Supervised Machine Learning: A Review of Training Algorithms for Financial Time-Series Forecasting

Said Jadid Abdulkadir, Sameer Ahmad Khan

INTRODUCTION

Chaotic forecasting is a huge problem in many real-world applications. Performance of chaotic time-series models is built on historic data which is used for model training with an accuracy aspect is solely applied to validate the performance of any time-series forecasting model. For more than a decade, in time-series forecasting, statistical models have been utilized. Nevertheless, their ability to forecast is limited to the nature of time-series applied which is of linear behaviour unlike real world problems which are non-linear. Box-Jenkins [1], Multi Regressions [2]-[3] and Exponential Smoothing [4]-[5] are examples of statistical methods applied in time-series forecasting. Non-statistical models which are non-linear models have also been applied in time-series forecasting. Examples include Fuzzy Logic [6], Genetic Algorithm [7], Support Vector Machines [8] and Artificial Neural Networks [9]-[11].

Forecasting chaotic time-series is classified as long term or short term. In the latter, Neural Network models have no loop that provides a feedback between the network output and the input regressor [12]. Only actual data samples are used in this model unlike the long-term forecasting model that utilizes the actual and target data as inputs to the model. The long-term forecast has a feedback loop that feeds the input for a number of time steps depending on the horizon set by the

forecaster [13]. The input regressor is constituted of replaced actual data from previously forecasted values of the time-series, nonetheless when the time steps tend to infinity the values in the input regressor consist of only estimated time-series values. The infinity process makes long term forecasting a much more complex dynamic modelling process than short term forecasting [14].

Forecaster's aim is to apply various approaches in successfully forecasting the applicable data using legal forecasting policies. "The central idea to successful chaotic time-series forecasting is achieving the best results using minimum required input data and the least complex model" [15]. With a focus on this idea it is evident that because of the complex nature of chaotic time-series forecasting, there is need for the application of dynamic forecasting models which would help in estimating future trends and reducing risks of decision making.

Problem Statement

For time series modelling, very powerful sequence models such as Recurrent Neural Networks are proposed. However, they do not enjoy widespread use because it is extremely difficult to train them properly due to the vanishing gradient problem [16]-[19]. In Recurrent Neural Network (RNN) trained over long sequences (e.g. minimum of 100-time steps) the gradients can easily vanish due to the magnitude of gradients being back propagated through the recurrent layers [16]-[19]. This problem affects the performance of recurrent networks in generating long term forecasts. In previous models [16]-[20] this problem persists which hinders the ability of the models in generating long term forecasts when applied on chaotic cases. Therefore, the proposed hybrid model provides procedures in enhancing learning of long-term dependencies with the aim of reducing the vanishing gradient problem as a result of the network being trained over long sequences (more than 100-time steps); hence generating robust long-term forecasts. This hypothesizes that the new non-linear hybrid model can cope with multifactor in chaotic time-series for robust multi-step-ahead forecasts as a result of enhanced learning in.

Based on current and previous stock prices, financial forecasting provides a way to the companies to visualize the trends and plans

Scaled Conjugate Gradient

Like other conjugate training algorithms this algorithm does not require line search at each iteration step. A time-consuming line search per learning iteration is avoided by using step size scaling mechanism. This mechanism differentiates it from other second order algorithms and makes it faster as compared to them. One of the limitations of this algorithm is that it requires more iteration to converge than the other conjugate gradient algorithms, but due to no line search the number of computations in each iteration is significantly reduced [33].

Conjugate Gradient backpropagation with Fletcher-Reeves Updates

The normal flow in conjugate algorithm in the first iteration is to start out by searching in the steepest descent followed by a line search to determine the optimal distance to move along the current search direction and this type of conjugate algorithm is computed as the ratio of the current gradient norm squared to the previous gradient norm squared [34].

Conjugate Gradient backpropagation with Powell-Beale Restarts

The search direction in all conjugate gradient algorithms is periodically reset to the negative of the gradient. In this algorithm the search direction is reset if between the current and the previous gradient even a little orthogonality is left.

Quasi-Newton Algorithms

Newton's algorithms give preferred and quick optimization over conjugate gradient algorithms. Newton's method preludes with the Hessian matrix of the performance index at the current values of the weights and biases, which enables it to converge faster than conjugate gradient methods. However, these algorithms tend to increase complexity, thereby increasing the computational time to feed Hessian matrix to forward neural networks. With each iteration, Hessian matrix is updated. Since these algorithms doesn't require the calculation of second derivatives, they are called quasi-Newton or secant method. They update an approximate Hessian matrix in each iteration of the algorithm.