

Redes Neurais Recorrentes Long Short-Term Memory (LSTM) Aplicadas a Previsão de Ações

Luis F. S. Souza¹, Bárbara D. do A. Rodriguez², Cristiana A. Poffal³

¹Instituto de Matemática Estatística e Física - IMEF, luisfernandessauceadosouza@furg.com

²Instituto de Matemática Estatística e Física - IMEF, barbararodriguez@furg.br

³Instituto de Matemática Estatística e Física - IMEF, cristianaandrade@furg.br

De acordo com [2] redes neurais recorrentes *long short-term memory* (LSTMs) emergiram como um modelo eficaz e escalável para diversos problemas de aprendizado relacionados a dados sequenciais. Métodos anteriores [3] para resolver esses problemas não escalavam para dependências de prazos prolongados. LSTMs, por outro lado, são gerais e eficazes na captura de dependências temporais de longo prazo.

Um esquema de bloco da arquitetura de LSTM mais popular (LSTM *vanilla*) pode ser visto na Figura 1. Ele apresenta três *gates* (entrada, esquecimento, saída), entrada do bloco, uma única célula (o Carrossel de Erro Constante), uma função de ativação de saída e conexões *peephole*. A saída do bloco é conectada recorrentemente de volta à entrada do bloco e a todos os *gates*.

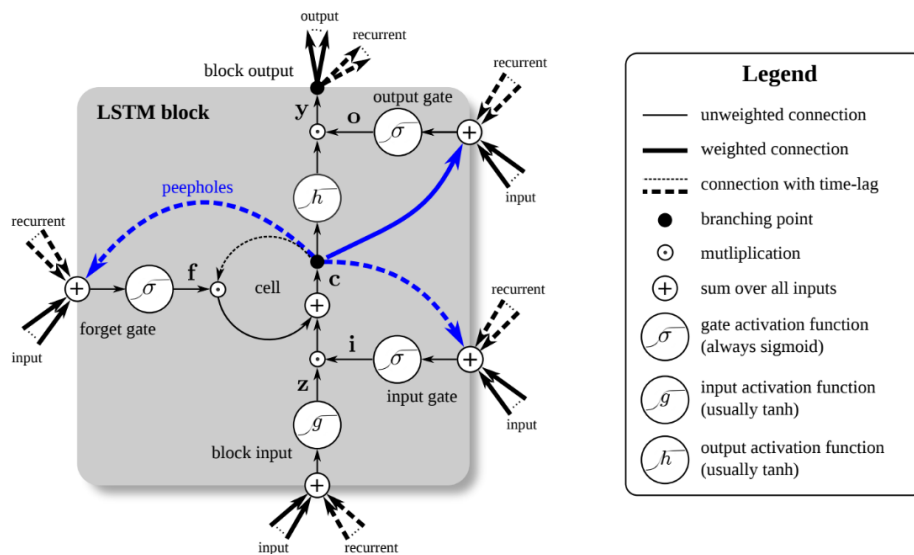


Figura 1. Esquema em blocos de uma rede LSTM. Fonte: Greff [2].

Na Figura 1, z representa a saída do bloco de *input*; i , do *input gate*; f , do *forget gate*; o , do *output gate*; c , da célula; e y , do bloco. Para essas saídas, são atribuídos pesos W e pesos recorrentes R e realizadas a propagação e retropropagação para suas atualizações. As equações de cada bloco, dado que \odot é o produto de Hadamard, são definidas como $\tilde{z}^t = W_z x^t + R_z y^{t-1} + b_z$, $z^t = g(\tilde{z}^t)$, $\tilde{i}^t = W_i x^t + R_i y^{t-1} + p_i \odot c^{t-1} + b_i$, $i^t = \sigma(\tilde{i}^t)$, $\tilde{f}^t = W_f x^t + R_f y^{t-1} + p_f \odot c^{t-1} + b_f$, $f^t = \sigma(\tilde{f}^t)$, $c^t = z^t \odot i^t + c^{t-1} \odot f^t$, $\tilde{o}^t = W_o x^t + R_o y^{t-1} + p_o \odot c^t + b_o$, $o^t = \sigma(\tilde{o}^t)$, $y^t = h(c^t) \odot o^t$.

Baseado no trabalho de [1], utilizando a arquitetura de rede com nós ocultos de estrutura $100 - 50 - 50 - 50 - 1$ com otimizador *RMSProp* e função perda de *Mean Square Error*, 150 épocas de treinamento e taxa de aprendizado 0,001, a rede é treinada com os preços baseados nos dados da biblioteca *yfinance* (<https://pypi.org/project/yfinance/>) de fechamento da ação WEGE3.SA do dia 01/01/2023 até 01/01/2024. Sendo, neste caso, os valores de entrada do modelo as datas de fechamento e os valores de saída os preços da ação. A partir dessa rede treinada, tentou-se prever o preço dessa ação do dia 01/01/2024 até 01/06/2024. Os resultados são apresentados na Figura 2.

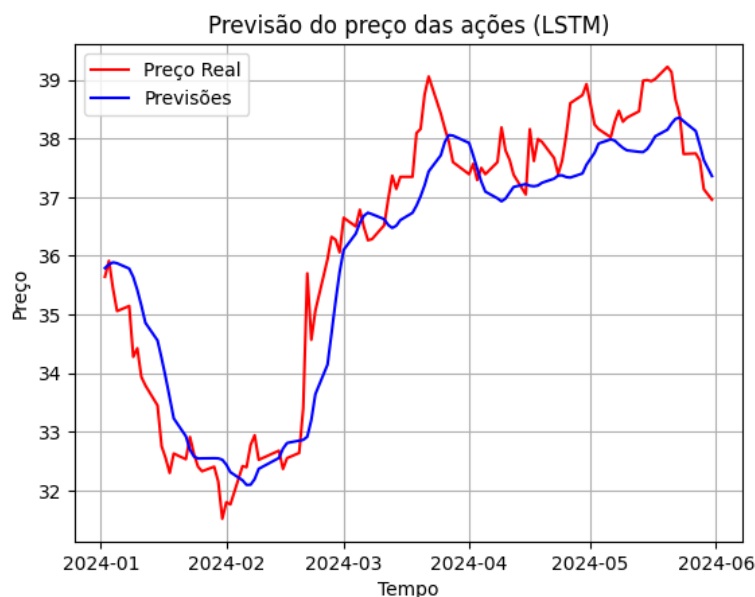


Figura 2. Preço real e previsão do preço de fechamento da ação WEG3.SA. Fonte: Dos Autores.

A estrutura LSTM mostrou-se eficiente para previsão dos preços de ações, mesmo com dados de longo prazo. O seu *Mean Absolute Error (MAE)* foi em torno de 0,65, significando que em média por dia se tem um erro de R\$ 0,65 na previsão do preço da ação.

REFERÊNCIAS

- [1] T. Balbi, *Introdução às redes LSTM: Prevendo valor de ações na bolsa*, <https://medium.com/datarisk-io/introduco-s-redes-lstm-prevendo-valor-de-acoes-na-bolsa-df270ca0cee5>, Acesso em: 16 jan. 2025, 2021.
- [2] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink e J. Schmidhuber, “LSTM: A Search Space Odyssey”, *Transactions on Neural Networks and Learning Systems*, v. 28, n. 10, pp. 2222–22 232, 2017. DOI: 10.1109/TNNLS.2016.2582924.
- [3] S. Hochreiter, Y. Bengio, P. Frasconi e J. Schmidhuber, “Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies”, em *A Field Guide to Dynamical Recurrent Networks*, J. F. Kolen e S. C. Kremer, editores. New York: IEEE Press, 2001, cap. 14, pp. 250–256, ISBN: 9780470544037.