



UNIVERSITY *of* OTAGO
TE WHARE WĀNANGA O OTĀGO

DUNEDIN NEW ZEALAND

A shape metric for evolving time series models

Peter Whigham
Colin Aldridge

**The Information Science
Discussion Paper Series**

Number 2000/10
April 2000
ISSN 1177-455X

University of Otago

Department of Information Science

The Department of Information Science is one of six departments that make up the Division of Commerce at the University of Otago. The department offers courses of study leading to a major in Information Science within the BCom, BA and BSc degrees. In addition to undergraduate teaching, the department is also strongly involved in postgraduate research programmes leading to MCom, MA, MSc and PhD degrees. Research projects in spatial information processing, connectionist-based information systems, software engineering and software development, information engineering and database, software metrics, distributed information systems, multimedia information systems and information systems security are particularly well supported.

The views expressed in this paper are not necessarily those of the department as a whole. The accuracy of the information presented in this paper is the sole responsibility of the authors.

Copyright

Copyright remains with the authors. Permission to copy for research or teaching purposes is granted on the condition that the authors and the Series are given due acknowledgment. Reproduction in any form for purposes other than research or teaching is forbidden unless prior written permission has been obtained from the authors.

Correspondence

This paper represents work to date and may not necessarily form the basis for the authors' final conclusions relating to this topic. It is likely, however, that the paper will appear in some form in a journal or in conference proceedings in the near future. The authors would be pleased to receive correspondence in connection with any of the issues raised in this paper, or for subsequent publication details. Please write directly to the authors at the address provided below. (Details of final journal/conference publication venues for these papers are also provided on the Department's publications web pages: <http://www.otago.ac.nz/informationsscience/pubs/publications.html>). Any other correspondence concerning the Series should be sent to the DPS Coordinator.

Department of Information Science
University of Otago
P O Box 56
Dunedin
NEW ZEALAND

Fax: +64 3 479 8311
email: dps@infoscience.otago.ac.nz
www: <http://www.otago.ac.nz/informationsscience/>

A Shape Metric for Evolving Time Series Models

P.A. Whigham¹ and C.H. Aldridge²

¹ Department of Information Science, University of Otago, New Zealand,
pwhigham@infoscience.otago.ac.nz,
<http://divcom.otago.ac.nz/sirc/Peterw/>

² Department of Information Science, University of Otago, New Zealand,
caldridge@infoscience.otago.ac.nz,
<http://divcom.otago.ac.nz/infosci/Staff/ColinA.htm>

Abstract. Most applications of Genetic Programming to time series modeling use a fitness measure for comparing potential solutions that treat each point in the time series independently. This non-temporal approach can lead to some potential solutions being given a relatively high fitness measure even though they do not correspond to the training data when the overall shape of the series is taken into account. This paper develops two fitness measures which emphasize the concept of shape when measuring the similarity between a training and evolved time series. One approach extends the root mean square error to higher dimensional derivatives of the series. The second approach uses a simplified derivative concept that describes shape in terms of positive, negative and zero slope.

1 Introduction

Multivariate time series data exists in many domains, including business, environmental, medical and engineering. Machine learning techniques have often been applied to these forms of data, either to explore the underlying properties of the described system or to produce models for forecasting. Population-based search methods, such as genetic algorithms (GAs), genetic programming (GP) and evolving neural networks have been successfully applied to a variety of time series tasks due to their ability to handle complex, non-linear problems [Zhang et al., 1997, Lee, 1999, Whigham and Crapper, 1999].

Evolutionary learning systems are fundamentally driven by the use of a fitness measure that biases the selection of individuals for reproduction. The metaphor of a fitness landscape [Eldredge, 1989, Jones, 1995], originally defined by evolutionary biologists, has been used to describe how the fitness function changes the shape and difficulty for searching through this landscape [Manderick et al., 1991]. Although it is clear that the appropriate selection of a fitness function is essential for the successful application of evolutionary learning techniques, little research has been done on determining the appropriate measures to be used for time series applications.

This paper is organised as follows. Section 2 presents the motivation for considering a revised form of fitness for time series analysis. Section 3 commences

by reviewing several standard measures of error used when evolving time series predictions. Two fitness measures that account for shape are then developed. Section 4 briefly gives some discussion and future work based on this preliminary study. Section 5 summarises the basic conclusions regarding these measures.

2 Motivation

In relation to time series, the fitness measure is intended to reflect how similar each member of the population is to the training data. The fitness measure allows a probabilistic selection of the population members for reproduction such that the fitter individuals are more likely to propagate into future generations. The field of data mining also uses the concept of similarity in relation to the mining of large temporal databases (see [Tansel et al., 1993] for examples). One goal of data mining is to extract objects from a database which have similar or dissimilar properties from some defined object. In relation to time series this translates into finding sequences that exhibit similar behaviour for a large subset of their length. Additional problems are often encountered in temporal data mining due to scaling, translation and outliers that make the problem of measuring similarity more difficult. This problem has been addressed by Keogh and Pazzani [Keogh and Pazzani, 1999] where they argue that similarity is a subjective measure not easily captured by a simple metric based on treating each point in a time series independently.

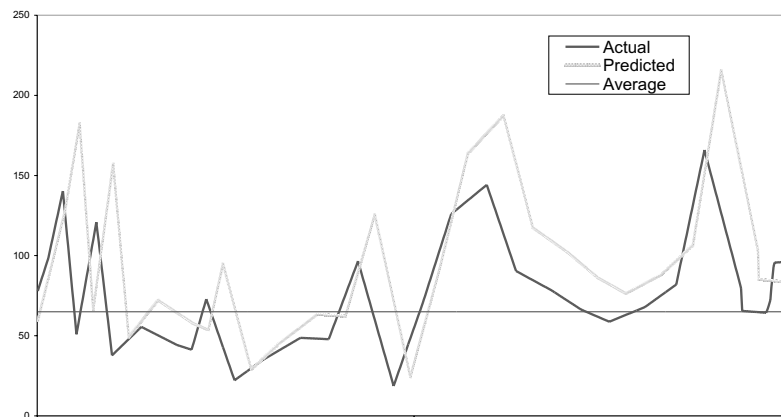


Fig. 1. The Straight Line (Average) has a lower RMSE than the Predicted curve.

This paper will argue that point-based metrics are not always the appropriate fitness measure when using evolutionary methods for developing models of time series behaviour because they ignore global information in terms of the form or structure of the complete time series. This lack of global information may bias the population evolution in such a way that intermediate structures that do capture the underlying patterns in the data have low fitness and therefore do not propagate. For example, Fig. 1 shows a time series (labeled Actual) and two other time series; one which appears quite similar in shape (Predicted) and the other (Average) which is a straight line representing the average of the Actual series. The Average series has a lower root mean square error (RMSE) than the Predicted sequence, whereas it seems clear from observation that this is not the correct ordering of similarity between the sequences. Further examples showing the unintuitive measure of similarity based on RMSE are given in [Keogh and Pazzani, 1999].

Consider the consequences of this mismatch when a population is evolving; a predicted sequence that has the correct shape but some outliers or scaling mismatch will not be selected even though it clearly has many of the features of the sequence being modeled. Given that the driving force behind evolutionary learning systems is selection pressure based on the fitness measure, an inappropriate metric will bias the search away from useful partial solutions.

3 What makes a time series temporal?

By its very nature a time series has an explicit ordering. This ordering extends the possible relationships between elements that do not exist with other (non-temporal) tables of data. It is this ordering which seems to be overlooked when measures of fitness(similarity) are considered with evolutionary systems. This can lead to population members being given inappropriate comparative fitness measures which in turn influence the future population dynamics and, ultimately, the resulting solution.

3.1 Non-Temporal Fitness Measures

This section describes several time series fitness measures often used in assessing forecasting techniques [Makridakis et al., 1983] and with GP approaches to time series prediction. The measures defined all have the property that they equal zero when there is a perfect fit between the predicted and actual series. The main disadvantage with each measure is that no account is taken of the inherent temporal nature of the series. Essentially the temporal ordering is discarded and the series is treated as a table of unrelated data items. There are many standard measures of time series error, however they are all basic variations of the following. Let P_t be the predicted value at time t . Let A_t be the actual value at time t . Let n be the number of data points in the time series. Let \bar{A} be the mean of the actual values.

$$RMSE(A, P) = \sqrt{\frac{\sum_{t=1}^n (P_t - A_t)^2}{n}} \quad (1)$$

$$MAPE(A, P) = \frac{\sum_{t=1}^n |(P_t - A_t) \div A_t|}{n} \quad (2)$$

$$COE(A, P) = \frac{\sum_{t=1}^n (P_t - A_t)^2}{\sum_{t=1}^n (A_t - \bar{A})^2} \quad (3)$$

The root mean square error (RMSE) is commonly used as a fitness measure when evolving time series models. RMSE tends to be used because large errors between points in the predicted and actual series are accentuated (due to squaring). This tends to produce partial solutions that model the larger values in the time series.

The mean absolute percentage error (MAPE) is a relative measure and is often used for comparison between different time series. MAPE gives more emphasis towards smaller values in the sequence since the absolute error between the predicted and actual values in the time series is divided by the actual value.

The coefficient of equation (COE) is similar to the RMSE however it is scaled based on the deviation of each value about the mean. COE also emphasises large values in the time sequence due to the squared error term between the actual and predicted points in the series.

3.2 Temporal Fitness Measures

The previous measures treat the time series as a table of unrelated data items. They do not consider the importance of neighboring values in the sequence. To explicitly take advantage of the temporal nature of time series the error metric must include information that is represented by the 'shape' of the time sequence. Differencing, a type of filter often used for removing trends in a time series, is an appropriate starting point for incorporating temporal relationships into an error (fitness) metric. For a time series $\{x_1, \dots, x_n\}$ a new series $\{y_1, \dots, y_{n-1}\}$ is formed using the difference operator [Chatfield, 1989]:

$$y_t = x_{t+1} - x_t = \nabla x_{t+1} \quad (4)$$

Equation 4 represents the absolute change in consecutive time steps; in other words, it is the first derivative of the time series curve for consecutive time steps. The second difference is defined as:

$$\nabla^2 x_{t+2} = \nabla x_{t+2} - \nabla x_{t+1} = x_{t+2} - 2x_{t+1} + x_t \quad (5)$$

The concept of differencing can now be extended to develop a similarity measure $F(A, P)$ between two time series A and P . So that the measure is appropriate for use with a GP system two properties are required; firstly, $F(A, P) \geq 0$ and, secondly, $F(A, P) = 0$ when the sequences A and P are identical. The

fitness measure, $F(A, P)$, will be based on the non-temporal RMSE (Eqn. 1) and extended using k -th order differencing. Measures based on MAPE or COE could also be developed in a similar manner.

Let the predicted and actual sequences be represented as $\{P_1, P_2, \dots, P_n\}$ and $\{A_1, A_2, \dots, A_n\}$. Let P_t and A_t be the predicted and actual values at time t . Define the zeroth order difference between sequences P and A at time t as:

$$\nabla^0(A, P)_t = A_t - P_t \quad (6)$$

The first order difference between P and A between the points t and $t + 1$ can now be defined as:

$$\nabla^1(A, P)_t = \nabla^0(A, P)_{t+1} - \nabla^0(A, P)_t \quad (7)$$

Eqn. 7 represents the difference between the slopes of the actual and predicted curves at t and $t + 1$, shown by expanding Eqn. 7 as follows:

$$\nabla^1(A, P)_t = (A_{t+1} - P_{t+1}) - (A_t - P_t) = (A_{t+1} - A_t) - (P_{t+1} - P_t) \quad (8)$$

In general, the k th order difference between A and P at time t may be stated as:

$$\nabla^k(A, P)_t = \nabla^{k-1}(A, P)_{t+k} - \nabla^{k-1}(A, P)_{t+k-1} \quad 1 \leq k \leq n - 1 \quad (9)$$

The previous difference equations can now be extended to incorporate the RMSE format. The RMSE for $k = 0$ is defined as:

$$RMSE^0(A, P) = \sqrt{\frac{\sum_{t=1}^n \nabla^0(A, P)_t^2}{n}} \quad (10)$$

The RMSE for the k th difference is defined as:

$$RMSE^k(A, P) = \sqrt{\frac{\sum_{t=1}^{n-k} \nabla^k(A, P)_t^2}{n - k}} \quad (11)$$

A fitness function $F^w(A, P)$, which measures the similarity between the time series $\{A_1, A_2, \dots, A_n\}$ and $\{P_1, P_2, \dots, P_n\}$ using a differencing window of length w , can now be defined as:

$$F^w(A, P) = \sqrt{\sum_{k=0}^w \sum_{t=1}^{n-k} \frac{\nabla^k(A, P)_t^2}{n - k}} \quad 0 \leq w \leq n - 1 \quad (12)$$

Note that $F^0(A, P)$ is equal to the original RMSE definition of Eqn. 1. Allowing w to increase produces a measure that incorporates more information about the shape of each sequence and therefore gives a more unique and informative measure of similarity between A and P .

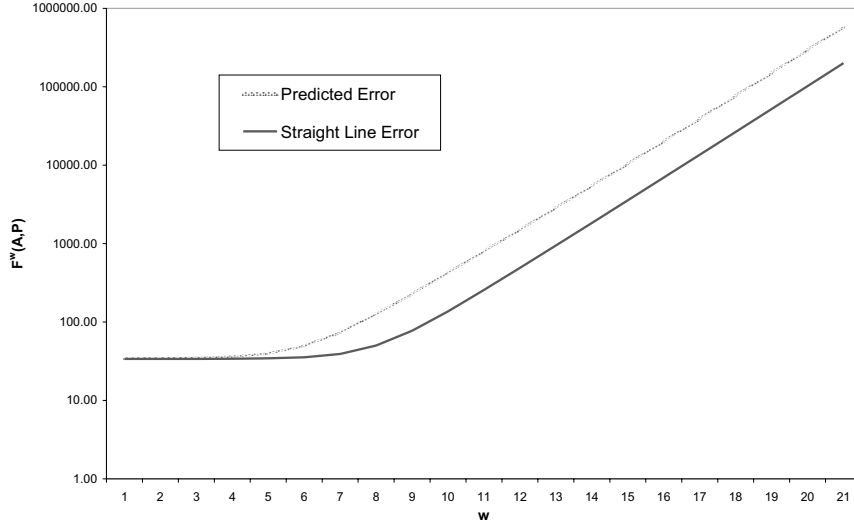


Fig. 2. $F^w(A, P)$ for predicted and straight line (average) series

3.3 Problems associated with $F^w(A, P)$

The original motivation for this work came about by recognising the inappropriate ordering of "fit" time series when using RMSE. This is clearly shown in Fig. 1. To explore whether $F^w(A, P)$ can offer a more appropriate similarity measure the fitness measure was calculated for the sequences shown in Fig. 1 for $0 \leq w < 20$. The resulting error for increasing w is shown in Fig. 2. There are several concerns with these results. One is the computational cost when calculating $F^w(A, P)$. The second point is that $F^w(A, P)$ rapidly rises for increasing w . Finally, Fig. 2 shows that the straight line continues to have a lower overall error than the predicted curve. Hence, although the notion of shape has been explicitly captured by $F^w(A, P)$, the squaring of the difference for each $\nabla^k(A, P)$ is rapidly dominated by one or more errors caused by a poor value for the predicted series.

3.4 An alternative $F^w(A, P)$ based on slope direction

The goal of this work was to develop a fitness measure that allowed global information about shape to be incorporated. Simplifying the error measurement can be achieved by classifying each value of the k th difference into one of three values; positive slope, negative slope and zero slope. The error for each difference is then the number of times that the sign of the difference does not match. This value is normalised by the number of points in the k th sequence to give a value between zero and one. A perfect match with shape for the k th difference will equal zero, while a complete mismatch with shape will equal one. Define

this error measurement for the k th difference as $\Delta^k(A, P)$. Fig. 3 shows the normalised shape error for increasing k for the predicted and average curves related to the actual curve of Fig. 1. Note that the average prediction (straight line) is now penalised due to its poor correlation with the slope of the actual curve. The predicted curve, that by observation appears to capture many of the shape aspects of the actual curve, has a lower error for most of the k differences.

The main objection to this approach is that it appears rather ad hoc compared with the original definition of error. For example, how should this error be combined with the original RMSE? Since the fitness measure should only go to zero when the two time series are identical one possible measure would be:

$$S^w(A, P) = F^0(A, P) + F^0(A, P) * \frac{1}{w} \sum_{k=1}^w \Delta^k(A, P) \quad 0 < w \leq n - 1 \quad (13)$$

Using Eqn. 13 the error $S^{n-1}(A, P)$ for the predicted curve is 43.9 versus the average (straight line) error of 67.5.

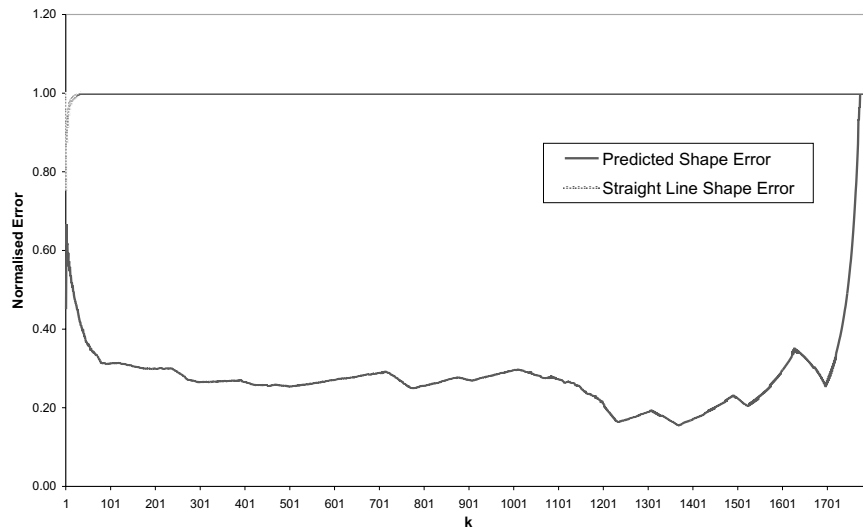


Fig. 3. Normalised error in slope for predicted and straight line (average) series

4 Discussion

This paper has presented two alternative approaches to incorporating shape knowledge into a fitness measure when evolving time series models. $F^w(A, P)$ has been shown to be too complex and produces fitness values that are too large

when w approaches the total length of the series. The simplified shape estimate, $S^w(A, P)$, appears to have some advantages in that it is less computationally demanding and produces values that are more amenable to use as a fitness measure.

This preliminary work indicates several lines of possible research. Based on Fig. 3 it is worth determining what properties of the time series are causing the predicted shape error to level out to approximately 0.30 when $k > 90$. This may have some relationship to seasonal patterns in the data, or to some high order correlations in the data. At present, work is being undertaken to use $S^w(A, P)$ to evolve time series models for varying values of w . The results of this work are expected to confirm that using a shape metric as part of the fitness measure will allow more accurate time series models to be developed.

5 Conclusion

This paper presents two fitness measures for comparing time series predictions. Problems associated with the calculation of $F^w(A, P)$ for large w led to a second metric that simplified the concept of shape to a positive, negative or zero slope measurement. This measure $S^w(A, P)$ has been shown to produce relative errors between sequences that align more with our intuitive notion of similarity. The obvious difficulty with this work is proving that $F^w(A, P)$ or $S^w(A, P)$ for $w > 0$ will in general help evolve better solutions than the original non-temporal error measurements. Each new time series will have different properties and be more suited to one of the temporal or non-temporal measures that have been presented. The most persuasive evidence of an improvement by using shape for fitness would be to evolve a comprehensively better model for a well-known problem than previously discovered. It is recommended that the fitness measures introduced here be applied by researchers in time series modelling and GP to ascertain whether this is the case.

References

- [Chatfield, 1989] Chatfield, C. (1989). *The Analysis of Time Series: An Introduction*. Chapman and Hall, 4 edition.
- [Eldredge, 1989] Eldredge, N. (1989). *Macroevolutionary Dynamics: Species, Niches and Adaptive Peaks*. McGraw-Hill.
- [Jones, 1995] Jones, T. (1995). *Evolutionary Algorithms, Fitness Landscapes and Search*. PhD thesis, Computer Science Dept., University of New Mexico, Albuquerque, New Mexico.
- [Keogh and Pazzani, 1999] Keogh, E. and Pazzani, M. (1999). Relevance feedback retrieval of time series data. In *The Twenty-Second Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*.
- [Lee, 1999] Lee, G. (1999). Genetic recursive regression for modeling and forecasting real-world chaotic time series. In Spector, L., Langdon, W., and Angeline, U. O. P., editors, *Advances in Genetic Programming 3*, chapter 17, pages 401–423. USA:MIT Press.

- [Makridakis et al., 1983] Makridakis, S., Whellwright, S., and McGee, V. (1983). *Forecasting: Methods and Applications*. John Wiley & Sons, Inc., 2 edition.
- [Manderick et al., 1991] Manderick, B., de Weger, M., and Spiessens, P. (1991). The genetic algorithm and the structure of the fitness landscape. In *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 143–149. Morgan Kaufmann Publishers.
- [Tansel et al., 1993] Tansel, A., Clifford, J., Gadia, S., Jajodia, S., Sesgev, A., and Snodgrass, R. (1993). *Temporal databases: theory, design and implementation*. Redwood City, Calif.:Benjamin/Cummings Pub. Co.
- [Whigham and Crapper, 1999] Whigham, P. and Crapper, P. (1999). Time series modelling using genetic programming: An application to rainfall-runoff models. In Spector, L., Langdon, W., O'Reilly, U., and Angeline, P., editors, *Advances in Genetic Programming 3*, chapter 5, pages 89–104. USA:MIT Press.
- [Zhang et al., 1997] Zhang, B.-T., Ohm, P., and Mühlenbein, H. (1997). Evolutionary induction of sparse neural trees. *Evolutionary Computation*, 5(2):213–236.