|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm A3C** – each learner thread | | | |
| *// Assume global shared parameters and and global episode counter* | | | |
| *// Assume thread-specific parameter vectors and* | | | |
| **while** | | | |
|  |  | | |
|  | Synchronize thread-specific parameters and | | |
|  |  | | |
|  | Get state | | |
|  | **while** not terminal and | | |
|  |  | Perform according to policy | |
|  |  | Receive reward and new state | |
|  |  |  | |
|  |  | | for terminal |
|  | for non-terminal |
|  | **for** | | |
|  |  |  | |
|  |  |  | |
|  |  |  | |
|  | Perform asynchronous update of using and of using | | |
|  | Reset gradients: and | | |