Computational Intelligence / Volume 33, Issue 2 / pp. 241-261

Original Article

Forecasting Financial Time Series with Grammar-Guided Feature Generation

Anthony Mihirana de Silva ✉, Richard I. A. Davis, Syed A. Pasha, Philip H. W. Leong

First published: 04 January 2016
https://doi.org/10.1111/coin.12083

Abstract

The application of machine learning techniques to forecast financial time series is not a recent development, yet it continues to attract considerable attention because of the difficulty of the problem that is compounded by the nonlinear and nonstationary nature of the time series. The choice of an appropriate set of features is crucial to improve forecasting accuracy of machine learning techniques. In this article, we propose a systematic way for generating rich features using context-free grammars. Our proposed methodology identifies potential candidates for new technical indicators that consistently improve forecasts compared with some well-known indicators. The notion of grammar families as a compact representation to generate a rich class of features is exploited, and implementation issues are discussed in detail. The proposed methodology is tested on closing price data of major stock market indices, and the forecasting performance is compared with some standard techniques. A comparison with the conventional approach using standard technical indicators and naive approaches is shown.

References

Brockwell, P. J., and R. A. Davis. 2002. *Introduction to Time-Series and Forecasting*. Springer: New York.

Google Scholar

Cheng, W., W. Wagner, and C. H. Lin. 1996. Forecasting the 30 Year US Treasury bond with a system of neural networks. *Journal of Computational Intelligence in Finance*, **4**: 10–16.

Google Scholar

Cortes, C., and V. Vapnik. 1995. Support-vector networks. *Machine Learning*, **20**(3): 273–297.

Web of Science® | Google Scholar

Ding, C., and H. Peng. 2003. Minimum redundancy feature selection from microarray gene expression data,. *In* Proceedings of the IEEE Computer Society Conference on Bioinformatics. CSB '03. IEEE Computer Society, Washington, DC. http://dl.acm.org/citation.cfm?id=937976.938050.

Google Scholar

Eads, D., K. Glocer, S. Perkins, and J. Theiler. 2005. Grammar-guided feature extraction for time-series classification. *In* Proceedings of the 9th Annual Conference on Neural Information Processing Systems (NIPS), Vancouver, BC.

Google Scholar

Edwards, R. D., J. Magee, and W. H. C. Bassetti. 2007. *Technical Analysis of Stock Trends*. CRC Press: Boca Raton, FL.

Google Scholar

Espinoza, M., J. Suykens, and B. De Moor. 2005. Load forecasting using fixed-size least squares support vector machines. *In Computational Intelligence and Bioinspired Systems*. Springer: Barcelona, Spain, pp. 1018–1026.

Google Scholar

Franke, J., W. Härdle, and C. M. Hafner. 2008. *Statistics of Financial Markets: An Introduction*. Springer: Berlin.

Google Scholar

Guo, H., L. B. Jack, and A. K. Nandi. 2005. Feature generation using genetic programming with application to fault classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **35**(1): 89–99.

PubMed | Web of Science® | Google Scholar

Guyon, Issabelle. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, **3**: 1157–1182.

Google Scholar

Hall, J. W. 1994. Adaptive selection of US stocks with neural nets. *In Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*. *Edited by* G. J. Deboek. Wiley: New York, pp. 45–65.

Google Scholar

Hanchuan, P., L. Fuhui, and D. Chris. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(8): 1226–1238.

PubMed | Web of Science® | Google Scholar

Hong, G., and A. K. Nandi. 2005. Breast cancer diagnosis using genetic programming generated feature. *In* IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, pp. 215–220.

Google Scholar

Huan, L., and Y. Lei. 2005. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, **17**(4): 491–502.

Web of Science® | Google Scholar

Huang, C. L., and C. Y. Tsai. 2009. A Hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications*, **36**(2): 1529–1539.

Web of Science® | Google Scholar

Huang, S. C., and T. K. Wu. 2008. Integrating GA-based time-scale feature extractions with SVMs for stock index forecasting. *Expert Systems with Applications*, **35**(4): 2080–2088.

Web of Science® | Google Scholar

Huang, Z., H. Chen, C. J. Hsu, W. H. Chen, and S. Wu. 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, **37**(4): 543–558.

Web of Science® | Google Scholar

Hyndman, R. J. 2010. Why every statistician should know about cross-validation. Available at: http: //robjhyndman.com/hyndsight/crossvalidation/. Accessed October 20, 2015.

Google Scholar

Hyndman, R. J., G. Athanasopoulos, S. Razbash, D. Schmidt, Z. Zhou, Y. Khan, and C. Bergmeir. 2013. forecast: Forecasting functions for time-series and linear models. http://CRAN.R-project.org/package=forecast. R package version 4.06.

Google Scholar

Il-Seok Oh, Jin-Seon Lee, and Byung-Ro Moon. 2004. Hybrid genetic algorithms for feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**(11): 1424–1437.

PubMed | Web of Science® | Google Scholar

Ince, H., and T. B. Trafalis. 2004. Kernel principal component analysis and support vector machines for stock price prediction, **3**. *In* Proceedings of IEEE International Joint Conference on Neural Networks, pp. 2053–2058.

Google Scholar

Islamaj, R., L. Getoor, and W. J. Wilbur. 2006. A feature generation algorithm for sequences with application to splice-site prediction. *In Knowledge Discovery in Databases: Pkdd 2006*. *Edited by* J. G. Carbonell, and J. Siekmann. Springer: Berlin, pp. 553–560.

Google Scholar

Jeffrey, A. R. 2013. quantmod: Quantitative financial modelling framework. http://CRAN.R-project.org/package = quantmod. R package version 0.4-0.

Google Scholar

Kamruzzaman, J., and R. A. Sarker. 2003. Forecasting of currency exchange rates using ANN: a case study, **1**. *In* Proceedings of the International Conference on Neural Networks and Signal Processing, IEEE, pp. 793–797.

Google Scholar

Kim, KJ. 2003. Financial time-series forecasting using support vector machines. *Neurocomputing*, **55**(1): 307–319.

Web of Science® Google Scholar

Kim, M., S. Min, and H. Ingoo. 2006. An evolutionary approach to the combination of multiple classifiers to predict a stock price index. *Expert Systems with Applications*, **31**(2): 241–247.

Web of Science® Google Scholar

Knaus, J. 2010. snowfall: Easier cluster computing (based on snow). http://CRAN.R-project.org/package= snowfall.R package version 1.84.

Google Scholar

Kohavi, R., and G. H. John. 1997. Wrappers for feature subset selection. *Artificial Intelligence*, **97**(1): 273–324.

Web of Science® Google Scholar

Krawiec, K., and B. Bhanu. 2005. Visual learning by coevolutionary feature synthesis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **35**(3): 409–425.

PubMed Web of Science® Google Scholar

Krollner, B., B. Vanstone, and G. Finnie. 2010. Financial time-series forecasting with machine learning techniques: a survey. *In* Proceedings of the European Symposium on Artificial Neural Networks—Computational Intelligence and Machine Learning, Bruges, Belgium, pp. 25–30.

Google Scholar

Lai, T. L., and H. Xing. 2008. *Statistical Models and Methods for Financial Markets*. Springer: New York.

Web of Science® | Google Scholar

Lai, R. K., Chin-Yuan Fan, Wei-Hsiu Huang, and Pei-Chann Chang. 2009. Evolving and clustering fuzzy decision tree for financial time-series data forecasting. *Expert Systems with Applications*, **36**(2): 3761–3773.

Web of Science® | Google Scholar

Lendasse, A., E. de Bodt, V. Wertz, and M. Verleysen. 2000. Non-linear financial time-series forecasting—application to the BEL 20 Stock Market Index. *European Journal of Economic and Social Systems*, **14**(1): 81–91.

Google Scholar

Liang, T., and A. Noore. 2004. A novel approach for short-term load forecasting using support vector machines. *International Journal of Neural Systems*, **14**(05): 329–335.

PubMed | Google Scholar

Lu, C., T. Lee, and C. Chiu. 2009. Financial time-series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, **47**(2): 115–125.

Web of Science® | Google Scholar

Markovitch, S., and D. Rosenstein. 2002. Feature generation using general constructor functions, **49**. *In* Machine Learning. MIT Press: Cambridge, MA, pp. 59–98.

Google Scholar

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2012. e1071: Misc functions of the Department of Statistics (e1071), TU Wien. http://CRAN.R-project.org/package=e1071.R package version 1.6-1.

Google Scholar

Mohandes, M. 2002. Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research*, **26**(4): 335–345.

Web of Science® | Google Scholar

Noorian, F., A. M. de Silva, and P. H. W. Leong. 2015. gramEvol: Grammatical evolution in R. *Journal of Statistical Software*. To appear.

Web of Science® | Google Scholar

Pachet, F., and P. Roy. 2009. Analytical features: a knowledge-based approach to audio feature generation. *EURASIP Journal on Audio, Speech, and Music Processing*, **2009**(1): 1–23.

Web of Science® | Google Scholar

Ritanjali, M., G. Panda, B. Majhi, and G. Sahoo. 2009. Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques. *Expert Systems with Applications*, **36**(6): 10097–10104.

Web of Science® | Google Scholar

Ritthof, O., R. Klinkenberg, S. Fischer, and I. Mierswa. 2002. A hybrid approach to feature selection and generation using an evolutionary algorithm. *In* U.K. Workshop on Computational Intelligence, Birmingham, UK, pp. 147–154.

Google Scholar

Robnik-Šikonja, Marko, and Igor Kononenko. 2003. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.*, **53**(1-2): 23–69.

Web of Science® | Google Scholar

Sapankevych, N., and R. Sankar. 2009. Time-Series Prediction using Support Vector Machines: A Survey. *IEEE Computational Intelligence Magazine*, **4**(2): 24–38.

Web of Science® | Google Scholar

Schabacker, R. W. 1930. *Stock Market Theory and Practice*. BC Forbes Publishing Company: New York.

Google Scholar

Shen, W., X. Guo, C. Wu, and D. Wu. 2011. Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. *Knowledge-Based Systems*, **24**(3): 378–385.

Web of Science® | Google Scholar

Sipser, M. 1997. Context-free grammars. *In Introduction to the Theory of Computation*. *Edited by* M. Lee. PWS Publishing: Boston, pp. 91–122.

Google Scholar

Tay, F. E. H., and L. Cao. 2001. Application of support vector machines in financial time-series forecasting. *Omega*, **29**(4): 309–317.

Web of Science® | Google Scholar

Tsay, R. S. 2005. *Analysis of Financial Time-series*, **543**. Wiley-Interscience: Hoboken, NJ.

Google Scholar

Vapnik, V. 1999. *The Nature of Statistical Learning Theory*. Springer: New York.

Google Scholar

Yang, J., and V. Honavar. 1998. Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems and Their Applications*, **13**(2): 44–49.

Web of Science® Google Scholar

Yu, L., S. Wang, and Lai Kin-Keung. 2005. Mining stock market tendency using GA-based support vector machines. *In* Lecture Notes in Computer Science 3828. Springer: Heidelberg Berlin, pp. 336–345.

Google Scholar

Zapranis, A. 2006. Testing the random walk hypothesis with neural networks. *In Artificial Neural Networks ICANN*, vol. 4132. *Edited by* S. Kollias et al., Lecture Notes in Computer Science. Springer: Berlin Heidelberg, pp. 664–671.

Google Scholar