

LONG SHORT-TERM MEMORY APLICADO A PREÇOS DE AÇÕES BRASILEIRAS: um estudo sobre o número de épocas no aprendizado de uma rede neural recorrente

CAMILO ILZO SHIMABUKURO

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA - CPS

NAPOLEÃO VERARDI GALEGALE

CENTRO PAULA SOUZA - FATEC/ETEC

JOSÉ ODÁLIO DOS SANTOS

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO (PUCSP)

LONG SHORT-TERM MEMORY APLICADO A PREÇOS DE AÇÕES BRASILEIRAS: um estudo sobre o número de épocas no aprendizado de uma rede neural recorrente

1. INTRODUÇÃO

Os resultados recentes obtidos por fundos quantitativos baseados em algoritmos de negociação, *Algorithmic Trading* (AT), contribuíram para impulsionar o uso de abordagens baseadas em *Machine Learning* (ML) e *Deep Learning* (DL) (IEEE SPECTRUM, 2017; HARVARD UNIVERSITY, 2017; HEATON; POLSON; WITTE, 2017; ARNOTT; HARVEY; MARKOWITZ, 2019 ; BUCHANAN, 2019; WIGGLESWORTH, 2020).

Abordagens baseadas em *Deep Learning* (DL) têm se mostrado promissoras em áreas como o reconhecimento de imagens e o processamento da linguagem natural por sua capacidade de extração de informações sobre dados não lineares, prescindindo do trabalho intensivo da seleção de características (*feature selection*) das técnicas de *Machine Learning* (ML) (BENGIO; COURVILLE; PASCAL, 2012; LECUN; BENGIO; HINTON, 2015).

Entretanto, revisões da literatura realizadas até 2015 apresentaram poucos estudos sobre abordagens de DL aplicadas em AT (LÄNGKVIST; KARLSSON; LOUTFI, 2014; BUCHANAN, 2019). No mercado de capitais brasileiro, tais publicações são ainda menos frequentes (CAVALCANTE et al., 2016). Por outro lado, nos últimos quatro anos, observa-se um número crescente de artigos sobre o tema (XIONG; NICHOLS; SHEN, 2016; BAO; YUE; RAO, 2017; FISCHER; KRAUSS, 2017; KRAUS; FEUERRIEGEL, 2017; TSANTEKIDIS et al., 2017; BUCHANAN, 2019; SIRIGNANO, 2019).

As redes neurais profundas sobrepõem camadas de representação de características com maior capacidade de modelar estruturas complexas de dados. No entanto, para a obtenção de resultados satisfatórios e aprendizado mais eficiente, dependem da redução da dimensionalidade, de uma criteriosa seleção das variáveis de entrada e dos parâmetros de rede (LÄNGKVIST; KARLSSON; LOUTFI, 2014; CHONG; HAN; PARK, 2017).

O processo de aprendizado requer uma adequada escolha de parâmetros, como a taxa de aprendizado, o tamanho do lote e o número de épocas. Utilizando-se técnicas de regularização como o *early stopping* para a escolha do número de épocas, pode-se melhorar a eficiência de aprendizado e até reduzir o risco de *overfitting* (PRECHELT, 1998; FISCHER; KRAUSS, 2017; SIRIGNANO, 2019).

Nesse contexto, pretende-se estudar o comportamento do erro de predição em função dos parâmetros de treinamento de uma rede neural. Assim, este trabalho busca responder à questão: como o erro de predição varia em função do número de épocas no treinamento de uma rede neural recorrente baseada em DL, sobre séries de preços de ações brasileiras?

Este trabalho tem por objetivo estudar o efeito do número de épocas sobre o erro de predição de uma rede neural recorrente baseada em DL, utilizando a abordagem *Long Short-Term Memory* sobre séries de preços de ações negociadas na bolsa de valores brasileira.

Contribui-se para cobrir lacuna da pesquisa recente sobre abordagens de *Deep Learning* aplicadas no segmento de negociação de ações por algoritmos, ao estudar formas de melhoria da eficiência (redução do tempo de aprendizado) e da eficácia (redução de *overfitting*) de ferramentas preditivas que visam a redução de incertezas e riscos, com benefícios para gestores de recursos, instituições financeiras e investidores.

Vale destacar, considerando-se a possibilidade de generalização dos resultados a outras áreas do conhecimento, o artigo de Bukhari e Jameel (2020), pesquisadores do MIT, sobre o efeito da temperatura na disseminação da COVID-19, recomendando estudos futuros utilizando

a rede neural recorrente LSTM como modelo no estado da arte para predição de eventos em séries temporais de dados.

2. FUNDAMENTAÇÃO TEÓRICA

Estratégias quantitativas rentáveis envolvem uma eficiente coleta e análise de informações, bem como o desenvolvimento de estratégias dinâmicas que otimizem os critérios de desempenho prescritos na presença da incerteza (GUO et al., 2017).

Para Gadre-Patwardhan, Katdare e Joshi (2016), a incerteza está entre os maiores desafios enfrentados por pesquisadores e gestores no campo das finanças ao introduzir um inevitável fator de risco que torna complexa a tomada de decisões.

Os preços das ações exibem passeio randômico, tipicamente não estacionário. Porém, sob certas condições e horizontes de tempo, podem apresentar algum grau de reversão à média, comportamento de tendência, ou regimes. Tais comportamentos observáveis em séries financeiras evidenciam a importância da previsão da mudança de tendências, regimes, ou do ponto de inflexão. Nessas situações podem ser utilizadas abordagens de ML como *Hidden Markov Model* (HMM), filtro de Kalman, Redes Neurais ou os *Markov Switching Models* (MSM), ainda que com menor capacidade preditiva (CHAN, 2009; LÄNGKVIST; KARLSSON; LOUTFI, 2014).

Entretanto, a aplicação do ML no contexto do AT apresenta também complexidades inerentes à microestrutura do mercado, à granularidade dos dados, com execuções parciais, cancelamentos de ordens, liquidez oculta, em que não se tem ao menos uma intuição de como a distribuição de liquidez dos livros de ordens limitadas se relaciona com movimentos futuros dos preços. Ainda que se reconheça no ML um *framework* poderoso, escalável e fundamentado para análise de dados e previsão, não se tem um caminho fácil para a rentabilidade (KEARNS; NEVMYVAKA, 2013).

Por outro lado, ainda que se consiga abordagens com boa aderência aos dados, Prado (2018) chama a atenção para o *overfitting* e o teste de modelos com base em dados históricos (*backtesting*), pois a pesquisa na área das finanças lida com conjuntos de dados pouco extensos e uma baixa razão sinal/ruído.

Não obstante, pesquisas recentes sobre reconhecimento de padrões apresentam uma tendência para a aplicação do DL à predição em séries temporais financeiras, utilizando abordagens de topologia não-linear das Redes Neurais Convolucionais (CNN) e Redes Neurais Recorrentes (RNN) (LÄNGKVIST; KARLSSON; LOUTFI, 2014; BAO; YUE; RAO, 2017), conforme revisão da literatura realizada por meio de buscas nas bases Scopus e Web of Science, utilizando os critérios apresentados no Quadro 1.

Quadro 1 – Critérios de busca sobre abordagens de DL aplicadas ao AT

Atributo	Critério
Expressão	("Deep Learning") AND ("Algorithmic Trading" OR "High Frequency Trading" OR "Systematic Trading" OR "Quantitative Trading" OR "Automated Trading" OR "Stock Market" OR "Stock Trading" OR "Stock Price" OR "Trading Systems")
Período	2000 a 2019
Idioma	Inglês
Publicação	Artigos publicados em periódicos e conferências
Domínios Excluídos	Artes, Humanidades, Psicologia, Medicina, Agricultura, Biociências, Química

Fonte: Autores

Os resultados da busca são apresentados na Tabela 1.

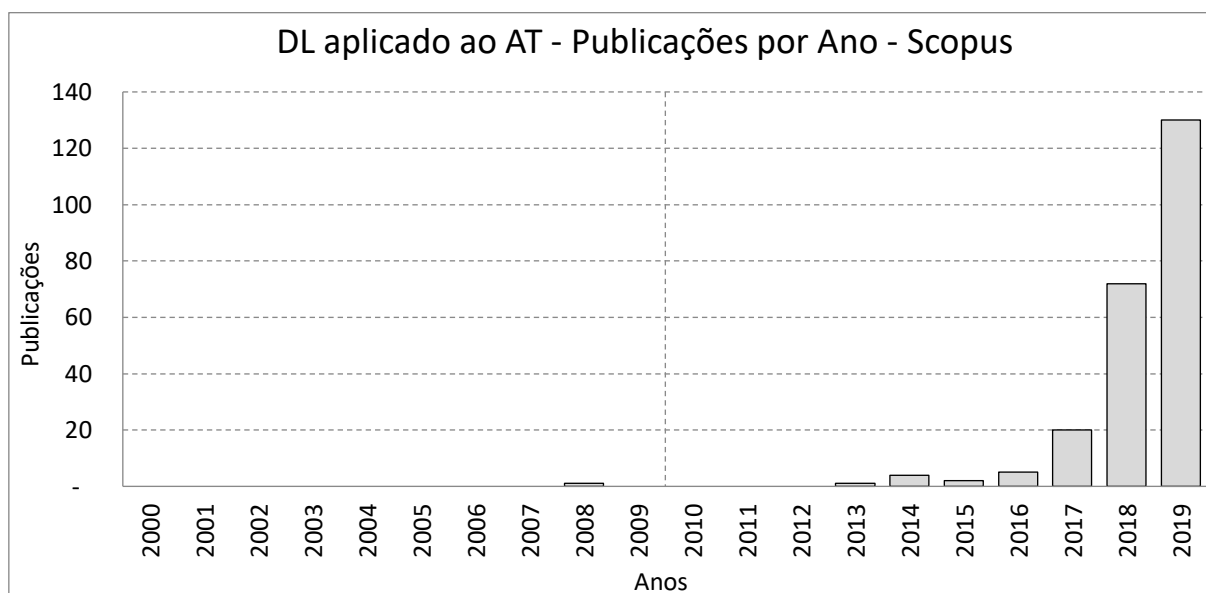
Tabela 1 – Publicações resultantes da busca nas bases Scopus e Web of Science

Base	Número de publicações
Scopus	235
Web of Science	129

Fonte: Scopus (2020) e Web of Science (2020)

A evolução anual das publicações na base Scopus é apresentada na Figura 1.

Figura 1 – Publicações sobre abordagens de DL aplicadas ao AT (Scopus)



Fonte: Scopus (2020)

Para a análise de conteúdo, foram avaliados o título, palavras-chave, resumo e conteúdo dos 50 artigos mais citados, selecionados conforme critérios do Quadro 2, e resultados na Tabela 2.

Quadro 2 – Critérios de seleção sobre abordagens de DL aplicadas ao AT

Tipo de Critério	Critério
Inclusão	Conteúdo se concentra em abordagens de DL aplicadas ao AT
	Conteúdo apresenta aspectos relevantes sobre abordagens de DL aplicadas ao AT
Exclusão	Conteúdo fora da área de interesse ou marginalmente relacionado ao tema de busca
	Documento não disponível para consulta online ou Documento duplicado

Fonte: Autores

Tabela 2 – Abordagens de mais frequentes

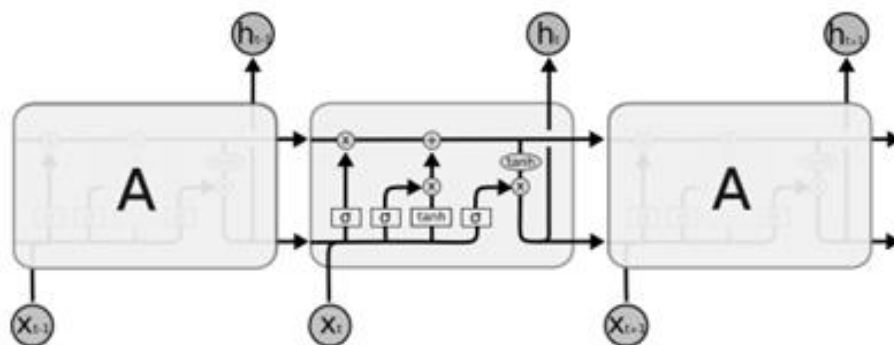
Abordagem	Número de artigos
Convolutional Neural Network – CNN	20
Long Short-Term Memory – LSTM	21

Fonte: Scopus (2020) e Web of Science (2020)

As CNNs apresentam grande número de estudos por sua capacidade de identificar relações não-lineares dos dados e de armazenamento da totalidade dos dados da microestrutura dos mercados, como os livros de ofertas limitadas (TSANTEKIDIS et al., 2017), todavia, não apresentam a capacidade de preservar a memória sobre dados passados (BAO; YUE; RAO, 2017).

A LSTM, uma rede neural recorrente, por sua vez, além de extrair características de relações não lineares, memoriza séries temporais longas, retém seletivamente informações passadas relevantes, descarta os dados não relevantes, como ruídos, pois seus nós fecham circuitos de realimentação recorrentes, resultando em característica de memorização extensa. Além disso, mantém a capacidade de aprendizado durante o treinamento, mantendo os gradientes da função de perda estáveis, propagando-se com menor risco de *vanishing gradients*, por meio de funções sigmóides e tangentes hiperbólicas. (HOCHREITER; SCHMIDHUBER, 1997; XIONG; NICHOLS; SHEN, 2016; FISCHER; KRAUSS, 2017; KIM; WON, 2018). A Figura 2 ilustra a recorrência da rede LSTM no tempo.

Figura 2 – Rede LSTM

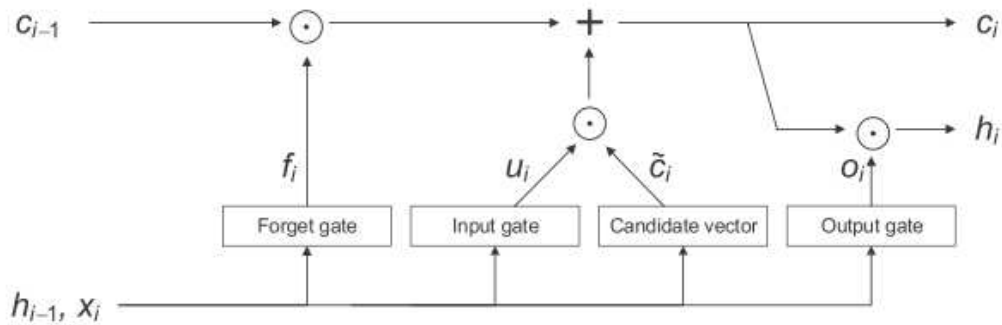


Fonte: Olah, 2020

O conceito central da unidade LSTM é permitir que os dados percorram toda a sua estrutura, recebendo ou eliminando informações operadas pelas comportas que controlam o conteúdo e o estado da unidade, conforme Figura 3 (KRAUS; FEUERRIEGEL, 2017; SELVIN et al., 2017).

A configuração de redes neurais utiliza estratégias para a otimização de hiperparâmetros, em que tentativas de escolha aleatórias se mostram mais eficientes que buscas em rede (*grid search*). Os hiperparâmetros são valores escolhidos que controlam o processo de aprendizado, diferentemente dos pesos, que são aprendidos durante o treinamento da rede neural (BERGSTRÄ; BENGIO, 2012).

Figura 3 – Detalhamento da estrutura interna de unidade LSTM



Fonte: Kraus, Feuerriegel (2017)

A taxa de aprendizado, por exemplo, determina o tamanho das mudanças nos pesos no processo de *backpropagation* na descida do gradiente. Baixas taxas tornam o aprendizado mais lento, enquanto taxas muito altas fazem com que a função de erro se altere descontroladamente, sem uma melhora contínua. Para aperfeiçoar esse processo utiliza-se um parâmetro de impulso (*momentum*) que verifica como as alterações passadas afetaram as mudanças nos pesos atuais, e realizar a próxima alteração de peso aproximadamente na mesma direção da anterior (ZHANG; PATUWO; HU, 1998; DUNIS; WILLIAMS, 2003).

Embora as redes neurais profundas com grande número de parâmetros sejam sistemas robustos de ML, o *overfitting* permanece um problema crítico. Grandes redes neurais apresentam processamento lento, dificultando lidar com o *overfitting* pela combinação ou co-adaptação das previsões em tempo de teste. Um procedimento para lidar com este problema é o *Dropout*, uma técnica em que se desconecta aleatoriamente algumas unidades neurais da rede durante o treinamento, prevenindo o excesso de co-adaptação entre elas (SRIVASTAVA et al., 2014). Outro hiperparâmetro utilizado no treinamento de redes neurais é o número de épocas, isto é, o número de vezes que o processo de aprendizado será realizado sobre o conjunto de dados de treinamento. Técnicas de ajuste do número de épocas e métodos de regularização, como o *early stopping*, podem melhorar a eficiência de aprendizado e reduzir o risco de *overfitting* (PRECHELT, 1998; FISCHER; KRAUSS, 2017; SIRIGNANO, 2019). Dentro do conjunto de técnicas de ajuste de parâmetros, pretende-se estudar o efeito do número de épocas sobre o erro da previsão.

3. PROCEDIMENTOS METODOLÓGICOS

Este trabalho é de natureza aplicada e abordagem qualitativa, e utiliza uma revisão da literatura, executada na seção 2-Fundamentação Teórica seguida de análise qualitativa, na seção 4-Resultados e Discussão, sobre o comportamento do erro de previsão de uma rede LSTM resultante da variação do número de épocas no processo de treinamento, por meio de medidas de dispersão e do coeficiente de determinação do erro de previsão.

Para realizar essa análise, foi utilizado um programa de computador que implementa uma rede LSTM (HOCHREITER; SCHMIDHUBER, 1997; FISCHER; KRAUSS, 2017; KARPATY, 2020; OLAH, 2020), em linguagem Python versão 3.7.4, disponível na plataforma GitHub no arquivo “Modelo LSTM-B3.py” (GITHUB, 2020; ROHRER, 2020; PALAI, 2020), e utiliza as bibliotecas de Numpy, Pandas, Scikit-Learn, TensorFlow, Matplotlib (GÉRON, 2017; CHOLLET, 2018), baseado na plataforma de distribuição

Anaconda v.1.9.7 e interface de desenvolvimento MS Visual Studio Code v.1.47.2 (PEDREGOSA et al., 2011).

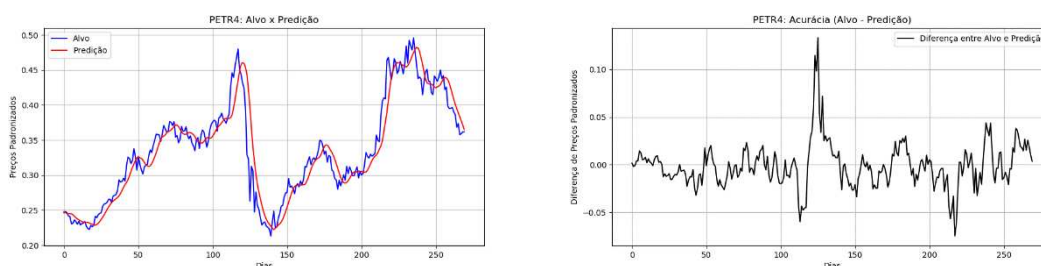
A rede é composta de duas camadas de 200 unidades neuronais cada, adequadas aos recursos computacionais à disposição, com a taxa de aprendizado fixada em 0,001, tamanho do lote de 50 e número de épocas variando de dez a cem com incrementos de dez épocas. Optou-se pelo otimizador de gradiente ADAM (*Adaptive Moment Estimation*) pela facilidade de implementação, eficiência computacional e adequação a dados não estacionários, gradientes esparsos e alta relação ruído/sinal (KINGMA, 2015; RUDER, 2017).

Como dados de entrada, foram utilizados os preços de fechamento das ações preferenciais de Petrobrás, PETR4, obtidos do site da B3 (B3, 2020), de 2008 a 2018, em arquivo .csv, disponibilizados na plataforma GitHub (GITHUB, 2020). Os dados foram ajustados para eliminar o efeito de desdobramentos ocorridos no período, e segmentados em: 80% para o treinamento, 10% para a validação, e 10% para os testes. Assim, das 2719 linhas de dados, 2160 são destinadas ao treinamento, outras 271 destinadas à validação e as 271 restantes destinadas aos testes. O tamanho da sequência de dados foi definido para 20 pontos de dados.

4. RESULTADOS E DISCUSSÃO

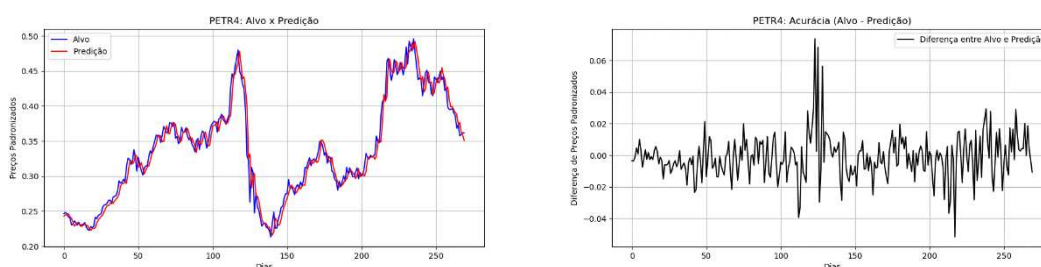
O programa foi executado mediante variação das épocas de dez em dez, partindo de dez e terminando em cem, mantendo-se os demais hiperparâmetros constantes. A título de ilustração, são apresentados os gráficos de predição e do erro (acurácia) emitidos pelo programa para épocas de dez, cinquenta e cem, conforme Figura 4, Figura 5 e Figura 6, sobre ações preferenciais da Petrobras, PETR4. Os demais gráficos intermediários estão disponíveis para verificação na plataforma GitHub (2020).

Figura 4 – Gráficos de Predição e Erro de PETR4 com treinamento de 10 Épocas



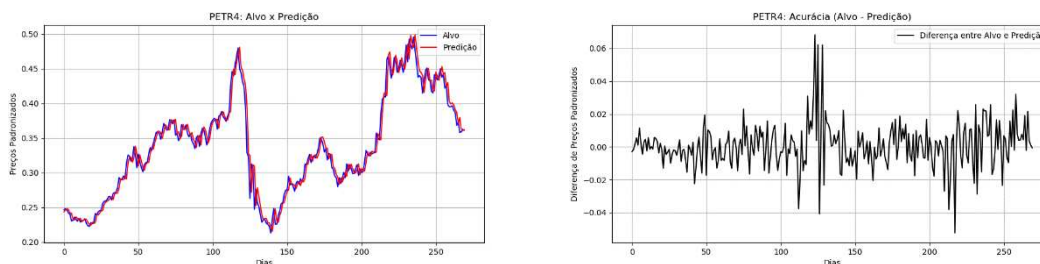
Fonte: Resultado da Pesquisa

Figura 5 – Gráficos de Predição e Erro de PETR4 com treinamento de 50 Épocas



Fonte: Resultado da Pesquisa

Figura 6 – Gráficos de Predição e Erro de PETR4 com treinamento de 100 Épocas



Fonte: Resultado da Pesquisa

A média, medidas de dispersão e o coeficiente de determinação R^2 , em função das épocas, são apresentadas na Tabela 3.

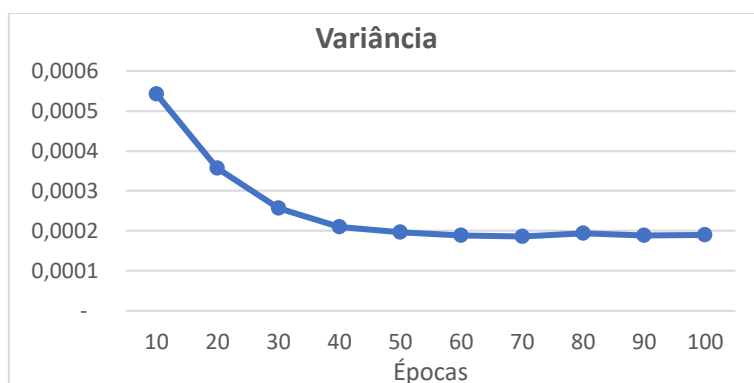
Tabela 3 – Média, Medidas de Dispersão e Coeficiente de Determinação

Épocas	Média	Variância	Desvio Padrão	R2
10	- 0,001260	0,000542	0,023290	0,895881
20	- 0,006160	0,000357	0,018892	0,929903
30	0,001015	0,000257	0,016028	0,950584
40	- 0,001234	0,000210	0,014484	0,959468
50	- 0,001052	0,000197	0,014042	0,962305
60	- 0,001320	0,000189	0,013736	0,963244
70	- 0,000841	0,000186	0,013620	0,963935
80	0,001911	0,000194	0,013913	0,962786
90	- 0,002884	0,000189	0,013735	0,963005
100	0,000733	0,000190	0,013768	0,963688

Fonte: Resultado da Pesquisa

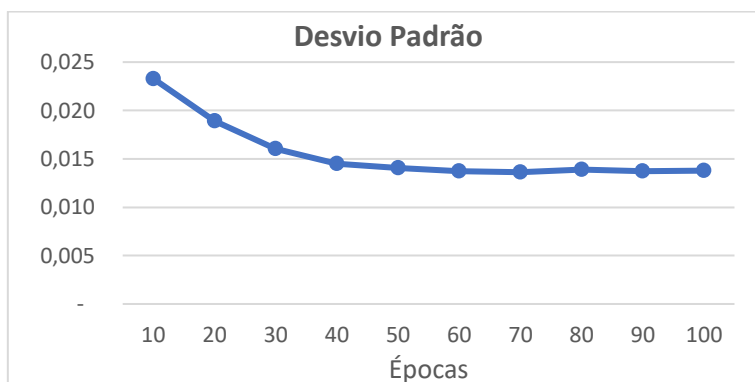
Nos gráficos da Figura 7 e da Figura 8, observa-se que as medidas de dispersão decrescem de forma significativa no intervalo entre 10 e 50 épocas. Porém, no intervalo entre 50 e 100 épocas, apresentam apenas uma discreta melhoria.

Figura 7 – Variância do erro de predição em função do número de Épocas



Fonte: Resultado da Pesquisa

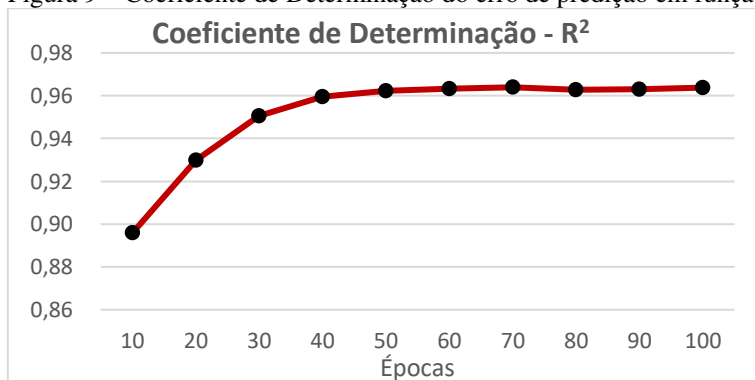
Figura 8 – Desvio Padrão do erro de predição em função do número de Épocas



Fonte: Resultado da Pesquisa

No gráfico da Figura 9, observa-se que o coeficiente de determinação aumenta no intervalo de 10 a 50 épocas, apresentando discreta melhoria a partir de 50 épocas.

Figura 9 – Coeficiente de Determinação do erro de predição em função do número de Épocas



Fonte: Resultado da Pesquisa

Tais resultados indicam que no intervalo de 10 a 50 épocas houve maior redução do erro de predição, e que de 50 a 100 épocas os resultados variam de forma quase assintótica, estabelecendo um ponto de partida em torno de 50 épocas para a rede LSTM desta pesquisa, visando a redução do tempo de processamento e a eficiência do treinamento.

5. CONSIDERAÇÕES FINAIS

A revisão da literatura identificou uma tendência recente de crescimento na produção científica sobre abordagens de *Deep Learning* aplicadas ao *Algorithmic Trading*. Até 2016, o tema objeto de pouco interesse, porém atingindo 72 publicações em 2018, e 130 publicações em 2019. As publicações resultantes justificaram a escolha da rede LSTM, não só pela frequência com que é utilizada em pesquisas recentes, como pelas suas características de realimentação recorrente, extração de características de relações não lineares dos dados, memorização de séries temporais longas, retenção seletiva de informações relevantes, mantendo a capacidade de aprendizado, com baixo risco de *vanishing gradients*.

Este trabalho investigou um dos hiperparâmetros de aprendizado de uma rede neural recorrente LSTM, o ajuste do número de épocas sobre o erro de predição, aplicado a séries de

preços de ações brasileiras, por meio da análise de medidas de dispersão e do coeficiente de determinação do erro de predição, identificando um intervalo de épocas de maior sensibilidade, entre 10 e 50 épocas em que o erro se reduz de forma mais rápida, e um intervalo em que a redução é próxima à assintótica, entre 50 e 100 épocas, estabelecendo um valor inicial para o treinamento em torno de 50 épocas, tornando o treinamento mais eficiente, reduzindo o tempo de processamento e buscando diminuir o risco de *overfitting*.

Em estudos futuros, pode-se utilizar uma análise multivariada para a determinação dos intervalos para os demais hiperparâmetros e estabelecer uma superfície de hiperparâmetros que estabeleça valores iniciais que minimizem o tempo de treinamento da rede. Recomenda-se, entretanto, estudos quantitativos sobre conjuntos de dados mais amplos e redes neurais de maiores dimensões para estudos sobre a redução de risco do *overfitting*.

Vale reiterar a possibilidade de aplicação dos estudos sobre a rede LSTM em outras áreas, como a predição em séries temporais de propagação de pandemias (BUKHARI; JAMEEL, 2020), bem como em séries climáticas e de mobilidade urbana.

REFERÊNCIAS

- ARNOTT, R.; HARVEY, C.R.; MARKOWITZ H. A Backtesting Protocol in the Era of Machine Learning. **The Journal of Financial Data Science**. Winter, 2019.
- B3 Preços históricos. <http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/series-historicas/> Acesso em 11 jul. 2020.
- BAO, W.; YUE, J.; RAO, Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. **PLoS ONE**. 2017.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**. 2012.
- BUCHANAN, B. G. Artificial intelligence in finance. **The Alan Turing Institute**. 2019
- BUKHARI, Q.; JAMEEL, Y. Will coronavirus pandemic diminish by summer? **MIT. SSRN**, 2020.
- CAVALCANTE, C. C. et al. Computational Intelligence and Financial Markets: A Survey and Future Directions. **Expert Systems with Applications**. v 55, 2016.
- CHAN, E. P. **Quantitative trading**. Wiley, 2009.
- CHOLLET, F. **Deep Learning with Python**. Manning Publications Co., 2018.
- CHONG, E.; HAN, H.; PARK, F. C. Deep learning networks for stock market analysis and prediction. **Expert Systems with Applications**, v. 83, 2017.
- DUNIS, C.; WILLIAMS, M. Applications of Advanced Regression Analysis for Trading and Investment. Cap 1 in DUNIS, C.; LAWS, J.; NAIM, P. **Applied Quantitative Methods for Trading and Investments**. Wiley, 2003.
- FISCHER, T.; KRAUSS, C. Deep learning with Long Short-Term Memory Networks for Financial Market Predictions. **European Journal of Operation Research**. v. 270. 2017.
- GADRE-PATWARDHAN, S.; KATDARE V. V.; JOSHI M. R. A Review of Artificially Intelligent Applications in the Financial Domain. Cap. 1 **Artificial Intelligence in Financial Markets**. Palgrave MacMillan, 2016.
- GÉRON, A. **Hands-On Machine Learning with Scikit-Learn**. O'Reilly, 2017.

GITHUB Dados de Mercado, Códigos Fonte e Resultados da Pesquisa. Disponível em: <<https://github.com/Camilollzo/Dados-de-Mercado-B3>>. Acesso em 19 jul. 2020.

GOODEFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016.

GUO, X.; LAI, T. L.; SHEK, H.; WONG, S. P. **Quantitative trading - Algorithms, Analytics, Data, Models, Optimization**. CRC Press. Taylor & Francis Group, 2017.

HEATON, J. B.; POLSON, N. G.; WITTE, J. H. Deep learning for finance: deep portfolios. **Applied Stochastic Models in Business Industry**. v 33, 2017.

HARVARD UNIVERSITY – An Economy of algorithms. Publ. 19-jan-2017. Disponível em: <<https://www.seas.harvard.edu/news/2017/01/economy-algorithms>> ou <<https://youtu.be/2JpyJdw3Vwk>> Acesso em 11 jul. 2020.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, 1997.

IEEE Spectrum. **Hedge Funds Look to Machine Learning, Crowdsourcing for Competitive Advantage**. 2017. Disponível em: <<https://spectrum.ieee.org/tech-talk/computing/software/fintech-trends-hedge-funds-look-to-machine-learning-crowdsourcing-for-competitive-advantage>> Acesso em 11 jul. 2020.

KEARNS, M.; NEVMYVAKA, Y. Machine Learning for Market Microstructure and High-Frequency Trading. **High-Frequency Trading: New Reality for Traders, Markets and Regulators**. Risk Books, 2013.

KARPATY, A. The Unreasonable Effectiveness of Recurrent Neural Networks. Disponível em <<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>> Acesso em 11 jul. 2020.

KIM, H. Y.; WON, C.H. Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH models. **Expert Systems with Applications**. 2018.

KINGMA, D. P. Adam: A Method for Stochastic Optimization. arXiv:1412.6980v9 [cs.LG]. **Conference Paper ICLR**, 2015.

KRAUS, M.; FEUERRIEGEL, S. Decision support from financial disclosures with deep neural networks and transfer learning. **Decision Support Systems**, v. 104, 2017.

LÄNGKVIST, M.; KARLSSON, L.; LOUTFI, A. A review of unsupervised feature learning and deep learning for time-series modeling. **Pattern Recognition Letters**. v. 42, 2014.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. **Nature**. Vol. 521, 2015.

MIT TECHNOLOGY REVIEW. **As Goldman Embraces Automation, Even the Masters of the Universe Are Threatened**. Nanette Byrnes. 7 fev. 2017. Disponível em <www.technologyreview.com/s/603431/as-goldman-embraces-automation-even-the-masters-of-the-universe-are-threatened/>. Acesso em 11 jul. 2020.

OLAH, C. **Understanding LSTM Networks**, [online] Disponível em: <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>> Acesso em 12 jul. 2020.

PALAI, U. **Algorithmic Trading**. Disponível em <<https://github.com/umeshpalai/AlgorithmicTrading-MachineLearning>>. Acesso 11 jul. 2020.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**. v. 12., pp. 2825-2830, 2011.

PRADO, M. L. **Advances in Financial Machine Learning**. John Wiley & Sons, 2018.

PRECHELT, L. Early Stopping – But When? **Neural Networks: Tricks of the Trade**. 2ed. c.2, 1998.

ROHRER, B. Disponível em <<https://brohrer.github.io/blog.html>> Acesso em 11 jul. 2020.

RUDER, S. An Overview of Gradient Descent Optimization Algorithms. **arXiv:1609.04747v2** [cs.LG]. 2017.

SCHUMAKER, R.P.; CHEN, H. A quantitative stock prediction system based on financial news. **Information Processing and Management**. 2009

SELVIN, S.; VINAYAKUMAR, R.; GOPALAKRISHNAM, E. A.; MENON, V. K.; SOMAN, K. P. Stock Price Prediction using LSTM, RNN, and CNN-Sliding Window Model. **Interntnl. Conf. Advances in Computing, Comm. and Informatics, ICACCI**. 2017.

SIRIGNANO, J. A. Deep Learning for Limit Order Books. **Quantitative Finance**. v.19, n.4, 2019.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A; SUTSKEVER, I; SALAKHUTDINOV, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**. v. 15, pp. 1929-1958. 2014

TSANTEKIDIS, A. et al. Forecasting stock prices from the limit order book using convolutional neural networks. **2017 IEEE 19th Business Informatics**, 2017.

WIGGLESWORTH, R. Volatility: how ‘algos’ changed the rhythm of the market. Financial Times. 9 Jan. 2019. Disponível em <<https://www.ft.com/content/fdc1c064-1142-11e9-a581-4ff78404524e>>. Acessado em 21 Jul. 2020.

XIONG, R.; NICHOLS, E. P.; SHEN, Y. Deep Learning Stock Volatility with Google Domestic Trends. **Stanford.edu arXiv:1512.04916v3** [q-fin.CP]. 2016.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with Artificial Neural Networks: The State of The Art, **International Journal of Forecasting**, 14, 35–62, 1998.