ARTICLE IN PRESS

Computers & Industrial Engineering xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Domain-driven actionable process model discovery

Bernardo Nugroho Yahya [a], Minseok Song [b,*], Hyerim Bae [c], Sung-ook Sul [d], Jei-Zheng Wu [e]

[a] Department of Industrial and Management Engineering, Hankuk University of Foreign Studies, Oedae-ro 81, 17035 Yongin, South Korea
[b] Department of Industrial & Management Engineering, POSTECH (Pohang University of Science and Technology), 77 Cheongam-Ro, 37673 Pohang, South Korea
[c] Department of Industrial Engineering, Pusan National University, Pusandaehakyo-ro 63 bongil 2, Geumjong-Gu, 46241 Busan, South Korea
[d] Total Soft Bank, Ltd., 66-39 Bansong-ro 513, Haeundae, 48002 Busan, South Korea
[e] Department of Business Administration, Soochow University, 56 Kueiyang Street, Section 1, Taipei 100, Taiwan, ROC

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
Process mining
Business process
Proximity score
Users' knowledge
Integer linear programming

A B S T R A C T

Process discovery is a type of process mining that constructs a process model from the event logs of an information system. The model discovered using process discovery techniques and the process as perceived by users will always differ in some ways and to some extents. In particular, less structured process, such as operational process in business and manufacturing, often result overly confusing, spaghetti-like, process models caused by the inherent complexity of the process. As a result, the mined model has many limitations for providing the users with explicit knowledge that can be directly used to influence behavior for the user's interest. Explicit knowledge, as later called by actionable knowledge, is an important representation on measuring the interestingness of mined patterns. This actionable knowledge, which is incorporated with users' background knowledge and based on some notions of actionable rules, can result an actionable process model. Undoubtedly, domain experts, who know the process well, play a key role to enhance the mined model into an actionable model by their involvements during the discovery process. This paper presents a discovery method to obtain an actionable process model that is based on both the event relation in the log and users' knowledge to improve the incompatibility of the traditional process mining approaches. Users can set their knowledge in terms of constraints. Unlike the existing approach, the proposed approach synthesizes the activity proximity and attempts to extract behavior satisfied by the constraints which may be hidden in the event logs for resulting an actionable process model. In addition, the proposed method is used in order to achieve a sound process model when the existence of the constraints does not satisfy the workflow soundness property. The method was implemented in the ProM framework and tested on a real process.

1. Introduction

A process model is a means of communication that facilitates stakeholders' understanding of complex business processes (Kawalek & Kueng, 1997). Stakeholders typically formalize a process model using traditional top-down *a priori* design approaches. Although they have their own understandings about process model flow, a real application might result in different sequences. Process discovery, a type of process mining, constructs a process model from information systems' event logs. Such a discovered model, most probably, will differ from the process as perceived by stakeholders, since the discovery techniques can automatically derive a process model based on the statistics hidden in the event logs. As a result, stakeholders, most of the time, find difficulties on understanding the extracted patterns from the discovered model. Since the extracted patterns from process discovery are mostly related to the technical interestingness for process analysts, the business interestingness, called as actionable model, is often neglected.

In the field of knowledge discovery, the involvement of users, i.e., stakeholders, takes place to examine the extracted patterns for their subjective measures, as a measure of interestingness (Silberschatz & Tuzhilin, 1996). In the subjective (user-oriented) point of view, the extracted patterns can be classified as *unexpected* – a pattern which is "surprising" to the user- and *actionability* – a pattern that the user can act on it to his advantage. Therefore, actionable model can be considered as a result the involvement of users to afford important grounds to business decision makers. There was a highlight on the topic of actionable knowledge discovery that it will be one of the Grand Challenges for extant and future data mining (Ankerst, 2002; Fayyad, Shapiro, & Uthurusamy,

* Corresponding author.
E-mail addresses: bernardo@hufs.ac.kr (B.N. Yahya), mssong@postech.ac.kr (M. Song), hrbae@pnu.ac.kr (H. Bae), sosul@tsb.co.kr (S.-o. Sul), jzwu@scu.edu.tw (J.-Z. Wu).

http://dx.doi.org/10.1016/j.cie.2016.05.010
0360-8352/

2003). Since then, actionable knowledge discovery had been enhanced to satisfy the real business needs (Cao & Zhang, 2007; Cao et al., 2010; Yang, Yin, Ling, & Pan, 2007) because the real world problems are usually highly constraint-based and tightly embedded in domain-specific business rules. Nevertheless, the role of users, i.e., stakeholders, in process mining has been given little attention so far.

Process mining is presumed as an automated process to discover a model using algorithms and tools without human involvement. Process mining, which includes aspects such as discovery, conformance and enhancement, have been proved to solve business and industrial applications (Cho, Song, & Yoo, 2015; Goedertier, Weerdt, Martens, Vanthienen, & Baesens, 2011; Rebuge & Ferreira, 2012; van der Aalst et al., 2007; Wang, Caron, Vanthienen, Huang, & Guo, 2014). This situation partly results from the scenario that process mining is based on event logs where process mining algorithms extract patterns from data based on specific conditions, e.g. a log that has only start and end. Although a mined model can show patterns such as "hidden" and "unused" flow, it still remains some questions about how to extract specific patterns, i.e., actionable patterns. Some efforts have been considered to improve the mined model using proprietary filters. For example, the work on process mining of test processes in semiconductor manufacturing mentioned the necessity to apply various filtering techniques to show different level of abstractions (Rozinat, de Jong, Gunther, & van der Aalst, 2009). However, filtering the log requires some works, e.g., it is not interactive, and potentially loses the connection to the previous stage of filtering (Rozinat et al., 2009). In addition, the used of filters is intended to reduce incorrect or corrupted information in logs, i.e., noise. Since the nature of noise is assumed as either infrequent or incomplete information, the filter techniques may be insufficient to tackle a positive noise, e.g., purchasing a diamond that occurs infrequently but invaluable, that may provide business interestingness. As a result, actionable patterns, which are a part of business interestingness, are often neglected since they are hidden in large quantities of data. Thus, the involvement of users on process mining plays a key role for generating an actionable model.

In general, process discovery algorithms and techniques only focus on the discovery of model satisfying expected technical significance (Gunther & van der Aalst, 2007; Weijters, van der Aalst, & de Medeiros, 2006). Some process discovery techniques involve users to filter and configure the mined model (Bergenthum, Desel, Lorenz, & Mauser, 2007; van der Aalst, Weijters, & Maruster, 2004); however, the results remain unsound and incompatible. Other techniques related to the constraint-based approach, artificial generated negative events (AGNEs) (Goedertier, Martens, Vanthienen, & Baesens, 2009) for example, offer detailed descriptions of automatic generated artificial negative events rather than generating a model using domain experts' constraints. The work on process discovery via precedence constraints using constraint programming (Greco, Guzzo, & Pontieri, 2012) intended to involve users for discovering a model. However, the aspect of business interestingness, i.e., users' interactivity to build the constraints, had been neglected. Fundamental work on process discovery is therefore necessary to cater for critical elements in real-world applications such as expert knowledge. This is related to algorithm innovation and performance improvement that will balance technical interestingness and business interestingness.

This paper proposes a new discovery technique, called proximity miner, that is designed to discover an actionable process model with the involvement of domain experts. The main contributions of this work are developing an interactive mode process discovery tool with the involvement of users and demonstrating the effectiveness and flexibility of the proposed approach in tackling real-world process mining. In addition, this approach not only utilizes pre-processing tasks, i.e. using filtering and mining as a back-and-forth cycle procedure, but also applies post-analysis that enables users to refine mined model with a set of constraints for generating actionable process model. Actionable process model in this paper is a process model that satisfies the constraints from a domain-specific expert in terms of their value in decision making.

To some extent, the objective of this approach is similar to that of process discovery via precedence constraints, however, its strategy and contributions differ in five ways. First, it introduces an integer linear programming (ILP)-based *domain-constraints* approach that entails less computation than constraint programming (Salvagnin, 2008). Second, it presents the concept of *categorization*, which is required when the log consists of several event types. Third, it introduces *proximity*, which guarantees the inclusion of both the observed and the "hidden" behavior of the event logs in our ILP model. The use of *proximity* ensures the production of a sound model when users' background knowledge is not available ("hidden") from the log-derived graph. Fourth, this approach uses an *interactive* interface that allows users to iteratively set the constraints for refining the control-flow factors. Fifth and finally, it can be considered *an actionable model builder* that does not require either any effort in drawing from scratch or back-and-forth cycle procedure between filtering and mining tasks. The proposed approach was tested by means of a case study conducted in the port logistics process domain, and was verified on the basis of quantitative and qualitative evaluations.

The remainder of this paper is organized as follows. Section 2 discusses the relevant process mining work available in the literature. Sections 3 and 4 present a running example of the proposed discovery algorithm and the overall methodology, respectively, in demonstrating how the ILP is combined with the concept of proximity. Section 5 analyzes the main features of our approach, its implementation in the ProM framework, and the evaluation results. Section 6 concludes the paper.

2. Related work

2.1. Process mining

Process mining is a recognized and well-known technique for mining process models from logs. It includes classes such as discovery, conformance and enhancement. Various approaches to the construction of process models from logs (i.e. process discovery) and the measuring of discrepancy between logs and a given predefined process model (i.e. conformance testing) have been developed over the past few years (de Medeiros, Weijters, & van der Aalst, 2007; Gunther & van der Aalst, 2007; Kawalek & Kueng, 1997; Rozinat, de Medeiros, Gunther, Weitjers, & van der Aalst, 2007; van der Aalst, 2011; van der Aalst et al., 2004; Weijters et al., 2006). Enhancement, which also needs a prior model, is used to enrich the mined model based on some performance data (van der Aalst, 2011). The concept has been applied to some industries such as health, logistics and telecommunications (Cho et al., 2015; Goedertier et al., 2011; Rebuge & Ferreira, 2012; van der Aalst et al., 2007; Wang et al., 2014).

However, most of the applicability-related work has been concerned with the general concept of process mining, with little focus on domains. This means that when analysts are eager to analyze a domain-specific problem, domain experts must be involved to interpret the result. Some approaches to cover specific behaviors in a process in terms of particular flows have been proposed, such as non-free choice (Wen, van der Aalst, & Wang, 2007), size of discovery problems (Carmona & Cortadella, 2014), probabilistic analysis (Cook, Du, & Liu, 2004), control dependencies with conditions (Hwang & Yang, 2002), activities' lifespans (Pinter & Golani, 2004),