

Applying Predictive Analytics in a Continuous Process Industry



Nitin Merh

Abstract In this paper an attempt is made to develop data driven models on pilot data set for predicting fault in machines of continuous process industry on various selected attributes using techniques of Multiple Linear Regression Model (MLR), Regression Tree (RT) and Artificial Neural Networks (ANN). Association rules are also derived from the available data set. Efforts are also made to predict total shutdown time of machines. These machines are used for manufacturing components machined for Heavy Commercial Vehicles (HCV), Light Commercial Vehicles (LCV), Multi Axle Vehicle (MA) and Tractors. To check the robustness of models a comparison is made between the results derived from various techniques discussed above. Performance evaluation is done on the basis of the errors calculated between the actual and predicted values of down time. Based on actual and predicted results various error scores are calculated to evaluate best model and check robustness of the models under study. Training and validation of the model is done using datasets collected from a manufacturing unit located at Pithampur industrial area near Indore, Madhya Pradesh, India. In the current paper an association is also developed between the attributes and occurrence of the fault. The developed model will be used on the bigger data set which will help the stakeholders of the organization for smooth functioning of the unit and for better governance in the organization. XLMiner is used for model development and simulations. After analysis results show that ANN, RT and Association Rule techniques are capable of capturing the data set.

Keywords Artificial neural network (ANN) · Regression tree (RT)
Association rule

N. Merh (✉)

Information Technology, SVKM'S NMIMS (Deemed to be University),
Indore 452001, Madhya Pradesh, India
e-mail: nitinmerh0812@gmail.com

© Springer Nature Singapore Pte Ltd. 2019

A. K. Laha (ed.), *Advances in Analytics and Applications*, Springer Proceedings
in Business and Economics, https://doi.org/10.1007/978-981-13-1208-3_10

105

1 Introduction and Literature Review

Fault prediction mainly deals with the fault that is likely to happen in the system on the basis of past and current states of the system. Fault prediction has attracted considerable attention across the world due to the growing demand for higher operational efficiency, safety in industrial systems and scheduling of shutdowns.

Modeling, data mining and machine learning are among the few areas of study for predictive analytics. Various statistical techniques are used for their implementation that analyze past data to make future prediction. Prediction helps organization in making right decision at right time by right person as there is always time lag between planning and actual implementation of the event. In a continuous work-flow or continuous process all outputs are treated similar. In such a case the process itself is divided in separate operations. Each unit flows among these operations, individually. In such kind of system the manufacturing of the standard products is carried out at a fixed rate. The mass production is carried on continuously for stock in anticipation of demand.

Principal component analysis (PCA) is a multivariate technique that analyzes data sets in which several inter-correlated quantitative variables exist. PCA is a mathematical technique which tries to find a set of uncorrelated variables among several correlated variables. Main goal of PCA is to extract important information from data set and to represent it as a set of new uncorrelated variables.

The PCA aims at reducing the number of variables of the dataset which define the dimensionality of the system, but the original variability in the data is retained and the complexity is reduced.

Thus PCA mainly explains the variance-covariance factor of a high dimensional system using only a set of few linear combinations of actual component variables.

Multiple Linear Regression (MLR) is one of the several prediction techniques used. It is applied on the dataset to understand the relationship between response and predictor variables or prediction of the response based on input variables. MLR uses following linear model for the selected dataset:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m + \varepsilon$$

where Y is the dependent or response variable. X_i : for $i = 1$ to m ; represents the independent or predictor variables, α_0 is intercept term and ε is the random error. α_i : for $i = 1$ to m are regression coefficients.

The input variables and predicted variables are main components of a decision tree. The nodes of the tree represent a test performed on an input variable and the prediction variables are the terminal nodes of the decision tree. A regression tree can be viewed as an adoption taken from decision trees.

Learning and pattern recognition problems can be solved by the use of Artificial Neural Networks (ANN). The learning process of ANN helps to find meaningful patterns in data (Afolabi and Olude 2007). Approximation of unknown functions, with no assumptions being made for the distribution of data, can be done by ANN,

that to for a desired accuracy (Sexton and Sikander 2001). Approximation of both types, linear and non-linear functions can be done by ANN, resulting in achievement of good performance. Because ANN learn and follow a non-parametric approach, they have gained popularity (Dacha 2007; Rimpley 1996).

The current study aims to find the predictive models for fault detection in the machines of continuous process industry. Selected predictive model/s will help stakeholders to take right decision at right time. It will also help in scheduling planned shutdowns and selecting key attributes responsible for forced or unscheduled shutdowns.

Indore CNC Pvt. Ltd. is a manufacturing unit located in Pithampur, Indore is selected as the organization under study. It is manufacturer of gear boxes which are supplied to heavy commercial vehicles, light commercial vehicles, multi axle vehicle and tractors manufactures located in Madhya Pradesh and other parts of the country. Due to unscheduled shutdowns/breakdowns inventory management, manpower planning and finances have suffered a lot. Scheduled delivery of the finished product to the clients is also hampered, which creates a bad name to the organization.

Samantha and Al-Balushi (2003) and Kankar et al. (2011) have demonstrated use of ANN for diagnosing faults in the manufacturing of rolling element bearings. The inputs which are used for ANN are time domain vibration signals of all bearings normal or defective used in the rotating machinery.

Artificial Intelligence (AI) has the capability to learn and acquire knowledge from facts, data and principles, which is then applied to a process. This capability of AI is used in engineering applications thus attracting many researchers and practitioners.

Key objective of the study is to construct predictive models for predicting main attributes of fault detection in continuous process industry. Next reason for undertaking this study is to compare trends and results of actual and predicted value generated by various models and finding the best model under study, to find out key factors responsible for the unscheduled shutdowns and to prescribe actions to be taken to reduce unscheduled shutdowns, to find out which types of errors occur together. Last objective is performance evaluation of models by statistical methods and by calculating and comparing various errors.

Methodology Used

In the current study an effort is made to develop models using predictive methods like Regression Tree (RT) and Artificial Neural Network (ANN) for predicting fault detection in a continuous process industry. Primary data for analysis is collected from Indore CNC located at Pithampur industrial area near Indore, Madhya Pradesh (M.P.).

At the first stage data is preprocessed, transformed, missing values are handled, outliers are identified and handled, data normalization and principal components are selected. After transformation of the data various selected techniques are used for predicting the fault and models are created.

Convergence, robustness and model evaluation has been done on the basis of the simulation results obtained by XLMiner. After the development of models using MLR, NN and RT the comparison of various forecasting errors have been calculated.

2 Data

Input data for all methods for developing predictive modeling is collected from manufacturing unit located at Pithampur, Indore, M.P., India for a period during 01.04.2015 to 30.09.2015. The collected data is used for the pilot study and on the basis of the results and inferences generated same can be applied for the larger dataset. Total 155 days sample is collected during period of 01.04.2015 to 30.09.2015. Few days were dropped due to holidays, shutdowns or when data was not generated. Data was collected for Tongtai-1 CNC machine used for manufacturing gear boxes. Time loss data was collected under various heads and attributes.

Pre-processing and Normalization of Data

Final dataset has been prepared after removing attributes having 0 values (No time loss), unary values. Attributes selected from Table 1 for developing model are as follows (Table 2).

Random partitioning has been done on the data set where 60% data is for training and 40% for validation.

For association rule analysis data set is converted in binary format where zero (0) represents non-occurrence of time loss and one (1) represents occurrence of time loss.

3 Data Analysis and Results

Main objective of the study is to find suitable predictive data driven model using various techniques. Neural Networks, Regression Tree, Multiple Linear Regression and Association Rule Mining are used to fit the data for developing the model. After developing models from above mentioned techniques, model with best results can be selected for final deployment on the bigger data set generated by not only Tongtai-1 machine but also on Taknio 86C and Hyundai machines which are installed in Indore CNC. A comparison can also be made on the performance on all machines based on the results generated.

Results of data analysis are as follows:

Dependent variable—Total (It is derived from calculating total down time occurred during a day).

Independent variables—Air Pressure Low Loss Time, APC Loss Time, ATC Loss Time, Magazine Problem Loss Time, Operator Door Problem Loss Time, Electrical Problem Loss Time, Power Cut Loss Time, Tool Broken Loss Time, Tool Fall Down Loss Time, Tool Grinding Loss Time, Coolant Rust Problem Loss Time, Fixture Work Loss Time, Setting Time Loss Time, House Keeping Loss Time, Insert Change Loss Time, Offset Given Loss Time, Spindle Chips Cleaning Problem Loss Time, Tool Proving Time Loss Time, Gauges Problem Loss Time, Inspection Time Loss Time, Rework Loss Time, Tools Not Available Loss Time, Insert Not Available Loss Time,

Table 1 Dimension wise attributes for Tongtai-1 machine

Mechanical	Electrical	Electronics or system	Age of machine	Human intervention	Training or efficiency	Quality	Planning
Air pressure low	Electrical problem	Sensor problem	Running time tool broken	Coolant rust problem	Bore adjust	Casting problem	Coolant problem
APC problem	Power cut	Servo alarm	Spindle over load problem	Data not filled		Excess load	Hydraulic oil low
ATC problem	Lubrication alarm	Spindle drive problem		Fixture work	House keeping	Gauges problem	Insert not available
B Axis clamping delay		X Axis drive problem	Tool broken	Operator late coming	Insert change	Inspection time	Load delay
Conveyor not working		Y axis drive problem	Tool fall down	Setting time	Lathe insert change	Lathe inspection	No helper
Coolant leakage		Z axis drive problem	Tool grinding		Machine cleaning	PDI work	No operator
Crane problem					Offset given	Rework	Production meeting
Magazine problem					Spindle chips cleaning problem	Tools not available	Single pallet
Mechanical problem					Tool proving time		Stud problem
Oil leakage							Tool problem
Operator door problem							
Telescopic cover							

Load Delay Loss Time, No Operator Loss Time, Stud Problem Loss Time, Servo Alarm, Spindle Drive Problem.

Variables dropped due to invalid inputs/unary values—Casting Problem Loss Time, X Axis Drive and Timeloss.

Results:

Artificial Neural Network:

Following parameters of NN were used for designing NN model (Tables 3, 4 and 5).

Table 2 Final selected attributes for Tongtai-1 machine

1	Air pressure low loss time	16	Offset given loss time
2	APC loss time	17	Spindle chips cleaning problem loss time
3	ATC loss time	18	Tool proving time loss time
4	Magazine problem loss time	19	Casting problem loss time
5	Operator door problem loss time	20	Gauges problem loss time
6	Electrical problem loss time	21	Inspection time loss time
7	Power cut loss time	22	Rework loss time
8	Tool broken loss time	23	Tools not available loss time
9	Tool fall down loss time	24	Insert not available loss time
10	Tool grinding loss time	25	Load delay loss time
11	Coolant rust problem loss time	26	No operator loss time
12	Fixture work loss time	27	Stud problem loss time
13	Setting time loss time	28	Servo alarm
14	House keeping loss time	29	Spindle drive problem
15	Insert change loss time	30	X axis drive

Table 3 Artificial neural network parameters

Parameters	
If input variables are normalized	Yes
Type of network architecture	Manual
Initial weights used/seed value	12,345
Number of hidden layers	1
Number of nodes in hidden layer 1	25
Number of epochs	30
Step size for gradient descent	0.1
Weight change momentum	0.6
Error tolerance	0.001
Weight decay	0
Cost function	Sum of squares
Activation function used at hidden layer	Standard
Activation function used at output layer	Standard

Table 4 Summary report of ANN training data set

Total sum of squared errors	Root mean square error	Average error
37887.84548	20.18406	4.336432

Table 5 Summary report of ANN validation data set

Total sum of squared errors	Root mean square error	Average error
423250.3804	82.62336	10.84254

Table 6 Regression tree parameters

Parameters	
If input variables are normalized	Yes
Minimum number of records in a terminal node	9
Maximum number of levels displayed in tree drawing	7
Do you want to draw full-grown tree?	Yes
Do you want to draw best-pruned tree?	Yes
Do you want to draw minimum-error tree?	No
Do you want to draw user-specified tree?	No

Primary investigation of the errors shows that NN model is capable to capturing the data set. It indicates that this model can be used on the bigger data set. Error is very less and most of the data set points are predicted correctly.

Regression Tree:

Following parameters of RT were used for designing regression tree (Tables 6, 7 and 8).

Good number of rules are generated with respect to validation pruned tree which will help in making decision regarding fault detection. Following are two key rules generated from regression tree:

Rule 1:

IF (sprindledriveproblem \leq 620 AND (settingtimelosstime \leq 53) AND (rework-losstime \leq 220) AND (toolsbrokenlosstime \leq 54.83) AND (inspectiontimeloss \leq 12.50) AND (nooperatotlosstime \leq 35) Then Down Time = 69.73.

Table 7 Summary report of training data using fully grown regression tree

Total sum of squared errors	Root mean square error	Average error
162489.2	41.79947	5.11896E-15

Table 8 Summary report of validation data using fully grown regression tree

Total sum of squared errors	Root mean square error	Average error
409415.2	81.26174	8.017533753

Table 9 Input data for association rule

Details of input data set	
Number of transactions in input data	155
Number of columns in input data	30
Number of items in input data	30
Number of association rules	12
Minimum support for association	16
Minimum confidence percentage	50.00%

Rule 2:

IF (sprindledriveproblem \leq 620 AND (settingtime losstime \leq 53) AND (rework- losstime \leq 220) AND (toolsbroken losstime \leq 54.83) AND (inspectiontime loss \leq 12.50) AND (nooperatof losstime $>$ 35) Then Down Time = 169.62 (With sub tree beneath).

From the above regression tree rules its can be concluded that down time can be reduced or avoided if necessary maintenance preventive measures can be taken. Other rules can also be derived and interpreted to reduce or avoid down time.

Association Rule:

Inputs:

See Table 9.

After applying association rule on the data set total twelve rules were generated. Following are the rules generated (Table 10).

Above table indicates that Consequents (C) are Servo Alarm and Insert Change Loss Time with various combinations of Antecedent (A). Prior maintenance of machine and man power training measures to be when any alarm is generated from any attribute so that machine down time can be avoided.

4 Findings and Interpretation of Results

Artificial Neural Networks—Model generated using given set of parameters, validation error and actual and predicted chart indicates that model is capable of capturing the inferences in the pilot data set (Tables 4 and 5 and Fig. 1). Hence we can conclude that it can be used on a larger dataset.

Regression Tree—Using this technique it is observed that machine is down primarily due to attributes in the order starting from root node spindle drive problem, setting time loss time, power cut loss time, rework loss time, tool broken loss time, inspection time loss, no operator loss time and insert changing loss time (Tables 7 and 8, Figs. 2 and 3). In case of reducing the unscheduled shut downs these attributes are to be controlled, maintenance and manpower is scheduled.

Table 10 Association rule

Confidence %	Antecedent (A)	Consequent (C)	Lift ratio
50.00	Tool grinding loss time	Servo alarm	1.192308
50.00	Tool grinding loss time and insert change loss time	Servo alarm	1.192308
51.61	Insert change loss time and inspection time loss time	Servo alarm	1.230769
83.78	Inspection time loss time	Insert change loss time	0.947919
88.89	Setting time loss time	Insert change loss time	1.005677
88.89	No operator loss time	Insert change loss time	1.005677
89.23	Servo alarm	Insert change loss time	1.009545
89.29	Offset given loss time	Insert change loss time	1.010167
92.00	Power cut loss time	Insert change loss time	1.040876
94.12	Inspection time loss time and servo alarm	Insert change loss time	1.064835
94.44	Tool grinding loss time	Insert change loss time	1.068532
94.44	Tool grinding loss time and servo alarm	Insert change loss time	1.068532

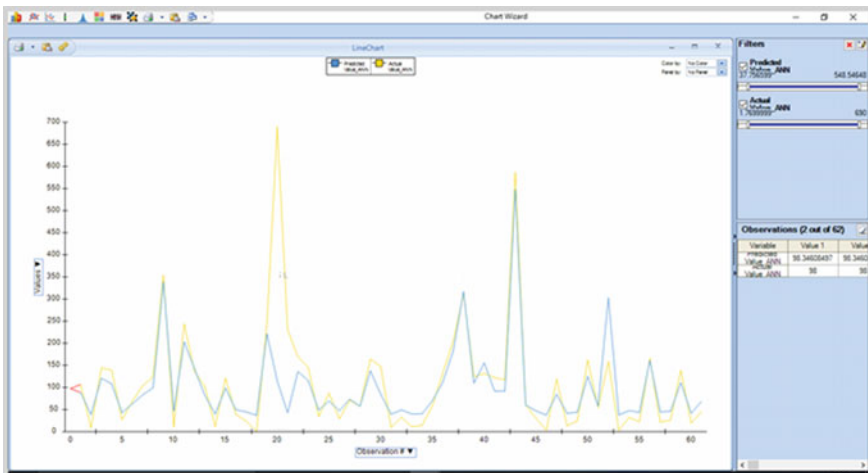


Fig. 1 Validation score actual and predicted values of total down time

(a) Best Pruned Regression Tree

(b) Fully Grown Regression Tree

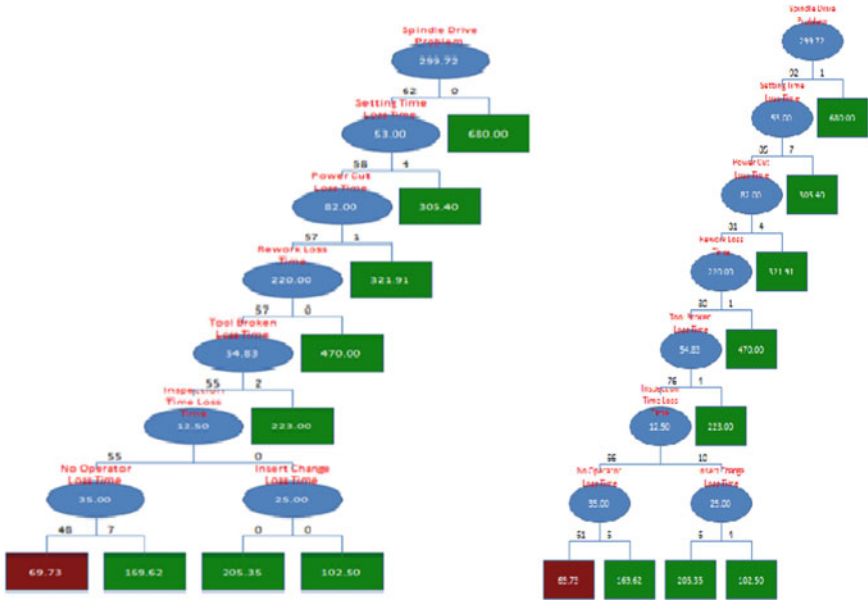


Fig. 2 a Best pruned regression tree, b fully grown regression tree

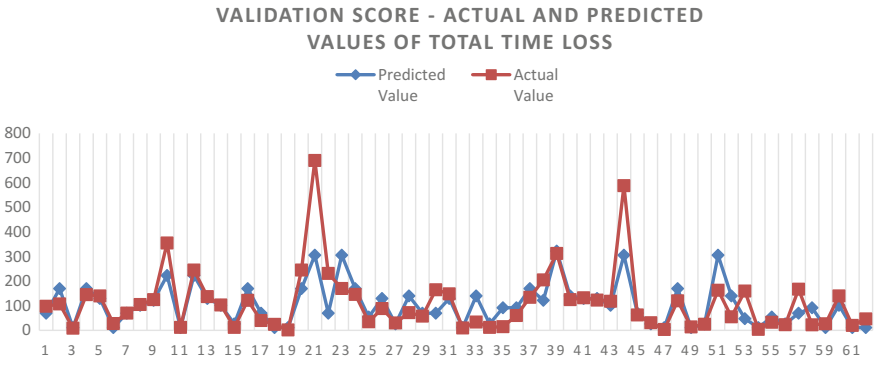


Fig. 3 Validation score—actual and predicted values of total down time using RT

Association Rule—After observing various rules generated by association rule with different antecedents servo alarm and insert change loss time are the consequents with a confidence percent between 50.00 and 94.44%. In all cases except one lift ratios are higher than 1 which indicates that rules can be accepted for decision making. This indicates that which type of loss time (error) are kept together in a basket (Table 10).

5 Conclusion

Applying various predictive techniques mentioned above it is observed that RT, NN and association rule are capable to predict and generate some meaningful results but MLR modeling technique was not able to predict due to over fitting problem and large set of unary values. It reflects that the data collected for the pilot study is not sufficient but same techniques will certainly generate good results when applied on larger data sets.

References

- Afolabi, M. O., & Olude, O. (2007). Predicting stock prices using a hybrid Kohonen self organizing map (SOM). In *Proceeding of 40th International Conference on Systems Sciences* (pp. 1560–1605).
- Dacha, K. (2007). Casual modeling of stock market prices using neural networks and multiple regression: A comparison report. *Finance India*, 21(3), 923–930.
- Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, 11, 2300–2312.
- Rimply, B. D. (1996). *Pattern recognition and neural networks*. Cambridge University Press.
- Samantha, B., & Al-Balushi, K. R. (2003). Artificial neural networks based fault diagnostics of rolling element bearings using time domain features. *Mechanical Systems and Signal Processing*, 17, 317–328.
- Sexton, R. S., & Sikander, N. A. (2001). Data mining using a genetic algorithm—Trained neural network. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 10(4), 201–210.