Model-based:

State-transition <T>

Q(s,a) =

Model-free

Q(s,a)

State:

Planet (1,0,0)

Two-step (daw):

unrewarded, rare transition

S0 -> A1 -> S2 -> R=0

Model-free Q(S0, A1) = decrease

Model-based Q(S0, A1) = p(S1|S0, A1) \* V(S1) + p(S2|S0, A1) \* V(S2)

V(S1) = 1- V(S2)

Real Env -> state transition

=

Observed Env ->

Ambiguity

Real Env S0 -> A1 -> S1

Observed Env 50% S0[S0] -> A1 -> S1[S1]

50% S0[S0] -> A1 -> S2[S1]

Variable

P(reward)

P(common) – uncertainty/stochasticity -> kim (90%, 50%)

P(obs = real) – ambiguity

P(switch reward) = 0.025

P(switch common) = 0.025 -> rodent plos comp bio

Research question

Q(MF)

Q(MB)

Q(combined)

Neural network -> MB learning(state-transition based learning)/MF learning,

High State-transition confidence -> MB

Low ST conf -> MF

Network

How is state-transition confidence represented

MB / MF network dynamics

Transition between MB/MF dynamics

Conf -> above

network training (curriculum training)

* Network architecture (LSTM, RNN)
* A2C
* P(reward) = 0.9, p(common) = 0.9
* P(reward) – 0.6~1
* P(common) – 0.6~1
* P(switch common)
* P(obs = real) = 0.9
* P(obs = real) – 0.5
* Confidence safebet
* …

analyze network behavior

* 4 bar
* Plos Comp (5 model)
* P(obs = real) = 0.9 -> MB
* P(obs = real) = 0.5 -> MF

analyze network dynamics

* Dynamic system analysis

Curriculum training

P(reward) = 0.9

P(transition) = 0.9

P(ambiguity) = 0.9

P(reward) = 0.6-1

P(transition) = 0.6-1

P(ambiguity) = 0.6-