青で停止することでエサが出るタスク
- 学習終盤　高確率でタスク成功



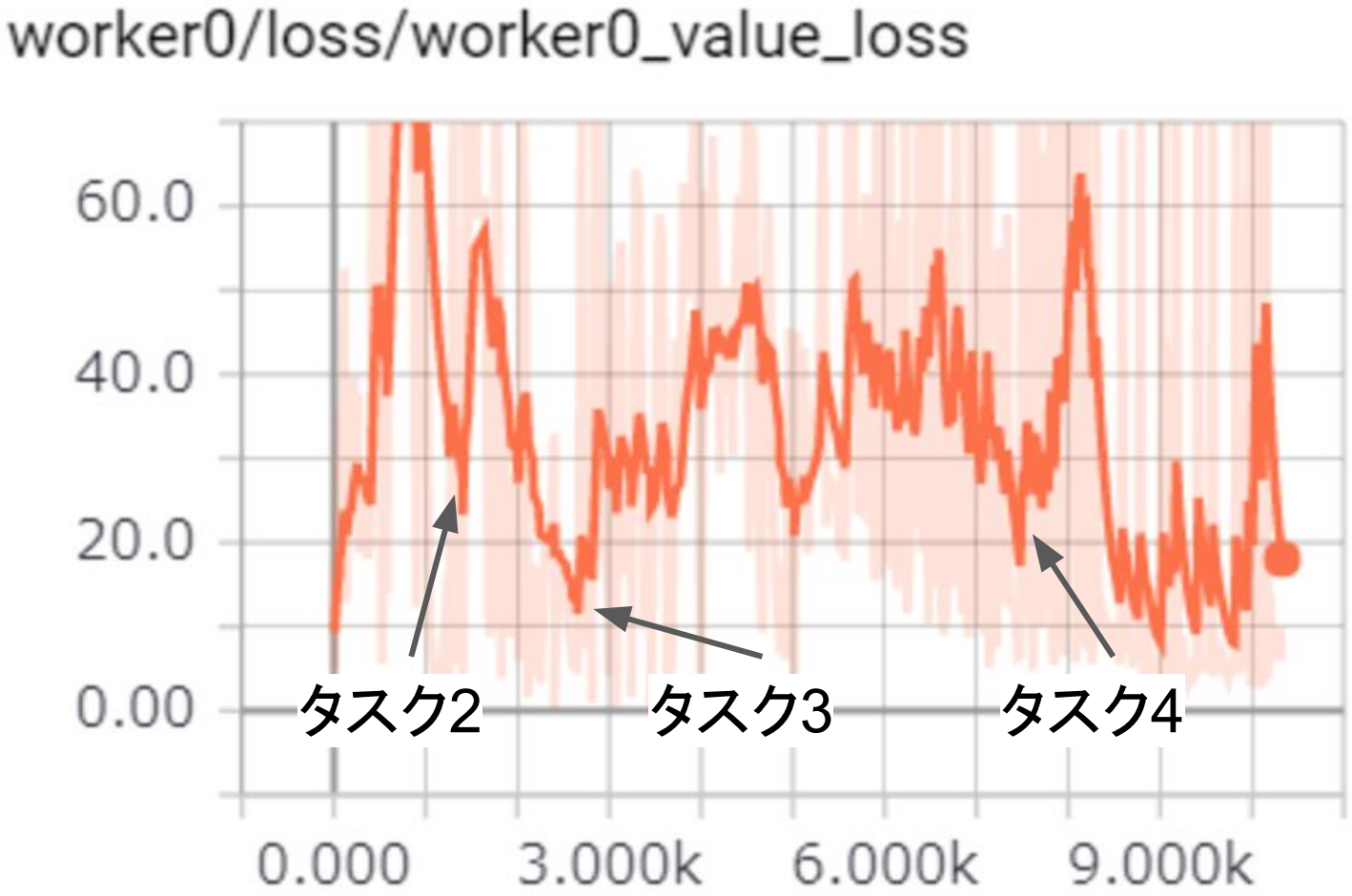
1次元迷路 1~4まで通して学習した時のLOSS



場所細胞のLOSS


worker1/loss/worker1_place_loss
1.800e+4
1.200e+4
6.00e+3
0.00
0.000
3.000k
6.000k
9.000k
ちゃんと下がってる！！

タスク切り替わり時の価値推定のLOSS



十字迷路のLOSS


grid
head
loss
place
value
Agent 1
Agent 2
Agent 3
Agent 4
worker0
worker0/loss/worker0_grid_loss
worker0/loss/worker0_head_loss
worker0/loss/worker0_loss
worker0/loss/worker0_place_loss
worker0/loss/worker0_value_loss
worker1
worker1/loss/worker1_grid_loss
worker1/loss/worker1_head_loss
worker1/loss/worker1_loss
worker1/loss/worker1_place_loss
worker1/loss/worker1_value_loss
worker2
worker2/loss/worker2_grid_loss
worker2/loss/worker2_head_loss
worker2/loss/worker2_loss
worker2/loss/worker2_place_loss
worker2/loss/worker2_value_loss
worker3
worker3/loss/worker3_grid_loss
worker3/loss/worker3_head_loss
worker3/loss/worker3_loss
worker3/loss/worker3_place_loss
worker3/loss/worker3_value_loss

十字迷路の場所細胞のLOSS

実装で苦労したところ

- Unityマルチ未対応問題
  - キャッシュを共有していたため環境を分離
  - Taskスキップモード搭載(A3C)

- Cherrypy へのPOST通信マルチクライアント未対応問題
  - Agentごとにロックを導入
  - より効率的なwsgiサーバを使用

- caffeモデル重い問題(GPU使用率低すぎ)
  - Chainerを捨ててTFに完全移行
  - AlexNetをTFで構築(読み込み&実行時間が大幅に短縮)
  - 複数GPU対応

- タスク1-8xがシリアライズにより読み込めない設定になっていた

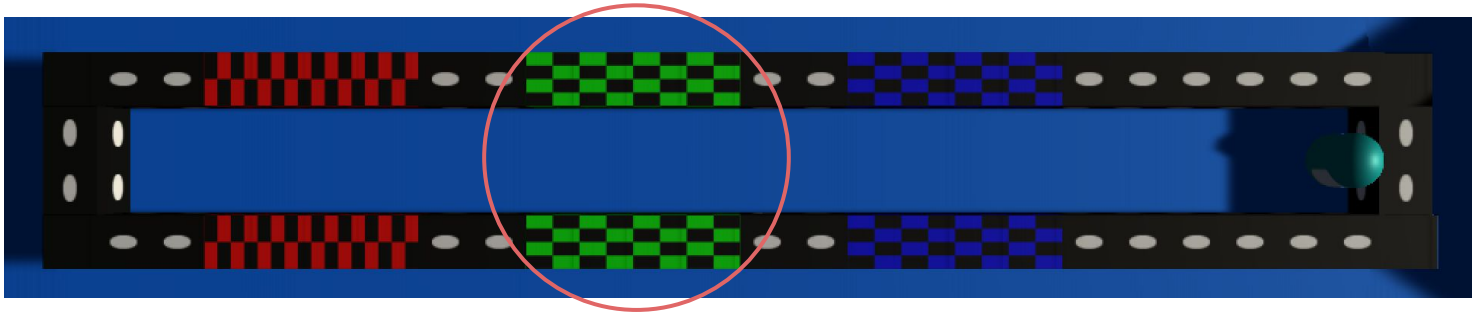
考察

実験結果と神経科学的妥当性

考察①: Task3について

LSTMを導入したにも関わらず**Task3**は難しかった

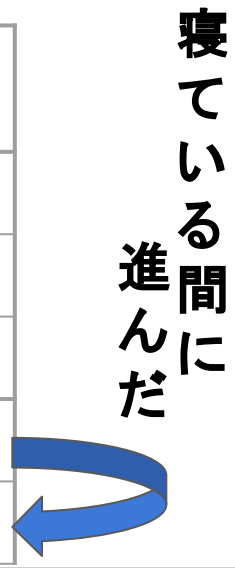
- 真ん中の難しさ
  - 赤と青は端の壁に衝突し、ある程度留まることができる
- 偶然の発見の難しさ
  - 通信遅延で止まり、偶然発見される可能性がある
  - Unity ⇔ A3Cの処理速度を改善しすぎた



考察②: 学習エピソード数

- 成功したタスクは何れも学習エピソード数が多い
  - 失敗エピソードが多いタスクも、学習を進めることで改善される可能性がある
- 成功した1-8aはエピソード数が極端に多い
  - 色なしの1次元迷路　ある位置で停止するとエサが出る
  - タスク自体が難しく、学習に時間がかかる

課題番号	成功エピソード数	失敗エピソード数	合計エピソード数 (成功+失敗)	学習状態
1-3 緑停止	49 (34%)	97 (66%)	146	学習時間不足
1-4 赤停止	53 (32%)	114 (68%)	167	学習時間不足
1-5 青停止	102 (53%)	90 (47%)	192	成功
1-8a 色なし	206(52%)	189(48%)	395	成功
1-8b 色なし	81(49%)	86(51%)	167	学習時間不足



考察③: Feature Extractorについて

- ImageNet 学習済みモデルをそのまま使用
  - ハッカソンの仕様によりFineTunningは行わなかった

問題点

- UnityのShading設定により1人称視点では色が薄い
  - 十分な特徴量が抽出できていない可能性

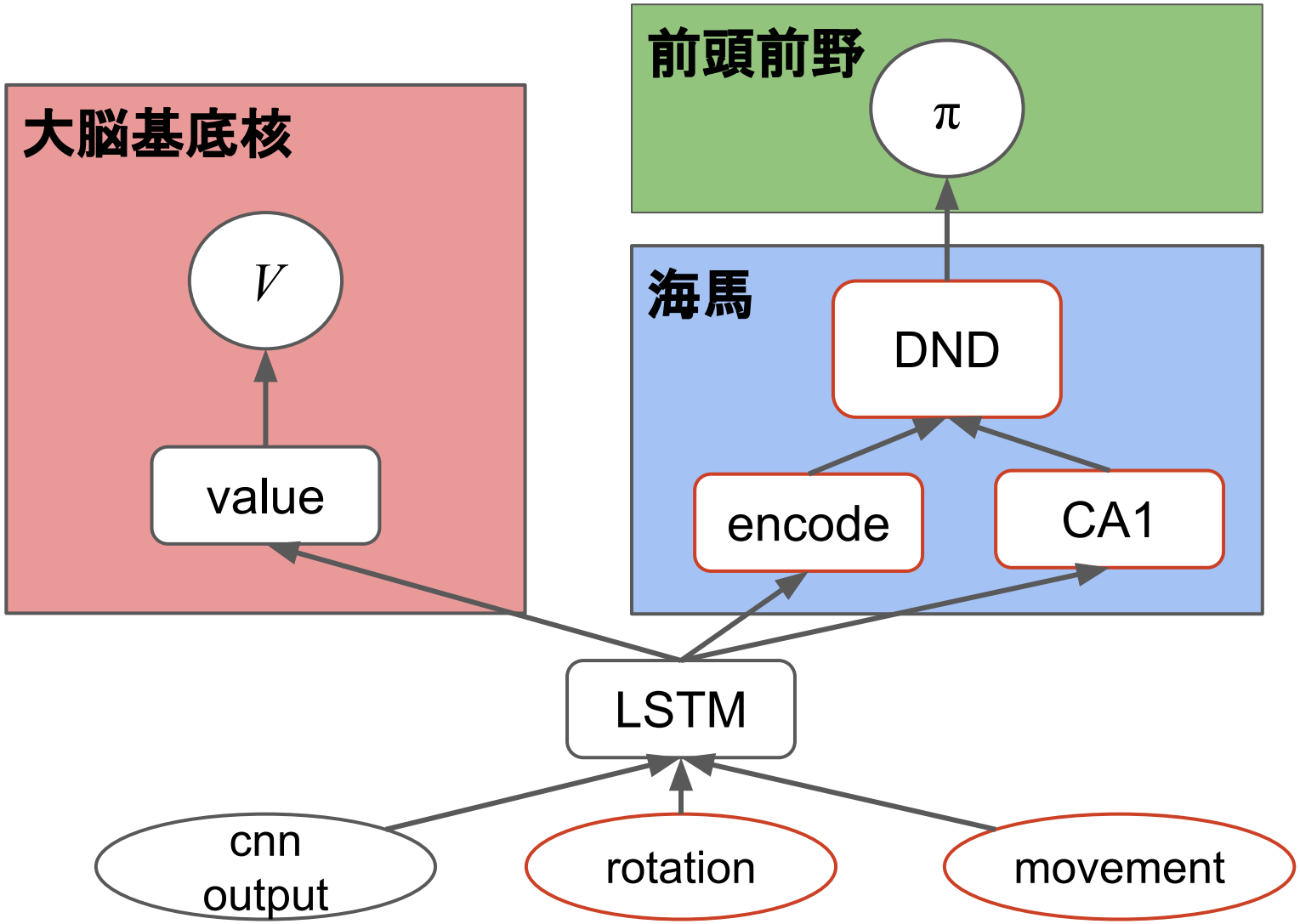
- 一般物体認識で学習したモデルをそのまま使用するのは無理がありそう
  - Feature ExtractorのFine Tunningも含めた学習により、精度が向上する可能性

神経科学的妥当性

		✔			✔
海馬内活動	リプレイ	✔	脳領域構造	CA1	✔
	プリプレイ			CA2	
	場所細胞	✔		CA3	✔
	グリッド細胞	✔		歯状回	
	頭部方向細胞	✔		嗅内皮質	✔
	シータ位相歳差			海馬支脚	
	スパース表現			Perirhinal Cortex	
	パターン補完			Postrhinal Cortex	
	細胞新生		その他	コネクトームの導入	✔
行動機能	自律的フェーズ変化			BiCAMON可視化	
	エピソード記憶	✔		その他	✔
	場所の再認	✔			
	記憶転送				
	ナビゲーション/空間認知	✔			
	Path integration				

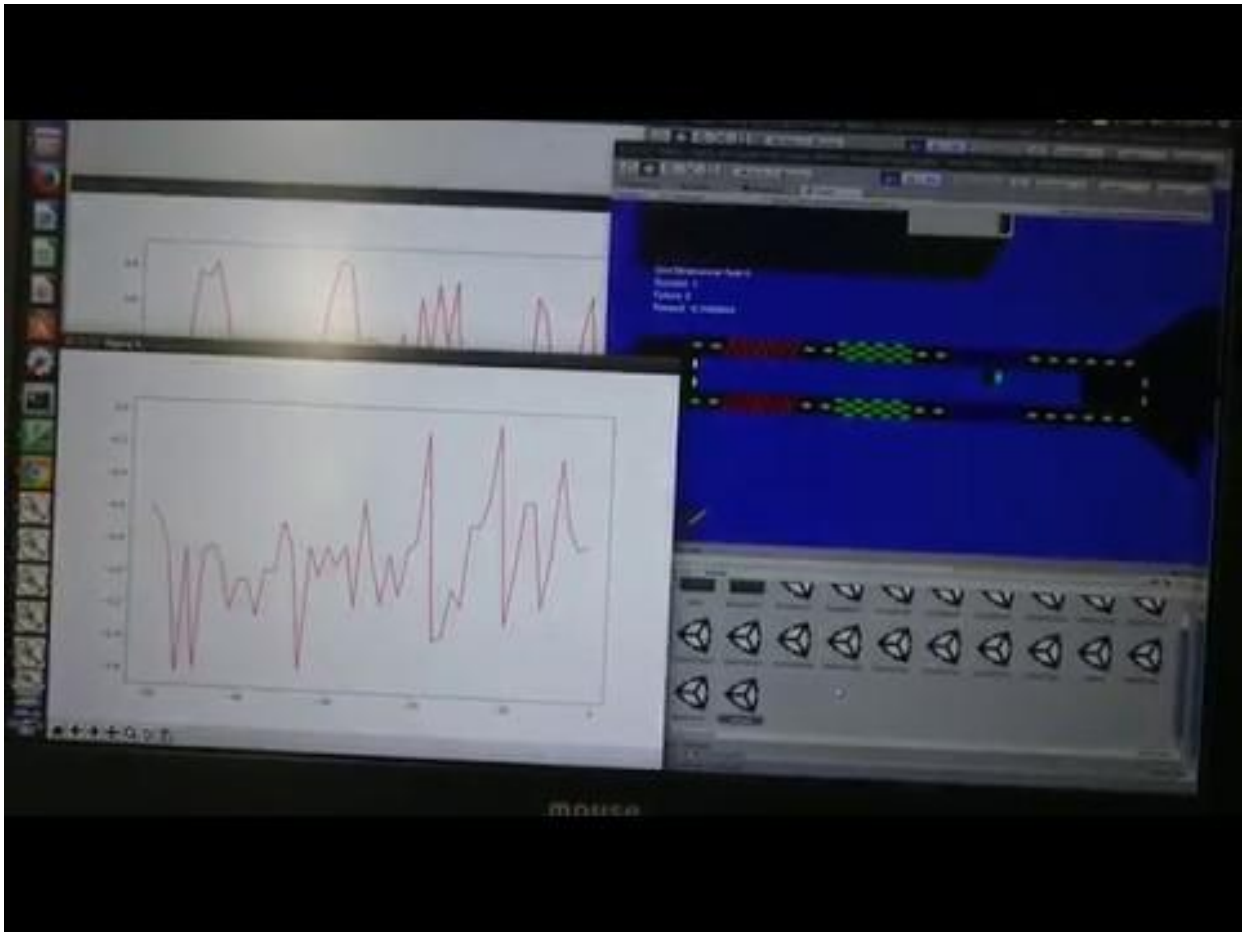
新規性

アーキテクチャ


前頭前野
π
大脳基底核
V
value
海馬
DND
encode
CA1
LSTM
cnn
output
rotation
movement

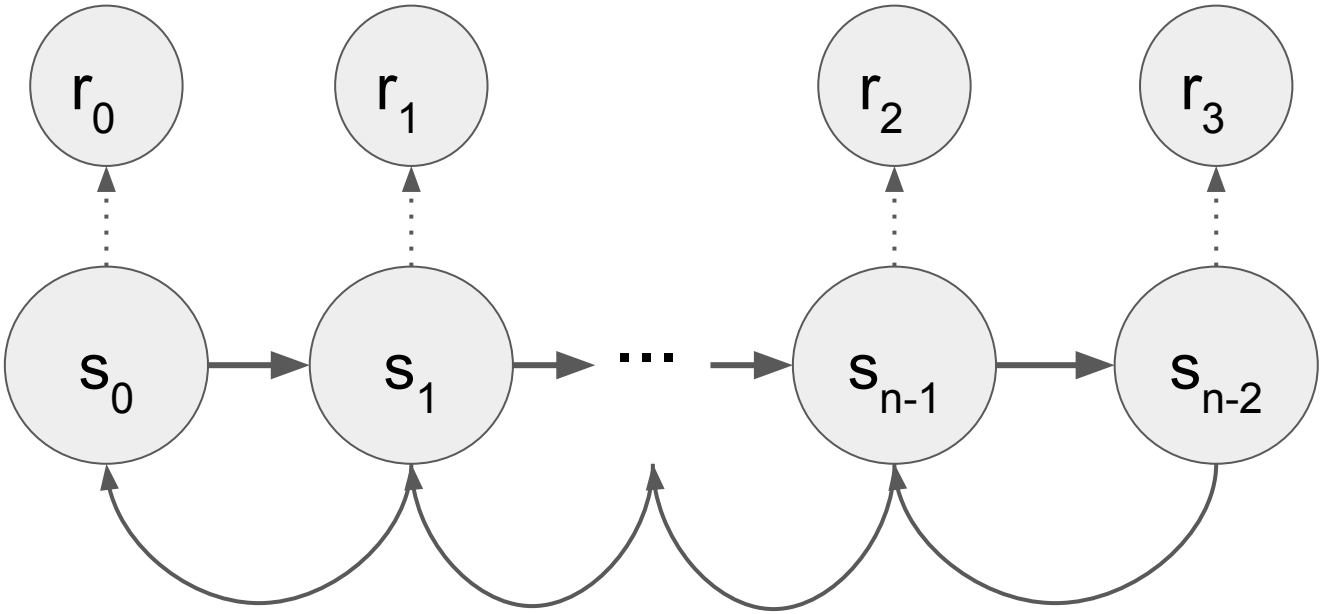
価値推定のグラフ

潜在的にどれくらい報酬をもらえる状態なのか予測 V(s)を可視化



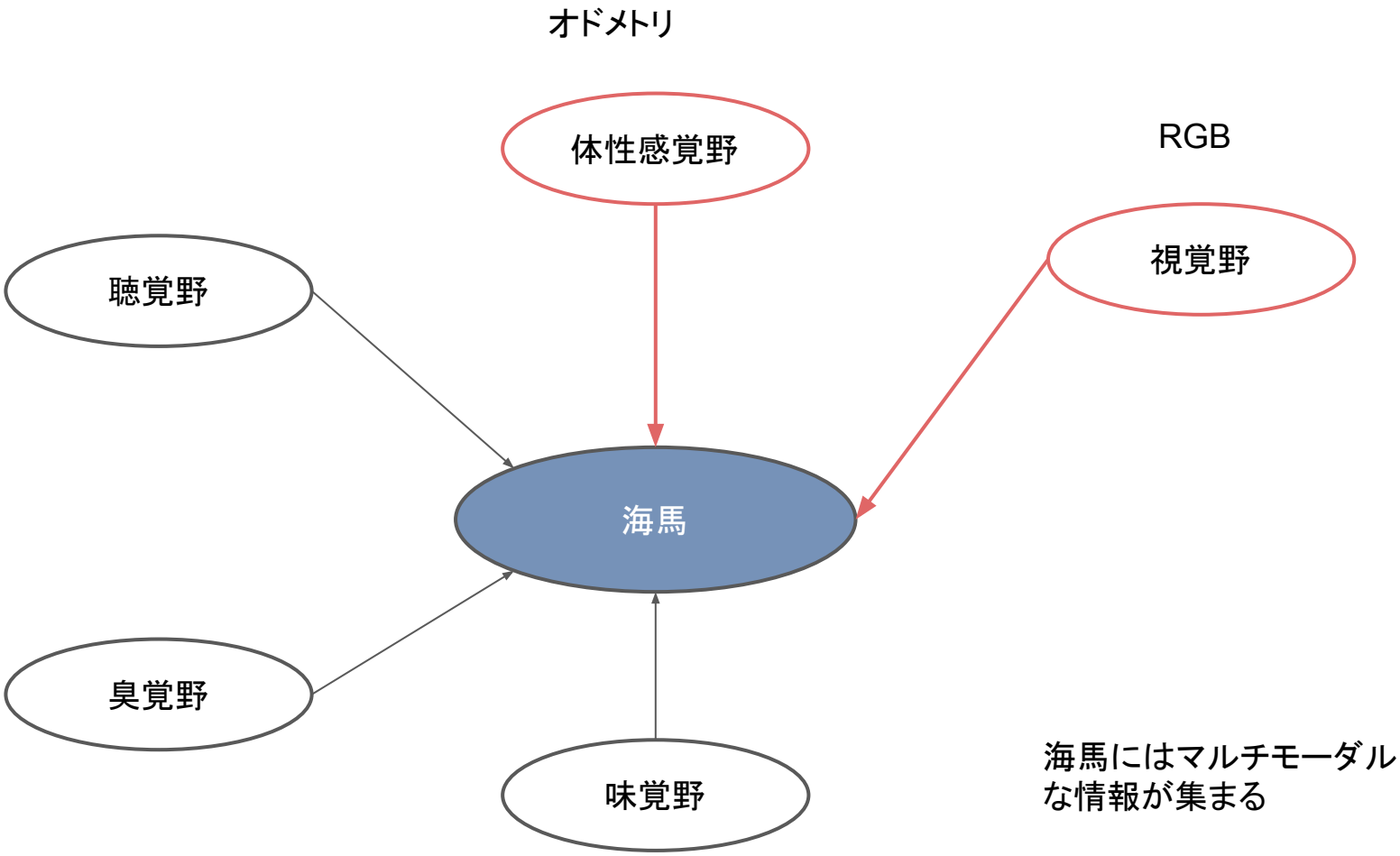
Reverse Replay

Experience Replayの代わりに50step毎に
Reverse Replay[David+ 06]を行なって長期的にどのくらい報酬をもらえるか(価値)を計算



Reverse Replay で計算した値を**エピソード記憶**に保存

海馬と感覚情報



海馬にはマルチモーダルな情報が集まる

まとめ

Future Worksとまとめ

Future Works

- Finetuning
  - AlexNet 及び 空間認知モジュール
- 可視化
  - 空間認知(Place, Head, Grid)のマッピング
- 歯状回のモデリング
  - 新生ニューロン = Progressive Neural Net

まとめ

Brain-inspired A3C

海馬を世界最先端の工学モデルに組み込む!

神経科学的:
空間認知 + エピソード記憶の機能をニューラルネットで学習

工学的:
A3C + NEC + 空間認知情報

結果:
時系列+空間特徴を活かし、
課題迷路タスクをある程度こなすことが出来た

APPENDIX

参考文献の表記

参考文献は以下のように表記されています

- 論文の場合: [Hinton+ 12]
  - APPENDIXのReferencesはAPA形式

- 本の場合: [Asakawa: 16]
  - APPENDIXのReferencesはAPA形式

- WEBページの場合: [Karpathy:: 17]
  - APPENDIXのReferencesはAPA形式

References

- [Mnih+ 15] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. Nature, 5

- [Minh+ 13] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

- [Minh+ 16] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. (2016, June). Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning* (pp. 1928-1937).

- [Pritzel+ 17] Pritzel, A., Uria, B., Srinivasan, S., Puigdomènech, A., Vinyals, O., Hassabis, D., ... & Blundell, C. (2017). Neural Episodic Control. arXiv preprint arXiv:1703.01988.

- [David+ 06] David J Foster and Matthew A Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature, 440(7084):680–683, 2006.

- [Wimmer+ 12] Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270-273.

References

- [Masami 07] 龍野正実. (2007). メモリーリプレイと記憶の固定化 . *生物物理*, *47*(6), 368-377

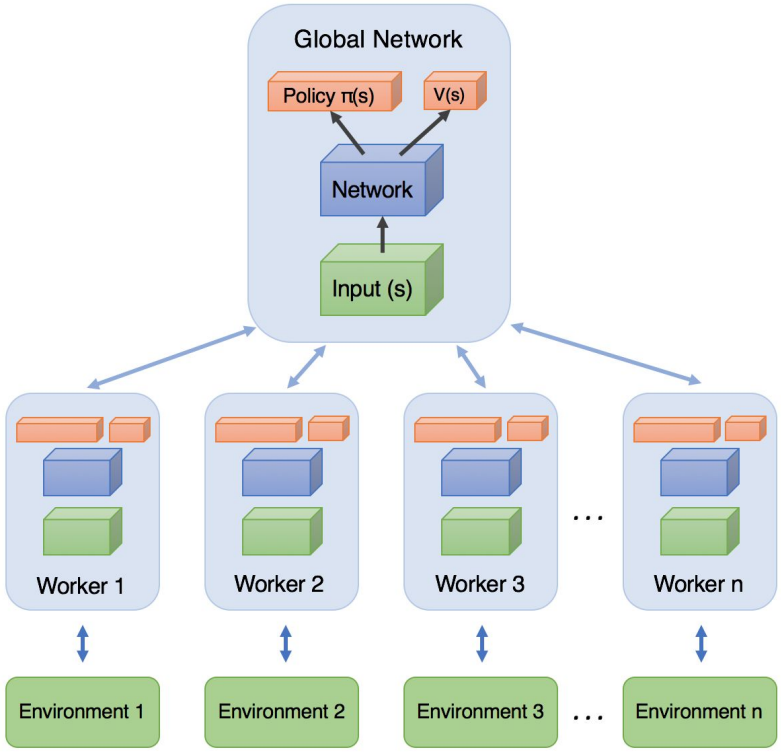
- [Bato+ 95] Barto, A. G. (1995). 1" 1 Adaptive Critics and the Basal Ganglia. *Models of information processing in the basal ganglia*, 215.

学習

A3Cによって学習を行う

Actor: **前頭前野モジュール**

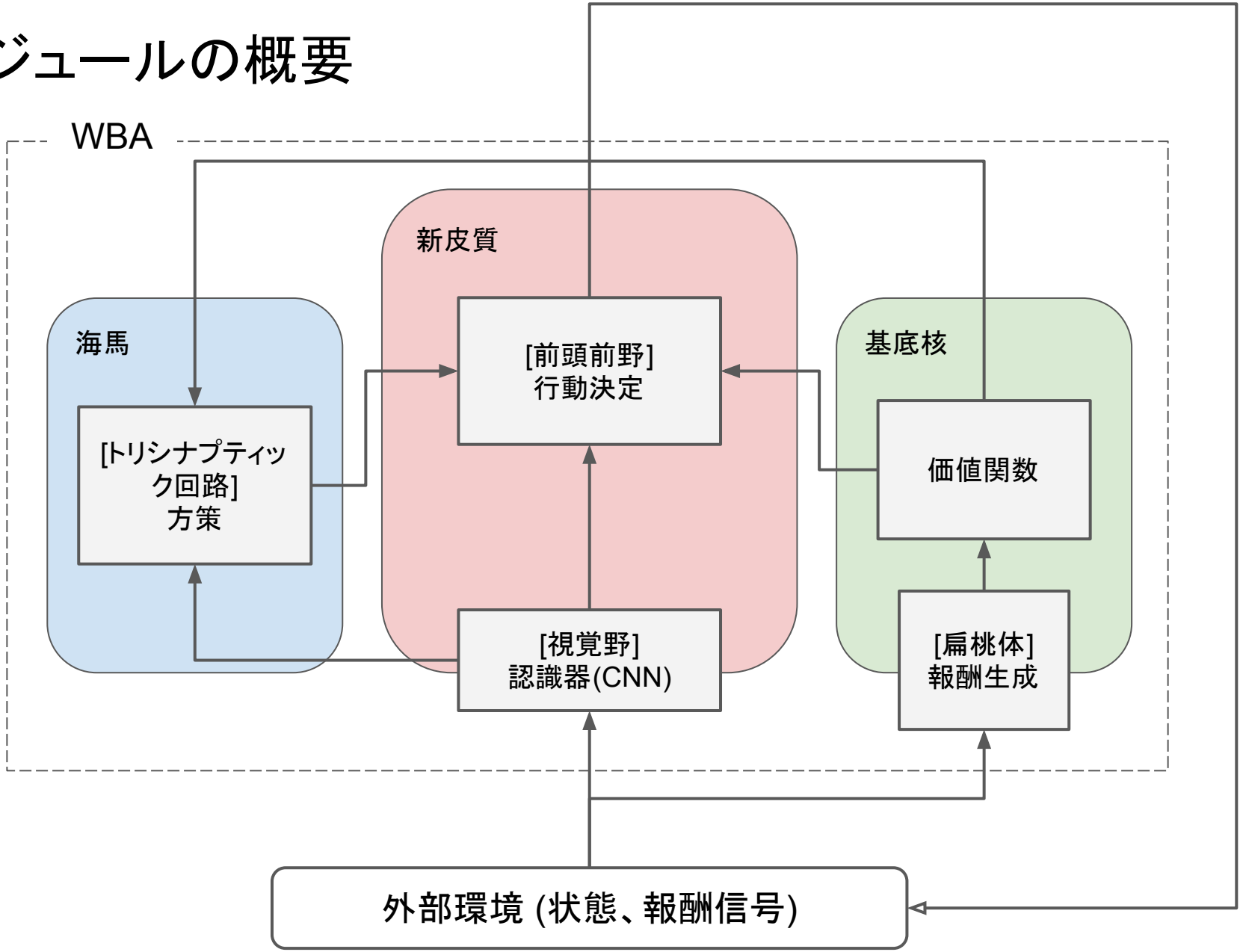
Critic: **海馬-基底核モジュール**



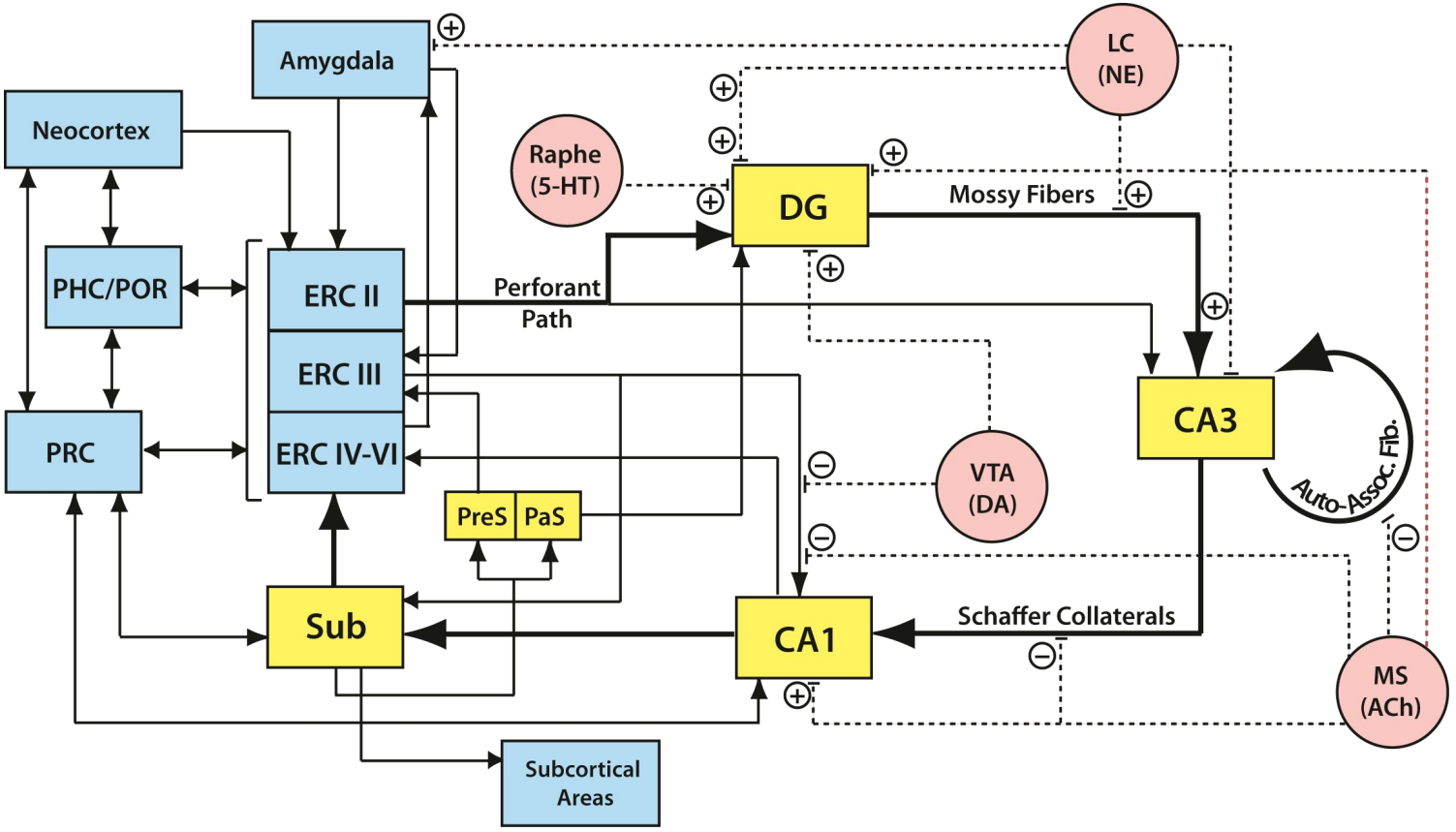
A3Cで学習することにより
RNNを使うことができる

時間情報の必要なタスクに有用

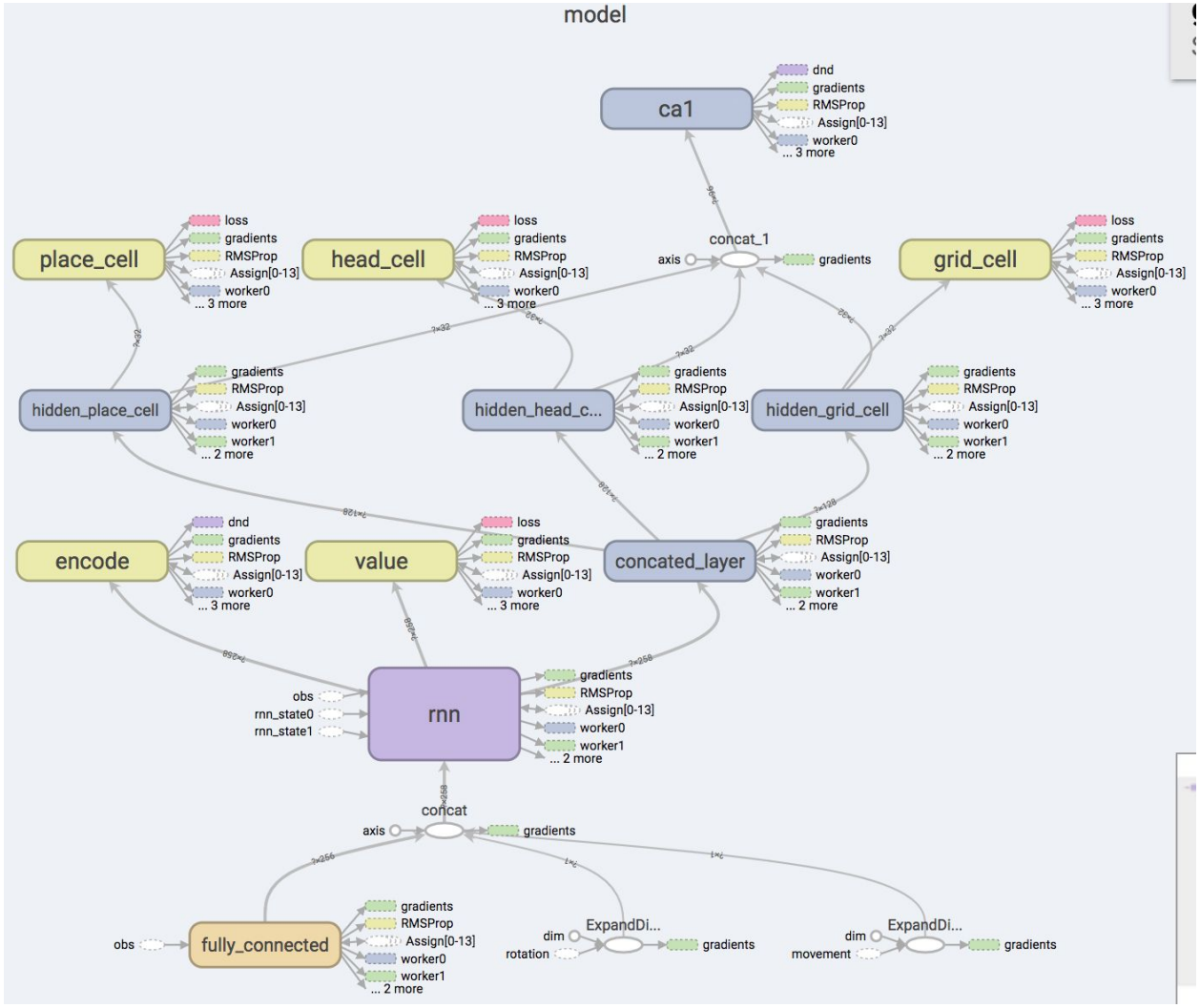
モジュールの概要


WBA
新皮質
海馬
基底核
[前頭前野]
行動決定
[トリシナプティック回路]
方策
価値関数
[視覚野]
認識器(CNN)
[扁桃体]
報酬生成
外部環境 (状態、報酬信号)

海馬の機能実装


Amygdala
LC
(NE)
Neocortex
Raphe
(5-HT)
DG
Mossy Fibers
PHC/POR
ERC II
Perforant
Path
ERC III
CA3
PRC
ERC IV-VI
VTA
(DA)
Auto-Assoc. Fib.
PreS
PaS
Sub
CA1
Schaffer Collaterals
MS
(ACh)
Subcortical
Areas

アーキテクチャ


model
ca1
dnd
gradients
RMSProp
Assign[0-13]
worker0
... 3 more
concat_1
axis
gradients
place_cell
head_cell
grid_cell
loss
gradients
RMSProp
Assign[0-13]
worker0
... 3 more
hidden_place_cell
hidden_head_c...
hidden_grid_cell
gradients
RMSProp
Assign[0-13]
worker0
worker1
... 2 more
encode
value
concated_layer
dnd
loss
gradients
RMSProp
Assign[0-13]
worker0
worker1
... 3 more
... 2 more
rnn
obs
rnn_state0
rnn_state1
gradients
RMSProp
Assign[0-13]
worker0
worker1
... 2 more
concat
axis
gradients
fully_connected
obs
gradients
RMSProp
Assign[0-13]
worker0
worker1
... 2 more
ExpandDi...
dim
rotation
gradients
ExpandDi...
dim
movement
gradients

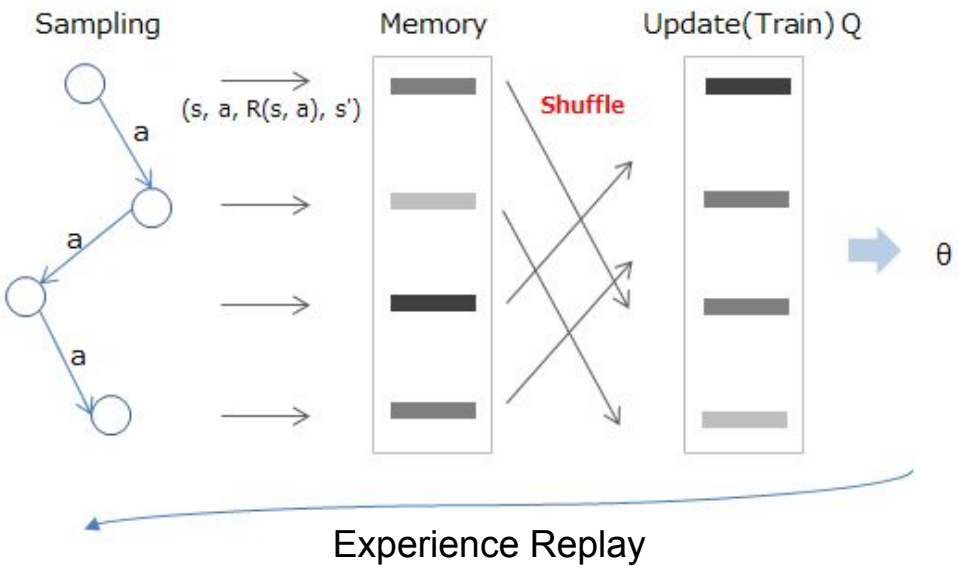
Concept

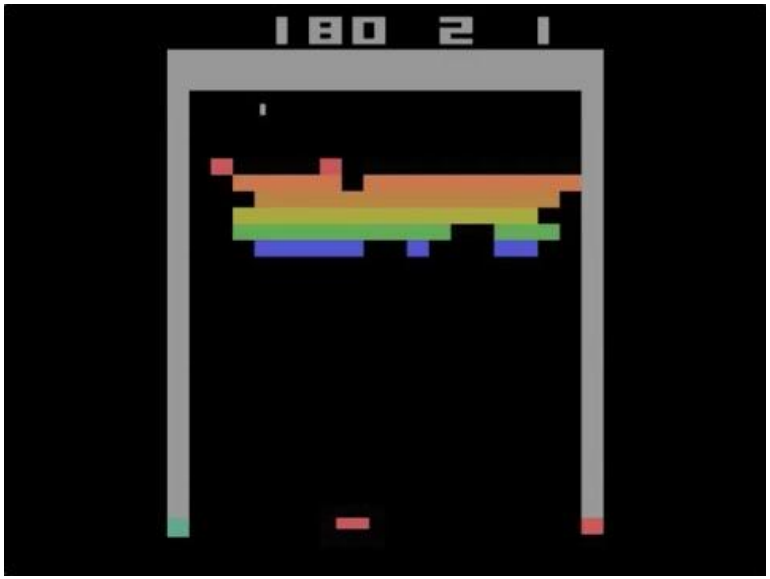
Brain-inspired A3C

Neuroscience ✖ ML

DQNで良いのか

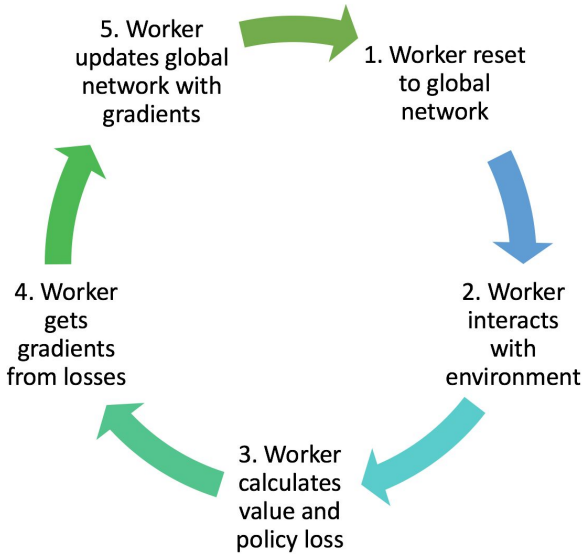
- DQNはもはや時代遅れ...?(2013年)
  - 諸刃の剣: Experience Replay
  - RNNが使えない
  - 連続値を扱いにくい

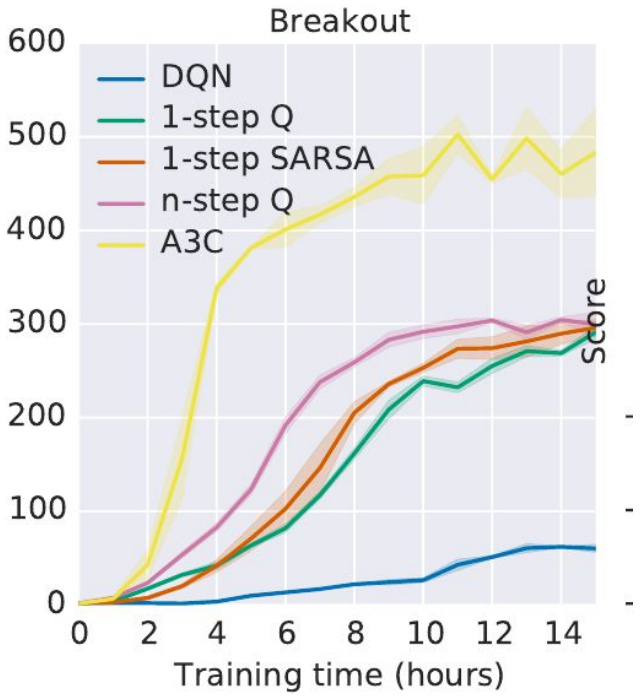




なぜA3Cなのか

- A3Cの躍進(2016年)
  - Asynchronous (複数のエージェントで非同期に更新)
  - Advantage (PGのベースラインとしてAdvantageを使用)
  - Actor-Critic (πとVを別々に学習)





1-3のプレイ動画