



НАУЧНЫЙ ЖУРНАЛ НАУКА И МИРОВОЗЗРЕНИЕ

АНАЛИЗ МЕТОДА ГЛУБОКОГО МУЛЬТИАГЕНТНОГО ОБУЧЕНИЯ НА ОСНОВЕ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ

Еремичев Константин Андреевич

студент, Московский государственный технический университет им. Н.Э. Баумана, РФ, г. Москва

Алфимцев Александр Николаевич

научный руководитель, д-р. техн. наук, профессор, Московский государственный технический университет им. Н.Э. Баумана, РФ, г. Москва

В рамках данной статьи будет описан процесс исследования концепции обучения агента при помощи генеративно-состязательных нейронных сетей.

На фоне быстро протекающего процесса всеобъемлящей автоматизации и информатизации большое распространение получил особый тип информационных систем, называемых нейронными сетями. Огромные массивы данных, требующие быстрой обработки и анализа, а также отсутствие ярко-выраженных взаимосвязей между ними приводят к тому, что в некоторых ситуациях классические алгоритмы обработки данных терпят неудачу: к примеру, написание качественной программы для распознавания образов с использованием процедурного подхода является практически неподъемной задачей. С данной задачей намного лучше справляются нейронные сети.

В современном мире существует большое количество различных типов нейронных сетей, а также моделей их обучения, на которые в данной статье и будет сосредоточено основное внимание. В рамках научного исследования был изучен сравнительно недавно описанный метод обучения генеративно-состязательных нейронных сетей. Для проведения анализа работоспособности и качественной оценки работы данного метода обучения была разработана тестовая обучающая модель для нейронной сети, имитирующей поведение игрового агента в компьютерной игре Starcraft, разработанной компанией Blizzard.

1. Исследовательская часть

В данной части статьи будут введены основные концепции, термины и понятия, необходимые для лучшего понимания сути дальнейшей работы. Отдельные подразделы будут посвящены математическим моделям, используемым при моделировании состязательных и игровых процессов, моделям обучения агентов, а также генеративно-состязательным сетям

1.1 Математическая модель процесса принятия решений Маркова

Одним из традиционных подходов, описывающих взаимодействие отдельных агентов со стационарной окружающей средой, является модель процесса принятия решений Маркова. В силу своей простоты и понятности данная модель будет использована при моделировании агентов в данной статье. Далее будет приведено ее краткое описание.

- Состояния среды;
- Вероятностное распределение, характеризующее способность среды переходить из одного состояния в другое;
- Действия, которые может выполнять агент в различных состояниях среды;
- Политики (вероятностные распределения, характеризующие поведение агента в том или ином состоянии среды);
- Функция вознаграждения.

Функция вознаграждения является одним из важнейших понятий процесса принятия решений Маркова, ее основная цель состоит в поощрении определенного поведения агента и штрафах агента за несоответствующее поведение. Процесс принятия решений в данной опирается на условие того, что в каждой, описываемой одноагентной системе существует политика, способная привести агента к максимальному вознаграждению, а целью каждого агента является нахождение данной оптимальной политики.

1.2 Основы обучения агентов

Одной из традиционных моделей, описывающих процесс выбора агентом наилучшей политики для максимизации получаемого вознаграждения, является модель обучения с подкреплением. Основной задачей обучения с подкреплением является выбор политики, которая сделает действия агента наиболее выгодными в данной среде при данных правилах взаимодействия. Чем больше получаемое вознаграждение, тем выгоднее считается подобранная политика.

Для качественного обоснования выбора наиболее выгодной политики данная модель вводит следующие понятия:

- Value function (от англ. функция значения) $V(s)$. Данная функция характеризует математическое ожидание вознаграждения, получаемого агентом за всю игру, в случае, если агент начнет ее из заданного состояния;
- Q function (кью функция) $q(s,a)$. Данная функция характеризует математическое ожидание вознаграждения, получаемого агентом за всю игру, в случае, если агент начнет взаимодействие со средой из заданного состояния и произведет в нем заданное действие. По данным функциям при помощи уравнения Белмана, есть возможность восстановить траекторию действий агента, приводящую к максимальному вознаграждению.

1.3 Введение в генеративно-сопоставительные сети

Модели, описанные в подразделе 1.2 могут быть неявно использованы для обучения агента, модель которого представляет собой нейронную сеть. На вход данной модели будут подаваться значения переменных среды, на выходах будут устанавливаться вероятности выбора того или иного действия при заданных состояниях среды. Таким образом задача обучения агента, не имеющего доступа к функции вознаграждения, по действиям эксперта превращается в задачу обучения нейронной сети для повторения поведения эталонного агента.