Мультиагентное обучение с подкреплением в задачах ИБ

Чернышов Юрий Юрьевич

к.ф.-м.н., доцент кафедры «АТиСУ» ИРИТ-РТФ УрФУ

заведующий лабораторией кибербезопасности NEO Lab ИРИТ-РТФ УрФУ

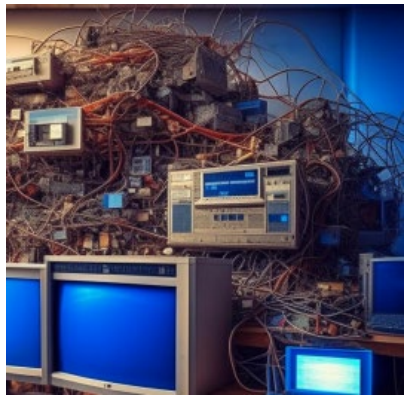
руководитель исследовательского центра Udv Group

RusCrypto
март 2024

Системы усложняются

Components

Data



Communications

SW Dev & Ops Tools

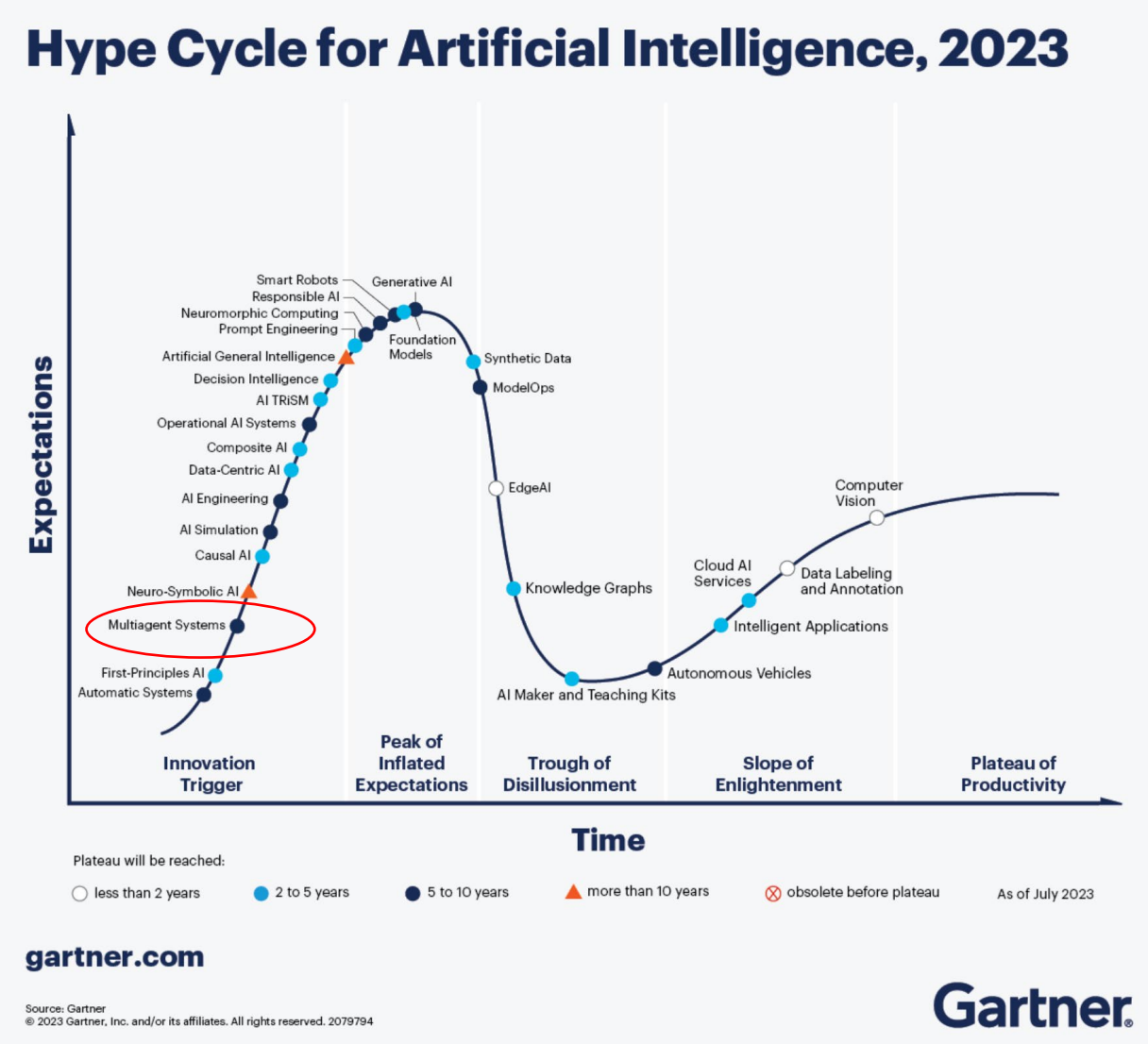
Ages

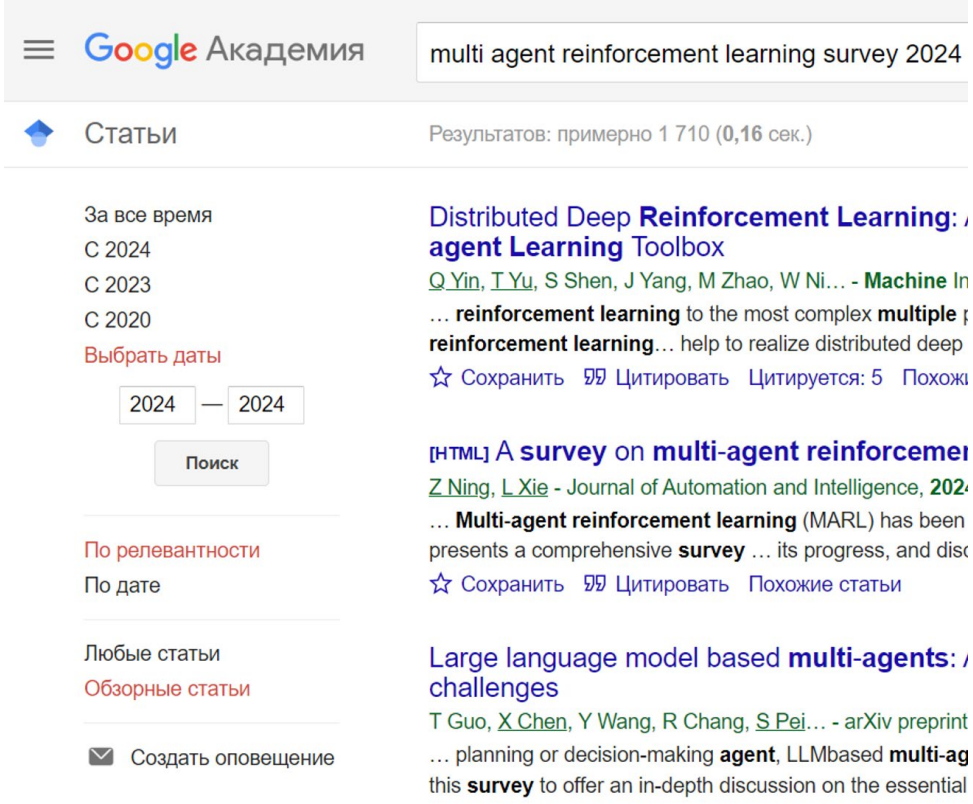
Threats

Regulation

Infrastructure

… и аналитики об этом знают !





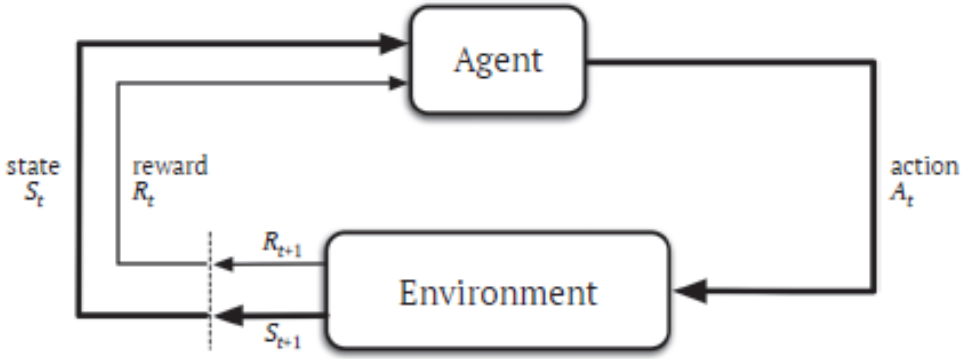
From Google Scholar

https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle

Обучение с подкреплением (Reinforcement Learning , RL)

Метод «проб и ошибок» (Trial and Error Learning, TE):

- Взаимодействие со средой
- Самообучение
- Нацеленность на получение награды



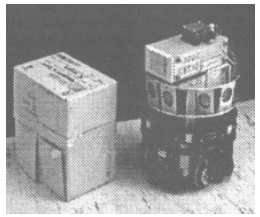
$$s_0,\ a_0,\ r_1,\ s_1,\ a_1,\ \ldots,\ r_n,\ s_n$$

$$Q^*(s,a) = \sum_{s'} \mathcal{T}(s',s,a)\left[r + \gamma \max_{a'} Q^*(s',a')\right]$$

Примеры применения RL

Designed a robot that can push cubes

Mahadevan, S., and Connell, J. (1992). Automatic programming of behavior-based robots using reinforcement learning. Artificial Intelligence, 55(2-3), 311-365.



Created a humanoid robot that can effectively solve the pole-balancing task

Schaal, S. (1997). Learning from demonstration. In Advances in Neural Information Processing Systems (pp. 1040-1046)

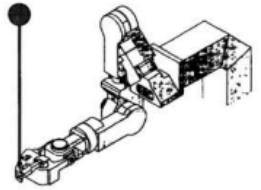
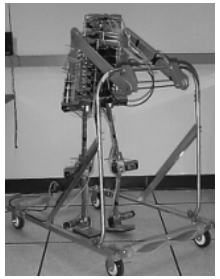


Figure 5: Sketch of SARCOS anthropomorphic robot arm

Made a biped robot that can learn to walk without any knowledge of the environment

Benbrahim, H., and Franklin, J. A. (1997). Biped dynamic walking using reinforcement learning. Robotics and Autonomous Systems, 22(3-4), 283-302

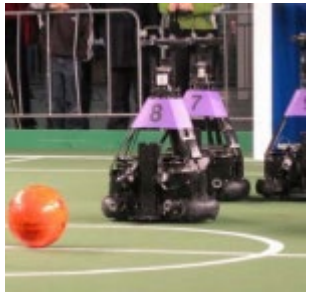


Trained a robot to play table tennis

Mulling, K., Kober, J., Kroemer, O., and Peters, J. (2013). Learning to select and generalize striking movements in robot table tennis. The International Journal of Robotics Research, 32(3), 263-279.

Built a soccer robot team

Riedmiller, M., Gabel, T., Hafner, R., and Lange, S. (2009). Reinforcement learning for robot soccer. Autonomous Robots, 27(1), 55-73



Q-learning has been applied to solve various real-world problems, but it is unable to solve high-dimensional problems where the number of calculations increases drastically with number of inputs.

Deep RL: все началось с успеха deep ML для “Atari games”

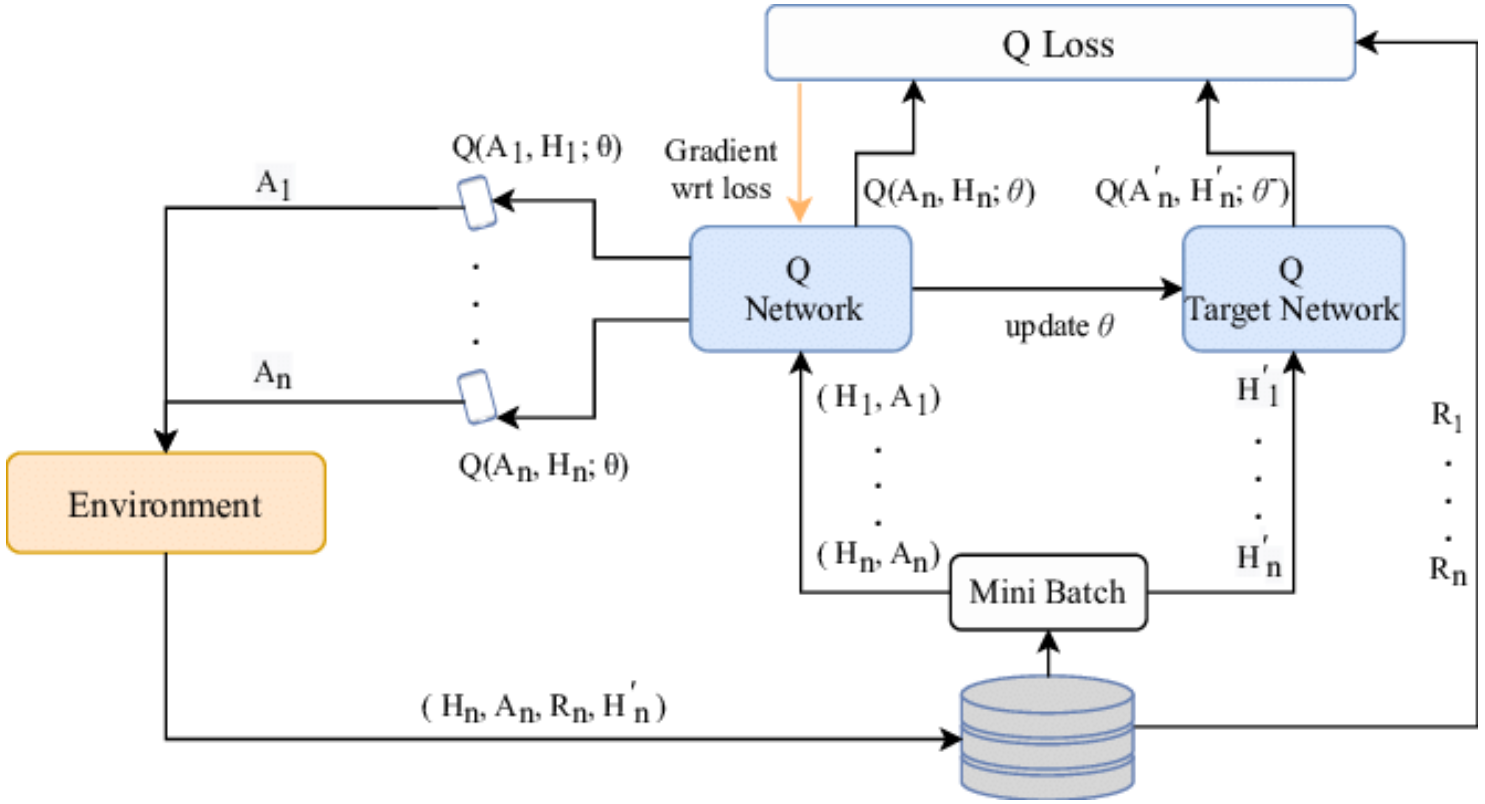
Breakout Atari game

- Состояния описываются массивом пикселей
- Действия: влево и вправо
- Награда: очки
- Конечное состояние: мяч упал

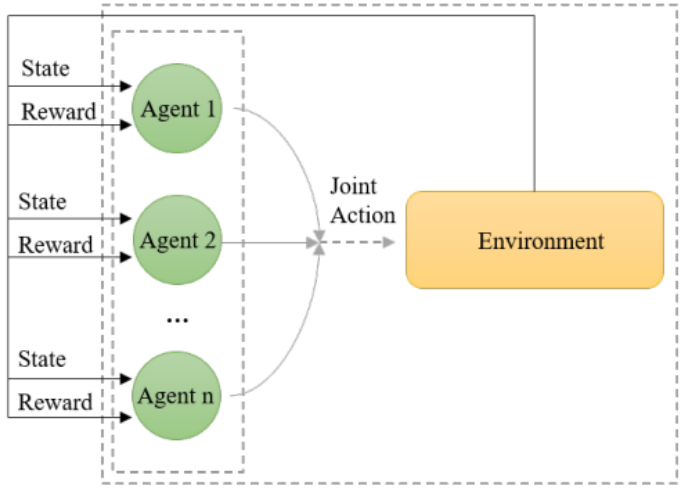
- DRL использует DL как аппроксиматор для высокоразмерных данных
- Deep Q-network by Mnih et al. (2013)
- DQN создает Q-values для всех действий в конкретном состоянии
- Применяется CNN
- Своего рода “policy deep network”
- Превзошел человека в 49 Atari games

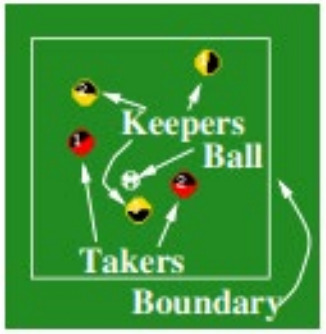
“Playing Atari with Deep Reinforcement Learning”. Mnih, Kavukcuoglu, Silver, Graves, Antonoglou, Wierstra, Riedmiller.
https://arxiv.org/abs/1312.5602

Neural networks, Replay memory, Online/Target Q -networks and other from DRL kitchen



Multi -agent Deep Reinforcement Learning (MADRL)





Multiagent domains as robotic soccer (Stone and Sutton, 2001, Balch, 1997)

Predator-and-prey pursuit games (Tan, 1993, De Jong, 1997, Ono and Fukumoto, 1996)

Плюсы

Агенты могут выполнять действия

Не требуется строгих ограничений среды

Минусы

Много ситуаций в реальном мире, когда недостаточно одного агента

Агент не знает о стратегиях других агентов, усложняется среда

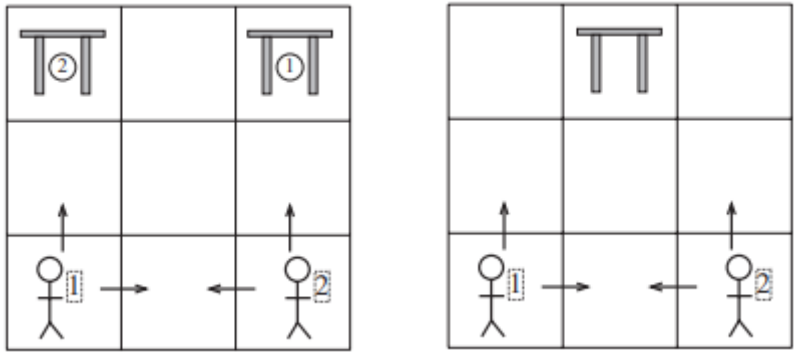
Простой пример Q-обучения с совместными действиями агентов

Hu and Wellman, 2000 Nash Q-Learning for General-Sum Stochastic Games. Taking into account joint actions of agents in Q-learning

Способ 1: Общая Q-table

Способ 2: Отдельные Q-learning

Способ 3: Для стохастической игры с общей суммой Q-функция вычисляется с дополнительной целью для агентов – добиться равновесия по Нэшу.

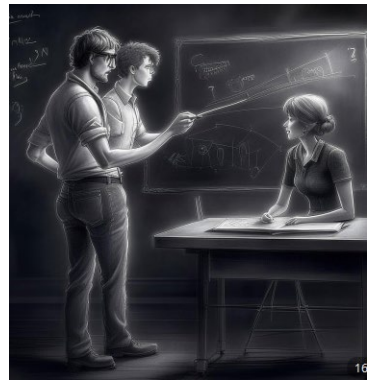


$$Q_{t+1}^{j}(s,a^{1},\ldots,a^{n}) = (1-\alpha_t)Q_t^{j}(s,a^{1},\ldots,a^{n}) + \alpha_t\left[r_t^{j} + \beta Nash Q_t^{j}(s')\right]$$

Сложности для многоагентного Deep RL

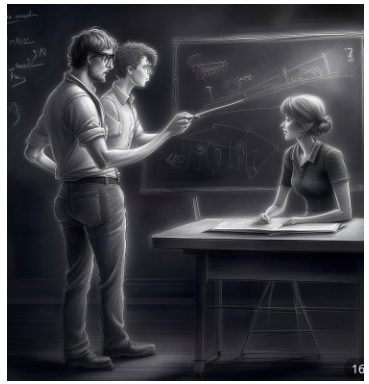
non-stationarity

partial observability



Interpretability

multi-agent training schemes



transfer knowledge between agents

«Leniency » против « Non-stationarity »

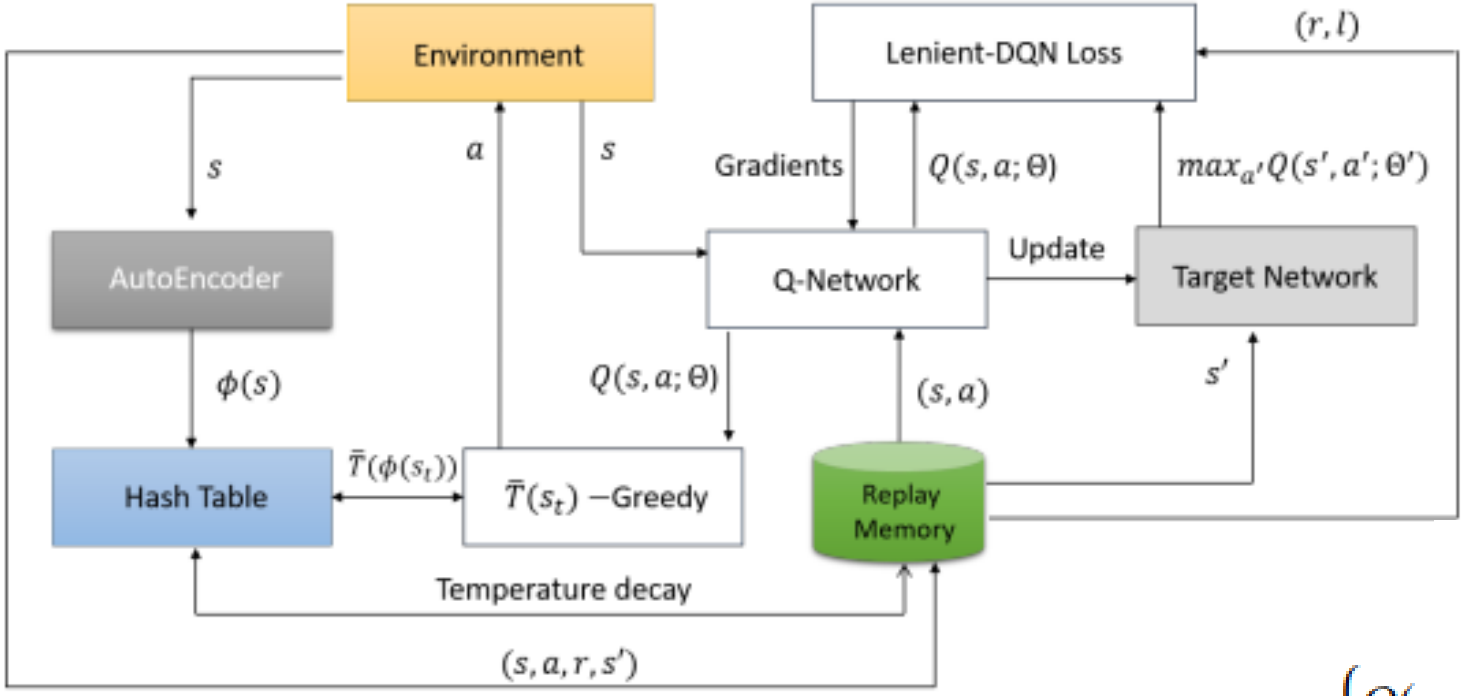


Fig. 7: Architecture of LDQN.

$$l(s_t, a_t) = 1 - e^{-K * T_t(s_t, a_t)}$$

$$Q(s_t, a_t) = \begin{cases} Q(s_t, a_t) + \alpha\delta & \text{if } \delta > 0 \text{ or } x > l(s_t, a_t). \\ Q(s_t, a_t) & \text{if } \delta \leq 0 \text{ and } x \leq l(s_t, a_t). \end{cases}$$

Lenient Multi-Agent Deep Reinforcement Learning. arXiv:1707.04402v2 [cs.MA] 27 Feb 2018

«Recurrent policy inference » против «Partial observability »

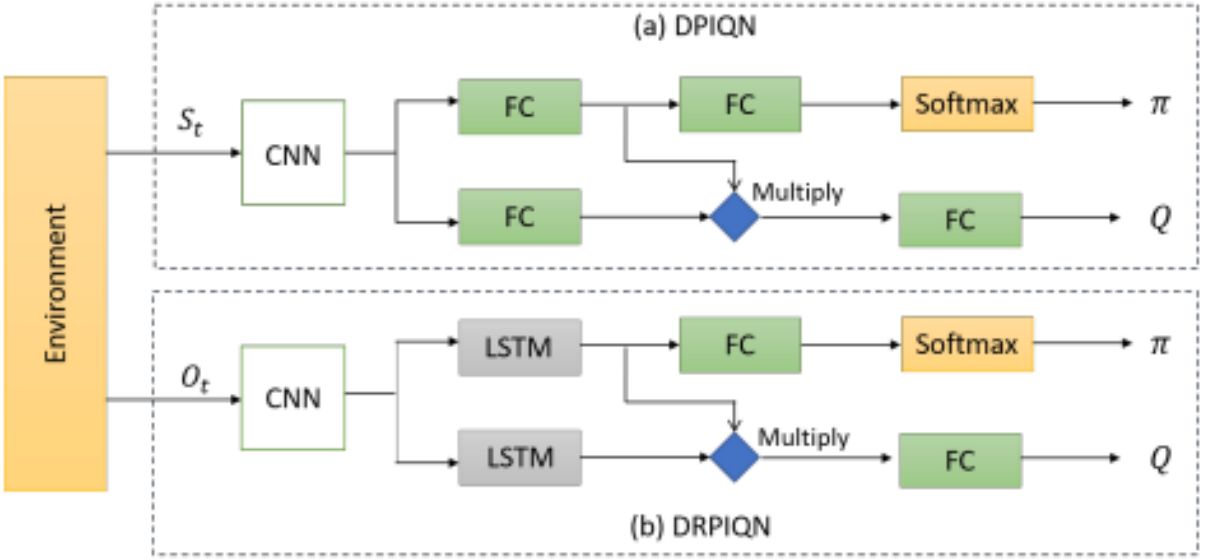
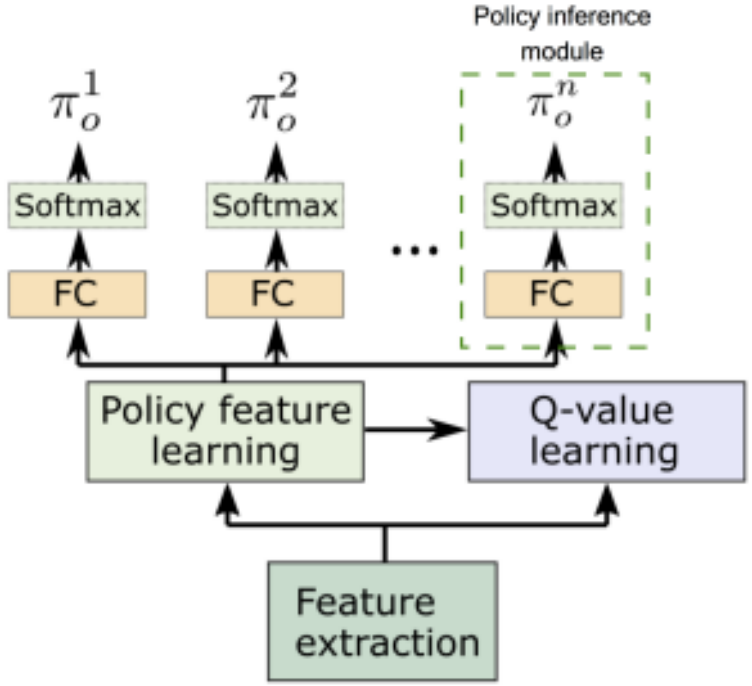


Fig. 8: Architecture of DPIQN and DRPIQN.



Hong, Z. W., Su, S. Y., Shann, T. Y., Chang, Y. H., and Lee, C. Y. (2018, July). A deep policy inference Q-network for multi-agent systems. In Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (pp. 1388-1396).

MAS training schemes: централизованное обучение и децентрализованное выполнение

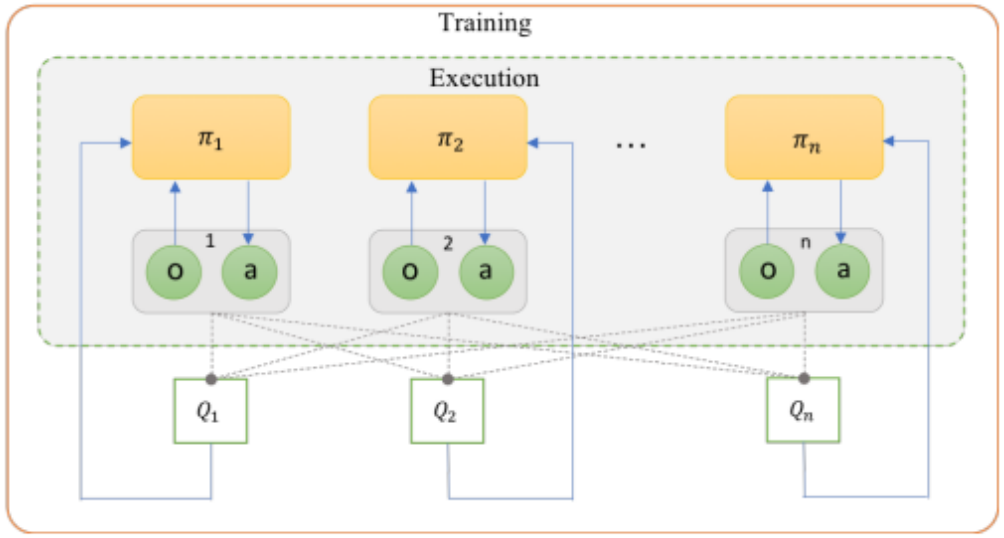
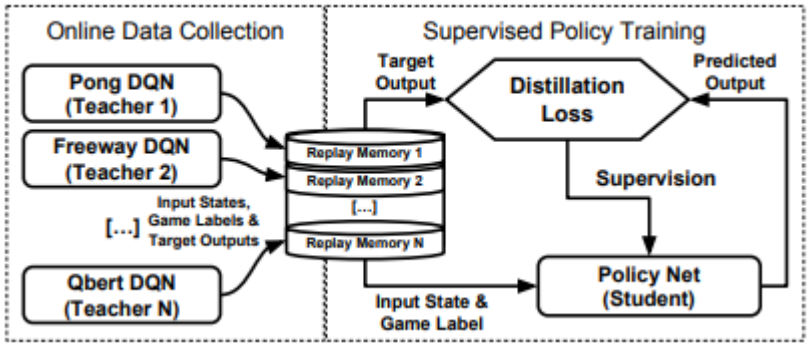


Fig. 9: Centralized learning and decentralized execution based MADDPG where policies of agents are learned by the centralized critic with augmented information from other agents' observations and actions.

Transfer Knowledge



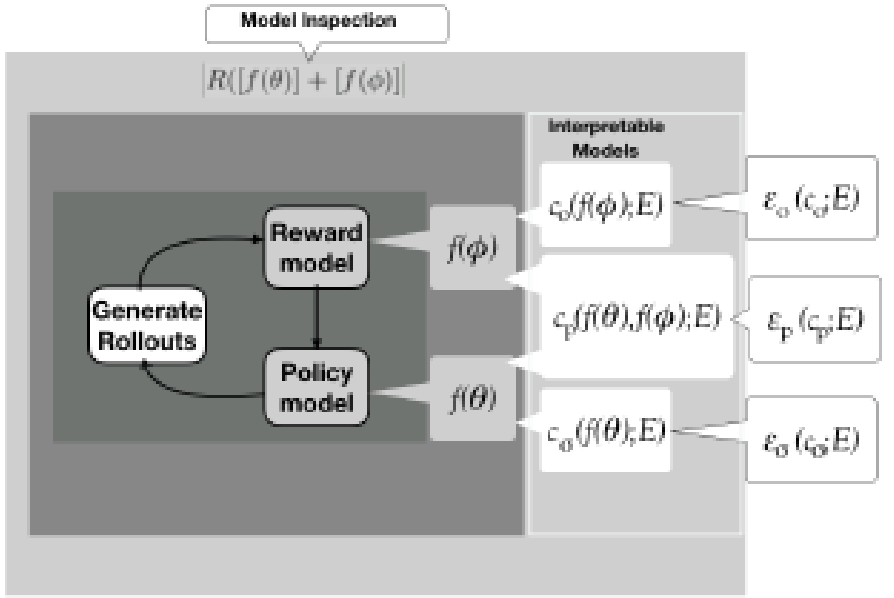
Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, O. P., and Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. In Advances in Neural Information Processing Systems (pp. 6379-6390).

Rusu, A. A., Colmenarejo, S. G., Gulcehre, C., Desjardins, G., Kirkpatrick, J., Pascanu, R., ... and Hadsell, R. (2015). Policy distillation. arXiv preprint arXiv:1511.06295.

Объяснимый DeepRL

Задачи

- Model inspection
- Policy explanation
- Reward explanation



Explainable Deep Reinforcement Learning: State of the Art and Challenges
https://arxiv.org/abs/2301.09937

Методы

- Distillation
- Mimic with more explainable model

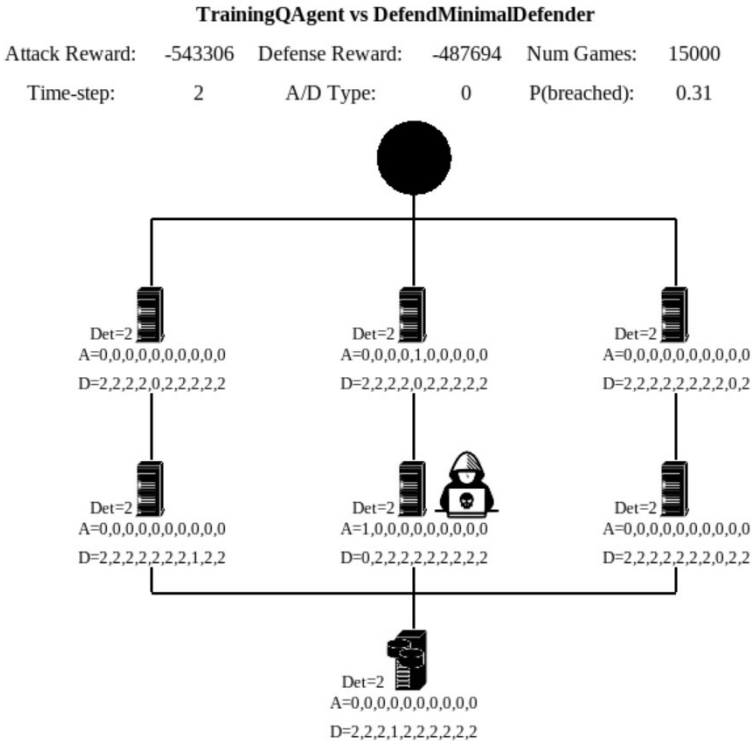
Примеры применения MAS DRL в кибербезопасности

gym-idsgame

Симулятор действий атак и защит в абстрактной игре «Intrusion network»

2-player Markov game

Включает подготовленные бейзлайны
q-функция



experiments > tests > test1.py

```python
import gymnasium as gym
from gym_idsgame.envs import IdsGameEnv
env_name = "idsgame-maximal_attack-v3"
env = gym.make(env_name)

print(env.reset())

```

```
(array([[0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
       [0., 0., 0., 0., 0., 0., 0., 0., 0., 1.]]), {})
```

Hammar, Stadler. Finding Effective Security Strategies through Reinforcement Learning and Self-Play
https://arxiv.org/abs/2009.08120 https://github.com/Limmen/gym-idsgame

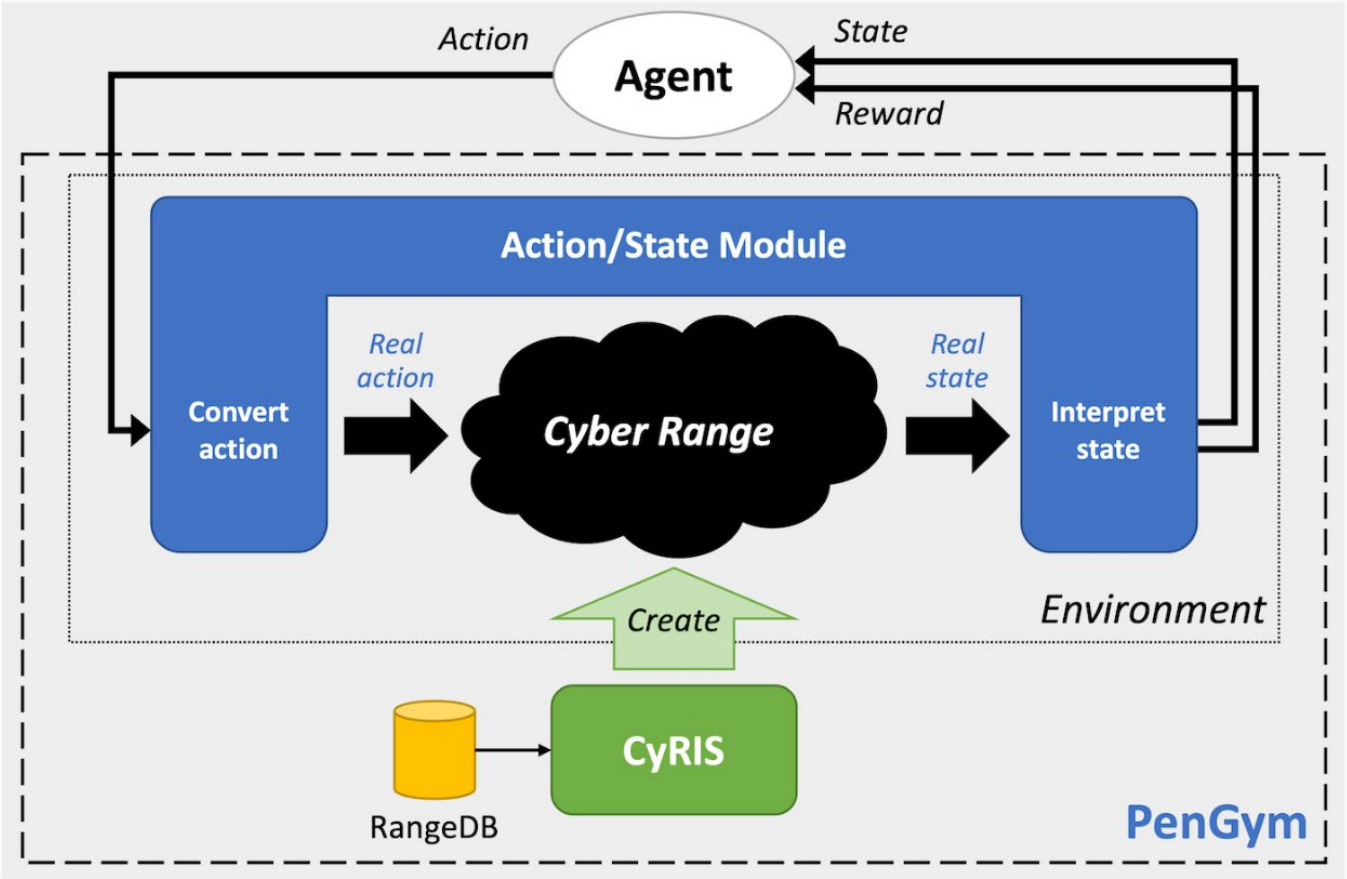
RL для тестирования на проникновение, среда PenGym

JAIST и KDDI

Nmap, Metasploit

Cyber Range Instantiation System

Network Attack Simulator



https://github.com/cyb3rlab/PenGym

Итоги, выводы, планы

Методы RL теоретически применимы для задач ИБ, но многое остается на уровне лабораторных экспериментов

Переход к практической реализации непростой, так как много сценариев и деталей

Полигоны и симуляторы решают исследовательские и учебные задачи, позволяют проверить гипотезы и выделить перспективные подходы

Возможные развития темы: применение больших языковых моделей, реализация объяснимости, маппинг событий высокого уровня на низкоуровневые действия

udv|group

Спасибо за внимание !

Вопросы ?

udv.group

Чернышов Юрий Юрьевич

к.ф.-м.н., доцент кафедры « АТиСУ» ИРИТ-РТФ УрФУ

заведующий лабораторией кибербезопасности NEO Lab ИРИТ-РТФ УрФУ

руководитель исследовательского центра UDV Group

LLM

PentestGPT: студент автоматизировал процесс взлома с помощью ChatGPTПодробнее: https://www.securitylab.ru/news/545263.php
https://github.com/GreyDGL/PentestGPT/blob/main/resources/README.md

Полезные ссылки

https://github.com/Limmen/awesome-rl-for-cybersecurity

CyRIS
https://github.com/cyb3rlab/CyRIS (старая версия тут: https://github.com/crond-jaist/cyris)
R. Beuran, C. Pham, D. Tang, K. Chinen, Y. Tan, Y. Shinoda, "Cybersecurity Education and Training Support System: CyRIS", IEICE Transactions on Information and Systems, vol. E101-D, no. 3, March 2018, pp. 740-749.