Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 52 (2016) 292 – 297

www.elsevier.com/locate/procedia

Changeable, Agile, Reconfigurable & Virtual Production

Use of Classification Techniques to Design Laser Cutting Processes

Hasan Tercan*, Toufik Al Khawli[b], Urs Eppelt[b], Christian Büscher[a], Tobias Meisen[a], Sabina Jeschke[a]

[a]Institute of Information Management in Mechanical Engineering (IMA) of RWTH Aachen University, Dennewartstraße 27, 52068 Aachen, Germany
[b]Department for Nonlinear Dynamics of Laser Processing (NLD) of RWTH Aachen University, Steinbachstraße 15, 52074 Aachen, Germany

* Corresponding author. Tel.: +49-241-80911-48; fax: +49-241-80911-22. *E-mail address:* hasan.tercan@ima.rwth-aachen.de

Abstract

In the field of manufacturing engineering, process designers conduct numerical simulation experiments to observe the impact of varying input parameters on certain outputs of a production process. The disadvantage of these simulations is that they are very time consuming and their results do not help to fully understand the underlying process. For instance, a common problem in planning processes is the choice of an appropriate machine parameter set that results in desirable process outputs. One way to overcome this problem is to use data mining techniques that extract previously unknown but valuable knowledge from simulation results. Our research examines the use of such techniques within the field of Virtual Production Intelligence (VPI). This paper proposes a novel approach for applying machine learning models, namely classification and regression trees, to design a laser cutting process. The evaluation shows that the models accurately identify regions in the multidimensional parameter space that increase the quality of the process (i.e. high cut quality). We implemented the models in the web-based VPI-platform, where the user is able to gain valuable insights into the laser cutting process with the aim of optimizing it.

Peer-review under responsibility of the scientific committee of the Changeable, Agile, Reconfigurable & Virtual Production Conference 2016
Keywords: Classification Tree, Regression Tree, Laser Cutting, Data Mining, Virtual Production Intelligence

1. Introduction

Nowadays, production companies in high-wage countries face the challenge of meeting individual customer requirements and rapidly changing market demands while keeping costs low. Providing a reliable but efficient production leads to growing complexities in the production processes [1]. Production planning and scheduling requires a large amount of human information processing and decision making. In particular in the field of manufacturing process planning, decisions involve the consideration of the effect of multidimensional parameters on preselected criteria of the manufacturing process. For instance, one common problem is the choice of an appropriate machine parameter set that results in desirable process outputs (e.g. high output quality or minimal energy consumption). Due to a high-dimensional domain space, the relationships between interdependent parameters and criteria is very difficult to achieve. In addition, they are very complex for the human mind to handle at a time.

In order to handle these problems, process designers make use of modern computational approaches for modeling and simulating manufacturing processes. The conventional technique is to perform several sets of simulation runs on the process, whereas each individual simulation is characterized by a high-dimensional set of parameters and several criteria [2]. The problem is that revealing the whole process behavior requires a very large number of time-consuming experiments. It is not feasible to run full numerical simulations throughout the whole parameter space at a reasonable computational cost. Because of that, experimental simulation runs are performed by appropriate Design of Experiment (DoE) techniques as well as other, experience-based procedures. Since simulations are based on discrete sets of process parameters, they can only cover partial aspects of the process and do not provide insights into the whole process. This is essential when it comes to extract useful information from the results and to understand the overall behavior of the underlying manufacturing process.

There exist different ways to cope with the computational complexity for analyzing processes and simulation results. At present, a paradigm shift is taking place towards a data driven simulation analysis by the use of data mining techniques. Data mining is the process of discovering interesting and yet unknown knowledge in data by the application of machine-aided algorithms [3]. In the context of manufacturing process planning, data mining models support the decision making process of engineers and thus help to gain competitive advantages of complex processes. Especially classification and regression techniques can identify process limits and reveal the impact of certain process parameters on the output. Process designers can use the gained knowledge to choose appropriate parameter sets

Peer-review under responsibility of the scientific committee of the Changeable, Agile, Reconfigurable & Virtual Production Conference 2016
doi:10.1016/j.procir.2016.08.001

for a given market demand (e.g. high quality output, process speed). For this purpose, the data mining results need to provide accurate insights in an intuitive way. Many existing solutions involve black box approaches that make it difficult to understand and contextualize the findings. Hence, within the scope of this paper, one key requirement is that the applied models are interpretable for the user. In addition to that, they need to accurately reflect the underlying manufacturing process and produce reliable results. In order to fulfill these requirements, this paper answers the following two research questions:

- How can data mining techniques in terms of classification and regression provide reliable decision support to optimize a manufacturing planning process?
- What are the necessary steps of evaluation and visualization to overcome the trade-off between interpretability and accuracy of the models?

The paper is organized as follows: Section 2 presents the related work. It briefly introduces the data mining process, including classification and regression trees that are investigated. Furthermore, it includes related work that deals with data mining solutions in the context of manufacturing simulation. Section 3 describes the laser cutting process for sheet metal. It provides the use case for an exemplary application of classification and regression trees in section 4. The evaluation results provide answers to the stated research questions. Finally, section 5 summarizes the main issues of this paper and gives an outlook on our future research.

2. Related work

2.1. Data Mining and Knowledge Discovery

In the narrow sense, data mining means the application of machine learning techniques for extracting certain patterns in data [4]. However, the solely application of algorithms does not suffice for gaining knowledge from data. For this purpose, different process models have been developed to provide an overview of essential steps that have to be taken for data mining. A popular model that is widely used within the data mining community is the process for knowledge discovery in databases (KDD) [5]. Given that the relevant data is collected, the KDD process basically consists of three large phases: pre-processing (e.g. data cleaning and transformation), modeling, and model evaluation (see Fig. 1). In a broader sense data mining is referred to be the knowledge discovery process that comprises all of these steps.

The modeling step of the data mining process involves the application of learning algorithms to reveal certain patterns. The algorithms are used to solve various types of problems. In this paper, two existing learning tasks are studied: classification and regression.

Classification is based on the following idea: given a certain problem domain, unknown relations can be identified on the basis of existing observations. Classification is used to learn a so called classifier on such a set of observations. It is applied on data which has many records, input variables (i.e. characteristics of records), and a categorical outcome variable (i.e. the target class) [6]. Given such a data set, the purpose is to train a

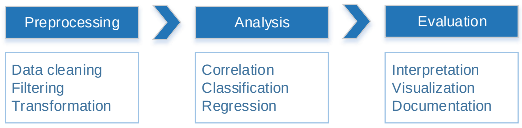

Fig. 1. Important steps of the knowledge discovery process [5]

model that analyzes the class label on the basis of the input variables. The found model can then be used to predict the classes of further records whose classes are not known yet.

Regression is similar to classification, but instead of a categorical class, it estimates a numerical output value [3,6]. Thus, regression is applied on numerical data only. Regression algorithms are generally used to analyze the relationship between a dependent variable (outcome) and several independent ones (attributes).

Once a data mining model has been learned on a training set, it is essential to evaluate its predictive performance on a separate test set. In doing so, the predicted outcomes of a model are compared with the real outcome of the test instances. Concerning the classification, the classification accuracy is one of the most widely used evaluation metrics [3].

$$Accuracy = \frac{\#\ correctly\ classified\ instances}{\#\ all\ instances} \quad (1)$$

There exist different metrics to evaluate the predictive performance of a numerical prediction. This paper proposes the mean absolute error (MAE). The lower this error is, the better the model predicts the output. Assuming that $p_1, \ldots, p_n$ are the predicted values of the test set and $a_1, \ldots, a_n$ are the actual ones, the MAE is calculated as follows [4]:

$$MAE = \frac{|p_1 - a_1| + \ldots + |p_n - a_n|}{n} \quad (2)$$

Various classification and regression models have been developed, each differing in certain aspects (e.g. functionality, performance, interpretability). In this paper, the main focus lies on tree-based learning models. The reason is that, in contrast to existing black-box models, they are understandable and interpretable by the user. This is one of the main requirements for the model. Furthermore, they reveal non-linear relationships between the inputs and the output variable. A tree-based model depicts a tree-structure which consists of a single root node, several internal nodes, edges and leafs [3]. A node represents a test on a certain attribute of the training set and thus divides the set into subsets. The edges that start from this node correspond to the outcome of this test. Each of them is linked to a child-node or a leaf node. A leaf specifies the class label (in classification) or the predicted numerical value (in regression). The prediction happens in a top-down manner, beginning from the root and walking down the tree.

Classification and regression trees divide the input data space into several segments, where each segment corresponds to a certain target value (i.e. class or predicted value). As a result, a trained model can not only be used for predicting unknown outcomes but also to analyze the relationships between specific output ranges and input data ranges in a visual manner. The visualization aspect makes them superior to other established