PREDICTION OF STRIDE INTERVAL TIME SERIES

Etienne Zahnd[1], *Jennifer S. Brach*[2], *Subashan Perera*[3] *and Ervin Sejdić*[1]

[1]Dep. of Electrical and Computer Engineering, University of Pittsburgh
[2] Dep. of Physical Therapy, University of Pittsburgh
[3]Div. of Geriatric Medicine, University of Pittsburgh

ABSTRACT

The power law in the frequency spectrum $S(f) = 1/f^{\beta}$ allows for a good representation of the various time evolution and complex interactions of many physiological processes. The spectral exponent β can be interpreted as the degree of fractal characteristic which in turn makes it some sort of biomarker that gives an idea of the relative health of an individual. The prediction of the $1/f^{\beta}$ time series can thus prove to be an asset in the medical field where forecasting the future health state of an individual can be important for rehabilitation purposes. The goal of this paper is to consider the accuracy of several time series prediction methods such as the neural networks, regression trees and bagged regression trees learning method. To test these methods we simulate stride intervals time series as $1/f^{\beta}$ processes. Our results show that the regression trees can accurately predict between five and fifteen points.

Keywords: Prediction, stride interval, time series, $1/f^{\beta}$ process, human gait, neural networks, regression trees.

1. INTRODUCTION

There are many reasons that explain for a decline in the locomotor system of the human body: disease, aging, trauma, genetic disorders to name a few [1, 2]. The locomotor system is mainly composed of the central nervous system and other physiological systems. A complex connection amongst the brain, the nervous system, and the muscles used when in balance allows gait [1, 2]. If the individual is healthy, a stable walking pattern is possible if the interaction between all the components that make the locomotor system function correctly.

As mentioned, the gait can be affected due to neurophysiological changes which alters the functionality of the locomotor system presented above. Diseases such as amyotrophic lateral sclerosis, or ALS, Parkinson's disease (PD), and even Huntington's disease (HD) are all neurodegenerative diseases that directly affect the locomotor system in an intense way [3, 4]. This can be observed in the increased stride duration [3, 4]. However, longer strides do not imply that the individual suffers from a neurodegenerative disease since it can be seen as well in a process that happens to everybody which is aging [5]. Determining the potential problems that an individual can present is highly complicated since the physiological systems dealt with are complex and highly non-linear [5, 6].

The main goal of this paper is to be able to predict the stride interval time series of the $S(f) = 1/f^{\beta}$ process using various prediction methods available, mainly the machine learning methods. To attain our goal, we will simulate stride intervals time series and examine the accuracy of the consider algorithms to accurately estimate typical stride interval characteristics such as mean and coefficient of variation.

2. PREDICTION METHODS

There exist multiple prediction methods, the ones used in this study were neural networks, where two different neural networks were used, and two decision trees.

2.1. Neural Network

The basic biological definition of a neural network is very similar to the definition of a neural network in machine learning. Regarding the latter, a neural network is a computer system modeled on the human brain and nervous system of the human body. It allows the studied system to learn from mistakes made and be allowed to improve the results through what is called the epochs [7, 8] . There exist multitudes of neural networks, the ones chosen in this study were the feedforward neural network and the layer recurrent neural network.

Feedforward Neural Network (FFNN) - The simplest of all artificial neural networks, the feedforward neural network has only one layer hidden and contrarily to the recurrent version, it does not cycle. The information only moves forward from the input nodes to the output nodes through the hidden layer [7].

Layer Recurrent Neural Network (LRNN) - As opposed to the neural network presented hereinabove, the layer recurrent neural network does actually form a cycle, allowing for a usage of their internal memory to process arbitrary sequences of inputs [8].

2.2. Trees

Trees are structures used to predict the response to inputs. This method is often used in machine learning [9] and it is based on the construction of a binary tree where the nodes represent the tests made on the inputs. In this study, the regression tree and the bagged decision tree have been chosen as the structures used for prediction. They both are very useful since they provide easy to understand predictions in all types of conditions [9, 10].

3. RESULTS

We simulated stride interval time series using the fact that these time series can be simulated using the the $1/f^{\beta}$ power law [11, 12]. In this study, the number of strides chosen was 300 as it is the average number of strides a healthy individual does in three minutes that is not considered to be active in sports [5]. We considered $0 \leq \beta \leq 2$ to mimic a spectrum of conditions. As for the neural networks, their parameters were set as follows: the number of nodes chosen for the FFNN are 150 as it seems that this outputs less errors than for any other number of nodes and for the LRNN, the hidden sizes are chosen to be equal to 15 and a short layer delay is added.

A MonteCarlo simulation was computed where a forecast of 5, 15, 25 and 35 points was done compared to the spectral exponent. Since the computation of these prediction methods is intensive, a hundred simulations were done per β value. These results are displayed in Fig. 1 for the coefficient of variation (CV) result and Fig. 2 for the mean values. As can be seen, the value of the spectral exponent β does have a huge impact on the mean and CV. The more the value of β increases, the more the mean and the value of the CV increase. When the number of forecasted points does not exceed 15 for both the mean and the CV, the original signal and the models have identical values. However, as the number increases, the CV of the models displays more discrepancy compared to the original signal. The neural networks offer far less accuracy, the LRNN varies greatly when forecasting 35 points and the FFNN does not

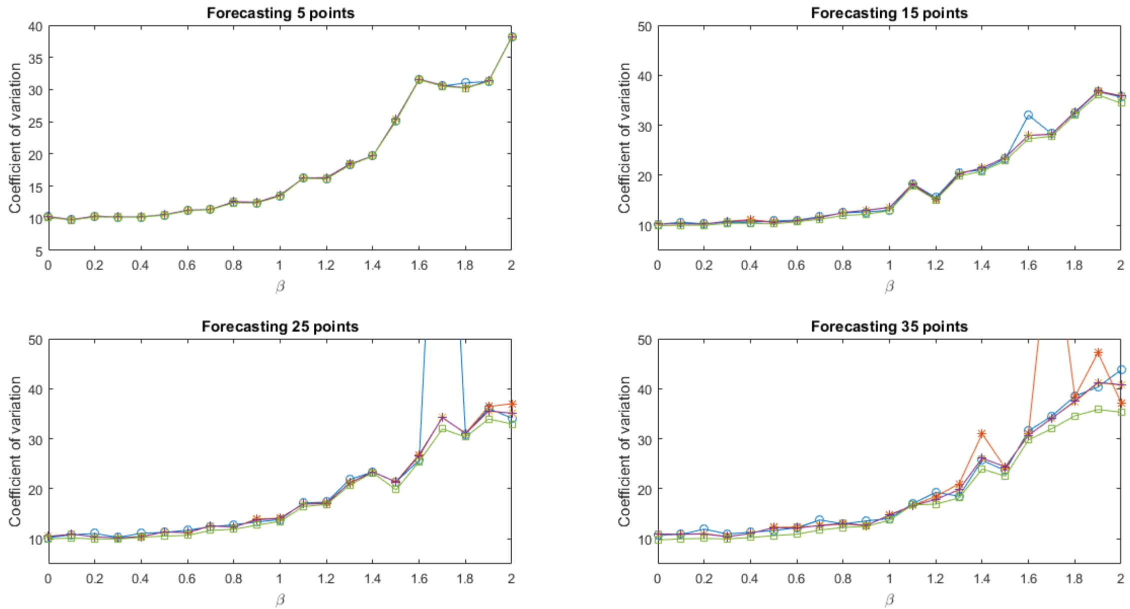

Fig. 1. MonteCarlo simulation result for the coefficient of variation of the forecasted signals depending on the β value. Signals: □ is the original signal, o is FFNN, * is LRNN, x is the regression tree, + is the bagged regression tree

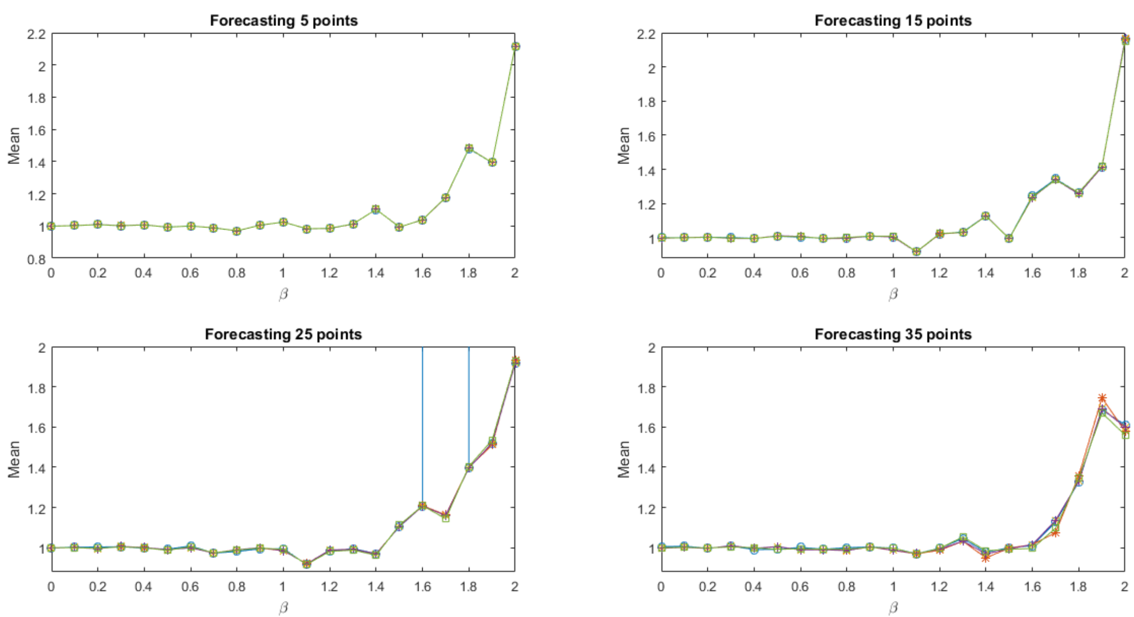

Fig. 2. MonteCarlo simulation result for the mean of the forecasted signals depending on the β value. Signals: □ is the original signal, o is FFNN, * is LRNN, x is the regression tree, + is the bagged regression tree

exhibit perfect forecasts as suggested by its CV calculations in Fig. 1. Both trees are more accurate than the neural networks, however, when exceeding 15 points, when $\beta > 1$, their CV values differ from the original signal's. The results for the mean in Fig. 2 show that a greater number of forecasted points will negatively impact the mean of the models as they will differ from the original signal's mean values at a given β. That difference only increases as the number of points increases as well.

4. DISCUSSION

In this study, we considered four machine learning methods to predict the stochastic process that is the $1/f^{\beta}$ function. The CV and the mean of the forecasted signals were generated in order to better visualize the accuracy of the results.

Since the $1/f^{\beta}$ process is a stochastic process, it was not surprising that predicting it would not give accurate results. When using the neural networks and the decision trees, the predictions do not display the exact variations of the white, pink and red noises. This resulted in choosing the regression trees as the best methods for prediction the stride interval time series. The goal being to obtain the value at one particular point in a future state, the extrapolation results show that more accurate predictions can be achieved when forecasting up to 15 points.

As shown, the accuracy of the forecasting is very variant depending on the value of the spectral exponent β. The closer its value is to 2, the less accuracy we are able to forecast. These may affect the results, especially for healthy individuals for which $\beta = 1$. The CV being a tool that measures the dispersion of a frequency distribution can give an idea of the accuracy of the result. It is shown that in this case both the regression trees are more accurate with a CV that is relatively similar to the original signal.

In this study, the prediction was done using the exact number of strides inputted and then it was possible to determine future outputs based on the predictions of the process at hand. However, this number varies depending on the health of an individual. A patient that would be associated with a β value different from 1 would have far fewer strides in the same time interval. Furthermore, the list of machine learning methods used in this study is limited to only four elements. Finally the parameters set for the neural networks are probably not restrictive enough for time series prediction.

In a future project, it could be interesting to consider these changes in stride interval between individuals. The prediction of the number of strides could also be greatly increased by choosing other methods that would be more suited to the problem at hand.

5. CONCLUSION

In this paper, four methods for predicting signals have been studied. It has been shown that the methods best suited for predicting the $1/f^{\beta}$ power law are both the regular and bagged regression trees. When using both for forecasting, it has displayed far better accuracy than the other methods up to 15 points at most.

6. ACKNOWLEDGMENT

This work was supported by the National Institute on Aging through the Pittsburgh Claude D. Pepper Older Americans Independence Center under Grant NIA P30 AG 024827.

7. REFERENCES

[1] A. Eke, P. Herman, J. Bassingthwaighte, G. Raymond, D. Percival, M. Cannon, I. Balla, and C. Ikrényi, "Physiological time series: distinguishing fractal noises

from motions," *Pflügers Archiv*, vol. 439, no. 4, pp. 403–415, 2000.

[2] A. Eke, P. Herman, L. Kocsis, and L. Kozak, "Fractal characterization of complexity in temporal physiological signals," *Physiological Measurement*, vol. 23, no. 1, pp. R1–R38, 2002.

[3] J. M. Hausdorff, C. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Is walking a random walk? evidence for long-range correlations in stride interval of human gait," *Journal of Applied Physiology*, vol. 78, no. 1, pp. 349–358, 1995.

[4] J. M. Hausdorff, P. L. Purdon, C. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations," *Journal of Applied Physiology*, vol. 80, no. 5, pp. 1448–1457, 1996.

[5] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C.-K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger, "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease," *Journal of Applied Physiology*, vol. 82, no. 1, pp. 262–269, 1997.

[6] J. B. Bassingthwaighte, "Physiological heterogeneity: fractals link determinism and randomness in structures and functions," *Physiology*, vol. 3, no. 1, pp. 5–10, 1988.

[7] G. Bebis and M. Georgiopoulos, "Feed-forward neural networks," *IEEE Potentials*, vol. 13, no. 4, pp. 27–31, 1994.

[8] M. W. Goudreau, C. L. Giles, S. T. Chakradhar, and D. Chen, "First-order versus second-order single-layer recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 5, no. 3, pp. 511–513, 1994.

[9] S. Cohen, L. Rokach, and O. Maimon, "Decision-tree instance-space decomposition with grouped gain-ratio," *Information Sciences*, vol. 177, no. 17, pp. 3592–3612, 2007.

[10] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[11] A. Schaefer, J. S. Brach, S. Perera, and E. Sejdić, "A comparative analysis of spectral exponent estimation techniques for $1/f^{\beta}$ processes with applications to the analysis of stride interval time series," *Journal of Neuroscience Methods*, vol. 222, pp. 118–130, 2014.

[12] E. Sejdić and L. A. Lipsitz, "Necessity of noise in physiology and medicine," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 2, pp. 459–470, 2013.