

Тема магістерської роботи: Застосування методів навчання з підкріпленням для балансування інвестиційних портфелів

Студент: Руденко Владислав Валерійович
Керівник: Олецкий Олексій Віталійович

Мета, об'єкт і предмет дослідження

Мета дослідження: порівняти RL-агентів зі статичними портфельними політиками та перевірити, чи дають інформаційно керовані бари вимірну перевагу над часовими барами.

Об'єкт: динамічне ребалансування криптовалютного портфеля.

Предмет: ефективність RL-агентів за різних просторів спостережень.

Дослідницьке питання: чи допомагають інформаційно керовані бари й складніші моделі покращити якість портфельної політики?

Відомі підходи до задачі

Основні підходи:

- класична портфельна оптимізація
- статичні ризикові політики, зокрема HRP
- портфельне навчання з підкріпленням
- часові та інформаційно керовані бари

Проблема: потрібне чесне порівняння в однаковому середовищі.



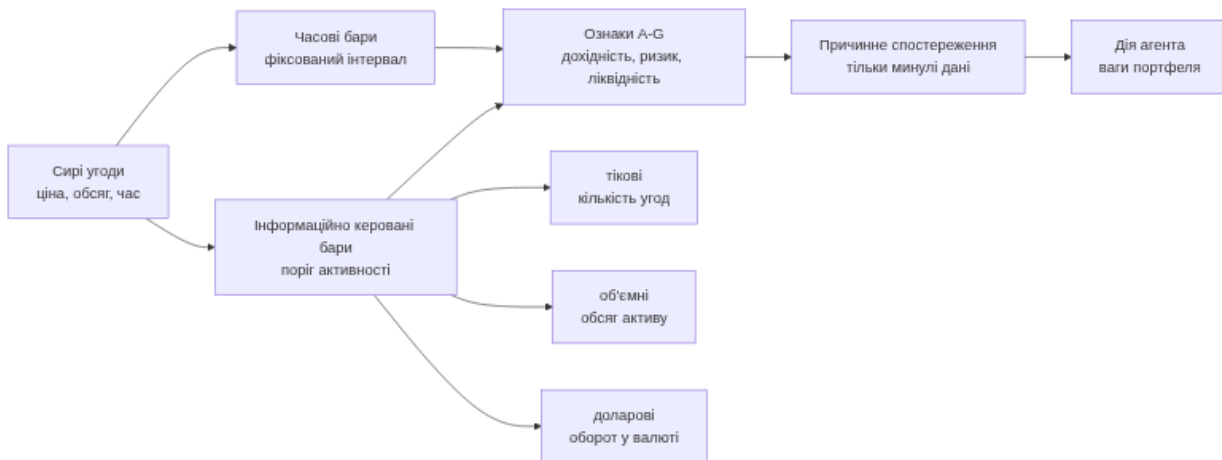
Простір спостережень: часові та інформаційні бари

Часові бари:

- закриваються за фіксований інтервал
- прості для інтерпретації
- можуть змішувати активні й спокійні періоди

Інформаційно керовані бари:

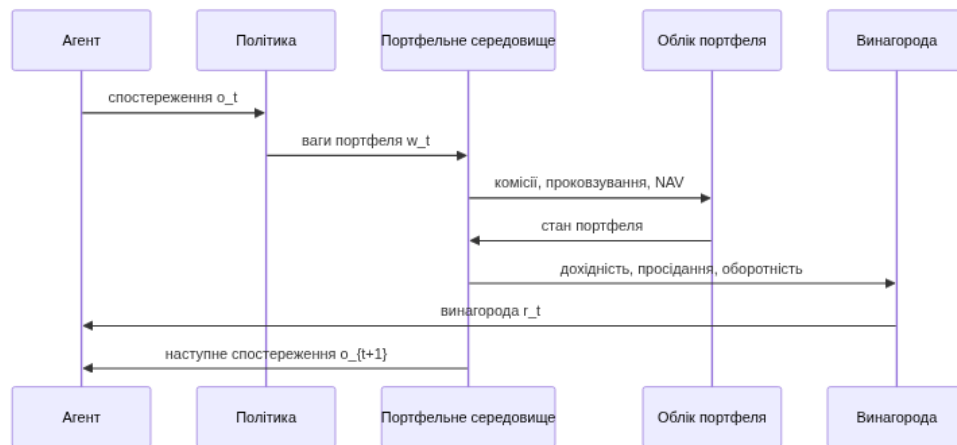
- **тікові:** кількість угод
- **об'ємні:** базовий обсяг
- **доларові:** торговий оборот



Як формується рішення агента

На кожному кроці:

- формується причинне спостереження
- агент обирає ваги портфеля
- середовище враховує ринок і витрати
- винагорода враховує дохідність, ризик і оборотність



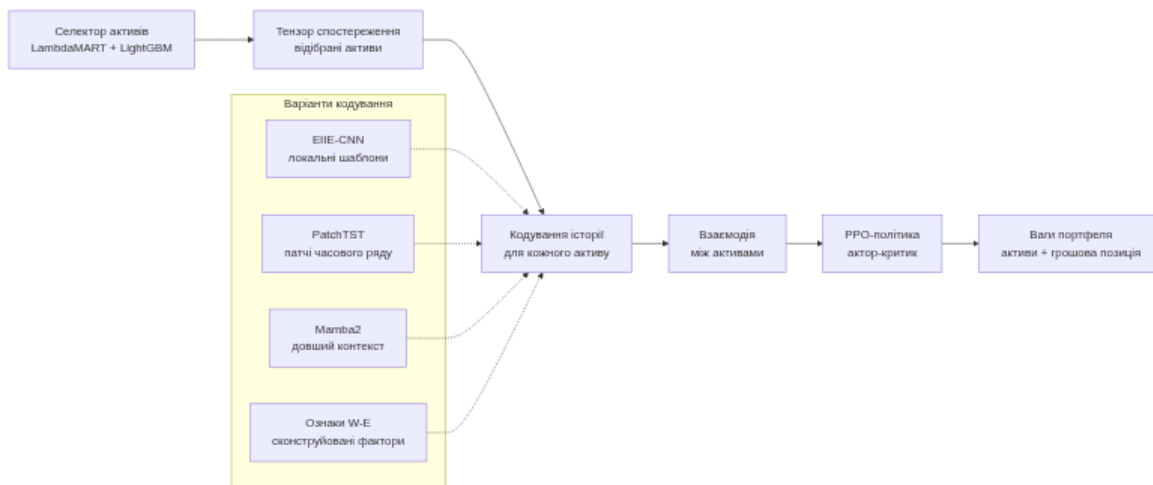
Інші приклади застосування в межах експерименту

Два рівні моделювання:

1. Селектор активів:
LambdaMART / LightGBM
ранжує кандидатів.
2. RL-алокатор: PPO навчає
політику портфельних ваг.

Блоки кодування:

- **ЕІЕ-CNN**: локальні часові шаблони
- **PatchTST**: патчі часового ряду
- **Mamba2**: довший послідовний контекст
- **iTransformer**: взаємодія між активами



Тренування та оцінювання

Протокол:

- послідовні історичні розбиття
- окремі навчальна, валідаційна й тестова частини
- однаковий селектор у порівнюваних запусках
- RPO: 65 536 кроків у базових конфігураціях

Метрики:

- коефіцієнт Шарпа
- максимальне просідання
- оборотність
- частка O7

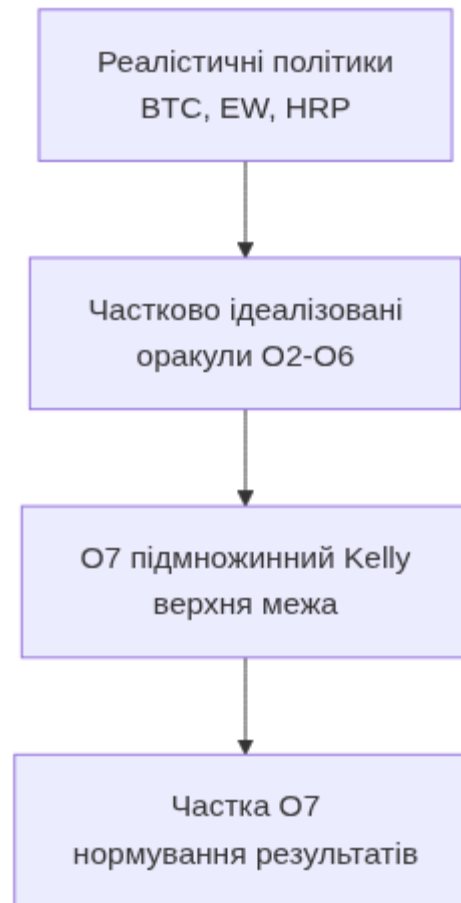
Реальна задача та базові орієнтири

Дані:

- Binance Spot
- USDT-пари
- 2017-2026
- денне ребалансування

Базові орієнтири:

- утримання BTC
- рівноваговий портфель
- HRP: ієрархічний розподіл ризику
- оракульна шкала O0-O7



Результат: простір спостережень

Перевірено:

Результат: лише часові бари проти скороченої багатобарової конфігурації.

- $\Delta \text{Sharpe} = -0.066$
- 95% CI: $[-0.130; -0.007]$
- виграші багатобарового варіанта: 40%

Висновок: скорочена багатобарова конфігурація не покращила якість у цій постановці.

Результат: порівняння моделей

Архітектурне порівняння:

- EIE-CNN: коефіцієнт Шарпа = -0.005
- Mamba2/iTransformer: коефіцієнт Шарпа = -0.002

Різниця: 0.003

Складніша модель має приблизно 14× оборотність

Висновок: простіший агент залишається конкурентним.

Результат: практичний орієнтир

Провідний RL-агент: коефіцієнт Шарпа = -0.002.

Найкраща статична політика HRP: коефіцієнт Шарпа = 0.217.

Практичний результат:

- отримано відтворюваний орієнтир
- HRP є сильним орієнтиром
- RL-постановка потребує сильнішого навчання й ознак

Наукова новизна та практична значимість

Науковий результат:

- контрольоване порівняння RL у різних просторах спостережень
- порівняння архітектур за однакових умов
- єдина шкала базових політик та оракулів

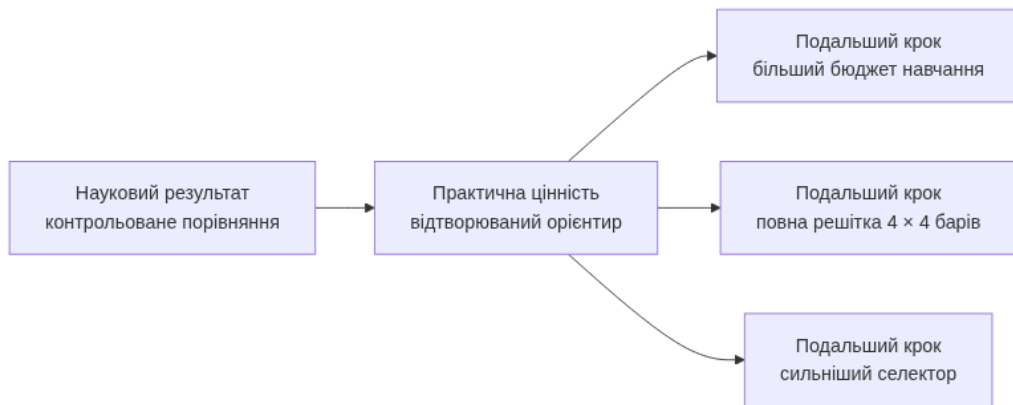
Практична значимість:

- відтворюваний протокол оцінювання
- сильний статичний орієнтир
- основа для наступних RL-експериментів

Перспективи розвитку

Перспективи:

- більший бюджет навчання
- повна 4×4 решітка барів
- сильніший селектор
- мікроструктурні ознаки



Дякую за увагу!

Питання?