

"actor"

policy improvement

control
policy

environment

$\{V(x)\}$

"critic"

policy evaluation

learning
feedback

