Hierarchical Intrinsically Motivated Agent based on Free Energy Principle

Petr Kuderov, AIRI, MIPT

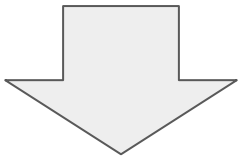
Biologically-plausible models and learning methods in AI

Cognitive sciences

- Goal: explain natural intelligence
- Limited by physiology
- Computational models is AI
- Computational cognitive models aren't as successful at generating complex behaviour as modern DeepRL models

Artificial Intelligence

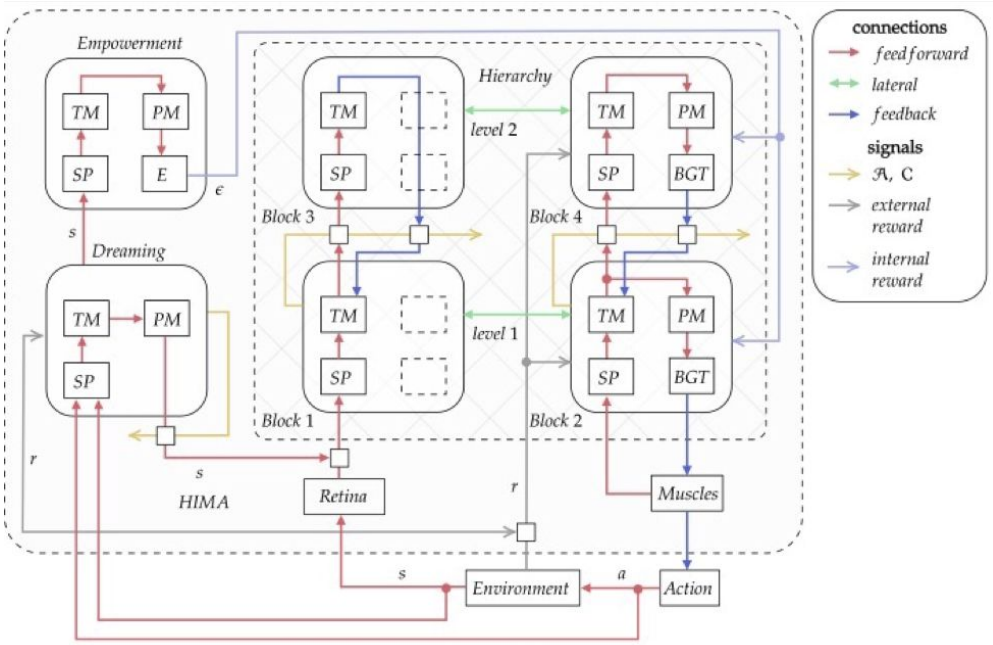
- Goal: build artificial intelligence
- Any solution works
- Wide range of complex benchmarks
- Weak compatibility with cognitive models prevents collaborations



Biologically-plausible approach in AI allows collaboration toward both goals

Hierarchical Intrinsically Motivated Agent (HIMAgent)

- Biologically plausible model of an autonomous agent
- Hierarchical experience organization and behavior control
- Learns by reinforcement signal
- Accumulates and reuses its experience to reach changing goals



Dzhivelikian, E., Latyshev, A., Kuderov, P. *et al.* Hierarchical intrinsically motivated agent planning behavior with dreaming in grid environments. *Brain Inf.* 9, 8 (2022). Link
Source code: https://github.com/AIRI-Institute/him-agent

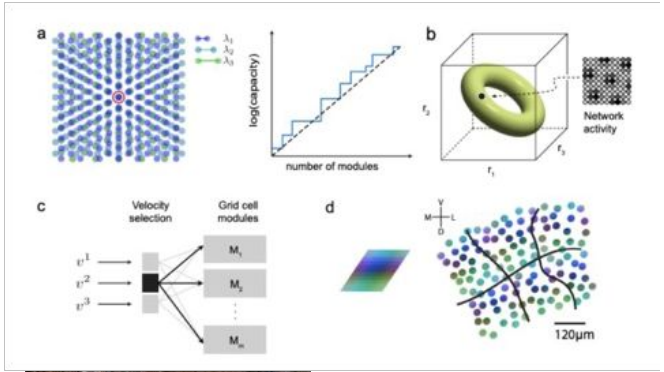
Related works

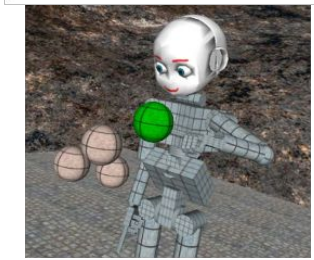
HTM framework:

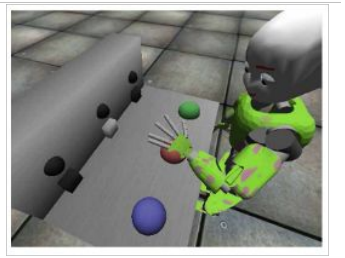
- Hole, K.J., Ahmad, S. A thousand brains: toward biologically constrained AI. SN Appl. Sci. 3, 743 (2021). https://doi.org/10.1007/s42452-021-04715-0
- Klukas M, Lewis M, Fiete I (2020) Efficient and flexible representation of higher-dimensional cognitive variables with grid cells. PLOS Computational Biology 16(4): e1007796. https://doi.org/10.1371/journal.pcbi.1007796

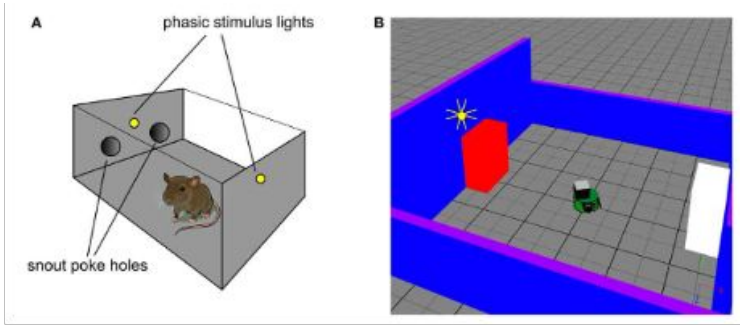
Basal ganglia, hierarchical memory, intrinsic motivation:

- Santucci, V.G., Baldassarre, G., Mirolli, M.: GRAIL: A goal-discovering robotic architecture for intrinsically-motivated learning (IEEE Transactions on Cognitive and Developmental Systems). IEEE Trans. Cogn. Dev. Syst. 8(3), 214–231 (2016)
- Bolado-Gomez, R., Gurney, K.: A biologically plausible embodied model of action discovery. Front. Neurorobot. 7(MAR), 1–24 (2013).
- Fiore, V.G. et al: Keep focussing: Striatal dopamine multiple functions resolved in a single mechanism tested in a simulated humanoid robot. Front. Psychol. 5(FEB), 1–17 (2014).
- Klyubin AS, Polani D, Nehaniv CL (2005) All else being equal be empowered. In: Capcarrère MS, Freitas AA, Bentley PJ, Johnson CG, Timmis J (eds) Advances in artificial life. Springer, Berlin, pp 744–753



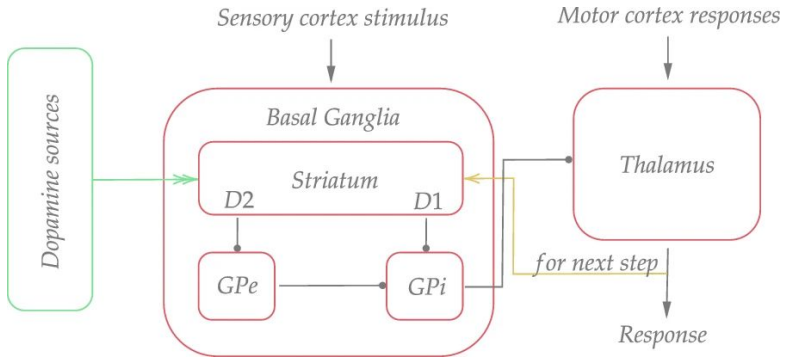


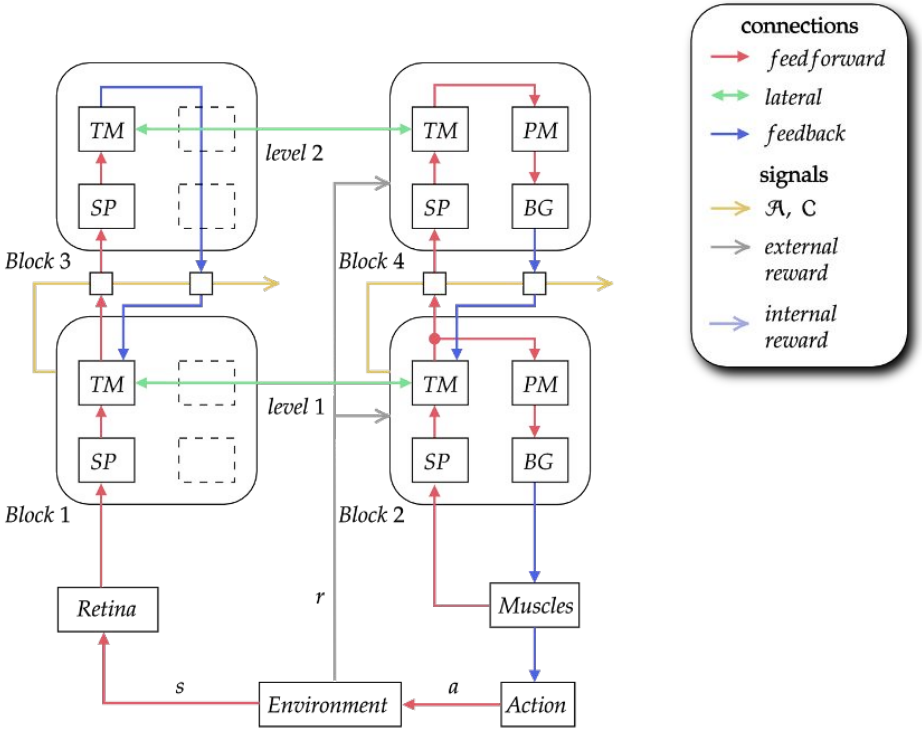




Hierarchical experience organization

- Associative temporal memory learns sparse distributed representation of state–actions
- Builds a compact model of the environment
- The basal ganglia model learns effective action policy on different levels of abstraction





The basal ganglia, cortex and thalamus cooperate to select behavior

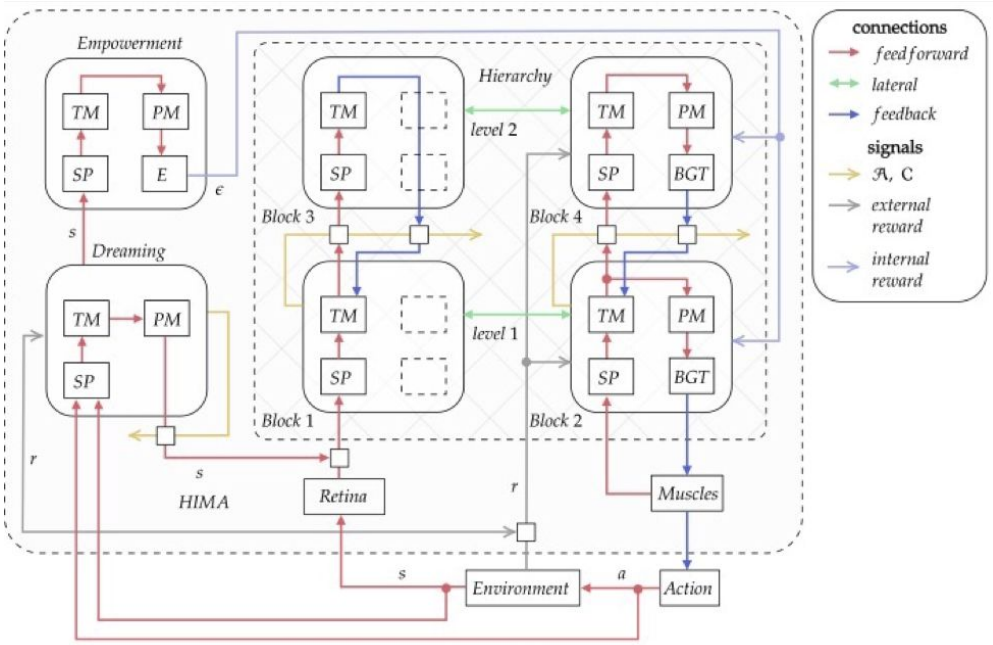
Learned model utilization

Generate an intrinsic motivation signal

- Based on Empowerment
- Drives the agent in the absence of the extrinsic signal
- Allows modulation between intrinsically motivated exploration and extrinsically motivated goal-directed behavior

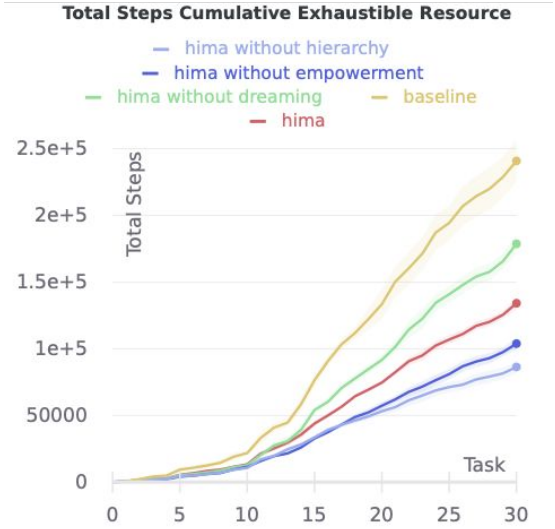
Act in imagination (aka dreaming)

- Refines value estimation for policies
- Speeds up the learning process

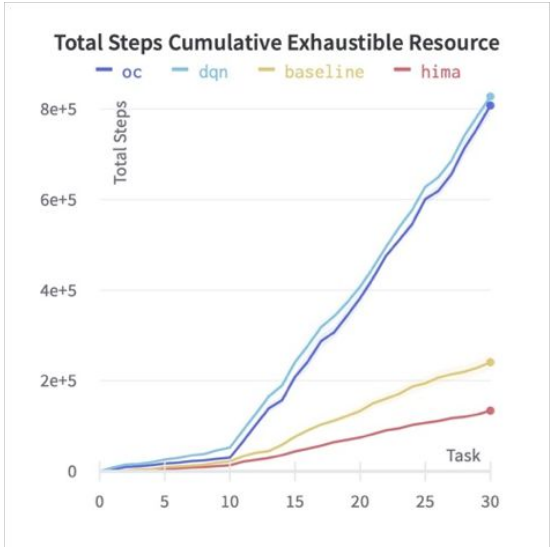


Klyubin AS, Polani D, Nehaniv CL (2005) All else being equal be empowered. In: Capcarrère MS, Freitas AA, Bentley PJ, Johnson CG, Timmis J (eds) Advances in artificial life. Springer, Berlin, pp 744–753

HIMAgent results in four rooms environment



HIMAgent ablation study



Comparison with Deep RL methods: DQN [1] and Option-Critic [2]

[1] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M.: Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602 (2013)
[2] Bacon, P.-L., Harb, J., Precup, D.: The option-critic architecture. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI'17, pp. 1726–1734. AAAI Press (2017)

What can be improved

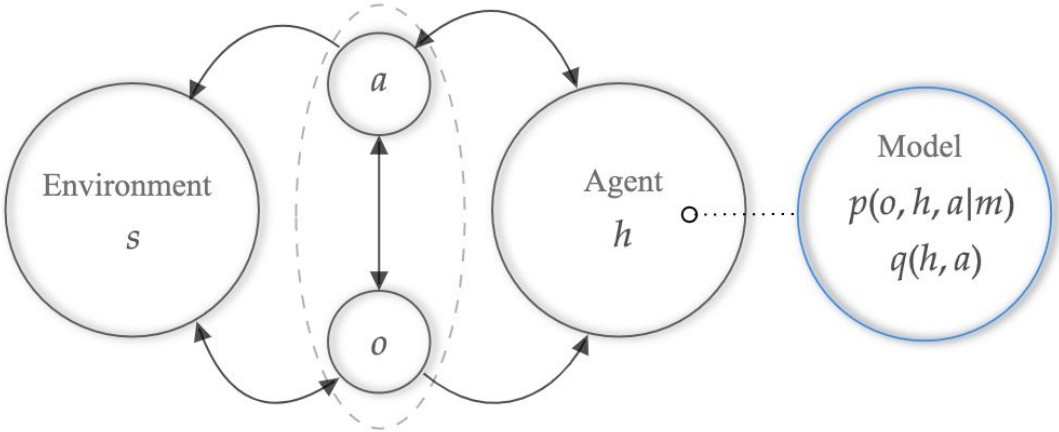
- HTM Temporal Memory is deterministic
  - predictions are deterministic and equiprobable
  - … but, world and behavior aren’t
  - solution: make TM probabilistic ⇒ Belief TM
- Motivation model lacks biological plausibility
  - how basic values incorporating goal-directed behavior are represented
  - how an alternation between exploratory and goal-directed behavior is controlled
  - how an alternation between conscious behavior [re-]planning and acting by habit is controlled
- Spatiotemporal processing discrimination ability is too strong
  - it is a feature of HTM TM, which is not useful for making spatiotemporal abstractions

What can be improved

- **HTM Temporal Memory is deterministic**
  - predictions are deterministic and equiprobable
  - … but, world and behavior aren't
  - solution: make TM probabilistic ⇒ Belief TM
- **Motivation model lacks biological plausibility**
  - how basic values incorporating goal-directed behavior are represented
  - how an alternation between exploratory and goal-directed behavior is controlled
  - how an alternation between conscious behavior [re-]planning and acting by habit is controlled
- Spatiotemporal processing discrimination ability is too strong
  - it is a feature of HTM TM, which is not useful for making spatiotemporal abstractions

Active Inference and Free Energy Principle may be used as a theoretical basis for a computational biologically-plausible model with such requirements

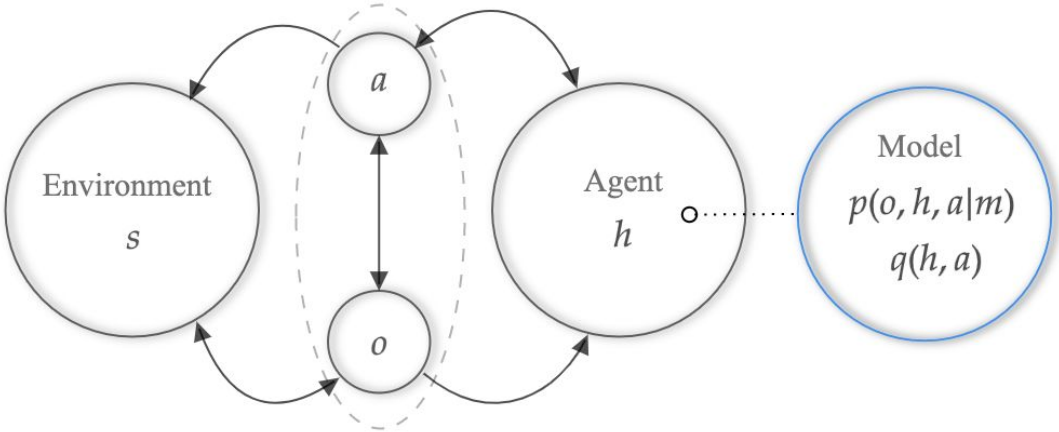
Active Inference



$$p(s, h|a, o) = p(s|a, o)p(h|a, o)$$

Autonomy condition

Active Inference



$$p(s, h|a, o) = p(s|a, o)p(h|a, o)$$

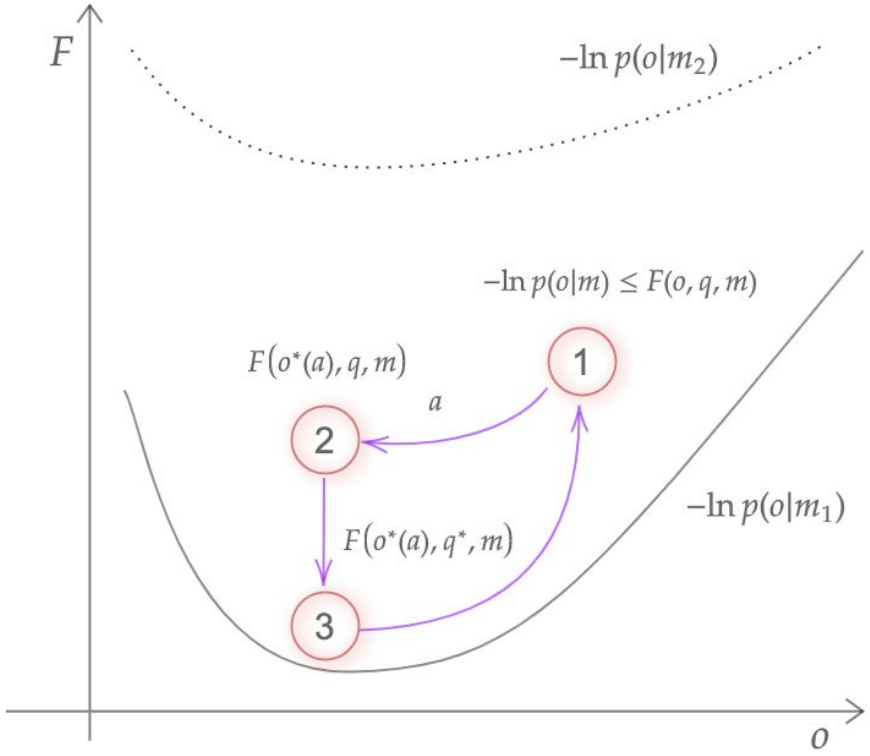
Autonomy condition

Vatiational free energy minimization

$$\underbrace{-\ln p(o)}_{\text{evidence}} \leq F(o, q, m) = -ELBO$$

$$F(o, q, m) = \int q(h, a) \ln \frac{q(h, a)}{p(o, h, a|m)} dh\, da =$$
$$= F[q||p(o, h, a|m)]$$

Active Inference

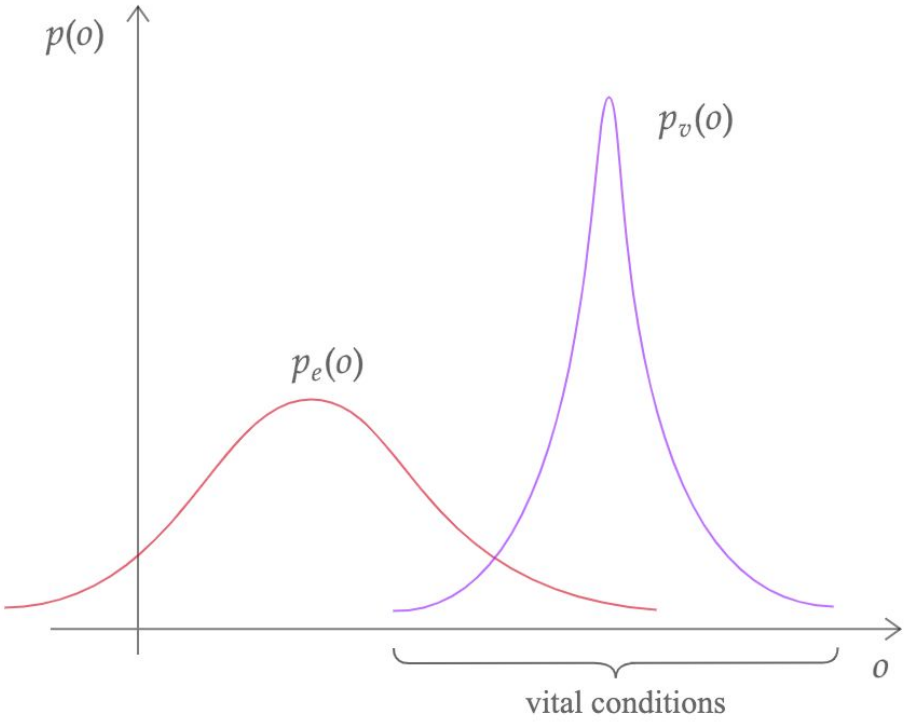


1-2: action $a \sim q(a)$

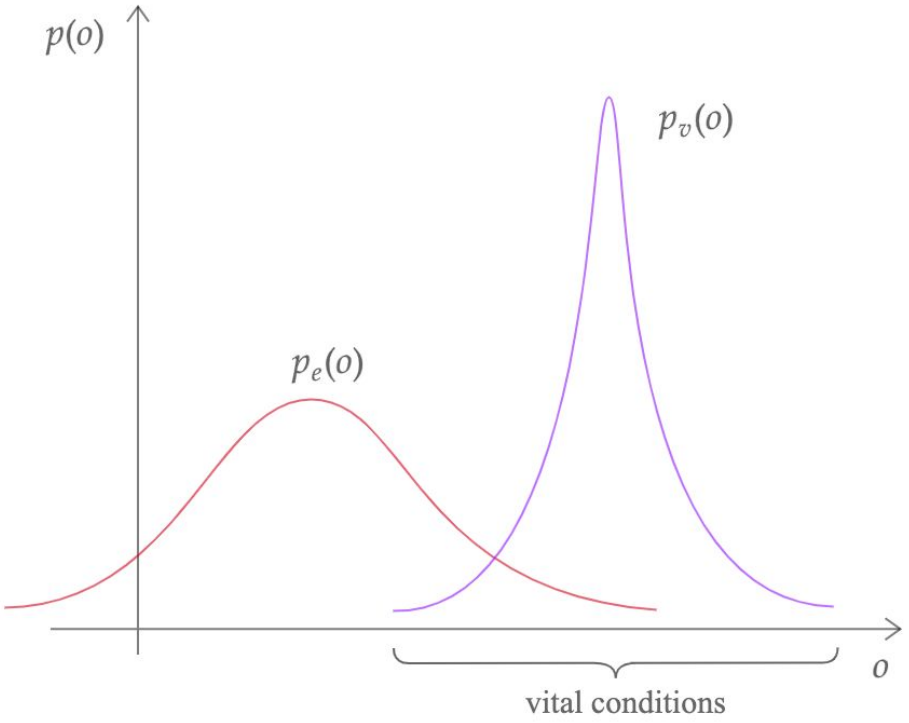
2-3: perception

3-1: environmental changes

Active Inference



Active Inference

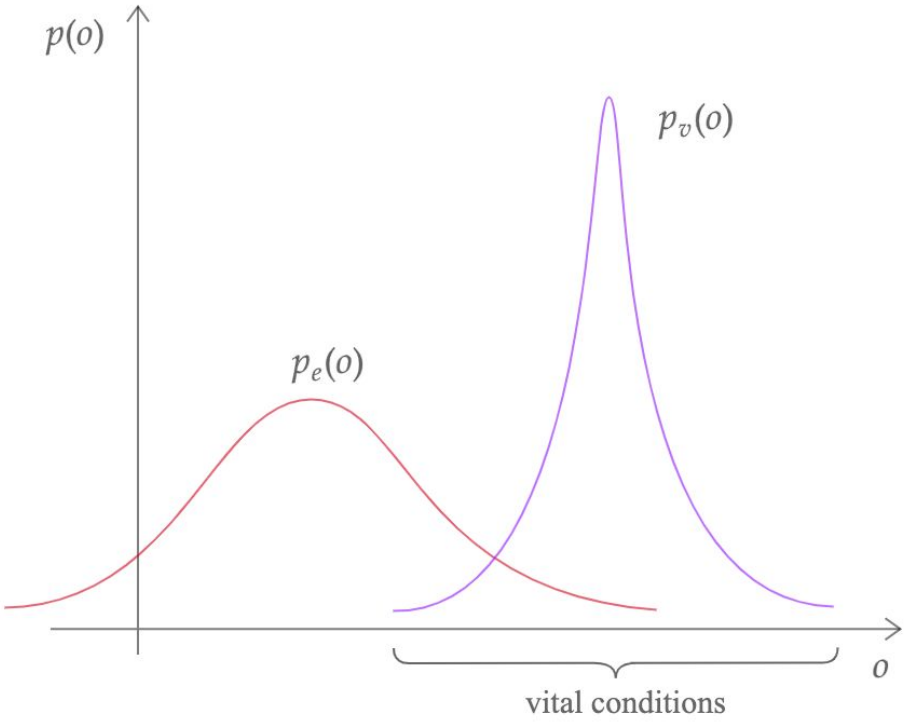


$$p(h, o, a|m) = p_e(h, o, a|m)e^{-G(a)}$$

$$o' = o(a),\ h' = h(a)$$

$$G(a) = \underbrace{-E_{q(o',h'|a)}[\ln p_v(o')p_e(o'|h')]}_{\text{goal}} \underbrace{- H[q(o'|a)]}_{\text{exploration}}$$

Active Inference



$$p(h, o, a|m) = p_e(h, o, a|m)e^{-G(a)}$$

$$o' = o(a),\ h' = h(a)$$

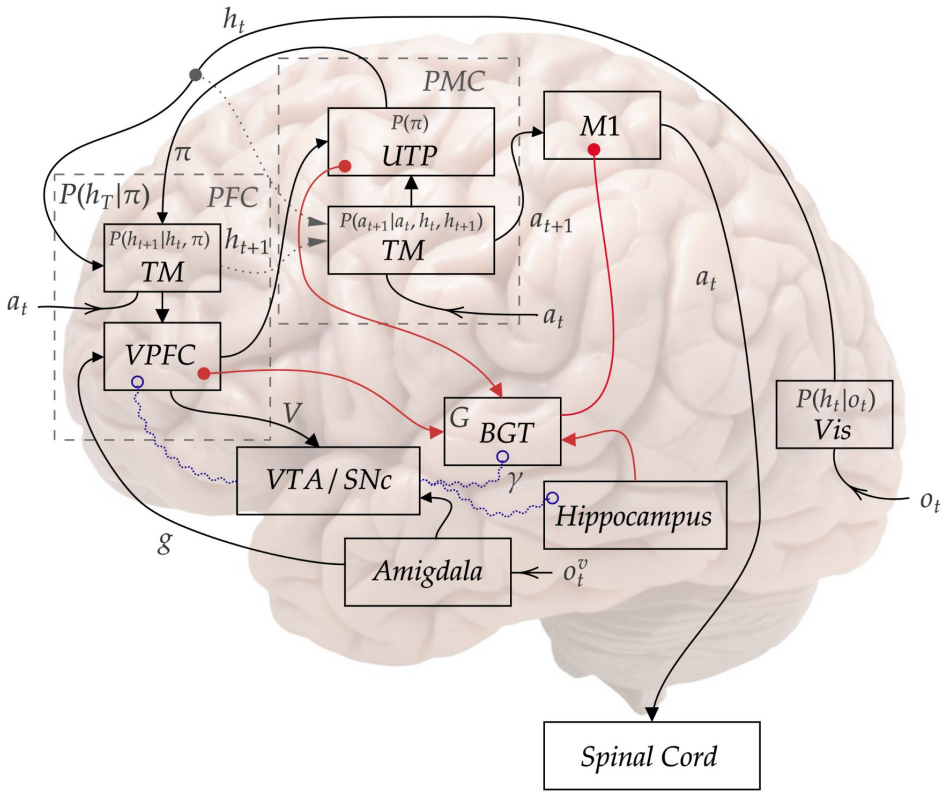
$$G(a) = \underbrace{-E_{q(o',h'|a)}[\ln p_v(o')p_e(o'|h')]}_{\text{goal}} \underbrace{- H[q(o'|a)]}_{\text{exploration}}$$

$$q^*(a) = \underset{q(a)}{\operatorname{argmin}}\, F[q||p]$$

$$q^*(a) \sim e^{-(F_h(a) + G(a))}$$

$$F_h(a) = F[q(o',\ h|\ a)||p_e(o,\ h|a)] - \text{passive inference}$$

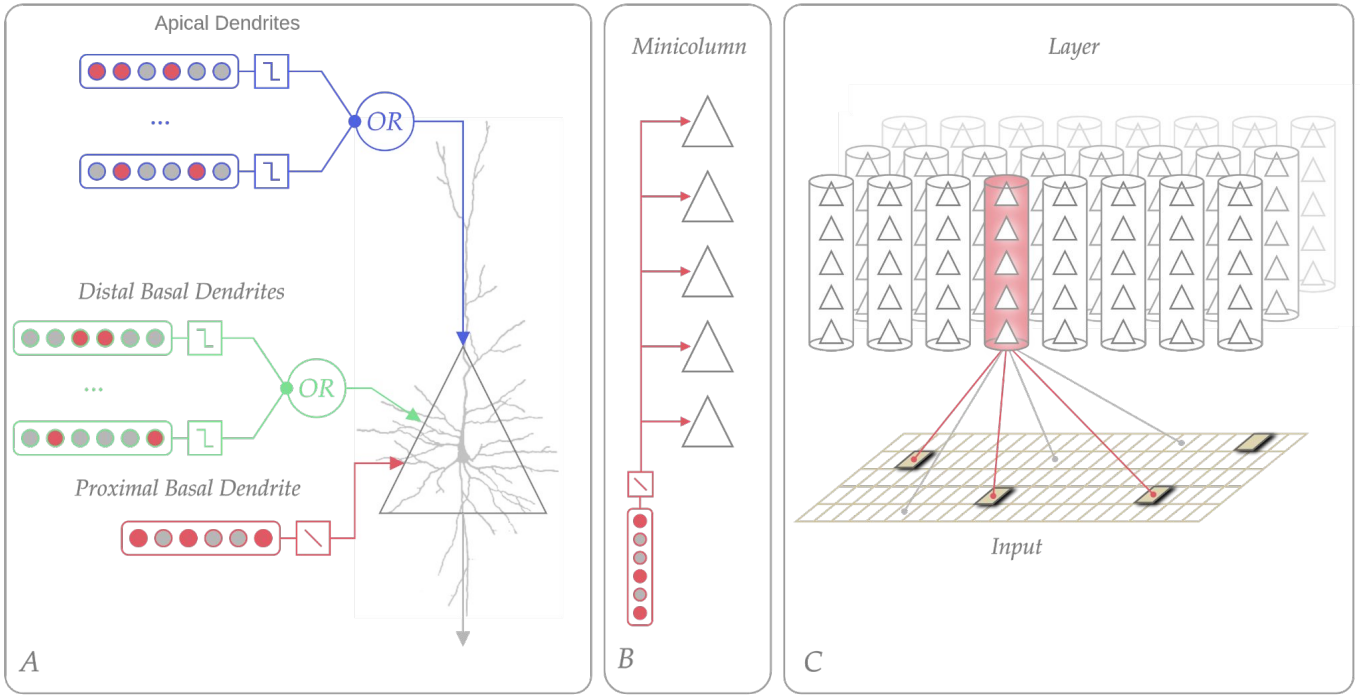
HIMAgent based on Free Energy Principle



Combines two approaches:

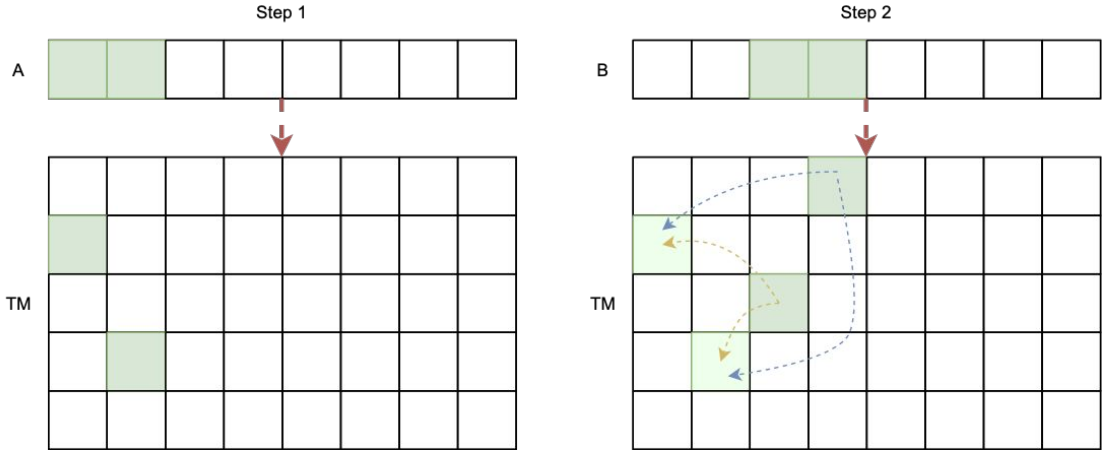
- top-down (Active Inference, psychology)
- bottom-up (neuroscience)

Hierarchical Temporal Memory Framework



Hawkins J, Ahmad S, Cui Y. A Theory of How Columns in the Neocortex Enable Learning the Structure of the World. Front Neural Circuits. 2017 Oct 25;11:81.

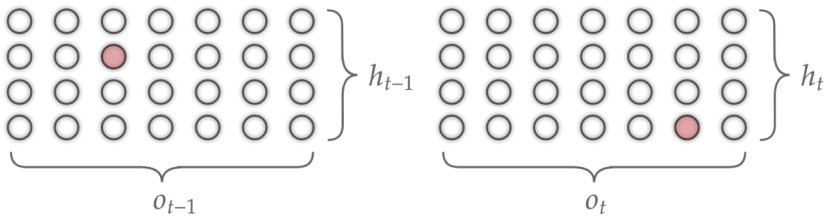
Temporal Memory



Biologically-plausible version of RNN

Utilizes the columnar structure of the neocortex to form hidden states.

Belief TM



$$-\ln p(o_t|h_{t-1}) \leq F[q(h_t)||p(h_t, o_t|h_{t-1})]$$

$$q(h_t) = p(h_t|o_t, h_{t-1})$$

$$p(h_t, o_t|h_{t-1}) = \underbrace{p(h_t|h_{t-1})}_{\text{learning part}} p(o_t|h_t)$$

Expected utility

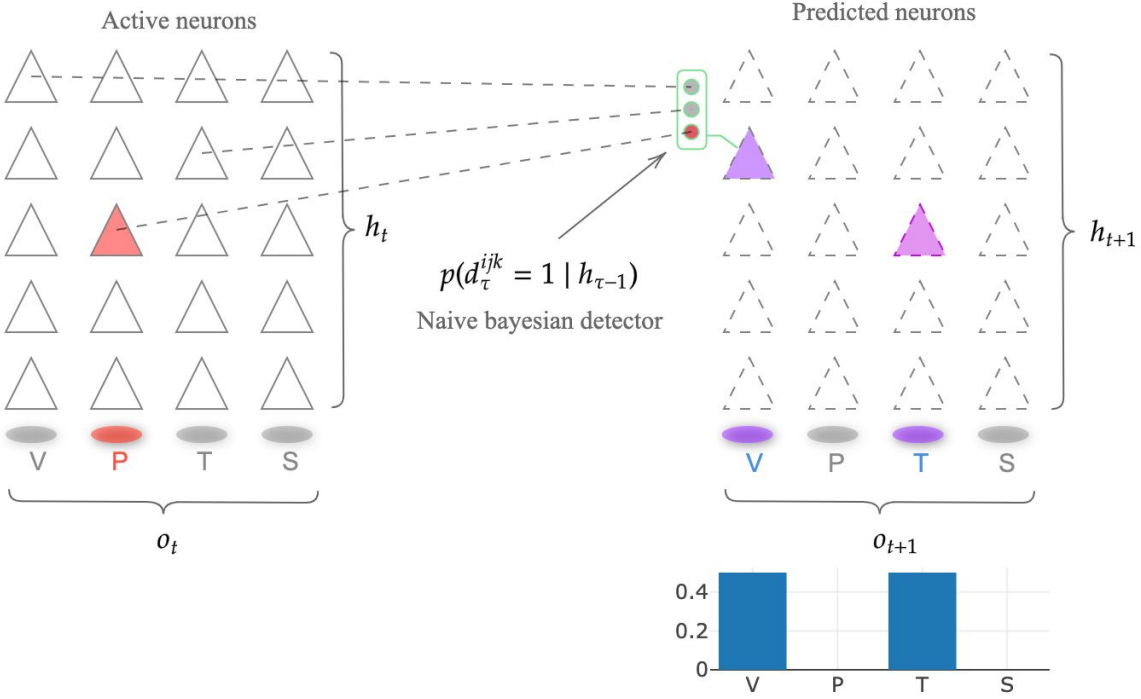
$$\mathcal{U}(\pi) = \sum_{\tilde{o}} q(\tilde{o} \mid \pi) \log p^*(\tilde{o})$$

Fixed strategy

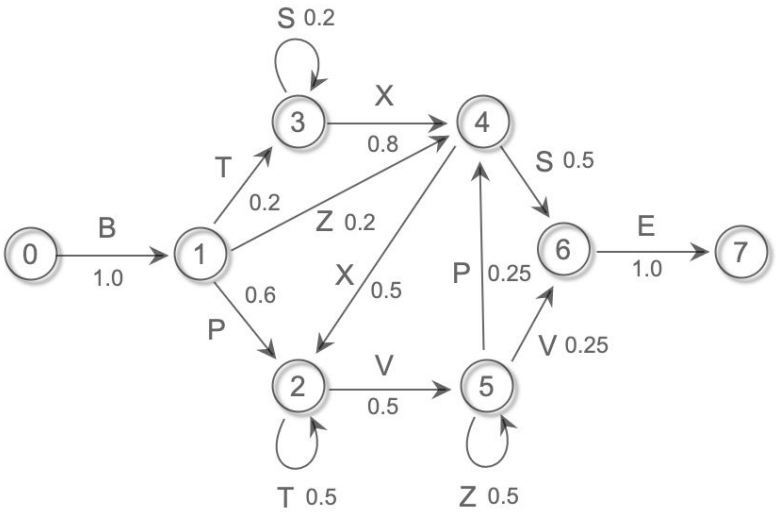
$$q(\tilde{o}) = \sum_{h_{1:T}} q(\tilde{o}, h_{1:T}) = \sum_{h_{1:T}} \prod_{\tau=t+1}^{T} p(o_\tau \mid h_\tau) p(h_\tau \mid h_{\tau-1}) q(h_t)$$

Determined by TM algorithm

learning

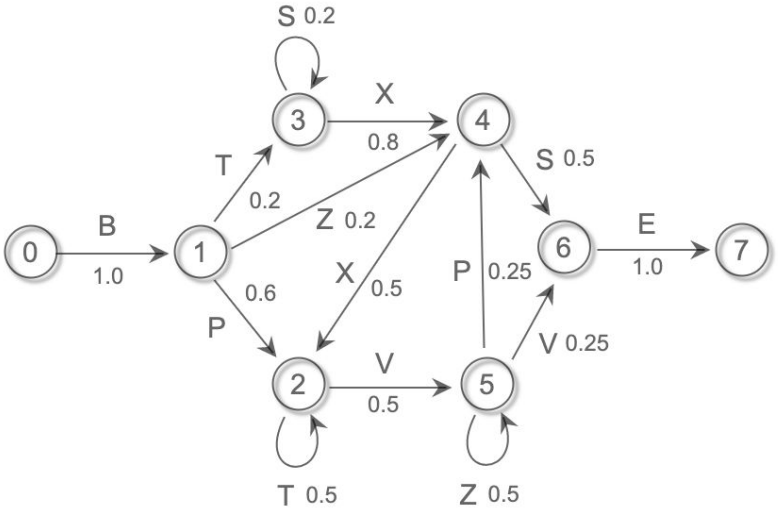


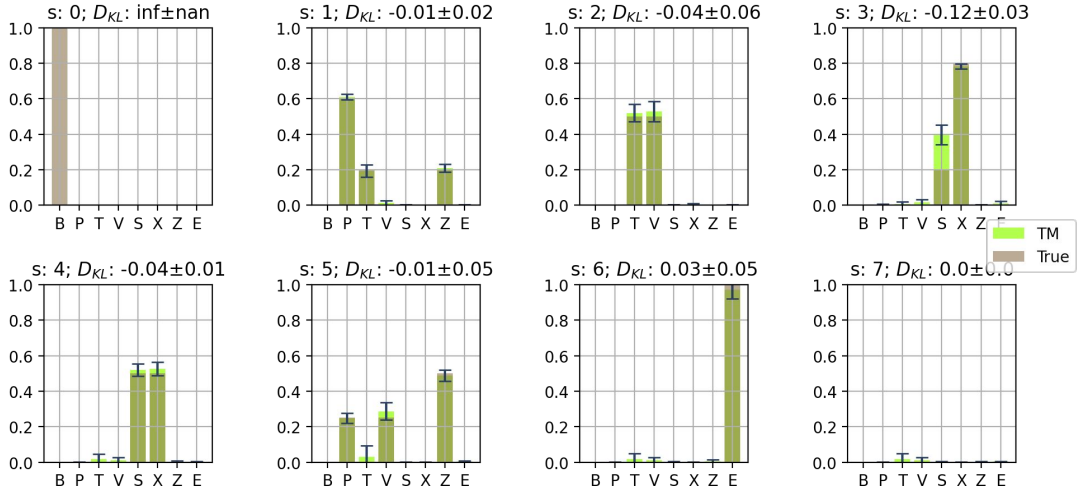
Markov hidden process example



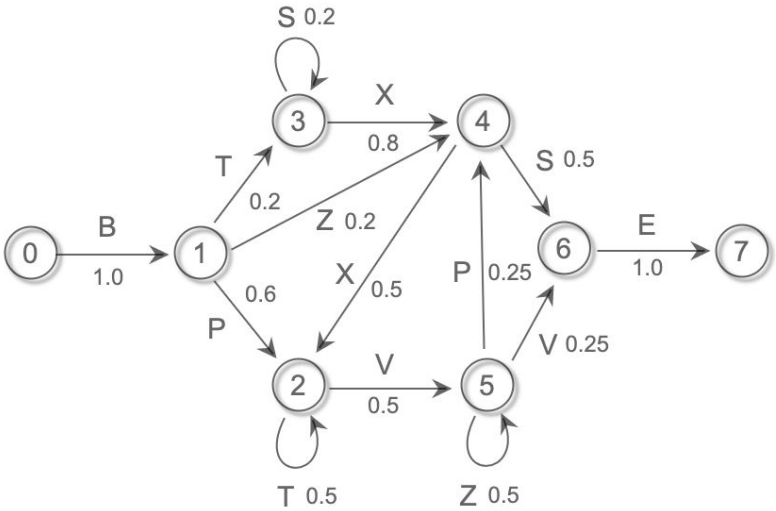
BTXXVPSE
BPVVE
BPTTTVPSE
…

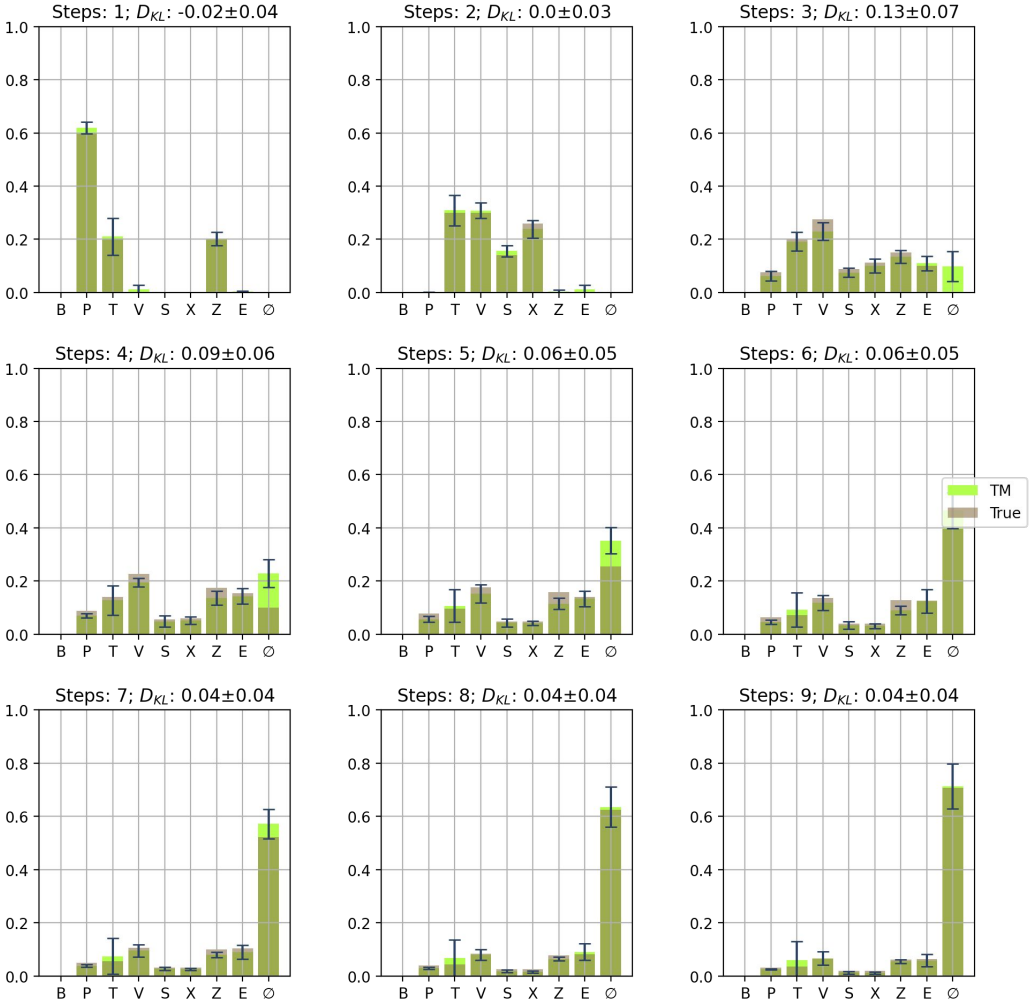
Belief TM


S 0.2
X
0.8
S 0.5
T
0.2
Z 0.2
B
1.0
E
1.0
0.6
X
0.5
P
0.25
P
V
0.5
V 0.25
T 0.5
Z 0.5
0
1
2
3
4
5
6
7

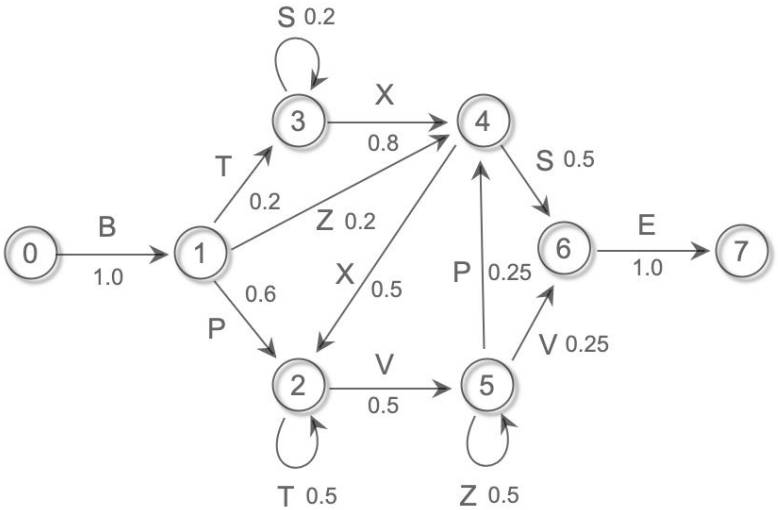

s: 0; D_{KL}: inf±nan
s: 1; D_{KL}: -0.01±0.02
s: 2; D_{KL}: -0.04±0.06
s: 3; D_{KL}: -0.12±0.03
s: 4; D_{KL}: -0.04±0.01
s: 5; D_{KL}: -0.01±0.05
s: 6; D_{KL}: 0.03±0.05
s: 7; D_{KL}: 0.0±0.0
B P T V S X Z E
TM
True

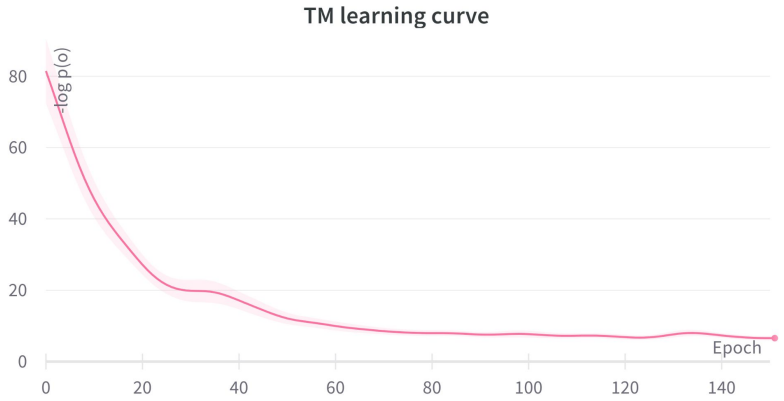
Belief TM

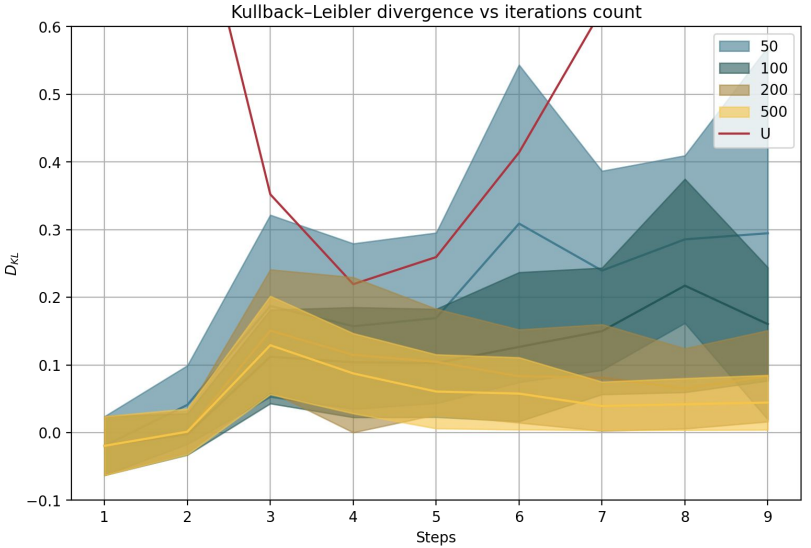

S 0.2
X
0.8
T
0.2
Z 0.2
S 0.5
B
1.0
E
1.0
X
0.5
P 0.25
0.6
P
V
0.5
V 0.25
T 0.5
Z 0.5
0
1
2
3
4
5
6
7


Steps: 1; D_{KL}: -0.02±0.04
Steps: 2; D_{KL}: 0.0±0.03
Steps: 3; D_{KL}: 0.13±0.07
Steps: 4; D_{KL}: 0.09±0.06
Steps: 5; D_{KL}: 0.06±0.05
Steps: 6; D_{KL}: 0.06±0.05
Steps: 7; D_{KL}: 0.04±0.04
Steps: 8; D_{KL}: 0.04±0.04
Steps: 9; D_{KL}: 0.04±0.04
B P T V S X Z E ∅
0.0 0.2 0.4 0.6 0.8 1.0
TM
True

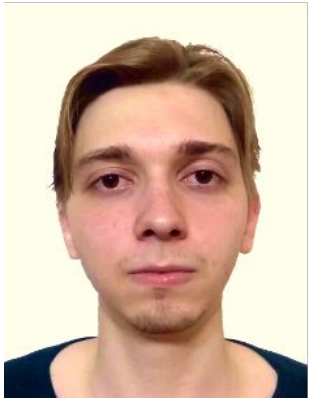
Belief TM


S 0.2
X
0.8
S 0.5
T
0.2
Z 0.2
B
1.0
E
1.0
0.6
X
0.5
P
0.25
P
V
0.5
V 0.25
T 0.5
Z 0.5
0
1
2
3
4
5
6
7


TM learning curve
-log p(o)
80
60
40
20
0
0
20
40
60
80
100
120
140
Epoch


Kullback–Leibler divergence vs iterations count
D_{KL}
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
1
2
3
4
5
6
7
8
9
Steps
50
100
200
500
U

Our team



Evgenij Dzhivelikian

MIPT

Artem Latyshev

MIPT

Petr Kuderov,

AIRI, MIPT

Alexandr I. Panov

AIRI, FRC CSC, MIPT

Thank you

Mail us: kuderov.pv@phystech.edu
Fork us: https://github.com/AIRI-Institute/him-agent