Cross-organizational collaborative workflow mining from a multi-source log

Qingtian Zeng [a,*], Sherry X. Sun [b], Hua Duan [a], Cong Liu [a], Huaiqing Wang [c]

[a] College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China
[b] Department of Information Systems, City University of Hong Kong, 83 Tat Chee Avenue, Hong Kong
[c] Department of Financial Math and Financial Engineering, South University of Science and Technology of China, Shenzhen, China

ARTICLE INFO

ABSTRACT

Today's enterprise business processes become increasingly complex given that they are often executed by geographically dispersed partners or different organizations. Designing and modeling such a cross-organizational workflow is a complicated, time-consuming process and requires that a designer has extensive experience. Workflow logs captured by different cross-organizational systems provide a very valuable source of information on how business processes are executed in reality and thus can be used to derive workflow models through process mining. In this paper, we investigate the application of process mining for workflow integration based on the concept of RM_WF_Net, a type of Petri net extended with resource and message factors. Four coordination patterns are defined for workflow integration. A process mining approach is presented to discover the coordination patterns between different organizations and the workflow models in different organizations from the running logs containing the information about resource allocation. A process integration approach is then presented to obtain the model for a cross-organizational workflow based on the model mined for each organization and the coordination patterns between different organizations.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

With the development of the Internet and distributed computing technologies, today's enterprise business processes become increasingly complex given that they are often executed by geographically dispersed partners or different organizations. For example, a typical multi-modal transportation business process may require the coordination of several business partners including the sender, the consignor, the carrier, the shipper, the buyer and other related partners. Designing and modeling such a cross-organizational workflow require a workflow designer to have lengthy discussions with the involved workers and managers of different organizations. Therefore, creating a workflow design is a complicated and time-consuming process and there are typically discrepancies between the actual workflow processes and the processes as perceived by the management [21].

Many enterprise information systems, such as enterprise resource planning (ERP), customer relationship management systems (CRM), and workflow management systems (WfMS), usually record information of the execution of business processes in their event logs [21]. Thus, process mining, also referred to as workflow mining, has recently been proposed to distil a structured process description from the event logs of those systems [9]. The goal of process mining is to analyze event logs so as to construct a model that best describes all the recorded instances of a workflow.

Recently, cross-organizational process mining has received more attention [18]. Although many papers on process mining have been published [3,5,6,20,21], it is difficult to apply existing approaches directly to discover a model for a cross-organizational workflow. This can be attributed to the fact that the running log of a cross-organizational workflow is usually distributed on different servers owned by different partners or different organizations, while most existing process mining approaches assume that a single server is used to manage all the operations of a process [3,5,6,20,21].

To address this problem, a process mining based integration approach is proposed in this paper to obtain cross-organizational workflow models. Unlike the existing studies on process mining [3,5,6,20,21], the event logs in this study are distributed on different servers located in different organizations. The event logs contain information about resource allocation and messages exchanged, which are two important coordination mechanisms between organizations. The concept of RM_WF_Net is proposed to represent the mined workflows. An RM_WF_Net is a type of extended Petri net that allows representation of resource allocation and messages exchanged in workflows. On the basis of the RM_WF_Net model, four different coordination patterns are defined for cross-organizational workflow integration. A process mining approach is presented to discover workflow models in different organizations from the event logs containing the information on resource allocation. To discover the coordination patterns between different organizations, a middleware has been implemented to integrate the workflow running logs between different organizations. Based on the logs integrated, the coordination patterns between different organizations can be obtained by using the process mining approach. According to the workflow models in different

organizations and the coordination patterns between different organizations, a process integration approach is presented to obtain the cross-organizational workflow models. Our approach can help an organization to identify from its own perspective the overall workflow model reflecting how various organizations collaborate. Moreover, when the event log data of all relevant workflows are available, our method can help derive a complete overview of a workflow that crosses multiple organizational entities. In the paper, a multi-modal transportation business process is studied as an example to validate the proposed approaches. However, it is only possible to obtain the overall workflow model from the individual perspective of each organization involved in the multi-modal transportation business process, and to understand the collaboration between the organization's private workflow and other workflows. In many tightly coupled application cases, the approaches proposed in this paper can be used to obtain the overall model for a cross-organizational collaborative workflow; consequently, an example of hospital workflow model integration is used to demonstrate the approaches proposed in the paper.

The remainder of the paper is organized as follows. Section 2 presents a review of the related work. Section 3 gives a framework for cross-organizational workflow mining and integration. An example of a cross-organizational workflow used to validate the proposed approach in the paper is also given in Section 3. Section 4 defines the RM_WF_Net model for a cross-organizational workflow and the coordination patterns for different organizations. Section 5 presents the workflow mining for one single organization. Section 6 presents the cross-organizational workflow integration using process mining. Section 7 presents an experiment evaluation for the proposed approach. Section 8 concludes the paper with our contributions and future research directions.

2. Related work

In this section, we compare our work with related work on process mining and process integration.

2.1. Process mining

Process mining has been proposed as a tool for analyzing business processes based on event logs [21]. Many papers on process mining have been published [20,21]. A survey of existing process mining approaches was provided by [20]. Most process mining methods start with event-based logs, i.e., sequences of instantaneous events. In event-based logs, it is possible to identify the start and end events to infer the duration of activities. In [1], methods were proposed for automatically deriving a formal model of a process from a log of events recording the executions of a process. However, these studies were limited to a sequential process structure. By using the same type of process logs as input, a workflow model based on Petri nets was derived to capture timing information such as minimal, maximal, and average time spent in different stages of a process [19]. The timing information is associated with a type of node called a "place" in the derived Petri net-based workflow model.

Recently, [22] demonstrated the application of process mining for the feasibility of conformance checking of service behavior, that is, comparing message logs with service behavior specifications to detect and to quantify deviations. In our previous work [2], a mining approach was proposed to discover the structural and temporal model for a workflow from its timed running log. According to the reachability graph of the discovered Petri net model, all running schemas of a workflow can be generated, defining the temporal constraints among activities. In [15], based on examination of the model resulting from process mining, we demonstrated how to determine the minimum time to finish a workflow and how to partition the workflow in order to achieve efficient server usage. Most process mining tools have focused on the control-flow perspective [1.19], and workflow systems are often built on client/server

architecture with the event log recorded and collected from a centralized log database. In the centralized paradigm, the mining algorithm is applied to the centralized log database to obtain the process model directly [3,5,20,21]. In a cross-organizational business process, information of the process execution is recorded and stored by individual databases belonging to different organizations. There is no centralized log database to store all the execution information of a workflow; thus, it is difficult to discover the model for a cross-organizational business process directly using existing process mining approaches.

2.2. Process integration

Business process integration and management is a critical element in enterprise business transformation [28], and business process integration is considered a crucial technique for supporting inter-organizational business interoperability [23]. Many studies have addressed process integration [7,12]. In [7], the authors analyzed the existing technologies and standards, and proposed a reference model for B2B workflow integration. In [12], a solution was proposed to the integration design of material flow management in an e-business manufacturing environment.

According to the studies addressing process integration [7,12], the approaches to integrating processes include, but are not limited to, model-driven [28], integration points based [23], contract-based [25], and workflow merge approaches [14]. A set of model-driven business integration and management methods, frameworks, supporting tools, and a runtime environment was developed to study the efficacy of model-driven business process integration and management [28]. In [23], the business process integration points were classified and a priority evaluation method between different points was presented. Based on the concept of the integration point, typical atomic integration patterns and some composite patterns were proposed. Ref. [25] presents contracts that encapsulate commitments laid down as a set of obligations to coordinate and control the interaction between business workflows. A business contract specification language was introduced to formally link the specification of business object based workflow systems. In [14], the concept of workflow merge was described and methods for merging business processes were proposed. Four merge categories were addressed: sequential, parallel, conditional, and iterative, and the corresponding algorithms for performing these four operations were reported.

As models for workflows, Petri nets have been used for process integration [4,8,17]. In [17], a modeling approach was presented for inter-organizational workflows in terms of Petri nets and the techniques were given to verify the correctness of the inter-organizational workflows. In [8], the Petri nets extended with time and color were proposed as a formalization for managing events, in which seven basic patterns were designed to capture the modeling concepts that commonly arise in supply chains. Moreover, a complete Petri net can be built from those patterns and analyzed by using dependency graphs and simulation [8]. In [4], open Petri nets were proposed as a suitable semantic model for workflows spanning different enterprises.

From the studies reviewed here [8,14,17,23,25,28], a model integration for workflow models is one of the approaches most addressed. In this paper, we also mainly discuss the model integration for workflow processes. Four coordination patterns are defined for cross-organizational workflows, including coordination with synchronized activities, coordination with messages exchanged, coordination with shared resources, and coordination with abstract procedures, which are used to integrate processes. To address this problem, a process mining based approach is introduced for cross-organizational process integration. To integrate processes, one important assumption is that there must be a model for each organization [8,14,17,23,25,28], otherwise it is impossible to integrate a process for a cross-organizational business workflow. To address this problem, process-mining technology is applied in this paper to obtain the workflow model within each