State Abstraction in MAXQ
Motivation

- State Abstraction helps limit the size of the problem
  - If we can represent a large number of identical states as one state, it keeps the value function simpler
- When can we safely abstract states into one representative state?
  - What group of features $Y$ is unnecessary to consider in a task or subtask?
Conditions for Safe State Abstraction

- Subtask Irrelevance
- Leaf Irrelevance
- Result Distribution Irrelevance
- Termination
- Shielding
Subtask Irrelevance

- A set of features $Y$ are irrelevant for a subtask if the probability distribution for any next state is independent of $Y$.
- In the taxi example, source and destination of the passenger are irrelevant; only the target $t$ and current position matter.
- Formally:
  - We partition the features of MDP $M$ into $X$ and $Y$ and consider $M_i$, an MDP subtask of $M$.
  - If 2 properties hold for any stationary abstract hierarchical policy $\pi$ on $M_i$, the features in $Y$ are irrelevant to subtask $i$:  
    - $P^{\pi}(s', N | s, j) = P^{\pi}(x', y', N | x, y, j) = P^{\pi}(x', N | x, j) \cdot P^{\pi}(y', N | y, j)$
    - For any pair of states $s_1 = (x, y_1)$ and $s_2 = (x, y_2)$ and any action $j$ in $M_i$, $V^{\pi}(j, s_1) = V^{\pi}(j, s_2)$.
Leaf Irrelevance

- The set $Y$ is irrelevant for a given primitive action $a$ if any two states differing only in $Y$ have the same expected reward.

- Example: this is true for all features for the primitive movement actions (North, South, East, West) in the taxi domain.
  - Since there is no immediate reward for any of them except the -1 movement, and we incur that no matter which way we move.

- Formally:
  - $Y$ is irrelevant for a primitive action $a$ if, given $s_1=(x,y_1)$ and $s_2=(x,y_2)$:
Result Distribution Irrelevance

- If a feature has no effect on the end state and reward of a subtask, it can be ignored for the purposes of the subtask.

- One example is the **Get** option, which ends in the same location no matter where it started.
  - This is only true in the undiscounted case.

- Formally:
  - $Y$ is irrelevant for the result distribution of action $j$ if, for all abstract policies executed on $M_i$, given $s_1=(x,y_1)$ and $s_2=(x,y_2)$:
    \[
    \forall s' \ P^{\pi}(s'|s_1,j) = P^{\pi}(s'|s_2,j)
    \]
Termination

- If $M_j$ is a suboption of $M_i$ such that whenever $M_j$ terminates, $M_i$ also terminates, then $C(i, s, j) = 0$, and so does not need to be represented.
  - This “funnels” large numbers of states into terminal states for $M_i$.

- The **Put** function will always succeed and terminate the episode from states where the taxi has picked up the passenger.
Shielding

- Given a subtask $M_i$ and a state $s$ such that all paths in the directed acyclic graph from the root to $M_i$ include a subtask that is terminated, then no $C$ values need to be represented for $M_i$, as they are states that do not make sense.

- The **Put** option does not need any $C$ values represented for states where the passenger is not in the taxi, as **Put** terminates if the passenger has not been picked up yet.