Data driven modeling & optimization of industrial processes

Stein, B. van

7 Conclusions and Outlook

Data science, as in the combination of analytics, machine learning, data mining and also optimization, plays an increasingly important role in industrial processes. Lots of sensor data, machine parameters, quality measurements, images and other signals are collected during these, usually complex, processes. Data mining can aid in extracting relevant information from this data to provide insight to domain experts. In addition, machine learning can aid the domain experts by predicting possible defects, quality indicators or machine parameters. Last but not least, optimization algorithms can be exploited to increase the overall productivity of a manufacturing process, increase the product quality or lower the cost.

Each of these data science components strongly interacts with each other, machine learning requires features which can be collected by data mining algorithms or by using the insight that data mining may provide. The optimization algorithms, on the other hand, can exploit the trained machine learning models to optimize the process efficiently, without requiring lots of evaluations on the real process.

In the context of the PROMIMOOC project these components are presented in a framework for controlling and optimizing industrial manufacturing processes. In Section 7.1 conclusions are drawn from the works in this dissertation and the contributions to the components and interlinking of components of this framework are discussed. Section 7.2 discusses what can be done in the (near) future regarding industrial process optimization.

7.1 Conclusions

To answer the research question, *"How can supervised and unsupervised machine learning techniques be utilized in complex industrial processes?"*, a framework for monitoring, controlling and optimizing an industrial manufacturing process is presented consisting of several components (Chapter 2). The aim of such a framework is three fold. The first objective is to provide the domain expert / process controller with relevant information about the current and possible future states of the process as quickly as possible. Secondly, to warn the domain expert in case of anomalies, and last but not least, to provide the domain expert with a near-optimal set of machine parameters to produce the next batch of products.

For such a framework to work efficiently, not only does it require algorithms with sufficient predictive power, it also requires algorithms that can work with high-dimensional data, in limited time and with a limited number of evaluations (in case of optimization algorithms). In addition, the framework needs to cope with data quality issues such as missing values. To solve this issue and as an answer to the research question, *"How can real-world data issues, such as missing data, be treated efficiently?"* , a novel algorithm for repairing missing values *Incremental Attribute Regression Imputation* (IARI) is proposed. In addition, a visualization technique to get more insight into missing value patterns is presented.

Outlier detection, or anomaly detection is an important component of the framework, as it is the first component that delivers crucial information to the domain expert. In most manufacturing processes the source material used for the final products come from different suppliers and vary in material properties. Outlier detection on these material properties is not straight-forward. Leading to the research question *"How can unsupervised learning be used for anomaly detection in high-dimensional industrial applications?"*. *Global Local Outliers in SubSpaces* (GLOSS) is specifically proposed to find outliers in high-dimensional data sets from a mixed distribution. GLOSS is able to find many relevant outliers in these material property data sets that existing state-of-the-art outlier algorithms are not able to find. By combining local and global neighborhood information, GLOSS works even when the dimensionality is huge and the outlier only shows outlying behaviour in a small subset of these dimensions.

To provide the domain expert with possible machine parameters to handle the incoming material, data driven models and optimization algorithms are required. One of the most used global optimization algorithms for expensive (in time or money) function evaluations, such as an expensive industrial process or time consuming simulator, is *Efficient Global Optimization* (EGO). EGO exploits an underlying Kriging model to find the next candidate solution, effectively replacing the expensive objective function with a much cheaper (but less accurate) surrogate model. The candidate solution is then evaluated on the expensive function and the result is used to update the underlying model. EGO uses the Kriging model because it does not only provide a predicted mean but also a prediction variance, the so-called Kriging variance. Unfortunately, Kriging has a major disadvantage, the training time complexity of the model is $O(n^3)$ and the memory complexity $O(n^2)$, which makes the model quickly infeasible as the data size increases. Leading to the two research questions *"How can predictive models such as Kriging be efficiently applied to a large amount of data?"* and the larger question; *"How can global optimization be applied in an industrial context, even in real-time?"*.

An answer and solution to these research questions, is the proposed *Cluster Kriging*. Cluster Kriging is a Kriging approximation algorithm where the data is being split into clusters or partitions. On each of these clusters a normal Kriging model is being trained. Using either an optimal weight scheme or membership probabilities, the different Kriging models' predictions and variances are combined into one final prediction and prediction variance. Using Cluster Kriging the time complexity of training the model goes down by a factor of k^2, where k is the number of clusters. When k is set proportional to the number of data points and k CPUs are provided, the training of the model even becomes linear. It is shown that the accuracy of Cluster Kriging is consistent and sometimes even outperforming the original Kriging algorithm on a large set of functions.

Another solution to the problem of the EGO with Kriging complexity is to not use Kriging as the underlying surrogate model. Unfortunately, many machine learning techniques do not have a prediction variance or prediction uncertainty available out of the box. In this dissertation an uncertainty measure is proposed, the k-NN uncertainty. This heuristic uncertainty measure is model independent and calculates an approximation of the prediction error using the error information of the

k-nearest neighbors. It is shown that EGO in combination with different models, such as artificial neural networks, support vector machines and random forests, using the k-NN uncertainty, performs equally well or better than using the original EGO with Kriging. The performance of the EGO algorithm is very dependent on the fit of the underlying model, being able to use different predictive models allows for choosing a better fitting model and therefor increasing the efficiency of the EGO algorithm.

7.2 Future work

This dissertation contributes by providing a set of specific solutions in the fields of missing value imputation, anomaly detection, Kriging and model-based optimization. In addition, an overview of a framework for monitoring and optimizing industrial manufacturing processes by using data mining, machine learning and optimization techniques is provided. Different implementations of the framework components are suggested and several issues in connecting these components are challenged. A lot of challenges are of course still present and the design choices made during this project will not always be optimal for each application.

Some of the open questions related to the research in this dissertation are:

Proper benchmarking of outlier detection algorithms: One of the most challenging topics in outlier detection is that the result of an outlier detection algorithm, most of the time, cannot be evaluated using a simple metric. There is no one definition of what an outlier or anomaly is. How a domain expert defines an anomaly depends highly on the domain, application and on the expert. To compare outlier detection algorithms, well defined benchmark data sets are essential. In current state-of-the-art research, classification data sets are misused to validate the performance of anomaly detection algorithms by detecting the least present class in the data set. This results in a biased comparison which favors clustering based algorithms that use the residual class as outlier identification. The development of a proper benchmark for these algorithms, including a wide variety of anomaly and data types, would aid the research field enormously.

Automatically configuring predictive models: The predictive models such as Cluster Kriging, Deep Neural Networks and Random Forests all have a few to many hyper-parameters. In the development of the framework for Tata Steel and BMW, most of these models' parameters where optimized by hand, grid search or simply by taking the most commonly used values from literature. Algorithm configuration is an upcoming research field that could provide very valuable insight and improvements in selecting and configuring these models for optimal performance.

Automatically optimizing artificial neural network architectures: Next to the hyper-parameters of the more classical predictive models, deep neural networks consist of many design choices. From the number of layers to the hyper-parameters of each layer, the preprocessing step of the input data and the activation functions. All of these design choices could possibly be automated by using efficient optimization techniques. Reducing the effort of finding a good predictive model and possibly increasing the final accuracy of these predictive models.