Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 63 (2017) 557 – 562

www.elsevier.com/locate/procedia

The 50th CIRP Conference on Manufacturing Systems

Context Aware Process Mining in Logistics

Till Becker*[ab], Wacharawan Intoyoad[ab]

[a]Production Systems and Logistic Systems, Department of Production Engineering, University of Bremen, Germany
[b]Bremer Institut für Produktion und Logistik, University of Bremen, Germany

* Corresponding author. Tel.: +49-421-218-50176; fax: +49-421-218-9850176. *E-mail address:* tbe@biba.uni-bremen.de

Abstract

Processes in manufacturing and logistics are characterized by a high frequency of changes and fluctuations, caused by the high number of participants in logistic processes and the variety of goods handled and services offered. This dynamic behavior particularly requires well documented processes, but at the same time it also complicates process documentation. Manually documented processes can, e.g., miss details of alternative process branches, and the continuing change in logistic processes renders the documentation quickly outdated. A possible solution approach is to automate the documentation of processes. This automated documentation, if based on transaction and master data from the IT systems connected to the logistic process, is called Process Mining. Being a subfield of Data Mining, Process Mining extracts sequences of activities from event logs in databases. Logistics has the opportunity to greatly benefit from the application of Process Mining, because the identification and tracking of goods in the supply chain involves many IT systems. However, the IT landscape in logistics is heterogeneous, because the data are scattered among different specialized systems for various purposes (e.g., warehousing, transportation planning, and billing) of different companies. Due to this lack of standardization, the data cannot simply be analyzed by a predefined routine. Therefore, additional information beside the stored data should be taken into consideration. Machine Learning offers the possibility to classify single items in large data sets and to categorize these items with regard to the context they are in. Adding context awareness to unstructured event data in logistics has the potential to improve the results of Process Mining. Our research investigates how to apply context awareness based on Machine Learning in Process Mining for logistic processes and demonstrates its performance in a logistic scenario.

Peer-review under responsibility of the scientific committee of The 50th CIRP Conference on Manufacturing Systems

Keywords: Logistics, Process Mining, Context Awareness

1. Introduction

Logistic and production systems are characterized by a high degree of complexity, dynamics, and uncertainty. The large numbers of participants in logistics processes as well as the variety of goods and service handled cause this complexity. At the same time, technological advances in combination with the globalization of activities in manufacturing lead to frequent innovation and shorter life cycles of products [1]. Therefore, logistic systems are required to adapt quickly to changes so that companies are able to gain competitive advantages [2], [3]. Enabling this reactivity is challenging due to the high level of dynamics and uncertainty in most logistic systems. Being aware of the actual logistic processes and deviations from the planned processes in a dynamic environment is essential for companies in order to gain additional flexibility [2]. As a result, the documentation of processes is likely to become obsolete and lead to a mismatch between the real process and the documented process.

Process Mining is a potential tool to solve the challenges of keeping the process documentation up to date. It is the technique to work on business processes by discovering, monitoring, and improving actual processes by extracting knowledge from event logs [4]. The knowledge from Process Mining can be used to optimize the current processes which leads to higher competitiveness, e.g., by cost reduction, management time reduction, and the elimination of errors [4]. Process Mining has been successfully deployed to gain insight knowledge of business processes in a variety of application domains such as healthcare, education, software development,

Peer-review under responsibility of the scientific committee of The 50th CIRP Conference on Manufacturing Systems
doi:10.1016/j.procir.2017.03.149

and also logistics and production [5] [6]. Nevertheless, most studies examine data from simulated scenarios or for a single case. This paper deploys the Process Mining technique to analyze a larger number of heterogeneous logistic processes provided in datasets in order to investigate to what extent Process Mining can be automated.

Previous studies of Process Mining focus on extracting event logs to discover processes, check conformance, and enhance the processes [7] [6] [5]. However, Process Mining can go beyond analyzing historic data. There is important context data which is relevant in a logistic process. In order to further develop the Process Mining technique, context awareness is added to discover specific contexts based on the particular task. The aim of this paper is to investigate to what extent context information can support Process Mining in heterogeneous processes in manufacturing and logistics. In the subsequent section, we briefly review the key ideas of Process Mining, context awareness, and clustering. Section 3 proposes and tests an approach to improve Process Mining by adopting clustering techniques in the data pre-processing to manage event logs of high volume and with high variability in the records [8].

2. Related Work

2.1. Process Mining

Business process modelling is a visual way of presenting operations in organizations, e.g., entities, activities, and their relations [9], [10]. It has been widespread used within organizations. A number of research activities is devoted to process modeling techniques and tools [11]. Process Mining is the link between data mining, business process modeling, and analysis [12]. Software vendors more and more start to integrate Process Mining functionality to their tools [12]. Process Mining has furthermore been recognized as a potential tool for analyzing and improving processes also in logistics and manufacturing [5]. It exploits the knowledge from event logs which are generated by information systems. Making use of this knowledge is beneficial for process improvement [4]. Most Process Mining research is applied in the areas of healthcare, software development process, and online education [13].

In logistics and manufacturing, business process modelling plays a vital role [4], [11]. As a consequence of the complexity and rapid changes of the logistics processes, existing process models quickly become obsolete and can no longer serve their initial purpose. Process Mining was already proposed to be integrated into a wider Process Maintenance Framework to deal with heterogeneous logistic processes and automated process model creation effectively [5]. Process Mining in manufacturing and logistics has been tested already with simulated logistic data. It has been proven to be able to mine the actual logistic process, yet it needs improvements for a regular practice such as how to manage heterogeneous data sources [5]. However, a comprehensive understanding and procedure of Process Mining in logistic and manufacturing is still missing.

2.2. Context-Awareness in Logistics

A context-aware system is a system that is able to provide suitable service or information with regard to a user's task [14], [15]. There are many authors who define the meaning of context differently. In this paper we follow the definition of Dey (1999) as "Context is characteristics which describe the situation of an entity. An entity refers to a person, place, or object which is related to the interaction between user and application" [16].

The logistics and manufacturing domains are ideal candidates to include context awareness features, because industrial activities happen in very heterogeneous environments with a multitude of information available beside the actual observed process. Context data can be time, location, and frequency of events as well as related communication, tools, devices, or operators.

The Process Mining technique utilizes event logs as the primary input data. These events log represent sequentially recorded events. Each event refers to a single activity (a step in some process) and it is associated with a particular case (a process instance) [12]. All Process Mining techniques use these activities and cases for analysis. However, event logs may store additional data such as timestamps, size of order, related persons, etc. [12] Furthermore, additional context information can be extracted from event logs when moving away from investigating single log entries and proceeding to merging information on a higher level, e.g. grouping similar events and counting their frequency of occurrence. In our paper, we include the frequency of a process and its overall cycle time as the relevant context information as part of a logistics and manufacturing process. In Section 3 we will illustrate the influence of these pieces of context information on Process Mining in logistics and manufacturing.

2.3. Clustering Analysis

The purpose of data clustering is to group a set of objects based on their similarity of attributes and their dissimilarity to other groups [17]. In Process Mining, clustering analysis can be applied in the data pre-processing step to handle large amounts of data [18]. Trace clustering in Process Mining based on the generic edit distance has been deployed in research on Process Mining in the healthcare domain. The results illustrate that clustering techniques can be effective to handle spaghetti-like processes in the healthcare domain [18].

In the logistics and manufacturing domain, event logs are complex, heterogeneous, and may contain outliers. Furthermore, the sequence of activities or events is sensitive. Generally, additional information other than the events themselves can be retrieved from event logs, i.e. the number of events per case, unique events, unique paths, and the similarity among cases. The clustering method we apply in this paper is k-medoids, which is considered as a suitable method for clustering processes in logistics and manufacturing.

The k-medoids clustering method is a variant of the popular k-means algorithm, and they both are classified as partitioning clustering algorithms [19], [20]. In contrast to k-means, k-medoids is able to work with nominal attributes, because it uses