



DINCON 2017

CONFERÊNCIA BRASILEIRA DE DINÂMICA, CONTROLE E APLICAÇÕES
30 de outubro a 01 de novembro de 2017 – São José do Rio Preto/SP

Previsões da Taxa de Câmbio de Bitcoin por meio de Técnicas de Aprendizado de Máquina

Dennys Ch. A. Mallqui¹

Programa de Pós-Graduação em Ciência da Computação, UFSCar, São Carlos, SP

Ricardo A. S. Fernandes²

Programa de Pós-Graduação em Ciência da Computação, UFSCar, São Carlos, SP

Resumo. Bitcoin é a criptomoeda mais aceita e com um valor de mercado em constante variação. Desta forma, torna-se possível o uso de técnicas baseadas em aprendizado de máquina para realizar previsões destas variações. Assim, este artigo propõe uma metodologia para a previsão da orientação da variação da taxa de câmbio de bitcoin em relação ao dólar, usando Redes Neurais Artificiais, Redes Recorrentes, Modelo Linear Generalizado e *Support Vector Machine*. São usadas as informações de *Blockchain* e também referentes à economia. Por meio dos resultados obtidos notou-se uma melhora em relação a estudos recentes (que só conseguem até 52% de acurácia), com valores de área sob a curva ROC entre 59,12% e 68,59%, bem como uma acurácia entre 58,68% e 67,36%.

Palavras-chave. Bitcoin, Previsão, Preço, Aprendizado, Máquina.

1 Introdução

O estudo das séries temporais que analisa um conjunto de dados (x_t) de um fenômeno determinado, observado num tempo (t), é realizado em diversos campos, como a engenharia, ciência, sociologia e economia [1]. Nas últimas décadas foram desenvolvidas diversas técnicas de previsão de séries temporais relacionadas a dados financeiros, as que incluem técnicas de aprendizado de máquina [1-6]. Porém, o fenômeno das criptomoedas gera grandes mudanças no mundo financeiro com a aparição, no ano de 2009, da moeda chamada *bitcoin* [7].

Bitcoin não precisa de uma instituição ou banco central que faça sua liberação e controle, pois suas transações são avaliadas e armazenadas na própria rede, em um formato *peer-to-peer* [5]. Assim, faz-se uso de *Blockchain*, por meio de mineradores que conseguem resolver problemas computacionais de avaliação, sendo recompensados pelo sistema com um número de *bitcoins* [8]. Esta ausência de controle é um dos fatores para seu sucesso [9].

O crescimento da aceitação de *bitcoin*, é refletido no número médio de transações diárias feitas, sendo 69.000 em 2014, depois 226.000 em 2016 e continua a subir para 292.000 (até maio de 2017). Segundo [10], *bitcoin* tem um comportamento similar com as moedas tradicionais,

¹ dennys.mallqui@dc.ufscar.br

² ricardo.asf@ufscar.br

porém com presença de uma maior volatilidade. Assim, torna-se uma tarefa importante a previsão de comportamento da taxa de câmbio de *bitcoin*.

O estudo da série temporal da taxa de câmbio de *bitcoin* é recente [11]. Em [12] usa-se um método de regressão Bayesiana para classificação binária, chamado *Latent Source Model*. Com base neste método e dados da *Blockchain* foi desenvolvida uma estratégia simples de comercialização (compra-se ao notar subida e vende-se ao notar uma baixa em intervalos de uma hora), na qual consegue quase duplicar o valor do investimento (89%), num período de 50 dias. Porém, segundo [11], não foi possível replicar o experimento de [12] com sucesso, pois não houve como determinar o período de testes usado.

Em [13] foi usado o modelo *Binomial GLM* com dados de *Blockchain* para realizar previsões da variação do preço diário de *bitcoin*. O estudo indica uma acurácia de 98,7% e, para intervalos de 10 minutos, a acurácia variou entre 50% e 55%. Porém, [11] indica que o estudo não garante o uso de dados de validação e, portanto, existe a possibilidade de *overfitting*.

Em termos de técnicas de aprendizado de máquina, em [4] são usadas Redes Neurais Artificiais, onde o método apresentou 55% de acurácia para períodos de uma hora usando dados de *Blockchain*. Também, [11] faz a comparação entre diferentes métodos obtendo até 52% de acurácia, onde os melhores resultados foram obtidos por meio das redes neurais recorrentes, *Long-Short Term Memory* (LSTM) e *Jordan*, usando os dados de *Blockchain*.

Conforme indicado por [4], os dados de *Blockchain* têm um limite para fazer previsões. Em [14], os autores buscam por novos atributos externos à *Blockchain*. Assim, é realizado um estudo de análise de sentimentos em comentários de Twitter, usando *SentiStrength*. Notou-se a existência de correlação cruzada da variação do preço de *bitcoin* (até -35%).

Em [3], além dos comentários de Twitter, os autores estudam a correlação da quantidade de pesquisas da palavra *bitcoin* no *website* Wikipédia e o valor do índice de mercado S&P 500, encontrando em ambos os casos correlações cruzadas significativas com o preço de *bitcoin*.

Finalmente, uma forte correlação cruzada do preço de *bitcoin* com o número de transações e taxas voluntárias pagas aos mineradores fora encontrada por [15]. Estes autores ainda mostram uma alta correlação cruzada com indicadores econômicos de referência mundial como são o preço do petróleo e do ouro. Finalmente, ainda é identificada correlação cruzada significativa com índices de mercado, tais como S&P 500, NASDAQ e DAX.

Baseado no contexto apresentado, este trabalho foca seus estudos na comparação de diferentes técnicas de aprendizado de máquina e apresenta uma acurácia maior na previsão da variação do preço de *bitcoin* com relação ao dólar. Portanto, serão usadas: Redes Neurais Artificiais, Redes Neurais Recorrentes, Modelo Linear Generalizado (GLM – do inglês *Generalized Linear Model*) e *Support Vector Machine* (SVM). As técnicas serão analisadas sob um conjunto de dados gerado num estudo anterior, através de técnicas de seleção dos atributos, considerando não apenas informações da *Blockchain*, mas também agregando dados de valores referentes à economia, como *commodities* e índices de ações.

2 Métodos

2.1 Conjunto de dados e pré-processamento

Foram obtidos dados que representam aproximadamente 7 anos (entre 2010 e 2017) para valores diários da taxa de câmbio de *bitcoin* (BTC/USD *Bitstamp*), através do *website* <http://bitcoincharts.com>. A informação inclui os valores de abertura, mínimo, máximo e de

fechamento da taxa de câmbio; e também o volume de bitcoins comercializados. Foram adicionados os dados diários de *Blockchain* obtidos via o website <http://quandl.com>, os quais incluem o número de transações realizadas, as receitas em *bitcoins* recebidas pelos mineradores dividido pelo número de transações, o *hashrate* que mede o valor de processamento necessário para avaliar as transações e o valor em *bitcoins* das taxas voluntárias entregues por usuários da rede aos mineradores para acelerar a validação de suas transações. Considerando estudos como [15, 16] foram adicionados atributos representativos da economia; que incluem os valores diários do preço do petróleo e ouro; e os índices de mercado S&P 500, NASDAQ e DAX.

Na etapa de pré-processamento, foi atribuída à classe o valor “1”, se o preço de fechamento de *bitcoin* num tempo (t) for maior ou igual ao dia anterior ($t-1$). Caso contrário, foi atribuído o valor “0”. Foi usada a metodologia similar à proposta em [16], para cada classe num tempo t , onde considerou-se como entradas, os dados históricos das variações dos dias anteriores até 7 dias ($t-1$ até $t-7$).

Também foi aplicada a média móvel ponderada (WMA – *Weighted Moving Average*) de 30 dias corridos aos valores de abertura, mínimo, máximo e fechamento do preço de *bitcoin*, os dados de *Blockchain* e os valores de referência do mercado.

Foram usados dois conjuntos de dados, o primeiro usado para comparar os resultados com os obtidos por [11], desde 19 de agosto de 2013 até 19 de julho de 2016 (1066 casos); e um conjunto maior desde 1 de abril de 2013 até 1 de abril de 2017 (1462 casos). No primeiro intervalo, foi considerado 80% dos dados para treinamento e os 20% restantes para validação, conforme empregado por [11]. Já o segundo intervalo foi composto por 75% dos dados para treinamento e os 25% mais recentes para validação.

2.2 Seleção de atributos

Segundo o indicado na seção anterior são considerados até 85 atributos possíveis de entrada. Portanto, visando reduzir a dimensionalidade da entrada, fez-se necessário o emprego de métodos para selecionar os atributos considerados mais relevantes.

Num estudo anterior feito pelos autores, usando os dados de treinamento e o software Weka 3.8.1, foram testadas distintas técnicas onde foram selecionados os 20 atributos mais relevantes segundo o índice de correlação (*corr*), algoritmos *relief* e ganho de informação (*infogain*). Também foi usada a técnica de componentes principais (PCA – *Principal Components Analysis*) e de seleção de subconjunto (*cfs*).

Como resultado do estudo anterior, o melhor conjunto de dados testados para o primeiro intervalo foi obtido pela técnica de ganho de informação e para o segundo intervalo foram selecionados os atributos pela técnica *relief*.

2.3 Uso de Técnicas de Aprendizado de Máquina

Foram comparadas diversas técnicas de aprendizado de máquina para fazer as predições: Redes Neurais Artificiais com arquitetura multicamada (MLP – *Multilayer Perceptron*), Redes Neurais Recorrentes com arquitetura *Elman-Jordan*, o Modelo Linear Generalizado e o modelo *Support Vector Machine*. Ressalta-se que a implementação foi realizada via o software Matlab.

2.3.1 Redes Neurais Artificiais

Foram utilizadas redes neurais artificiais com arquitetura multicamada (MLP – *Multilayer Perceptron*), devido a sua flexibilidade e bons resultados em outros estudos como [16].

Foi selecionada a função de transferência tangente hiperbólica para as camadas neurais intermediárias. No caso da camada neural de saída foram testadas as funções de transferência sigmoide logística e a função exponencial normalizada (*softmax*).

Os algoritmos de treinamento empregados foram o de Levenberg-Marquardt e a gradiente de descendente com *momentum* e taxa de aprendizado adaptativa (*traingdx*). Assim, diversas configurações com uma, duas e três camadas neurais intermediárias foram testadas, com combinações de 5, 10, 15, 20, 25, 30 e 35 neurônios com um número de épocas variando entre 20 até 500 (no caso de treinamentos usando *trainglx* foi testado até 1000 épocas).

2.3.2 Redes Neurais Recorrentes

Representam uma versão de redes neurais que fazem uso de sinais de saída das camadas neurais intermediárias ou das camadas de saída como entradas, gerando informações de “contexto” para a previsão. Para o estudo foi usada uma rede recorrente com arquitetura Jordan que é usada no estudo de [11].

Como o caso das redes com arquitetura MLP, a rede recorrente foi configurada usando a função de transferência tangente hiperbólica para as camadas ocultas e as funções de saída testadas foram a logística e a exponencial normalizada (*softmax*). Os métodos de treinamento testados foram Levenberg-Marquardt e *traingdx*.

Igualmente, foram testadas configurações de uma, duas e três camadas neurais intermediárias de 5, 10, 15, 20, 25, 30 e 35 neurônios em cada. O treinamento foi realizado com 20 até 500 épocas (no caso de uso do método *traingdx* foi testado até 1000 épocas).

2.3.3 Modelo Linear Generalizado

É uma generalização do modelo de regressão linear muito usado em estudos de séries temporais, a qual foi testada devido aos bons resultados reportados pelo estudo de [13].

O modelo apresenta três componentes, a saber: a distribuição de probabilidade da variância da variável dependente, o preditor que afeta as variáveis dependentes e a função de enlace da variável dependente com as independentes. Foi usada a distribuição binomial, o preditor linear e funções de enlace logarítmica, logística e a complementar logística.

3.3.4 Support Vector Machine

É um modelo da família de algoritmos que são implementados em problemas de classificação, reconhecimento, regressão e previsões de séries temporais [17], sendo a ideia básica a construção de um hiperplano que possa separar os exemplos de variação positiva e negativa do tipo de câmbio de *bitcoin*.

Foi usada como *kernel* a função de base radial e polinomial. Foram usadas diversas funções de transformação como a logística, logística inversa e *ismax*. Também foram testados diversos valores positivos de escala do *kernel*.

4 Resultados

A efetividade das previsões foi medida usando a área sob a curva ROC (*Receiver Operator Characteristic*) devido aos diferentes intervalos de dados não estarem balanceados. Também foi agregada a acurácia para fazer as comparações com outros estudos.

A Tabela 1 apresenta os resultados das previsões feitas no primeiro intervalo com cada algoritmo testado. Assim, é mostrado o modelo com maior área sob da curva ROC obtido para cada classe de algoritmo, bem como a informação da acurácia associada e os principais dados de configuração que permitem gerar tal modelo. É importante lembrar que o primeiro intervalo foi selecionado para realizar a comparação com o estudo de [11].

Tabela 1: Resultados para o primeiro intervalo.

Algoritmo	Área sob da curva ROC	Acurácia	Dados de configuracao
Rede Neural Artificial	68.59%	58.68%	No. de camadas: 3 Neurônios em cada: 10, 30, 5 Função de transferência das camadas ocultas: tangente hiperbólica Função de transferência da camada de saída: sigmoide logística Época de treinamento: 20 Método de treinamento: Gradiente de descenso com momentum e taxa de aprendizado adaptativa (traingdx)
Rede Neural Recorrente	64.32%	67.36%	No. de camadas: 3 Neurônios em cada: 20, 30, 15 Função de transferência das camadas ocultas: tangente hiperbólica Função de transferência da camada de saída: sigmoide logística Época de treinamento: 100 Método de treinamento: Gradiente de descenso com momentum e taxa de aprendizado adaptativa (traingdx)
Modelo Linear Generalizado	58.78%	60.56%	Distribuição: Binomial Preditor: Linear Função de enlace: Logarítmica
Support Vector Machine	59.12%	63.85%	Função de kernel: Polinômica Ordem de polinômio: 6 e 7 Escala de kernel: 1 Função de pontuação: Logística

Devido à base de dados não estar balanceada, ter uma maior área sob a curva ROC não significa necessariamente ter uma maior acurácia. Neste caso, apresenta-se 51% de casos positivos nos dados de treinamento e 58% de casos positivos nos dados de validação.

Na Tabela 2 são mostrados os resultados das previsões para o segundo intervalo, o qual possui maior quantidade de dados observados e também maior porcentagem de dados de validação (25%). Portanto, os valores de área sob da curva ROC e acurácia são menores em comparação com o primeiro intervalo.

Cabe mencionar que no segundo intervalo de dados também há um desbalanceamento, com 52% das amostras contendo variação positiva nos dados de treinamento e 61% dos casos com variação positiva nos dados de validação.

Neste estudo, usando os mesmos dados para treinamento e validação que [11] e com a estrutura de atributos proposta consegue-se melhores resultados de previsão usando os algoritmos de Redes Neurais Artificiais com arquitetura multicamada (três camadas neurais intermediárias) com uma área sob a curva de 68,59% e uma acurácia de 58,68%, sendo maior em comparação com os 52% obtidos pelo estudo em referência [11].

Tabela 2: Resultados para o segundo intervalo.

Algoritmo	Área sob da curva ROC	Acurácia	Dados de configuracao
Rede Neural Artificial	61.85%	54.52%	No. de camadas: 2 Neurônios em cada: 15, 15 Função de transferência das camadas ocultas: tangente hiperbólica Função de transferência da camada de saída: sigmoide logística Época de treinamento: 50 Método de treinamento: Levenberg-Marquardt
Rede Neural Recorrente	58.29%	62.20%	No. de camadas: 3 Neurônios em cada: 30, 25, 5 Função de transferência das camadas ocultas: tangente hiperbólica Função de transferência da camada de saída: sigmoide logística Época de treinamento: 100 Método de treinamento: Gradiente de descenso com momentum e taxa de aprendizado adaptativa (traingdx)
Modelo Linear Generalizado	53.63%	61.37%	Distribuição: Poisson Preditor: Linear Função de enlace: Logarítmica
Support Vector Machine	50.87%	55.34%	Função de kernel: Polinômica Ordem de polinômio: 5 Escala de kernel: 1 Função de pontuação: Logística

Também observa-se que os resultados usando o Modelo Linear Generalizado e *Support Vector Machine* apresentam os piores indicadores, muito longe do indicado em [13, 17].

5 Conclusão

Observa-se que além do algoritmo mencionado na seção anterior que apresenta os melhores resultados, os outros algoritmos também conseguem superar os resultados do estudo em referência, considerando que a forma e estrutura dos atributos selecionados são mais adequadas, as quais usam atributos de mercado para melhorar as capacidades de previsão, destacando o índice DAX que é usado em ambos intervalos.

Porém, não é suficiente para fazer uma correta previsão diária. Assim, também haveria a possibilidade de que os critérios de variação do preço de *bitcoin* acontecessem em períodos menores ao diário, sendo assim, seria necessário usar uma maior granularidade em próximos experimentos. Há ainda a possibilidade de complementar as informações de previsão com dados externos como os comentários de redes sociais ou outras fontes mencionadas nos estudos de [3, 14].

Agradecimentos

Ao apoio financeiro dado pela CAPES, CNPq e FAPESP.

Referências

- [1] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. Published by Springer, (2016).
- [2] E. M. Azoff, *Neural Network Time Series Forecasting of Financial Markets*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., (1994).

- [3] I. Georgoula, D. Pournarakis, C. Bilanakos, D. N. Sotiropoulos, and G. M. Giaglis, "Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices," *SSRN Electron. J.*, no. October 2015, Greece, (2015).
- [4] A. Greaves and B. Au, "Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin," *Social and Information Network Analysis*. Stanford University, (2015).
- [5] J. Iglesias de Ussel, "Bitcoin: a new way to understand Payment Systems," Doctoral dissertation, Massachusetts Institute of Technology, (2015).
- [6] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," *Neurocomputing*, vol. 10, no. 3, 215–236, (1996).
- [7] M. Swan, *Blockchain: Blueprint for a new economy*, O'Reilly Media, Inc., (2015).
- [8] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," (2008).
- [9] I. Miers, C. Garman, M. Green, and A. D. Rubin, "Zerocoin: Anonymous distributed e-cash from bitcoin," *Proc. - IEEE Symp. Secur. Priv.*, 397–411, (2013).
- [10] K. H. McIntyre and K. Harjes, "Order Flow and the Bitcoin Spot Rate," *Appl. Econ. Financ.*, vol. 3, no. 3, 136–147, (2016).
- [11] S. McNally, "Predicting the Price of Bitcoin Using Machine Learning", Ph.D. Thesis, Dublin, National College of Ireland, (2016).
- [12] D. Shah and K. Zhang, "Bayesian regression and Bitcoin," *2014 52nd Annu. Allert. Conf. Commun. Control. Comput. Allert. 2014*, vol. 2, no. 1, 409–414, (2014).
- [13] I. Madan, S. Saluja, and A. Zhao, "Automated Bitcoin Trading via Machine Learning Algorithms," Department of Computer Science, Stanford University, 1–5, (2014).
- [14] M. Matta, I. Lunesu, and M. Marchesi, "The Predictor Impact of Web Search Media On Bitcoin Trading Volumes," in *Knowledge Discovery, 7th International Joint Conference on* (Vol. 1, 620-626), IEEE, (2015).
- [15] S. Vassiliadis, P. Papadopoulos, M. Rangoussi, T. Konieczny, and J. Gralewski, "Bitcoin Value Analysis Based on Cross-Correlations," vol. 22, (2017).
- [16] L. A. Laboissiere, R. A. S. Fernandes, and G. G. Lage, "Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks," *Appl. Soft Comput.*, vol. 35, 66–74, (2015).
- [17] Y. Kara, M. Acar Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange," *Expert Syst. Appl.*, vol. 38, no. 5, 5311–5319, (2011).