Journal of Manufacturing Systems 43 (2017) 214–224

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical Paper

Interpretative identification of the faulty conditions in a cyclic manufacturing process

Dominik Kozjek[a,*], Rok Vrabič[a], David Kralj[b], Peter Butala[a]

[a] Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva 6, Ljubljana, Slovenia
[b] ETI Elektroelement d.d., Obrezija 5, Izlake, Slovenia

ARTICLE INFO

Article history:
Received 28 September 2016
Received in revised form 24 February 2017
Accepted 3 March 2017
Available online 22 March 2017

Keywords:
Production process
Fault identification
Root cause analysis
Decision support
Big Data

ABSTRACT

The intensive development of information and communication technologies in recent years has led to an increase in data size and complexity. Conventional approaches, with associated methods of analysis based on descriptive and inductive statistics, may no longer be suitable for extracting the valuable information that is hidden in the available data.

Computer-controlled manufacturing systems are becoming rich sources of data. Plastic injection moulding and die casting systems are typical examples of such manufacturing systems where the parts are produced by repeating the same sequence of steps that make up a manufacturing cycle. For each cycle, similarly structured data is generated.

In this work a method for systematic data analysis for cyclic manufacturing processes is presented. The proposed data-analysis method integrates well-known heuristic algorithms, i.e., decision trees and clustering, with the purpose of identifying types of faulty operating conditions. The result of the analysis is an interpretable model for decision support that can be used for fault identification, to search for root causes, and to develop prognostic systems. A holistic approach of applying the proposed data-analysis method, along with suggestions and guidelines for implementation, is presented. A case study is presented in which the proposed method is applied to real industrial data from a plastic injection-moulding process.

1. Introduction

With the development of technology, the complexity of manufacturing systems is steadily increasing. In turn, the complexity of generated data is also increasing, which is reflected in incompleteness and inaccuracy of data, heterogeneous and dynamically changing data structures, large and fast increasing data volumes, etc. All these facts make analysis of manufacturing data for extracting useful information and knowledge very challenging.

For the above reasons, advanced online and real-time monitoring, identification and prognostics are becoming increasingly difficult. Insufficient data quality, lack of holistic databases that integrate operational and decision-support data, and a lack of models for analysis are some of the recent, most pressing challenges in industry [1]. In the future, new applications for management and use of data will need to be provided [2].

Current issues related to data management have led to the formation of what is known as the concept of Big Data. Big Data is generally related to the interplay of information- and communication-technology developments that enables an increase in storage capacities and computing power, as well as the advanced integration of traditional and novel methods for the acquisition, storage, integration, analysis, modelling, visualisation, and other kinds of management of data that is too large or too complex to allow the extraction of value using traditional approaches [1–6]. Data complexity can arise due to diversity, variability, data distribution, lack of structure, shortcomings, non-credibility, speed, or automation requirements [7,8]. A typical Big Data system is divided into four stages that sequentially form the value chain: (1) generation, (2) acquisition, (3) storage, and (4) analysis [9,10]. Each of these phases includes elements, techniques, tools, methods and concepts, which conjointly form efficient systems for extracting the value from large and complex data.

Heuristic methods of analysis turn out to be very efficient for fault-diagnosis problems in complex manufacturing processes, where the states are described by complex combinations of many parameters. Recently, in the manufacturing domain, a lot of research was carried out regarding fault-diagnostics problems by

* Corresponding author.
E-mail addresses: dominik.kozjek@fs.uni-lj.si (D. Kozjek), rok.vrabic@fs.uni-lj.si (R. Vrabič), david.kralj@eti.si (D. Kralj), peter.butala@fs.uni-lj.si (P. Butala).

http://dx.doi.org/10.1016/j.jmsy.2017.03.001
0278-6125/

incorporating machine-learning methods, data mining, and other evolving techniques, targeting the problems of sensing and describing multiple and dynamic faults, monitoring, real-time processing, searching for the best features, failure-condition identification, early prediction of failures, and assessing the severity of failures [11].

Many of so-far untouched challenges and opportunities remain for the development and introduction of useful heuristic techniques in manufacturing systems that would improve the process quality through an intelligent use of the available data [12].

The paper presents a holistic data-analysis method for the interpretative identification of faulty conditions in a cyclic manufacturing process (IIFC). The objective here is to demonstrate the usability of large volumes of data generated during a widespread type of the so called cyclic manufacturing processes, such as plastic injection moulding, for extracting valuable information and new knowledge models for better understanding of the manufacturing process and its faulty operating conditions. The proposed approach uses well known data-mining methods and addresses the arising issues of increasing data size and complexity.

Cyclic-manufacturing processes are defined as manufacturing processes in which the parts are produced repeating the same sequence of steps that forms the manufacturing cycle. The data that is generated in these processes describes the state of the process for each cycle. This is suitable for applying machine-learning and other advanced techniques of data analysis, because minimal effort is required to define the common features for each cycle due to the repetitive nature of the cycle's steps. Although machine-learning techniques are well known, relatively little effort has been made towards an effective implementation in real, every-day cases in manufacturing systems.

The proposed IIFC method is intended to systematically analyse the empirical data that are usually generated in the cyclic manufacturing process. The aim of applying the IIFC method is to improve the quality of an already well-optimised process through the identification of rare faulty operating conditions, and by learning about their types and characteristics. A two-phase data-analysis workflow is suggested. In the first phase, rules,[1] describing the process conditions are extracted with the use of a decision-tree heuristic algorithm and the expert knowledge of the observed manufacturing process. Bit vectors (containing only 0 and 1) are obtained, describing the combination of rules that are true (1) or false (0) for each faulty cycle of the manufacturing process. In the second phase, different types of faulty conditions are revealed and described using a clustering technique and the extracted rules. The output of IIFC analysis is an interpretative model, which is understandable to people such as operators, process engineers, plant supervisors, etc., that interact daily with the observed manufacturing process.

The paper is structured as follows. Section 2 describes the proposed IIFC method. The IIFC method's role in a general workflow of data-analytics, a holistic description of the combination and the sequence of steps that form the workflow of the IIFC method, including suggestions for how to organise and apply the data as well as the properties related to implementation and applicability of the IIFC method, are given. The content of this paper continues with an application of the proposed method in a real industrial scenario of plastic injection moulding. The procedure for applying the method is presented together with the results and an interpretation that demonstrates the usability in a real industrial environment.

[1] E.g., "*Temperature of a cylinder exceeded 251 °C.*", "*Day of the week is Saturday.*", etc.

2. Interpretative identification of faulty conditions

The proposed IIFC method suggests a combination and sequence of data-analysis steps together with guidelines for applying them to usually available data of cyclic-manufacturing processes.

The role of the proposed data-analysis steps of the IIFC method according to the general workflow of data analytics is shown in Fig. 1. Heterogenous data can be generated at different locations by work systems, the manufacturing execution system, etc. Using appropriate methods for data acquisition, we can reduce the data size and dimensionality, with a tendency to minimise the loss of information. Large amounts of data are then managed by advanced databases and cloud-computing systems that enable high throughput, reliability, and availability [9,10].

The object of the IIFS analysis is a set of manufacturing process cycles. A cycle can be *normal*, *faulty* or *other*. The method assumes that the cycle can be faulty for various reasons, e.g., deviations from the required dimensional tolerances of the products, an unplanned machine stop, etc. Other cycles are those that do not belong to the group of normal cycles nor to the group of faulty ones, e.g., cycles that occur in the vicinity of the faulty ones, cycles with missing data, etc. Each cycle c_i is described in terms of a unique ID, a date and time, a work system, process parameters, etc. (Eq. (1)).

$$\begin{aligned} c_i = \{ & c_i^{id}, c_i^{dateTime}, c_i^{workSystem}, \ldots, c_i^{processParameter1}, c_i^{processParameter2}, \\ & c_i^{processParameter3}, \ldots, c_i^{inputMaterial}, \ldots, c_i^{machineAlarm1}, \\ & c_i^{machineAlarm2}, \ldots, c_i^{outsideHumidity}, c_i^{outsideTemperature}, \\ & c_i^{workingShift}, \ldots \} \end{aligned} \tag{1}$$

The approach of IIFC method is (1) to describe each faulty cycle with a combination of (a) *fault-specific rules*,[2] i.e., rules that describe faulty operating conditions, and (b) *other rules*,[3] i.e., rules that are not necessarily related to the physics of the manufacturing process or faults, but their inclusion could potentially lead to the discovery of the root causes for the emergence of faulty operating conditions, (2) to reveal similar subgroups of faulty cycles based on combinations of fault-specific rules and (3) to create detailed descriptions of these identified subgroups based on all extracted rules.

Data that are usually available are in different ways related to the manufacturing process and the faulty conditions, thus they need to be treated differently when extracting the rules which are used for the identification and the description of the faulty conditions. The data types (*id*, *dateTime*, *workSystem*, etc.) are of different kinds, e.g., integers, real values, descriptors, etc., and are classified into four groups: (1) metadata ($G1$), (2) process-specific ($G2$), (3) fault-specific ($G3$) and (4) other ($G4$) data. Metadata ($G1$) are used to support computing operations and enables efficient data storage and retrieval. From process-specific ($G2$) and fault-specific ($G3$) data, fault-specific rules which are used to identify different types of faulty operating conditions, are extracted. To describe properties of identified faulty-condition types, fault-specific rules, together with other rules extracted from other ($G4$) data, are used. Classification of input data types is presented in Table 1 and explained below.

Metadata ($G1$). Group $G1$ includes the data that identify the manufacturing cycle time- and location-wise. The unique ID of the cycle, the date and time of the cycle, the work system, and the cycle quality (*normal*/*faulty*/*other*), etc. are the types of data that belong to this group. A unique ID is given to a cycle in the step of data storage. A

[2] E.g., "*Temperature of a cylinder exceeded 251 °C.*"

[3] E.g., "*Day of the week is Saturday.*"