Predição da Tendência de Valorização do Bitcoin Usando Técnicas de Aprendizado de Máquina

Bitcoin trend prediction Using Machine Learning Techniques

Antonio Ricardo A. Brasil, Luiz Alberto Pinto, Karin S. Komati
Programa de Pós-Graduação em Engenharia de Controle e Automação (ProPECAut)
Instituto Federal do Espírito Santo Campus
Serra ES-010, Km-6,5 - Manguinhos,
Serra/ES, Brasil
anribrasil@gmail.com, luiz.pt@ifes.edu.br, kkomati@ifes.edu.br

RESUMO

A moeda virtual Bitcoin surgiu em meados de 2008 e atualmente movimenta mais de 200 mil transações por dia. Um dos maiores interesses dos investidores é saber se o preço sofrerá uma queda ou alta, para que assim, possam realizar compras ou vendas. Neste trabalho foram usadas técnicas de aprendizado de máquina para predição da tendência de preço da Bitcoin em determinado dia, considerando a utilização dos classificadores Floresta Aleatória, Rede Bayesiana Gaussiana, Rede Neural Perceptron de Múltiplas Camadas. Os resultados preliminares são promissores e como trabalho futuro, propõe-se o estudo dos métodos de fusão de classificadores na predição da tendência diária (alta ou queda) do preço da Bitcoin. Para avaliação dos resultados serão utilizadas as métricas: Precisão, Revocação, Acurácia, RMSE e Medida F1.

Palavras-chave

Bitcoin, Aprendizado de Máquina, Rede Neural Perceptron de Múltiplas Camadas.

ABSTRACT

Bitcoin emerged in mid-2008 and currently handles more than 200,000 transactions per day. One of the biggest interests of investors is to know at what point the price will fall or rise, so that they can make purchases or sales. In this work, machine learning techniques were used to predict Bitcoin's price trend on a given day, considering the use of the classifiers Random Forest, Gaussian Bayesian Network and Multilayer Perceptron. The preliminary results are promising and as future work, it is proposed to study methods of classifiers fusion in the prediction of the daily tendency (high or fall) of the Bitcoin price. To evaluate the results will be used the metrics: Precision, Recall, Accuracy, RMSE and F1 Measure.

CCS Concepts

• **Applied Computing→Electronic Commerce;** *Digital Cash* • **Computing methodologies→***Machine learning*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SBSI 2018, June 4th–8th, 2018, Caxias do Sul, Rio Grande do Sul, Brazil.
Copyright SBC 2018.

Keywords

Bitcoin, Machine Learning, Multilayer Perceptron.

1. INTRODUCTION

A moeda Bitcoin surgiu em meados de 2008 por meio do artigo "Bitcoin: *A Peer-to-Peer Electronic Cash System*" [13] e tem como objetivo ser um sistema eletrônico descentralizado de moeda virtual. O termo *Peer to Peer* na Bitcoin se refere às trocas efetuadas online sem a necessidade de um intermediador, dessa forma, toda transação da Bitcoin é efetuada sem a necessidade de uma instituição financeira. Atualmente, para obter uma Bitcoin, o usuário poderá comprá-la por meio de mercados de moedas virtuais, ou poderá fazer a mineração, por meio da disponibilização de processamento de sua máquina. Como a quantidade de moedas é limitada, sua disponibilidade vem diminuindo, tornando cada vez mais difícil minerar uma moeda Bitcoin.

Para garantir a veracidade das transações da Bitcoin, foi criada a *blockchain*, que é uma estrutura de dados que representa uma entrada de contabilidade financeira, ou um registro de todas as transações executadas em determinado endereço da Bitcoin. Desse modo, cada transação é digitalmente assinada com um código, tendo como objetivo garantir a sua autenticidade e a garantia de que ninguém adultere e ataque a originalidade de uma transação, mantendo a integridade da rede Bitcoin [16].

2. APRESENTAÇÃO DO PROBLEMA

Atualmente, a Bitcoin movimenta mais de 200 mil transações por dia, com o preço unitário de R$34.200,00 (referente ao dia 22/02/2018)[1] e possui uma capitalização de mais de 9 bilhões de dólares[2]. Investidores têm sido atraídos pela forte valorização do preço da moeda [17], mas há o risco da volatilidade. Volatilidade é a possibilidade de queda ou alta do preço do ativo, muitas vezes de forma significativa, em um determinado período de tempo. Por meio da Figura 1 é possível constatar a ocorrência da alta volatilidade do preço da Bitcoin desde abril de 2016. Nesse gráfico, onde o eixo y é o preço do Bitcoin e o eixo x o tempo, é possível visualizar que no decorrer de janeiro de 2018 a abril de 2018 houve uma alta queda no preço, indo de quase 20 mil dólares para cerca de 8 mil dólares. Além de interferências

[1] http://bitcoin.com

[2] http://blockchain.info

ocasionais no mundo, como a proibição da Bitcoin em transações virtuais como em alguns países[3].

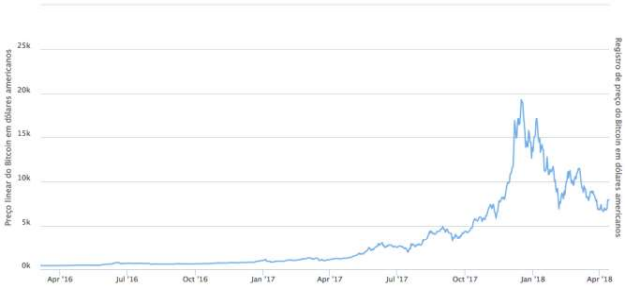

Figure 1. Volatilidade do Preço da Bitcoin de abril de 2016 à abril de 2018 [10].

Um dos maiores interesses dos investidores é ter uma predição desta tendência de valorização (se o preço sofrerá uma queda ou alta), para que assim, possam realizar compras ou vendas, processo denominado no mercado de ações como *trading*. Dessa forma, um sistema inteligente para predição da tendência dos preços da Bitcoin pode ser uma vantagem competitive no Mercado financeiro, auxiliando na tomada de decisões de investidores e análise de riscos da aplicação da moeda virtual. Neste trabalho, serão estudadas a aplicação de técnicas de Aprendizado de Máquina no histórico de preços da Bitcoin para predição da tendência de sua Valorização.

Para a modelagem deste cenário, será usada a notação proposta por Gareth et al. [7]. Dado um determinado dia, a moeda Bitcoin inicia com um preço X, e finaliza com um preço Y. Esta variação será modelada por uma variável binária, descrita na Equação 1, onde 0 (zero) representa que naquele dia o preço da Bitcoin fechará com uma alta, e 1 (um) com uma baixa.

$$B = \begin{cases} 0 & se\ alta \\ 1 & se\ baixa \end{cases} \quad (1)$$

Desse modo, é possível representar o modelo matemático por meio de uma Equação $C = f(A)$, onde A é a entrada: o vetor de atributos correspondente ao dia que se inicia, e f é a função que aplica o algoritmo de aprendizado de máquina nos atributos, tendo como retorno C, a resposta do classificador que é uma saída binária, cujos valores foram definidas em B.

3. PROPOSTA DE SOLUÇÃO

Os resultados preliminares da aplicação dos classificadores: Floresta Aleatória [3], Rede Bayesiana Gaussiana [9] e Rede Neural Perceptron de Múltiplas Camadas [14] foram promissores.

A proposta deste trabalho é aplicar métodos de fusão de classificadores para aumentar as métricas de acurácia obtido pelas técnicas de aprendizado de máquina usadas separadamente e independentemente. A fusão assume que cada classificador tem conhecimento sobre todo o espaço de características e agrega a opinião coletiva de todos. Pretende-se usar os seguintes métodos de fusão de classificadores [8]: voto majoritário ponderado e a fusão por Bayes. O voto majoritário ponderado, basicamente, atribui pesos aos classificadores, que serão avaliados, assim é possível dar maior importância ao classificador que possui maior precisão nos acertos. Já o método fusão por Bayes utiliza de uma matriz de confusão para identificar os acertos de cada classificador.

4. PROJETO DE AVALIAÇÃO DA SOLUÇÃO

A base de dados que será utilizada nos experimentos está disponível no próprio site da BitCoin[4], que contém as transações desde o ano de 2011, de hora em hora. A base de dados está em formato CSV e contém as informações da data e hora, valor de abertura, maior preço do dia, menor preço do dia, preço de fechamento do dia, preço médio da Bitcoin do dia, além de outras informações.

Será usada a validação *Holdout*, na qual a base de dados será dividida em um percentual para dados de teste e o restante para o treinamento dos algoritmos. Os experimentos serão realizados com os seguintes percentuais para teste e treino: 20% e 80%; 30% e 70%; e 40% e 60%.

Para avaliar os resultados serão utilizadas as métricas de Precisão, Revocação, Acurácia, RMSE (*Root Mean Square Error*, em português Raiz do Erro Médio Quadrático) e a Medida F1. A precisão é o número de dados previstos acertadamente de uma determinada classe e a revocação representa o número de dados de uma classe que foram previstos na classe correta. A medida F1 é uma média harmônica ponderada da Precisão e da Revocação, onde seu melhor valor é 1 (um) e o seu pior valor é 0 (zero). O RMSE avalia o desempenho do classificador considerando o desvio padrão da diferença entre o valor predito e o valor real da amostra.

5. ATIVIDADES JÁ REALIZADAS

Uma das atividades realizadas foi um estudo do estado da arte, a seguir citam-se os artigos considerados mais relevantes.

Um dos recentes trabalhos [6] utilizou as características da rede *blockchain* para prever o preço da Bitcoim, aplicando técnicas de classificação SVM (*Support Vector Machine* ou Máquina de Vetor de Suporte), Regressão Linear e Redes Neurais com duas camadas, obtendo as acurácias respectivas de 53,7%, 54,3%, 55,1%, entretanto os autores concluíram que a sua previsibilidade é limitada, pois os preços são em sua maior parte ditado por trocas, cujo comportamento está fora do *blockchain*.

No trabalho de Georgoula et al. [5], além do uso do SVM como classificador, foi empregada a análise de sentimentos sobre as consultadas na Wikipédia, mostrando que o grau de interesse público na Bitcoin tem um efeito positive na determinação do seu preço.

No trabalho de Madan, Saluja e Zhao [12] foram comparadas as tendências de preço do Bitcoin com dados obtidos no *Google Trends* e o volume de *tweets* positivos postados na rede social Twitter. Dessa forma, o trabalho encontrou uma correção cruzada positive nestes websites, especialmente entre o preço Bitcoin e os dados do Google Trends.

O trabalho de Devavrat e Kang [15] utilizou a regressão Bayesiana para prever a variação de preço do Bitcoin. De acordo

[3] http://fortune.com/2018/01/17/china-bitcoin-cryptocurrency-crackdown/

[4] http://bitcoincharts.com/charts

com os resultados de seus experimentos de simulação, foi possível duplicar os investimentos em um tempo inferior a 60 dias.

O trabalho de Amjad e Shah [1] propôs uma análise da Bitcoin em tempo real para prever os preços futuros para execução de *trades* (compra e venda de Bitcoins). Os pesquisadores constataram que alguns métodos clássicos de previsão de séries temporais, como o ARIMA, produziram previsões precárias, e que a proposta do trabalho que usava algoritmos simples escaláveis obteve um alto retorno do investimento em Bitcoin, com precisão em torno de 60% a 70%.

5.1 Experimentos Preliminares

Foram realizados experimentos preliminares usando uma base de dados com as transações de 2011 a 2017, que continha mais de 80.000 linhas. Após a obtenção da base de dados, foi necessário realizar um pré-processamento para a remoção das linhas com valores iguais a zero ou valores nulos. Após a etapa de pré-processamento, foi criada a variável *B*, conforme a Equação 1.

Os classificadores utilizados nos experimentos preliminares foram: Floresta Aleatória [3], Rede Bayesiana Gaussiana [9] e Rede Neural Perceptron de Múltiplas Camadas [2]. Os resultados obtidos são apresentados na Tabela 1, onde é possível notar que o algoritmo que obteve maior precisão foi a Rede Neural Perceptron de Múltiplas Camadas, obtendo 70% na medida F1.

Tabela 1. Resultado dos algoritmos aplicados.

	Precis.	**Revoc.**	**F1**
Floresta Aleatória	69%	70%	70%
Rede Bayesiana Gaussiana	63%	67%	59%
Rede Neural Perceptron de Múltiplas Camadas	71%	70%	70%

A precisão de 70% atingida nos resultados parciais deste trabalho foram melhores que o resultado do trabalho de Greves e Au [6] e Madan, Saluja e Zhao [11], que obtiveram respectivamente 53% e 55% de precisão.

6. CONCLUSÕES

Os resultados obtidos são promissores, e são motivadores para continuar os estudos para melhorar a acurácia da predição. A proposta é usar duas técnicas de fusão de classificadores: voto majoritário ponderado e a fusão por Bayes. O principal objetivo em combinar múltiplos classificadores é produzir uma decisão final que seja melhor que aquela produzida por um único classificador.

A proposta de uso de fusão de classificadores não foi detectada no estudo inicial nos artigos correlatos de predição de tendência de valorização da Bitcoin e portanto, acredita-se que esta será a principal contribuição científica do estudo.

Como trabalho futuro, é possível a aplicação de mais classificadores e mais técnicas de fusão. Outra possibilidade seria a análise de outras moedas virtuais, por exemplo, a *Ethereum*, uma moeda virtual que está em ascensão nos últimos anos. Essa análise seria feita por meio da obtenção e utilização dos dados de transações da moeda e um estudo seria avaliar se o seu crescimento continuará nos próximos anos, uma vez que ela sofreu uma valorização no seu preço similar à da Bitcoin [4].

7. REFERENCES

[1] M. Amjad and D. Shah. Trading Bitcoin and online time series prediction. In O. Anava, A. Khaleghi, M. Cuturi, V. Kuznetsov, and A. Rakhlin, editors, *Proceedings of the Time Series Workshop at NIPS 2016, volume 55 of Proceedings of Machine Learning Researc*h, pages 1–15. PMLR, 09 Dec 2017.

[2] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[3] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[4] C. M. Cap. Crypto-currency market capitalizations. *coinmarketcap*. com, 2014.

[5] 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. 2015

[6] A. Greaves and B. Au. Using the Bitcoin transaction graph to predict the price of Bitcoin. 2015.

[7] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[8] L. I. Kuncheva. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.

[9] D. D. Lewis. Naive (bayes) at forty: The independence assumption in information retrieval. *In European conference on machine learning*, pages 4–15. Springer, 1998.

[10] W. M. LLC. Gráfico do histórico de preço do bitcoin. http://buybitcoinworldwide.com/pt-br/preco, 2018. Acessado em 12 de abril de 2018.

[11] I. Madan, S. Saluja, and A. Zhao, Automated Bitcoin trading via machine learning algorithms, 2015.

[12] M. Matta, I. Lunesu and M. Marchesi. Bitcoin spread prediction using social and web search media. In *UMAP Workshops*, 2015.

[13] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, 2008.

[14] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and M. Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems,* 8(7):579-588, 2009.

[15] D. Shah and K. Zhang. Bayesian regression and Bitcoin. In *Communication, Control and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, pages 409-414. IEEE, 2014.

[16] M. Swan. *Blockchain: Blueprint for a new economy.* "O'Reilly Media, Inc.", 2015.

[17] J. Young. Billionaire investor novogratz: Institutional investors will soon adopt Bitcoin. https://cointelegraph.com/news/billionaire-investor-novogratz-institutional-investors-will-soon-adopt-bitcoin, 2017.