International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

Forecasting with Machine Learning

Dr. Osman Mohamed Abbas*

Email: phd.sudan@gmail.com

Abstract

For years, people have been forecasting weather patterns, economic and political events, sports outcomes, and more. In this paper we discussed the ways of using machine learning in forecasting, machine learning is a branch of computer science where algorithms learn from data. The fundamental problem for machine learning and time series is the same: to predict new outcomes based on previously known results. Using the suitable technique of machine learning depend on how much data you have, how noisy the data is, and what kind of new features can be derived from the data. But these techniques can improve accuracy and don't have to be difficult to implement.

Keywords: forecasting; machine learning; accuracy.

1. Introduction

Predicting the future is one of the most relevant and challenging tasks in applied sciences. Building effective predictors form historical data demands computational and statistical methods for inferring dependencies between past and short-term future values of observed values as well as appropriate strategies to deal with longer horizons [17]. Machine learning is being incorporated into solutions in every walk of life - home thermostats, health monitoring systems, equipment maintenance, marketing software, etc. In the latest generation of products, machine learning is adding intelligence pretty much everywhere you look [29]. Data is driving this trend. More data is available than ever before, but tools are needed to take advantage of it. Machine learning that allows the computer to "learn" from data even without rules-based programming nicely filling this need for improved analysis [55]. Machine learning is the study of computational methods for improving performance by mechanizing the acquisition of knowledge from experience. Expert performance requires much domain-specific knowledge, and knowledge engineering has produced hundreds of AI expert systems that are now used regularly in industry.

----

* Corresponding author.

Machine learning aims to provide increasing levels of automation in the knowledge engineering process, replacing much time-consuming human activity with automatic techniques that improve accuracy or efficiency by discovering and exploiting regularities in training data. The ultimate test of machine learning is its ability to produce systems that are used regularly in industry, education, and elsewhere [41].

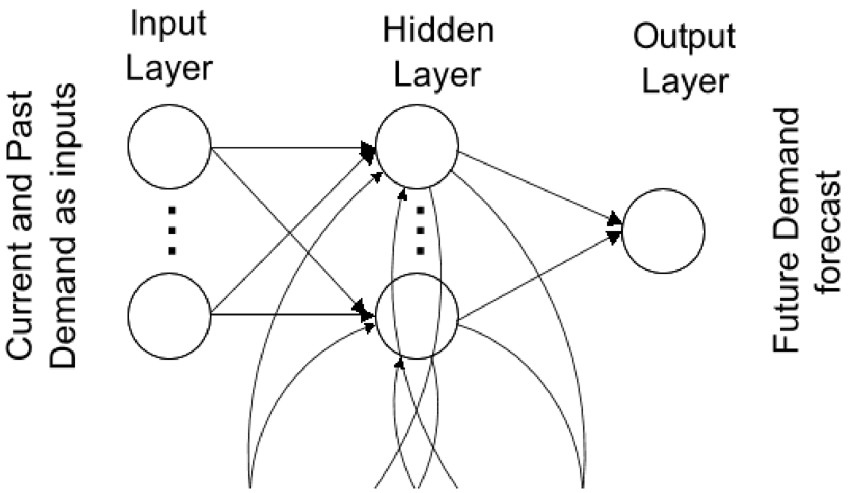

Figure 1: Future Demand Forecasting

2. Machine Learning

The machine learning field, which can be briefly defined as enabling computers make successful predictions using past experiences, has exhibited an impressive development recently with the help of the rapid increase in the storage capacity and processing power of computers. Together with many other disciplines, machine learning methods have been widely employed in bioinformatics. The difficulties and cost of biological analyses have led to the development of sophisticated machine learning approaches for this application area. In this chapter, we first review the fundamental concepts of machine learning such as feature assessment, unsupervised versus supervised learning and types of classification [11]. Machine learning is the subfield of computer science that, according to Arthur Samuel in 1959, gives "computers the ability to learn without being explicitly programmed." Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data– such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), learning to rank, and computer vision [26]. As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis [30]. However, an increasing emphasis on the logical, knowledge-based approach caused a rift between AI and machine learning. Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation, expert systems had come to dominate AI, and statistics was out of favor. Work on symbolic/knowledge-based learning did continue within AI, leading to inductive logic programming, but the more statistical line of research was now outside the field of AI proper, in pattern recognition and information retrieval [18].computer science around the same time. This line, too, was continued outside the AI/CS field, as "connectionism", by researchers from other disciplines including Hopfield, Rumelhart and Hinton. Their main success came in the mid-1980s with the reinvention of back propagation [22]. Machine learning, reorganized as a separate field, started to flourish in the 1990s. The field changed its goal from achieving artificial intelligence to tackling solvable problems of a practical nature. It shifted focus away from the symbolic approaches it had inherited from AI, and toward methods and models borrowed from statistics and probability theory. It also benefited from the increasing availability of digitized information, and the possibility to distribute that via the Internet [33]. Machine learning and data mining often employ the same methods and overlap significantly, but while machine learning focuses on prediction, based on known properties learned from the training data, data mining focuses on the discovery of (previously) unknown properties in the data (this is the analysis step of Knowledge Discovery in Databases). Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve learner accuracy. Much of the confusion between these two research communities (which do often have separate conferences and separate journals, ECML PKDD being a major exception) comes from the basic assumptions they work with: in machine learning, performance is usually evaluated with respect to the ability to reproduce known knowledge, while in Knowledge Discovery and Data Mining (KDD) the key task is the discovery of previously unknown knowledge. Evaluated with respect to known knowledge, an uninformed (unsupervised) method will easily be outperformed by other supervised methods, while in a typical KDD task, supervised methods cannot be used due to the unavailability of training data [12]. Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of some loss function on a training set of examples. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances (for example, in classification, one wants to assign a label to instances, and models are trained to correctly predict the pre assigned labels of a set of examples). The difference between the two fields arises from the goal of generalization: while optimization algorithms can minimize the loss on a training set, machine learning is concerned with minimizing the loss on unseen samples [40].

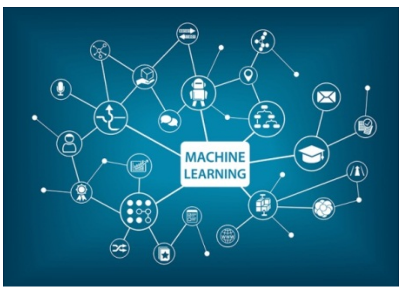

Figure 2: Machine Learning

3. Forecasting

Forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends. A commonplace example might be estimation of some variable of interest at some specified future date. Prediction is a similar, but more general term [39]. Forecasting has applications in a wide range of fields where estimates of future conditions are useful. Not everything can be forecasted reliably, if the factors that relate to what is being forecast are known and well understood and there is a significant amount of data that can be used very reliable forecasts can often be obtained. If this is not the case or if the actual outcome is effected by the forecasts, the reliability of the forecasts can be significantly lower [17]. Climate change and increasing energy prices have led to the use of Egain Forecasting for buildings. This attempts to reduce the energy needed to heat the building, thus reducing the emission of greenhouse gases. Forecasting is used in Customer Demand Planning in everyday business for manufacturing and distribution companies [1]. While the veracity of predictions for actual stock returns are disputed through reference to the Efficient-market hypothesis, forecasting of broad economic trends is common. Such analysis is provided by both non-profit groups as well as by for-profit private institutions (including brokerage houses and consulting companies [9]. Forecasting foreign exchange movements is typically achieved through a combination of chart and fundamental analysis. An essential difference between chart analysis and fundamental economic analysis is that chartists study only the price action of a market, whereas fundamentalists attempt to look to the reasons behind the action. Financial institutions assimilate the evidence provided by their fundamental and chartist researchers into one note to provide a final projection on the currency in question [23]. Forecasting has also been used to predict the development of conflict situations. Forecasters perform research that uses empirical results to gauge the effectiveness of certain forecasting models. However research has shown that there is little difference between the accuracy of the forecasts of experts knowledgeable in the conflict situation and those by individuals who knew much less [33]. Similarly, experts in some studies argue that role thinking does not contribute to the accuracy of the forecast. The discipline of demand planning, also sometimes referred to as supply chain forecasting, embraces both statistical forecasting and a consensus process. An important, albeit often ignored aspect of forecasting, is the relationship it holds with planning. Forecasting can be described as predicting what the future will look like, whereas planning predicts what the future should look like. There is no single right forecasting method to use. Selection of a method should be based on your objectives and your conditions [48].

Figure 3: Forecasting

4. Forecasting with Machine Learning Techniques

Machine learning comes with its own specific set of concerns. Feature engineering, or the creation of new predictors from the data set is an important step for machine learning and can have a huge impact on performance. This engineering can be a necessary way to address the trend and seasonality issues of time series data. In addition, some models encounter issues with how well they fit the data. It is possible that they can both over fit the available data and underperform on new data, or they can under fit and miss the underlying trend [55]. Time series and machine learning approaches do not need to exist in isolation from each other. They can be combined together in order to give you the benefits of each approach. Time series does a good job at decomposing data into trended and seasonal elements. This analysis can then be used as an input into a machine learning model, which can incorporate the trend and seasonal information into its algorithm, giving you the best of both worlds [50]. In the last two decades, machine learning models have drawn attention and have established themselves as serious contenders to classical statistical models in the forecasting community . These models, also called black-box or data driven models , are examples of nonparametric nonlinear models which use only historical data to learn the stochastic dependency between the past and the future. For instance, Werbos found that Artificial Neural Networks (ANNs) outperform the classical statistical methods such as linear regression and Box-Jenkins approaches. A similar study has been conducted by Lapedes and Farber who conclude that ANNs can be successfully used for modeling and forecasting nonlinear time series. Later, other models appeared such as decision trees, support vector machines and nearest neighbor regression. Moreover, the empirical accuracy of several machine learning models has been explored in a number of forecasting competitions under different data conditions (e.g. the NN3,NN5, and the annual ESTSP competitions creating interesting scientific debates in the area of data mining and forecasting [32]. Time series is a sequence of historical measurements of an observable variable at equal time intervals. Time series are studied for several purposes such as the forecasting of the future based on knowledge of the past, the understanding of the phenomenon underlying the measures, or simply a succinct description of the salient features of the series. In this chapter we shall confine ourselves to the problem of forecasting. Forecasting future values of an observed time series plays an important role in nearly all fields of science and engineering, such as economics, finance, business intelligence, meteorology and telecommunication. An important aspect of the forecasting task is represented by the size of the horizon. If the one-step forecasting of a time series is already a challenging task, performing multi-step forecasting is more difficult because of additional complications, like accumulation of errors, reduced accuracy, and in-creased uncertainty [44]. The forecasting domain has been influenced, for a long time, by linear statistical methods such as ARIMA models. However, in the late 1970s and early1980s, it became increasingly clear that linear models are not adapted to many real applications. In the same period, several useful nonlinear time series models were proposed such as the bilinear model , the threshold autoregressive model and the autoregressive conditional heteroscedastic (ARCH) model (see and for a review). However, the analytical study of non-linear time series analysis and forecasting is still in its infancy compared to linear time series [29]. Two main interpretations of the forecasting problem on the basis of historical dataset exist. Statistical forecasting theory assumes that an observed sequence is a specific realization of a random process, where the randomness arises from many independent degrees of freedom interacting linearly [4]. However, the emergent view in dynamical systems theory [23,17] is that apparently random behavior may be generated by deterministic systems with only a small number of degrees of freedom, interacting nonlinearly. This complicated and aperiodic be-havior is also called deterministic chaos [36]. The embedding formulation in (5) suggests that, once a historical record is available, the problem of one-step forecasting can be tackled as a problem of supervised learning. Supervised learning consists in modeling, on the basis of a finite set of observations, the relation between a set of input variables and one or more output variables, which are considered somewhat dependent on the inputs. Once a model of the mapping (5) is available, it can be used for one-step forecasting. In one-step forecasting, the previous values of the series are available and the forecasting problem can be cast in the form of a generic regression problem [34]. Forecasting one-step-ahead consists then in predicting the value of the output when a subset of past observed values (also denoted as query) is given. Machine learning provides a theoretical framework to estimate from observed data a suit-able model of the time dependency [45]. The Nearest Neighbor method is the most trivial example of local approximation applied to the problem of time series forecasting. This method consists in looking through the data set for the nearest neighbor of the current state and predicting that the current state will evolve in the same manner as the neighbor did [27]. The Lazy Learning (LL) is a lazy and local learning machine which automatically adapts the size of the neighborhood on the basis of a cross-validation criterion. The major appeal of Lazy Learning is its divide-and-conquer nature: Lazy Learning reduces a complex and nonlinear modeling problem into a sequence of easily manageable local linear problems, one for each query. This allows to exploit, on a local basis, the whole range of linear identification and validation techniques which are fast, reliable, and come with a wealth of theoretical analyses, justifications, and guarantees [37]. Local learning appears to be an effective algorithm not only for one-step but also for multi-step forecasting. This section discusses some works which used local learning techniques to deal specifically with the long term forecasting problem. In the authors proposed a modification of the local learning technique to take into account the temporal behavior of the multi-step forecasting problem and consequently improve the results of the recursive strategies. In particular modified the PRESS criterion (10) by introducing an iterated version of the leave-one-out statistic. They showed that the iterated PRESS outperforms a non-iterated criterion by assessing the generalization performance of a local one-step predictor on a horizon longer than a single step, yet preserving nice properties of computational efficiency [46].

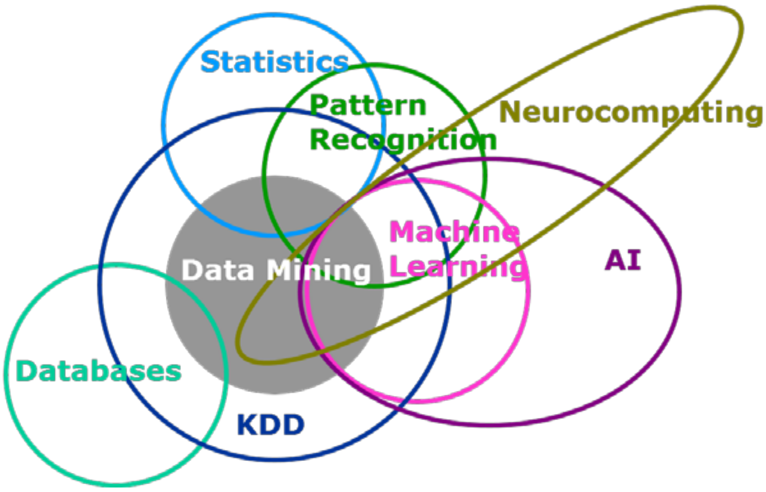

Figure 4: Forecasting Techniques

5. Conclusion

The goal of learning is to find a model which is able to generalize, i.e. able to return good predictions for input values independent of the training set. In a nonlinear setting, it is possible to find models with such a complicate structure that they have null training errors. Are these models good? Typically NOT. Since doing very well on the training set could mean doing badly on new data. This is the phenomenon of overfitting. Using the same data for training a model and assessing it is typically a wrong procedure, since this returns an over optimistic assessment of the model generalization capability. However, when it comes to forecasting there is no silver bullet and what works best may be problem specific. One downside of using machine learning methods for forecasting problems (or any non-parametric model for that matter) is that we can't quantify the uncertainty in our predictions in terms of frequentist confidence or Bayesian credible intervals. This problem can perhaps be partly mitigated by using the block bootstrap to get bootstrapped confidence intervals. If your ultimate goal is more explanatory rather than predictive in nature, you may find that more classical models like state-space models will give you better bang for your buck. Bayesian dynamic linear models (DLMs) in particular work nicely here, because of their flexibility and ease of interpretation

References

[1] Ahmed, N.K., Atiya, A.F., El Gayar, N., El-Shishiny, H.: An empirical comparison of machine learning models for time series forecasting. Econometric Reviews 29(5- 6) (2010)

[2] Allen, D.M.: The relationship between variable selection and data agumentation and a method for prediction. Technometrics 16(1), 125–127 (1974)

[3] Alpaydin, E.: Introduction to Machine Learning, 2nd edn. Adaptive Computation and Machine Learning. The MIT Press (February 2010)

[4] Anderson, T.W.: The statistical analysis of time series. J. Wiley and Sons (1971)

[5] Atkeson, C.G., Moore, A.W., Schaal, S.: Locally weighted learning. AIR 11(1-5), 11–73 (1997)

[6] Ben Taieb, S., Bontempi, G.: Recursive multi-step time series forecasting by per- turbing data. In: Proceedings of IEEE-ICDM 2011(2011)

[7] Ben Taieb, S., Bontempi, G., Atiya, A., Sorjamaa, A.: A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. ArXiv e-prints (August 2011)

[8] Ben Taieb, S., Bontempi, G., Sorjamaa, A., Lendasse, A.: Long-term prediction of time series by combining direct and mimo strategies. In: Proceedings of the 2009 IEEE International Joint Conference on Neural Networks, Atlanta, U.S.A., pp. 3054–3061 (June 2009)

[9] Ben Taieb, S., Sorjamaa, A., Bontempi, G.: Multiple-output modelling for multi-step-ahead forecasting. Neurocomputing 73, 1950–1957 (2010) Machine Learning Strategies for Time Series Forecasting 75

[10] Ben Taieb, S., Bontempi, G., Sorjamaa, A., Lendasse, A.: Long-term prediction of time series by combining direct and mimo strategies. In: International Joint Conference on Neural Networks (2009)

[11] Birattari, M., Bontempi, G., Bersini, H.: Lazy learning meets the recursive least- squares algorithm. In: Kearns, M.S., Solla, S.A., Cohn, D.A. (eds.) NIPS 11, pp. 375–381. MIT Press, Cambridge (1999)

[12] Bontempi, G.: Local Learning Techniques for Modeling, Prediction and Control. PhD thesis, IRIDIA-Universit´e Libre de Bruxelles (1999)

[13] Bontempi, G.: Long term time series prediction with multi-input multi-output local learning. In: Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP 2008, Helsinki, Finland, pp. 145–154 (February 2008)

[14] Bontempi, G., Birattari, M., Bersini, H.: Lazy learners at work: the lazy learning toolbox. In: Proceeding of the 7th European Congress on Intelligent Techniques and Soft Computing, EUFIT 1999 (1999)

[15] Bontempi, G., Birattari, M., Bersini, H.: Local learning for iterated time-series prediction. In: Bratko, I., Dzeroski, S. (eds.) Machine Learning: Proceedings of the Sixteenth International Conference, pp. 32–38. Morgan Kaufmann Publishers, San Francisco (1999)

[16] Bontempi, G., Ben Taieb, S.: Conditionally dependent strategies for multiple-step- ahead prediction in local learning. International Journal of Forecasting (2011) (in press, corrected proof)

[17] Casdagli, M., Eubank, S., Farmer, J.D., Gibson, J.: State space reconstruction in the presence of noise. PHYD 51, 52–98 (1991)

[18] Cheng, H., Tan, P.-N., Gao, J., Scripps, J.: Multistep-Ahead Time Series Predic- tion. In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) PAKDD 2006. LNCS (LNAI), vol. 3918, pp. 765–774. Springer, Heidelberg (2006)

[19] Crone, S.F.: NN3 Forecasting Competition, http://www.neural-forecasting-competition.com/NN3/index.html (last up- date May 26, 2009) (visited on July 05, 2010)

[20] Crone, S.F.: NN5 Forecasting Competition, http://www.neural-forecasting-competition.com/NN5/index.html (last up- date May 27, 2009) (visited on July 05, 2010)

[21] Crone, S.F.: Mining the past to determine the future: Comments. International Journal of Forecasting 5(3), 456–460 (2009); Special Section: Time Series Monitor-ing

[22] Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50(4), 987–1007 (1982)

[23] Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. Physical Review Letters 8(59), 845–848 (1987)

[24] Farmer, J.D., Sidorowich, J.J.: Exploiting chaos to predict the future and reduce noise. Technical report, Los Alamos National Laboratory (1988)

[25] De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting. International Journal of Forecasting 22(3), 443–473 (2006)

[26] De Gooijer, J.G., Kumar, K.: Some recent developments in non-linear time series modelling, testing, and forecasting. International Journal of Forecasting 8(2), 135– 156 (1992)

[27] Guo, M., Bai, Z., An, H.Z.: Multi-step prediction for nonlinear autoregressive mod- els based on empirical distributions. In: Statistica Sinica, pp. 559–570 (1999)

[28] Hand, D.: Mining the past to determine the future: Problems and possibilities. International Journal of Forecasting (October 2008) 76 G. Bontempi, S. Ben Taieb, and Y.-A. Le Borgne

[29] Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: data mining, inference and prediction, 2nd edn. Springer (2009)

[30] Hsu, W., Lee, M.L., Wang, J.: Temporal and spatio-temporal data mining. IGI Pub. (2008)

[31] Ikeguchi, T., Aihara, K.: Prediction of chaotic time series with noise. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sci-ences E78-A(10) (1995)

[32] Kline, D.M.: Methods for multi-step time series forecasting with neural networks. In: Peter Zhang, G. (ed.) Neural Networks in Business Forecasting, pp. 226–250. Information Science Publishing (2004)

[33] Lapedes, A., Farber, R.: Nonlinear signal processing using neural networks: predic- tion and system modelling. Technical Report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM (1987)

[34] Lendasse, A. (ed.): ESTSP 2007: Proceedings (2007)

[35] Lendasse, A. (ed.): ESTSP 2008: Proceedings. Multiprint Oy/Otamedia (2008) ISBN: 978-951-22-9544-9

[36] Lorenz, E.N.: Atmospheric predictability as revealed by naturally occurring ana- logues. Journal of the Atmospheric Sciences 26, 636–646 (1969)

[37] Mat´ıas, J.M.: Multi-output Nonparametric Regression. In: Bento, C., Cardoso, A., Dias, G. (eds.) EPIA 2005. LNCS (LNAI), vol. 3808, pp. 288–292. Springer, Heidelberg (2005)

[38] McNames, J.: A nearest trajectory strategy for time series prediction. In: Pro- ceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling, pp. 112–128. K.U. Leuven, Belgium (1998)

[39] Micchelli, C.A., Pontil, M.A.: On learning vector-valued functions. Neural Com- put. 17(1), 177–204 (2005)

[40] Mitchell, T.M.: Machine Learning. McGraw-Hill, New York (1997)

[41] Owen, S.: Mahout in action. Manning (2012)

[42] Packard, N.H., Crutchfeld, J.P., Farmer, J.D., Shaw, R.S.: Geometry from a time series. Physical Review Letters 45(9), 712–716 (1980)

[43] Palit, A.K., Popovic, D.: Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications. Advances in Industrial Control. Springer- Verlag New York, Inc., Secaucus (2005)

[44] Poskitt, D.S., Tremayne, A.R.: The selection and use of linear and bilinear time series models. International Journal of Forecasting 2(1), 101–114 (1986)

[45] Price, S.: Mining the past to determine the future: Comments. International Jour- nal of Forecasting 25(3), 452–455 (2009)

[46] Priestley, M.B.: Non-linear and Non-stationary time series analysis. Academic Press (1988)

[47] Saad, E., Prokhorov, D., Wunsch, D.: Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. IEEE Transactions on Neural Networks 9(6), 1456–1470 (1998)

[48] Schuster, H.G.: Deterministic Chaos: An Introduction. Weinheim Physik (1988)

[49] Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., Lendasse, A.: Methodology for long-term prediction of time series. Neurocomputing 70(16-18), 2861–2869 (2007)

[50] Sorjamaa, A., Lendasse, A.: Time series prediction using dirrec strategy. In: Verley-sen, M. (ed.) European Symposium on Artificial Neural Networks, ESANN 2006, Bruges, Belgium, April 26-28, pp. 143–148 (2006)

[51] Sorjamaa, A., Lendasse, A., Verleysen, M.: Pruned lazy learning models for time series prediction. In:

European Symposium on Artificial Neural Networks, ESANN 2005, pp. 509–514 (2005) Machine Learning Strategies for Time Series Forecasting 77

[52] Takens, F.: Detecting strange attractors in fluid turbulence. In: Dynamical Systems and Turbulence. Springer, Berlin (1981)

[53] Tiao, G.C., Tsay, R.S.: Some advances in non-linear and adaptive modelling in time-series. Journal of Forecasting 13(2), 109–131 (1994)

[54] Tong, H.: Threshold models in Nonlinear Time Series Analysis. Springer, Berlin (1983)

[55] Tong, H.: Non-linear Time Series: A Dynamical System Approach. Oxford Univer- sity Press (1990)

[56]. Tong, H., Lim, K.S.: Thresold autoregression, limit cycles and cyclical data. JRSS B 42, 245–292 (1980)

[57] Tran, T.V., Yang, B.-S., Tan, A.C.C.: Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems. Expert Syst. Appl. 36(5), 9378–9387 (2009)

[58] Weigend, A.S., Gershenfeld, N.A.: Time Series Prediction: forecasting the future and understanding the past. Addison Wesley, Harlow (1994)

[59] Werbos, P.J.: Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, Cambridge, MA (1974)