Разработка и исследование алгоритмов обучения нейронных сетей с подкреплением в области игровой индустрии

Д.И. Ульянов[1], Д.А. Савельев[1,2]

[1]*Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086*
[2]*Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001*

Аннотация

В работе проведена проверка гипотезы о том, что искусственные нейронные сети с одинаковой архитектурой, обученные разным игровым аспектам, показывают результаты, превосходящие универсального агента, обладающего той же архитектурой в рамках своего аспекта. Обучение с подкреплением нейронных сетей производилось по алгоритму глубокого Q-обучения в трех вариациях игровой среды для экономического и боевого аспектов, а также для стандартных условий игры. Полученные результаты подтверждают гипотезу в рамках заявленной игровой среды.

Ключевые слова

Q-обучение, глубокие нейронные сети, StarCraft 2

1. Введение

Как правило, положением вещей в игровой среде управляет один агент, часто представляемый искусственной нейронной сетью [1]. Но в реальности в одиночку сложно эффективно управлять сложными системами, и наблюдается разделение на подсистемы. Соответственно, логичным видится разделить аспекты взаимодействия с игровой средой для повышения эффективности и улучшения получаемого результата.

Сформулируем гипотезу следующим образом: нейронные сети с одинаковой архитектурой, обученные различным аспектам игровой среды, показывают результат, качественно превосходящий универсального агента с той же архитектурой в рамках каждого рассматриваемого аспекта. В данной работе проведено исследование по проверке выше сформулированной гипотезы в игровой среде StarCraft 2 с искусственной нейронной сетью в качестве реализации агентов.

2. Архитектура искусственной нейронной сети и редактирование карты

Игровая среда StarCraft 2 позволяет работать с данными двух типов: данные, предоставляемые интерфейсом обучения с подкреплением и изображения игровой среды от лица игрока [2]. Учитывая данный факт, разобьём нейронную сеть на три составные части: выделение признаков изображения игровой среды, анализ признаков от интерфейса игровой среды и объединение этих данных с принятием решения о действии. Части взаимодействуют таким образом, что при получении данных, каждый набор отправляется в свой сегмент сети. После обработки выходы передаются в скреплённом виде на вход сегмента, ответственного за окончательное принятие решения.

Функция ошибки распространения имеет следующий вид:

$$loss = \frac{1}{2}\left(\alpha(r(s',a) + \gamma \max_a Q\left(s',a,\hat{\theta}\right)) - Q(s,a,\theta)\right), \quad (1)$$

где θ – веса сети; $\hat{\theta}$ – веса зафиксированной сети; s' – следующее состояние; s – текущее состояние; a – действие, α – скорость обучения; γ – скидочный фактор (влияние уже закреплённой политики при формировании новой).

За основу игрового уровня для экспериментов была взята карта Simple64. Для выработки и оценки боевого аспекта агента, из неё были изъяты все элементы экономического аспекта: здания, возможности добывать ресурсы и проводить улучшения имеющихся юнитов. Для экономического аспекта была убрана возможность создавать и улучшать боевые единицы всех типов. Условие победы в экономическом аспекте – большее суммарное число имеющихся ресурсов по итогу 50 минут игровой партии.

3. Обучение сетей и проведение экспериментов

Перед формированием глубокой Q-функции согласно алгоритму Q-обучения, необходимо провести подготовку на записях партий [3]. Для этой цели был сгенерирован набор из 50 записей игры ботов с максимальным уровнем сложности друг против друга.

Предобученную нейронную сеть продолжим обучать следующим образом: на протяжении партии против бота на максимальной сложности с вероятностью 0.05 совершается случайное действие, а с вероятностью 0.95 – сгенерированное на основе входа нейронной сети. Во втором случае состояние, в которое перейдёт агент, помещается в специальный буфер, а нейронная сеть дублируется, и веса дубликата фиксируется. Из этого же буфера берётся состояние, взятое случайное количество шагов назад и рассчитывается ошибка между результатом зафиксированной и активной нейронных сетей, которая применяется для обновления весов согласно формуле (1). Результаты эксперимента по сравнению универсального и боевого агентов представлены в таблице 1.

Таблица 1

Доли побед агентов над ботами различных уровней сложности по результатам 100 партий, %

Агент/Бот	Лёгкий уровень	Средний уровень	Сложный уровень
Универсальный	100	91	8
Боевой	100	95	89

При проведении 100 партий универсального и боевого агента по правилам предыдущего эксперимента, доля побед боевого агента составила 61%.

Экономический агент, с учётом заявленных для него условий победы над универсальным агентом, имеет долю побед равную 100%, что говорит о неоспоримом преимуществе над универсальным агентом.

4. Заключение

На основе анализа полученных результатов можно сделать вывод, что искусственные нейронные сети, обладающие одинаковой архитектурой и обученные различным аспектам игры, действительно превосходят по качественным показателям универсальную сеть в рамках своего аспекта для игровой среды StarCraft 2.

5. Литература

[1] Taylor, M.E. Reinforcement learning agents providing advice in complex video games / M.E. Taylor // Connection Science. – 2014. – Vol. 26(1). – P. 45-63.

[2] The StarCraft II API [Электронный ресурс]. – Режим доступа: https://github.com/Blizzard/s2client-proto (2.10.2020).

[3] Liu, R. The effects of memory replay in reinforcement learning / R. Liu // Annual Allerton Conference on Communication, Control, and Computing. – 2018. – Vol. 56. – P. 478-485.