Deep Reinforcement Learning applied to Statistical Arbitrage Investment Strategy on Cryptomarket

Presented by:
Gabriel Vergara Schifferli

UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

7 Septiembre, 2023

Name of the student:
Gabriel Vergara[1,2]

Thesis advisor:
Werner Kristjanpoller[1]

Thesis coadvisor:
Pedro Gajardo[2]

External Examiner:
Javier Mella[3]

[1] Departamento de Industrias, Universidad Técnica Federico Santa María, Chile.
[2] Departamento de Matemática, Universidad Técnica Federico Santa María, Chile.
[3] Universidad de los Andes, Chile

Contenido

Introducción

Introducción

- Interés en el desarrollo de nuevas estrategias de inversión en ambientes de extrema volatilidad.
- Implementación inteligencia artificial (DRL) para la toma de decisiones de inversión.
- Estudiar estrategias ajustadas por riesgo en el mercado Crypto.
- Al mejor de nuestro conocimiento no existen investigaciones sobre estrategias de arbitraje mediante técnicas de DRL.
- Se propone un nuevo esquema de inversión a través del DRL.

Investigaciones relacionadas

- Deng et al. (2016) introduce un sistema de trading utilizando un framework de deep direct RL con representación difuza sobre los retornos pasados.
- Wu et al. (2020) introducen un método de trading adaptivo utilizando DRL junto utilizando GDQN (Gated Deep Q-learning) y GDPG (Gated Deterministic Policy Gradient)
- Liu et al. (2021) proponen un sistema de alta frecuencia basado en DRL utilziando LSTM como política para el algoritmo PPO sobre Bitcoin.
- Pelger et al. (2021) generalizan el modelo de arbitraje mediande PCA y IPCA combinando los portafolios con técnicas de deep learning para generar estrategias de inversión óptimas.

Objetivos

Objetivos

Objetivo Principal

- Utilizar métodos de Aprendizaje Reforzado Profundo (DRL) para la generación de estrategias de arbitraje en el mercado de criptodivisas.

Objetivos Específicos

- Proponer una metodología unificada combinando elementos del arbitraje estadístico innovando en la implemetación de DRL.
- Reducir el riesgo generando retornos ajustados por riesgo superiores al mercado.
- La metodología es robusta ante fricciones de mercado, e.g. costos de transacción.
- Determinar que las acciones del agente no son aleatorias, mas bien fundamentadas.

Preliminares

Statistical Arbitrage

Características Principales.

Para operar un esquema de arbitraje se requiere de :

- Múltiples Activos
- Alta Volatilidad
- Alta correlación o dirección de movimiento conjunto $\Rightarrow$ **Cointegración**
- Reversión a la media

Publicaciones más citadas.

- Distancia: Gatev et al. 2006
- PCA: Avellaneda et al. 2010
- Cointegración (COIN): Galenko et al. 2012
- COINMAN: Yu y Rengjie 2017

VECM: Vector Error Correction Model

Considerando precios logarítmicos $\mathbf{p}_t = (p_{i,t}, \dots p_{n,t})'$

y un modelo $VAR(k)$ sobre $\mathbf{p}_t$ reescrito como:

$$\Delta \mathbf{p}_t = \mu + \Sigma_{i=1}^{k-1} \Gamma_i \Delta \mathbf{p}_{t-1} + \Pi \mathbf{p}_{t-1} + \varepsilon_t \tag{1}$$

Donde $\Gamma_i \in \mathbb{R}^{n \times n}$, $\Pi \in \mathbb{R}^{n \times n}$ y $\varepsilon_t \overset{iid}{\sim} N(0, \Lambda)$.

Si $rank(\Pi) = n$, entonces $\mathbf{p}_t$ es estacionario.

Si $rank(\Pi) = 0$ entonces $\Pi = 0$ implica que $\Delta \mathbf{p}_t$ es un proceso $VAR(k-1)$ y no hay vectores de cointegración.

Si $1 \leq rank(\Pi) = r \leq n-1$, entonces existen $n \times r$ matrices de rango r , $\mathbf{A}$ y $\mathbf{B}$, tal que Π el modelo (1) se puede expresar como:

$$\Pi = \mathbf{A}\mathbf{B}' \tag{2}$$

y $\mathbf{b}_1' \mathbf{p}_t, \dots, \mathbf{b}_r' \mathbf{p}_t$ son estacionarios, donde $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_r)$.

Arbitrage Portfolios

Entonces, para los log-precios de cada activo $\mathbf{p}_t$, las r columnas de $\mathbf{B}$ pueden ser utilizadas para formar r portafolios (COIN).

Los portafolios se generan a partir de los vectores de cointegración normalizados.

Dado el vector de cointegración $\mathbf{b}_i = (b_i^1, \ldots, b_i^n)'$ se define la posición direccional para cada activo:

$$k \in L_i \iff b_i^k \geq 0, \quad \forall i = 1, \ldots r \tag{3}$$

$$k \in S_i \iff b_i^k < 0, \quad \forall i = 1, \ldots r \tag{4}$$

luego, los conjuntos L_i y S_i definen la posición dirección para cada activo del protafolio de cointegración formado por el vector $\mathbf{b}_i$.

Markov Decision Process

- El agente interactúa con la naturaleza en cada intervalo de tiempo
- En cada momento t el Agente observa el estado S_t y realiza una acción A_t a travéz de la política π
- En función de la acción, se observa un nuevo estado S_{t+1} y se genera una recompensa R_{t+1}
- La interación continua entre el agente y la naturaleza produce una trayectoria de **estado-acción-recompensa** : $\tau = (S_0, A_0, R_1, S_1, A_1, R_2, \dots)$

El **objetivo** del agente es maximizar la **recompensa acumulada** G_t:

$$G_t = \sum_{k=t+1}^{T} \gamma^{k-t-1} R_k \tag{5}$$

donde $\gamma \in [0, 1)$ corresponde al factor de descuento.

Deep Reinforcement Learning

- Implementación de técnicas de aprendizaje profundo para modelación
- Paradigma distinto al aprendizaje supervisado y no supervisado
- Utilización de redes neuronales como parametrización de funciones generales
  - política π : $\pi(\cdot|\theta)$
  - función Valor: $V^{\pi}(s) = \mathbb{E}\left[\sum_{h=0}^{\infty} \gamma^h R_{t+h} | s_t = s\right]$
  - Q - Valor : $Q^{\pi}(s,a) = \mathbb{E}\left[\sum_{h=0}^{\infty} \gamma^h R_{t+h} | s_t = s, a_t = a\right]$
- Algoritmos de entrenamiento por backpropagation:
  - PPO : Optimización directa de la política
  - DQN : Optimización directa del Q-Valor
  - A2C : doble red , una determina la política y otra evalúa la acción

Modelo

Portafolios de Arbitraje

Se construyen dos portafolios en función del signo del coeficiente de cointegración obteniendo posiciones direccionales contrarias.

Así, se construyen r portafolios neutrales. Para cada vector $\mathbf{b}_i$ se tiene su costante de normalización $l_i = \sum_{k \in L_i} |b_i^k|$ y $s_i = \sum_{k \in S_i} |b_i^k|$.

Por lo tanto los pesos se determinan como:

$$W_i^{(k)} = \begin{cases} b_i^{(k)}/l_i & \text{if } b_i^{(k)} \geq 0 \\ b_i^{(k)}/s_i & \text{if } b_i^{(k)} < 0 \end{cases} \tag{6}$$

Estos r portafolios son *dollar-neutral* y el portafolio $\mathbf{P}$ se construye como una equiponderación de todos:

$$\mathbf{P} = \frac{1}{r} \sum_{k=1}^{r} \mathbf{W}_k, \quad \mathbf{W}_k = (W_k^1, \ldots, W_k^n) \quad , \mathbf{P} = \mathbf{A} - \mathbf{B} \tag{7}$$

Reward Function

Con los portafolios **A** y **B**, la acción correponde a la posición direccional

(largo-**A** y corto-**B** o viceversa) con $A_t \in \{-1, 1\}$.

Dado el retorno de la acción: $r_t^A = \ln P_t^A - \ln P_{t-1}^A$ and $r_t^B = \ln P_t^B - \ln P_{t-1}^B$, por lo tanto

$$R_t = A_{t-1}\left[e^{r_t^A} - e^{r_t^B}\right]. \tag{8}$$

Se busca resolver el problema

$$\max_{\Theta} U_T\{R_1 \dots R_T|\Theta\}$$

$U_T(\cdot)$ es la recompensa del periodo, donde

$$U_T = \sum_{t=1}^{T} R_t$$

Trading Game

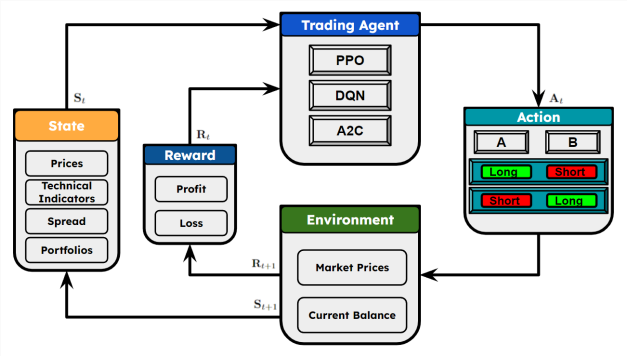


Figure 1: Esquema de interacción del sistema DRL para la aplicación de trading. Los estados corresponden a aquellos definidos a través de indicadores técnicos, volatilidades y otros generados a partir de activos sintéticos y spreads.

Entrenamiento del Modelo

Configuración de entrenamiento

- Step 1: Definir una ventana de tiempo historica de 6 días para estimar los portafolios de cointeración.
- Step 2: Usar la relación de cointegración y la data historica para la construcción de los portafolios **A** y **B**. El rango de cointegración se determina según el test de cointegración de Johansen a un nivel del 90%.
- Step 3: Con los activos sintéticos y la historia construír los estados del mercado a traves de señales (indicadores técnicos varios).
- Step 4: Definir un horizonte de 1 día para operar con los portafolios obtenidos.

Data

- Ventana temporal: 2020-11-01 00:00:00 until 2022-10-11 05:30:00 (frecuencia de 30 min)
- 6 días (288 t) para construcción de portafolios 1 día (48 t) de operación
- Más de 30.000 escenarios para entrenamiento

1. **BTC**: Bitcoin
2. **ETH**: Ethereum
3. **BNB**: BNB
4. **XRP**: Ripple
5. **ADA**: Cardano
6. **SOL**: Solana
7. **DOT**: Polkadot
8. **BCH**: Bitcoincash
9. **LTC**: Litecoin
10. **AVAX**: Avalanche
11. **ALGO**: Algorand
12. **AAVE**: Aave
13. **UNI**: UniSwap
14. **CAKE**: PancakeSwap

Training

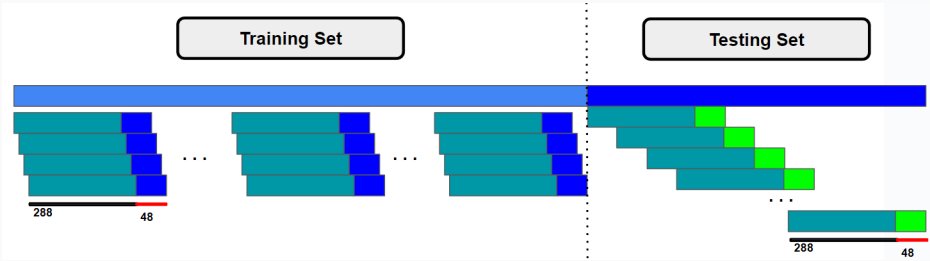
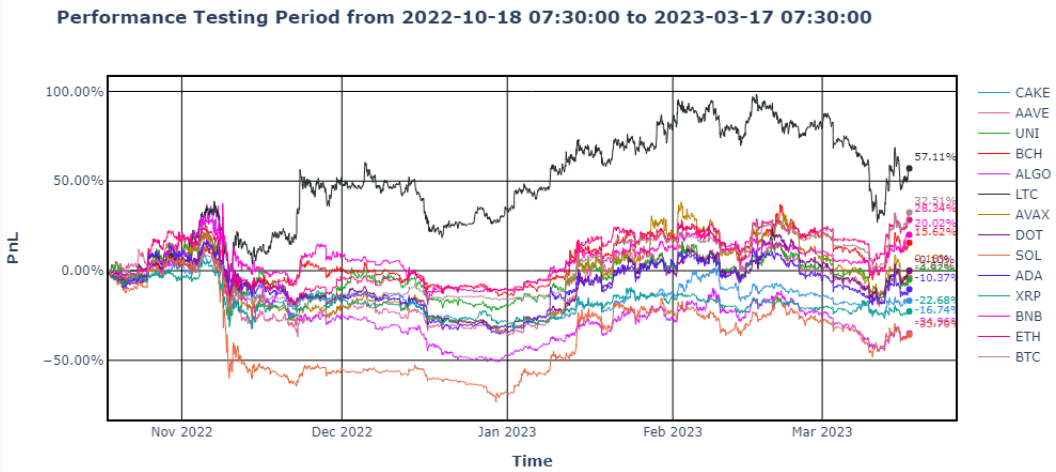


Figure 2: Esquema de segmentación de escenarios de entrenamiento y prueba. En el conjunto de entrenamiento hay superposición de escenarios, mientras que en el conjunto de prueba, no ocurre superposición y no se encuentra presente información utilizada para el entrenamiento.

Periodo de Prueba



Resultados

Risk and Performance Measures

Para evaluar las diferentes estrategias se utilizaron las siguientes métricas:

Riesgo :

- VaR : Value at Risk 5%
- ES : Expected Shortfall 5%
- σ : Desviación estándar
- MDD : Máximal Drawdown
- AS : Aumann & Serrano (2008) economic risk index (GHYP distribution)

Rendimiento ajustado por riesgo:

- Calmar
- Sharpe
- EPM: Economic Performance Measure

Rendimiento

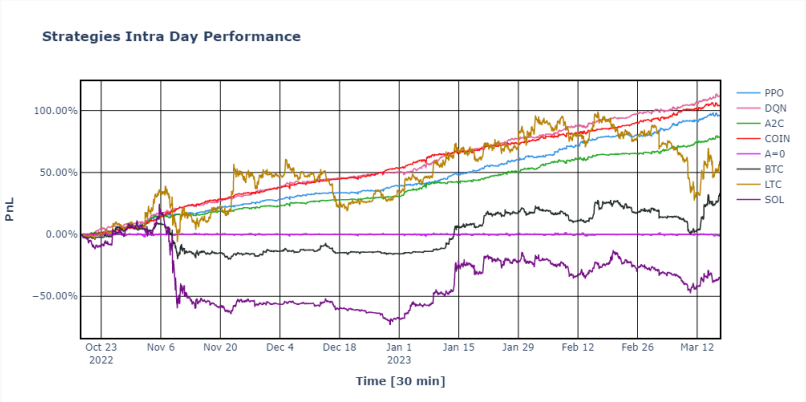


Figure 3: Profit & Loss a frecuencia de 30-minutos. Se incluye estrategia neutral A=0, los diferentes agentes DRL utilizados (PPO, DQN, y A2C), el benchmark COIN junto con referencias de mercado.

Rendimiento: Fricciones de mercado

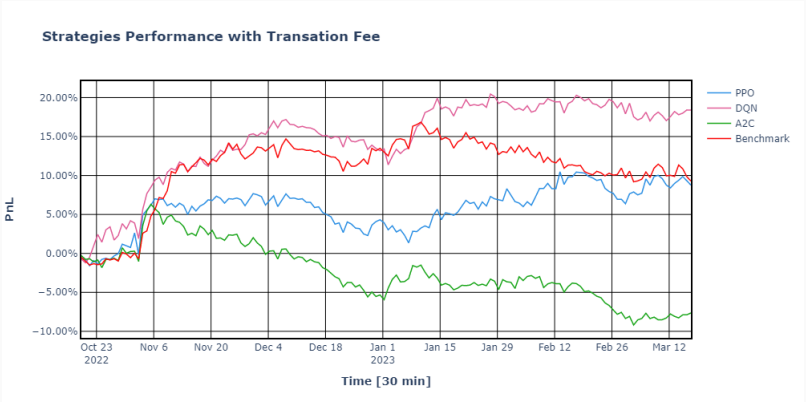


Figure 4: Profit & Loss de resultados diarios aplicando un costo por transacción de 0.02% por operación. La estrategia benchmark corresponde a COIN.

Resumen Métricas diarias

	Max D.	ES 5%	VaR 5%	σ	A-S
PPO	5.87[%]	-1.39[%]	-0.99[%]	0.82[%]	0.0036
DQN	4.96[%]	-1.44[%]	-1.09[%]	0.77[%]	0.0030
A2C	14.57[%]	-1.23[%]	-1.08[%]	0.77[%]	0.0032
COIN	6.56[%]	-1.22[%]	-1.10[%]	0.76[%]	0.0037

	Calmar	Sharpe	EPM	R[%]
PPO	0.0103	0.0738	0.1558	8.77
DQN	0.0239	0.1533	0.4004	18.39
A2C	-0.0035	-0.0670	-0.1430	-7.63
COIN	0.0096	0.0834	0.1597	9.29

Comparativa mercado

	MDD	ES 5%	VaR 5%	σ	AS	Calmar	Sharpe	EPM	R[%]
A = 0	3.22%	-1.67%	-1.26%	0.83%	0.00475	0.003	0.013	0.051	1.69
BTC	26.30%	-5.79%	-3.95%	2.78%	0.01558	0.007	0.068	0.048	32.51
LTC	34.82%	-9.31%	-6.50%	4.59%	0.02902	0.009	0.066	0.040	57.11
SOL	74.78%	-16.32%	-9.94%	7.19%	0.03629	-0.003	-0.040	-0.124	-35.70

- A = 0 : Control de posición
- BTC : Benchmark de mercado
- SOL : Peor rendimiento
- LTC : Mejor rendimiento

Conclusiones

Conclusiones

1. Se propone un método unificado de generación de portafolios de arbitraje, representación de mercado y toma de decisiones.
2. El método porpuesto se basa en DRL y se prueba con diferentes algoritmos.
3. Se propone un innovador método de entrenamiento basado en la generación de escenarios con horizonte fijo.
4. La estrategia obtiene buenos resultados en un ambiente de extrema volatilidad reduciendo el riesgo en gran medida.
5. Los resultados se mantienen positivos frente a fricciones de mercado (costos de transacción).
6. Las acciones del agente son fundamentadas generando decisiones coherentes.

Deep Reinforcement Learning applied to Statistical Arbitrage Investment Strategy on Cryptomarket

Defensa de título profesional de Ingeniera Civil Matemática y grado de Magíster en Ciencias de la Ingeniería Industrial

Gabriel Vergara Schifferli

7 Septiembre, 2023

Departmento de Matemáticas & Departamento de Industrias
Universidad Técnica Federico Santa María