Transparency of Task Dependencies in Enviroment of Reinforcement Learning

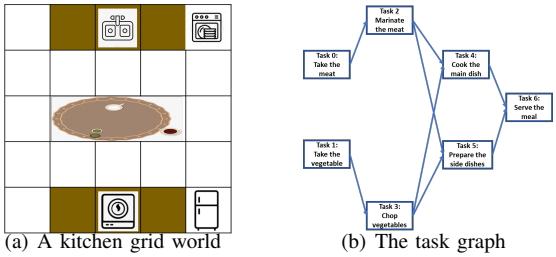
I. Introduction

II. Preliminaries and Notation

Markov decision process (MDP) is a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R})$ where $\mathcal{S}$ indicates a set of states, $\mathcal{A}$ indicates a set of actions, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ indicates a transition function, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ indicates a reward function and $\pi : \mathcal{S} \to \mathcal{A}$ indicates a policy. Markov process can be denoted as a tuple ($\mathcal{S}$, $\mathcal{T}$).

Definition II.1. *A network $(\mathcal{V}, \mathcal{E})$ is a directed acyclic graph (DAG) where $\mathcal{V} = \{v_1, v_2, ..., v_m\} \subset \mathcal{S}$ is a finite set of nodes and $\mathcal{E} = \{(v_i, v_j) | i, j = 1, 2, ..., m\}$ is a set of directed edges over $\mathcal{V}$.*

In this paper, we use nodes to represent tasks in an RL environment, and we call such a task network or a task graph. As an example scenario we use Figure 1 to illustrate how tasks can be interdependent. In this scenario, an agent (or robot) can carry out 7 tasks, such as *take the meat out* (Figure 1(b)), *cook the main dish*, etc. In order to do so, the agent first needs to travel to some specific locations, such as the fridge's location (see Figure 1(a)) or the sink's location (for washing vegs).



(a) A kitchen grid world (b) The task graph

Fig. 1. An illustration scenario for tasks in RL

A task network (V, E) may have several initial task nodes, which are tasks that do not dependent on any other tasks, such as *take the meat out* or *take the vegetable*. A task can only be executed when all of its predecessors have been completed. For example, in the task graph in Figure 1, both task 0 and task 1 are initial task nodes and either task can be executed as the first task. Nodes with no predecessor tasks are referred to as root nodes, such as task 0 and task 1, and nodes with no successor tasks are referred to as leaf nodes, such as task 6. Further details of the scenario can be found in the appendix **??**

Definition II.2. *Let $h = (s_1, a_1, \ldots, s_t, a_t) \subset \mathcal{H}$ be a trajectory of a policy. Let $H = [h_1, h_2, \cdots, h_n]^T$ be an $1 \times n$ metrix, containing n trajectories, where $[\cdot]^T$ indicates matrix transposition.*

Definition II.3. *Let O be a task in $\mathcal{V}$ and h be a policy. O is said to be complete in h for timesteps $[i, \cdots, t]$ if either O has no predecessor tasks or every predecessor task of O has been complete before timestep i.*

Definition II.4. *Let $(s_{t_i}, a_{t_i}, s_{t_i+1})$ (resp. (s_{t_i}, s_{t_i+1})) be the completion marker for task i in a trajectory of MDP (resp. MP). If $(s_{t_i}, a_{t_i}, s_{t_i+1})$ exists in trajectory h (resp. h'), then this indicates that task i has completed at time t_t in h (resp. h'), where the only difference between h and h' is that h' does not explicitly specific action a_t.*

Note that the marker for task i completion may not be unique.

Definition II.5. *Let M_o^a be a after-relationship vector for task O where the elements in M_o^a denote all the tasks that can be executed after O is complete. Let M^a be the matrix for all the m tasks' after-relationship vectors, then M^a is represented as:*

$$M^a = [M_1^a, M_2^a, \cdots, M_m^a]^T = \begin{bmatrix} M_{11}^a & M_{12}^a & \cdots & M_{1m}^a \\ M_{21}^a & M_{22}^a & \cdots & M_{2m}^a \\ \vdots & \vdots & \ddots & \vdots \\ M_{m1}^a & M_{m2}^a & \cdots & M_{mm}^a \end{bmatrix}$$

As an example, the after-relationship vector for each task in Figure 1 is shown below. In a after-relationship vector such as for task O_0, value 1 in the vector indicates that that corresponding task shall be executed after task O_0. Note, we have given each task an order number, as shown in Figure 1, such as task 0 is for 0, task 1 is for 1, task 3 is for 3 $\cdots$, and all the after-relationship has the same ordering.
After-relationship vector of each task is listed as follows:
Task O_0: the after vector M_0^a:[0,1,1,1,1,1,1], the after set X_0^a:{1,2,3,4,5,6}
Task O_1: the after vector M_1^a:[1,0,1,1,1,1,1], the after set X_1^a:{0,2,3,4,5,6}
Task O_2: the after vector M_2^a:[0,1,0,1,1,1,1], the after set X_2^a:{1,3,4,5,6}
Task O_3: the after vector M_3^a:[1,0,1,0,1,1,1], the after set X_3^a:{0,2,4,5,6}
Task O_4: the after vector M_4^a:[0,0,0,0,0,1,1], the after set X_4^a:{5,6}
Task O_5: the after vector M_5^a:[0,0,0,0,1,0,1], the after set X_5^a:{4,6}
Task O_6: the after vector M_6^a:[0,0,0,0,0,0,0], the after set

X_6^a:{ }
The after matrix of task graph is as follows:

$$M^a = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Definition II.6. *Let M_o^b be a before-relationship vector for task O where the elements in M_o^b denote all the tasks that can be executed before O is complete. Let M^b be the matrix for all the m tasks' before-relationship vectors, then M^b is represented as:*

$$M^b = [M_1^b, M_2^b, \cdots, M_m^b]^T = \begin{bmatrix} M_{11}^b & M_{12}^b & \cdots & M_{1m}^b \\ M_{21}^b & M_{22}^b & \cdots & M_{2m}^b \\ \vdots & \vdots & \ddots & \vdots \\ M_{m1}^b & M_{m2}^b & \cdots & M_{mm}^b \end{bmatrix}$$

As an example, the before-relationship vector for each task in Figure 1 is shown below. In a before-relationship vector such as for task O_0, value 1 in the vector indicates that that corresponding task shall be executed before task O_0. Here, we follow the same order numbers for tasks as before, such as task 0 is for $0 \cdots$.
Before vector of each task is listed as follows:
Task O_0: the before vector M_0^b:[0,1,0,1,0,0,0], the before set X_0^b:{1,3}
Task O_1: the before vector M_1^b:[1,0,1,0,0,0,0], the before set X_1^b:{0,2}
Task O_2: the before vector M_2^b:[1,1,0,1,0,0,0], the before set X_2^b:{0,1,3}
Task O_3: the before vector M_3^b:[1,1,1,0,0,0,0], the before set X_3^b:{0,1,2}
Task O_4: the before vector M_4^b:[1,1,1,1,0,1,0], the before set X_4^b:{0,1,2,3,5}
Task O_5: the before vector M_5^b:[1,1,1,1,1,0,0], the before set X_5^b:{0,1,2,3,4}
Task O_6: the before vector M_6^b:[1,1,1,1,1,1,0], the before set X_6^b:{0,1,2,3,4,5}
The before matrix of task graph is as follows:

$$M^b = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Definition II.7. *Let M_o^u be a free-relationship vector for task O where the elements in M_o^u denote all the tasks that can be executed either after or before O is complete. Let M^u be the matrix for all the m tasks' free-relationship vectors, then M^u is represented as:*

$$M^u = [M_1^u, M_2^u, \cdots, M_m^u]^T = \begin{bmatrix} M_{11}^u & M_{12}^u & \cdots & M_{1m}^u \\ M_{21}^u & M_{22}^u & \cdots & M_{2m}^u \\ \vdots & \vdots & \ddots & \vdots \\ M_{m1}^u & M_{m2}^u & \cdots & M_{mm}^u \end{bmatrix}$$

As an example, the free-relationship vector for each task in Figure 1 is shown below. In a free-relationship vector such as for task O_0, value 1 in the vector indicates that that corresponding task shall be executed after and before task O_0.
Free vector of each task is listed as follows:
Task O_0: the free vector M_0^u:[0,1,0,1,0,0,0]
Task O_1: the free vector M_1^u:[1,0,1,0,0,0,0]
Task O_2: the free vector M_2^u:[0,1,0,1,0,0,0]
Task O_3: the free vector M_3^u:[1,0,1,0,0,0,0]
Task O_4: the free vector M_4^u:[0,0,0,0,0,1,0]
Task O_5: the free vector M_5^u:[0,0,0,0,1,0,0]
Task O_6: the free vector M_6^u:[0,0,0,0,0,0,0]

The free matrix of task graph is as follows:

$$M^u = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Note: The free vector of task i M_i^u is the form of the intersection of before set X_i^b and after set X_i^a of task i.

Definition II.8. *Let $\overline{M_o^b}$ be a predecessors-relationship vector for task O where the elements in $\overline{M_o^b}$ denote all the tasks that can **only** be executed before O is complete. Let $\overline{M^b}$ be the matrix for all the m tasks' predecessors-relationship vectors.*

As an example, the predecessors-relationship vector for each task in Figure 1 is shown below. In a predecessors-relationship vector such as for task O_0, value 1 in the vector indicates that that corresponding task shall be executed before task O_0.
Predecessors vector of each task is listed as follows:
Task O_0: the predecessors vector $\overline{M_0^b}$:[0,0,0,0,0,0,0]
Task O_1: the predecessors vector $\overline{M_1^b}$:[0,0,0,0,0,0,0]
Task O_2: the predecessors vector $\overline{M_2^b}$:[1,0,0,0,0,0,0]
Task O_3: the predecessors vector $\overline{M_3^b}$:[0,1,0,0,0,0,0]
Task O_4: the predecessors vector $\overline{M_4^b}$:[1,1,1,1,0,0,0]
Task O_5: the predecessors vector $\overline{M_5^b}$:[1,1,1,1,0,0,0]
Task O_6: the predecessors vector $\overline{M_6^b}$:[1,1,1,1,1,1,0]
The predecessors matrix of task graph is as follows:

$$\overline{M^b} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Note: predecessors set and before set are different relationships. The former indicates all tasks that are preceding this task, i.e., must be executed before this task, and the latter indicates all tasks that can appear before this task, i.e., can be executed before this task. If a task cannot be executed after this task i, then this task can only be executed before this task, i.e. this task is one of the predecessors of task i.

Definition II.9. *Let $\overline{M_o^a}$ be a successor-relationship vector for task O where the elements in $\overline{M_o^a}$ denote all the tasks that can **only** be executed after O is complete. Let $\overline{M^a}$ be the matrix for all the m tasks' successor-relationship vectors.*

As an example, the successor-relationship vector for each task in Figure 1 is shown below. In a successor-relationship vector such as for task O_0, value 1 in the vector indicates that that corresponding task shall be executed after task O_0.
Successor vector of each task is listed as follows:
Task O_0: the successor vector $\overline{M_0^a}$:[0,0,1,0,1,1,1]
Task O_1: the successor vector $\overline{M_1^a}$:[0,0,0,1,1,1,1]
Task O_2: the successor vector $\overline{M_2^a}$:[0,0,0,0,1,1,1]
Task O_3: the successor vector $\overline{M_3^a}$:[0,0,0,0,1,1,1]
Task O_4: the successor vector $\overline{M_4^a}$:[0,0,0,0,0,0,1]
Task O_5: the successor vector $\overline{M_5^a}$:[0,0,0,0,0,0,1]
Task O_6: the successor vector $\overline{M_6^a}$:[0,0,0,0,0,0,0]
The successor matrix of task graph is as follows:

$$\overline{M^a} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$