PREDICTING ASSEMBLY QUALITY OF COMPLEX STRUCTURES USING DATA MINING
Predicting with Decision Tree Algorithm

Ekaterina S. Ponomareva, Kesheng Wang, Terje K. Lien
Department of Production and Quality Engieering, Norwegian University of Science and Technology, NTNU,, Norway. Email; ekaterina.s.ponomareva@ntnu.no.

Abstract: Our research aims at obtaining the relevant factors that cause the decrease in quality of assembly parts. Decision tree technique was employed to induce useful information hidden within a vast collection of data. The major objective of this study was to classify the existing data into certain types of segmentations and then predict the behaviour of a ball joint assembly. The intervals of the rolling time and achieved rolling force leading to occurrence of the high moment values of the ball joint during testing stage have been found.

Key words: Prediction, machine learning, decision trees, assembly.

1. INTRODUCTION

Many of automobile parts are being assembled from different components on the assembly lines. The components had to be machined with certain degree of precision before being assembled into the final product. To maintain the desired quality of the assembly product, the methods for electronic monitoring and data acquisition at the assembly stage have to be developed. The establishment of good models for the assembly processes will permit to make reliable performance predictions from observable data.

In order to make such performance predictions data collected from production process could be helpful. Automatic knowledge acquisition techniques have been developed to address this problem. Inductive learning is an automatic technique for knowledge acquisition. The inductive approach produces a structured representation of knowledge as the outcome of

learning. Induction involves generalising a set of examples to yield a selected representation that can be expressed in terms of a set of rules, concepts or logical inferences, or a decision tree.

In this paper, decision trees have been used to extract predictive intervals of roll forming operation parameters in order to reduce the possibility of appearance of high values of friction moment after testing the ball joint assembly part.

2. PROBLEM DESCRIPTION

As a case of quality control tasks in manufacturing by the use of data mining we will consider the control arm of a front wheel suspension of an automobile system. The control arm is made of aluminium. A critical part is the ball joint that is an assembly part. It is integrated in the housing on one of the ends of the control arm. The ball joint is connected to the nut part that supports the shaft of the wheel by a bearing. The ball joint consists of 8 parts, which need to be assembled together (Fig. 1).

The manufacturing process is automated, but there are still many potential sources of failures that lead to defective products. Inappropriate dimension matching, slight variations in surface roughness or variations in degree of deformation of the roll forming joining process can lead to out of tolerance friction forces in the ball joint.
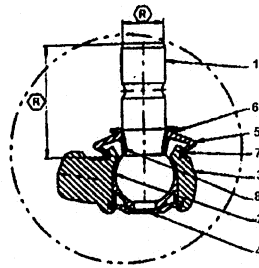

Figure 1. Ball joint assembly part: 1 – ball stud; 2 – plastic liner; 3 – housing; 4 – cap; 5 – seal; 6 – clamping ring;7 – clip ring; 8 – sleeve.

When ball stud – liner assembly is mounted in a housing **3** at the one of the ends of the control arm, the cap **4** of a special form is mounted on the top of the plastic liner **2** inside the top the housing **3**. The rim of the housing **3** is then rolled over the cap **4**, by applying force **F**, to fasten the housing. After this has been done, the test to check friction moment of the joint is carried out. In case, when friction moment has been registered higher or lower than the expected values, such parts considered being defective, the pallet is marked with special code 10, which means that these parts have a failure.

The marked pallets with those parts go through the assembling line without taking any further actions on them, until the end of the line, where these parts are discarded. The parts with satisfying results after the test are then assembled with other components: seal, clamping ring, clip ring, sleeve.

The quality control problem considered here is to reduce the possibility of appearance of high or low values of friction moment. The testing of the quality of assembled products occurs after the roll forming operation, when rim of the housing **3** is rolled over the cap **4,** by applying force **F,** to fasten the housing. Roll forming operation stage can be explained as follows: when the housing with other assembled components is ready for roll forming, rolling tool moves down until it meets the rim of the housing **3**, so the roll forming begins. Tool goes down applying force **F**, sensors register rolling time (RT) for how long tool is moving down, and achieved rolling force (RF).

If we assume that all the components of assembly part have been assembled in right order and without any flaws and defects, the only cause of appearance of high and low moment values of assembled components in testing stage can be unacceptable values of rolling time and achieved rolling force during roll forming operation.

Data-mining is one of the important techniques of information technology and known to be effective in dealing with the discovery of hidden knowledge, unexpected patterns and new rules from data (Adriaans & Zantinge, 1998). In the past few years, data-mining has also demonstrated enormous benefit in production (Milne, Drummond, & Renoux, 1998). Decision trees as one of the basic algorithms in Data Mining, can be used to build classification or regression models by recursive partitioning of data.

For data mining it is important to have good tool which combines advanced modeling technology with ease-of-use, helping to discover the interesting and valuable relationships within the data. One such tool is data mining tool *Clementine*.

3.   DECISION TREES

A decision tree algorithm begins with the entire set of data, splits the data into two or more subsets according to the values of one or more attributes, and then repeatedly splits each subset into finer subsets until the split size reaches an appropriate level. The entire modeling process can be represented in a tree structure and the model generated can be summarized into a set of 'if–then' rules. Classification trees are similar to regression models in that they have a single dependent variable and multiple independent variables. Different form regression models, they can be translated directly into sets of logical if–then rules for predicting the conditional probability distribution of

the dependent variable from the values of the independent variables. In addition to these, they can complement regression models by discovering additional useful patterns.

Ultimately, the advantages of decision trees include the following: (1) Non-parametric. (2) Easy to interpret. (3) Automatic interaction detection. (4) More informative outputs. Decision tree algorithms can be used to test conditional independence relations among variables in large data sets.

4.   DATA ACQUISITION AND PROCEDURES

Data of the roll forming operation from the control arm assembly process have been collected. The file contains fields with Rolling Time (RT) and achieved Rolling Force (RF) for each control arm after roll forming operation with corresponding values of friction moment after the quality control test. Data have been cleaned up from noises and unwanted values. In addition the test procedure has been done on the 1782 instances.

The data contain the index number for every instance, time taken to roll over the rim of the aluminium housing **3** (Fig. 1), achieved rolling force, resulting friction moment, and a status of the quality control testing. The status is identified as follows: the acceptable limitations of the friction moment for the ball joint integrated into the aluminium housing are from 2 Nm to 8 Nm, in case, if values of friction moment are inside the limitation – the Status is "Normal"; in case when values of friction moment appear to be lower than 2 Nm – the Status is "Low"; in case when values of friction moment are higher than 8 Nm – the Status is "High".

There are hundreds of thousands data collected. Since the phase of data preparation takes a long time, it was decided to use first 1782 data points as a sample data set for the first iteration of data mining analysis. These samples contain only some records of friction moment with high moment values, and do not contain records of friction moment with low friction moment values.

4.1   Decision Trees Generation

To obtain the knowledge from the data, a decision tree is employed to determine a definition for each class.

The procedure consists of two phases for achieving our objective. Phase 1 is the tree-growing stage that is constructed by repeated splits of subsets of data into two descendant subsets. The fundamental concept relies on measuring for goodness of split (i.e. to select each split so that the data in each of the subsets are purer than those in the parent node). In this stage, the entropy function is used to measure each node's impurity (Breiman,

Friedman, Olshen, & Stone, 1984). The constructed tree in Phase 1 may be too complicated. Therefore, pruning this tree to remove branches with little statistical validity is the main work in Phase 2. In the tree pruning process, the cost-complexity pruning is used to reduce the size of the tree.

The entire procedure of generating decision tree has been done using data mining tool *Clementine 9.0*. For the construction of decision tree we used algorithm C4.5. It is an advanced version of famous algorithm ID3. In particular, C4.5 uses an improved criterion for the best attribute selection and a more sophisticated method of probability estimation. The system C4.5 includes an option that turns the tree into rules that are further tuned, which can bring further improvements. The tree is presented in Figure 2.
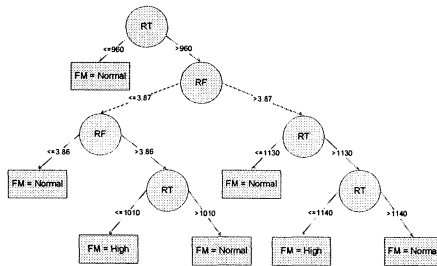

Figure 2. Decision tree for the prediction problem.

5.   RESULTS

The particular interest is paid on to the values and rules associated with high moment values (FM = High). The set of rules for appearance of high moment values is shown below:

Rule 1:   IF $960 < RT \leq 1010$ And $3.86 < RF \leq 3.87$ THEN FM = High
Rule 2:   IF $1130 < RT \leq 1140$ And $RF \leq 3.87$ THEN FM = High
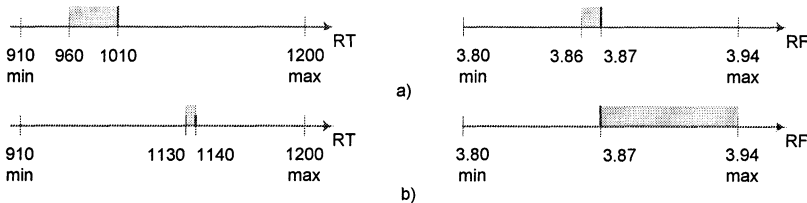Graphically these rules can be presented as displayed in Figure 3.

Figure 3. Graphical representation of the decision rules.

This is the most important result that has been extracted from decision tree. Rolling time and achieved rolling force are critical parameters, which cause the high values of friction moment of the ball joint. The main result of the decision tree induction is the identification of critical intervals for the rolling time and achieved rolling force causing the presence of high moment values.

Now it is possible to make a suggestion to avoid operation under following conditions: 1) if rolling time is in the interval (960 - 1010] and achieved rolling force is in the interval (3.86 - 3.87]; 2) if rolling time is in the interval (1130 - 1140] and achieved rolling force is in the interval [3.87 – 3.94).

6.   CONCLUSIONS

In this paper, our research shows that we can obtain the relevant factors that cause the decrease in quality of assembly parts using Data Mining. Decision trees were employed to induce useful information and knowledge hidden in a vast amount of data. We have classified the existing data into certain types of segmentations and then predict behaviour of the ball joint assembly part. From practical views, results of this study show that it is desirable to avoid operation under certain conditions in order to improve the quality of the product. The most important factor is extraction of intervals of the critical parameters.

The intervals of the rolling time and achieved rolling force leading to occurrence of the high moment values of the ball joint during the testing stage have been found, further studies to find critical intervals of the same parameters for the occurrence of the low moment values have to be done.

7.   REFERENCES

Adriaans, P., & Zantinge, D., (1998), *Data Mining*, Harlow: Addison-Wesley.

Breiman, L., Friedman, J., Olshen, R., & Stone, C., (1984), *Classification and regression trees*, Belmont, CA: Wadsworth International Group.

Chen, M.-S., Han, J., & Yu, P. S., (1996), Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866–883.

Fayyad, U., & Stolorz, P., (1997), Data mining and KDD: Promise and challenge. *Future Generation Computer Systems*, 13(2), 99–115.

Li, X. B., Sweigart, J., Teng, J., Donohue, J., & Thombs, L., (2001), A dynamic programming based pruning method for decision trees. *Journal on Computing*, 13(4), 332–344.

Milne, R., Drummond, M., & Renoux, P., (1998), Predicting paper making defect on-line using data mining, *Knowledge-Based Systems*, 11, 331– 338.