Requirements towards Effective Process Mining

Matthias Lohrmann and Alexander Riedel

KPMG AG Wirtschaftsprüfungsgesellschaft,
IT Advisory,
{mlohrmann,ariedel}@kpmg.com

Abstract. Process mining is prominent contemporary research topic. This paper describes requirements to be fulfilled for effective practical adoption on the basis of application scenarios and a sample project.

1   Introduction

Business process management (BPM) methods have achieved broad acceptance to support cooperation requirements within and between organizations [1]. In this context, process mining has been a prominent research issue in recent years [2,3]. From our perspective, its purpose lies in complementing or even substituting *a-priori design models* with *ex-post enactment models* of business processes. In terms of practical application, this is very promising considering the following issues:

- Design models are subject to an "enactment bias", i.e. deviations between to-be model and as-is instantiations, or not available at all.
- As opposed to design models, enactment models can be enriched with actual enactment statistics covering, e.g., the prevalence of certain enactment patterns or process variants, cycle times, or actually contributing individuals or roles. Note that this leads to a "wider" view of process mining than what is adopted in parts of related literature – respective tools, however, generally integrate these aspects.

This position paper aims at illustrating potentials of and requirements for process mining techniques and tools from the perspective of an audit and advisory firm. Its contribution lies in highlighting topics of practical relevance for future research. In the following sections, we thus describe general application scenarios (cf. Section 2) and more specific findings from a real-world pilot project executed with an industrial client (cf. Section 3). Both lead to requirements for further development as discussed in Section 4.

Note that our affiliation with an audit firm requires us to refrain from citing concrete tools and suppliers in this paper.

2   Application Scenarios

This section describes application scenarios for process mining enabling to deduct functional requirements. The scenarios have been selected on the basis of the topics that have emerged as the most relevant from discussions with our clients.

Application Scenario 1 (Process Optimization in Functionally Structured Organizations). In most organizations, function-oriented organizational structures (e.g., the procurement, logistics, and accounting functions) are still prevalent in comparison to process-oriented structures (e.g., purchase-to-pay). In this context, end-to-end (E2E) process optimization is a very important, but challenging issue. As an example, consider the capturing of supplier order data. This is often seen as tedious work by the procurement function, which, however, leads to huge additional workload in accounting. Process mining can make these issues transparent by analyzing process patterns independently of organizational borders. Thus, continuous end-to-end process improvement can be fostered.

Application Scenario 2 (Compliance Management and Identification of Deficiencies). For many real-world processes, compliance deficiencies cannot be rigorously precluded through appropriate process design. Moreover, the capability to clearly retrace process execution often constitutes a compliance requirement in itself (e.g., §238 of the German *Handelsgesetzbuch*). Process mining techniques allow to analyze execution logs with respect to non-compliant characteristics (e.g., occurrence of non-approved payment instances). Here, their particular advantage lies in the automated analysis of large numbers of process instances. Compared to the "traditional" risk-oriented audit approach, they allow to execute continuous or even full-scope audits [4]. In some cases, even the systematic circumvention of compliance requirements through certain process patterns can be uncovered (cf. Section 3).

Application Scenario 3 (Internal Benchmarking). Benchmarking refers to the comparison of good practices and performance indicator values (e.g., cycle times, unit costs, and transaction volumes per unit of capacity) between peer organizations [5]. From our experience, the so-called *internal* benchmarking, i.e. the comparison within one group, for instance between business units or production plants, has emerged as particularly useful since it reduces the issue of (perceived or real) limitations to the comparability of peers. In this context, process mining can play an important role. Typical tools combine the assessment of practices (i.e., enactment patterns or process variants) with performance indicators. It is particularly suited for internal benchmarking because in that case the probability of equal process-aware information systems (PAISs) is higher. As an additional advantage with regard to implementation, it makes use of readily available process log data instead of additional surveys, thus reducing the cost of benchmarking.

Application Scenario 4 (Capacity Management). Process mining allows to identify "bottlenecks" in transactional processing. This is achieved by identifying roles or users with particular high transactional volumes or cycle times in terms of time lag between "receipt" and "completion" events. The respective capacities can then be subject to specific management.

3   Pilot Application and Findings

This section shortly discusses findings from a pilot project conducted with a client. The objectives of the project primarily corresponded to Application Scenario 1 as described above, although aspects of other scenarios were addressed as well. Example 1 describes the sample process.

Example 1 (Sample Process: Purchase-to-Pay). The purchase-to-pay process covers process steps that are briefly summarized in Business Process Model and Notation (BPMN, [6]) in Figure 1. The model has been simplified for reasons of space, but amended with reference numbers we use to allocate findings. The total mining sample covered 28,345 process instances with ca. 400 event types incurred at a foreign subsidiary in a timeframe of three months after data cleansing. Data cleansing involved removing all instances that were not started and completed (with defined start and end events) within the given timeframe.[1]
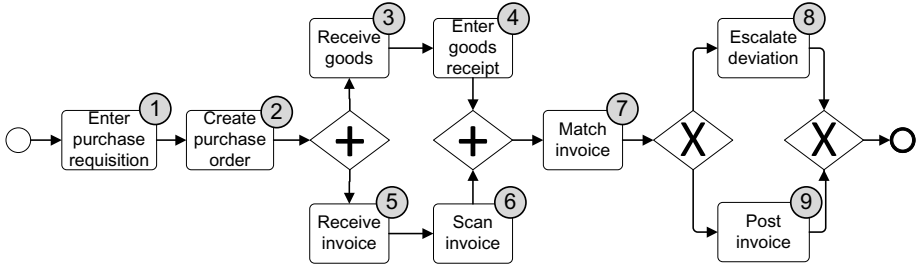
Fig. 1. Sample Process: Purchase-to-Pay

Due to its transactional character, its well-understood semantic content with a high degree of standardization, its occurrence in basically all industries, and its broad support in PAISs, the purchase-to-pay process is particularly well-suited to illustrate the potentials of process mining. The application of two exemplary process mining tools enabled comparing a broader range of functionality. Major findings described in the following illustrate which topics should be covered (i.e., *discoverable*) by effective process mining techniques and tools:

1. Up to two thirds of total project effort can be incurred in data integration and cleansing.
2. 72% of invoices are below the threshold value for mandatory purchase order creation, leading to missing purchasing data.
3. Missing purchasing data (Ref. 2 in Figure 1) leads to an 88% failure rate in automated invoice matching (Ref. 7).
4. 86% of postings (Ref. 9) occur on five out of a total of 116 accounts.

---

[1] Note that this constitutes an inherent weakness since issues leading to pending instances (e.g., "deadlocks") are systematically hidden from further analysis.

5. Transactional volume shifts significantly between months, which leads to issues with capacity management.
6. There have been more than 60,000 segregation of duty violations (e.g. one user entering both purchase order and payment approval).
7. In ca. 3% of cases, purchase orders have been created (Ref. 2) immediately before or after an invoice became overdue, which might point to a compliance violation.

4 Conclusion

Based on the application scenarios and the sample case discussed, we conclude that the following requirements are of particular relevance to the effective practical application of process mining:

Requirement 1 (Data Integration). To enable broad application, the integration of (process log) data from various sources is of critical importance. This issue is currently cumbersome with most tools and includes recurring and automated data staging as well as cleansing (e.g. Finding 1).

Requirement 2 (Compliance Rules Modeling). Tools should include a modeling facility for compliance rules such as the segregation of duties [7]. Automated analysis with respect to violations should be feasible (e.g. Application Scenario 2, Findings 6/7).

Requirement 3 (Pattern Analysis). The analysis of process variants including their cumulative frequency has emerged as the most useful aspect for management discussions. Tools should support this analysis including enrichment of patterns with performance indicators (e.g. Application Scenario 1, Finding 2).

Requirement 4 (Approximation of Manual Effort). Manual effort incurred in process activities cannot be measured by mining execution events, but approximation facilities should be provided by tools (e.g. Application Scenario 3, Finding 5).

Requirement 5 (Automated Regression Analysis). Statistical regression analysis, i.e. the analysis of relations between dependent and independent variables should be automated. This would, for instance, be useful to identify bottlenecks or for lower level internal benchmarking (e.g. between cost centers, Finding 4).

Requirement 6 (Sample Delineation). A methodical solution for delineating a sample of process instances is still missing. For instance, delineation on the basis of completed instances can result in blanking out precisely critical cases (e.g. Finding 1).

Requirement 7 (Visualization of Variants). For management discussions, process variants should be visualized effectively (e.g. regarding the difference between two patterns) using innovative solutions (e.g. Application Scenarios 1-4).

References

1. van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M.: Business Process Management: A Survey. In: van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 1–12. Springer, Heidelberg (2003)
2. van der Aalst, W.M.P., Weijters, A.J.M.M.: Process mining: a research agenda. Computers in Industry 53(3), 231–244 (2004)
3. Li, C., Reichert, M., Wombacher, A.: Mining business process variants: Challenges, scenarios, algorithms. Data & Knowledge Engineering 70(5), 409–434 (2011)
4. Rezaee, Z., Sharbatoghlie, A., Elam, R., McMickle, P.L.: Continuous auditing: Building automated auditing capability. Auditing 21(1), 147–163 (2002)
5. Camp, R.C.: Benchmarking: The search for industry best practices that lead to superior performance. Quality Press (1989)
6. The Object Management Group: Business Process Model and Notation: Version 2.0 (2011), http://www.omg.org/spec/BPMN/2.0
7. Ly, L.T., Knuplesch, D., Rinderle-Ma, S., Göser, K., Pfeifer, H., Reichert, M., Dadam, P.: SeaFlows Toolset – Compliance Verification Made Easy for Process-Aware Information Systems. In: Soffer, P., Proper, E. (eds.) CAiSE Forum 2010. LNBIP, vol. 72, pp. 76–91. Springer, Heidelberg (2011)