

Zuikio klajonės

Kurso „Dirbtinis intelektas“ projektinė užduotis

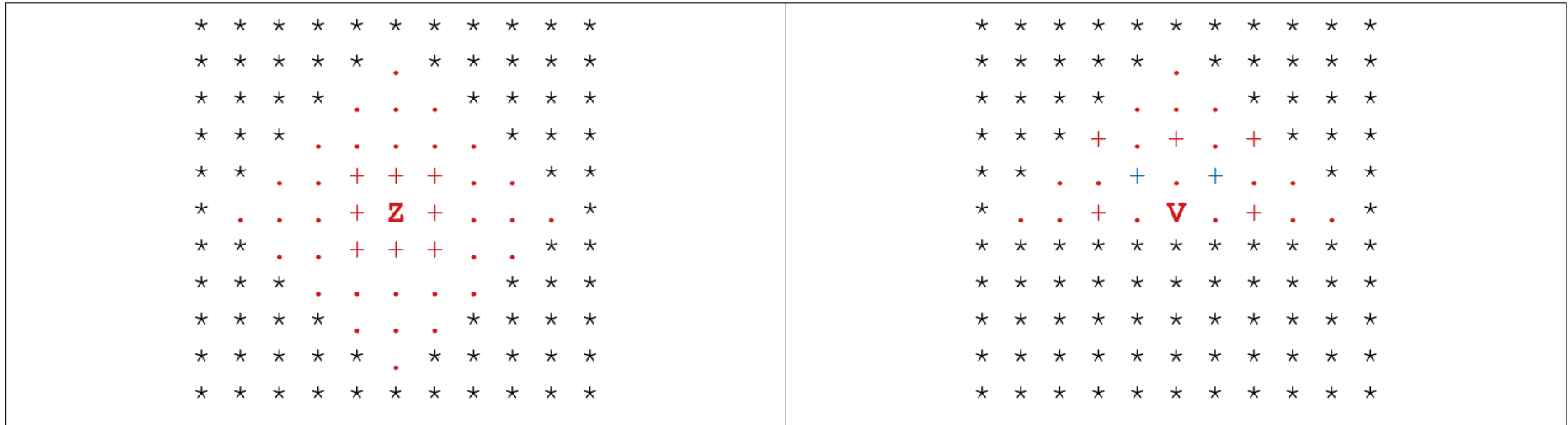
Atliko: Donatas Liupševičius
Rokas Silkinis

Užduoties sąlyga

Apibrėžimai

- Agentas (vadinkime jį Zuikiu) juda stačiakampiame lauke.
 - Kas ėjimą Zuikis pajuda į vieną iš 8 kaimyninių langelių. Zuikis negali likti vietoje.
 - Zuikis mato 4 (Manheteno) langelių atstumu visomis kryptimis (žr. pav. kairėje).
 - Zuikis neskiria kompasu krypties ir nemoka braižyti žemėlapių, bet gali įsiminti, kaip jam sekėsi vienoje ar kitoje situacijoje, apibrėžtoje pagal regėjimo lauko turinį.
- Lauke yra keli agentai – priešininkai (pavadinkime juos Vilkais).
 - Vilgai mato 4 (Manheteno) langelių atstumu, bet ne už savęs (žr. pav. dešinėje).
 - Kol Vilkas nemato Zuikio, jis juda įstrižai per vieną laukelį ir atsimuša nuo sienų.
 - Kol Vilkas mato Zuikį, jis juda jo kryptimi (iš 5 galimų) per lygiai du laukelius.
 - Kai Vilkas pameta Zuikį, jis vėl pasirenka įstrižą kryptį (iš dešinės) ir eina toliau.
 - Vilgai atminties neturi.
- Lauke dar yra atsitiktinai išdėstytų objektų – Morkų.
 - Viena Morka suteikia Zuikiui M energijos.
 - Vidutinis atstumas tarp morkų yra $0,7M-0,9M$.
 - Suvalgyta Morka iš naujo padedama atsitiktiniame langelyje.
- Zuikis pradeda ėjimus turėdamas N energijos, kur N – langelių skaičius lauke.
 - Vienas ėjimas kainuoja 1 energijos vienetą.
 - Susitikimas su Vilku kainuoja $N/4$ energijos, bet Zuikis perkeliamas per 4 langelius centro kryptimi. Jei langelyje atsидuria Zuikis, Morka ir Vilkas, Morka lieka nesuvalgyta.
 - Praradęs visą energiją Zuikis baigia darbą.

Užduoties sąlyga



Tikslas: išmokti išgyventi ir surinkti kiek įmanoma daugiau energijos.

- *min:* pagal išankstinę strategiją išgyventi daugiau ėjimų, negu lauke yra langelių.
- *max:* gyvai sudaryti regėjimo lauku pagrįstą modelį, kuris leistų Zuikiui išgyventi ir klestėti, t. y. sėkmingai išvengti Vilkų ir rinkti Morkas.
 - Vizualizuoti Zuikio klajones.
 - Išnagrinėti, kaip rezultatai priklauso nuo M ir N reikšmių. Kokia pradinė energijos reikšmė reikalinga, kad jūsų agentas išmoktų išgyventi pirmojo ir vienintelio paleidimo metu?

Sprendimo metodika

- Agento mokymui naudota **Q-mokymosi** (angl. ***Q-learning***) metodika.
- Q-mokymasis – vienas iš mašininio mokymosi metodų, skirtų gauti optimalią agento **strategiją**.
- Q-mokymosi tikslas – rasti optimalią strategiją apskaičiuojant optimalias **Q-vertes** (angl. ***Q-values***, ***action-values***) kiekvienai būsenos-veiksmo porai. Kitaip tariant, ieškoma optimali **kokybės funkcija** (kitai vadinama **veiksmo naudingumo funkcija**).
- **Belmano optimalumo lygtis:**

$$q_*(s, a) = E \left[R_{t+1} + \gamma \max_{a'} q_*(s', a') \right] \qquad q_*(s, a) = \max_{\pi} q_{\pi}(s, a).$$

Sprendimo metodika

- Agentas pradeda žaidimą nieko nežinodamas apie aplinką (visos Q-vertės lygios 0).
- Algoritmas iteraciškai atnaujinama Q-vertes kiekvienai būsenos-veiksmo porai naudojant Belmano optimalumo lygtį iki tol, kol Q-funkcija konverguoja iki optimalios Q-funkcijos.
- Optimali strategija yra sukonvergavusių Q-verčių Q-lentelė.
- Kaip pasirinkti patį pirmą veiksmą ir kaip žinoti, kad agentas iš tiesų turi daug informacijos apie aplinką? Įvedamos **tyrinėjimo** ir **eksploatacijos** sąvokos (***epsilon-greedy*** strategija).
- Q-vertės atnaujinimas:

$$q^{new}(s, a) = (1 - \alpha)q(s, a) + \alpha \left(R_{t+1} + \gamma \max_{a'} q(s', a') \right)$$

Rezultatai

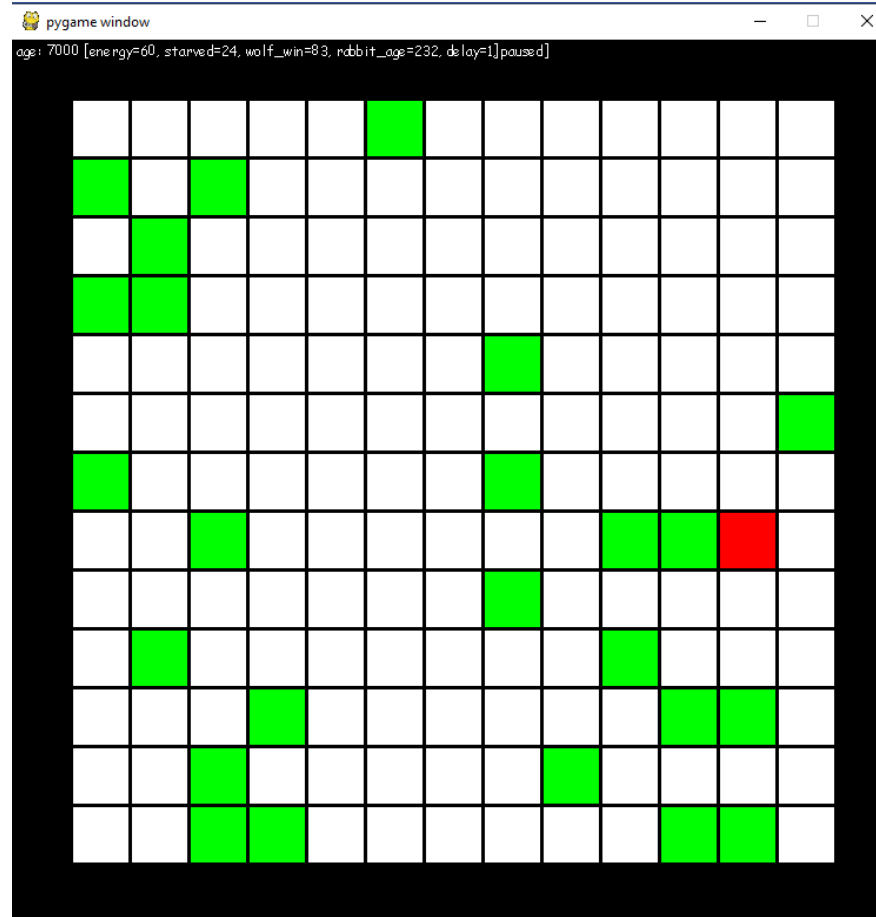
$N = 169$ (13×13)

$M = 8$

Strategija su $age = 400\,000$

1 Vilkas

25 Morkos



$starved = 24$, $wolf_win = 83$, $age = 7000$, $\epsilon = 1$

Rezultatai

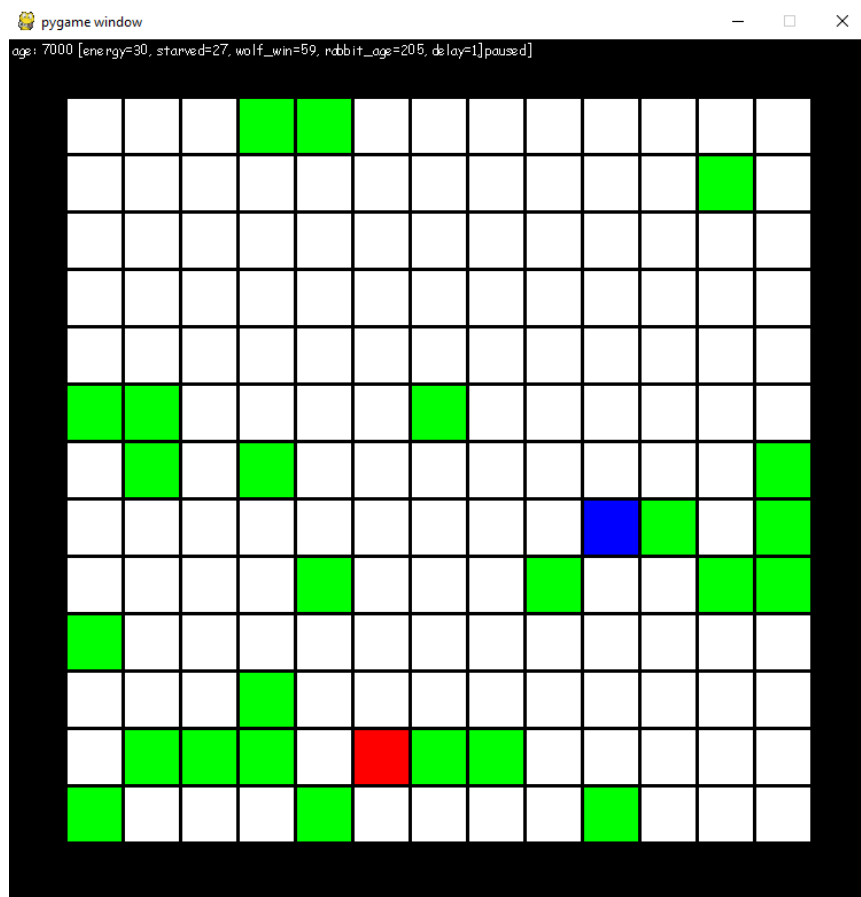
$N = 169$ (13×13)

$M = 8$

Strategija su $age = 400\,000$

1 Vilkas

25 Morkos



$starved = 27$, $wolf_win = 59$, $age = 7000$, $\epsilon = 0.5$

Rezultatai

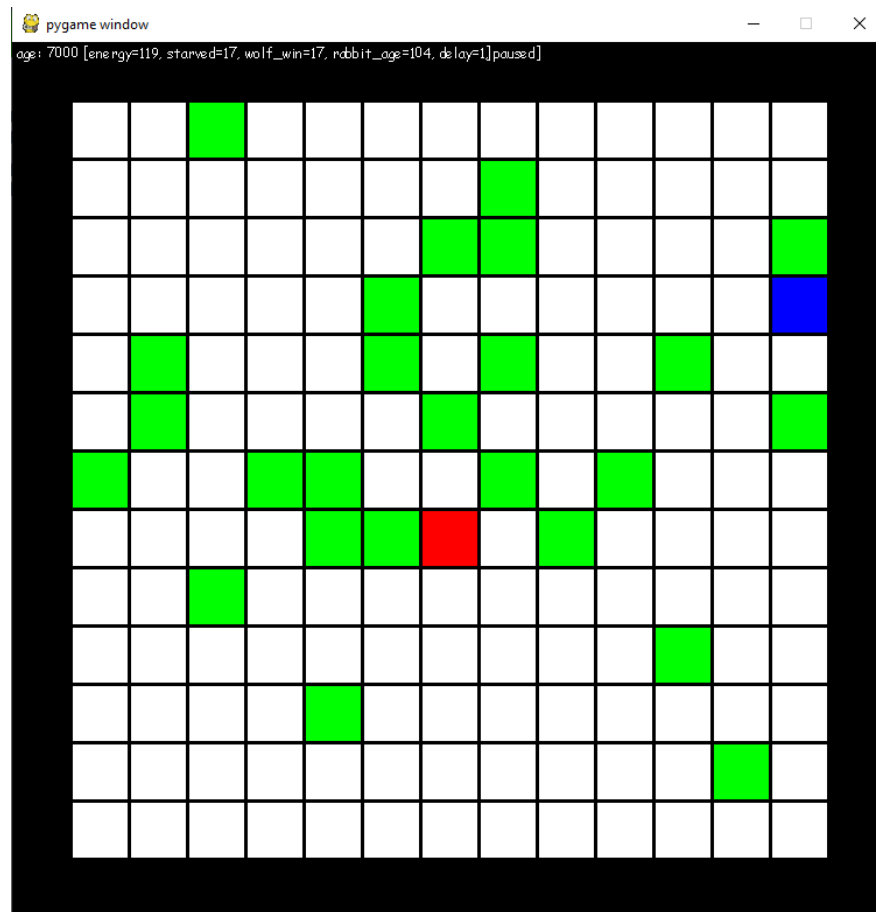
$N = 169$ (13×13)

$M = 8$

Strategija su $age = 400\,000$

1 Vilkas

25 Morkos



$starved = 17, wolf_win = 17, age = 7000, \epsilon = 0.1$

Rezultatai

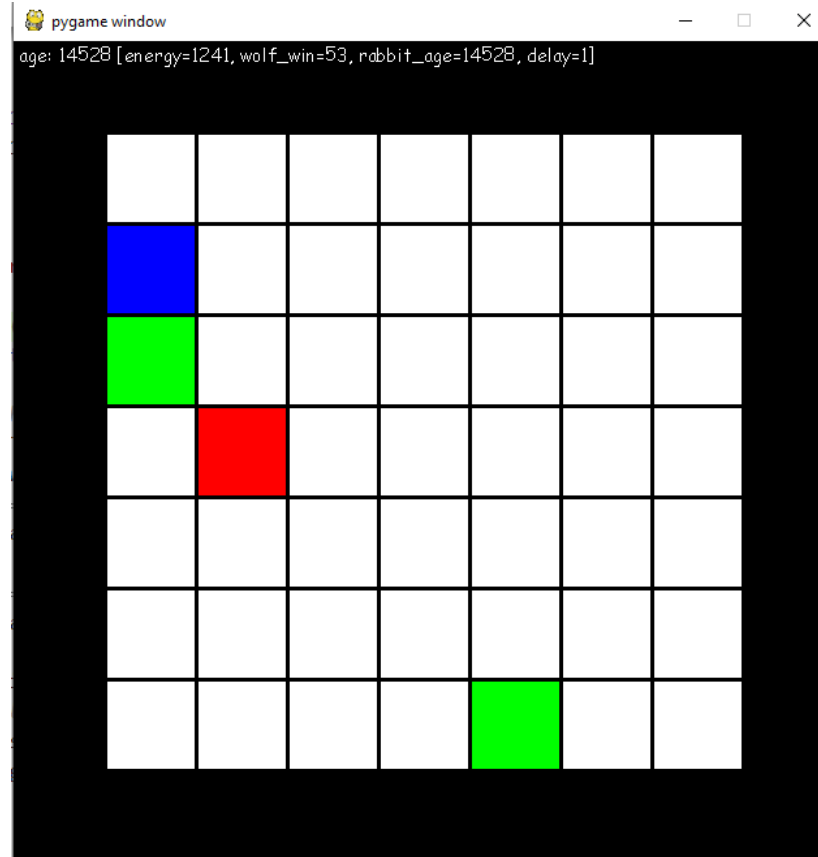
$N = 49 (7 \times 7)$

$M = 5$

Strategija su $age = 2\,200\,000$

1 Vilkas

2 Morkos



$starved = 0$, $wolf_win = 53$, $age = 14528$, $\epsilon = 0.1$