

8-10-2023

Multi-Source Data Integration in Smart Manufacturing Environments

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

Adir Even
IEM, adireven@bgu.ac.il

Follow this and additional works at: https://aisel.aisnet.org/treos_amcis2023

Recommended Citation

Azulay, Ofir; cohen, moshe; and Even, Adir, "Multi-Source Data Integration in Smart Manufacturing Environments" (2023). *AMCIS 2023 TREOs*. 89.
https://aisel.aisnet.org/treos_amcis2023/89

This material is brought to you by the TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2023 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Multi-Source Data Integration in Smart Manufacturing Environments

TREO Talk Paper

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

Adir Even

Ben-Gurion University of the Negev
adireven@bgu.ac.il

Abstract

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. The quest for smart production involves the rapid progress in information technologies (IT) that support data collection, storage, sharing and analysis. "Industrial Internet of Things" (IIOT) technologies, for example, enable the collection of real-time data from a variety of "smart" sensors embedded in the production lines, and the use of this data for process control, individual adjustment of product items, and preventive maintenance activities.

Our study deals with a major challenge for smart manufacturing environments – integration of data collected from several different sources during different stages of the production process. In many production environments, data integration becomes a major challenge as production batches are not necessarily uniformly coded, data is collected separately by different machines, and structured differently at different stages of the process. The challenge of data integration raises key questions that will direct our research - first, should data be represented, organized and stored in a multi-source production process, in a manner that will permit analysis of the entire process? Second, once data is integrated – who should it be analyzed, given the high variability between the process configurations, data structures and formats, and the broad variety of optimization goals?

Our research will be carried out in collaboration with a large Nanofabrication facility. The center carries out manufacturing processes, primarily for research and development (R&D) purposes. The center operates nano-technology manufacturing machines, which have collected vast amounts of data over the years, via various sensors. Each machine collects data separately, under different protocols and at different formats and sampling rates. Due to this lack of uniformity, the large-scale datasets are currently not integrated and their utilization for process optimization and decision support is from being satisfactory.

In response to those challenges faced by Nanofabrication facility, which are common to many modern manufacturing environments, this study will examine the question of data integration from multiple sources and its joint analysis in two stages. The first stage will focus on data organization and storage - the development of a generic data model, adapted to the needs of the center and other similar manufacturing environments, which will allow uniform collection of data from different machines under a common structure and format, however with certain flexibility for support the unique data characteristics of each source. A preliminary examination of data modeling highlighted limitation of the commonly-used relational database technologies for supporting such requirements; hence, the need to explore in alternative technologies that are not necessarily based on tabular representation, such as the Wide Column database. After the integration of data resources, the second research stage will focus on using them for analyzing production processes, identifying factors that affect process efficiency and outcomes, and detecting the impact of data quality defects.