

СРАВНЕНИЕ ЭФФЕКТИВНОСТИ ОБУЧЕНИЯ АГЕНТОВ В СОРЕВНОВАТЕЛЬНЫХ ЗАДАЧАХ ИГРОВОЙ ИНДУСТРИИ

И.Д. Курмакаев¹

¹Самарский университет, г. Самара
e-mail: ilham24.04.05@mail.ru

Аннотация

В работе исследуется специфика применения алгоритмов обучения с подкреплением (РРО) для создания неигровых персонажей в соревновательных видеоиграх с неполной информацией. Разработана тестовая среда с механиками скрытого аукциона и управления ресурсами. Проведена оценка эффективности обученной нейросетевой модели в сравнении с базовым эвристическим алгоритмом (Baseline). Результаты подтверждают высокую адаптивность агентов в конкурентных условиях и формируют базу для дальнейших исследований многоагентного самообучения (Self-Play).

Ключевые слова: машинное обучение, обучение с подкреплением, соревновательные задачи, неигровые персонажи, Unity ML-Agents.

Введение

Современная игровая индустрия активно внедряет методы машинного обучения для создания интеллектуальных агентов, способных составить конкуренцию игроку. Особую сложность представляют соревновательные игры с неполной информацией, где агенту необходимо не только оценивать математическое ожидание собственных действий, но и прогнозировать поведение оппонента в условиях скрытых данных [1, 2].

Целью данной работы является исследование эффективности обучения с подкреплением (Reinforcement Learning) в асимметричных соревновательных задачах и оценка качества выработанных агентом стратегий на базе платформы Unity ML-Agents.

1. Методология и описание соревновательной среды

Для проведения экспериментов была спроектирована пошаговая симуляция карточной игры, основанная на механике закрытых ставок. Соревновательный процесс разворачивается между двумя сущностями, обладающими общим пулом ресурсов (очками здоровья). Раунд состоит из слепого аукциона за право применения карты с варьирующимися свойствами (нанесение урона, восстановление здоровья, комбинированный эффект).

Для оценки эффективности нейросетевой модели был разработан базовый противник (Baseline) на основе эвристического автомата состояний, математически рассчитывающий оптимальную ставку в зависимости от текущего уровня здоровья и типа разыгрываемой карты. Нейросетевой агент получал информацию о состоянии среды через вектор наблюдений размерностью 4 (здоровье участников, сила и тип карты).

2. Оценка эффективности обучения и анализ результатов

Обучение модели проводилось с применением алгоритма Proximal Policy Optimization (PPO) [3]. Наличие в среде скрытой информации (ставки противника) и вариативности карт создавало риск сходимости агента к локальному оптимуму (пассивной стратегии с нулевыми ставками). Для решения проблемы был применен метод Curriculum Learning, а также скорректирован коэффициент исследования (beta).

В ходе эксперимента агент обучался противостоять базовому эвристическому алгоритму. График изменения метрики средней накопленной награды (Cumulative Reward) продемонстрирован на рисунке 1.

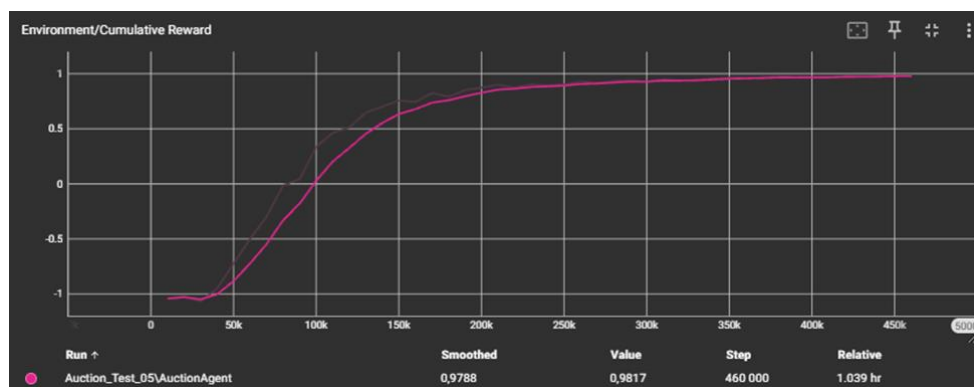


Рисунок 1 – Динамика средней награды в процессе обучения

Представленный график иллюстрирует процесс адаптации агента к конкурентной среде. После этапа первичного исследования (до 100 000 шагов), характеризовавшегося высокой дисперсией и отрицательной наградой, алгоритм PPO позволил сформировать контр-стратегию. Начиная с 150 000 итераций, эффективность агента стабилизировалась на отметке 0.97.

Для верификации результатов был проведен симуляционный турнир (1000 партий) в режиме вывода (Inference Only). Анализ собранных логов показал, что ML-агент одержал победу в 99.4% матчей. Модель научилась безошибочно классифицировать ценность карт (в частности, применяя максимальную агрессию при выпадении карт «Вампиризм»), превосходя жестко заданную логику оппонента.

Заключение

Результаты исследования доказывают высокую эффективность применения методов обучения с подкреплением в соревновательных игровых задачах. Агент успешно выработал стратегии адаптации к логике противника в условиях неполной информации.

Вместе с тем, обучение против статического алгоритма ограничивает эмерджентность поведения агента. Перспективным развитием данного проекта является применение технологии симметричного многоагентного самообучения (Self-Play) [4]. Исключение эвристического алгоритма и обучение агентов в конкурентной борьбе друг против друга потенциально позволит выработать более сложные поведенческие паттерны, недостижимые при текущем подходе.

Список использованных источников

1. Champandard, A. J. AI Game Development: Synthetic Creatures with Learning and Reactive Behaviors / A. J. Champandard. – Indianapolis : New Riders, 2003. – 720 p.
2. Sutton, R. S. Reinforcement Learning: An Introduction / R. S. Sutton, A. G. Barto. – Cambridge : MIT Press, 2018. – 552 p.
3. Proximal Policy Optimization Algorithms / J. Schulman, F. Wolski, P. Dhariwal [et al.] // arXiv.org. – 2017. – URL: <https://arxiv.org/abs/1707.06347> (дата обращения: 10.02.2026).
4. Unity: A General Platform for Intelligent Agents / A. Juliani, V. Berges, E. Teng [et al.] // arXiv.org : [сайт]. – 2018. – URL: <https://arxiv.org/abs/1809.02627> (дата обращения: 06.03.2026).

COMPARISON OF AGENTS TRAINING EFFICIENCY IN COMPETITIVE TASKS OF THE GAMING INDUSTRY

I.D. Kurmakaev¹

*¹Samara University, Samara
e-mail: ilham24.04.05@mail.ru*

Abstract

The paper investigates the efficiency of reinforcement learning (PPO) algorithms for non-player characters in competitive video games with imperfect information. A test environment with a blind auction mechanic was developed. The evaluation of the trained neural network model against a baseline heuristic algorithm confirms the high adaptability of ML agents in competitive conditions, forming a basis for future self-play research.

Keywords: machine learning, reinforcement learning, competitive tasks, non-player characters, Unity ML-Agents.