論文の内容の要旨

Thesis Summary

論文題目 Framework for data-driven process improvement and operations support in biopharmaceutical drug product manufacturing
(バイオ医薬品製剤製造におけるデータ駆動型プロセス改善・運転支援のためのフレームワーク)

氏 名　カゾーラ　ジオエレ

Globally, in developed countries, the aging population carries several consequences, such as manufacturing, services, taxes, and healthcare. In parallel, the cost of R&D persistently levitates, whereas the cost of manufacturing remains constant; therefore, large pharmaceutical companies have to face the increasing competition of generic producers. To contrast the resulting decrease of sales margin, big pharmaceutical companies are required to improve the performance of their processes, e.g., by achieving reduced production time, and down time as well as enhanced capacity.

Various Process Systems Engineering (PSE) approaches, which comprise modeling simulation and optimization have been applied to achieve the optimal design in the synthesis of pharmaceutical processes. A further concept, namely digitalization or Industry 4.0 (I4.0), is the current major driver of the improvement of processes and the technological revolution in the industry. More rigorous approaches and introspective studies on the incorporation of well-established PSE methodologies with the novel digital revolution approaches are necessary to unlock the real potential of I4.0 in improving pharmaceutical manufacturing processes.

The thesis presents a framework, the application of which assists the uncertainty-conscious and data-driven decision-making in the process improvement and operations support in Good Manufacturing Practice (GMP)-controlled biopharmaceutical manufacturing. The framework consists of three main parts: data preprocessing (I), process performance assessment (II), and predictive maintenance decision-making (III). The framework was applied in an industrial case study, where manufacturing records generated from a change over-process operated in a facility belonging to F. Hoffmann-La Roche Ltd.

In biopharmaceutical Drug Product (DP) manufacturing, sterile filling plants are usually non-dedicated, which implies the necessity of change-over operations. The sterile filling process comprises washing and sterilizing of empty glass containers, filling of these containers with a drug solution, sealing of the containers, and finally the visual inspection. Start-up and changer-over are support processes, which enable the switch from a product/format to another by maintaining the sterility and purity required for the manufacturing of high-quality products. An example of a support process is the Cleaning-In-Place and Sterilizing-In-Place (CIP/SIP), which involves cleaning the product-contacting surface of the filling system such as pipes, tanks, and filling needles by removing product residues and particles, sterilizing it, and finally drying it. Support processes are extremely time intensive, approximately half of the production timeline; in fact is the underlying reason of this case study.

In the first part of the framework, a new algorithm for transforming raw data recorded in biopharmaceutical manufacturing into functional data in an automated manner is presented. The algorithm consists of seven activities, including process task identification by natural language recognition, model training for selecting and filtering of noise—e.g., data not belonging to commercial batches—from the raw data, and clustering the process into single batches using semi-supervised clustering. The remaining activities ensure the compatibility of the resulting dataset with the remaining part of the framework.

The raw data string was treated similar to a DNA strain during sequencing: first, the process recipe was randomly cut into primers, which are sequences of strings and fragments of the raw data string; the heterogeneity coefficients between the primers were calculated. The heterogeneity coefficient and the primer size were then used to train decision tree-classifiers based on the human perception of data noise. The training of the classification model resulted in the identification of the noise in the data with an F-score of 0.99. The filtering of the noise could shrink the data set to 60% of the original size; 40% of the data recorded did not

contain process relevant information but was recorded because of GMP regulations and a non-cost-efficient monitoring strategy. The noise-free data were clustered using a constraint k-means algorithm, which allowed separation of every single batch. Last, the data points were classified into three categories, namely, normal process, failure, and repetition. The resulting dataset was used in the subsequent two parts of the framework to identify the improvement potentials and to predict imminent failures.

In the second part of the framework, an uncertainty-conscious methodology was presented; it can assess process performance and facilitate process improvement in biopharmaceutical DP manufacturing. The work was described as a six-activity model using IDEF0, which are "define key performance indicators (KPIs) (A1)," "create an initial process performance model (A2)," "collect and adapt data (A3)," "characterize the process performance model (A4)," "identify tasks to improve (A5)," and "perform what-if analysis (A6)".

The KPI was defined as the process runtime (A1) and was modeled as the sum of the duration of the process tasks, remedying operations (A2). The historical records were imported (A3 and A4). A two-loops stochastic global sensitivity analysis (GSA) employing a partial rank correlation coefficient (PRCC) was used to select the tasks that highly affect the KPI. The GSA aimed to assess the process design in an operational environment, so the task duration was defined as the changeable parameter. To incorporate operational uncertainty in the analysis, the outer loop propagated the operation uncertainty by random sampling Monte Carlo simulation (MCS). In the inner loop, the PRCC for every task was calculated with Latin hypercube sampling-MCS (LHS-MCS) using the performance model. The feasibility indicator was defined as 0 to 1, 0 being complete infeasibility and 1 being perfect feasibility; the feasibility process modification reflects industrial know-how. Finally, the results of PRCC and the feasibility analysis were combined (A5) to identify tasks that showed high impact and feasibility as improvement potentials. The suggestions of industry experts were implemented in a what-if analysis with the process model (A6); the result showed 120 h potential time-savings.

In the third part of the algorithm, an intelligent algorithm, which is based on Industry 4.0 and big data approaches was presented. The algorithm predicts the failure status of the process from physical sensors in real-time. A decision tree (DT) model was trained to identify failed batches from successful ones; historical sensor data were transformed by principal component analysis (PCA) to reduce the dimensionality of the system. A retraining loop maintained the quality of the prediction over time because the algorithm is based on

machine learning and the process continuously evolves; the algorithm resulted in a decision after the analysis of the risk on the performance in case of action. An integrated failure prediction and decision-making tool was used to support the decision of stopping a batch before a failure occurs. The deployment of the algorithm on the real-world data results in the potential time saving of approximately 100 hours per month.

The thesis proposes a data-driven framework for supporting the decision-making in process and operation improvement for biopharmaceutical manufacturing. The framework consists of three main steps: the integration of existing industrial databases, assessment of the process performance and identification of the task to improve, and imminent failure mitigation by plant predictive maintenance. The thesis aims to provide a novel and rigorous tool/approach that is applicable in an industrial environment, to solve long-term challenges, such as decision-making regarding maintenance policies and process improvement as well as daily operation challenges, such as downtime reduction. The result of the application of the framework in the industrial case study was implemented in the commercial facility; a reduction of the process runtime was achieved. In future works, the framework will be integrated in the manufacturing operations after further generalization. Moreover, the framework will be expanded to other pharmaceutical processes, such as sterile filling, packaging, transportation. Last, the impact on sustainability will be included in the decision-making of process modification.