

3_sim_exp.py

1. For all bots which are currently at a node & idle

Update:

- true idleness of all nodes in graph (true) = $+\Delta t$
- expected idleness of all edges in the graph (expect) = $+\Delta t$

Calculate: here, learning rate (α)= 0.1, discount factor (γ)= 0.95

- Value function all edges where bots are present (Q) =

$$Q^{new}(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \underbrace{\left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \right)}_{\text{new value (temporal difference target)}}$$

temporal difference

- Reward (r_t) = $\log(|expect - true|)$ and
 $r_t = 0$ if [expect = true]

- Softmax of Value function = value_exp =
$$\frac{e^{Q_i * expect[i]}}{\sum_{j=i}^k e^{Q_j * expect[j]}}$$

 (summation over all edges)

Set:

- True idleness of nodes where bots are present = 0
- Expected idleness of nodes wrt the corresponding bot = 0

NOTE: So effectively, we are calculating expected idleness perceived by the bot as the time elapsed since this bot last visited the node.

This is the **AGENT model** where the bot is calculating expected idleness by seeing the time elapsed since its last visit. Hence there is **no notion of memory** (estimating the expected idleness based on previous visits)

The name 3_sim_exp indicates => sim = simple estimation of expected idleness
 (AGENT model)

exp = including the expected idleness into the
 softmax calculation

2. For a bot deciding the next node to visit

Set:

- True idleness of the node where the bot is present = 0
- Expected idleness of the node wrt the corresponding bot = 0

Decision Making: here, we chose $\epsilon=0.1$

- With $(1 - \epsilon)$ probability, check all neighbours and visit the one with highest value of = $[value_exp]$
- With ϵ probability, go to a random node

-----END-----

DRAWBACK:

- Even when $|expect - true| = 1$, reward = 0.