Information Sciences 352–353 (2016) 98–120

Contents lists available at ScienceDirect

Information Sciences

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

Evaluating and predicting overall process risk using event logs

A. Pika[a,*], W.M.P. van der Aalst[b,a], M.T. Wynn[a], C.J. Fidge[a], A.H.M. ter Hofstede[a,b]

[a] Queensland University of Technology, Brisbane, Australia
[b] Eindhoven University of Technology, Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 13 July 2015
Revised 24 January 2016
Accepted 4 March 2016
Available online 8 March 2016

Keywords:
Event log
Process risk evaluation
Mining process risk
Overall process risk

ABSTRACT

Companies standardise and automate their business processes in order to improve process efficiency and minimise operational risks. However, it is difficult to eliminate all process risks during the process design stage due to the fact that processes often run in complex and changeable environments and rely on human resources. Timely identification of process risks is crucial in order to insure the achievement of process goals. Business processes are often supported by information systems that record information about their executions in event logs. In this article we present an approach and a supporting tool for the evaluation of the *overall process risk* and for the prediction of process outcomes based on the analysis of information recorded in event logs. It can help managers evaluate the overall risk exposure of their business processes, track the evolution of overall process risk, identify changes and predict process outcomes based on the current value of overall process risk. The approach was implemented and validated using synthetic event logs and through a case study with a real event log.

1. Introduction

In order to deliver desirable outcomes in an efficient manner and minimise operational risks companies often standardise and automate their business operations. Business processes are exposed to different risks that can jeopardise the achievement of process goals in terms of cost, timeliness or the quality of outputs [14]. It is not possible to consider all process execution scenarios and to eliminate all operational risks during the process design stage [27,30] as processes are often executed in complex [35] and changeable [2] environments. Multiple cases (i.e., process instances) are often processed in parallel, e.g., multiple loan applications can be processed in a bank at the same time. The same resources may be involved in execution of these process instances. Human resources tend to make mistakes and their productivity levels can vary. Human factors are considered to be "unequivocally the single most important element that can affect project success" [31]. A case can be affected by events, either happening in the case itself or external to the case [35]. For example, when a resource is busy processing a complex case, he may neglect other cases assigned to him, hence causing delays. Sometimes an event that is not risky on its own can be risky in combination with other events, e.g., consider the arrival of an urgent case – one such case will typically not be a problem, but when multiple urgent cases arrive within a short period of time, outcomes (e.g., quality and duration) of all cases active during this period may be affected.

* Corresponding author.
E-mail address: anastasiia.pika@hdr.qut.edu.au (A. Pika).

http://dx.doi.org/10.1016/j.ins.2016.03.003
0020-0255/

Furthermore, it can be very costly to address all known process risks during the process design stage. For example, in order to decrease the likelihood of human mistakes, the result of every completed task instance can be checked by a human resource not involved in its execution ("four-eyes principle" [39]). However, this strategy is very time-consuming and is too costly to be applied in non-critical processes. Although it is not always possible to build mechanisms into a process that address all known process risks during the process design stage, it is important to evaluate the level of overall process risk during the execution of a process.

Effective risk management is crucial for organisations [8]. ISO Guide 73:2009 defines risk as the "effect of uncertainty on objectives" where effect is "a deviation from the expected – positive and/or negative" [13]. Risk identification is a critical task and one of the most challenging in the risk management process [29]. Although many risk management approaches provide high-level guidelines about risk identification strategies, they do not provide any guidelines on how to operationalise them [21,29]. Recently a few approaches have been proposed that allow us to identify some process-related risks [3,5,24,25]. Existing approaches typically analyse characteristics of individual process instances and predict risks on the process instance level, though "the handling of cases is influenced by a much broader context" [35]. They also do not consider the fact that process risk can change over time. As managing risks in individual process instances can be very costly, for non-critical processes companies may be more interested in managing overall process risk rather than dealing with risks in individual process instances. For example, a manager may wish to be alerted only when overall process risk reaches some predefined threshold rather than receiving "delay likelihood" notifications for individual process instances.

Up to now insufficient attention has been paid to the problem of evaluation of overall process risk [4], i.e., the risk that threatens the achievement of overall process goals (e.g., completing the majority of cases within a given time period or within a given budget) and can be caused by different process-related risk factors in multiple process instances. Hence, our first research question in this paper is: *how can we evaluate overall process risk at a given point in time considering different risk factors across all running process instances?* A method for evaluation of overall process risk at different points in time can help managers identify changes in overall process risk. In the case of a significant increase of overall process risk, managers may wish to look more closely at business operations in order to investigate sources of the risk increase and take corresponding actions. For example, they can discover that the overall risk increased because specific tasks in a process are often repeated or delayed. Consequently, they may decide to provide training to resources that execute these tasks or to hire additional employees.

When multiple events that increase process risks (we refer to them as risky process behaviours) happen in a process within a short period of time, many cases that are active during this period may not achieve their goals (e.g., in terms of time or quality). Let us consider as an example a service desk organisation. Within a few hours many urgent requests may be lodged, a few complex incidents may be re-opened and a special type of expertise may be required for some urgent cases. During such periods resources may become abnormally busy. Due to the fact that the resources need to solve many urgent or complex issues it is likely that many cases that are active during this period may be delayed or completed with mistakes. Our second research question we tackle in this paper is: *how can we predict aggregate process outcomes based on the current value of overall process risk?* Examples of aggregate process outcomes include: the percentage of active cases that will not produce a process outcome of a high quality or the number of active cases that will be delayed. Such a prediction may help managers to mitigate significant risks, e.g., when they learn that many running process instances are likely to be delayed or completed with mistakes, they may prioritise tasks or decide to outsource some work.

Companies often use information systems to support executions of their business processes. These information systems record information about process executions, e.g., process tasks that were executed and resources involved in the tasks. Such information can be converted to event logs. Information systems may record different data attributes, e.g., case complexity or task urgency, or these may be derived from raw data. In this paper we present an approach and a supporting tool that allows to (1) evaluate overall process risk and (2) predict aggregate process outcomes based on the analysis of information about process executions recorded in event logs (Fig. 1). As the first step, we need to be able to capture risky behaviours of a process. An input to our approach is a process model that models desired process behaviour, i.e., behaviour that does not threaten the achievement of process goals. We consider deviations from such process model as risky process behaviours. Then, we need to identify such risky process behaviours in process instances. To tackle this issue we use an existing technique for replaying an event log on a process model [6]. We devise two measures of overall process risk at a given point in time that are based on the identified risky process behaviours. To track the evolution of the overall process risk over time we generate a time series of the overall process risk values, visualise them and annotate with automatically detected change points to facilitate the analysis for a user. Time series are often used to model the evolution of different phenomena [16]. (An example of the overall process risk time series is depicted in Fig. 15.) To tackle our second research question, we generate overall process risk time series for all processes whose behaviours affect outcomes of a given process as described above; we then extract aggregate process outcome time series from an event log, learn a regression model from past process executions (using extracted overall process risk time series and aggregate process outcome time series), and predict aggregate process outcome based on the current value of overall process risk. The approach is implemented as a plugin of the process mining framework ProM[1]. We demonstrate the approach using an illustrative example and evaluate it using synthetic event logs. We also present a case study on the overall process risk evaluation using a real event log.

[1] www.promtools.org.