

DETERMINING ACTION SELECTION POLICIES OF AN EXECUTION DEVICE

ABSTRACT

Disclosed herein are methods, systems, and apparatus for generating an action selection policy for a software-implemented application that performs actions in an environment that includes an execution device supported by the application and one or more other devices. One method includes, for each action among possible actions in a state of the execution device in a current iteration, obtaining a regret value of the action in the state of the execution device in a previous iteration; and computing a parameterized regret value of the action in the state of the execution device in the previous iteration; determining a respective normalized regret value for each of the possible actions in the previous iteration; determining, from the normalized regret values, an action selection policy of the action in the state of the execution device; and controlling operations of the execution device according to the action selection policy.

FIG. 5