Available online at www.sciencedirect.com
ScienceDirect
Procedia CIRP 67 (2018) 8 – 12
www.elsevier.com/locate/procedia

11th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRP ICME '17

Approach for defining rules in the context of complex event processing

Julia Pielmeier[a,]*, Stefan Braunreuther[a], Gunther Reinhart[a,b]

[a]*Fraunhofer Research Institution for Casting, Composite and Processing Technology (IGCV), Provinostr. 52, 86153 Augsburg, Germany*
[b]*Institute for Machine Tools and Industrial Management (iwb), Technical University Munich, Boltzmannstr. 15, 85748 Garching, Germany*

* Corresponding author. Tel.: +49 821 90678-178; fax: +49 821 90678-199. E-mail address: julia.pielmeier@igcv.fraunhofer.de

Abstract

The vision of 'Industrie 4.0' and the Internet of Things (IoT) is based on the connection of smart products and smart machines equipped with sensors and actuators. The digitalization of industrial processes leads to the production of data streams. In this context, real-time analytics is becoming more and more important for business applications as a result of the need to deal with the growth of data and to react instantly to changes in the data streams. Complex event processing (CEP) is an efficient methodology to enable processing and real-time analysis of streams of data. The main focus of CEP is the detection of patterns in data streams. Therefore, a set of rules has to be predefined. These rules are characterized by various parameters. Defining the optimal values for these parameters is challenging. In current CEP systems, experts have to define the rule patterns. In this paper we suggest three ways to define rules: manual by domain experts, semi-automated by rule mining, or optimization. However, not all of these three ways can be applied to a production scenario or use case. Thus, we compare these approaches and match them with the appropriate production scenario.

Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Intelligent Computation in Manufacturing Engineering

Keywords: Process control; Pattern recognition; Optimization

1. Introduction

Within the digitalization of production processes, increasingly large quantities of data are being produced while the data complexity is rising.

In order to gain useful information from the generated data masses, employees have to be supported with technology as the data complexity goes beyond human comprehension.

For this purpose, automatic methods have been developed to analyze various data. These methods and systems ensure the processing and analysis of the data so that the data can be interpreted by people. In this context real-time analytics is becoming more and more relevant for businesses and social applications as a result of the need to deal with the growth of data and the need to respond and react instantaneously to data triggers [1]. Such data is also referred to as "fast data". *Harvard Business Review* states that "large enterprises have spent heavily on managing large volumes and disparate varieties of data for analytical purposes, but they have devoted far less to managing high velocity data" [2]. This can be noticed irrespective of the fact that data is expected to double every two years for the next decade. The large amounts of new data being produced by intelligent devices are primarily responsible for this development. "That's a problem, because high velocity data provides the basis for real-time interaction and often serves as an early-warning system for potential problems and systemic malfunctions" [2].

The term "real-time system" signifies the requirement for IT systems to process occurring events within a specified time interval [1]. Important characteristics of real-time systems for the use in industrial fields are:

- Low Latency
- High Availability
- Horizontal Scalability

In order to achieve a near real-time monitoring, control of production and logistics processes, intelligent processing and analyzing of data is required. As a result of this development, Complex Event Processing (CEP) has established for analyzing extensive data streams in near real-time. CEP are methods, techniques, and tools to process events in real time [3]. The main purpose of CEP is the detection of complex event patterns from data streams such as sensors.

The definition of the rule patterns for matching the events based on the temporal, semantic, or spatial correlations is the

Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Intelligent Computation in Manufacturing Engineering
doi:10.1016/j.procir.2017.12.167

central task of CEP systems. CEP engines are completely based on these rules, which are mostly specified manually by domain experts. The complexity for defining and deriving rules depends on the production process and scenario. Domain experts can define rules in some cases [4]. However, defining the optimal values for the rules is challenging. Thus the expert-based rule specification is a limiting factor that complicates the integration and distribution of CEP [5].

On the basis of this observation it becomes clear that alternative approaches have to be pointed out for the rule definition. In this paper we have a look at the state of the art of this range of topics and suggest three ways to define rules based on the research. Further, we compare these approaches and match them with the appropriate production scenario. Finally, we show the election and application of one approach for a production use case.

2. State of the Art

2.1. Real-Time System and Analytics

The term real-time system signifies the timeliness of the information flows. "Real-time" does not explicitly apply to a period that is as short as possible. According to DIN 44300, real-time is defined as " the operation of a computer system in which programs for processing data are constantly ready for operation in such a way that the processing results are available within a predetermined period of time. Depending on the application, the data may be generated according to a temporally random distribution or at predetermined times" [6].

This definition shows that it is not determinant whether a process delivers its results as quickly as possible. The decisive factor is that the correct values are available within a process-dependent defined time span [7]. A precise determination of the time span results from the process under consideration and can therefore not be set to a specific value.

2.2. Event

Many real-time IT systems are based on the concept of events. According to Luckham et al. [8], an event can be defined as "everything that happens or is considered to be happening." For example, changes in sensor values or production failures can be considered as events. Furthermore, the incidents can be real or simulated as virtual events [8].

Event-driven computing can provide a timely response and a good throughput while leveraging availability and scalability properties of the underlying distributed infrastructure [1].

For the field of production processes, real-time applications involve sensors that detect and report events, which in turn need to be analyzed in order to detect patterns that signify opportunities or threats. For other applications it is essential to monitor events capturing actions of users and detect relevant patterns associated with their activities and subsequently record this information or generate alerts [1].

2.3. Domain Experts

CEP is working on the basis of predefined rules and patterns, which have to be defined and expressed in an executable Event Pattern Language (EPL). The domain experts have knowledge about the relations and dependences within the production processes. However, it is challenging for them to express these complex events in form of event patterns and communicate with technical CEP experts [9].

Therefore, one way to solve this problem is to use modeling methods in order to derive event patterns from the relations within the event model and moreover to translate the knowledge into an EPL [10].

The relevant hierarchical and temporal relationships and limitations need to be described in a user friendly way on the one hand but also largely standardized in order to derive rules and patterns for the event processing on the other.

Eligible modeling languages for event-oriented systems are the Business Process Model and Notation (BPMN) version 2.0, the Decision Model and Notation (DMN), the Event-driven Process Chain (EPC), and activity diagrams of the Unified Modeling Language (UML).

However, since none of these modeling approaches meet the requirements of an event-oriented system, specific modelling notations for CEP have been developed (EPMN, CEPMN, EDM). For more information see Vidačković [11], Gabriel et al. [12], and Krumeich et al. [13]. All these modeling notations concentrate on the automatic transformation from a business process model to an executable EPL.

2.4. Data Mining based Approach

"Data mining has the goal of extracting knowledge from data" [14]. Runkel defines knowledge as "interesting patterns that are generally valid, not trivial, new, useful and understandable" [15].

Basically, there are four main application classes within data mining. Clusters want to find matches or similarities in data. The classification is often used for adding new data to existing groups. The numerical prediction can be understood as a more precise classification, which has a forecast value output as a target, however. Lastly, there is the frequently used association analysis with the aid of which interactions and connections can be found between the different data.

Mehdiyev et al. present an approach for using machine learning techniques for the determination of rule patterns [4]. They use rule-based classifiers as machine learning algorithms in order to replace experts in generating rule patterns and conducted an empirical study to investigate the applicability of rule induction algorithms for sensor data. The performance was measured using various error measures, classification accuracy, and Kappa values. The result is that rule-based classifiers can be used for detecting rule patterns in CEP systems. The highest performance, in terms of classification accuracy, error rates, and Kappa values was obtained by the PART algorithm [4].