



UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

FERNANDO DANILO DE MELO

Otimização de Portfólio através de técnicas de Aprendizado por Reforço

São Paulo

2021

FERNANDO DANILO DE MELO

Otimização de Portfólio através de técnicas de Aprendizado por Reforço

Versão original

Projeto de pesquisa para exame de qualificação apresentado à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo como parte dos requisitos para obtenção do título de Mestre em Ciências pelo Programa de Modelagem de Sistemas Complexos.

Área de concentração: Aprendizado de Máquina e Economia

Orientador: Prof. Dr. Camilo Rodrigues Neto

São Paulo

2021

Sumário

1	Introdução	3
2	Metodologia	5
2.1	<i>Problema de Pesquisa</i>	5
2.2	<i>Hipótese</i>	5
2.3	<i>Objetivo</i>	6
2.4	<i>Dados</i>	6
2.5	<i>Implementação</i>	7
2.5.1	A Tarefa	7
2.5.2	Agente e o estado	8
2.5.3	Recompensa	8
2.5.4	Deterministic Policy Gradient	9
2.5.5	Policy Networks	9
	REFERÊNCIAS	11

1 Introdução

Segundo Smith (2020), depois que estabelecida a divisão do trabalho, o produto gerado por cada trabalhador nem sempre atende a maioria de suas necessidades, sendo preciso a realização de trocas entre partes, porém esse método é pouco prático, devido a inconsistência que pode haver entre as necessidades de cada parte.

Para solucionar esse problemas várias sociedades passaram a utilizar metais como moeda de troca, porém outros problemas apareceram, como a necessidade de pesagem e verificar a autenticidade do metal, com isso surgiu a necessidade de se cunhar valor no metal, dando origem a moeda como conhecemos, e as Casas da Moeda que visavam controlar a qualidade e a quantidade de moedas.

Com a criação do conceito de moeda, se tornou mais fácil o acúmulo da mesma, com esse valor excedente espera-se auferir uma rentabilidade sobre ele, essa parte é o que chamamos de capital.

Mais de meio século depois, Mill (2019) apresenta um conceito parecido, que todo estoque acumulado do trabalho é chamado de capital, porém apresenta a distinção entre dinheiro e capital.

Onde o dinheiro é o produto do capital investido, que pode ser utilizado para bens pessoais, ou reinvestido na forma de capital para produção, gerando um aumento/melhoria na produção e conseqüentemente lucro.

Se mantivermos o dinheiro parado, ele estará exposto a efeitos de desvalorização, como por exemplo, da inflação. Dessa maneira, é necessário buscar formas eficazes de comprometer esse dinheiro. Uma possibilidade seria renunciar à posse dele agora, por um valor maior no futuro. Essa alocação de capital, buscando um retorno maior em uma data seguinte, é o que chamamos de investimento (REILLY; BROWN, 2011).

Com dinheiro acumulado, é necessário decidir como investi-lo. Esse processo de alocação pode ser de diversas maneiras, como por exemplo em um ativo único, sem diversificação, ou buscando dividir o dinheiro em vários ativos, criando assim um portfólio de ativos. Markowitz (1952) assume que o investidor busca sempre maximizar o retorno, diminuindo o risco o máximo possível, ou seja, o investidor busca construir uma carteira que combine os ativos de maneira a chegar na melhor relação risco-retorno (MARCUS; BODIE; KANE, 2013).

Em seu trabalho Markowitz introduz o conceito de média-variância, onde calculamos a relação risco-retorno através da variância dos retornos, dessa maneira diversificação do portfólio através de ativos não correlacionados, ajuda a diminuir o risco sistemático e a variância do portfólio como um todo, tornando-o mais lucrativo e com menores variações que um ativo por si só.

Mas isso levanta questões como: os ativos de carteira são a melhor seleção possível, considerando o risco-retorno? No caso de não ser, qual seria a melhor carteira? A que mantém o risco e potencializa o retorno ou a que preserva o retorno e minimiza o risco?

2 Metodologia

2.1 *Problema de Pesquisa*

Considerando técnicas de aprendizado de máquina em um conjunto de dados amplo, conseguimos criar um algoritmo completamente autônomo na construção, otimização e manutenção de um portfólio? Ele tem potencial para ter um retorno melhor que modelos clássicos?

2.2 *Hipótese*

O Mercado financeiro pode ser considerado um Sistema Complexo devido as características que o definem, esse viés multifacetado abre a possibilidade para estudos com diversas abordagens, uma das maneiras mais comuns de aborda-lo é através de técnicas estatísticas que tragam previsibilidade no movimento dos ativos buscando vantagem competitiva de maneira a auferir lucros mediante a sua negociação.

Os modelos clássicos, partem do cálculo do retorno esperado em uma janela de tempo considerando dados históricos, entretanto no começo dos anos 70 (FAMA, 1970), apresenta a Hipótese do Mercado Eficiente, onde defende que o preço de mercado das ações é uma avaliação racional do valor da empresa dadas as informações disponíveis. De acordo com essa teoria, o mercado é eficiente em termos de informação, ou seja, o valor de uma ação reflete toda informação disponível sobre a companhia.

Segundo, (MARCUS; BODIE; KANE, 2013), isso implica que o preço do ativo é suscetível as notícias, que por sua vez não podem ser previstas, tendo assim um comportamento randômico. Considerando essa hipótese, a única coisa que pode mudar o valor de uma ação são as informações que alteram a percepção do valor da empresa para o mercado.

Nos últimos anos, temos gerado dia após dia uma grande quantidade de dados estruturados e não estruturados, um fenômeno que chamamos de "Big Data", com isso o maior desafio deixa de ser obter o dado, mas a transformação do dado bruto em informação. Em paralelo a isso, temos uma maior acessibilidade de poder computacional e a evolução de ambientes Cloud, onde temos Infraestrutura com precificação sobre demanda acessível a todos, proporcionando escalabilidade quase que infinita, é importante salientar o surgimento

de arquiteturas que possibilitam armazenar, processar e extrair essa grande quantidade de dados, como Apache Hadoop¹ e Apache Spark², (DAVENPORT; BARTH; BEAN, 2012).

Podemos destacar, alguns fundos, que de maneira constante conseguem a mais de 30 anos manter um retorno médio de 30%³ através de métodos quantitativos.

Podemos então levantar a hipótese, que o grande número de dados gerados, em conjunto com técnicas de modelagem estatística e poder computacional, se torna possível o desenvolvimento de estratégias, com atuação autônoma, que consigam gerar retornos maiores que os modelos clássicos e acima da média do mercado, contrariando a Teoria do Mercado Eficiente.

2.3 Objetivo

O objetivo desse trabalho, é propor uma abordagem automatizada através da criação de um agente utilizando técnicas de Aprendizado por Reforço, de maneira que ele seja capaz, sem intervenção humana de escolher os ativos, distribuir os pesos, e executar operações de compra e venda que resultem em portfólios melhores que os modelos clássicos.

Um dos pontos importantes desse trabalho, é agregar dados de análise de técnica na série histórica, que permita descrever de maneira mais assertiva o estado que o agente se encontra, melhorando o aprendizado e as ações a serem tomadas.

2.4 Dados

Os dados utilizados nesse trabalho, serão coletados através do pacote *pandas_datareader* que através de uma API se conecta ao Yahoo Finance trazendo informações, como preço de abertura, fechamento, máxima, mínima e volume negociado no dia. Serão utilizados dados de ativos que compõe os principais índices das bolsas Americanas, Nasdaq Composite, SP500 e Dow Jones Industrial Average, da bolsa Brasileira serão considerados ativos que componham o índice Ibovespa.

Para construção das métricas de análise técnica, iremos utilizar o pacote *Technical Analysis Library*, que constrói as métricas através das informações de cada ativo.

¹ <https://hadoop.apache.org/>

² <https://spark.apache.org/>

³ <https://www.wsj.com/articles/the-making-of-the-worlds-greatest-investor-11572667202>

2.5 Implementação

Iremos utilizar a implementação sugerida por (JIANG; XU; LIANG, 2017) que está descrita nas subseções abaixo, porém iremos propor duas alterações, a primeira é a retirada do valor de transação, pois algumas corretoras no Brasil vem zerando o custo desse serviço nos últimos anos ⁴. A segunda, e principal proposta, é a incorporação de dados de análise técnica em conjunto com técnicas de Aprendizado por Reforço, temos que definir conceitos básicos, como o Ambiente, o Agente, Recompensa e o Estado.

2.5.1 A Tarefa

Nessa tarefa, o agente precisa a cada período T redistribuir os pesos os ativos, iremos considerar $T = 3$ dias. Por ser um ambiente contínuo, iremos considerar que o preço fechamento em t é igual o preço de abertura em $t + 1$ e que o ativo pode ser comprado por esse valor.

Podemos construir um vetor que contenha todos preços de fechamento para t , v_t . Com isso podemos construir um vetor com o preço relativo dos ativos:

$$y_t = \left(1, \frac{v_{1,t}}{v_{1,t-1}}, \frac{v_{2,t}}{v_{2,t-1}}, \dots, \frac{v_{m,t}}{v_{m,t-1}} \right)^T \quad (1)$$

Onde cada elemento de y_t é a razão do fechamento do ativo no período t em relação ao período anterior $t - 1$.

Podemos também definir o vetor com os pesos dos ativos no portfólio, como:

$$w = (w_1, w_2, \dots, w_m)^T \quad (2)$$

Onde w é vetor, que contem a contribuição do ativo de todos ativos até m para o portfólio p .

Dessa maneira podemos descrever que o valor do portfólio p no momento t é:

$$p_t = p_{t-1} y_t \cdot w_{t-1} \quad (3)$$

A taxa de retorno na forma logarítmica, é definida como:

$$r_t = \ln \frac{p_t}{p_{t-1}} \quad (4)$$

⁴ <https://www.infomoney.com.br/mercados/rico-zera-taxa-de-corretagem-para-acoes-e-xp-reduz-valor-em-75-para-plataforma-digital/>

Para o instante final t_f :

$$p_f = p_0 \prod_{t=1}^{t_f+1} y_t \cdot w_{t-1} \quad (5)$$

Onde podemos observar que o valor do portfólio final p_f é o valor inicial p_0 multiplicado pelo produto dos valores de cada ativo em cada momento t pelo peso no momento anterior $t - 1$, até o momento $t_f + 1$ onde liquidamos todas posições.

2.5.2 Agente e o estado

O agente ira executar transações de compra e venda no momento t , iremos considerar que essa transação é ínfima em relação ao mercado como um todo, não impactando o mercado futuro, porém, a transação impacta recompensa do momento $t + 1$. No instante t o agente executa transações de venda e compra, buscando redistribuir os pesos w_t do portfólio.

Logo podemos descrever a ação do agente como:

$$a_t = w_t \quad (6)$$

Como a escolha feita em w_t é recompensada somente no instante seguinte r_{t+1} , a decisão tomada anteriormente impacta a decisão atual, utilizamos esse efeito, considerando w_{t-1} como parte do estado:

$$s_t = (X_t, w_{t-1}) \quad (7)$$

Onde X_t é a matriz de característica que ajudam a descrever o estado s_t .

2.5.3 Recompensa

No algoritmo de Actor-Critic Policy Gradient, são utilizadas duas redes neurais, uma para aproximar a política e outra a recompensa, porém iremos utilizar o método proposto por onde definimos uma função explícita para a recompensa, como pode ser visto abaixo:

$$R(s_1, a_1, \dots, s_{t_f}, a_{t_f}, s_{t_f+1}) = \frac{1}{t_f} \sum_{t=1}^{t_f+1} r_t \quad (8)$$

Ou seja, é a média dos retornos logarítmicos, onde podemos expressar de maneira precisa o retorno acumulado R , e o retorno imediato $\frac{r_t}{t_f}$.

2.5.4 Deterministic Policy Gradient

Utilizando uma estratégia de somente explorar, podemos mapear de maneira clara uma ação através da política em um estado s , podemos utilizar um método gradiente, como da equação ??, porém é necessário definir um equação de métrica de performance, para esse fim iremos utilizar a função da recompensa, como:

$$J_{[0,t_f]}(\pi_\theta) = R(s_1, \pi_\theta(s_1), \dots, s_{t_f}, \pi_\theta(s_{t_f}), s_{t_f+1}) \quad (9)$$

Onde θ é o conjunto de parâmetros, $a_t = \pi_\theta(s_t)$ e $J_{[0,t_f]}$ nossa métrica de performance. Substituindo 9 em ??, temos:

$$\theta = \theta + \alpha \nabla \widehat{J_{[0,t_f]}}(\pi_\theta) \quad (10)$$

Os parâmetros são inicializador de maneira aleatória, e com a equação acima, eles são atualizados a cada passo em direção do gradiente.

2.5.5 Policy Networks

Será construída uma rede Neural que busca representar nossa política π_θ . A rede pode ser construída de diversas maneiras, iremos utilizar 3 arquiteturas distintas, CNN (Convolutional Neural Network), RNN (Recurrent Neural Network) e LSTM (Long Short Term Memory), em todas as redes o conjunto de dados será o vetor de característica X_t , e na saída utilizamos uma função softmax, que gera o vetor de pesos w_t .

A abordagem para construção dessa rede é uma das inovações do artigo que estamos utilizando como base para implantação, onde a rede tem caminhos diferentes e independente para cada ativo m que se conectam na saída softmax, porém os parâmetros são compartilhados entre cada caminho, podemos considerar cada caminho como uma rede menor de cada ativo m , idêntica, mas independente entre elas, essas redes são chamadas de Identical Independent Evaluators (IIE) e a topologia de Ensemble of IIE (EIIE).

Essa topologia apresenta bons resultados na tarefa de otimizar portfólio, em uma abordagem tradicional, onde a rede pra todos ativos é treinada em conjunto, uma performance histórica negativa pode impactar a escolha de determinado ativo, porém treinando redes separadas os ativos podem ser avaliados de maneira segregada, tornando mais fácil identificar seu potencial recente.

Para o treinamento, são criados lotes de dados, que são entregues ao modelo organizados em termos da linha temporal. Devido a sua natureza, o mercado financeiro irá de maneira continua ceder dados para o conjunto de dados. Quando o agente executar suas ações pra t , esse novo lote será incluído aos outros para treino. Depois que executadas a ações para $t + 1$, ele irá selecionar aleatoriamente N_b mini lotes para retreino. Um lote é selecionado com uma distribuição de probabilidade geométrica $P_\beta(t_b)$:

$$P_\beta(t_b) = \beta(1 - \beta)^{t-t_b-n_b} \quad (11)$$

Onde $\beta \in (0, 1)$ é a probabilidade da taxa de decaimento, dando enfase aos eventos recentes e n_b o número de período em cada mini lote. Esse esquema recebe o nome de Online Stochastic Batch Learning (OSBL).

Referências

DAVENPORT, T. H.; BARTH, P.; BEAN, R. How 'big data' is different. MIT Sloan Management Review, 2012. Citado na página 6.

FAMA, E. F. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, [American Finance Association, Wiley], v. 25, n. 2, p. 383–417, 1970. ISSN 00221082, 15406261. Disponível em: <http://www.jstor.org/stable/2325486>. Citado na página 5.

JIANG, Z.; XU, D.; LIANG, J. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*, 2017. Citado na página 7.

MARCUS, P. A. J.; BODIE, P. Z.; KANE, A. *Investments*. McGraw-Hill Education, 2013. ISBN 9780077861674. Disponível em: <https://books.google.com.br/books?id=eJ71mgEACAAJ>. Citado 2 vezes nas páginas 3 e 5.

MILL, J. S. *Princípios de Economia Política-Stuart Mill*. [S.l.]: LeBooks Editora, 2019. Citado na página 3.

REILLY, F.; BROWN, K. *Investment Analysis and Portfolio Management*. Cengage Learning, 2011. ISBN 9780538482387. Disponível em: <https://books.google.com.br/books?id=CfB-qTXqRWEC>. Citado na página 3.

SMITH, A. *A Riqueza das Nações-Adam Smith: Vol. I*. [S.l.]: LeBooks Editora, 2020. v. 1. Citado na página 3.

Glossário

Apache Hadoop é um framework de código aberto para armazenamento e processamento para grandes volumes de dados.

Apache Spark é um framework de código aberto de processamento para grandes volumes de dados com foco em performance.