Information Systems 37 (2012) 117–141

Contents lists available at SciVerse ScienceDirect

Information Systems

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

Process diagnostics using trace alignment: Opportunities, issues, and challenges

R.P. Jagadeesh Chandra Bose [a,b,*], Wil M.P. van der Aalst [a]

[a] Department of Mathematics and Computer Science, University of Technology, Eindhoven, The Netherlands
[b] Philips Healthcare, Veenpluis 5–6, Best, The Netherlands

ARTICLE INFO

Available online 7 September 2011

Keywords:
Diagnostics
Conformance
Alignment
Execution patterns
Process mining

ABSTRACT

Business processes leave trails in a variety of data sources (e.g., audit trails, databases, and transaction logs). Hence, every process instance can be described by a trace, i.e., a sequence of events. Process mining techniques are able to extract knowledge from such traces and provide a welcome extension to the repertoire of business process analysis techniques. Recently, process mining techniques have been adopted in various commercial BPM systems (e.g., BPM|one, Futura Reflect, ARIS PPM, Fujitsu Interstage, Businesscape, Iontas PDF, and QPR PA). Unfortunately, traditional process discovery algorithms have problems dealing with less structured processes. The resulting models are difficult to comprehend or even misleading. Therefore, we propose a new approach based on *trace alignment*. The goal is to align traces in such a way that event logs can be explored easily. Trace alignment can be used to explore the process in the early stages of analysis and to answer specific questions in later stages of analysis. Hence, it complements existing process mining techniques focusing on discovery and conformance checking. The proposed techniques have been implemented as plugins in the ProM framework. We report the results of trace alignment on one synthetic and two real-life event logs, and show that trace alignment has significant promise in process diagnostic efforts.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Today's information systems are recording an abundance of event logs. Events may be scattered over various data sources, e.g., the database of a hospital containing information about patients, the audit trails of a BPM (business process management) system, the transaction log of an airline's web service, and the job lists of a printer. By ordering events belonging to each process instance, we obtain a collection of *traces* also referred to as an *event log*. Process mining techniques can extract non-trivial knowledge and interesting insights from these event logs and exploit these for further analysis [1]. For example, it is possible to discover a process model from an event log. Process mining is not limited to control-flow and may also be used to discover other perspectives, e.g., the social network of the people working on the process or decision trees explaining the decision points in the process. Process mining is not limited to discovery and also includes conformance checking (identifying points where process instances deviate from expected and/or normative behavior) and model enhancement (e.g., enriching models based on event logs).

The lion's share of process mining research has been devoted to control-flow discovery. In this paper, we focus on complementary techniques to *diagnose processes* that are based on the *alignment of traces*. This work is inspired by practical experiences using classical control-flow discovery techniques. We (our research group) have applied process mining in over 100 organizations and our experiences show

* Corresponding author at: Department of Mathematics and Computer Science, University of Technology, Eindhoven, The Netherlands.
Tel.: +31 40 2475909.
E-mail addresses: j.c.b.rantham.prabhakara@tue.nl, jcbose@gmail.com (R.P.J.C. Bose).

0306-4379/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.is.2011.08.003

that processes tend to be less structured than expected. Moreover, event logs tend to be far from complete, i.e., just a fraction of the possible behavior indeed occurs. Therefore, it may be too ambitious to construct a process model that also says something about unseen cases, i.e., future process instances that will leave traces that were not seen before. If processes are less structured and event logs are incomplete, then it is better to *carefully inspect the event log by grouping and aligning the traces found in the event log*. First, we group similar traces in clusters [2]. Second, we visualize these clusters by aligning the traces. *By aligning traces we can see the common and frequent behavior, and distinguish this from the exceptional behavior.*

Our approach is inspired by biological sequence alignment [3]. Sequence alignment is an essential tool in bioinformatics that assists in unraveling the secondary and tertiary structures of proteins and molecules, their evolution and functions, and in inferring the taxonomic, phylogenetic or cladistic relationships between organisms, diagnoses of genetic diseases, etc. [4,5]. Process mining also deals with sequences, i.e., traces of events stored in event logs. Multiple sequence alignment has been a subject of extensive research in computational biology for over three decades. There are still many open problems, e.g., dealing with ever increasing size and complexity of the data sets, misalignments and alignment errors [6,7], obtaining accurate alignments of non-coding and non-transcribed sequences [8], integrating disparate sources of knowledge when performing alignments (knowledge-based multiple sequence alignment) [9,10], etc. Moreover, there are various new challenges when adopting biological sequence alignment to trace alignment in the context of business processes, e.g., traces in an event log in process mining need not be from a coherent set of cases and can be of different lengths, scoring matrices of substitution and insertion/deletion of activities need to be defined/derived, the size of the alphabet (number of activities) can be large of the order of a few tens or hundreds, etc. The topic of trace alignment has not been explored before. We will show that it can be used to answer a variety of questions and that it is a welcome addition to the repertoire of process mining techniques.

To illustrate the importance of trace alignment, consider the dotted chart [11] in Fig. 1. Every line corresponds to a process instance and every dot corresponds to an event. The color of the event indicates the activity that was executed. The horizontal dimension describes time. Note that the dots are not aligned. Hence, it is difficult to see common patterns among different cases. Events are positioned based on their timestamps rather than the activity that was executed and similarities to events in other instances. Fig. 2 shows the effect of trace alignment. Both dimensions are used in a similar fashion as before, i.e., every row corresponds to a process instance and time increases from left to right. However, now the horizonal position is based on *logical* time rather than *real* timestamps. When two rows have the same activity name in the same column, then the corresponding two events are very similar and are therefore aligned. Note that the same activity can appear in multiple columns. By reading a row from left to right, we can see the sequence of activities (i.e., the trace) that was executed for a process instance. Process instances having the same trace can be grouped into one row to simplify the diagram. Clearly, Fig. 2 is much more informative than Fig. 1 if one is interested in commonalities and differences among traces.

The challenge is to find an alignment that is as simple and informative as possible. For example, the number of columns and gaps should be minimized while having as much consensus as possible per column. Obtaining such high quality alignments turns out to be a challenging problem.

The remainder of this paper is organized as follows. In Section 2, we introduce the basic notations used in the paper and list the questions we would like to answer. Section 3 introduces the concept of trace alignment and discusses the techniques for finding alignments. In Section 4, we propose a framework for finding alignments over a set of traces. In Section 5, we present techniques for

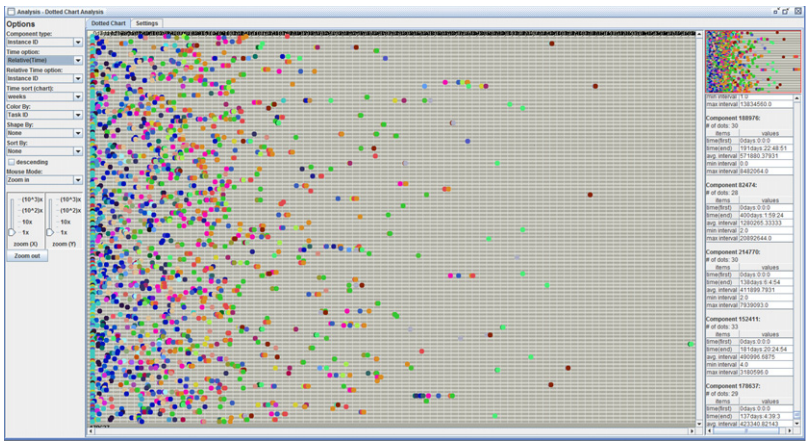

Fig. 1. Dotted chart showing all events in one glance. Consider a dot at position (x,y) having a color c. This dot refers to an event for the process instance corresponding to y (i.e., a horizonal line) that occurred at time x (i.e., time increases from left to right). The color describes a property of the event; in this case the activity that was executed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Download English Version:

https://daneshyari.com/en/article/397435

Download Persian Version:

https://daneshyari.com/article/397435

Daneshyari.com