Universidade Federal de Pernambuco
Centro de Informática

Computer Science Graduation

An Evolutionary Approach for Parameter Selection on Time series Forecasting Hybrid Models

Student: Ruy Brito Barbosa - rbb3@cin.ufpe.br
Advisor: Paulo Salgado Gomes de Mattos Neto - psgmn@cin.ufpe.br
Area: Artificial Intelligence

Recife, 22nd of August, 2019
Summary

Introduction 3

Goals 5

Schedule 6

Signatures 7

Possible evaluators 8

Bibliographic references 9
Introduction

Time series forecasting is comprehended as the prediction of future observations of quantifiable phenomena, given past observations through time. When the underlying generating process of the phenomenon cannot be formally described, approximative models are employed and used to extrapolate the time series into the future. Many occurrences in the literature focus on proposing new strategies for creating these models or improving existing ones. Improvements in this area may represent significant impact on several fields such as finance or climate analysis.

One of these models which has been widely used is the Autoregressive Integrated Moving Average (ARIMA). Despite being constrained to capture only the linear components of a time series, the simplicity of its statistical properties and the Box-Jenkins methodology in the model building process [1] have made it a popular choice over the past three decades.

On the other hand, Artificial Neural Networks (ANN) have been extensively employed on capturing the nonlinear components, being the first attempt reported by Lapedes and Farber [2]. The choice for ANNs is usually due to its capability of approximating any continuous measurable function with arbitrarily desired accuracy [3, 4]. Many further references in the literature describe successful applications and detailed analysis of ANNs usage on forecasting.

Support Vector Machines (SVM) applied to regression as proposed by Drucker, Harris et al. [5] have been also applied on forecasting as robust models. Thissen et al. [6] and Gui et al. [7] have employed the first occurrences of SVMs on time series forecasting. This approach maps samples from original space to high-dimensional space, through kernel tricks, and then constructs a linear decision function to achieve linear regression. The SVR can achieve better experimental results than the previous ARIMA model when time series data is very complex.

Hybrid approaches for time series forecasting combine different strategies to achieve better performance. Zhang [8] introduces the topic combining the ARIMA model with an ANN. Zhang’s Hybrid model assumes there is an additive correlation between linear and nonlinear components. The forecasting is performed in two steps, as to the neural network is assigned the task of predicting the error of the linear model forecasting. Empirical results suggests hybrid models can improve performance metrics of separate pure models.

Khashei and Bijari [9] have questioned Zhang’s primary assumption of an addictive relationship between linear and nonlinear components. Their hybrid approach, which is also based on ARIMA and ANN models, assigns to the neural network the task of deriving the correlation between these components. Empirical comparative results suggest they can improve the forecasting performance by generalizing this correlation.
The recent improvements in the time series forecasting field are significant. Yet there are challenges regarding the choice of parameters on the hybrid model design process. It has been shown that the choice of SVM hyperparameters has significant effect on performance metrics [10].

On the linear forecasting step of the aforementioned models, most of the parameters are derived from the statistical properties of the series. However, the choice of nonlinear prediction parameters represents a complex task. Trivial grid-search with cross-validation is not practical because of the large domain of possible parameter values, which is further exacerbated by the lack of prior knowledge on the data.
Goals

Understanding the challenges represented by the parameter selection on hybrid models for time series forecasting, the goal of this work is to present an evolutionary approach to approximate the optimal solution. The structure of the parameters to be optimised is derived from the strategy.

Assuming the ARIMA and the SVM as chosen strategies to be combined on the hybrid model, this work aims to design a Genetic Algorithm (GA) focused on optimising input parameters (lagged observations) and hyperparameters. By the end of the study, we aim to:

- Have a better understanding of the impact of parameter selection on time series forecasting hybrid models performance.
- Compare the performance of our evolutionary optimised hybrid models with previous examples from the literature.
- Understand which other variables on the hybrid model design process could also be optimised with more robust solutions.
## Schedule

<table>
<thead>
<tr>
<th></th>
<th>Aug</th>
<th>Aug</th>
<th>Sept</th>
<th>Sept</th>
<th>Oct</th>
<th>Oct</th>
<th>Nov</th>
<th>Nov</th>
<th>Dec</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Thesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Signatures

Ruy Brito Barbosa
Student

Paulo Salgado Gomes de Mattos Neto
Advisor
Possible Evaluators

George Darmiton da Cunha Cavalcanti - gcdc@cin.ufpe.br
Luciano de Andrade Barbosa - luciano@cin.ufpe.br
Bibliographic References


