DATA MINING APPROACH FOR QUALITY PREDICTION OF INJECTION MOLDING PROCESS THROUGH STATISTICA SVM, KNN AND GC & RT TECHNIQUES

Dr. E V Ramana
Professor, Department of Mechanical Engineering,
Guru Nanak Institutions Technical Campus, Hyderabad, India

S Sapthagiri
Associate Professor, Department of Mechanical Engineering,
Guru Nanak Institutions Technical Campus, Hyderabad, India

P Srinivas
Associate Professor, Department of Mechanical Engineering,
Guru Nanak Institutions Technical Campus, Hyderabad, India

ABSTRACT

Plastic injection molding (PIM) is one of the most important production methods of manufacturing plastic products. The complexity involved in injection molding process is high due to influence of the many process variables on the quality of products. Data mining techniques are useful to uncover the hidden relationships existing in process data collected during injection molding for Quality Prediction. This paper presents a data mining approach for the quality prediction of the plastic injection molding process using Support Vector Machines (SVM), K–Nearest Neighbors (KNN) and General Classification and Regression Trees (GC&RT) techniques available in STATISTICA software. The injection molding process parameters that are considered for analysis are barrel temperatures, injection speeds and pressures at various zones of injection molding machine.

Key words: Data mining; Quality Prediction; Support Vector Machines; K-Nearest Neighbors; Injection molding process.

1. INTRODUCTION
An injection molding process is one of the commonly used methods of manufacturing plastic products. The plastic pellets are melted by the supply of heat energy and then plastic in liquid state is forced into the mold of desired shape under pressure. The plastic material is held in mold under pressure until it is completely solidified. The complexity involved in injection molding is very high due to the difficulty in manipulation of large number of process parameters influencing the quality of the product in real time. A number of approaches including Genetic Algorithms (GA), Fuzzy logic, Case Based Reasoning (CBA), Artificial Neural Networks (ANN) etc. have been tried to determine the appropriate parameter setting for injection molding [1-9].

The trial and error approach in the manipulation of enormous number of process parameters is very time consuming and turns out to be costly. An attempt has been made to build data mining models making use of Support Vector Machines (SVM), K–Nearest Neighbors (KNN) and General Classification and Regression Trees (GC&RT) techniques available in STATISTICA software on the historical injection molding process data. These models were trained on the process data set in order to predict the quality of products for a given set of process parameters.

STATISTICA SVM is basically a technique that separates cases of different class labels. It supports both regression and classification tasks capable of handling multiple continuous and categorical variables. STATISTICA KNN can be used for both classification and regression tasks. The model is defined by a set of instances (examples) with known outcomes. The outcome of new instance can be estimated by finding K-instances (examples) which are closest in distance to the new instance. The STATISTICA GC&RT shall be used in building classification and regression trees for the prediction of categorical predictor variables and continuous dependent variables. GC&RT are useful in the rapid classification of new observations and interpretation of results is simple [7].

2. DATA SET
In this research work, injection molding process data has been collected at various stages, from the machine and log books. Sample data (80 records) related to various process attributes such as barrel temperatures, injection pressures and injection speeds in five zones, cycle time, injection fill time, injection hold time, hold pressure and speed, refilling and cooling time have been collected. The barrel temperatures, injection pressures and injection speeds at five zones in injection molding machine were considered as process attributes for analysis since there is no variation in other attribute data.

3. SUPPORT VECTOR MACHINES (SVM)
The categorical independent variables (Barrel temperatures, Injection Speeds and Injection Pressures at five zones) have been given as input to SVM algorithm. The sample data has been divided into train (sample size: 60) and test (sample size: 20) samples by using a sample variable and by labeling the each record (case) as Train or Test. The training by Classification SVM of Type 1 involves the minimization of the error function. C, Capacity constant represents one of the important parameters being used to handle non-separable data inputs. The value for C has to be chosen carefully to avoid over fitting of data. The cross validation algorithm has estimated capacity constant C value as 1.

RBF (Radial Basis Function) kernel has been chosen in constructing feature space of the SVM model. The $\gamma$ parameter value of 0.666 has been used for the RBF kernel chosen [7]. Cross validation tab as shown in Figure 1 has been selected to apply cross-validation algorithm to obtain best estimates of training parameters of SVM model chosen. The maximum number of iterations is chosen as 1000 and stops at error as 0.001 in the training tab to sufficiently train the SVM model.
The number of support vectors generated by model is fifteen and cross validation accuracy of 100% is obtained for both train and test samples. The summary of results obtained from support vector machines is presented in Figure 2.

The confusion matrix showing predicted class labels of test cases is shown in Figure 3. SVM model is able to correctly predict the class labels of all test cases representing acceptable products and products rejected due to poor color distribution and short shot. The confusion matrix showing the predictions made by for overall sample has been given in Figure 4. The model predicted class labels correctly for both training and test cases.
4. K –NEAREST NEIGHBORS (KNN)

STATISTICA KNN has been used for classification task in the research work. The model was built on training cases defined with known outcomes. The outcome of new case can be estimated by finding K-neighbors which are closest in distance to this case. Euclidean metric was selected to measure the distance between new case and cases from training sample. The value for K has to be chosen with care since it can highly influence the quality of predictions [7]. The appropriate value of K is estimated by checking cross validation option in the dialog box as shown in Figure 5.

Number of nearest neighbors is estimated by cross validation algorithm as 1 and cross validation accuracy was found to be 91.66. Cross validation accuracy is declining when the value of K exceeds 1 as depicted in Figure 6. The summary of results of KNN model is presented in Figure 7.
5. GENERAL CLASSIFICATION AND REGRESSION TREES (GC&RT)

The STATISTICA GC&RT model has been applied in this work for classification task. The results can be summarized and interpreted easily in the form of trees. The tree generated by GCR&T model has been presented in Figure 8. The classification matrix showing the observed and predicted values for the class label has been displayed in Figure 9.
The importance plot of dependent variable: Class showing the degree of importance of different process variables involved in injection molding has been displayed in Figure 10.

**Figure 8** GCR&T Tree

**Figure 9** Classification Matrix – GCR&T

The importance plot of dependent variable: Class showing the degree of importance of different process variables involved in injection molding has been displayed in Figure 10.

**Figure 10** Importance Plot
The parallel coordinate plot is helpful to identify typical patterns of predictor values of a particular terminal node. Parallel coordinate plots of a node 2, node 4 and node 5 represent accepted products and defective products rejected due to poor color distribution and short shot respectively along with predictor values. Parallel coordinate plots of typical nodes 2 and 4 are displayed in Figure 11 and Figure 12 respectively.

![Figure 11 Parallel Coordinate plot in node 2](image)

![Figure 12 Parallel Coordinate plot in node 4](image)

### 6. CONCLUSION

The data mining models that were built on the data set shall be used in predicting acceptable products and products rejected due to short shot and poor color distribution. Ten fold cross validation method has been used to evaluate the prediction accuracies of SVM and KNN models. The prediction accuracies of SVM and KNN models were found to be 100% and 91.66% respectively implies that the SVM model is more reliable. Barrel temperatures at zone 1 to zone 5, injection speed at zone 2 and zone 3 and injection pressure at zone 2 were found to be more important in predicting the class label as given by GCR&T Tree. Parallel coordinate plots along with their associated range of predictor values are helpful in decision making.
Data Mining Approach for Quality Prediction of Injection Molding Process Through Statistica SVM, KNN and GC & RT Techniques

making during the process parameter setting in order to manufacture products with acceptable quality and avoiding defective ones. All models predicted the class labels of all three cases mentioned above correctly with overall accuracy of 100% as shown in Figure 13.

![Figure 13](image)

**Figure 13** Color maps of predicted class frequencies relative to observed class frequency for class

ACKNOWLEDGEMENT

The Principal Investigator gratefully acknowledges UGC/SERO for the financial support provided to this research work under Minor Research Project (No. 6216/15) entitled “Comparative Study of Data Mining Tools for Knowledge Extraction, Process Control and Quality Improvement”.

REFERENCE


