Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 124 (2017) 224–231

Procedia Computer Science

www.elsevier.com/locate/procedia

4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

Scalable indexing algorithm for multi-dimensional time-gap analysis with distributed computing

Riska Asriana Sutrisnowati[a], Bernardo Nugroho Yahya[b], Hyerim Bae[c,*], Iq Reviessay Pulshashi[c], Taufik Nur Adi[c]

[a] *Pusan National University Dong-Nam Grand ICT R&D Center, South Korea*
[b] *Department of Industrial & Management Engineering, Hankuk University of Foreign Studies, South Korea*
[c] *Department of Industrial Engineering, Pusan National University, South Korea*

Abstract

Data has been become increasingly and abundantly available as industries have applied numerous information systems to automate business process execution. Process mining focuses on discovering knowledge from historical data repositories to support better decision making, mostly from a single perspective. Multi-perspective, or multi-dimensional process mining becomes an open issue in process mining working groups, since the nature of event logs vary according to its domain, and a single process model might not be justified. Notwithstanding the previous work and the multi-dimensional process mining approaches developed therein, the contents of iterative indexing method and platform-dependent computational issues cause problems on scalability and usability respecting real world implementation. In response to such problems, the present study formulated a scalable indexing algorithm for multi-dimensional process analysis with distributed computing. A new solution is applied wherein we index only attributes inside the selected events and show only a reduced graph of long-duration gaps between events. The implementation is done with an independent online analytical tool. Additionally, case study of an actual port is provided to illustrate and alidate our method.

Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017.

Keywords: process mining, multi-dimensional process view, distributed computing, visualization, port logistics

1. Introduction

Process mining has matured as a researched discipline for discovery of process-perspective knowledge. It generally relies on event logs, which are collections of events or instances recorded during process execution. A process instance is an actual occurrence of a process, which itself is a series of executed activities as defined in the process model. That instance can be identified in application-specific perspectives. Those instances conceive patterns and significant knowledge that can be displayed in the form of a process model. The discovered model, however, represents

* Corresponding author. Tel.: +82-51-510-2733 ; fax: +82-51-512-7603.
E-mail address: hrbae@pusan.ac.kr

1877-0509
Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017
10.1016/j.procs.2017.12.150

only a single perspective on a process in a particular domain. Meanwhile, real-world practices consisting of multiple perspectives significantly influence the behavior of a process.

Port logistics is a complex process wherein a variety of equipment, activities and procedures are executed daily based on an intricate plan. The activities in the process are interconnected, such that if one activity is executed late, the succeeding activities can be delayed. Such delays can incur container-handling operation efficiency problems and, therefrom, significant costs. The initial time-gap analysis lists the time gap between events and inspects those events showing time-gap values exceeding the normal execution time period. Often the cause of tardiness is hidden. For example, a container cannot be handled at its scheduled time, because it has to wait for the yard crane to complete its previous job sequence. Separate analysis, unfortunately, is far from trivial, and poses a challenge to comprehensive process mining analysis. Here, it becomes apparent that it is essential to analyze a container handling operation as a whole. Suppose that we have two perspectives, a container and yard crane; we need to look at the container process and, at the same time, we have to look at what actually happens in the yard crane process.

1.1. Objectives and Contributions

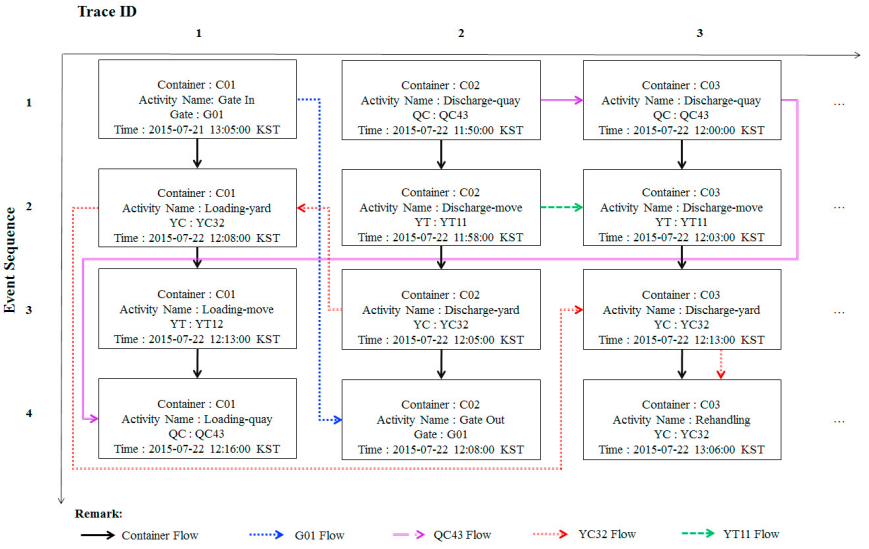

Fig. 1. Illustration of multi-dimensional perspective for event logs

As the result of industries' application of information systems and automation techniques to run their business process, there has been an explosion of data. Especially in the logistics sector, increasing vessel size has lead to increasing numbers of containers to be shipped [5]. The resultant information explosion has rendered the previous approaches to become obsolete. A solution is necessary to ensure the applicability of the proposed approach to real-world practices. This study, accordingly, aims to develop a scalable indexing algorithm for multi-dimensional time-gap analysis. In addition, to ensure applicability in the real-world practices, distributed computing is utilized to generate the index. The plugin, finally, is implemented in the online analytical tool known as BAB (Best Analytics in Big Data) [11].

The contributions of this paper are as follows. First, it enriches the literature on multidimensional process mining by developing a scalable indexing method for analysis. Second, the implementation using distributed computing assists practitioners' analysis of data in a reasonable way. Third, the demonstrated experimentation shows the applicability of the proposed approach to business environments such as logistics. Although the running example is a port-logistics-related process, it is applicable to other field (e.g., semiconductor manufacturing) as well.

This paper is organized as follows. Section 2 explains the multi-dimensional perspective and the distributed indexing schema, respectively. Section 3 shows our implementation, experimental result and case study. Section 4 briefly covers the related works, and, finally, section 5 draws conclusions.