REINFORCEMENT LEARNING
WPROWADZENIE

BY TOOPLOOX AI

Jeremi Kaczmarczyk
Rafał Nowak
Piotr Semberecki

REINFORCEMENT LEARNING

“Reinforcement Learning - jest to uczenie co zrobić - jak dopasować sytuacje do akcji aby zmaksymalizować numeryczny sygnał nagrody
- Reinforcement Learning: An Introduction 2nd ed”

STATE - STAN

$$s, s', s_t$$

Stan określamy symbolem s i jest to obecna sytuacja w jakiej znajduje się środowisko. Jako s' oznaczamy stan będący rezultatem stanu s, natomiast s_t jest to stan dla danego kroku.

ACTION - AKCJA

$$a, a_t$$

Znajdując się w stanie s możemy wykonać akcję a.
Akcja powoduję zmianę stanu z s do s'.

REWARD - NAGRODA

$$r, r_t$$

Po wykonaniu akcji otrzymujemy nagrodę r od środowiska. Nagrodę otrzymujemy po każdym kroku i niekoniecznie jest pozytywna. Projektowanie sygnału nagrody ma kluczowe znaczenie przy rozwiązywaniu problemów Reinforcement Learningiem.

POLICY - POLITYKA

$$\pi, \pi(s), \pi(a|s)$$

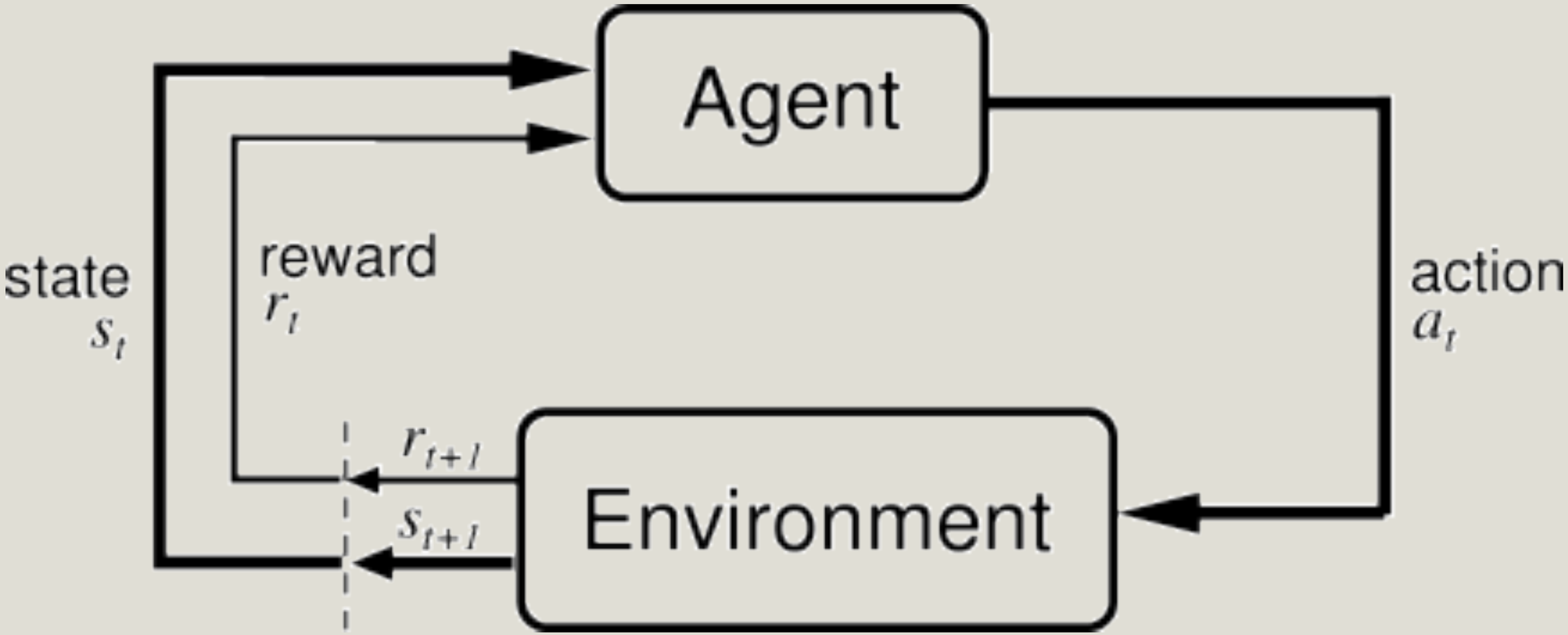
Polityka definiuje zachowanie w danym momencie. Jest to funkcja parametryzowana zwykle stanem lub parą stan-akcja. W przypadku prostych problemów jest to zwykle słownik. Zwraca akcję którą agent powinien wykonać w danym stanie albo prawdopodobieństwa wykonania każdej z akcji.

VALUE - WARTOŚĆ

$$v_\pi(s), q_\pi(s, a)$$

Wartość określana jest w stosunku do danej polityki π. Określana jako v jeśli jest to wartość dla stanu, albo q jeśli dla pary stan-akcja. Jest to numeryczna wartość określająca jak dobrze jest być w danym stanie albo inaczej jaka jest średnia skumulowana nagroda możliwa w danym stanie lub dla danej pary stan-akcja.

AGENT - ŚRODOWISKO



TAXI

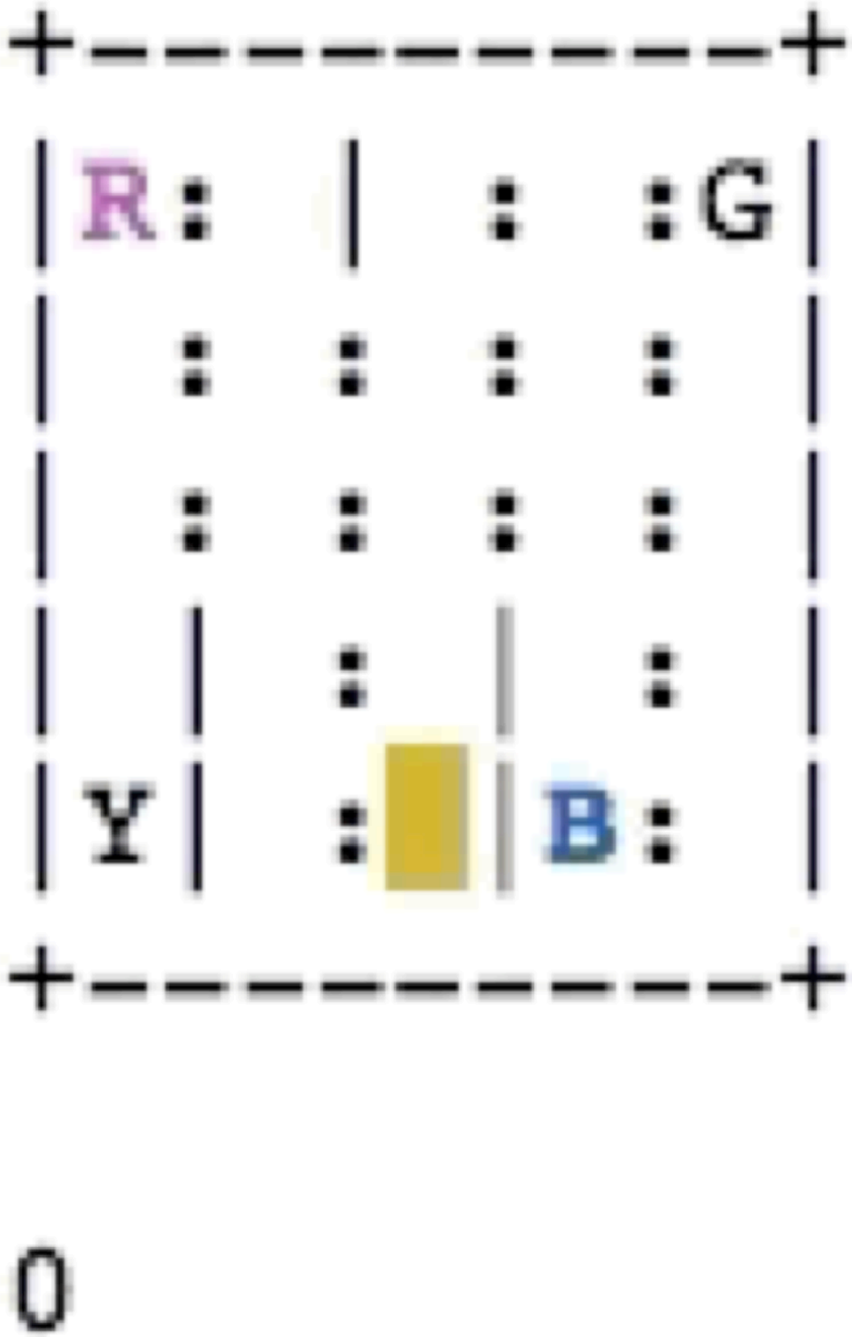
» : możemy przejść, przez | nie

» R, G, Y, B miejsca podnoszenia / zostawienia pasażerów

» +20 nagrody za sukces

» -10 nagrody za nielegalne podniesienie / zostawienie

» -1 nagrody za każdy ruch



0

MARKOV DECISION PROCESS - MDP