Event Correlation in Non-Process-Aware Systems

Ricardo Pérez-Castillo[1], Barbara Weber[2], Ignacio García-Rodríguez de Guzmán[1], Mario Piattini[1] and Jakob Pinggera[2]

[1] Instituto de Tecnologías y Sistemas de Información (ITSI)
University of Castilla-La Mancha
Paseo de la Universidad 4 13071, Ciudad Real, Spain
{ricardo.pdelcastillo, ignacio.grodriguez, mario.piattini}@uclm.es

[2] University of Innsbruck
Technikerstraße 21a, 6020, Innsbruck, Austria
{barbara.weber, jakob.pinggera}@uibk.ac.at

Abstract. Since business processes supported by traditional systems are implicitly defined, correlating events into the appropriate process instance is not trivial. This challenge is known as the event correlation problem. This paper presents an adaptation of an existing event correlation algorithm and incorporates it into a technique to collect event logs from the execution of traditional information systems. The technique first instruments the source code to collect events together with some candidate correlation attributes. Secondly, the algorithm is applied to the dataset of events to discover the best correlation conditions. Event logs are then built using such conditions. The technique has been semi-automated to facilitate its validation through an industrial case study involving a writer management system and a healthcare evaluation system. The study demonstrates that the technique is able to discover the correlation set and obtain well-formed event logs enabling business process mining techniques to be applied to traditional information systems.

Keywords. Process Mining, Event Correlation, Event Model, Case Study.

1 Motivation

Many companies today carry out a vast amount of daily operations through their information systems without having ever done their own enterprise modeling. Business process mining is a well-proven solution used to discover the underlying business process models that are supported by existing information systems. Business process discovery techniques employ event logs, which are recorded by process-aware information systems, as input. However, there are a wide variety of traditional systems without any in-built mechanism to collect events (representing the execution of business activities). To enable the application of process mining techniques to traditional information systems, mechanisms for collecting events from non-process-aware systems have been proposed. Unfortunately, business processes supported by

traditional systems are implicitly defined, correlating events into the appropriate process instance is not trivial. This challenge is known as the event correlation problem.

2 Proposal

The main proposal is a technique to obtain event logs from traditional information systems[1], which consists of four main stages: (1) The first stage is aimed at instrumenting traditional systems so that such systems can collect events during their execution. Experts identify some information such as candidate correlation attributes, whose runtime values will then be collected together with each event. This approach incorporates a set of guidelines for business experts to facilitate the instrumentation of source code. Additionally, it proposes a metric based on package dependencies to suggest experts which parts of source code should be instrumented. (2) In the second stage, the modified system is executed and events are progressively recorded. As a result, events and their respective attributes are then stored in a database in an intermediate format. (3) The third stage applies an adaptation of the algorithm proposed by *Motahari-Nezhad et al.* to the event datasets in order to discover the necessary set of attributes and conditions to correlate events. Finally, (4) the last stage applies an algorithm taking the correlation set into account to correlate each event with its corresponding process instance. As a result, a standard-format event log is obtained from the traditional system.

3 Empirical Validation Results

The most important contribution of this paper is an empirical study conducted to demonstrate the feasibility of the technique as well as its application in the industry. The study was applied with two industrial, traditional information systems: *AELG-Members*, an author management system and *CHES*, a system for collecting patient reported outcome (PRO) data. The study collected thousands of events to be analyzed after and obtain different correlation sets according to different parameters. The correlation sets were then used to generate different event logs, which were in turn used to discover business processes. Finally, the mined business process models were compared with the reference models. The analysis of the study's results shows that the technique is able to obtain event logs from traditional systems. However, the correlation set depend on the amount of collected events and the beta factor (which determines the way in which process instances are built). Although, more collected events can be good to obtain better correlation sets, the time spent on discovering such information grows according to a quadratic function. As a consequence, the work-in-progress is currently focusing on obtaining correlation sets more efficiently (to reduce the response time of the discovering algorithm) as well as the effectiveness (to provide accurate correlation sets with fewer events).

[1] Pérez-Castillo, R., B. Weber, I. García Rodríguez de Guzmán, P. Mario, and J. Pinggera, Assessing Event Correlation in Non-Process-Aware Information Systems. Software and Systems Modeling Journal, 2013. In Press.