Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 00 (2019) 000–000

www.elsevier.com/locate/procedia

CIRP Manufacturing Systems Conference 2019

Application of feature selection methods for defining critical parameters in thermoplastics injection molding

Olga Ogorodnyk[a]*, Ole Vidar Lyngstad[b], Mats Larsen[c], Kristian Martinsen[a]

[a]*Department of Manufacturing and Civil Engineering, Norwegian University of Science and Technology, Teknologivegen 22, Gjøvik, 2815, Norway*
[b]*Department of Materials Technology, SINTEF Manufactring, Enggata 40, 2830, Raufoss, Norway*
[c]*Department of Production Technology, SINTEF Manufactring, Enggata 40, 2830, Raufoss, Norway*

* Corresponding author. Tel.: +47-48630583. *E-mail address:* olga.ogorodnyk@ntnu.no

Abstract

Thermoplastics injection molding is a manufacturing process used for mass-production of plastic parts. The process includes four main stages during which material used goes through complicated thermo-mechanical changes. In order to make the process more controllable and repeatable it is, at first, necessary to understand which parameters are the most important ones. The following paper describes how application of statistical feature selection methods, such as Information gain and ReliefF, allows to identify which injection molding parameters have a greater influence on the final part quality. The article gives short description of the above-mentioned methods and shows what were results of their application on dataset obtained from 160 machine runs, during which 41 machine and process parameters were logged.

Keywords: Feature selection; Important injection molding parameters; Injection molding; Information Gain; ReliefF

1. Introduction

Injection molding is an important process for mass production of different kinds of plastic parts often with both complex geometry and high precision [1]. The process includes four main stages: plasticization, injection, cooling and ejection [2]. At first, plastic pellets are melted with help of a reciprocating screw, afterwards plastic melt is injected into a mold with cavity in form of a produced part with help of injection pressure, next plastic melt cools down and solidifies inside of the mold until it is opened and the part is taken out.

To achieve high process repeatability and quality of final products, it is important to set proper values of machine and process parameters in the beginning of the production process [3, 4]. Often these parameters are identified using trial and error method, which is time and resources consuming [3]. Instead, application of intelligent methods, such as Machine Learning (ML), for determination of the process parameter values, as well as process monitoring and control has been proven useful [5]. There are examples of prediction models built with help of Artificial Neural Networks (ANN) to predict shrinkage [6], flash [7], weight and length [8] and to classify final parts quality [5, 9]. In addition, Genetic Algorithm (GA) was used to minimize warpage [10], Decision Trees (DT) to predict parts quality [9], hybrid ANN/GA to minimize warpage [10] and to construct an inverse prediction model of injection molding process [11]. However, to apply these methods, it is important to have significant amount of data and to understand which parameters should and should not be included into a model.

Feature selection has been proven an effective strategy for preprocessing and preparing data for various machine learning problems [12]. *"Often, pre-processing of the datasets takes place for two main reasons: 1) reduction of the size of the dataset in order to achieve more efficient analysis, and 2) adaptation of the dataset to best suit the selected analysis method"* [13]. Various studies show that some of the features or parameters can be removed without significant decrease of model's quality [14, 15]. Moreover, if a feature is redundant removing it might lead to increase of the model's quality.

Feature selection has been an active field of research for decades and has been widely applied to many fields such as bioinformatics, text mining, intrusion detection and industrial applications to name a few [13, 14]. For example, Shao, Paynabar [16] are proposing to use *k*-folds cross validation to select important features important for specific manufacturing processes and tune parameters for quality monitoring. In [17] selection of parameters for prediction model of manufacturing lead times is performed using feature selection. Tirkel [18], on the other hand, used both manual (human experts) and statistical feature selection procedures in order to choose the most relevant parameters for cycle time prediction model used in water fabrication. Verron, Tiplica [19] present a method based on mutual information for selection of the most important features used in the fault diagnosis model. The following paper, in its turn, describes a process of applying Information Gain (InfoGain) and ReliefF feature selection algorithms to a dataset, which includes data from 160 injection molding machine runs with varied process parameters, where each cycle 41 machine and process parameters were logged. Feature selection, in this case, is applied to decrease number of injection molding parameters included in parts quality classification model.

2. Methodology

The next steps are proposed to be followed in order to select the most relevant parameters to be included in injection molded part quality prediction model:

- *Data acquisition*. Logging of data from sensors installed in the injection molding machine by its manufacturer.
- *Data preprocessing*. Necessary preprocessing, that might include normalization of data or filtering of outliers.
- *Application of feature selection algorithm*. Use of chosen feature selection algorithm to score previously acquired machine and process parameters to define which of them contain more information about the process in general.
- *Review results of the algorithm's work*. Since all of the parameters, which are logged from the injection molding machine are understandable for a human expert, it is suggested to look through results of the feature selection algorithm to confirm its results.

2.1. Experiment

All of the data used in this study was acquired from sensors installed by the machine manufacturer in "ENGEL insert 130" vertical injection molding machine with CC300 control unit, no additional sensors were used. Focus part was an ISO 527-2 type 1A dog bone specimen with 170 mm length, 20 mm width and 4 mm thickness shown on Fig. 1, material used was high-density polyethylene.

Design of experiment (DOE) for gathering of parameter data was created with help of Latin Hypercube method in ModeFRONTIER [20]. DOE consists of 32 different combinations of the following parameters: holding pressure, holding pressure time, backpressure, cooling time, injection speed, screw speed, barrel temperature and temperature of the mold. Each combination was launched 5 times, resulting in 160 machine runs and 160 corresponding data samples. As suggested by Scikit-Learn (open source Python programming language library for ML methods) "Machine Learning Map" [21], it is not recommended to use any of the machine learning methods if the amount of data points per parameter, included into prediction model, is less than 50. However, described case uses number of samples that is in 3.2 times bigger than that. Out of these machine runs, 101 were classified as the ones resulting in low quality parts and 59 in high quality ones. Classification of quality was based on visual inspection of the specimens and their weight (if weight was less than specified value, the part was classified as low quality).

In total 41 machine and process parameters were logged during each machine run. In addition to 41 logged, parameters from the DOE were added resulting in 47 different features in total. List of the parameters is provided in Table 1. After application of feature selection methods, classification models for further prediction of such specimens' quality were created using Artificial Neural Networks and Decision Trees methods. More information about this study can be found in [9].

Fig. 1. Dog bone specimen.

Before applying ML, feature selection methods, namely Information Gain and ReliefF, were used to understand which parameters out of 41 logged are statistically more significant, contain more information about the process and are more relevant to be included in the quality of injection molded parts classification model. These methods were chosen as they often show good results when used for further building of classification models [13, 22].

2.2. Information Gain

Information Gain is considered a standard attribute/parameter quality measure used in classification problems. It is defined as amount of information, obtained from a parameter for determining class in a classification problem [23]. In other words, InfoGain measures how much information certain parameter can provide about a class [24]. In the context of decision trees it is also sometimes called mutual information.

To calculate information gain, one must be familiar with a concept of entropy, which is defined as the average amount of information necessary to determine the outcome among m possible outcomes [23]. Another definition is amount of uncertainty present in distribution of events of a random variable X [25]. As a result, if an event is likely to happen, then entropy is low, as there is little uncertainty. To calculate entropy equation (1) can be used, where $P(X_j)$ is probability of outcome X_j:

$$H(X) = -\sum_{j=1}^{m} P(X_j) log_2 P(X_j) \qquad (1)$$

Information gain, on the other hand can be calculated using (2), where H_C is class entropy and $H_{C|A}$ is a conditional class entropy given the value of attribute/parameter A.

$$InfoGain(A) = H_C - H_{C|A} \quad (2)$$

At the same time, $InfoGain(A) \geq 0$ and $\max(InfoGain(A)) = H_C$. In general, information gain is a useful measure for deciding whether parameter is relevant to be included in a classification model, as it provides numerical value of amount of information shared between two random variables or gained about one random variable by watching another one [12]. However, it also has its drawbacks. If a parameter or an attribute used takes large number of different values, this attribute has a high mutual information. However, in case of classification problems such an attribute is not relevant, as it uniquely identifies each separate data sample and is unlikely to provide necessary generalization. In addition, inclusion of such parameter into a model might lead to overfitting. This is why it is a good idea to allow a human expert to review results of the algorithm's work before proceeding to a classification model creation.

When large number of such attributes is present, it is recommended to use Information Gain ratio instead, that can be calculated using (3), where H_A is entropy of an attribute A:

$$GainR(A) = \frac{Gain(A)}{H_A} \quad (3)$$

2.3. ReliefF

Feature selection measures such as Information Gain, Gini index, X^2 statistics, distance measure, etc. *"evaluate quality of an attribute independently of the context of other attributes"* [23]. These measures are called myopic measures, as they are unable to detect importance of attributes if they have high level of interdependency.

ReliefF, as well as its simpler variant Relief, on the other hand, take context of other attributes into account and are able to detect relevance of attributes even if they have a strong mutual dependence. Basic idea of Relief and ReliefF algorithms is hidden in considering not only difference in parameters' values and classes, but also distance between examples or data samples. *"Distance is calculated in the attribute space, therefore similar examples are close to each other and dissimilar are far apart"* [23]. If similarity of data samples is taken into account, the context of other parameters is also considered.

Relief algorithm searches for the nearest example from the same class and a nearest example from the opposite class, while ReliefF searches for k nearest examples from each class for each data sample from a random subset of data samples, and then weights contributions of different classes with their prior probabilities [23]. The next step is updating quality of each of parameters or attributes based on ability of an attribute to distinguish examples from different classes. When scoring the parameters, *diff* function is used, equation for calculating difference of attribute values for two examples is shown in (4) for both discrete and continuous parameter values.

$$diff(i, t_j, t_k) = \begin{cases} \frac{|v_{i,j} - v_{i,k}|}{Max_i - Min_i}, A_i \text{ is continuous} \\ 0, v_{i,j} = v_{i,k} \text{ and } A_i \text{ is discrete} \\ 1, v_{i,j} \neq v_{i,k} \text{ and } A_i \text{ is discrete} \end{cases} \quad (4)$$

Unlike Relief, ReliefF is able to deal with missing and incomplete data with help of generalization of $diff$ function for cases when one or both examples have unknown value of attribute A_i. In addition, looking for k nearest examples for each class instead of one, allows using ReliefF for multi-class problems. $ReliefF_{score} \in [-1; 1]$, however, values that are less than zero mean that the attribute is irrelevant.

3. Discussion and results

Information Gain and ReliefF are very different feature selection methods. The first one is based on entropy, does not take into consideration interconnections between attributes and might fail to generalize if an attribute used takes large number of different values. The second method, on the other hand, is based on distance between data samples from different classes and thanks to this, considers mutual dependence between parameters if it is present and is able to deal with attributes that have many different values.

Both methods were applied to 160 data samples with 47 machine and process parameters acquired during the above described experiment. To apply the methods WEKA software (Waikato Environment for Knowledge Analysis) developed by Waikato University of New Zealand was used [26]. Corresponding attribute scores and ranks assigned to parameters by InfoGain and ReliefF feature selection algorithms can be seen in Table 1.

Table 1. InfoGain and ReliefF scores for parameters from the study.

Parameter name	Info Gain rank	InfoGain score	ReliefF rank	ReliefF score
Parts counter	1	0.6694	16	0.0779
Shot counter	2	0.6694	18	0.0775
Counter of good parts	3	0.6694	17	0.0779
Cushion size after holding pressure	4	0.5685	2	0.2266
Smallest size of cushion	5	0.556	3	0.2242
Average cushion value	6	0.519	4	0.2113
Holding pressure (set)	7	0.4778	1	0.2579
Average temperature in zone 2 of nozzle	8	0.3812	11	0.0932
Barrel temperature (set)	9	0.2735	10	0.0939
Injection time	10	0.2312	22	0.047
Plasticizing number	11	0.2266	20	0.0624
Screw speed (set)	12	0.1999	9	0.0996
Maximum screw speed	13	0.1881	15	0.0793
Specific pressure value at switchover	14	0.1731	21	0.0531
Plasticizing time	15	0.1263	29	0.0034
Maximum Speed	16	0.121	7	0.1117
Tool temperature (set)	17	0.0663	5	0.2039
Injection work	18	0.0663	24	0.0445

Holding pressure time (set)	19	0.0495	8	0.1042
Injection speed (set)	20	0.0432	14	0.0818
Machine date	21	0.016	13	0.08625
Current station (set)	22	0	45	0
Waiting delay (set)	23	0	31	0
Last cycle time	24	0	27	0.0123
Shot volume (set)	25	0	38	0
Decomposition after plasticization time (set)	26	0	30	0
Switchover volume (set)	27	0	47	-0.0044
Switchover time (set)	28	0	44	0
Backpressure (set)	29	0	6	0.1372
Maximum injection time (set)	30	0	35	0
Maximum ejector position (set)	31	0	34	0
Cooling time (set)	32	0	12	0.0877
Maximum cushion smallest (set)	33	0	39	0
Last ejector position	34	0	26	0.0139
Maximum plasticizing time (set)	35	0	43	0
Flow number (set)	36	0	42	0
Last cooling time	37	0	19	0.0673
Clamping force at switchover	38	0	28	0.0091
Counter of bad parts	39	0	23	0.0463
Plasticizing delay time (set)	40	0	32	0
Minimum injection time (set)	41	0	37	0
Closing force	42	0	25	0.015
Ideal cushion value (set)	43	0	33	0
Injection pressure limit (set)	44	0	36	0
Minimum ejector position (set)	45	0	41	0
Minimum plasticizing time (set)	46	0	40	0
Machine time	47	0	46	0

As it is possible to see from Table 1 there are 26 parameters that have zero information gain scores and 16 parameters that have zero or negative ReliefF scores. These differences might occur due to above-mentioned differences in the algorithms. ReliefF, for example, gives a non-zero score to “last cycle time”, “backpressure (set)”, “cooling time (set)”, “last ejector position”, “closing force”, “clamping force at switchover”, “counter of bad parts” and “last cooling time”, while these eight parameters have zero information gain scores. Apart from this, both methods show similar results and give non-zero scores to similar set of machine and process parameters. It is also important to mention that some of parameters (for example, “shot volume”) that have received zero score from both algorithms were remained unchanged during the experiment run, and thus were considered irrelevant for classification model by the feature selection algorithms. Due to this, it is worth noting that depending on data that is processed and parameters that are varied, feature selection methods might give higher or lower score to certain parameters.

Results of the InfoGain application were used in [9] and it has been shown that models created by both ANN and DT *“show increase in accuracy after removing features that do not contain much information about the process”*.

At first, models based on results of the InfoGain were built. 26 features that have zero InfoGain score were removed, as well as “machine time”, “shot counter”, “counter of good parts”, “counter of bad parts”, “parts counter” and “machine date” due to their irrelevance. Accuracy of classification models trained with help of J48 decision trees algorithm and Multilayer Perceptron Artificial Neural Network increased from 96,25% (47 features, ANN) and 95,875% (47 features, DT) to 99,375% (17 features, ANN) and 98,75% (17 features, DT).

Secondly, models considering results of the ReliefF were created. Here, 18 attributes with zero score were removed, as well as parameters identified by a human expert as irrelevant. This resulted in parameter list consisting of 24 parameters that are highlighted in bold in Table 1. J48 decision trees algorithm resulted in 98,75% accuracy (24 features), while ANN had 97,5% (24 features). As it is easy to see, use of feature selection leads to increase in prediction accuracy of models in both cases. It is also important to mention, that in all models values of ROC area and F-measure do not show presence of overfitting.

Some parameters that have received the highest scores from information gain algorithm, excluding those identified as irrelevant by a human expert, are: “cushion size after holding pressure”, “smallest size of cushion”, “average cushion value”, “holding pressure (set)”, “average temperature in zone 2 of nozzle” and “barrel temperature (set)”. ReliefF gives some of the highest scores to the following features: “cushion size after holding pressure”, “smallest size of cushion”, “average cushion value”, “holding pressure (set)”, “tool temperature (set)”, and “backpressure (set)”.

4. Conclusions and future work

This study described application of two feature selection methods, namely, Information Gain and ReliefF. The algorithms were used in order to identify the most relevant injection molding parameters out of 47 logged and included into DOE for their further use in parts quality classification model. The dataset is based on 160 injection molding machine runs, where DOE with 32 combinations of process parameters was used to conduct the experiment and acquire process data. Each of these 32 parameter combinations was launched five times.

InfoGain method results included 26 parameters with zero score, while ReliefF algorithm gave 18 parameters zero or negative score, identifying them as irrelevant. Distinctive results can be explained by differences in the above-mentioned algorithms and quality measures they use to score the assessed parameters. It is also worth mentioning that depending on which parameters are varied during experiment, the algorithms might give higher or lower scores to certain features. Parameters that have received some of the highest scores from both methods are: “cushion size after holding pressure”, “smallest size of cushion”, “average cushion value”, “holding pressure (set)”, “average temperature in zone 2 of nozzle”, "barrel temperature (set)", "tool temperature (set)" and "backpressure (set)".

Based on the InfoGain and ReliefF algorithm results, it has been confirmed that removing features with zero scores increases quality of classification models built by Artificial Neural Networks and Decision Trees methods.

Acknowledgements

This research is funded by Norwegian Research Council as part of the "MegaMould" project.

References

[1] Zhou X, Zhang Y, Mao T, Ruan Y, Gao H, Zhou H. Feature extraction and physical interpretation of melt pressure during injection molding process. 2018.
[2] Ogorodnyk O, Martinsen K. Monitoring and Control for Thermoplastics Injection Molding A Review. Procedia CIRP. 2018;67:380-5.
[3] Kitayama S, Yokoyama M, Takano M, Aiba S. Multi-objective optimization of variable packing pressure profile and process parameters in plastic injection molding for minimizing warpage and cycle time. The International Journal of Advanced Manufacturing Technology. 2017;92(9-12):3991-9.
[4] Gao H, Zhang Y, Zhou X, Li D. Intelligent methods for the process parameter determination of plastic injection molding. 2018:1-11.
[5] Charest M, Finn R, Dubay R, editors. Integration of artificial intelligence in an injection molding process for on-line process parameter adjustment. Systems Conference (SysCon), 2018 Annual IEEE International; 2018: IEEE.
[6] Altan M. Reducing shrinkage in injection moldings via the Taguchi, ANOVA and neural network methods. Materials & Design. 2010;31(1):599-604.
[7] Zhu J, Chen JC. Fuzzy neural network-based in-process mixed material-caused flash prediction (FNN-IPMFP) in injection molding operations. The International Journal of Advanced Manufacturing Technology. 2006;29(3-4):308-16.
[8] Chen W-C, Fu G-L, Tai P-H, Deng W-JJESwA. Process parameter optimization for MIMO plastic injection molding via soft computing. 2009;36(2):1114-22.
[9] Ogorodnyk O, Lyngstad OV, Larsen M, Wang K, Martinsen K, editors. Application of Machine Learning Methods for Prediction of Parts Quality in Thermoplastics Injection Molding. International Workshop of Advanced Manufacturing and Automation; 2018: Springer.
[10] Yin F, Mao H, Hua L. A hybrid of back propagation neural network and genetic algorithm for optimization of injection molding process parameters. Materials & Design. 2011;32(6):3457-64.
[11] Tsai K-M, Luo H-J. An inverse model for injection molding of optical lens using artificial neural network coupled with genetic algorithm. Journal of Intelligent Manufacturing. 2017;28(2):473-87.
[12] Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: A data perspective. 2018;50(6):94.
[13] Jović A, Brkić K, Bogunović N, editors. A review of feature selection methods with applications. Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on; 2015: IEEE.
[14] Liu H, Motoda H, Setiono R, Zhao Z, editors. Feature selection: An ever evolving frontier in data mining. Feature Selection in Data Mining; 2010.
[15] Banin S, Dyrkolbotn GO. Multinomial malware classification via low-level features. Digital Investigation. 2018;26:S107-S17.
[16] Shao C, Paynabar K, Kim TH, Jin JJ, Hu SJ, Spicer JP, et al. Feature selection for manufacturing process monitoring using cross-validation. 2013;32(4):550-5.
[17] Pfeiffer A, Gyulai D, Kádár B, Monostori LJPC. Manufacturing lead time estimation with the combination of simulation and statistical learning methods. 2016;41:75-80.
[18] Tirkel I, editor Cycle time prediction in wafer fabrication line by applying data mining methods. 2011 IEEE/SEMI Advanced Semiconductor Manufacturing Conference; 2011: IEEE.
[19] Verron S, Tiplica T, Kobi AJEaoai. Fault diagnosis of industrial systems by conditional Gaussian network including a distance rejection criterion. 2010;23(7):1229-35.
[20] ARBURG. Host computer system (ALS) 2018 [Available from: https://www.arburg.com/en/products-and-services/injection-moulding/production-management/host-computer-system-als/.
[21] Scikit-Learn. Choosing the right estimator [Available from: http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html.
[22] Abusamra HJPCS. A comparative study of feature selection and classification methods for gene expression data of glioma. 2013;23:5-14.
[23] Kononenko I, Kukar M. Machine learning and data mining: introduction to principles and algorithms: Horwood Publishing; 2007.
[24] Haralampieva V, Brown G. Evaluation of Mutual information versus Gini index for stable feature selection. 2016.
[25] Shannon CE. A mathematical theory of communication. Bell system technical journal. 1948;27(3):379-423.
[26] Zhou X, Zhang Y, Mao T, Zhou HJJoMPT. Monitoring and dynamic control of quality stability for injection molding process. 2017;249:358-66.