

Simulacija robota u skladištu

Problem

- Potrebno je da "robot" prenese kutije do oblasti označene na mapi.
- Problem je rešavan u dva okruženja gde robot i kutije dobiju nasumičnu poziciju na mapi:
 1. Robot prebacuje sve kutije do ciljne oblasti na sredini mape
 2. Robot prebacuje samo određene kutije do ciljne oblasti koja je na vrhu mape
- Kutije se prebacuju tako što robot "zakači" kutije za sebe i nosi ih do cilja.

Pristup rešenju

- Za rešavanje problema se koristi reinforcement learning način učenja, konkretno deep Q network algoritam.
- Neuronska mreža kao funkcija aproksimator predviđa na osnovu stanja (ulaza) koje akcije robot treba da preduzme (izlazi).
- Ulazi su informacije poput pozicija robota i kutija i zlazi smer u kom robot treba da se kreće (vertikalno, horizontalno, dijagonalno).
- Robot dobija nagrade ako se približio ciljnoj kutiji, ili se kutija približila cilju, dok za obrnutu situaciju dobija negativnu nagradu.

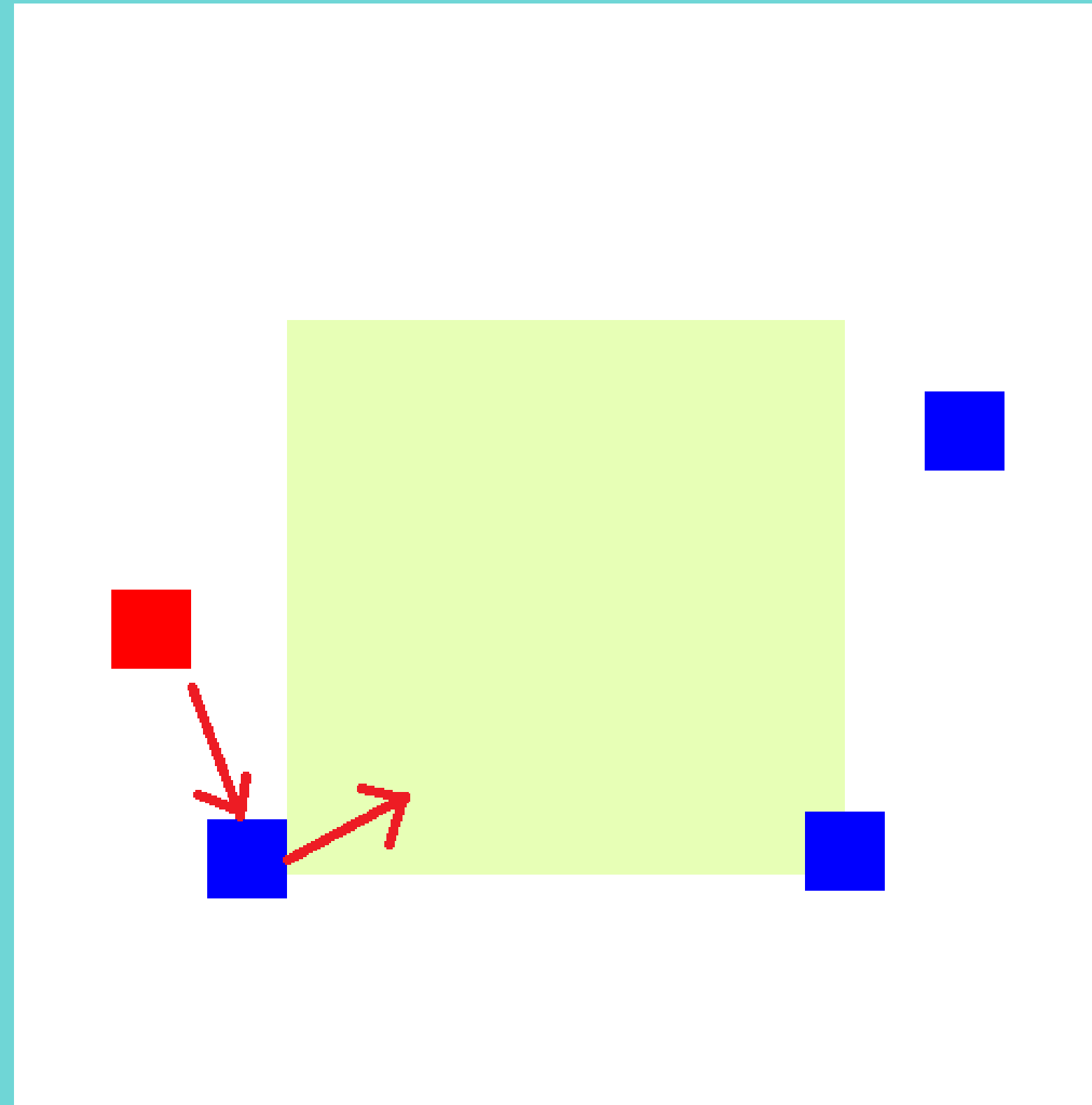
Prvo okruženje

- Robot može da prebacuje N kutija do ciljne oblasti.
- Neuronska mreža koja se koristi ima 7 ulaza, dva sloja od 256 neurona i 8 izlaza.
- Ulaz u neuronsku mrežu: x,y pozicija robota, trenutne kutije, centra ciljne pozicije, i da li je robot zakačio kutiju za sebe ili ne.
- Izlazi isti kao što je navedeno iznad.

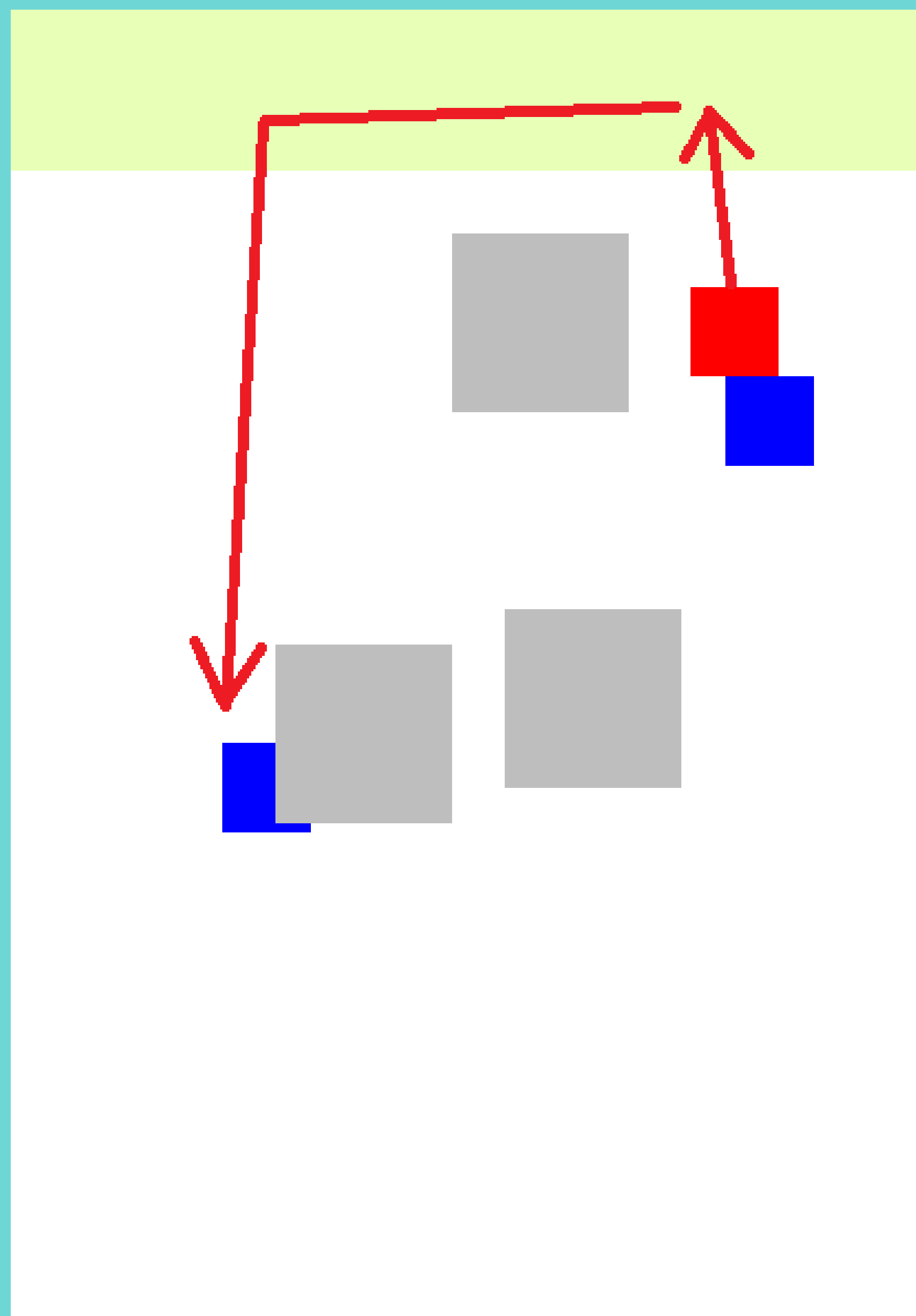
Performanse

- Robot u najvećem broju slučajeva prenese sve kutije do cilja, ne uzimajući nelogične putanje i kretajući se ka cilju.
- U nekim slučajevima robot krene samo da se preključuje između akcija "napred-nazad" i zbog toga ne uspe da prenese sve kutije.

1.



2.



Drugo okruženje

- Robot iz grupe kutija treba da prenese samo određene kutije.
- Neuronska mreža koja se koristi ima 13 ulaza, tri sloja od 256 neurona i 8 izlaza.
- Ulaz u neuronsku mrežu: x, y pozicije robota, kutija koje trebaju da se preskoče i oznake da li je ciljna kutija (gore, levo...), 1 ako jeste 0, ako nije.
- Izlazi isti kao pre.

Performanse

- Robot uspeva da prenese tražne kutije do ciljne oblasti, najčešće preskakajući kutije koje treba preskočiti.
- Ako je pozicija kutija komplikovanija, robot nekad prenese i jednu-dve kutije koje ne treba.
- Kao i u prvom okruženju, ponekad se desi da robot samo preključuje između dve akcije.

Proces treniranja

- Treniranje neuronske mreže se odvijalo brzo i u zavisnosti od kompleksnosti okruženja dovoljno je bilo 10-20 minuta da se nauči šablon ponašanja koji u većini slučajeva rešava problem.
- Problem je bio što malo posle tog perioda treniranje počinje samo da pogoršava ponašanje robota.
- Problem je nastao zbog pojave u reinforcement learning-u koja se zove "catastrophic forgetting".
- Dešava se jer robot počne dobro da se ponaša, ali posle nekog vremena izgubi u svojoj memoriji koje su akcije dobre, a koje loše.
- Problem je delimično rešen čuvanjem podataka sa početka treniranja i postepenim smanjivanjem learning rate-a tokom treniranja.