PhD Research Plan

Title: Process mining of ERP Systems

Santiago Aguirre*

saguirre@javeriana.edu.co

*PhD student at the School of Engineering of Pontificia Universidad Javeriana.

Advisor: Carlos Parra, PhD.

carlos.parra@javeriana.edu.co

Abstract:

Having been widely adopted in the business world, ERP systems contain huge amounts of data related to the actual execution of business processes. However, few works have addressed process mining as applied in real ERP environments. These systems have a particular way of recording activities which results in an unclear display of business processes in event logs [1]. Such recording mode certainly contrasts with those of BPMS or workflow based systems, whose systematical storage of events facilitates the application of process mining techniques [2]. Thus, the main objective of the current research project is to improve the usability and understandability of process mining techniques, by developing a methodology for their application in ERP contexts, detailed in terms of specific implementation tools and step by step coordination.

**Key words:** process mining, data mining, ERP, process modeling, process redesign, methodology.

1. Research Question:

It has been 15 years since Cook and Wolf [3] attempted the first process mining practical applications. Since then, several works have been conducted in this field, most of them focusing on the development of new algorithms for the automatic discovery of business processes. In this sense, great results have been achieved, as shown by the alpha [4], heuristics [5] and genetic algorithms [6], which overcome some of the issues stated as research challenges for process mining by Van der Alst [7] in 2004.

Researchers have also applied process mining in real world situations, in order to test the newly developed algorithms on ProM-like software. Most of these applications have been conducted in workflow systems [8,9,10,11,12], where processes are automatically stored, thus facilitating the application of process mining. Nevertheless, little research has been conducted on these
techniques as applied to ERP systems, and most of the publications taking on this issue [13,14,15] coincide in the need for further research. In consequence, the first process mining manifesto [2] considered the problem of finding, merging and cleaning event data as the first challenge for process mining, and classified ERP event logs at a maturity level of three, where five is the best level, only achieved by BPMS and workflow systems.

On the other hand, there have been some methodological developments around process mining, some focusing on process diagnostics [16,17], and others combining process mining with simulation techniques [18]. However, none of them focuses on ERP systems. The process mining manifesto [2] also highlights the importance of the integration of process mining with other methodologies and types of analysis as simulation and visual analytics. This led to stating the current research question:

How can organizations with ERP systems apply process mining for analyzing their business processes in order to improve them?

Specifically, this research answers the following three research questions:

RQ1. What methodological approach should be followed to apply process mining in organizations supported by ERP systems?

RQ2. What procedures should be followed for the extraction and processing of event logs from ERP systems, when intended for the application of process mining?

RQ3. Taking into consideration the objectives pursued by the process redesign project, what process and data mining techniques should be employed?

2. Background:

Table 1 summarizes the most relevant references on process mining as applied to ERP systems, as well as those proposing methodological approaches to the application of process mining.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Summary / relation to the research</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Ingvaldsen, J. E. Semantic Process Mining of Enterprise Transaction Data. Doctoral dissertation. Norwegian University of Science and Technology. 2011</td>
<td>In this thesis the author targeted challenges in all phases of a process mining project on a SAP ERP system, from the preliminary steps of gathering and transforming raw data, to techniques for describing process quality at the business level. The author proposes the incorporation of ontologies and search technologies both to harmonize scattered event log fragments and to provide a simple and exploratory environment for business level process analysis.</td>
</tr>
<tr>
<td>b) Jansen-Vullers, M., W.M.P. Van der Aalst, and M. Rosemann, Mining configurable enterprise information</td>
<td>This paper shows an alternative process mining procedure for logs containing only data on the number of times that a process’ steps have been executed (frequencies). To be able to</td>
</tr>
</tbody>
</table>
Reference | Summary / relation to the research
--- | ---
| Data & Knowledge Engineering, 2006. 56(3):pp. 195-244. | mine such event logs the authors used SAP process workload and applied linear programming and Configurable Event-driven Process Chains (C-EPCs) to determine process flow. The authors conclude that it is necessary to do more research on other ERP systems’ logs that might differ from SAP. |
c) Jans, M., et al., A business process mining application for internal transaction fraud mitigation. Expert Systems with Applications, 2011. 38(3): p. 13351-13359. | In this paper the authors present the application of process mining to address internal transaction fraud. The diagnosing methodology proposed by Boscaya [16] was applied to a case study consisted in a SAP supported procurement process. The authors conclude that process mining could be used in real scenarios, but it is important to improve tools, methodologies and visualization techniques. |
d) Maruster, L. and L. Van Beest, Redesigning business processes: a methodology based on simulation and process mining techniques. Knowledge and Information Systems, 2009. 21(3): p. 267-297 | In order to support process redesign, the authors propose a methodology that combines simulation and process mining. The proposed methodology is applied to three case studies, one of them operating with an ERP system. This method constitutes an important step towards the application of process mining in real scenarios and can be complemented with tools and techniques for extraction and cleanup of event logs, since the proposal is not very detailed in that regard. |
e) Bozkaya, M., J. Gabriels, and J. Werf. Process Diagnostics: A Method Based on Process Mining. In Proceedings International Conference, on Information, Process, and Knowledge Management. 2009. Cancun, Mexico. | Based on process mining, the authors introduce a process diagnostic methodology comprising the following steps: log preparation, log inspection, flow control analysis, performance analysis and role analysis. By now, it has already been applied to another case study [14] and could be complemented for it to be used with ERP systems. |
f) Rebuge, A. and D. Ferreira, Business process analysis in healthcare environments: A methodology based on process mining. Information Systems. 2012. 37(2): p. 99-116 | The authors propose a methodology for process analysis in healthcare and describe it as an extension of the work of Bozcaya et al. [16]. Focused on cluster analysis, this technique is aimed at discovering event log behavior patterns, and for such purpose it was applied to the emergency service of a hospital, as presented in the paper. |

3. Significance:

The main contribution of this research project will be a methodology aimed at promoting the use and extension of process mining by improving its understandability for, and usability by, non-experts. The projected users of this methodology are the staff of any organization operating an ERP and needing to conduct a process redesign project by means of process mining techniques. The specific contributions are:

C1. A methodology that includes activities, tools and techniques for applying process mining to ERP-supported business processes.

C2. A procedure for extracting, cleaning, merging, and analyzing event logs, tables and fields of the ERP system, intended for the application of process mining.
C3. A characterization of process and data mining techniques, such that it corresponds to process redesign project objectives.

4. Research design and methodology:

Figure 1 shows the procedure selected to achieve the expected research results.

**Fig 1. Research procedure**

**Figura 1. ERP data and event log analysis**

**Data sources.** The method is intended for organizations supporting their processes with Integrated Information Systems (ERPs), out of which those aimed by this project are SAP R/3, PeopleSoft and Oracle e-business suite, which are the most widely used ones, thus covering a greater market share.

**Validation of results.** The validation of the methodology will be carried out in at least two real business processes supported by different ERP systems. The evaluation criteria include:

- Level of achievement of *process redesign project* goals.
- Understandability and usability perceived by the staff of the organizations participating in the study.
- Findings resulting from the application of process mining as compared to other process analysis techniques such as statistical process control and time studies, among others.

5. Research stage:

The application of data mining techniques to the analysis of the procurement process of a private university supported by PeopleSoft was addressed by the author through a research project in 2011 [19]. The main findings are the determination of the variables that influence the cycle time, and the probability that, as a result of the application of decision trees, the process is executed within a certain time limit. A combination of methodologies was used in the project, namely CRISP-DM and BPtrends (respectively, data mining and process redesign methods), together with a series of simulation techniques.

The next step will analyze the data structure and process log of SAP R/3, initially in an academic implementation described in [20], and then in a real business.
6. References:


