Úvod do učenia posilňovaním

(Strojové učenie II)

M. Mach

Katedra kybernetiky a umelej inteligencie, FEI, TUKE

január 2021 - február 2023

Súčasti strojového učenia

Charakteristika učenia posilňovaním

- Cieľom je naučiť sa vhodne reagovať na prostredie s úmyslom dosiahnuť nejaký cieľ
- Učenie prebieha na základe interakcií s prostredím
  - interakcia prebieha súčasne s učením
  - agent je ponechaný sám na seba – použitý pokus-omyl princíp, učenie z úspechov aj neúspechov
  - agentove akcie ovplyvnia budúce agentove možnosti
- Spätná väzba – vo forme odmeny/pokuty
  - zriedkavá/častá – (nie) je generovaná po každej akcii
  - oneskorená/okamžitá – je relevantná (nielen) pre poslednú akciu (ale pre sekvenciu akcií)
    - nie je daná ne/správnosť jednotlivých akcií

CIT Centrum inteligentných technológií

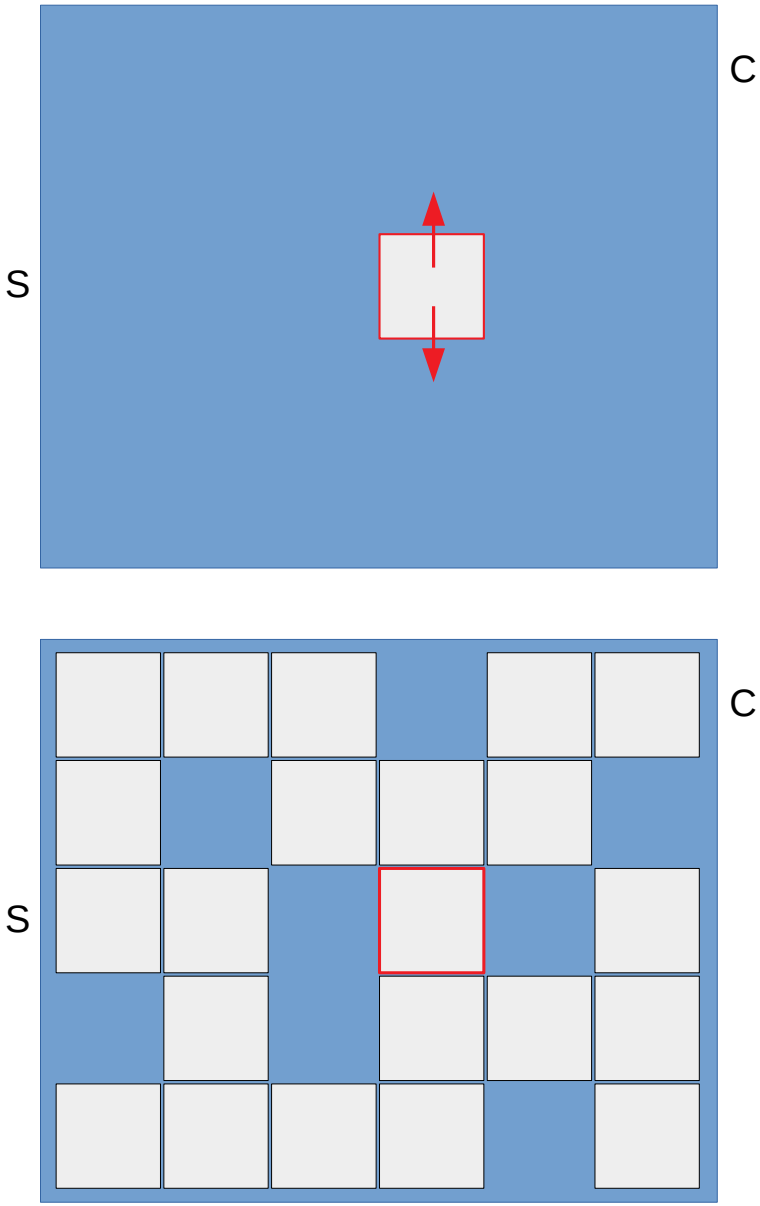
Príklady učenia posilňovaním

- Manévry leteckej akrobacie
- Hranie hier
  - doskové hry
  - počítačové hry
- Činnosť robota
- Správa investičného portfólia
- Liečba pacienta
- Autonómne vozidlo

Elementy učenia posilňovaním

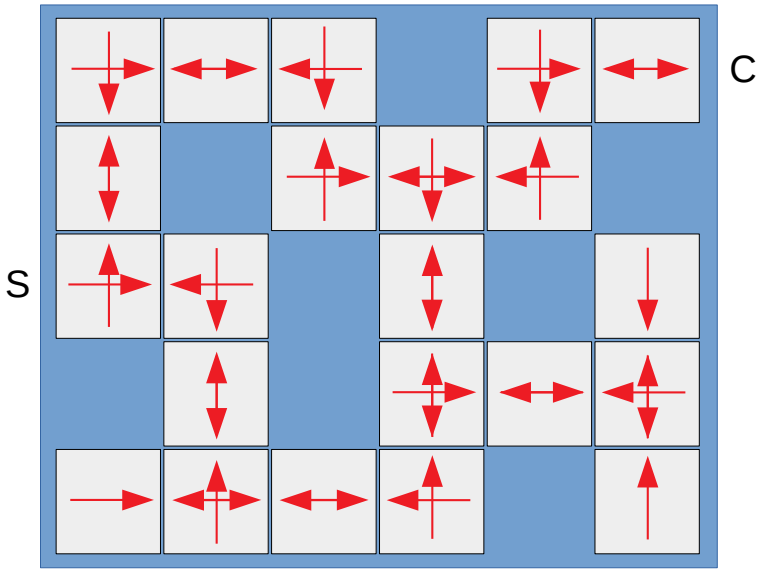
- Stav
  - informácia o aktuálnom statuse prostredia, je funkciou histórie zmien prostredia
- Akcia a prechod
  - informácia o možnej zmene prostredia
- Politika (voľba akcií)
  - voľba zásahu agenta do prostredia (ovplyvňovanie prostredia prostredníctvom zvolenej akcie)
- Odmena (spätná väzba)
  - definuje cieľ (zisk čo najväčšej kumulatívnej odmeny musí byť ekvivalentný splneniu cieľa)
- Model - nepovinné
  - umožňuje predikciu chovania prostredia ako reakciu na možné akcie agenta

Stav



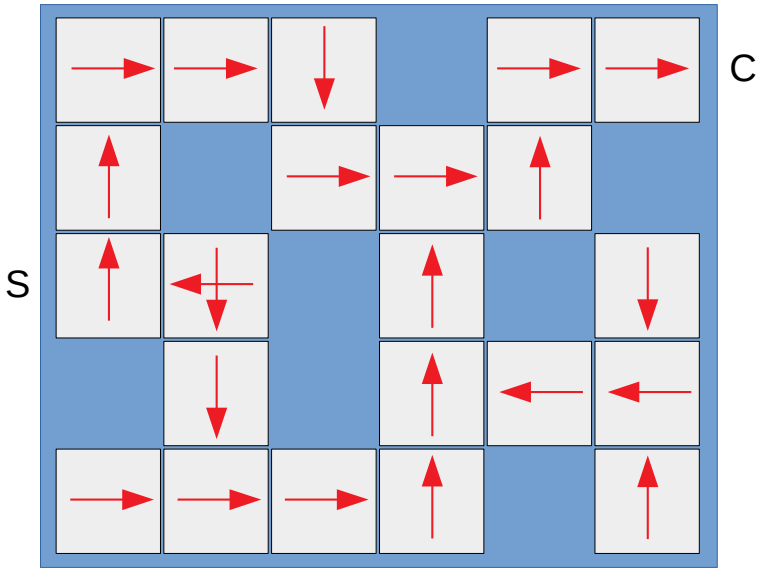
- Typy stavov
  - Stav prostredia S_t^e
  - Agentov stav prostredia S_t^a
  - Pozorovanie prostredia O_t
- Plná pozorovateľnosť
  - agent priamo pozoruje stav prostredia $O_t = S_t^e = S_t^a$
- Čiastočná pozorovateľnosť
  - $S_t^e \neq S_t^a$
  - agent musí konštruovať S_t^a z histórie pozorovaní
  - $S_t^a = (P[S_t^e = s_1], \ldots, P[S_t^e = s_n])$

Akcie



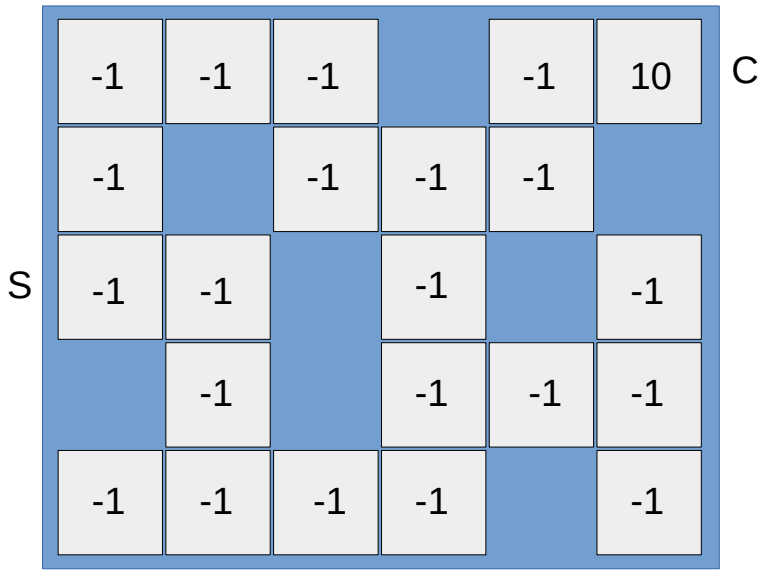
- Typy akcií
  - diskrétne / spojité
  - fyzické / mentálne
  - nízkoúrovňové / vysokoúrovňové
- Dostupnosť
  - všetky v každom stave
  - rôzne podmnožiny v rôznych stavoch
- Následok akcie
  - prechod zo stavu do stavu (aj zotrvanie) $T : S \times A \rightarrow S$
  - prechod deterministický alebo stochastický

Politika



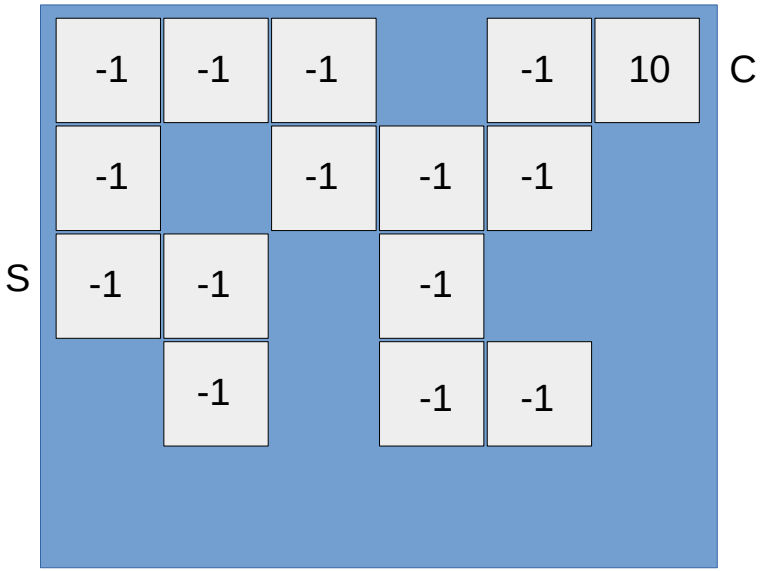
- Definuje chovanie agenta
- Mapovanie $\pi : S \rightarrow A(s)$
- Deterministcká $a = \pi(s)$
- Stochastická
  $\pi(a|s) = P[A_t = a|S_t = s]$

Odmena



- Prijímaná po každom zásahu do prostredia
- Numerická hodnota reprezentujúca spätnú väzbu
  - indikuje ako dobre si agent viedol v poslednom kroku
  - Hodnota môže byť kladná (odmena), nulová (neutrálna) alebo záporná (trest)
- Môže byť deterministická alebo stochastická
- Cieľom je maximalizovať kumulatívnu odmenu

Model

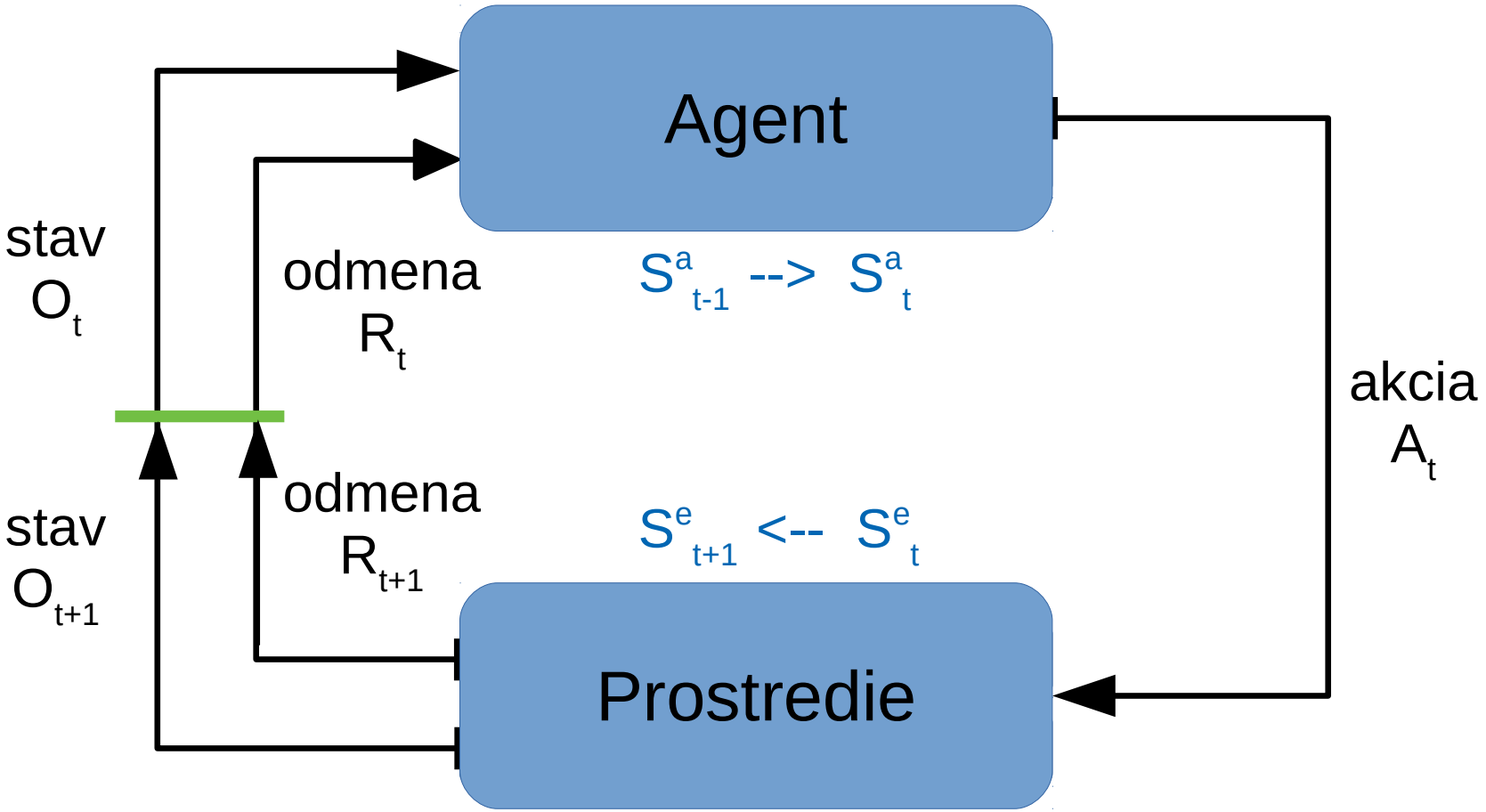


- Agentova reprezentácia chovania prostredia
- Model umožňuje predikovať
  - dynamiku (ako akcia zmení aktuálny stav)
  - odmenu (ako bude akcia v danom stave odmenená)
- Model nemusí byť perfektný
- Agentovi model môže byť daný, môže sa ho naučiť alebo ho nemusí mať

Sekvenčné prijímanie rozhodnutí

- Cieľom je aby agent selektoval akcie takým spôsobom, ktorý by maximalizoval budúcu kumulatívnu odmenu
- Agent v kroku t
  - získava odmenu R_t
  - obdrží pozorovanie O_t
  - updatuje svoju reprezentáciu S_t^a
  - vyberie a vykoná akciu A_t
- Prostredie v kroku t
  - na základe akcie A_t zmení svoj stav na S_{t+1}^e
  - emituje informáciu o novom stave O_{t+1}
  - emituje informáciu o odmene R_{t+1}
- Čas je inkrementovaný po kroku prostredia pred krokom agenta

Model interakcie agenta s prostredím



Rozhranie agent-prostredie

- Fyzické rozhranie medzi prostredím a agentom (človekom či robotom)
- Možný posun bližšie k agentovi
  - výber akcie robota - agent
  - vykonanie akcie - prostredie
- Poloha rozhrania je daná tým, čo agent ovláda alebo vie

Učenie agenta

- Učenie politiky z vlastnej skúsenosti s prostredím
  - metóda pokus-omyl
  - učenie "za behu"
  - snaha nestratiť príliš z kumulatívnej odmeny počas učenia
- Aby agent vedel posúdiť účinky akcií, musí ich vyskúšať
  - deterministické prostredie s malým počtom možností - možné úplné prehľadanie
  - stochastické prostredie - opakované skúšanie pre konvergenciu
    - k skutočným hodnotám stredných hodnôt odmien
    - k skutočným hodnotám prechodovej matice
  - ak má (úplný) model prostredia, môže využiť simulácie, inak pracuje s reálnym prostredím

Explorácia vs exploatácia

- Agent vyskúšal nejakú podmnožinu možností - je schopný z nich vybrať tie najlepšie. Môže
  - fungovať ďalej podľa takto získanej politiky
  - skúšať ďalšie z nepreskúmaných možností
- Exploatácia
  - využívanie známej informácie pre maximalizáciu kumulatívnej odmeny
- Explorácia
  - získavanie viac informácií o prostredí
- Príklady: hranie hier, návšteva reštaurácie, ťažba surovín
- Dôležité kombinovať oba princípy

Taxonómia agentov