Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский университет «Высшая школа экономики»

Факультет Санкт-Петербургская школа физико-математических и компьютерных наук
Машинное обучение и анализ данных
(название ОП)
магистратура
(уровень образования)

О Т Ч Е Т
по *проектной* практике
(указать вид практики)
проект
(тип практики (наименование ЭПП)
_____ Симуляция фондового рынка на RL агентах _____
(название проекта)

Выполнил студент гр.__ММО221С

Шахвалиева Юлиана Сергеевна
(ФИО)

_____
(подпись)

Проверил:

Шпильман Алексей Александрович ___
(должность, ФИО руководителя ЭПП)

_____
(подпись)

__*15.06.2023*_____
(дата)

СОДЕРЖАНИЕ

ОБЩЕЕ ОПИСАНИЕ ПРОЕКТА

- Руководитель проекта: Шпильман Алексей Александрович
- Тип проекта: исследовательский
- Место работы по проекту: НИУ ВШЭ в Санкт-Петербурге

СОДЕРЖАТЕЛЬНАЯ ЧАСТЬ

1.1 Общее описание проекта

Цель данного проектного задания – создание мульти-агентной среды, на которой будут торговать агенты, совокупное поведение которых должно имитировать реальную биржу. Для оценки степени схожести созданной симуляции с реальной биржей вводится дискриминатор, которому на вход подается как сгенерированный, так и настоящий стаканы. Цель дискриминатора – отличить стаканы, цель всех агентов – научиться генерировать стаканы, максимально похожие на реальные, чтобы дискриминатор ошибался. У каждого отдельного агента не стоит цель приумножения имеющегося капитала, так как среда закрытая, а значит является игрой с нулевой суммой.

В рамках данной работы сделаны следующие допущения:

1. Время дискретно, агенты совершают действия в каждый момент времени одновременно;
2. Биржа работает без перерывов;
3. На рынке существует один актив;
4. Биржа допускает подачу только одной заявки, то есть не принимается заявка от агента, если не закрыта предыдущая, при этом есть возможность отменить ранее поданную заявку.

Каждый агент подает заявку на покупку/продажу определённого количества актива по фиксированной цене. Если в какой-то момент времени на бирже появляются совпадающие заявки, то есть один агент готов купить по одной цене, а другой готов продать по этой цене, то заявки исполняются биржей, при этом меняется баланс активов и денежных средств агентов, участвовавших в заявке.

Среда принимает на вход список действий каждого агента и возвращает состояние, награды и флаги завершения эпизода для каждого из агентов. С помощью этой информации агенты обучаются алгоритмом PPO. Начальный капитал – количество денег и активов – каждого агента определяется случайно. Эпизод для каждого агента завершается в случае, если и активы, и денежные средства равны нулю.

1.2 Пространство состояний

Пространство состояний для каждого агента состоит из двух частей: информация об агенте и текущий стакан, которые одинаков для всех агентов.

В информацию об агенте входят: количество имеющихся у него ресурсов - денег и активов, - флаг наличия заявки на бирже, количество дней, которое заявка находится на бирже, а также цена и объем заявки. В случае, если у агента нет поданной заявки, то возвращаются служебные символы.

Стакан – это список из k заявок на продажу, и k заявок на покупку, где k – гиперпараметр. Каждая заявка описывается ценой и объемом. Если объем меньше нуля, то это заявка на продажу, иначе – на покупку. Если в данный момент заявок какого-либо типа меньше, чем k, то заявки дополняются нулями до нужного количества. В случае, когда заявок больше k, то берутся k заявок с наименьшей ценой, то есть заявоки, цена которых ближе всего к равновесной цене на рынке.

Стакан отсортирован следующим образом: сначала идут k заявок на покупку (с положительным значением объема) в порядке убывания цены, после идут k заявок на продажу (с отрицательным значением объема) в порядке возрастания цены.

1.3 Пространство действий

У каждого агента есть три возможных действия:

- подать заявку;
- отменить ранее поданную заявку;
- пропустить шаг.

Последнее действие добавлено с целью приближения симуляции к реальному миру, где не все торговые агенты совершают действие единовременно, тем не менее, у них всегда есть доступ к текущему положению дел на фондовом рынке.

Пространство действий континуально, выражено в трех числах в диапазоне $[-1, 1]$. Список таких действий попадает от агентов среду. Далее среда преобразует эти числа в конкретные действия из вышеприведенного списка.

Преобразование происходит следующим образом: номер наибольшего по модулю числа определяет тип действия: 0 – подача заявки, 1 – пропуск хода, 2 – отмена заявки. В случае, если максимальное по модулю число находится на нулевой позиции, то есть должна быть подана заявка, то остальные два числа используются для определения цены и объема.

1.4 Шаг среды и функция награды

Шаг среды состоит из 3 этапов, и на каждом агент получает награду. Итоговая награда, возвращаемая средой, - сумма наград на каждом шаге. Рассмотрим подробнее реализацию одного шага среды:

1) Преобразование действия и его проверка

На этом этапе на вход поступает от каждого агента список из трех чисел, которые преобразуются в действие способом, описанном в предыдущем разделе. После преобразования происходит проверка доступности действия: не пытается ли агент отменить заявку, которой не было или подать вторую заявку, хватает ли ресурсов у агента для исполнения данной заявки. В случае не выполнения данных требований, заявка не попадает в стакан, а агенту возвращается штраф в размере -10.

Если действие – пропуск хода, то агенту назначается награда в размере -2, дальше это действие не проходит. Если действие – отмена заявки, то заявка удаляется из стакана, а агенту возвращается награда в размере -1, в следующие этапы это действие также не проходит.

Агентам, подавшим заявки на покупку или продажу, которые успешно прошли проверку, назначается награда в размере $+2$, и эти заявки помещаются в стакан.

2) Оценка дискриминатора

На данном этапе происходит взаимодействие с дискриминатором. Дискриминатор – нейронная сеть для бинарной классификации, которая пытается отличить реальный стакан от сгенерированного. Реальный стакан берется из заранее подготовленных данных.

Данные были скачаны с сайта Московской Биржи, они предоставляют собой образцы недельных исторических данных «Full Order Book - тип А» для тестирования и анализа глубины и ликвидности своих рынков. Данные содержат все сделки и все заявки за период 1 – 5 сентября 2014 года по акциям Сбербанка. Данные выглядят следующим образом:

	DATE	NO	SECCODE	BUYSELL	TIME	ORDERNO	ACTION	PRICE	VOLUME	TRADENO	TRADEPRICE
0	20140901	2	SBER	S	100000000	2	1	74.18	1500	NaN	NaN
1	20140901	3	SBER	S	100000000	3	1	74.75	1500	NaN	NaN
2	20140901	33	SBER	B	100000000	33	1	73.16	500	NaN	NaN
3	20140901	57	SBER	B	100000000	57	1	71.33	6000	NaN	NaN
4	20140901	83	SBER	S	100000000	83	1	74.80	70	NaN	NaN
5	20140901	84	SBER	S	100000000	84	1	74.40	60	NaN	NaN
6	20140901	85	SBER	S	100000000	85	1	74.00	50	NaN	NaN
7	20140901	86	SBER	S	100000000	86	1	73.60	40	NaN	NaN
8	20140901	87	SBER	B	100000000	87	1	72.80	10	NaN	NaN
9	20140901	101	SBER	B	100000000	101	1	72.60	3300	NaN	NaN

Они были разбиты на две части: заявки на покупку и заявки на продажу. Так как данные представляют собой временной ряд, то реальный стакан формировался следующим образом: брался случайный срез из заявок на покупку длины k, и соответствующий ему срез из заявок на продажу. Данные срезы сортировались по возрастанию/убыванию цены тем же способом, что и в сгенерированном стакане.

На вход дискриминатору подается 7 реальных стаканов и 1 сгенерированный. В зависимости от того, счел ли дискриминатор сгенерированный стакан реальным, агенты награждаются. Награждение на этом этапе происходит равномерно: каждому агенту одинаковая награда, вне зависимости от совершенного им действия. Это происходит потому, что дискриминатор оценивает не качество конкретного агента, а то, как все агенты взаимодействуют между собой.

Награда рассчитывается следующим образом:

$$\text{награда} = 10 * (2 * p - 1),$$

где p – вероятность того, что стакан реальный, полученная на выходе дискриминатора.

Таким образом в случае, если агентам удалось обмануть дискриминатора, их награда будет положительная, а зависит она от степени уверенности агента. В случае, если дискриминатор наверняка уверен ($p = 1$), что сгенерированный стакан – реальный, то все агенты получат максимально возможную награду: $+10$.

Так как на первых итерациях агенты инициализированы случайно и сгенерированные стаканы сильно отличаются от реальных, то дискриминатор очень быстро обучается. Это приводит к невозможности дальнейшего обучения агентов. Для

решения данной проблемы к данным, попадающим на вход дискриминатора, добавляется случайный шум.

3) Торговля

На данном этапе происходит исполнение заявок, находящихся в стакане. Алгоритм, следующий: перебираем все заявки в стакане в случайном порядке, фиксируем исполняемую заявку. Для нее пробегаем по всему стакану, пытаемся исполнить заявку по цене из стакана. То есть, если текущая заявка на продажу по одной цене, а в стакане нашлась заявка на покупку по цене, которая не больше, то происходит торговля по цене из стакана, если текущая заявка на покупку, то алгоритм аналогичный. При этом меняются ресурсы у агентов, участвующих в торговле.

Агенты, чьи заявки были исполнены награждаются по следующей формуле:

$$\text{награда} = \frac{10}{\text{кол} - \text{во дней на бирже}}$$

Таким образом, если заявка будет исполнена в день подачи, то агент получит $+10$ в качестве награды. Те заявки, которые не удалось исполнить, остаются в стакане, при этом у них инкрементируется счетчик дней, проведенных на бирже.

1.5 Описание результатов проекта

Целью данного проекта являлось создание симуляции фондового рынка на RL агентах. Для достижения поставленной цели была разработана пользовательская мульти-агентная RL- среда, реализующая работу биржевого стакана. В среду был добавлен функционал для визуализации ее работы. Также был построен дискриминатор, призванный оценивать качество получаемой симуляции.

На данной среде были обучены RL-агенты с помощью алгоритма PPO. Были проведены эксперименты по настройке гиперпараметров, архитектур нейронных сетей, в ходе которых удалось симулировать поведение фондовой биржи с помощью торговых агентов.

ЗАКЛЮЧЕНИЕ

Во время работы над проектом были развиты и приобретены следующие навыки и компетенции:

- Проектирование и осуществление всех этапов жизненного цикла проекта, начиная от разработки ТЗ и заканчивая тестированием полученной симуляции;
- Проектирование и разработка пользовательской RL-среды;
- Обучение мульти-агентной RL-среды;
- Работа с генеративно-состязательной концепцией.