SARSA

SARSA is a reinforcement learning technique used for decision-making in dynamic and uncertain environments, similar to Q-learning. In SARSA, an agent learns a policy by interacting with the environment and updates its knowledge based on the state-action-reward-state-action sequence. Unlike Q-learning, which is an off-policy method, SARSA is an on-policy algorithm, meaning the action used for updating the Q-value is based on the current policy.

In SARSA, the agent maintains a Q-table (or Q-function) that stores Q-values representing the expected cumulative rewards for taking specific actions in given states. SARSA has been effectively applied to a range of decision-making problems, including autonomous navigation, game playing, and task allocation.

In the context of task scheduling, SARSA can learn an optimal policy for assigning tasks to computing resources by taking into account the current system state, task characteristics, and environmental factors. The Q-table in SARSA also utilizes a state-action pair to index a Q-value, denoted as Q(s,a), where s is the state and a is the action. The Q-table is dynamically updated depending on the reward or punishment associated with a given state-action pair.

The Q-value update rule for SARSA is given as:

Q(t+1)(st,at)=Q(st,at)+α(rt+γQ(st+1,at+1)−Q(st,at))

where:

γ is the discount factor within [0,1],

rt is the reward or punishment,

α is the learning rate within [0,1], calculated as:

α=1−0.9×t/epochs​

The value of the parameter α is crucial in determining whether the agent explores or exploits:

If α is close to 1, recently acquired data is given more importance, leading to more exploration.

If α is closer to 0, the agent focuses on exploiting the current knowledge.

The discount factor γ influences how much importance is given to future rewards. In this case, γ is set to 0, meaning the agent prioritizes immediate rewards.

The reward rt is defined as:

rt=1,if the current action improves the solution

rt=−1,otherwise

This framework allows SARSA to dynamically learn from interactions with the environment and progressively refine its decision-making policy over time.