Predictive maintenance of production equipment based on neural network autoregression and ARIMA

Patrícia Ramos (patricia.ramos@inescporto.pt)
INESC TEC, ISCAP, Instituto Politécnico do Porto
Rua Dr. Roberto Frias, 378 4200-465, Porto, Portugal

José Manuel Oliveira (jmo@inescporto.pt)
INESC TEC, Faculdade de Economia, Universidade do Porto
Rua Dr. Roberto Frias, 378 4200-465, Porto, Portugal

Paula Silva (psilva@inescporto.pt)
INESC TEC, Rua Dr. Roberto Frias, 378 4200-465, Porto, Portugal

Abstract

This paper presents a predictive study applied to a manufacturing equipment in order to predict malfunctions, and consequently enabling predictive maintenance practices. ARIMA forecasting methods are successfully compared with neural networks models, both used over data obtained from a monitoring system that continuously keeps track of the relevant equipment parameters. The results show that both models could detect the discs replacement, however The ARIMA model forecasts quite well the increasing of the distance between the discs before and after the replacement which is not the case for the NN model.

Keywords: Manufacturing equipment, Predictive maintenance, ARIMA, Neural network autoregression

Introduction

Globalization and competitiveness in existing markets currently cast an increasingly demanding challenge for organizations. The delivery of the product or service desired by the customer is becoming less a differentiating factor, but a matter of survival. The client demands that the product is produced according to the desired characteristics to the first, with guaranteed quality and on time. This increasingly challenge, driven by the need to continuously optimize the quality of products, made maintenance began to be treated in a different way. Maintenance, here, is seen as the set of technical and administrative actions designed to maintain acceptable conditions in manufacturing facilities and equipment to ensure regularity, quality and safety in production, with minimal total costs. Intelligent methods for collecting and organizing data and predict potential failures will contribute greatly to the effectiveness of the machine preventive/predictive maintenance.

From visual inspection, which is the oldest method, yet still one of the most powerful and widely-used, predictive maintenance has evolved to automated methods that use advanced signal processing techniques based on pattern recognition, including neural networks, fuzzy logic, and data-driven empirical and physical modelling (Hashemian, 2011). As equipment begins to fail, it may display signs that can be detected by human senses (eyes, ears and noses) or by sensors that are currently available to identify the onset of equipment degradations and failures. Integrating these sensors with the predictive maintenance techniques can avoid unnecessary equipment replacement, save costs, and improve process safety, availability and efficiency.

The prediction of failures and maintenance actions of industrial machines is a problem with interesting characteristics. We need to forecast certain rare events, which are supposed to be dependent on the recent values of a set of time series values. These time series describe the recent values of a set of sensors that monitor several aspects of the industrial machines. For each task being handled by these machines (a kind of working context), the sensors are expected to have a certain typical behavior. Deviations from this typical behavior are good indicators of a foreseen failure or some maintenance action.

A large number of different approaches have been used to develop models for predictive maintenance, including Data Mining and Statistical Inference Methods particularly nonparametric techniques (Bohoris and Leitão, 1991) (Lopes et al., 2010). However, the process of model development is, to a great extent, manual. Indeed, time series forecasting models need different preprocessing tasks in order to identify the existing seasonalities and impact factors (Makridakis et al., 1998; Pena et al., 2001). Also, data mining algorithms need to address different preprocessing tasks and parameter tuning (Mendes-Moreira et al., 2012). Therefore, in order to develop predictive maintenance models for machines operating in diverse environments requires a significant amount of expensive human effort. Furthermore, it is hard to ensure that the models remain reliable over time in dynamic environments. Despite the various solutions available for the detection of potential failures, predictive maintenance derived from a correct failure prediction is not yet a reality.

The Box-Jenkins approach to modeling ARIMA processes provides a convenient framework to find an appropriate statistical model which can be used to make forecasts (Box et al., 1994). Zhao et al.(2007) estimates an ARMA model to forecast faults in a semiconductor ATM factory, and there are other successful examples.

Neural networks are a class of flexible nonlinear models that can discover patterns adaptively from the data. Theoretically, it has been shown that given an appropriate number of nonlinear processing units, neural networks can learn from experience and estimate any complex functional relationship with high accuracy. Empirically, numerous successful applications have established their role for pattern recognition and forecasting (Zhang and Qi, 2005).

The first step for addressing a predictive study is the construction of a good quality data set. Such data set should provide the models with examples of the rare events we are trying to forecast. Without a history of failures and maintenance activities it is not possible for models to forecast these events. The ideal data set should consist of a time tagged sequence of observations of the machine state. The second step is to forecast the future values of the sensors of the machine. Regarding this we will use ARIMA and neural networks models. The final step of the predictive study is to detect failures and maintenance actions based on the forecasts of the sensors future values.

The rest of the paper is organized as follows. The following section gives a brief description of the forecasting models used in this work: ARIMA and neural networks models. Next, we present the data and the methodology followed to develop the work.

The obtained results are then presented. Finally, we highlight the main conclusions and ideas for future work.

Forecasting models

ARIMA modeling

ARIMA is one of the most versatile linear models for forecasting time series. It has enjoyed great success in both academic research and industrial applications during the last four decades. The class of ARIMA models is broad. It can represent many different types of stochastic seasonal and non-seasonal time series such as pure autoregressive (AR), pure moving average (MA) and mixed AR and MA processes (Chatfield, 2000). The theory of ARIMA models has been developed by many researchers and its wide application was due to the work by Box and Jenkins (1994) who developed a systematic and practical model building method. The non-seasonal ARIMA model denoted as ARIMA$\left(p,d,q\right)$ has the following form (Brockwell and Davis, 1991):

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)\left(1-B\right)^d\, y_t=c+\left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right)\varepsilon_t, \tag{1}$$

where B is the backward shift operator, d is the degree of first differencing involved, $\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)$ and $\left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right)$ are the regular autoregressive and moving average polynomials of orders p and q respectively, $c=\mu(1-\phi_1-\cdots-\phi_p)$ where μ is the mean of $(1-B)^d y_t$ process and ε_t is a normally distributed white noise process with mean 0 and variance σ^2. The roots of polynomials $\phi_p(B)$ and $\theta_q(B)$ should lie outside a unit circle to ensure causality and invertibility. For $d\geq 1$, $c=0$ is usually assumed because a quadratic or a higher order trend in the forecast function is particularly dangerous (Shumway and Stoffer, 2011).

The main task in ARIMA forecasting is selecting an appropriate model order, that is the values of p,q and d. Usually the following steps are used to identify manually a tentative model (Wei, 2005):

(1) Plot the time series, identify any unusual observations and choose the proper variance-stabilizing transformation. A series with nonconstant variance often needs a logarithm transformation. More generally to stabilize the variance a Box-Cox transformation may be applied;

(2) Compute and examine the sample ACF (autocorrelation function) and the sample PACF (partial autocorrelation function) of the transformed data (if a transformation was necessary) or of the original data to further confirm a necessary degree of differencing. Because variance-stabilizing transformations such as the Box-Cox transformations require positive values and differencing may create some negative values, variance-stabilizing transformations should always be applied before taking differences;

(3) Compute and examine the sample ACF and sample PACF of the properly transformed and differenced series to identify the orders of p and q by matching the patterns in the sample ACF and PACF with the theoretical patterns of known models.

(4) After identifying a tentative model the next step is to estimate the parameters in the model. This quite complex task is usually performed by a software package. After identifying an appropriate model the residuals from the model should be checked (Ljung and Box, 1978).

Neural network autoregression

Neural networks are the most versatile nonlinear models that can represent both nonseasonal and seasonal time series (Chu and Zhang, 2003). The most important capability of neural networks compared to other nonlinear models is their flexibility in modeling any type of nonlinear pattern without the prior assumption of the underlying data generating process. The most popular neural network model for time series forecasting is the three-layer feedforward network model which can be written as:

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-i} + \beta_{0j} \right) + \varepsilon_t \tag{2}$$

where m is the number of input nodes, n is the number of hidden nodes, f is a sigmoid transfer function such as the logistic: $f(x) = \frac{1}{1+\exp(-x)}$, $\{\alpha_j, j = 0,1,\ldots,n\}$ is a vector of weights from the hidden to output nodes and $\{\beta_{ij}, i = 0,1,\ldots,m;, j = 1,\ldots,n\}$ are weights from the input to hidden nodes, α_0 and β_{0j} are weights of arcs leading from the bias terms which have values always equal to 1. The neural network expressed in (2) is equivalent to a nonlinear AR model.

For a time series forecasting problem NN model building is equivalent to determining both the number of input nodes and the number of hidden nodes. The input nodes are the past lagged observations through which the underlying autocorrelation structure of the data can be captured. Identifying the proper autocorrelation structure of a time series can be done by examining the sample ACF and the sample PACF. Although the NN universal approximation theory indicates that a good approximation may require a large number of hidden nodes, only a small number is often needed in real applications (Chu and Zhang 2003).

Experiments with different architectures are often performed to identify an appropriate neural network model. In fact, the available data is often divided into three portions. The first training part is used for model training, i.e. parameter estimation, while the second validation part is for model selection. The last test sample is then used for true forecasting evaluation.

Data and Methodology

The predictive maintenance work presented in this paper was carried out over a refiner located in one of the Sonae Indústria factories. Sonae Indústria is the current world leader in the production of wood derivative panels, being the owner of 27 factories distributed by 3 continents. Their products cover the most usual wood derivative panels, namely particleboard, medium density fiberboard (MDF), hardboard and oriented strand board.

The refiner is used to mechanically separate wood fibers from pieces of wood that are introduced between two metallic discs, one of them rotary, separated by few millimeters. Figure 1 illustrates the refiner components.

The data considered in this work consist on a large set of time series each describing a key sensor of the machine under study (the refiner). These sensors describe relevant properties of the refiner and their evolution through time being able to provide hints on future failures or maintenance needs.

Our work focus on the maintenance of the defibrator discs, in particular on their replacement. Consequently, the sensors analyzed were Sensor 11 which measures the defibrator infeed screw motor current (in amperes) and Sensor 20 which measures the distance (in millimeters) between the two discs of the defibrator. The Sensor 11 is used to detect the periods when the refiner is turned off. These periods were not considered in the time series analysis of this work, i.e., we assume that the refiner is working continuously. Thus, the Sensor 20 measurements were only considered when Sensor 11 is not down.

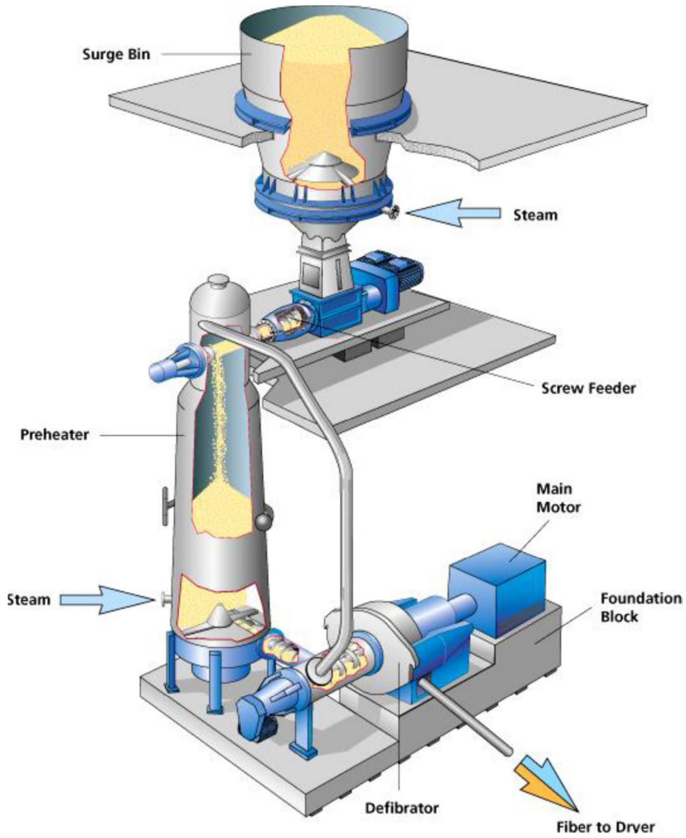

Figure 1 – Refiner components.

The monitoring time started on 2008-01-02 04:55:54 GMT and finished on 2011-11-30 23:59:55 GMT. Figure 2 shows the observations from Sensor 11, already without the periods where the refiner is turned off. This data set comprises 7062589 observations, being each measurement taken each 8/9 seconds. Thus, the data set corresponds to a total period of around 671 days where the refiner is continuously working.

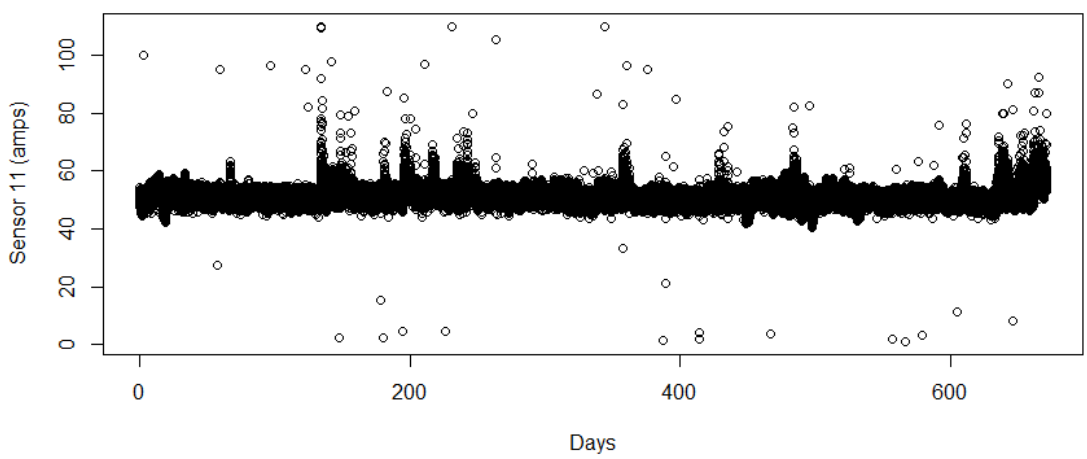

Figure 2 – Measurements of Sensor 11 during 671 working days.

Figure 3 shows the Sensor 20 measurements taken on the exact time instants of the Sensor 11 data set, depicted in Figure 2. It is clear from Figure 3 that the Sensor 20 data set is composed by cycles of values between 10 and 30 mm. Each cycle corresponds to a pair of discs of the defibrator that are then replaced by new ones, after a working period. When the discs are replaced by new ones its distance inside the defibrator decreases which is detected in the measurements of Sensor 11 by a jump down.

Taken the historical data of discs replacements given by Sensor 20, the purpose of this work is to predict when the next replacement of the defibrator discs should occur, to avoid possible faults in the refiner derived by the use of improper discs.

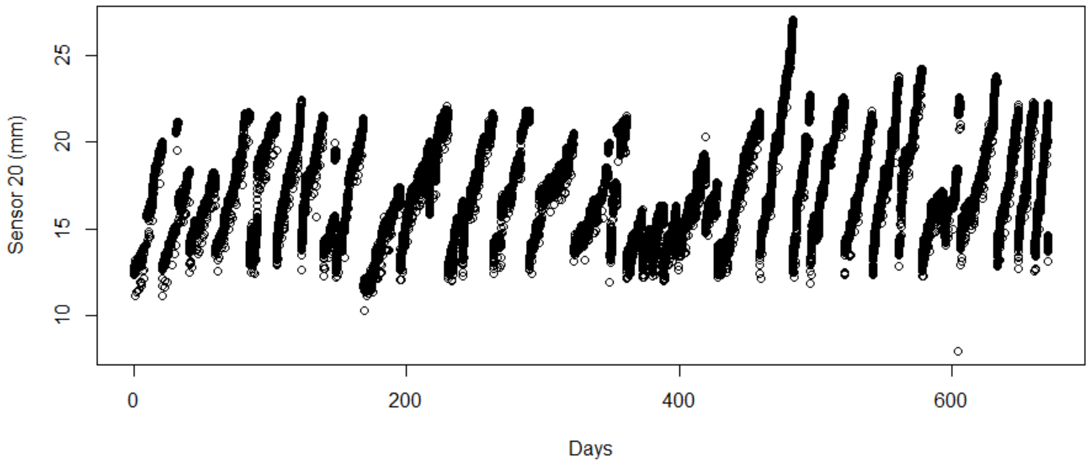

Figure 3 – Measurements of Sensor 20 during 671 working days.

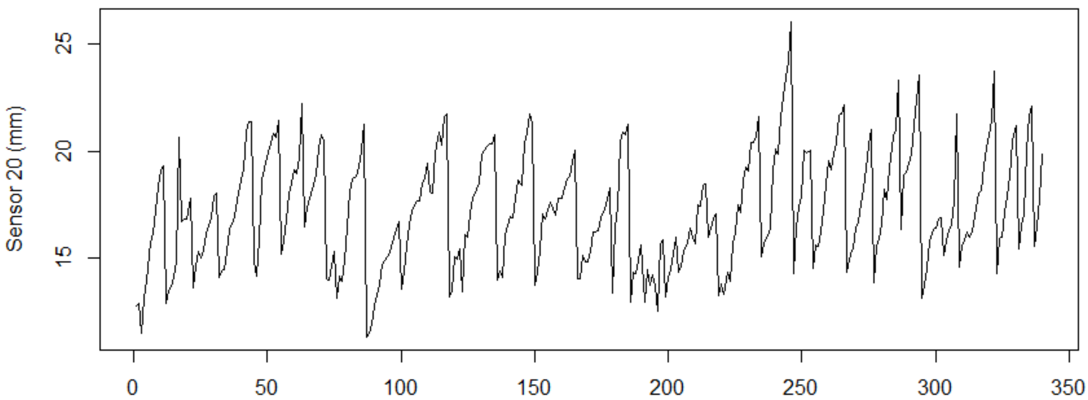

Figure 4 – Measurements (one each two days) of Sensor 20 used in the analysis.

Figure 4 shows the plot of the Sensor 20 measurements (one each two days) that were analyzed in this work using ARIMA and NN (340 observations). It can be seen that the time series has a cyclic behavior and that the cycles have no fixed length. This plot also indicates that the time series is stationary in the mean but may not be stationary in variance, so a Box-Cox transformation should be investigated.

Figure 5 shows the sample ACF and the sample PACF for the measurements of Sensor 20 time series. It can be seen that the sample ACF decays very slowly and that the sample PACF has a large spike at lag 1 and another spike at lag 37. This large autocorrelation at lag 37 suggests that a tentative ARIMA model may be an AR(37), and that 37 input nodes should be tried when considering the NN architecture.

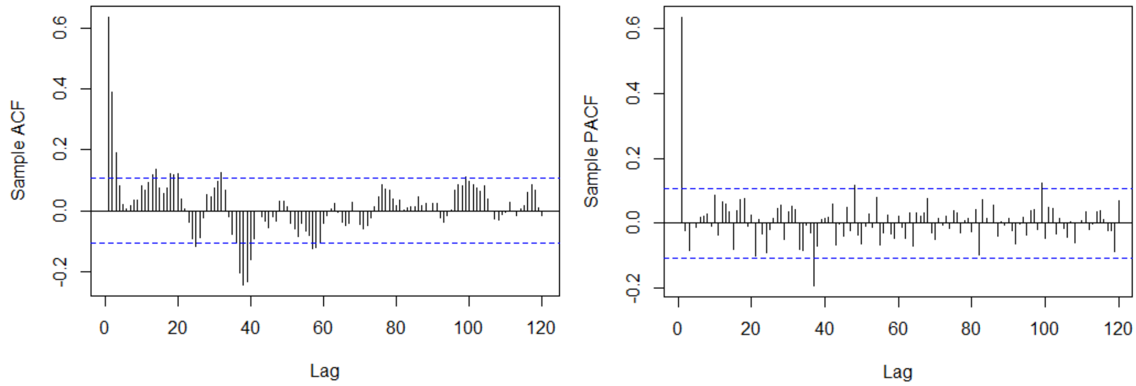

Figure 5 – Sample ACF and Sample PACF for the Sensor 20 measurements.

It was also our purpose to compare the forecasting performance of ARIMA models and NN models. It is important to evaluate forecast accuracy using genuine forecasts. That is, it is not valid to look at how well a model fits the historical data. The accuracy of forecasts can only be determined by considering how well a model performs on data that were not used when fitting the model. When comparing different models, it is common to use a portion of the available data for fitting – the in-sample data, and use the rest of the data to measure how well the model is likely to forecast on new data – the out-of-sample data. Following the common practice, we use the cross-validation approach to select the best ARIMA model and the best NN architecture (Arlot and Alain, 2010). That is, the in-sample data are further split into a training set and a testing set. The training set is used to estimate the model parameters and the testing set is used to choose the final model.

In this study, the first 333 observations of the time series are used for model fitting and selection (in-sample data) and the last 7 observations are used for forecast evaluation (out-of-sample data). The last 7 observations of in-sample data are used as validation and testing sample and the rest of observations are used for model estimation (326 observations). The model with the best performance in the testing sample is selected as the final model for further evaluation in the out-of-sample. All model comparisons are based on the results for the out-of-sample.

In the case of NN models we consider 6 different levels of input nodes: 1, 2, 3, 36, 37 and 38, and 12 hidden node levels from 2 to 24 with an increment size of 2. Thus, a total of 72 different networks are experimented in the model building process. We also investigate whether the Box-Cox transforming of the data would enhance neural network's capability of modeling the cyclic variation.

In the case of ARIMA we consider the models where p can take the values 1, 2, 3, 36, 37 and 38, q can take the values 0, 1, 2, 3, and 4, and c can take the values 0 or 1 giving a total of 60 models. To investigate the stabilization of the variance the Box-Cox transformation was also applied. The model with the minimum RMSE value on the forecasts of the testing sample that passed the Ljung-Box test with a significance level of 5% was selected from all fitted ARIMA models.

Results

The time series analysis was carried using the statistical software R programming language (R Development Core Team, 2013) and the specialized package forecast (Hyndman, 2008).

To evaluate and compare the forecasting performance of the two types of models, we use three overall error measures in this study: the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). These are the most commonly used forecast error measures among both academics and practitioners.

Table 1 summarizes the ARIMA and NN modeling results for the in-sample data. It can be seen that neural networks are able to model and forecast better than ARIMA in the in-sample judged by the three performance measures; although it should be emphasized that these results should not be used for forecast evaluation. In the testing set the RMSE, the MAE and the MAPE are 39%, 29% and 19% smaller, respectively.

Another important observation is that the Box-Cox transformation was not important for improving NN and ARIMA's ability to model and forecast the cyclic behavior of the Sensor 20 time series.

Table 1 – Comparison results for model building (in-sample data).

Model	RMSE	MAE	MAPE
Training set			
ARIMA	1.894	1.312	8.010
NN	1.603	1.105	6.649
Testing set			
ARIMA	2.963	2.208	11.174
NN	1.796	1.561	9.101

The out-of-sample forecasting comparison between neural networks and ARIMA is presented in Table 2. The results of this table show that the out-of-sample forecasting performance of ARIMA models evaluated via RMSE, MAE and MAPE is better than NN models. The RMSE, the MAE and the MAPE are 39%, 38% and 35% smaller, respectively. Sometimes, different accuracy measures will lead to different results as to which forecast method is best. However, in this case, all the results point to ARIMA as the best of the methods for this data set. These results also emphasize that a model which fits the data well does not necessarily forecast well.

Table 2 – Out-of-sample forecasting comparison between neural networks and ARIMA.

Model	RMSE	MAE	MAPE
ARIMA	2.181	1.896	9.814
NN	3.553	3.048	15.142

To see the individual point forecasting behavior, we plotted the actual data versus the forecasts from both NN and ARIMA models in Figure 6. We find that both NN and ARIMA models have the capability to forecast the discs replacement (observation 337). However the jump down in the ARIMA forecast is more prominent then the jump down in the NN forecast. The ARIMA model forecasts quite well the increasing of the distance between the discs before and after the replacement which is not the case for the NN model. It can be seen a decreasing of the distance between the sensors in the last two measurements that does not correspond to the reality.

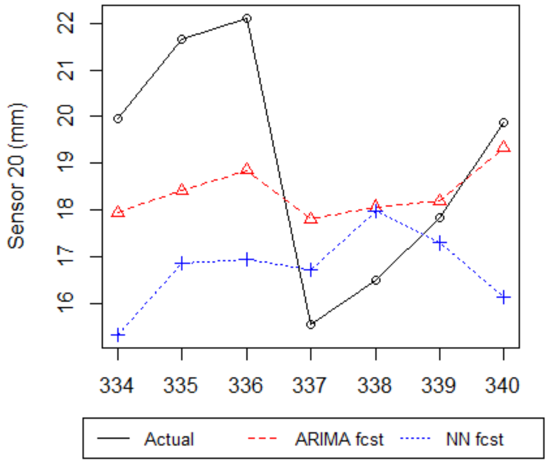

Figure 6 – Out-of-sample forecasting comparison for the Sensor 20 time series (14 days).

Conclusions

Manufacturing enterprises, particularly SMEs, are quickly evolving according to market and products fast changes. Manufacturing machines are a core enabling technology in a number of key European industrial sectors which have common requirements for increased product customization and improved competitiveness in terms of reduced cost, shorter delivery times and improved quality. In order to pursue these increasing needs, manufacturing machines should be more and more reliable and available. The greater integration between the machine performances and the related parameter (technical, environmental and process) becomes also a crucial requirement that the user of the machine has difficulties to grasp and control. Intelligent methods for collecting and organizing data and predict potential failures will contributes greatly to the effectiveness of the machine preventive/predictive maintenance. Despite the various solutions available for the detection of potential failures, predictive maintenance derived from a correct failure prediction is not yet a reality.

In this work we compare the forecasting performance of ARIMA and NN models to detect failures and maintenance actions based on the forecasts of the sensors future values. We concluded that both NN and ARIMA models have the capability to forecast the discs replacement detected in Sensor 20. However the jump down in the ARIMA forecast is more prominent then the jump down in the NN forecast. The ARIMA model forecasts quite well the increasing of the distance between the discs before and after the replacement which is not the case for the NN model.

Acknowledgments

Project “NORTE-07-0124-FEDER-000057” is financed by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT).

This work was also co-supported by the PRODUTECH-PSI: Novos Produtos e Serviços para a Indústria Transformadora project from QREN – ON.2 program.

References

Arlot, S. and Alain, C. (2010), “A survey of cross-validation procedures for model selection”, *Statistics Surveys*, 4, 40–79.

Box, G.E.P., Jenkins G.M. and Reinsel G.C. (1994), *Time series analysis: forecasting and control*, Prentice-Hall, Englewood Cliffs, New Jersey.

Bohoris, G. and Leitão, A.L.F. (1991), "Proportional Hazards Modelling Versus Two-Sample Tests: A Case Study", *Quality and Reliability Engineering International*, 7, 393-402.

Brockwell, P.J. and Davis, R.A. (1991), *Time series: theory and methods*, Springer-Verlag, New York.

Chatfield, C. (2000), *Time-series forecasting*, Chapman & Hall/CRC, Boca Raton.

Chu, C.W. and Zhang, P.G.Q. (2003), "A comparative study of linear and nonlinear models for aggregate retail sales forecasting", International Journal of Production Economics. 86, 217–231.

Hashemian, H.M. (2011), "State-of-the-Art Predictive Maintenance Techniques", IEEE Transactions on Instrumentation and Measurement, 60 (1), 226-236.

Hyndman, R.J. (2008), "forecast: Forecasting functions for time series. R package version 4.06", http://cran.rstudio.com/.

Hyndman, R.J. and Athanasopoulos, G. (2013), *Forecasting: principles and practice*, http://otexts.com/fpp/.

Ljung, G.M. and Box, G.E.P. (1978), "On a measure of lack of fit in time series models", *Biometrika*, 65, 297–303.

Lopes I., Pires L. and Bastos P. (2010), "A Decentralized Predictive Maintenance System Based on Data Mining Concepts", *Business Sustainability*, 254-258.

Makridakis, S., Wheelwright, C. and Hyndman, R.J. (1998), *Forecasting: methods and applications*, Wiley.

Pena, D., Tiao, G.C. and Tsay, R.S. (2001), *A course in time series analysis*, John Wiley & Sons, New York.

R Development Core Team (2013), *R: A language and environment for statistical computing*, http://www.R-project.org/. R version 3.0.0.

Shumway, R.H. and Stoffer, D.S. (2011), *Time series analysis and its applications: with R examples*, Springer, New York.

Wei, W.S. (2005), *Time series analysis: univariate and multivariate methods*, Addison Wesley.

Zhang, G. and Qi. M. (2005), "Neural network forecasting for seasonal and trend time series", European Journal of Operational Research 160: 501-514.

Zhao, J., Xu, L. and Liu, L. (2007) "Equipment Fault Forecasting Based on ARMA Model", *Proceedings of the IEEE Conference on International Conference on Mechatronics and Automation*, August 5 - 8, 2007, Harbin, China.