Udacity 2018: Capstone Project

Forecasting time series with machine learning

with applications to currency exchange rates

Author: Oscar Javier Hernandez

August 30, 2018

1 Definition

1.1 Project Overview

A set of data that is indexed by time is known as a time series. They appear in many different fields, such as statistics, physics, finance, economics, biology, or even business [1]. Because of their wide applicability, it is important to generate accurate forecasts of time series data. These forecasts are generated using specific mathematical models or algorithms which are trained on a subset of the past values of a given time series. For the purpose of simplifying future discussions, we will adopt the following notation for a time series, denoted $X(t)$ or X_t, as

$$\{X(t); t = 0, 1, ...\}. \tag{1}$$

Where t denotes the time-index of the series. One of the simplest models for a time series is the ARIMA (Auto regressive integrated moving average) model. This model is denoted as ARIMA(p, q, d), and assumes that the time series X_t has the form

$$X_t = \mu + \epsilon_t + \sum_{i=1}^{p} \phi_i L^i \left[(1 - L)^d\right] X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}, \tag{2}$$

where $\{\phi_i | i = 1, ..., p\}$, $\{\theta_i | i = 1, ..., q\}$ are model parameters and L is the lag operator defined as $LX_t = X_{t-1}$. The term ϵ_t denotes the error terms, assumed to be independent, identically distributed random variables sampled from a zero-mean, normal distribution. The value μ denotes the average of this model. ARIMA models can be applied to make forecasts of stationary time series (defined as a time series whose mean, variance and auto correlation does not change over time), or to a time series that can be transformed into a stationary time series. However, there are other state-of-the-art machine learning methods that can be used to model time series methods. Which will the main goal of this project.

One important type of financial time series is the exchange rate between different currencies (Fig. 1). An exchange rate, is the rate at which one currency will be exchanged for another. There are many factors that can influence this rate, such as balance of payments, interest rate levels, inflation levels and other economical factors which are beyond the scope of this project [3].

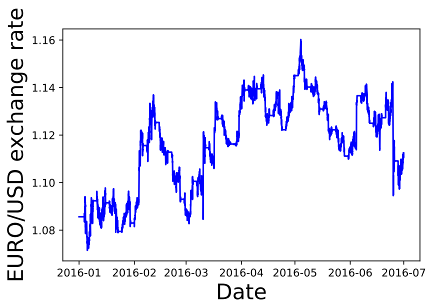

Figure 1: The exchange rate from EUR to USD from Jan 2016 to Jul 2016.

1.2 Project Statement

The main objective of this project will be to use classical and more recent machine learning techiques to make forecasts of different time series with the goal of applying the best methods to predict currency exchange rates. The simplest model that we will use is the ARIMA model, defined in the previous section as the baseline model, along with a linear regression model. The ARIMA model has been shown to be adequate in estimating the exchange rates of certain currencies Ref. [2]. We will then use different neural networks architectures such as the feed forward and recurrent networks, as in Ref. [5, 4, 6], to make predictions of time series and use our baseline models and root mean square differences to quantify and compare the performance of our different architectures.

1.3 Metrics

There are several metrics that we can use to evaluate the predictions of our models Ref. [1], however, for our project we will focus on two commonly used metrics. We will first define some terminology, the forecast error, e_t, is given by

$$e_t = X_t - F_t, \tag{3}$$

where X_t is the value of the time series at time step t, and F_t is the forecasted value at the same time step. The three metrics that we will use for our project are

1. The mean square error (MSE)
   - $\text{MSE} = \frac{1}{N} \sum_{t=1}^{N} e_t^2$
2. The root mean squared error (RMSE)
   - $\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} e_t^2}$

In these three cases, the smaller the value of the MAPE, MSE, and MAE, then the better the model.

2 Analysis

2.1 Data Exploration

2.2 Exploratory Visualization

2.3 Algorithms and Techniques

2.4 Benchmark

Using the

3 Methodology

3.1 Data Preprocessing

4 Results

4.1 Model Evaluation and Validation

4.2 Justification

5 Conclusion

5.1 Free-Form Visualization

5.2 Reflection

5.3 Improvement

References

[1] R. Adhikari and R. K. Agrawal, An Introductory Study on Time Series Modeling and Forecasting, 2013, [arXiv:1302.6613] https://arxiv.org/abs/1302.6613.

[2] Research Journal of Finance and Accounting www.iiste.org ISSN 2222-1697 (Paper) ISSN 2222-2847 (Online) Vol.7, No.12, 2016 http://iiste.org/Journals/index.php/RJFA/article/viewFile/31511/32351.

[3] P. J. Patel, N. J. Patel and A. R. Patel, IJAIEM 3, 3, 2014. http://www.ijaiem.org/volume3issue3/IJAIEM-2014-03-05-013.pdf

[4] B. Oancea, S. Cristian Ciucu, Proceedings of the CKS 2013, [arXiv:1401.1333] https://arxiv.org/abs/1401.1333.

[5] T. D. Chaudhuri and I. Ghosh, Journal of Insurance and Financial Management, Vol. 1, Issue 5, PP. 92-123, 2016, [arXiv:1607.02093] https://arxiv.org/abs/1607.02093.

[6] Pant, N. (2017, September 07). A Guide For Time Series Prediction Using Recurrent Neural Networks (LSTMs). Retrieved from https://blog.statsbot.co/time-series-prediction-using-recurrent-neural-networks-lstms-807fa6ca7f.

[7] Vincent, T. (2018). ARIMA Time Series Data Forecasting and Visualization in Python — DigitalOcean. [online] digitalocean.com. Available at: https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3 [Accessed 29 Jul. 2018].

[8] C. Esteban, S. L. Hyland and G Rätsch, https://github.com/ratschlab/RGAN [arXiv:1706.02633].