

ÖNVEZETŐ AUTÓZÁS DUCKIETOWN KÖRNYEZETBEN

K ÉPÜLET

ENDRÉSZ BALÁZS – F10RLU

MONORI J. BENCE – PVUZ1Z

WENESZ DOMINIK – NBMU7U

CÉLKITŰZÉSEK – MOTIVÁCIÓ

- Önvezetés: érdekes probléma mechatronikai szempontból
- Megismerkedés a Reinforcement Learninggel
- AI beats Chrome Dino.
- Kísérletezés a Q learninggel, egy algoritmust alapjaitól felépíteni
- Cél: Q-learning sikeres implementálása

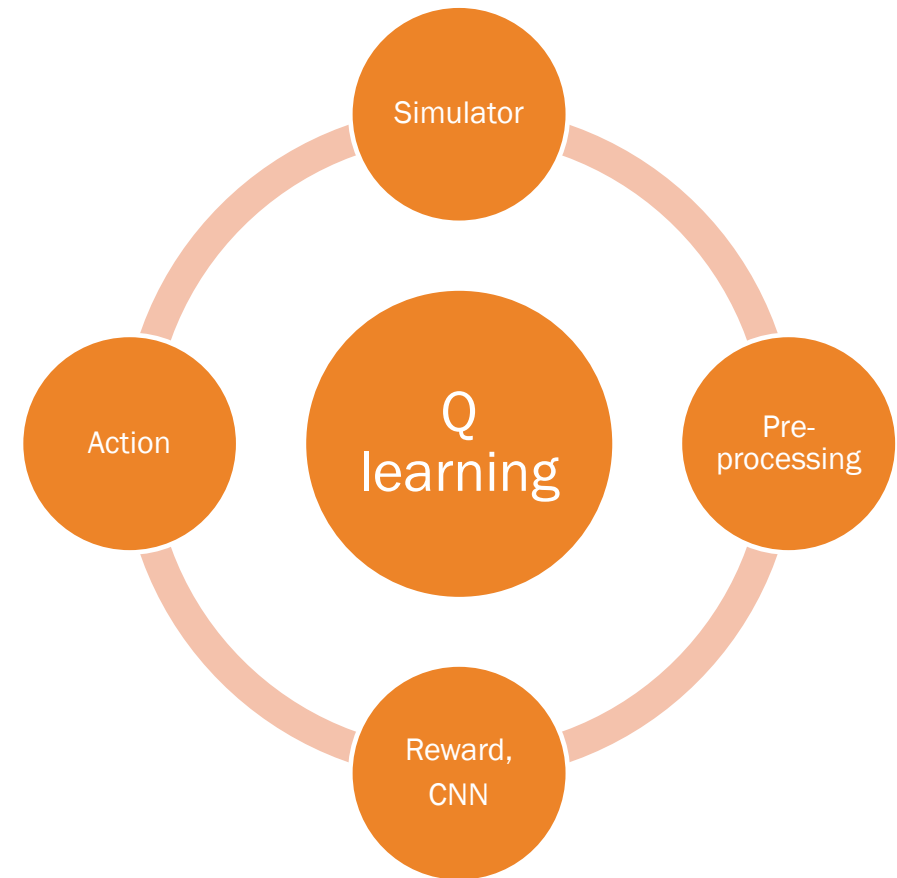


ELŐZŐ MEGOLDÁSOK - INFORMÁCIÓSZERZÉS

- Alul látható kutatási eredmények voltak a fő kiinduló pontok
 - Saját magunk által implementált algoritmussal akartunk dolgozni
 - Q-learning kapcsán tanulmányok olvasása
-
- Péter Almási, Róbert Moni, and Bálint Gyires-Tóth. Robust reinforcement learning-based autonomous driving agent for simulation and real world. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–8, 2020.
 - András Kalapos, Csaba Gó, Róbert Moni, and István Harmati. Sim-to-real reinforcement learning applied to end-to-end vehicle control. In 2020 23rd International Symposium on Measurement and Control in Robotics (ISMCR), pages 1–6, 2020.

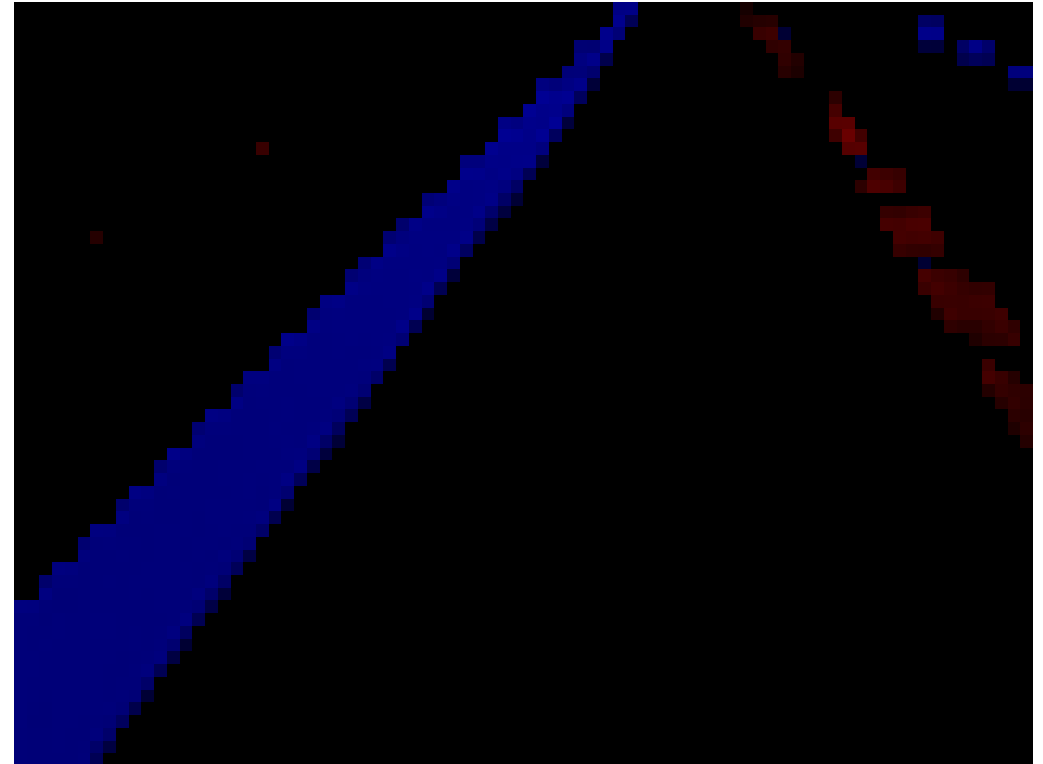
RENDSZERTERV

- Minden lépésben elmentjük a képernyőképet
- A kacsára ráukjuk a 'szemüveget'
- Egy konvolúciós hálóval tanítjuk a rendszert
- Rewardot számolunk
- Meghatározzuk a legjobb kimenetet
- Visszaadjuk ezt a rendszernek



A 'SZEMÜVEG'

- A hatékony tanítás érdekében a képeket feldolgozzuk
- Szegmentáljuk a felfestett vonalakat – R/B csatorna
- Átméretezés, normalizálás: 80x60 pixel
- Bővítési lehetőség: akadályok szegmentálása



- C.K. Chow and T. Kaneko. Boundary detection of radiographic images by a threshold method. In Satoshi Watanabe, editor, *Frontiers of Pattern Recognition*, pages 61–82. Academic Press, 1972.

A CNN HÁLÓZAT

- Vizuális környezet: konvolúciós háló
- Max pooling hatása: nem feltétlenül javít a hálón
- Hiperparaméteroptimalizálás: mekkora legyen a háló
 - LeNet: LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), 1995.

Layer		
Name	Shape	Param
Conv2D	(58,78,32)	896
MaxPooling2D	(29,39,32)	0
Conv2D	(27,37,32)	9248
MaxPooling2D	(13,18,32)	0
Conv2D	(11,16,64)	18496
MaxPooling2D	(5,8,64)	0
Flatten	(2560)	0
Dense	(128)	327808
Dense	(3)	387

Q LEARNING

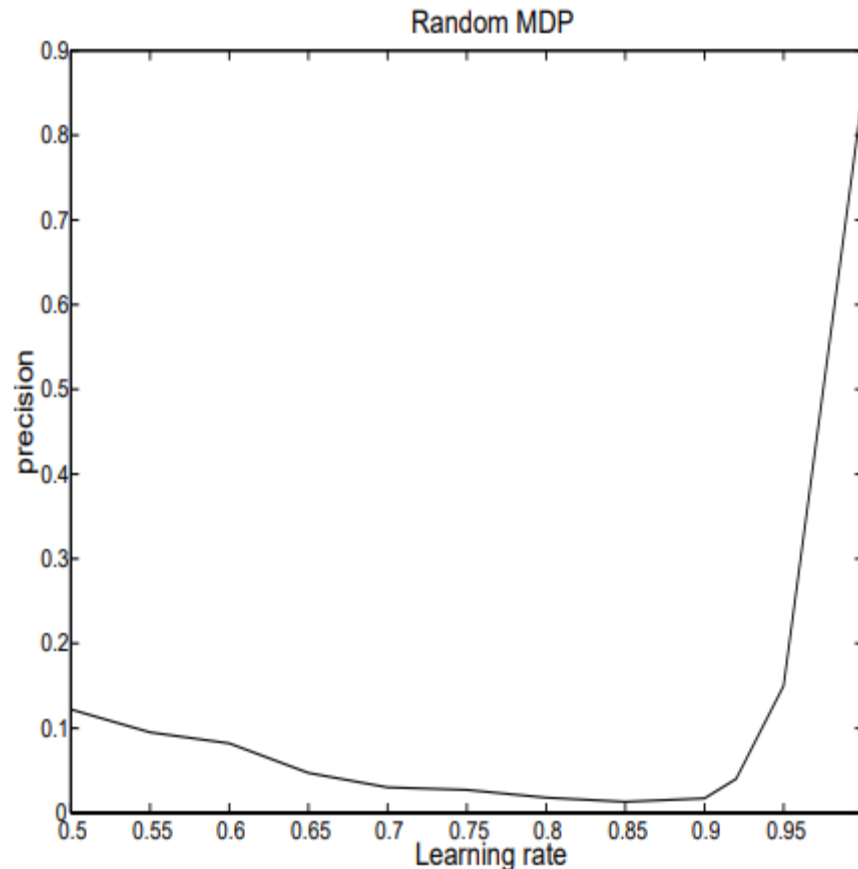
- Egy szimulátoron belül dolgozunk: gym-duckietown
- Q learning alapja: reward function: r_t
- $Q' = Q + \alpha \cdot (r_t + \gamma \cdot \max Q_{est} - Q)$
 α – learning rate; γ – discount factor
- Epsilon-greedy policy
- Experience replay

- Christopher JCH Watkins and Peter Dayan. Q-learning. Machine learning, 8(3-4):279–292, 1992.



Andrej Andrejevics Markov
(matematikus, 1856–1922)

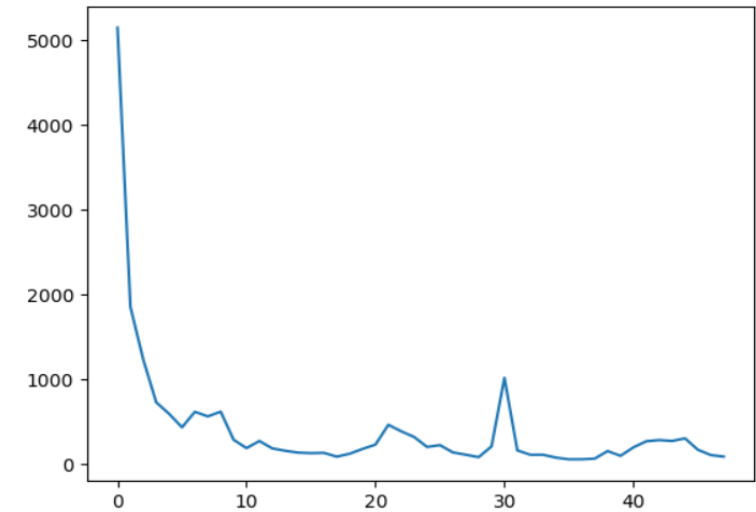
NEHÉZSÉGEK



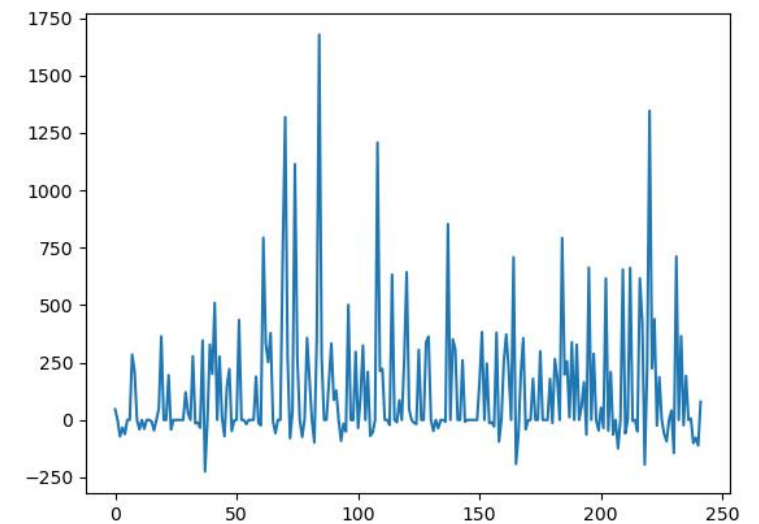
- Q learning: saját implementálás
- Magasfokú sztochaszticitás, a modellnek folyamatosan számolni kell
- Nem feltétlen optimális a DT környezetben
- Hiperparaméteroptimalizálás
learning rate, discount factor
- Eyal Even-Dar, Yishay Mansour, and Peter Bartlett. Learning rates for q-learning. Journal of machine learning Research, 5(1), 2003.

EREDMÉNYEK

- Loss function: konvergál
- Reward: törekedik a nagy rewardra
- Fejlesztési lehetőségek:
 - Előre megírt algoritmusok
 - Hiperparaméteroptimalizálás
 - További challengek



loss function



reward

ÖNVEZETŐ AUTÓZÁS DUCKIETOWN KÖRNYEZETBEN

K épület:

- Endrész Balázs
- Monori J. Bence
- Wenesz Dominik

2021. 12. 14.

KÖSZÖNJÜK A FIGYELMET!

