

Avaliação dos diferentes tipos de redes LSTM para predição de ações na bolsa de valores

Gabriel Souto¹, Bruna Capistrano¹, Matheus Matias¹,
Jorge Soares¹, Eduardo Ogasawara¹, Lucas Giusti¹

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

{gabriel.souto, bruna.capistrano, matheus.matias}@aluno.cefet-rj.br,
jorge.soares@cefet-rj.br, eogasawara@ieee.org, lucas.giusti@aluno.cefet-rj.br

Resumo. *Redes Neurais Profundas são modelos valiosos na tarefa de aprendizagem. Neste trabalho, propomos o uso do método multicamadas conhecido como Long Short-Term Memory. Aplicamos três modelos diferentes (LSTM Vanilla, Stacked e Convolutacional) para a mesma série de ações. Essa escolha foi feita devido à lacuna na literatura ao comparar quais modelos LSTM podem ser usados na predição de séries temporais. Os resultados encontrados comprovaram-se uma alternativa ao que pretendemos mostrar, no sentido de um trabalho comparativo com os melhores modelos LSTM.*

Abstract. *Deep Neural Networks are valuable models in the task of learning tasks. In this work, we propose using the multilayer method known as Long Short-Term Memory. We apply three different models (LSTM Vanilla, Stacked and Convolutional) for the same stock series. This choice was made due to the gap in the literature when comparing which LSTM models to use in time series prediction. The results found proved to be an alternative to what we intend to show, in the sense of comparative work with the best LSTM models.*

1. Introdução

O *day trade* se relaciona com a ideia de compra e venda de ações com o objetivo de obter lucros em um curto espaço de tempo, geralmente no mesmo dia. Ser bem sucedido no *day trading* é difícil, devido a isso investidores precisam tomar diversos cuidados. O que provoca um enorme fascínio nas pessoas em geral é o fato do *day trading* proporcionar ganhos consideráveis no conforto de casa. No entanto, os investidores de varejo possuem certa dificuldade em ganhar dinheiro com o *day trading*. Segundo o estudo realizado por Barber et al. [2014] apenas 1% dos *day traders* ganham dinheiro de forma consistente. O estudo examinou negócios ao longo de um período de 14 anos, de 1992 a 2006. *Day traders* dependem das flutuações das ações ou do mercado para obter seus lucros, eles preferem aquelas ações que flutuam muito ao longo do dia, comprando ações quando ela estiver subindo, ou vendendo no mesmo dia de acordo com um sentimento de mercado, por exemplo.

As previsões de séries temporais no valor de ações são uma tarefa árdua. Isto se deve à natureza ruidosa, volátil e dinâmica das séries temporais financeiras. A utilização de modelos de aprendizado de máquina tem sido explorada de modo a mitigar as dificuldades deste desafio [Hyndman and Athanasopoulos, 2018; Tsay, 2010; Bustos and Pomares-Quimbaya, 2020; Obthong et al., 2020].

Alguns trabalhos na literatura fornecem comparações de redes neurais com outros métodos mais convencionais de predição como em [Saad et al., 1998]. Recentemente, modelos *LSTM* (*Long Short-Term Memory*) têm sido utilizados para predição de um determinada série temporal. As redes *LSTM Vanilla*, umas das arquiteturas de *LSTM* mais utilizadas nos dias de hoje, foram inicialmente introduzidas por [Wu et al., 2018]. As redes *LSTM Stacked* são um modelo composto de várias camadas de *LSTM*, esta abordagem permite que o estado oculto em cada nível opere em diferentes escalas de tempo [Selvin et al., 2017]. As redes *Convolutional LSTM* foram desenvolvidas para leitura de dados de espaço-temporais bidimensionais, mas podem ser adaptadas para uso com predição de séries temporais univariadas [Shi et al., 2015].

Neste trabalho, utilizamos para a tarefa de predição uma das mais avançadas arquiteturas de redes neurais: as redes de memória de longo prazo. As *LSTM* são o estado da arte para tarefas de aprendizagem de sequência, como reconhecimento de escrita à mão, reconhecimento de fala ou predição de série temporal [Fischer and Krauss, 2018; Hochreiter and Schmidhuber, 1997]. Todos estes modelos foram aplicados para um conjunto de ações único da bolsa de valores da B3 da empresa Vale.

Este trabalho está organizado da seguinte forma: na seção 2 é apresentado em detalhes cada uma das redes *LSTM* mencionadas acima, e os detalhes utilizados na implementação. A explicação dos parâmetros utilizados, e os resultados comparativos dos modelos são apresentados, sendo melhores elucidados na seção 3. Por último, na seção 4, estão presentes as conclusões e os trabalhos futuros apresentados.

2. Métodos

2.1. Problema de predição do valor de ações

A utilização de métodos de *deep learning* para predição do valor de ações tem sido objeto de diversos estudos [Fischer and Krauss, 2018]. Dados referentes ao valor das ações são sensíveis devido à possibilidade de serem afetados por fatores imprevisíveis de natureza política, econômica e social. De acordo com Tsay [2010], as séries temporais de dados financeiros são não-estacionárias, ou seja, não têm um equilíbrio estável e sua variância não é constante. Desta forma, ainda que os valores das ações apresentem tendências, os dados têm a característica de serem sazonais e não-lineares.

Nesse sentido, os modelos tradicionais de predição não são confiáveis o suficiente para prever o valor das ações. Ademais, o volume elevado dos dados dessa natureza, bem como sua não linearidade, demandam modelos capazes de identificar os padrões escondidos e dinâmicas subjacentes. Desta forma, algoritmos de *Deep Learning* podem ser utilizados, dada a sua eficácia na verificação desses padrões a partir de um processo de autoaprendizagem, fornecendo uma boa predição ao analisar as interações e os padrões ocultos nos dados [Selvin et al., 2017].

2.2. LSTM

As redes neurais têm sido utilizadas em diversas aplicações para modelar e prever problemas relacionados à predição de séries temporais em diferentes áreas de conhecimento [Siami-Namini et al., 2019].

As redes neurais recorrentes (*RNNs*) são projetadas para processar dados sequenciais com mais eficiência, levando em consideração a natureza sequencial dos dados. Em

particular, redes *LSTM* têm sido utilizadas na predição de séries temporais e em dados econômicos e financeiros, devido à sua capacidade de reter informações anteriores [Siami-Namini et al., 2019].

As redes *LSTM* são um modelo especial de *RNN* com recursos adicionais para memorizar a sequência de dados [Siami-Namini et al., 2019]. As redes recorrentes diferem das redes *feed-forward* tradicionais no sentido de que não têm apenas conexões neurais em uma única direção, assim, os neurônios podem passar dados para uma camada anterior ou para a mesma camada. Nesse caso, os dados não fluem de uma forma única, e o efeito prático para isso é a existência de memória de curto prazo, além da memória de longo prazo que as redes neurais já possuem em decorrência do treinamento ([Nelson et al., 2017]).

As redes *LSTM* são compostas por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída [Fischer and Krauss, 2018]. As camadas são conjuntos de células nos quais os fluxos de dados são capturados e armazenados. Cada célula contém *gates* que permitem que os dados nela contidos sejam descartados, filtrados ou adicionados para as células seguintes ([Siami-Namini et al., 2019]).

As células *LSTM* possuem *gates*, *input*, *forget* e *output*, que, no estado da célula, permitem inserir informações úteis, descartar informações inúteis e extrair informações úteis do estado da célula para gerar uma saída, respectivamente. Este fluxo de informação controlado dentro da célula permite que a rede memorize múltiplas dependências de tempo com características diferentes ([Lindemann et al., 2021]).

2.3. Variações de LSTM

Vanilla LSTM. A variante do *LSTM Vanilla* é o modelo mais simples e também o mais comumente usado na literatura [Wu et al., 2018]. Este modelo possui uma camada oculta e uma camada de saída, que é utilizada para realizar a predição. A vantagem das redes neurais *LSTM* é que elas oferecem suporte nativo para sequências. O modelo *LSTM* lê um passo de tempo da sequência por vez e constrói uma representação interna do estado que pode ser usada como um contexto aprendido para fazer uma predição [Wu et al., 2018].

Stacked LSTM. Este modelo consiste em várias camadas *LSTM* ocultas empilhadas. Uma camada *LSTM* requer uma entrada tridimensional, e as células *LSTMs* por padrão produzirão uma saída bidimensional como uma interpretação do final da sequência. Neste caso, é necessário realizar um tratamento para ter uma saída 3D da camada *LSTM* oculta como entrada para a próxima camada [Wu et al., 2018].

Convolutional LSTM. Os dados de entrada que representam períodos mais extensos podem ser filtrados e reduzidos com base em operações de convolução que são incorporadas em redes *LSTM* ou diretamente na estrutura de células *LSTM*. Em outras palavras, a leitura convolucional dos dados de entrada é incorporada diretamente a cada unidade *LSTM* [Nelson et al., 2017]. Os *gates* recebem os recursos gerados como novas entradas. Eles são uma representação reduzida que captura apenas as informações mais relevantes para que a eficiência do mecanismo de atualização do estado da célula seja aprimorada. Essa abordagem tem como objetivo aprimorar a precisão da predição de dependências de longo prazo com base em um processamento mais eficiente de sequências de entrada, projetando os dados em um espaço de recurso de dimensão inferior [Lindemann et al., 2021].

2.4. Métricas utilizadas para avaliação (SMAPE)

O método de erro percentual médio absoluto (MAPE) é um cálculo de avaliação usado para medir a exatidão ou precisão de uma predição. Usando o método MAPE, obtém-se a diferença entre o valor real (A_t) e o valor previsto (F_t). Esta técnica facilita o entendimento e a visualização do tamanho do erro da predição. Quanto menor o valor percentual MAPE, mais precisa é a predição [Kurniawan et al., 2021]. O MAPE pode ser representado pela fórmula:

$$M = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Uma das desvantagens do MAPE é que a métrica penaliza mais os erros positivos do que os negativos. Com isso, o uso do MAPE "simétrico" surge como uma solução [Hyndman and Athanasopoulos, 2018].

Substituindo o denominador do MAPE pela média entre o valor atual e o previsto [(Valor Real + Valor Previsto) / 2] e deixando o numerador do MAPE igual, obtém-se o MAPE simétrico (sMAPE). O sMAPE sempre será diferente de zero quando o valor previsto ou real for diferente de zero [Hyndman and Athanasopoulos, 2018]. O sMAPE pode ser definido da seguinte forma:

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|F_t - A_t|}{(A_t + F_t)/2}$$

3. Avaliação Experimental

Para o presente estudo, foi utilizada uma série temporal univariada. No caso, a série escolhida para ser estudada foi o histórico de preço das ações VALE3 na B3, de janeiro de 2014 até maio de 2021, conforme a Figura 1.



Figura 1. Histórico do preço mínimo durante o pregão da B3 das ações VALE3

3.1. Protocolo experimental

Visando uma predição mais robusta e confiável, foi utilizada a técnica de *rolling-origin*, que consiste em decompor a série em janelas móveis de treino e teste, no qual são atualizados sucessivamente a origem da predição, sendo produzidas novas previsões para cada

origem. Tal abordagem nos permitiu utilizar o conceito de janelas móveis, no qual, a cada iteração com uma nova origem de predição, uma observação é removida do início da série e outra observação é adicionada ao final da série, deixando-a com o tamanho constante [Hyndman and Athanasopoulos, 2018]. O tamanho das janelas utilizado nos treinamentos e predições das redes foi variando de 1 a 30 para as redes *Vanilla* e *Stacked* e de 2 a 30 (valores pares) para a rede *Convolutional*, devido a natureza da sua entrada.

Das 1828 observações contidas na série temporal, foram utilizadas as primeiras 1207 para treino e as outras 621 observações para testes de predição.

3.2. Comparação dos métodos

A Tabela 1 apresenta uma visão comparativa dos métodos propostos (*Vanilla*, *Stacked* e *Convolutional*). Vale destacar que os três métodos foram executados com janelas de tamanho 30 e as redes neurais foram treinadas por 1000 épocas.

Tabela 1. Desempenho das Redes LSTM para predição do valor de ações

Janelas	Vanilla		Stacked		Convolutional	
	Treino (s)	SMAPE	Treino (s)	SMAPE	Treino (s)	SMAPE
1	82.91	1.66	66.57	1.74		
2	56.74	1.76	87.32	1.77	125.92	1.69
3	64.66	2.02	114.88	2.03		
4	72.14	1.79	139.94	2.01	136.36	1.65
5	111.42	2.04	159.14	2.29		
6	126.62	2.31	169.43	1.99	161.04	1.75
7	134.69	1.84	211.13	1.95		
8	147.02	2.37	237.93	2.66	165.65	1.87
9	165.56	1.93	255.75	2.04		
10	210.5	2.18	280.81	2.03	185.5	2.09
11	218.63	2.76	301.76	2.20		
12	159.67	2.82	320.89	2.36	212.73	2.25
13	164.31	3.25	357.83	4.26		
14	171.78	2.07	342.72	2.84	239.3	1.81
15	181.18	2.89	394.25	2.96		
16	190.30	2.39	414.19	3.08	277.53	2.16
17	209.27	1.94	749.58	4.04		
18	216.08	2.24	865.53	3.87	278.81	1.86
19	240.38	2.70	929.90	2.10		
20	254.01	2.53	922.71	5.07	306.32	1.79
21	263.74	1.94	607.21	4.98		
22	267.39	4.90	565.89	1.94	329.23	2.24
23	293.44	3.33	554.93	3.13		
24	310.69	3.34	542.62	3.63	355.69	1.86
25	316.98	3.14	558.28	2.73		
26	312.93	19.48	572.16	2.33	380.76	2.04
27	329.86	200	598.68	3.57		
28	350.36	7.46	620.63	2.29	407.53	2.07
29	346.51	1.78	640.37	2.99		
30	358.76	2.84	647.11	9.36	429.71	3.46

A Tabela 1 caracteriza a comparação entre as redes *LSTM Vanilla*, *Stacked* e *Convolutional*, mostrando o tempo de treinamento gasto pela rede, e a métrica de avaliação do modelo.

3.3. Resultados

Foi observado que as redes *Vanilla* possuem o menor tempo de treinamento. A *Convolutional LSTM* apresentou o melhor resultado para a métrica de avaliação. Considerando o tamanho das janelas no tempo, observou-se que, para até quatro janelas, as previsões foram também melhores.

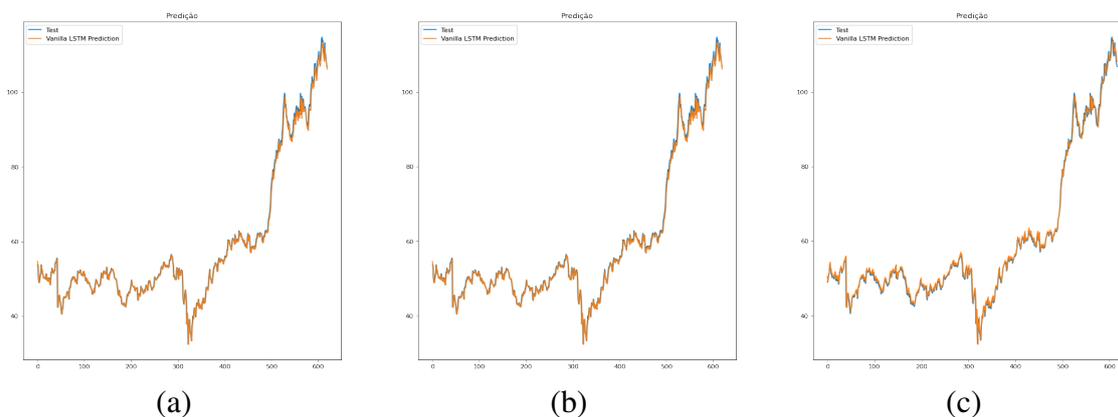


Figura 2. Predição da rede *Vanilla* com uma única janela de tempo (a); Predição da rede *Stacked* com uma única janela de tempo (b); Predição da rede *Convolutional* com janela de tempo de quatro observações (c)

A Figura 2 representa a predição das redes *Vanilla*, *Stacked* e *Convolutional* que obtiveram os melhores valores para a métrica avaliada. A partir deste cenário, é possível observar que a seleção do número de passos de predição no tempo impacta no desempenho preditivo dos modelos. O erro aferido foi elevado para experimentos executados para mais de quatro etapas preditivas.

4. Conclusão

A partir deste artigo, analisou-se o desempenho de três modelos de redes neurais *LSTM* na predição de valores de ações, a saber: *Vanilla LSTM*, *Stacked LSTM* e *Convolutional LSTM*. Entre os modelos analisados, o *Convolutional* apresentou os melhores resultados considerando a métrica de avaliação escolhida, o sMAPE, com um erro de 1.65%. Já com relação ao tempo de treinamento da rede, verifica-se que o modelo *Vanilla*, em geral, foi executado mais rapidamente para todas as janelas de tempo utilizadas no experimento. Ademais, é possível identificar que os resultados alcançados para a métrica selecionada são melhores quando o experimento foi executado para números menores de janelas no tempo.

Foi possível observar que o número de passos no tempo impacta no desempenho dos modelos. Vale ressaltar que, neste estudo, o número de células *LSTM* foi fixo, de cinquenta células nas camadas. Além disso, conforme presente na tabela 1, também não foi explorado a possibilidade de variação do número de épocas de treinamento das redes.

Agradecimentos

Os autores agradecem à CAPES (código 001), à FAPERJ e ao CNPq pelo financiamento parcial do trabalho.

Referências

- B.M. Barber, Y.-T. Lee, Y.-J. Liu, and T. Odean. The cross-section of speculator skill: Evidence from day trading. *Journal of Financial Markets*, 18(1):1–24, 2014.
- O. Bustos and A. Pomares-Quimbaya. Stock market movement forecast: A Systematic review. *Expert Systems with Applications*, 156, 2020.
- T. Fischer and C. Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669, 2018.
- S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
- Rob J. Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, May 2018. ISBN 978-0-9875071-1-2.
- F. Kurniawan, D.E. Herwindiati, and M.D. Lauro. Raw Paper Material Stock Forecasting with Long Short-Term Memory. In *2021 9th International Conference on Information and Communication Technology, ICoICT 2021*, pages 342–347, 2021.
- B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich. A survey on long short-term memory networks for time series prediction. In *Procedia CIRP*, volume 99, pages 650–655, 2021.
- D.M.Q. Nelson, A.C.M. Pereira, and R.A. De Oliveira. Stock market’s price movement prediction with LSTM neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2017-May, pages 1419–1426, 2017.
- M. Obthong, N. Tantisantiwong, W. Jeamwatthanachai, and G. Wills. A survey on machine learning for stock price prediction: Algorithms and techniques. In *FEMIB 2020 - Proceedings of the 2nd International Conference on Finance, Economics, Management and IT Business*, pages 63–71, 2020.
- E.W. Saad, D.V. Prokhorov, and D.C. Wunsch II. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on Neural Networks*, 9(6):1456–1470, 1998.
- S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V.K. Menon, and K.P. Soman. Stock price prediction using LSTM, RNN and CNN-sliding window model. In *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017*, volume 2017-January, pages 1643–1647, 2017.
- X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 2015-January, pages 802–810, 2015.
- S. Siami-Namini, N. Tavakoli, and A. Siami Namin. A Comparison of ARIMA and LSTM in Forecasting Time Series. In *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, pages 1394–1401, 2019.
- Ruey S. Tsay. *Analysis of Financial Time Series*. Wiley, July 2010. ISBN 978-0-470-64455-3.
- Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing*, 275:167–179, 2018.