

# Белот бот с Deep Q-learning

Рая Литова 5МИ0800410

---

# Съдържание

1. Основни функционалности
2. Архитектура
3. Реализация
4. Проблеми и решения
5. Демо

# Основни функционалности

1

## Симулация на игра

Пълна имплементация на правилата на Белот (раздаване, валидни ходове, определяне на победител в трик, точкуване).

2

## Обучение

Агентът се обучава чрез изиграване на хиляди игри срещу себе си, оптимизирайки своята стратегия чрез награди (rewards).

3

## Конзолна игра срещу AI

Конзолен интерфейс, позволяващ на потребител да играе в отбор с AI срещу двама опоненти

4

## Запазване и зареждане

Възможност за съхраняване на обучените модели във файлове.

5

## Unity проект

Пълна имплементация на играта Белот, заедно с комуникация със сървър, на който се намира AI.

# Архитектура

1

## Logic layer

Дефинира правилата на играта

2

## State management

Съхранява информация за текущата ръка, изиграните карти и резултата.

3

## NN layer

Дефинира архитектурата на AI.

4

## RL layer

Обработват логиката на обучението и вземането на решения.

5

## Execution layer

Входна точка за стартиране на обучение или игра. Или сървър, приемащ заявки от Unity приложението.

# Реализация

## Използвани технологии и библиотеки

Python  
PyTorch  
NumPy  
Collections

## DQN алгоритъм

Decision Making  
Epsilon-Greedy Strategy  
Experience Replay

## Модел на данните

Собствена ръка (32 бита)  
Изиграни карти (32 бита)  
Текущо раздаване (32 бита)  
Обявление (6 бита)  
Текущ лидер (4 бита)

## Трениране

Self-play със случайно раздаване и обявление

# Проблеми и решения

Агентът играе силни карти безсмислено, когато отборът му губи и не се стреми към запазване на покритие на високи карти.

Добавено допълнително наказание за изиграване на карти със стойност  $\geq 10$  точки, ако разиграването не се печели от текущия играч или неговия партньор.

Невронната мрежа предлага невалидни ходове

Имплементирана е "маскировка", която задава стойност минус безкрайност на Q-стойностите за нелегални ходове, принуждавайки AI да избира само от позволените карти.

Агентът не се стреми към печелене на последната ръка в разиграването

Допълнителна награда при печелене на последната ръка.

# Проблеми и решения

Бавно учене на комплексни правила

Въвеждане на бонус от +50/-50 точки за крайна победа/загуба в цялото разиграване, за да се насърчи дългосрочната стратегия.

Агентът не разбира от стандартни сигнали на играча

Допълнителни награди при спазване на стандартни разигравания при сигнали и наказания в противен случай.

**ДЕМО**

Благодаря за  
вниманието!