Information Sciences 328 (2016) 558–576

Contents lists available at ScienceDirect

Information Sciences

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

Process mining with token carried data

Chuanyi Li [a,b], Jidong Ge [a,b,*], Liguo Huang [c], Haiyang Hu [b,d], Budan Wu [b], Hongji Yang [e], Hao Hu [a], Bin Luo [a]

[a] State Key Laboratory for Novel Software Technology, Software Institute, Nanjing University, Nanjing 210093, China
[b] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[c] Department of Computer Science and Engineering, Southern Methodist University, Dallas, TX 75275-0122, USA
[d] School of Computer, Hangzhou Dianzi University, Hangzhou 310018, China
[e] Centre for Creative Computing (CCC), Bath Spa University, England, UK

ARTICLE INFO

Article history:
Received 7 November 2014
Revised 20 June 2015
Accepted 31 August 2015
Available online 8 September 2015

Keywords:
Process mining
Process discovery
Workflow net
Petri net
Token
Token log

ABSTRACT

Process mining is to discover, monitor and improve real processes by extracting the knowledge from logs which are available in today's information systems. The existing process mining algorithms are based on the event logs where only the executions of tasks are recorded. In order to reduce the pre-processing efforts and strengthen the mining ability of the existing process mining algorithms, we have proposed a novel perspective to employ the data carried by tokens recorded in token log which tracks the changes of process resources for process mining in this study. The feasibility of the token logs is proved and the results of pairwise *t*-tests show that there is no big difference between the efforts that are taken by the same workflow system to generate the token log and the event log. Besides, a process mining algorithm (τ) based on the new log is proposed in this paper. With algorithm τ, the mining efficiency as well as the mining capability is improved compared to the traditional event-log-based mining algorithms. We have also developed three plug-ins on top of the existing workflow engine, process modeling and mining platforms (YAWL, PIPE and ProM) for proving the feasibility of token log and realizing the token log generation and algorithm τ.

1. Introduction

When process mining was initiated, people were often curious about the difference and commonality between data mining and process mining. Data mining is aimed to extract or mine knowledge from a large corpus of data [7], while process mining targets at mining process models from data represented by process execution workflows. Therefore, data is the basis for both data mining and process mining. However, data is stored differently in these two cases. Data employed in data mining are often stored in databases, data warehouses, World Wide Web or other information repository [7], while data used in process mining are usually stored in logs of capturing the system workflows.

Logs used in traditional process mining research are called event logs [26]. Event logs are used to record and observe the process workflows from the control-flow perspective. Events represent execution of tasks. When the event logs are used to mine processes, not only three assumptions have to be satisfied [26] (see Section 2.3) but also a significant amount of efforts need to be taken in pre-processing the logs. To eliminate the three assumptions and reduce the pre-processing efforts in mining processes

* Corresponding author at: State Key Laboratory for Novel Software Technology, Software Institute, Nanjing University, Room B924, Feiyimin Building, #22 Hankou Road, Gulou District S, Nanjing 210093, China. Tel.: +86 13813968571.
E-mail address: gjdnju@163.com, gjd@nju.edu.cn, gjd@software.nju.edu.cn (J. Ge).

http://dx.doi.org/10.1016/j.ins.2015.08.050
0020-0255/

on event logs, we here propose a new type of log named token log as the basis of process mining, which enables us to discover the process from a novel perspective of resources produced and/or consumed in the process. A token in Petri nets [15,17] is passed between transitions as the input or output of a transition. The token logs record the producer and consumer of tokens. The token logs can be directly acquired from the workflow management or other related systems just like the event logs. Section 3 defines and provides the details of the token logs.

Using token logs as the basis of process mining, we are able to save a significant amount of pre-processing effort for mining because we no longer need to distinguish different execution instances of a system and to generate the task traces, which have to be done in event-log-based process mining (see Section 2.3). The causal dependencies and parallel relationships among tasks can be directly observed from the token log. Besides, additional special structures could be discovered and the mining efficiency is improved as well. The major contributions we made in this paper are:

(1) We propose to use token logs, which can be easily generated from the information systems on the fly during the generation of event logs, as the basis of process mining. Extra data which have never been leveraged by the existing process mining methods are inserted into token logs to enhance both efficiency and ability of process mining.
(2) A novel process mining algorithm τ leveraging the data carried by token logs is proposed.
(3) Plug-ins for generating token logs and for implementing the mining algorithm are developed on top of the existing open source workflow engine and process mining tools.

Although one may argue that combining all the existing event-log-based mining algorithms (α [26], α^+ [4], α^{++} [29], β [30], λ [28]) may deliver the comparable ability for mining certain special structures as our token-log-based mining algorithm, Three major overheads resulted by combining multiple algorithms may compromise the mining efficiency:

(1) Different event logs are used by various mining algorithms. For example, β algorithm marks a type for each event and λ algorithm adds the post tasks into events, while there are no type or post tasks in the event log used by α algorithm. Hence the efforts for generating different kinds of event logs have to be taken into account.
(2) Each mining algorithm needs to be executed independently to visit all the entries in the event log for this algorithm. Combining the results of multiple mining algorithms requires multiple iterations of event logs for different algorithms, which duplicates the efforts of process mining.
(3) Models discovered by different mining algorithms may be different or even inconsistent. Then subsequent manual effort will be added to resolve the difference and inconsistencies.

Our token-log-based mining algorithm successfully avoids these overheads, which makes full use of the extra data recorded by the management systems on the fly with the generation of event logs and mines process models with the special structures correctly and more efficiently.

The rest of this paper is laid out as follows. Section 2 presents the background of process mining, Petri nets, WF-net and event log. Section 3 gives a detailed introduction to the token log, which leads the way to Sections 4 and 5. Sections 4 and 5 respectively describe the objectives of using token logs and the mining algorithm τ using token logs with a case study. Section 6 evaluates the τ algorithm and compares τ with α [26]. Section 7 discusses the limitation of applying τ. Section 8 summarizes the related work. Section 9 concludes the paper and envisages our future work.

2. Background

This section discusses the background of process mining. Section 2.1 gives an overview of process mining. Section 2.2 introduces the process modeling languages for the business processes used in this paper. Section 2.3 reviews the event logs.

2.1. Process mining

The term process mining is used for the method of distilling a structured process description from a set of real executions in [26]. Process mining is useful for at least in two aspects. First of all, it could be used as a tool to find out how people and/or procedures really work. Second, process mining could be used for delta analysis, i.e., comparing the actual process with some predefined process [26]. Based on the fact that (i) the advancements of multi-core and parallel technology resulted in a spectacular growth of the digital universe [2] (ii) the growth of a digital universe that is well-aligned with processes in organizations makes it possible to record and analyze more events [3], process mining has become one of hot topics in workflow technology.

Later in [22], the intension of term *process mining* is officially broadened to the idea of *process discovery*, *conformance checking* and *enhancement* [18,20,22]. Process discovery [21] takes an event log as input and produces a model without using any *a priori* information and this is the most prominent process mining technique. Conformance checking [18] aims to identify whether the reality, as recorded in the event logs, conforms to the model and vice versa. The idea of process enhancement is to extend or improve an existing process model leveraging information of the actual process execution recorded in event logs. This paper is focused on process discovery.

2.2. Petri nets and WF-net

Workflow nets (WF-nets) [25] are used in this paper for modeling workflow processes and the WF-net is a kind of Petri net. The graph in Fig. 1 is a WF-net and also a Petri-net.