Comparison of feature selection methods for machine learning based injection molding quality prediction

Cite as: AIP Conference Proceedings **2289**, 020052 (2020); https://doi.org/10.1063/5.0028546
Published Online: 30 November 2020

Alexander Schulze Struchtrup, Dimitri Kvaktun and Reinhard Schiffers

Comparison of Feature Selection Methods for Machine Learning based Injection Molding Quality Prediction

Alexander Schulze Struchtrup[1, a)], Dimitri Kvaktun[1] and Reinhard Schiffers[1]

[1] University of Duisburg-Essen, Institute of Product Engineering, Lotharstr. 1, 47057 Duisburg, Germany

[a)] Corresponding author: alexander.schulze-struchtrup@uni-due.de

Abstract. Predicting the quality of injection molded parts based on process data has been dealt with in research for some time. However, existing approaches did not prevail in the industry to date due to lacking integrity. An important step in building a successful quality model using machine learning methods is the selection of suitable process features. This paper describes the various methods and steps for feature selection, including filters, wrappers and embedded methods, and compares the resulting feature sets with regard to the achievable goodness of the quality models that were created using seven different supervised machine learning algorithms for regression.

INTRODUCTION

Thermoplastics injection molding is one of the most relevant processes in the field of plastics processing. It is a discontinuous process that allows the automatic, highly reproducible production of molded parts with complex geometry [1]. Although injection molding machines have been improved with regard to both, mechanical precision and control techniques over the last couple of decades [2], internal and external perturbations may still negatively affect the quality of the molded parts.

Therefore, quality assurance plays an important role plastics in processing companies. Despite the efforts taken, scrap production is usually detected only with delay in a sample based quality inspection, while bad parts produced between two samples may stay completely unnoticed. To overcome these drawbacks, research efforts have been made, to predict the quality of the molded parts directly from machine and process data using machine learning algorithms, e.g. [3-6]. Although the named approaches yielded relatively good results, they did not prevail in industry to date, although corresponding products are available. To some extent this is due to the fact that the process of robustly building a good quality model requires many steps, including data generation, selection and pre-processing, feature extraction, construction and selection as well as learning and adapting suitable models including hyper-parameter optimization, which, in the past approaches had to be carried out mainly manually and resulted in a lot of effort.

One of the most important steps is the process of feature selection, since it affects the achievable quality of the quality prediction model. The purpose of this paper is to provide an overview over the most relevant methods for feature selection and evaluate their performance applied to thermoplastics injection molding data.

FUNDAMENTALS OF FEATURE SELECTION FOR SUPERVISED LEARNING

The field of machine learning can be divided into supervised, unsupervised and reinforcement learning. Predicting the quality of injection molded parts is a task that requires supervised learning methods. These are learning a model of the relationships between input variables (process data) and output variables (quality data). If we are dealing with continuous quality features, e.g. a part's weight, we are dealing with regression, while in the case of discrete features, e.g. part ok / not ok, we are dealing with classification. In this paper, we use the following seven supervised learning machines to the build regression models used for evaluating the feature selection methods: multiple linear regression,

Proceedings of PPS2019 Europe-Africa Regional Conference of the Polymer Processing Society
AIP Conf. Proc. 2289, 020052-1–020052-5; https://doi.org/10.1063/5.0028546
Published by AIP Publishing. 978-0-7354-4019-7/$30.00

artificial neural networks, support vector machines, Gaussian-process regression, k-nearest neighbors, binary decision tress (CART) and ensembles (bagging and boosting) of binary decision trees.

Most of the learning machines incorporate one or more hyper-parameters controlling the learning process and model complexity. The choice of theses hyper-parameters can significantly affect the resulting model's quality [7]. In our case we use Bayes-optimization for automatic hyper-parameter optimization. As already mentioned, also the features selected for learning have a significant impact on model quality. While it is generally possible to select features for quality prediction manually, the quality of the outcome is strongly dependent on individual experience. To lower the selection effort and to guarantee a reasonable quality of the selected feature set, algorithms for automatic feature selection have been developed. These may be divided into the following three groups [8]: filter methods, wrapper methods and embedded methods.

The three categories mainly differ with regard to the question whether they require and in which way they use a learning machine for the feature selection process: Filter methods are heuristic approaches using a performance evaluation metric in combination with a relevance criterion for feature (set) evaluation, therefore no learning algorithm is needed. In contrast, wrapper methods use the learning machine(s) as a black box and select features based on the resulting model quality (goodness of fit). Embedded methods also use the learning machine(s) for feature selection, however the selection process is using a selection criterion specific to the learning algorithm and therefore cannot be transferred to or used with other learning machines. Since it is in most cases computationally infeasible to evaluate all possible feature sets, all of these methods are used in interaction with a search strategy that decides which feature or combination of features is to be evaluated.

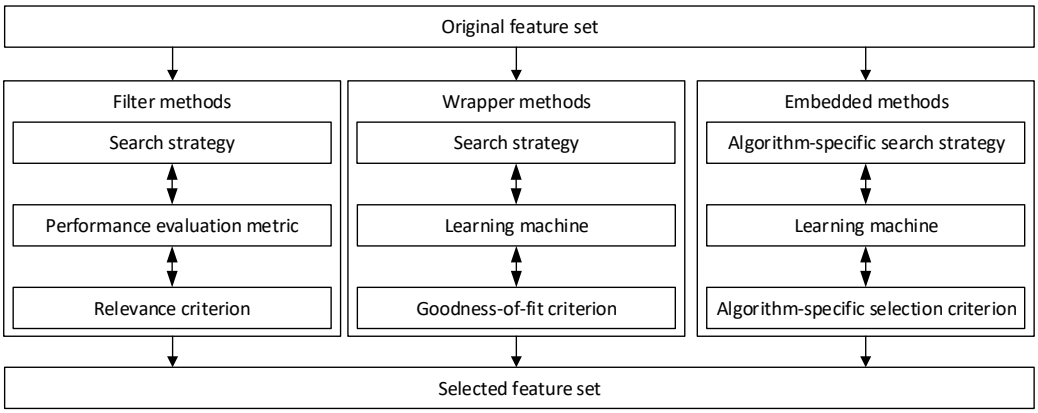

FIGURE 1. General structure of feature selection using filter, wrapper and embedded methods

FEATURE SELECTION ALGORITHMS

Search Strategies

To identify the best combination of features for a task like quality prediction, one could think of testing each possible combination, which is referred to as exhaustive search [8]. Given the usually available number of features to choose from (e.g. provided by the machine's actual values), and the fact that a number of n features yields $2^n - 1$ combinations, it becomes obvious, that this strategy quickly turns out computationally infeasible. Therefore, so-called greedy search strategies have been developed, which lower the computational effort while still yielding good results. The following search strategies are applied in this paper: sequential forward selection, sequential backward selection, sequential-floating forward selection, sequential-floating backwards selection and genetic algorithm.

In the sequential forward selection, the algorithm starts with an empty feature set. In the first step, each feature's relevance is evaluated with regard to the quality criterion and the highest rated feature is added to the feature set. In the subsequent step, the performance evaluation metric (cf. chapter below) for each of the remaining features in combination with the already chosen feature is calculated and the feature is chosen, which yields the best performance. This procedure continues until a predefined number of features (in this paper we use 10) is chosen. The procedure of the sequential backward selection is similar to those of the sequential forward selection, however the sequence is inverted. Despite their computational efficiency, there may be cases, when the named sequential methods suffer from

the so-called "nesting effect", meaning that features which have once been selected cannot be deselected later on. This issue can be addressed using sequential floating forward and backward selection methods respectively. The sequential floating forward selection algorithm is similar to the non-floating one, however after each executed forward step it includes a backward selection which deselects features in case this improves the feature set's performance evaluation metric. Correspondingly, the sequential floating backward selection includes a forward selection after each backward step. The previously introduced feature selection algorithms have in common that they are deterministic, so that when applied multiple times to the same data, they will provide the same results. Another approach is to use a genetic algorithm for feature selection. Here, the information whether or not a feature is part of a certain feature set is encoded in a vector containing only ones (part of current feature set) and zeros (not part of current feature set). In the beginning, a set of vectors is randomly created which are then altered using the concepts of selection, mutation and crossover. This way, the genetic algorithm works like an optimization. Due to relatively low computational effort, the genetic algorithm works well especially for larger feature numbers [9].

Performance Evaluation Metrics

Filter methods use a performance evaluation metric (PEM) to evaluate and compare the fitness of features or feature sets. This usually means taking into account both the relevance of a feature with regard to the quality criterion and the redundancy among the features. Obviously, we are looking for features with have a high relevance for the quality criterion and a low redundancy to the features already chosen. Two metrics taking both aspects into account are *correlation based feature selection* (CFS) [10] (cf. eq. 1) and *minimum redundancy – maximum relevance* (MRMR) [11] (cf. eq. 2):

$$M_{CFS} = \frac{k \cdot \overline{r_{cf}}}{\sqrt{k + k \cdot (k-1) \cdot \overline{r_{ff}}}} \qquad (1)$$

$$M_{MRMR} = \frac{|S| \cdot \sum_{t \epsilon S} I(\boldsymbol{x}_t, \boldsymbol{y})}{\sum_{s,t \epsilon S} I(\boldsymbol{x}_t, \boldsymbol{x}_s)} \qquad (2)$$

In the equations above, M_{CFS} and M_{MRMR} are the CFS- and MRMR-metrics respectively. k is the number of features included in the current feature set, while $\overline{r_{cf}}$ and $\overline{r_{ff}}$ represent the average feature to quality criterion and feature to feature relevance. $|S|$ indicates the overall number of features, $\boldsymbol{x}_t$ and $\boldsymbol{x}_s$ are the feature vectors of features t and s, and $\boldsymbol{y}$ is the quality criterion's vector.

Relevance criteria

The introduced performance evaluation metrics make use of a relevance criterion that quantifies the dependence between two features. As already mentioned, relevance criteria are not only applied to calculate the relevance of process features regarding the quality feature but also to measure redundancy to the other process features. In this paper we are using three different relevance criteria, which we explain in the following: Pearson's correlation, Spearman's rank correlation and information gain.

Pearson's correlation coefficient is probably the most well-known way to measures a linear correlation between two variables. However, since the injection molding process includes also nonlinear interactions, Pearson's correlation might be a suboptimal choice to measure the dependence between process and quality features. Spearman's rank correlation coefficient solves this issue to some extent as it does not correlate the feature values themselves but their rank, so also nonlinear monotonic dependencies are covered. A somewhat different approach to measure relevance comes from the field of information theory. The information gain measures how much information (negative of entropy) is gained when a certain feature is added to a feature set [12]. Since this criterion uses the features probability distributions, continuous features (as present in regression) have to be discretized in beforehand.

EXPERIMENTS

For the feature selection and quality prediction both process and quality data are generated using a *KraussMaffei PX120-380* fully-electric injection molding machine. The part is a plate specimen mold with two cavity pressure

sensors. We use the part weight as quality feature, which is weighed with a *Sartorius Entris 153I-1S* lab balance, upon which they are automatically placed by the IMM's linear robot *LRX50*. The available process features are directly taken from the machines actual value protocol. For data generation, six different process states are induced, all of them earlier or later occurring in a real-world injection molding production: start-up, stable process, downtimes, regrind material, regrind material + adaptive process control (APC) and DOE (central composite design). At each process state, 1000 injection molding cycles are carried out, except for the DOE with 860 cycles due to its predefined structure.

Before the actual quality model generation, the feature sets have to be determined. To achieve this, each of the five introduced search strategies is combined with the two performance evaluation metrics using each of the three relevance criteria. This results in a total of thirty feature selection combinations. For building quality prediction models, the seven supervised learning algorithms named above are applied including Bayes-hyper-parameter-optimization. To evaluate the models' performances, they have to be validated on data which have not been used in the learning phase. In this paper, we use five-fold cross validation. In every fold, 80 % of the data is used for learning and 20 % for validation, so taking together all folds, each injection molding cycles is used exactly once for validation and four times for learning. The average coefficient of determination achieved in the validations is used as goodness-of-fit criterion.

RESULTS

Figure 2 shows the coefficients of determination of the models learnt using the thirty combinations of search strategy, performance evaluation metric and relevance criterion for feature selection. Each bar represents the average value obtained in 42 combinations of process states and learning algorithms. The displayed error bars indicate the standard deviations of the results from these combinations. Analyzing the results, it is of interest, how the chosen search strategy, performance evaluation metric and relevance criterion as well as their interactions affect the goodness of the learnt models. In general, a rather mediocre level of coefficients of determination can be observed, which is due to averaging all datasets and learning algorithms, since also those yielding poor results are included.

Regarding the search strategy, sequential forward and sequential-floating forward selection yield better results than their backward pendants. Comparing the two forward selection methods, the differences between the models performances are negligible, while the sequential-floating backward selection performs a little bit better than the simple backward version. This might be due to a small nesting effect occurring in case of backward selection. The results of the genetic algorithm are in between those of forward and backward selection methods.

Regarding the performance evaluation metrics, the CFS-metric yield better results than the MRMR-metric. It is also obvious, that there is a higher variance within the MRMR-metric, especially in combination with the backward selection methods. The relevance criteria do not show significant differences when used with the CFS-metric. In combination with the MRMR-metric, the variation is much larger. However there is no clear best or worst criterion. The MRMR/Information gain combination stands out to some extent since it sometimes performs worst (with sequential backward selection) and sometimes best (with sequential-floating backward as well as sequential forward selection).

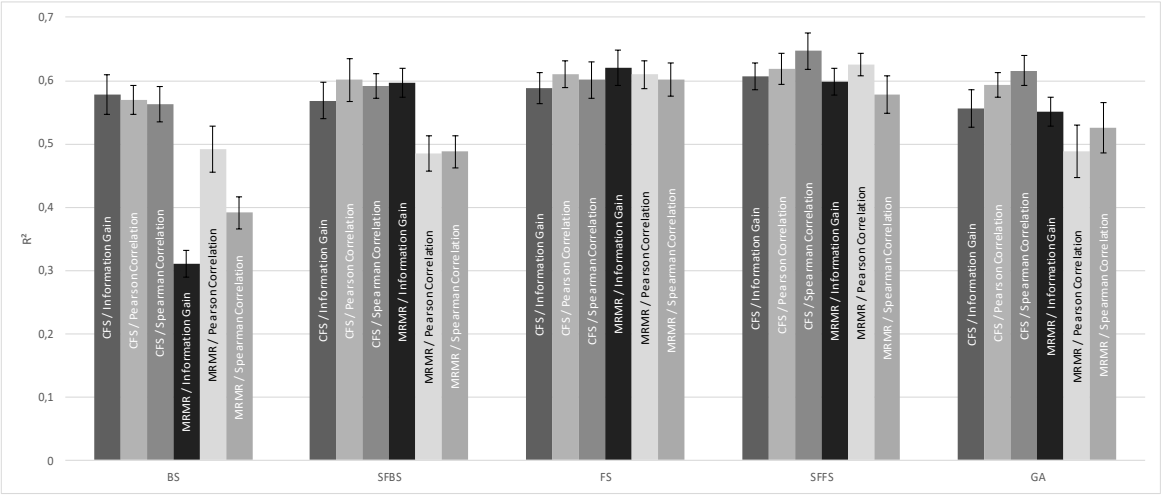

FIGURE 2. Coefficients of determination of the quality prediction models built with feature sets from different combinations of search strategy, performance evaluation metric and relevance criterion.

CONCLUSIONS AND OUTLOOK

In this paper, we have compared several methods for feature selection. Based on the theoretical fundamentals of filter, wrapper and embedded methods, we evaluated different filter methods in detail, taking into account the choice of search strategy, performance evaluation metric and relevance criterion. From the results one can derive that choosing sequential forward selection in combination with CFS-metric and Pearson's correlation is reasonable with regard to both model quality and computational effort.

In general, the displayed standard deviations are rather large, so for future research, one has to look more detailed into the interaction of feature selection, dataset and learning algorithm. In this case, the usage of additional datasets is advisable to ensure a sound statistical basis. Also, one should focus on those datasets with a high variance allowing proper results as the others are less likely to be used for building quality prediction models.

REFERENCES

1. T. Osswald, L. Turng and P. Gramann, Injection molding handbook, 2nd edn., (Carl Hanser publishers, Munich, 2008).
2. R. Schiffers, " Verbesserung der Prozessfähigkeit beim Spritzgießen durch Nutzung von Prozessdaten und eine neuartige Schneckenhubführung," PhD thesis, University of Duisburg-Essen, 2009.
3. J. Häussler, " Eine Qualitätssicherungsstrategie für die Kunststoffverarbeitung auf der Basis künstlicher Neuronaler Netzwerke," PhD thesis, University of Duisburg-Essen, 1994.
4. R. Vaculik, „Regelung der Formteilqualität beim Spritzgießen auf Basis statistischer Prozessmodelle," PhD thesis, RWTH Aachen, 1996.
5. B. Ribeiro, „Support vector machines for quality monitoring in a plastic injection molding process," IEEE Trans. Sys., Man, Cyb., **35** (3), pp. 401-410, (2005).
6. Y. Yang and F. Gao, "Injection Molding Product Weight: Online Prediction and Control Based on a Nonlinear Principal Component Regression Model," Poly. Eng. Sci. **46** (4), pp. 540-548, (2006).
7. M. Claesen and B. De Moor, "Hyperparameter search in machine learning," 11th Metaheuristics international conference, (Agadir, Marokko, 2015).
8. I. Guyon, S. Gunn, M. Nikravesh and L. Zadeh, *Feature Extraction – Foundations and Applications*, (Springer Publishing, Berlin, 2006).
9. W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection" in *Handbook of pattern recognition and computer vision* edited by C. Chen, L. Pau and P. Wang (World Scientific Publishing, Singapore, 1993), pp. 88-107.
10. M. Hall, "Correlation-Based Feature Selection for Machine Learning," PhD thesis, University of Waikato, 1999.
11. C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data", IEEE Comp. Soc. Bioinf. Conf., pp. 523-528, (Stanford, USA, 2003).
12. Z. Zhang, "Estimating mutual information via kolmogorov distance," IEEE Trans. Inf. Theo. **53** (9), pp. 3280-3282, (2007).