11_ses_exp.py

1. <u>For all bots which are currently at a node & idle</u>

Update:

- true idleness of all nodes in graph (true) = +Δt
- Store all the true idleness values at this time stamp
- Expected idleness of nodes at which bots are currently present is calculated by performing simple exponential smoothing (ses) of true idleness for a particular edge (now, expected idleness is function of edge not node)
- We have chosen last 5 values at any instant and α = 0. 4 for SES.

<u>Calculate</u>: here, learning rate (α)= 0.1, discount factor (γ)= 0.95

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

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

- Reward (r_t)= $log(\,|expect \;-\; true|\,)$and

  $r_t = 0$ if [expect = true]

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

  (summation over all edges)

Set:

- True idleness of nodes where bots are present = 0

OBSERVATION model: bot will calculate the expected idleness as an average of all the past true idleness it has seen when it last visited the node while travelling **along that particular edge**.

The name 11_ses_exp indicates => ses = SES to estimate expected idleness (OBSERVATION 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

Decision Making: here, we chose ε=0.1

- With $(1 - \varepsilon)$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.