

A data-driven machine learning algorithm for financial market prediction

Mahmut Bağcı

Department of Software Development, School of Advanced Science, Yeditepe University, Istanbul, Turkey

Received: 13 June 2021, Accepted: 16 June 2021

Published online: 5 July 2021.

Abstract: A machine learning algorithm, that is based on Gabor transform (spectrogram) and singular value decomposition, is built to classify and predict financial market trends. The reliability of the classifier is demonstrated by randomly generated time series, and effectiveness of the trend prediction algorithm is examined by the practical implementation of the NASDAQ-100 stock markets. It is manifested that price pattern of stock markets can be recognized and trend of the considered stock markets can be predicted effectively via the developed algorithm.

Keywords: Stock market prediction, Machine learning, Pattern recognition, Data-driven methods.

1 Introduction

Developments of computer sciences in the last two decades trigger the study of machine learning (ML) algorithms and application of them across all science fields and industrial areas such as health care, transportation or financial services. Nowadays, the ML methods have a significant role in financial markets. It is important to analyze the trend of stock market to identify investment strategies with high return albeit low risk [1]. Recently, data driven machine learning methods are widely applied for stock market prediction and algorithmic trading in the financial markets [2]. These prediction algorithms are developed by different techniques such as artificial neural networks [3], decision trees [4], vector machine learning [5] or genetic algorithms [6].

Financial markets generate huge amount of electronic data and investors can reach this data almost instantly. Accordingly, almost all machine learning approximations involve processing the huge data that are generated by price series of financial markets. Thus, reduction of the data size is a critical part of an effective prediction algorithm [1], [2].

In this study, a machine learning classifier is constructed to recognize and predict the trend of stock markets. The pattern classification (or recognition) of the nonlinear stock market price series are realized by the application of singular value decomposition (SVD) to Gabor transform (spectrogram) of price series [2]. The study is outlined as follows: In section 2, the numerical methods will be given for the classification of the price series and randomly generated time series will be used to test the effectiveness of the method. Then, In section 3, the developed classification algorithm will be implemented with real stock price series from NASDAQ-100. In section 4, results of the study will be summarized.

2 The Classification algorithm

In order to predict the trend of any stock market, the Gabor transform of stock market time series are taken to generate spectrograms, and SVD reduction of spectrograms are applied to create a library of the most dominant SVD modes. After the library is built, the spectrogram of examined stock market series will be calculated and the modes of this spectrogram are matched with library modes to classify the corresponding time series.

2.1 The Gabor transform (Spectrogram)

A spectrogram shows a time series in both the time and spectral domain and it can be obtained by the Gabor transform [2]. These spectrograms are expected to be unique for each stock market price series. Thus, they can be applied as the basis of a pattern classification (or recognition) algorithm for determining the trend of considered stock market. As an example, stock market price series of Apple Inc. (with ticker symbol AAPL in NASDAQ-100) and corresponding spectrogram are shown in Fig. 1.

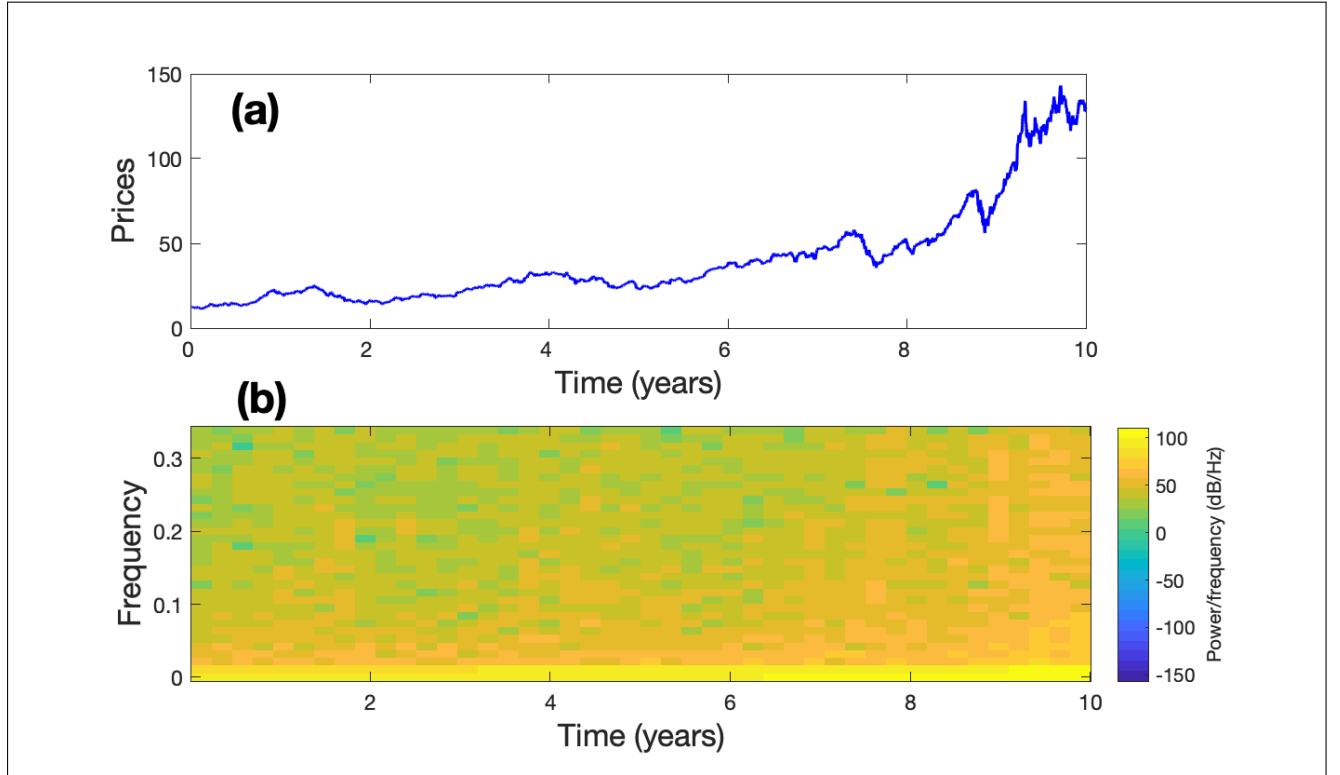


Fig. 1: (a) Stock price series of AAPL and (b) corresponding spectrogram.

2.2 Pattern classification and recognition

To determine a stock market trend with library entries, at first, the spectrogram (S_k) is computed for a large number of stock market price series where k ranges from 1 to M , and SVD is applied to each spectrogram:

$$S_k = U_k \Sigma_k V_k^* \quad (1)$$

where the SVD modes are denoted by

$$U_k = [u_{k1} u_{k2} \dots u_{kn}]. \quad (2)$$

For each k value, the first m ($m < n$) modes (low-rank) approximation of U_k are picked up and stored as the modes library U_L such that

$$U_L = [\tilde{U}_1 \tilde{U}_2 \dots \tilde{U}_M] \quad (3)$$

where the k -th sub-library \tilde{U}_k contains the first m (dominant) modes of U_k :

$$\tilde{U}_k = [u_{k1} u_{k2} \dots u_{km}]. \quad (4)$$

After the library is built, the same procedure is carried out for the examined market stock price series and the first m SVD modes of the spectrogram are kept. These modes and library modes are matched by L_1 -norm search to find a vector

$$a = \arg \min_a \|a\|_1 \tag{5}$$

subject to

$$U_L \cdot a = u_{m1}. \tag{6}$$

The L_1 -norm minimization generates a sparse vector a , i.e. only a small portion of the elements are non-zero. The elements of vector a act as a classifier for determining the sub-library that stock market trend falls into. Thus if the largest element falls into the i -th sub-library, the recognized trend is T_i .

To test performance of the classification algorithm, library of the stock market prices are constructed by randomly generated 3000 time series in Matlab. It is assumed that each time series (T_i) is sampled at 5000 points with a high volatility, and the initial price is normalized to 1. For each spectrogram of the time series, the first 10 SVD modes are kept. One of the test time series (blue line) and the recognized time series (red line) is shown in Fig. 2(a). Furthermore, pattern recognition performance of the classifier is examined by 500 test series, and relative error values are displayed in Fig. 2(b) where the relative error defined as

$$\text{relative error} = \frac{\max(\text{error})}{\max(\text{price}) - \min(\text{price})}. \tag{7}$$

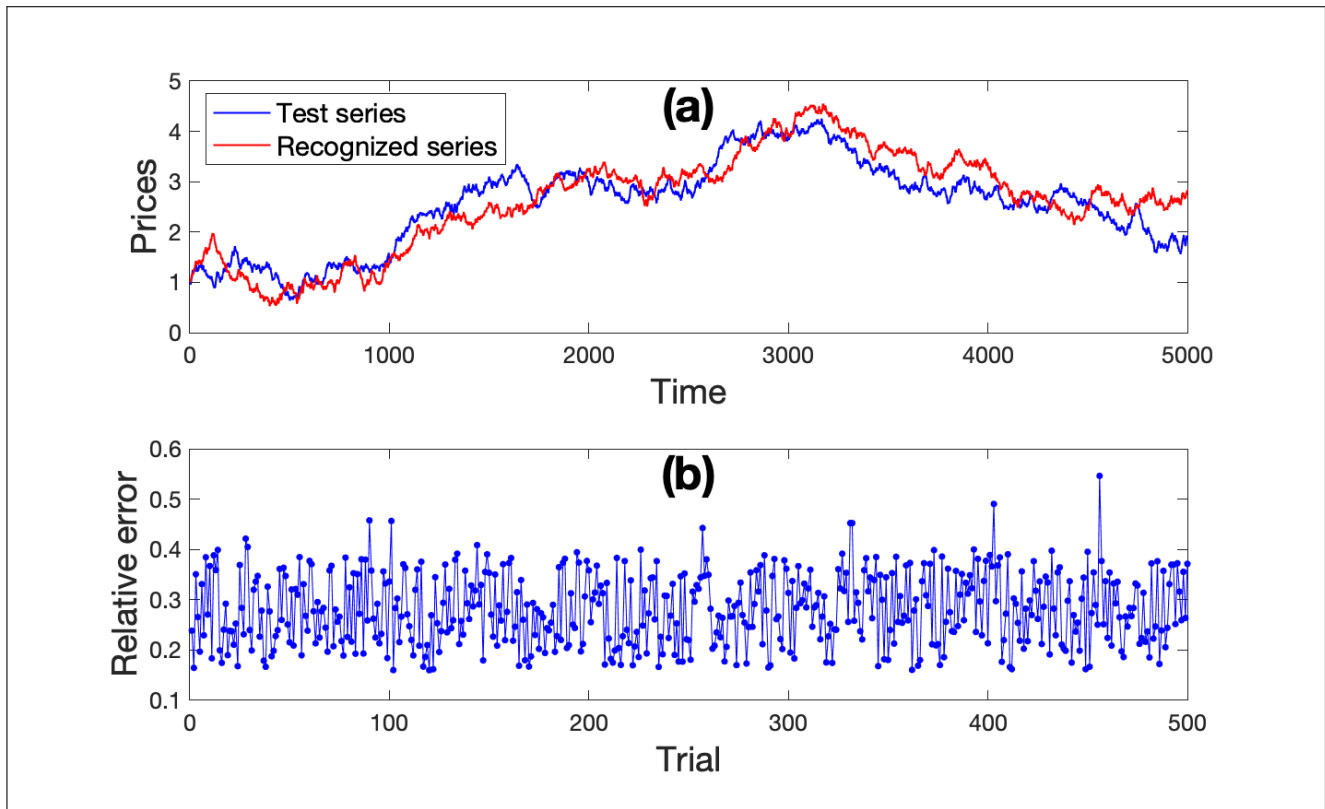


Fig. 2: (a) A test time series (blue line) and the recognized time series (red line). (b) Relative error of the classifier for randomly generated 500 test series.

It can be seen that the relative error is smaller than 0.4 for the most of test time series. It is also observed that the trend of test series are consistent with the recognized series even when the relative error are considerably large. Thus pattern recognition can be accomplished for high-volatility stocks.

3 Trend prediction

In order to predict trend of any stock market, a library of modes is created by the SVD reduction of all stock market price series, and the considered stock market price series is matched with library modes asynchronously. This asynchronous comparison is done as follows: the library modes are generated from the time series between Jan.1, 2015 and Dec. 31, 2018 for all stock markets, and modes of the considered stock market is generated from price series between Jan. 1, 2016 and Dec. 31, 2019. Then trend of the examined stock market is recognized and trend of the recognized price series between Jan. 1, 2019 and Dec.31, 2019 is expected (predicted) to be trend of the examined stock market between Jan. 1, 2020 and Dec. 31, 2020.

To test the prediction algorithm, a library of modes is constructed for 91 stock markets of the NASDAQ-100. The initial price is normalized to 1 for each stock price series, and the first 10 dominant modes are stored in the library for each stock. Then, as a practical implementation, real stock market prices of Alphabet Inc. (the company formerly known as Google with ticker symbol GOOGL) and the predicted price series for the time period between Jan. 1, 2020 and Dec. 31, 2020 are shown in Fig. 3. It can be seen that the real (blue line) and predicted (red line) price series are consistent (see Fig. 3 right panels).

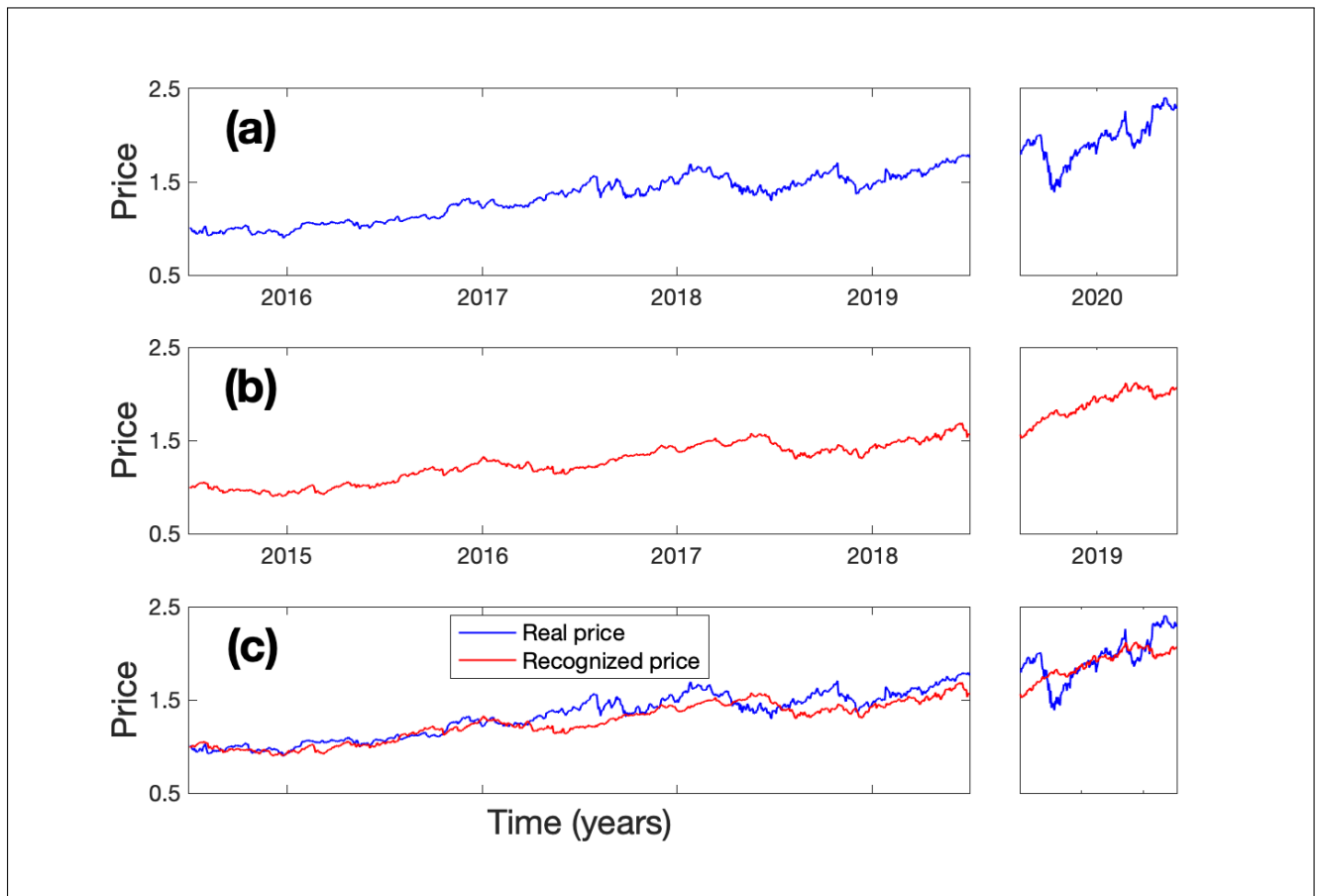


Fig. 3: (a) Real stock price of GOOGL; (b) Price series that is recognized by the classification algorithm; (c) Recognized price series is compared with the real price series.

Similarly, to examine performance of the developed machine learning algorithm, the relative errors of predictions are displayed in Fig. 4 for 91 stock markets of the NASDAQ. It is observed that, for the most of cases, the difference (relative error) between the real stock price series and predicted series are smaller than 0.5. In particular, the best and the worst prediction cases are shown by 'b' and 'c' points in Fig. 4 (a) (with red circles), and the real price series and predicted price series are plotted for these stock markets in Fig. 4 (b) and (c), respectively. It is also noted that, the increasing or decreasing trends over the testing period are truly estimated for 73 of 91 stock markets by the developed machine learning classifier. These practical results demonstrate success of the developed method in stock market trend prediction.

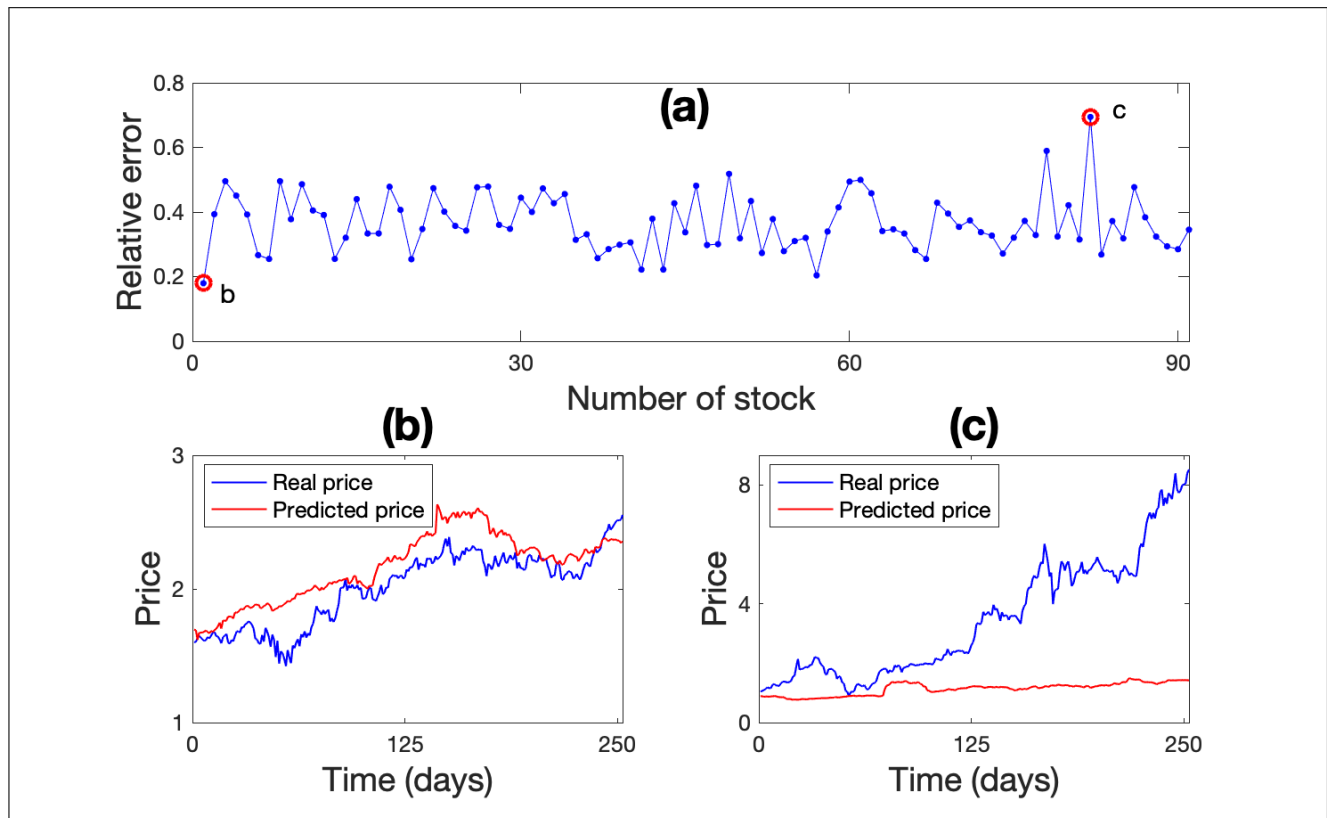


Fig. 4: (a) The relative errors of predictions for 91 stock markets. The real (blue line) and predicted (red line) price series (b) for the best and (c) for the worst prediction cases.

4 Conclusions

A machine learning algorithm, that is based on Gabor transform (spectrogram) and singular value decomposition, is constructed to classify and predict financial market trends. The reliability of the classifier is demonstrated by randomly generated time series and effectiveness of the trend prediction algorithm is examined by the practical implementation of NASDAQ-100 stock markets. It has shown that pattern of the stock price series can be recognized and trend of the considered stock prices can be predicted effectively via the developed algorithm.

It is important to note that the only input of the algorithm is the stock price series and this fact makes the algorithm independent from the exchange market and the number of stock markets that are considered. Furthermore, the algorithm can be applied to different time intervals with any frequency time series value.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors have contributed to all parts of the article. All authors read and approved the final manuscript.

References

- [1] Strader, T. J., Rozycki, J. J., Root, T. H. and Huang, Y.-H. J., "Machine learning stock market prediction studies: Review and research directions", *Journal of International Technology and Information Management*, 28(4), 63-83, 2020.
- [2] Kutz, J. N., "Data-Driven Modeling and Scientific Computation", Oxford University Press, New York, 2013.
- [3] Kim, K. J., and Lee, W. B., "Stock market prediction using artificial neural networks with optimal feature transformation. *Neural computing and applications*", 13(3), 255-260, 2004.
- [4] Basak, S., Kar, S., Saha, S., Khaidem, L. and Dey, S. R., "Predicting the direction of stock market prices using tree-based classifiers", *The North American Journal of Economics and Finance*, 47, 552-567, 2019.
- [5] Yu, L., Chen, H., Wang, S., and Lai, K. K., "Evolving least squares support vector machines for stock market trend mining", *IEEE Transactions on evolutionary computation*, 13(1), 87-102, 2008.
- [6] Kim, H. J., and Shin, K. S., "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets", *Applied Soft Computing*, 7(2), 569-576, 2007.