



Факультет Санкт-Петербургская школа физико-математических и компьютерных наук

Машинное обучение и анализ данных

Симуляция фондового рынка на RL агентах

Подготовила: Шахвалиева Юлиана Сергеевна

Руководитель: Шпильман Алексей Александрович

Введение в предметную область

Фондовый рынок — совокупность экономических отношений по поводу выпуска и обращения ценных бумаг между его участниками

Биржевой стакан — это таблица лимитных заявок. Каждая заявка содержит цену и количество

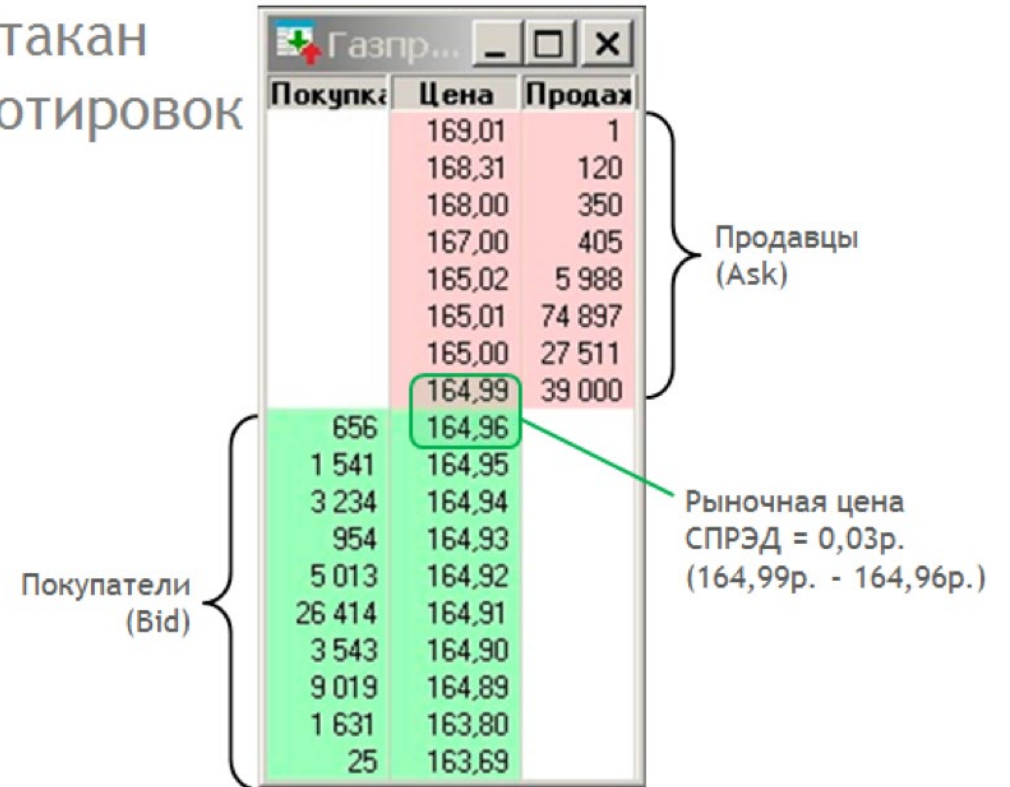
Обучение с подкреплением (RL) - это метод машинного обучения, в котором наша система (агент) обучается методом проб и ошибок

Важность изучения

Изучение фондового рынка и его симуляция имеют **практическое и исследовательское значение**:

- Инвесторы могут тестировать свои торговые стратегии на созданной симуляции
- Исследователи могут исследовать реакцию рынка на различные события, например, на гиперинфляцию

Стакан котировок



Покупка	Цена	Продажа
	169,01	1
	168,31	120
	168,00	350
	167,00	405
	165,02	5 988
	165,01	74 897
	165,00	27 511
	164,99	39 000
656	164,96	
1 541	164,95	
3 234	164,94	
954	164,93	
5 013	164,92	
26 414	164,91	
3 543	164,90	
9 019	164,89	
1 631	163,80	
25	163,69	

Продавцы (Ask)

Покупатели (Bid)

Рыночная цена
СПРЭД = 0,03р.
(164,99р. - 164,96р.)

Цель и задачи проекта

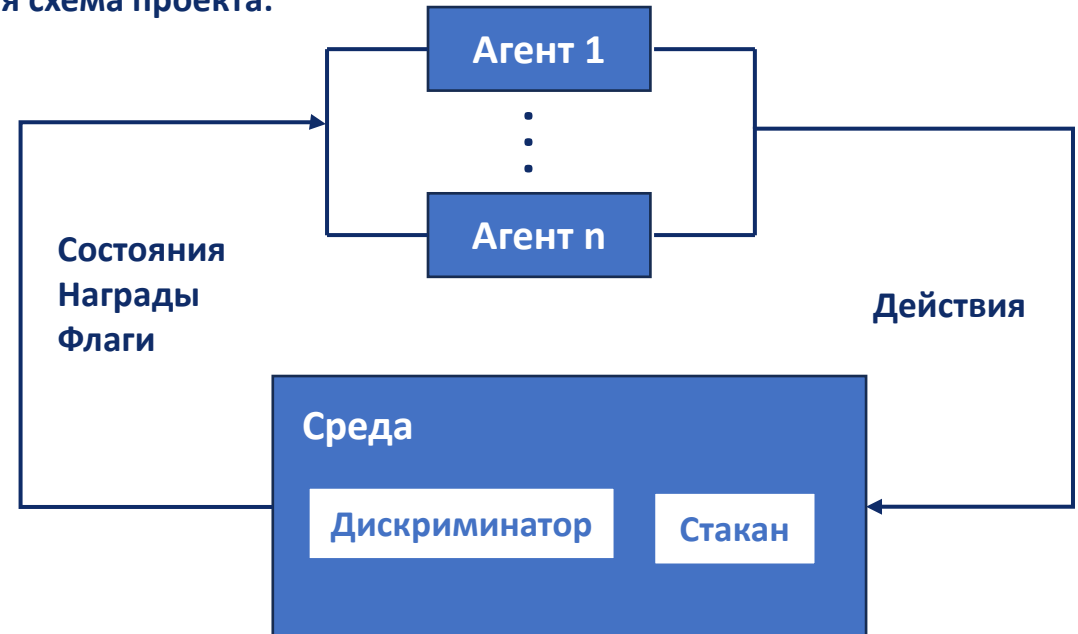
Цель: создание мульти-агентной среды, на которой будут торговать агенты, совокупное поведение которых должно имитировать реальную биржу

Оценка степени схожести созданной симуляции с реальной биржей проводится с помощью **дискриминатора**

Задачи:

- **Определение** пространства действий, состояний, а также функции награды
- **Реализация** логики работы **биржевого стакана**
- **Сбор и обработка данных** для дискриминатора
- **Реализация торговых агентов** и рендеринга среды
- **Обучение** и демонстрация работы симуляции

Общая схема проекта:



Допущения:

- **Время дискретно**, агенты совершают действия в каждый момент времени одновременно
- Биржа работает **без перерывов**
- На рынке существует **один актив**
- Биржа допускает подачу только **одной заявки**



Пространство действий

У каждого агента есть **три возможных действия**:

0. подать заявку
1. пропустить ход
2. отменить ранее поданную заявку

Пространство действий **континуально**, выражено в трех числах в диапазоне $[-1, 1]$

Преобразование в действия:

- Тип действия: номер наибольшего по модулю числа
- Если «подать заявку» - остальные два числа используются для определения цены и объема (если > 0 , то покупка, иначе – продажа)

Пространство состояний

▪ Информация об агенте

- флаг наличия заявки на бирже
- количество итераций, проведенных на бирже
- цена и объем заявки
- ресурсы

Если у агента нет поданной заявки, то возвращаются служебные символы

▪ Текущий стакан

- k заявок на покупку (объем > 0)
- k заявок на продажу (объем < 0)

k – длина стакана

Если заявок не хватает, то стакан дополняется нулями



Шаг среды и функция награды

1 Преобразование действия + проверка

Действие	Награда	Реакция среды
Отмена несуществующей заявки	-5	Штраф
Повторная заявка	-5	Штраф
Нехватка ресурсов	-5	Штраф
Пропуск хода	-3	Возврат стакана
Отмена существующей заявки	-1	Удаление заявки из стакана
Подача заявки	+1	Добавление заявки в стакан

2 Оценка дискриминатора

- Нейронная сеть для бинарной классификации
- На вход 7 реальных стаканов + 1 сгенерированный
- Данные с сайта Московской биржи по всем сделкам и заявкам по акции Сбербанка
- Зашумление входных данных
- Награда:

$$10 * (2 * p - 1)$$

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

3 Торговля

- Для каждой заявки проходимся по всему стакану и пытаемся ее исполнить по цене из стакана
- Если текущая заявка на продажу по цене m , а в стакане нашлась заявка на покупку по цене $\leq m$, то происходит торговля
- Награда при закрытии заявки:

5

кол — во итераций на бирже

Агенты получают награду на каждом шаге. Итоговая награда – сумма наград на каждом шаге среды.



Обучение и демонстрация

- Начальное количество денег и активов каждого агента **инициализируется случайно**
- Эпизод заканчивается, когда **все ресурсы агента равны нулю**
- При обучении проводится **фиксированное количество итераций торговли**, после чего происходит обновление агентов
- Агенты обучались с помощью **алгоритма Proximal Policy Optimization (PPO)**, так как он подразумевает всего 2 нейронные сети, тем не менее показывает высокие результаты

Результаты

- Спроектирована и разработана **мульти-агентная RL-среда**
- Разработана **визуализация работы** среды с помощью pygame
- **Обучена симуляция** на 10 агентах, с длиной стакана 3

Биржевой стакан		
Продажа	Цена	Покупка
	100.0	1000
	41.66	866
	37.03	175
211	95.98	

