



Escola de Artes, Ciências e Humanidades
da Universidade de São Paulo

SIN5006 - Inteligência Computacional

**Trabalho de avaliação da
disciplina**

1o sem 2025 - aluno Ivan F. Gancev (16298145)

Professora Patrícia R. Oliveira

Dados do artigo selecionado

Título: Stock Price Manipulation Detection using Variational Autoencoder and Recurrence Plots

Autores:

- *Khaled Safa*
 - Department of Computer Science, NTIC Faculty University of Constantine2, Abdelhamid Mehri, Constantine, Algeria
- *Ammar Belatreche*
 - Department of Computer and Information Sciences, Fac. of Engineering and Environment, Northumbria University, Newcastle Upon Tyne, UK
- *Salima Ouadfel*
 - Department of Computer Science, NTIC Faculty University of Constantine2, Abdelhamid Mehri, Constantine, Algeria

Publicação: 2024 International Joint Conference on Neural Networks (IJCNN)

Link: <https://doi.org/10.1109/IJCNN60899.2024.10650234>

Objetivo do artigo

O mercado de ações demanda constante necessidade de aprimoramento dos mecanismos que identificam cenários de manipulação ilícitas em seus preços. Isso interfere diretamente na confiança dos investidores e atualizações de normas feitas por seus reguladores.

O artigo escolhido desenvolve um método de identificação de manipulações em preços de ações usando *Recurrence Plots* e *beta Variational Autoencoders*, usando redes neurais convolucionais. Este modelo demonstrou alta eficácia na identificação de manipulações usando dados do projeto LOBSTER, que reuniu dados oficiais da NASDAQ das empresas Amazon, Google, Intel, Microsoft e Apple.

Breve descrição das técnicas

Recurrence Plots

- ❖ Introduzido por **Eckmann** e outros em 1987
- ❖ É uma técnica avançada de análise de dados não lineares. É uma visualização (ou um gráfico) de uma matriz quadrada, na qual os elementos da matriz correspondem aos momentos em que um estado de um sistema dinâmico se repete.
- ❖ Tecnicamente, o RP revela todos os momentos em que a trajetória do espaço de fase do sistema dinâmico visita aproximadamente a mesma área no espaço de fase.
- ❖ Ref:
<http://www.recurrence-plot.tk/glance.php>

Variational Autoencoder

- ❖ Por **Diederik P. Kingma** e **Max Welling**
- ❖ É um modelo gráfico probabilístico composto por uma rede de codificadores, uma rede de decodificadores e uma função de perda. O codificador mapeia os dados de entrada em uma distribuição com uma média e um desvio padrão, então a variável latente é amostrada dessa média e desvio padrão e o decodificador aprende a reconstruir os dados através da representação oculta por *backpropagation* dos erros da função de perda.
- ❖ Ref: <https://arxiv.org/pdf/1312.6114>

Conjuntos de dados

O projeto LOBSTER (Limit Order Book System - The Efficient Reconstructor) possui um conjunto de dados contendo o **livro de ofertas** e o **histórico de ordens** de 5 ativos da bolsa de NASDAQ. Os dados do artigo são os de 1o nível de 21/jun/2012. Os ativos são:

- Amazon: **AMZN**
- Apple: **AAPL**
- Google: **GOOG**
- Intel: **INTC**
- Microsoft: **MSFT**

Ref: <https://lobsterdata.com/info/DataSamples.php>

Códigos-fonte

Os códigos-fonte usados no artigo não estão disponíveis

Resultados

O artigo faz um comparativo no final sobre a ingestão proposital de manipulações nos preços e compara os resultados de alarmes falsos e AUC (Area Under the Curve) que mostram resultados relevantes do modelo proposto pelo artigo. Os modelos comparados foram:

- Adaptative Hidden Markov Model with Anomaly States (**AHMMAS**)
- Kernel Principal Component Analysis (**KPCA**)/Multidimensional Kernel Density Estimation (**MKDE**)

AUC			
	Proposed approach	AHMMAS	KPCA-MKDE
APPLE	0.8500	0.5344	0.9206
AMZN	0.8721	0.5152	0.9602
MSFT	0.9145	0.6711	0.9143
INTC	0.9376	0.5169	0.8732
GOOG	0.8805	0.5119	0.8996

False Alarm Rate			
	Proposed approach	AHMMAS	KPCA-MKDE
APPLE	1.24	7.83	1.07
AMZN	1.58	9.22	1.22
MSFT	2.5	0.52	0.71
INTC	1.32	1,15	0.54
GOOG	2.16	0.5	1.62

Ideia para o projeto

Etapa preliminar (30/abr):

- Criar a implementação para *Recurrence Plots* com os dados usados no artigo

Apresentação final (18/jun):

- Criar a implementação para *Variational Autoencoders* para o mesmo conjunto de dados
- Comparar os resultados obtidos com os resultados demonstrados no artigo



EACH



Obrigado!