Simulating annealing and neural networks for chaotic time series forecasting

George Atsalakis and Katerina Tsakalaki

Technical University of Crete, Greece
atsalak@otenet.gr ; kattsak_89@hotmail.com

Abstract: This paper examines how neural networks that use simulating annealing for training is relative to linear and polynomial approximations to forecast a time series that is generated by the chaotic Mackey-Glass differential delay equation. The forecasting horizon is one step ahead. A series of regressions with polynomial approximators and neural networks that use genetic algorithms and simulating annealing for training are taking place and compare the multiple correlation coefficients. The experimental results confirm that neural networks using simulating annealing algorithm perform well as a global search algorithm. Furthermore, it is shown that using the genetic algorithms to determine their values can improve the forecasting effectiveness of the resulting model when applied to a chaotic time series problem.

Keywords: simulating annealing forecasting, chaos forecasting, Mackey-Glass forecasting, time series forecasting, neural network forecasting.

1 Introduction

Chaotic theory is developing in a new way that influences the world around us and consequently also influences our ways of approaching, analyzing and solving problems. It is not surprising that one of the central models in the chaos literature, the Hénon-Heiles model, is presented in a paper with the title "The applicability of the third integral of motion: Some numerical experiments." Numerical experiments in 1964 were the basis for many significant changes in astronomy in the decades that followed. In 1963 Edwin Lorenz, in his pioneering work on "Deterministic Nonperiodic Flow", proposed a more prominent title for chaotic modelling, by including the term "deterministic". His work spearheaded numerous studies on chaotic phenomena (Skiadas, 2009). On the one hand, according to Wikipedia, simulated annealing (SA) is a generic probabilistic metaheuristic for the global optimization problem of locating a good approximation to the global optimum of a given function in a large search space. It is often used when the search space is discrete (e.g., all tours that visit a given set of cities). For certain problems, simulated annealing may be more efficient than exhaustive enumeration— provided that the goal is merely to find

---

an acceptably good solution in a fixed amount of time, rather than the best possible solution. On the other hand, time series prediction is a very important practical problem with a wide variety of applications ranging from economic and business planning to weather forecasting and signal processing and control. A difficulty that characterizes a chaotic time series is that if the data is not generated by a high dimensional process, it should have short-term predictability and so the use of linear forecasting models is not appropriate. This led to the development of several mathematical tools, such as neural networks and neuro-fuzzy systems which deal with nonlinearity and nowadays they are widely used by many researchers. More specific, artificial neural networks (ANNs) have received more and more attention in financial time series forecasting in recent years. This popularity springs from their capability of performing highly complex mappings on nonlinear data. Nonetheless, they have some significant drawbacks such as the lack of any restrictive assumptions about the functional relationships between the predictor variables and the predicated variable, the difficulty to deal with qualitative information and the 'black box' syndrome. On the other hand, fuzzy inference systems incorporate human knowledge by using the if-then rules and expertise for inference and decision making. However, the disadvantage of fuzzy logic is the lack of self learning capability. This is the reason why the integration of these two approaches is preferred in order to overcome the disadvantages not only of the neural networks but also of the fuzzy systems and results in neuro-fuzzy system models. Moreover, many forecasting algorithms have also been developed in order to approximate initially, general continuous functions, such as polynomial approximation, local linear approximation, radial basis functions and neural networks. However, these algorithms still present some limitations as far as the power of prediction is concerned and this is due to the irregularity of chaotic behaviour related to the complication of geometric structures that chaotic attractors possess and the sensitive dependence on initial conditions in chaotic systems. This study is examining the predictability of a simulated annealing algorithm that is used to training an neural network, as far as a time series generated by the chaotic Mackey-Glass differential delay equation. The results are compared with linear and polynomial approximations. The rest of the paper is organized as follows: Section 2 reviews related research and Section 3 discusses the proposed methodology of simulated annealing. Section 4 presents the models and Section 5 reports the empirical findings, while Section 6 includes the conclusions and some further discussions about the future research in this sector.

2 Related research

Many researchers have worked on the chaotic Mackey-Glass differential delay equation and have forecasted time series using different methods including artificial neural networks, fuzzy logic, stochastic models, simulated annealing and even integration of two or more methods. Related researches are the

following: Skiadas C. Rompogianakis G. Atsalakis G., (2001) with the paper titled "Chaotic Aspects of a Generalized Rational Model and Application in Innovation Management" , Atsalakis, G., Skiadas C. and Braimis I., (2007) with the paper titled "Probability of trend prediction of exchange rate by neuro-fuzzy techniques", L.P. Maguire, B. Roche, T.M. McGinnity and L.J. McDaid (1998) have forecasted a chaotic time series using a fuzzy neural network, Atsalakis G., Skiadas C. and Nezis D., (2008) have forecasted chaotic time series using a neural network, J. Doyne Farmer and John J. Sidorowich (1987) have forecasted chaotic time series using a forecasting technique, George G. Szpiro (1997) has forecasted chaotic time series using genetic algorithms, L. Studer and F. Masulli (1996) have forecasted chaotic time series using a neuro-fuzzy system, Ajoy Kumar Palit and D. Popovic (1999) have forecasted chaotic time series using neuro-fuzzy approach, Liang Zhao and Yupu Yang (2009) have used PSO-based single multiplicative neuron model for time series prediction, Hirotaka Inoue, Yoshinobu Fukunaga and Hiroyuki Narihisa (2001) have used efficient hybrid neural network for chaotic time series prediction, Rahib H. Abiyev (2006) has forecasted time series using a fuzzy wavelet neural network model, Junhong Nie (1994) has forecasted time series using a fuzzy-neural approach, M.R. Hassan, B. Nath and M. Kirley (2006) have forecasted time series using HMM based fuzzy model, Mohammad Assaad, Romuald Bone and Hubert Cardot (2006) have forecasted chaotic time series using boosted recurrent neural networks, Ding Gang, Zhong Shi-Sheng and Li Yang (2008) have forecasted time series using a wavelet process neural network, Catherine Vairappan, Hiroki Tamura, Shangce Gao and Zheng Tang (2009) have forecasted time series using batch type local search-based adaptive neuro-fuzzy inference system (ANFIS) with self-feedbacks, Xieping Gao and Fen Xiao (2004) have forecasted chaotic time series using multiwavelet networks, Cui Wan, Zhao Zhu, Chang Chun, Bao Wen, Xing Liu and Jun Hua (2005) have forecasted chaotic time series using support vector machines for fuzzy rule-based modeling, W.K. Wong, Min Xia and W.C. Chu (2010) have forecasted time series using an adaptive neural network model, Hong-Wei Wang, Hong Gu and Zhe-Long Wang (2005) have forecasted chaotic time series based on SVD matrix decomposition, Muhammad Ardalani-Farsa and Saeed Zolfaghari (2010) have forecasted chaotic time series with residual analysis method using hybrid Elman-NARX neural networks, Ping Liu and Jian Yao (2009) have forecasted chaotic time series using least square support vector machine based on particle swarm optimization, H.J. Song, C.Y. Miao, Z.Q. Shen, W. Roel, D.H. Maja and C. Francky (2010) have forecasted chaotic time series using neural networks, Hongwei Wang and Hong Gu (2009) have forecasted chaotic time series based on neural network with Legendre polynomials, Yuehui Chen, Feng Chen and Qiang Wu (2007) have forecasted time series using an artificial neural networks based dynamic decision model, F. Pan, H. Zhang and M. Xia (2009) have forecasted time series using a hybrid forecasting model and Meiying Ye (2007) has forecasted chaotic time series using LS-SVM with simulated annealing algorithms.

3 Simulated Annealing

This paper considers the development of neural network that uses a simulated annealing algorithm in order to forecast a time series generated by the chaotic Mackey-Glass differential delay equation. Simulated annealing is a stochastic search method, which does not rely on the use of first- and second-order derivatives, but starts with an initial guess Ω_0 and proceeds with random updating of the initial coefficients until a "cooling temperature" or stopping criterion is reached. This method starts with a candidate solution vector, Ω_0, and the associated error criterion, Ψ_0. A shock to the solution vector is then randomly generated, Ω_1, and the associated error metric, Ψ_1 is calculated. If the error metric decreases the new solution vector is always accepted. However, since the initial guess Ω_0 may not be very good, there is a small chance that the new vector, even if it does not reduce the error metric, may be moving in the right direction to a more global solution. So with a probability $P_{(j)}$ conditioned by the Metropolis ratio $M_{(j)}$ the new vector may be accepted, even though the error metric actually increases. The rationale for accepting a new vector Ω_i even if the error Ψi is greater than Ψ_{i-1}, is to avoid the pitfall of being trapped in a local minimum point. According to Robinson (1995), simulated annealing consists of running the accept/reject algorithm between the temperature extremes. As the temperature $T_{(j)}$ cools, changes are more and more likely to be accepted only if the error metric decreases and with gradually decreasing temperature, the algorithm becomes "greedy". Simulated annealing is a random search that moves to a better minimum point, escaping from a likely local minimum rather than a global search and this is the reason why the best one has to do after the convergence to a given point is to see if there are better minimum points in the neighbourhood of the initial minimum. Moreover, the current state of the system or coefficient vector $\hat{\Omega}_j$, depends only on the previous state $\hat{\Omega}_{j-1}$, and a transition problem $P_{(j-1)}$ and is thus independent of all previous outcomes. This system has the Markov chain property and as Haykin points out, an important property of this system is asymptotic convergence, even though resort to finite-time approximation of the asymptotic convergence properties does not guarantee the finding of the global optimum with probability one (McNelis, 2005).

4 Models Presentation

This paper proposes a chaotic time series model, which predicts a time series one step ahead and is generated by the following Mackey-Glass time-delay differential equation.

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t)$$

The time series value was obtained by applying the conventional fourth-order Runge-Kutta algorithm. This model shows how efficient simulated annealing is relative to linear and polynomial approximations. Figure 1 depicts the Mackey-Glass chaotic time series.
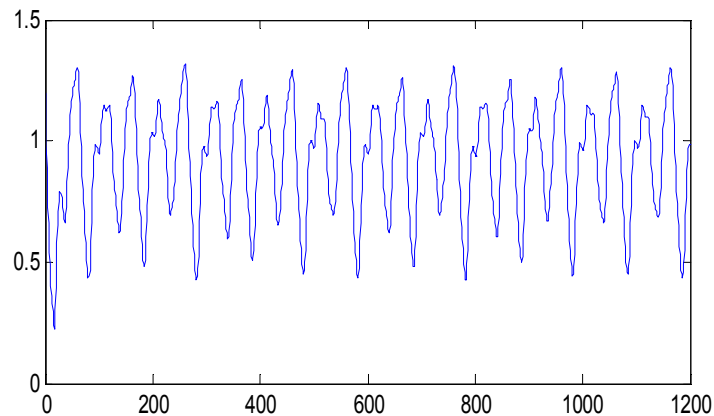

Figure 1. Mackey-Glass chaotic time series

Due to the fact that the time series is chaotic, there is no clearly defined period. Additionally, in time series prediction known values of the time series up to the point in time are used to predict the value at some point in the future. A series of regressions with polynomial approximators and neural networks combined by simulating annealing model is taking place and the multiple correlation coefficients are compared. In this paper, the neural network that is selected uses simulated annealing for training.

 Apart from a neural network that uses simulated annealing for training, which was analyzed in section 2, this paper includes the use of linear regression model, power polynomial-order 2 approximation, orthogonal-order 2 approximation (Tchebeycheff, Hermite, Legendre and Lagueree polynomials) and a simple neural network with two neurons and one layer, which uses genetic algorithms for training. Specifically, the linear regression model seeks for a set of

parameters for the regression model to minimize the sum of squared differences between the actual observations y and the observations predicted by the linear model, $\hat{y}$. In contrast to the linear regression model, a polynomial expansion around a set of inputs x with a progressively larger power P is capable of approximating to a given degree of precision any unknown but continuous function $y = g(x)$.[5] and the parameters here are neither limited in number, nor do they have a straightforward interpretation, as the parameters do in linear models. The orthogonal polynomials, such as the Tchebeycheff, Hermite, Legendre and Lagueree polynomials, unlike the typical polynomial based on raising the variable x to powers of higher order, they are based on sine, cosine or alternative exponential transformations of the variable x and they have proven to be more efficient approximators than the power polynomial. Finally, the genetic algorithm is an evolutionary search process, which reduces the likelihood of landing in a local minimum by starting with a population N^* (an even number) of random vectors. The next step is to select two pairs of coefficients from the population at random, with replacement and evaluate the fitness of these four coefficient vectors, in two pair-wise combinations, according to the sum of squared error function. Coefficient vectors that come closer to minimizing the sum of squared errors receive better fitness values and are retained for "breeding" purposes. Then, crossover takes place in which the two parents "breed" two children and following this operation, each pair of parent vectors is associated with two children coefficient vectors. If crossover has been applied to the pair of parents, the children vectors will generally differ from the parent vectors. The fifth step is mutation of the children where with some small probability, which decreases over time, each coefficient of the two children's vectors is subjected to a mutation. The last step is the election tournament, in which the four members of the "family" engage in a fitness tournament with the children being evaluated by the same fitness criterion used to evaluate the parents. The two vectors with the best fitness, whether parents or children, survive and pass to the next generation, while the two with the worst fitness value are extinguished. The above process is repeated, with parents i and j returning to the population pool for possible selection again, until the next generation is populated by N^* vectors and the convergence is evaluated by the fitness value of the best member of each generation. Once the next generation is populated, elitism can be introduced where all the members of the new generation and the past generation are evaluated according to the fitness criterion. If the best member of the older generation dominates the best member of the new generation, then this member displaces the worst member of the new generation and is thus eligible for selection in the coming generation (McNelis, 2005).

5 Results

Table 1 summarizes the results for the goodness of fit or R^2 statistics for this base set of realizations. Linear model, second-order polynomials are compared with simple neural networks with two neurons and one layer trained by genetic algorithms and simulating annealing.

Table 1: Goodness of fit or R^2

Approximation	R^2
Linear	0.65
Polynomial-Order 2	0.89
Tchebeycheff Polynomial-Order 2	0.89
Hermite-Order 2	0.89
Legendre-Order 2	0.89
Lagueree-Order 2	0.89
Neural Network: FF, 2 neurons, 1 layer-genetic algorithm	0.98
Neural Network: FF, 2 neurons, 1 layer-simulated annealing	0.99

This table shows several important results as far as the prediction of a chaotic time series is concerned. First, there are definite improvements in abandoning pure linear approximation. Second, the power polynomial and the orthogonal polynomials give the same prediction results and so there is no basis for preferring one over the other. Third, the neural network, a very simple neural network genetically evolved, is superior to the polynomial expansions and delivers a very good result. However, this section clearly demonstrates the effectiveness of the proposed neural network, a very simple neural network using simulated annealing for training for the prediction of the Mackey-Glass time series. This neural network prevails among all polynomial expansions and the genetically evolved neural network and delivers an excellent result, indicating that this neural network is much more precise relative to the other methods across a wide set of realizations.

6 Conclusion

This paper presents a neural network with two neurons and one layer, which uses simulated annealing to forecast the chaotic Mackey-Glass time series. The model is developed using Matlab software and it is compared with polynomial expansions and a genetically evolved neural network with two neurons and one layer. The results of the prediction are very satisfactory, indicating that this model can predict well as far as chaotic time series modeling is concerned. This research shows that neural networks in general are designed to mimic very well the ability of the human brain to process data and information and comprehend

patterns and have the ability to analyze and solve business problems and implement those solutions, resulting in being a really helpful tool for forecast purposes. Moreover, according to Paul Coddington from Northeast Parallel Architectures Center at Syracuse University, simulated annealing and its use to predict the chaotic Mackey-Glass time series have the following advantages, which make it an attractive option for optimization problems where heuristic methods are not available:

a) It is relatively easy to code, even for complex problems.

b) It can deal with arbitrary systems and cost functions.

c) It statistically guarantees finding an optimal solution.

d) It generally gives a 'good' solution.

 Yet, further research is recommended by using various time series data in order to reduce the long training times and improve the forecast results.

References

1. Atsalakis G., Skiadas C. and Nezis D., (2008), Forecasting Chaotic time series by a Neural Network, XIIIth International conference on Applied Stochastic Models and Data Analysis, Greece.
2. Skiadas C., Rompogianakis G. Atsalakis G., (2001), Chaotic Aspects of a Generalized Rational Model and Application in Innovation Management, Accounting, Management, Marketing, The Works of Scientific Session of Petru Maior University, Publishing House of Petru Maior University, Tîrgu-Mureş, volume II, pg. 398-405.
3. Atsalakis G., Skiadas C. and Braimis I., (2007), Probability of trend prediction of exchange rate by neuro-fuzzy techniques, XIIth International conference on Applied Stochastic Models and Data Analysis, Greece.
4. L.P. Maguire, B. Roche, T.M. McGinnity, L.J. McDaid, (1998), Predicting a chaotic time series using a fuzzy neural network, Information Sciences 112, 125-136.
5. Christos H. Skiadas, Charilaos Skiadas, (2009), Chaotic Modelling and Simulation Analysis of Chaotic Models, Attractors and Forms, CRC Press Taylor & Francis Group, New York.
6. Paul D. McNelis, (2005), Neural Networks in Finance: Gaining Predictive Edge in the Market, Elsevier Academic Press, New York.
7. Robinson, Guy (1995), "Simulated Annealing" Webpage: www.npac.syr.edu/ copywrite/pcw/node252.
8. J. Doyne Farmer and John J. Sidorowich (1987), Predicting chaotic time series, Physical Review Letters, Vol. 59, No. 8, 845-848
9. George G. Szpiro (1997), Forecasting chaotic time series with genetic algorithms, Physical Review Letters, Vol. 55, No. 3, 2557-2568
10. L. Studer and F. Masulli (1996), On the structure of a neuro-fuzzy system to forecast chaotic time series, Neuro-Fuzzy Systems, 1st International Symposium on Neuro-Fuzzy Systems, AT'96, 103-110

11. Ajoy Kumar Palit and D. Popovic (1999), Forecasting chaotic time series using neuro-fuzzy approach, International Joint Conference on Neural Networks, IJCNN'99, 1538-1543, Vol. 3

12. Liang Zhao and Yupu Yang (2009), PSO-based single multiplicative neuron model for time series prediction, Expert Systems with Applications, Vol. 36, Issue 2, Part 2, Pages 2805-2812

13. Yoshinobu Fukunaga and Hiroyuki Narihisa (2001), Efficient hybrid neural network for chaotic time series prediction, Lecture Notes in Computer Science, Vol. 2130, 712-718

14. Rahib H. Abiyev (2006), Time series prediction using fuzzy wavelet neural network model, Lecture Notes in Computer Science, Vol. 4132, 191-200

15. Junhong Nie (1994), A fuzzy-neural approach to time series prediction, Neural networks, IEEE World Congress on Computational Intelligence, 3164-3169, Vol.5

16. M.R. Hassan, B. Nath and M. Kirley (2006), HMM based fuzzy model for time series prediction, IEEE International Conference on Fuzzy Systems, 2120-2126

17. Mohammad Assaad, Romuald Bone and Hubert Cardot (2006), Predicting chaotic time series by boosted recurrent neural networks, Lecture Notes in Computer Science, Vol. 4233, 831-840

18. Gang, Zhong Shi-Sheng and Li Yang (2008), Time series prediction using wavelet process neural network, Chinese Physics B, Vol. 17, No. 6, 1674-1056

19. Catherine Vairappan, Hiroki Tamura, Shangce Gao and Zheng Tang (2009), Batch type local search-based adaptive neuro-fuzzy inference system (ANFIS) with self-feedbacks for time-series prediction, Neurocomputing, Vol. 72, Issues 7-9, Pages 1870-1877

20. Xieping Gao and Fen Xiao (2004), Multiwavelet networks for prediction of chaotic time series, IEEE International Conference on Systems, Man and Cybernetics, 3328-3332, Vol. 4

21. Cui Wan, Zhao Zhu, Chang Chun, Bao Wen, Xing Liu and Jun Hua (2005), Prediction of the chaotic time series using support vector machines for fuzzy rule-based modeling, http://en.cnki.com.cn/Article_en/CJFDTOTAL-WLXB200507008.htm

22. W.K. Wong, Min Xia and W.C. Chu (2010), Adaptive neural network model for time-series forecasting, European Journal of Operational Research, Vol. 207, Issue 2, Pages 807-816

23. Hong-Wei Wang, Hong Gu and Zhe-Long Wang (2005), Fuzzy prediction of chaotic time series based on SVD matrix decomposition, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, 2493-2498, Vol. 4

24. Muhammad Ardalani-Farsa and Saeed Zolfaghari (2010), Chaotic time series prediction with residual analysis method using hybrid Elman-NARX neural networks, Neurocomputing, Vol. 73, Issues 13-15, Pages 2540-2553

25. Ping Liu and Jian Yao (2009), Application of least square support vector machine based on particle swarm optimization to chaotic time series prediction, IEEE International Conference on Intelligent Computing and Intelligent Systems, 458-462

26. H.J. Song, C.Y. Miao, Z.Q. Shen, W. Roel, D.H. Maja and C. Francky (2010), Design of fuzzy cognitive maps using neural networks for predicting chaotic time series, Neural Networks, Vol. 23, Issue 10, Pages 1264-1275

27. Hongwei Wang and Hong Gu (2009), Prediction of chaotic time series based on neural network with Legendre polynomials, Lecture Notes in Computer Science, Vol. 5551, 836-843

28.Yuehui Chen, Feng Chen and Qiang Wu (2007), An artificial neural networks based dynamic decision model for time-series forecasting, International Joint Conference on Neural Networks, IJCNN 2007, 696-699

29.F. Pan, H. Zhang and M. Xia (2009), A hybrid time-series forecasting model using extreme learning machines, Second International Conference on Intelligent Computation Technology and Automation, ICICTA'09, 933-936

30.Meiying Ye (2007), Prediction of chaotic time series using LS-SVM with simulated annealing algorithms, Lecture Notes in Computer Science, Vol. 4492, 127-134