6-14-2024

Data Integration and Analysis For Smart Manufacturing Process Optimization

Adir Even
IEM, adireven@bgu.ac.il

Ofir Azulay
Ben-Gurion University of the Negev, ofirazu@post.bgu.ac.il

moshe cohen
Ben-Gurion University of the Negev, moshe8@post.bgu.ac.il

DATA INTEGRATION AND ANALYSIS FOR SMART MANUFACTURING PROCESS OPTIMIZATION

TREO Paper

Adir Even, Ben-Gurion University of the Negev, Beer-Sheva, Israel, adireven@bgu.ac.il

Ofir Azulay, Ben-Gurion University of the Negev, Beer-Sheva, Israel, ofirazu@post.bgu.ac.il

Moshe Cohen, Ben-Gurion University of the Negev, Beer-Sheva, Israel, moshe8@post.bgu.ac.il

Abstract

The concept of Smart Manufacturing reflects, among other principles, data collection and analysis toward manufacturing processes optimization. Smart manufacturing commonly mandates the integration of data collected from multiple independent sources toward joint analysis. This research collaborates with a large Nano-Fabrication facility that produces Nanotechnology-based prototypes. Machines at this facility collect vast amounts of data, applying different data structures, formats, and sampling rates. Given this lack of uniformity, joint analysis of data collected at different manufacturing stages is practically impossible. Motivated by this challenge, this research aims to develop a methodology for data collection, integration, and storage in complex manufacturing environments. The methodology requires formulation of data structures and integration procedures that can handle data sources' variability and inconsistencies. The benefits of forming a unified infrastructure will be demonstrated by using the integrated data for detecting relationships and dependencies between variables that reflect various manufacturing stages, toward process optimization and improvement.

Keywords: Smart Manufacturing, Manufacturing Data, Data Integration, Process Optimization.

1 Research Settings and Motivation

The concept of Smart Manufacturing relies on the collection and analysis of data accumulated during the production process, and its use for optimizing processes, informing decisions, and improving outcomes quality and their fit to individual consumer needs (Jamwal at al., 2021). The quest for smart production involves the rapid progress in information technologies (IT) that support data collection, storage, sharing and analysis (Ehrlinger et al., 2018; Ehrlinger et al., 2022). For example, "Industrial Internet of Things" (IIOT) technologies enable the collection of real-time data from a variety of advances sensors embedded in the production lines, and the use of this data for process control, individual adjustment of product items, and predictive maintenance (Jaskó et al., 2020; Kim et al., 2022).

Our research collaborates with a large Nano-Fabrication facility that specializes in manufacturing Nanotechnology-based prototypes, primarily for research and development (R&D) purposes. This facility operates a broad variety of machines that collect vast amounts of data, using various sensors. Data acquisition consistency cannot be enforced, as machines are purchased from different vendors; hence, apply different data protocols and representation. This lack of consistency prevents the integration of data resources and their analysis toward process optimization and decision support; hence, substantially hinders efforts to apply smart manufacturing principles in this facility.

Our study deals with a major challenge for smart manufacturing environments – the need to integrate the data collected from several different sources during different stages of the production process. Many production environments do not necessarily apply uniform coding of production batches and collect data separately by different machines at inconsistent structures and sampling rates. The challenge of data integration raises key questions that direct our research: First, how should data be represented, organized, and stored in a multi-source production process, in a manner that would permit the analysis of a manufacturing process as a whole? Second, once data is integrated, how should it be analyzed, given high process configurations variability, and possibly a broad variety of optimization goals?

Guided by these questions, our research will be conducted in two stages - A) Data Infrastructure: The development of modeling methods and appropriate infrastructure for data reconfiguration and storage and B) Data Analysis: Demonstration of the use of well-integrated manufacturing data resources for guiding process improvement efforts and supporting decision making.

2 Data Infrastructure

A preliminary investigation revealed some significant quality defects that are common to manufacturing data – missing value, lack of conformance to value-domains, latencies and others. Data quality defects could be attributed to various structure-models, storage methods, and coding errors. The absence of uniform coding impedes the merging of data that would be essential for analysis and reporting, while decentralized storage complicates retrieval and integration, and leads to missing or underutilized data. Such data quality defects are common in complex manufacturing environments with decentralized data collection and limit analytical capabilities substantially.

The recognition of potential data-quality deficiencies as such motivated the exploration of modelling techniques for data reconfiguration and storage. The preliminary model developed by this study was defined in a generic form that may fit multiple manufacturing environments; and then applied to address the more specific needs of the Nano-Fabrication facility. To support researchers' needs at the facility, the integrative data model consolidates all the data produced during the production processes, both manually and by manufacturing machines, and subsequently permits access and retrieval for analysis.

The model guides uniformity in the collection of datasets from different machines by setting high-level structure and format requirements, but still permits certain flexibility to reflect the unique data characteristics of each source. A preliminary examination of data modeling requirements highlighted some limitations of the commonly used relational database servers (RDBMS) technologies for supporting common data manufacturing characteristics; hence, motivated evaluation of alternative technologies, such as the Wide-Column databases, that do not necessarily adhere to tabular method.

The model was embedded with a newly implemented system for manufacturing process documentation. Thus cloud-based system documents manual process configuration, as well as the data collected automatically by machines at the facility. The manual setup of processes involves the definition of process-stages order, configuration of machines and materials, setting parameter values, and documentation of other process-related activities. The system applies various functionalities that aim at simplifying data entry, enhancing usability and reducing errors, toward efficiency and accuracy in documenting the production process. The data collection comprises two entity types: A) Static entities hold permanent configurational data established during system setup. B) Dynamic entities capture real-time data, reflecting ongoing operations and the current state of the production process.

3 Data Analysis

The later data analysis phase will focus on demonstrating the use of the integrated manufacturing data to extract insights that can help enhance the production process. The assumption is that effective data that reflects an entire multi-stage process may help identify inefficiencies and failure points and link them to machines' setup and operation procedures. Uncovering relationships between variables at different stages is crucial for gaining insights into the production process, facilitating an understanding of how adjustments impact material properties, and enable attainment of specific product characteristics.

Another scope of analysis is developing methods for detecting the presence of data quality (DQ) defects and measuring their magnitude. The potential damage of DQ defects is obvious in manufacturing environments such as the facility that this research collaborates with; However, as the collected data does not necessarily follow a predefined structure, traditional DQ detection and measurement methods cannot be applied, and alternatives must be developed and explored. Furthermore, when attempting to assess how DQ defects affect data analysis outcomes, our research will have to deal with issues of high variability among manufacturing-process outcomes (R&D prototypes, in our case), and relatively small sample sizes per batch.

4 Research Progress and Future Steps

Research progress has been notable, particularly with the completion of the implementation of the cloud-based information system at the institute. This assimilation effort entailed comprehensive training sessions for the institute's researchers on the methodology of data entry into the system. The iterative nature of the assimilation process facilitated adjustments to the system based on emerging requirements, while researchers gained proficiency in its usage. Currently, with the system fully operational, data regarding the production process is systematically accumulated, providing a rich source for analysis. Currently, our focus lies in establishing data integration protocols (ETL - Extraction, Transformation, Loading) for retrieving accumulated data from the cloud-based system to a local data server, where data will be restructured for further analysis.

The research team aims to develop tools and methods to understand parameter relationships affecting the production process. We will focus on designing mathematical approaches to identify DQ defects and create tools for pinpointing production process failures, enhancing performance and enabling precise parameter adjustments. Advanced analytics techniques, such as machine learning, will be used to extract insights from manufacturing data, contributing to the advancement of smart manufacturing practices.

References

Ehrlinger, L. et al. (2022). “Data Integration, Management, and Quality: From Basic Research to Industrial Application,” *Database and Expert Systems Applications - DEXA 2022 Workshops. Communications in Computer and Information Science*, 1633, 167–178.

Ehrlinger, Lisa, Werth, B., & Woe, W. (2018). “QuaIIe: a data quality assessment tool for integrated information systems,” *The Tenth International Conference on Advances in Databases, Knowledge, and Data Applications*, Nice, France.

Jamwal, A., Agrawal, R., Sharma, M., & Giallanza, A. (2021). “Industry 4.0 technologies for manufacturing sustainability: A systematic review and future research directions,” *Applied Sciences* 11 (12), 5725.

Jaskó, S., Skrop, A., Holczinger, T., Chován, T., & Abonyi, J. (2020). “Development of manufacturing execution systems in accordance with Industry 4.0 requirements: A review of standard- and ontology-based methodologies and tools,” *Computers in Industry* 123 (2020), 103300.

Kim, S., Ricardo, P., Caballero, I., & Lee, D. (2022). "Organizational process maturity model for IoT data quality management,” *Journal of Industrial Information Integration* 26 (2022), 100256.