

Q-learning - umelá inteligencia na obzore?

Ing. Michal CHOVANEC
Fakulta riadenia a informatiky

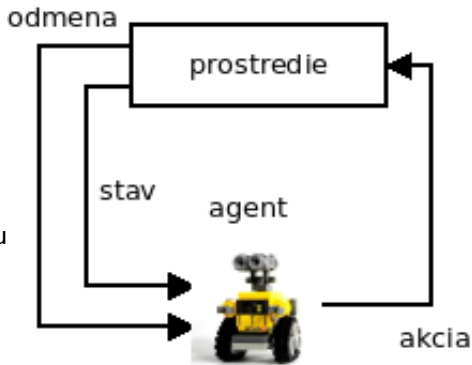
Apríl 2016

- Reinforcement learning
- Q-learning algoritmus
- Možnosti aproximácie



Reinforcement learning

- Zistenie stavu
- Výber akcie
- Vykonanie akcie
- Prechod do ďalšieho stavu
- Získanie odmeny alebo trestu
- Učenie sa zo získanej skúsenosti



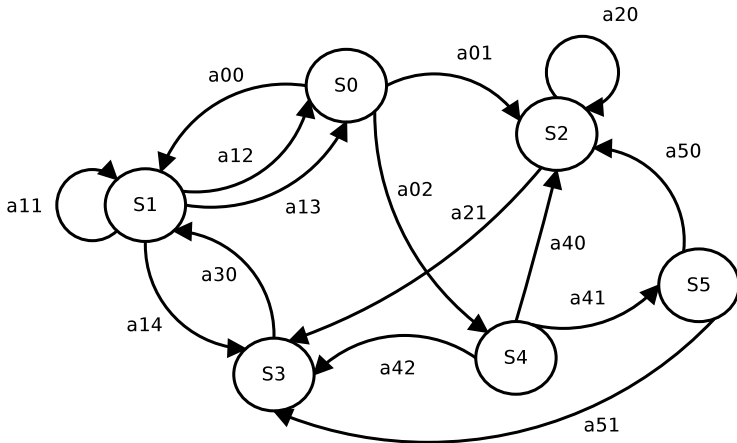
Definuje sa čo robiť, nie ako to robiť

- vďaka odmeňovacej funkcií
- agent sa môže naučiť všetky detaily problému

Lepšie konečné riešenie

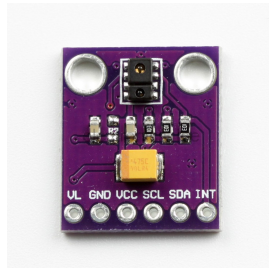
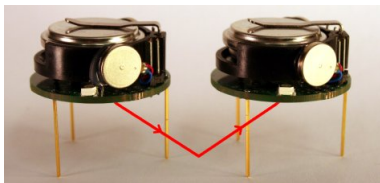
- založené na skutočnej skúsenosti, nie skúsenosti programátora
- treba menej ľudského času na nájdenie dobrého riešenia

Markovov rozhodovací proces



Experiment s jedným stavom - nanoQ learning

Plánovanie pohybu robota - aký krok má robot vybrať? Dostupné akcie : vľavo, vpravo, vpred, (vzad)

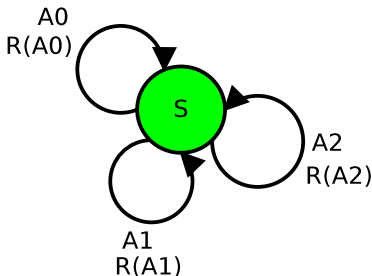


Robot s jednoduchým senzorom

- smer nie je známy
- známa je len vzdialenosť
- šum

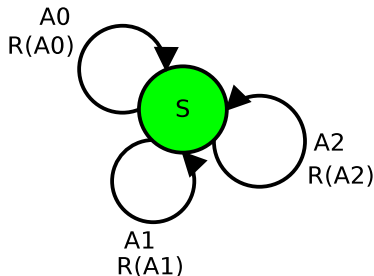
Experiment s jedným stavom - nanoQ learning

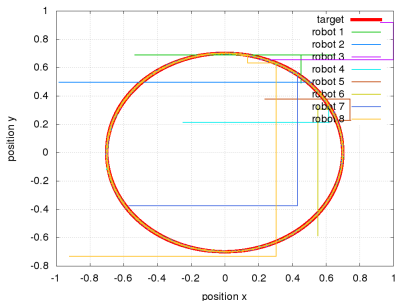
- Najjednoduchší prípad Q-learning algoritmu
- Reward zadaný dvoma hodnotami :
 - situácia sa zlepšuje +1
 - situácia sa zhoršuje -1
- Voliteľná prevdepodobnosť $p \in \langle 0, 4 \rangle$ náhodnej zmeny rewardu - šum



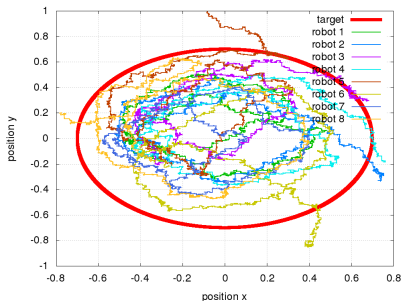
$$Q_n(A(n)) = R(n) + \gamma \max_{a'(n-1) \in \mathbb{A}} Q_{n-1}(a'(n-1)) \quad (1)$$

$$a(n) = \begin{cases} a(n-1) & \text{ak } Q_{n-1}(a(n-1)) > 0 \\ \text{random}() & \text{inak} \end{cases} \quad (2)$$

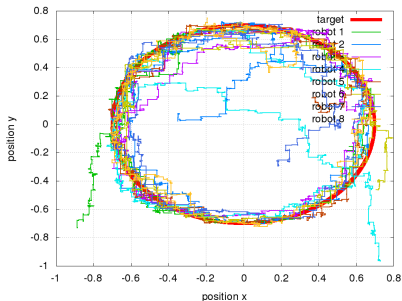




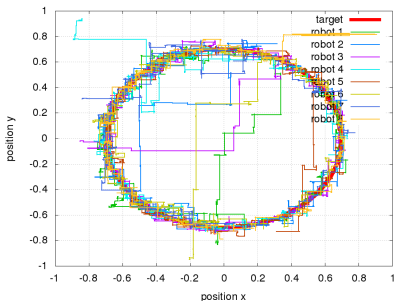
Obr. : Dráha robotov pre
 $\gamma = 0.7p = 0.0$



Obr. : Dráha robotov pre
 $\gamma = 0.7p = 0.4$



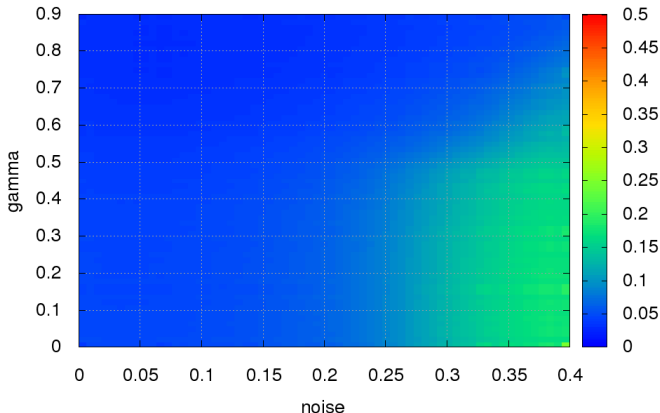
Obr. : Dráha robotov pre
 $\gamma = 0.9p = 0.4$



Obr. : Dráha robotov pre
 $\gamma = 0.98p = 0.4$

Výsledky - zhrnutie

Komplexné vyšetrenie závislosti γ a p . Znázornenie funkcie vzdialenosti od pohyblivého cieľa po 25000 krokoch robota, priemer z 32 robotov



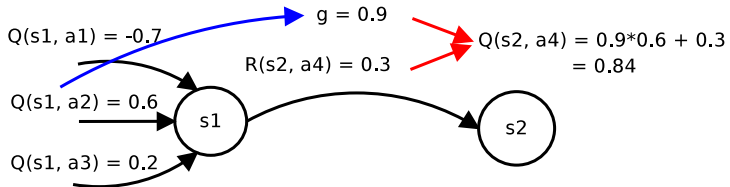
Daná je funkcia ohodnotení

$$Q(s(n), a(n)) = R(s(n), a(n)) + \gamma \max_{a(n-1) \in \mathbb{A}} Q(s(n-1), a(n-1))$$

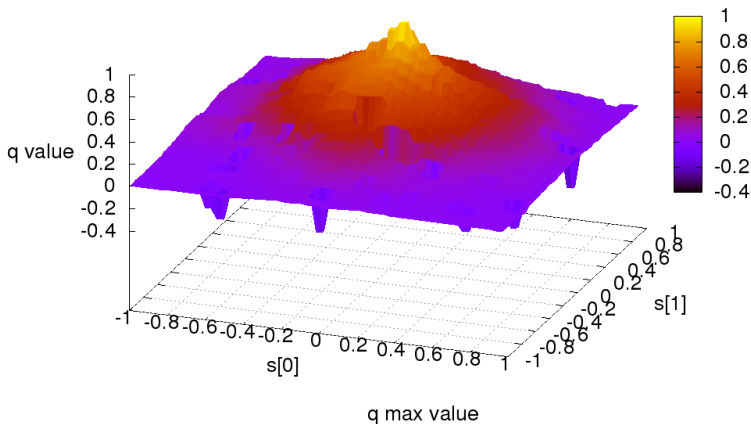
kde

- $R(s(n), a(n))$ je odmeňovacia funkcia s hodnotami v $\langle -1, 1 \rangle$,
- $Q(s(n-1), a(n-1))$ je odmeňovacia funkcia v stave $s(n-1)$ pre akciu $a(n-1)$,
- γ je odmeňovacia konštanta a platí $\gamma \in (0, 1)$.

Odmeňovacia funkcia



Rozsiahly stavový priestor - ukážka požadovaného tvaru funkcie



Implementačné problémy

Problémy tabuľkovej interpretácie $Q(s(n), a(n))$:

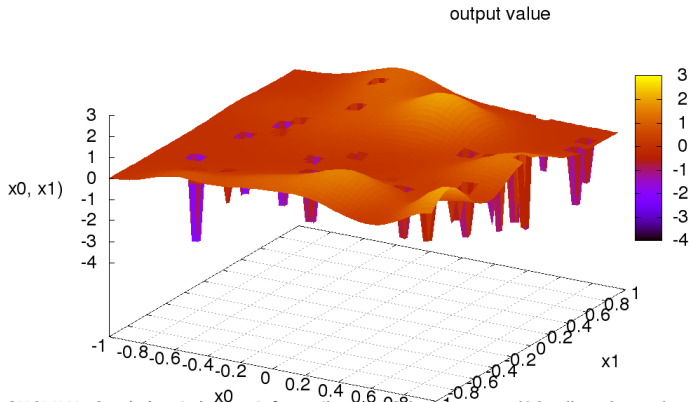
- pre veľké n_s alebo n_a narastajú pamäťové nároky,
- o nevyplnených $Q(s(n), a(n))$ nevieme povedať nič,
- pre rozsiahle stavové priestory ťažko vypočítateľné,
- ako aproximovať $Q(s(n), a(n))$?

Je možné zostaviť neurónovú sieť, ktorá sa dá naučiť lokálne?

Experiment s novou bázickou funkciou

Experimentálne sa zítali typické rysy funkcie pre

$$Q_n(s(n), a(n)) = R(s(n), a(n)) + \gamma \max_{a(n-1) \in \mathbb{A}} Q_{n-1}(s(n-1), a(n-1))$$



Experiment s novou bázickou funkcíou

Peak and Hill funkcia

$$P_i(s(n), a(n)) = \begin{cases} r_{ai} & \text{if } s(n) = \alpha_i^1 \\ 0 & \text{inak} \end{cases} \quad (3)$$

$$H_j(s(n), a(n)) = w_{aj} e^{-\beta_{aj} \sum_{i=1}^{n_s} (s_i(n) - \alpha_{aji}^2)^2} \quad (4)$$

$$Q(s(n), a(n)) = \sum_{i=1}^I P_i(s(n), a(n)) + \sum_{j=1}^J H_j(s(n), a(n)) \quad (5)$$

kde

α_j^1 sú oblasti kde $H_j(s(n))$ nadobúda nenulové hodnoty

α_{ji}^2 sú oblasti pre ktoré $f_j(s(n), a(n))$ nadobúda maximum

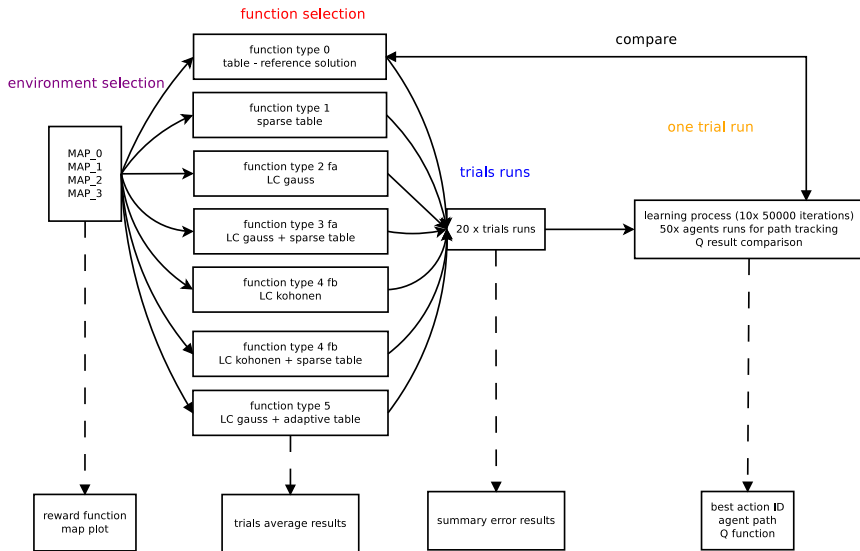
r_{ai} je hodnota okamžitej odmeny $R(s(n))$ v tomto stave

w_{aj} je váha a zobovedá veľkosti maxima resp. minima pre fukciu

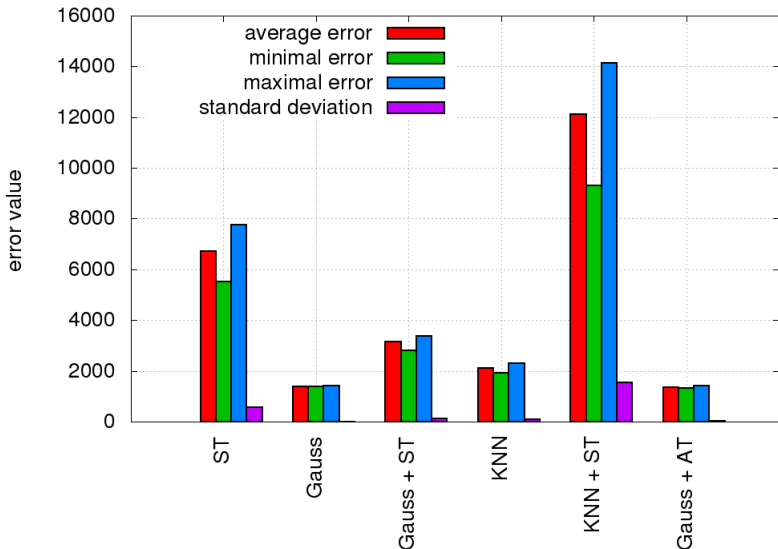
β_{aj} je strmosť, a platí $\beta > 0$

I a J sú počty bázických funkcií

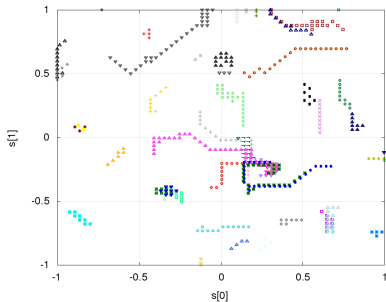
Schéma priebehu experimentov



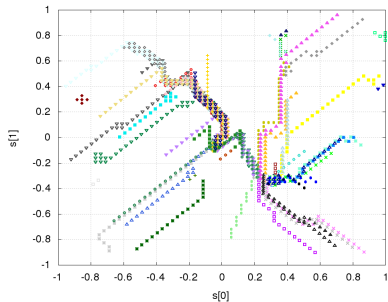
Porovnanie s ostatnými



Porovnanie s ostatnými

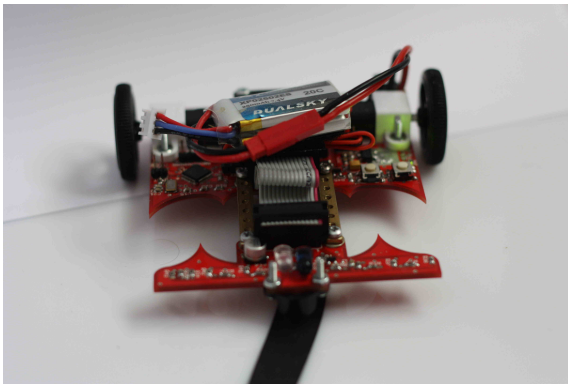


Obr. : Dráha robotov, funkcia 2 - Gauss



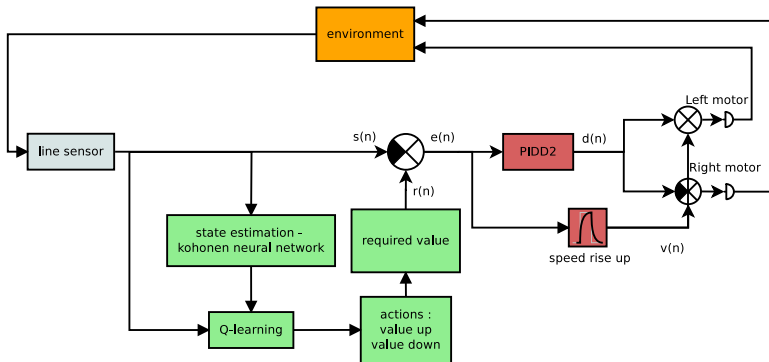
Obr. : Dráha robotov, funkcia 6 - Peak and Hill

Doplnkový experiment



Doplnkový experiment

Bloková schéma robota



Ďakujem za pozornosť

michal.chovanec@yandex.ru

https://github.com/michalnand/q_learning

