Process Constraint Discovery Based on Regulatory Documents and Process Execution Logs (Extended Abstract)

Karolin Winter[1,2]

[1]*Department of Informatics, Technical University of Munich, Germany*
[2]*Faculty of Computer Science, University of Vienna, Austria*

Abstract

Despite digitalizing business process compliance is more vital than ever due to the continuously increasing amount of regulatory documents and process execution data a full stack digitalized compliance management is still out of reach. This thesis aims at bridging this gap by providing support for the *modeling* as well as *discovery* phase of business process compliance.

Keywords

Process Compliance, Natural Language Processing, Instance Spanning Constraints, Process Mining

1. Introduction

Business process management has become a key success factor for businesses over the past years and targets the design, enactment, management and analysis of business processes. Business processes describe workflows within companies, i.e., consist of a number of events, activities, contain decision points and can involve actors or physical objects [1]. Besides capturing workflows, business processes usually have to comply to constraints which emerge from regulatory documents and are the subject of business process compliance. Despite digitalizing business process compliance is more vital than ever due to the continuously increasing amount of regulatory documents and process execution data generated by, e.g., process aware information systems, a full stack digitalized compliance management is still out of reach. This thesis aims at bridging this gap by providing support for the *modeling* as well as *discovery* phase of business process compliance. In the following, the contributions of the thesis are described along these two phases complemented by a summary of overarching contributions.

2. Business Process Compliance – Modeling Phase

Within the modeling phase, constraint discovery from regulatory documents as well as determining their relation to business processes is envisioned.

2.1. Research Gaps

Several state-of-the-art approaches have addressed the problem of extracting business process models and constraints from natural language text (cf., e.g., [2, 3]). Most of them consider so-called process descriptions as input which fundamentally differ from regulatory documents. Yet, regulatory documents constitute the primary source of constraints [4] and therefore, this thesis addresses the question *How to extract and represent constraints from regulatory documents in an automated way?* In order to estimate whether constraints were captured correctly during the modeling phase, the thesis addresses the question *How to assess compliance between regulatory documents and process models?*

2.2. Contributions

Text fragmentation and constraint discovery [5]   A *method chain* based on existing text mining techniques in combination with a novel idea of applying text fragmentation to improve the results of, e.g., clustering algorithms was developed. It enables the retrieval of relevant constraints from regulatory documents and is feasible to extract main characteristics related to planning and implementing of regulatory documents as well as best practices in existing standards and guidelines.

Constraint grouping and determination of constraint relations [6]   After constraints have been discovered from regulatory documents, the thesis presents three options for *grouping* them based on, e.g., topics or stakeholders by employing natural language processing techniques. This facilitates and accelerates the installation and maintenance of regulatory documents and related processes, shortens the consulting procedure, and reduces the overall implementation costs. Moreover, it is crucial to enable a content-wise comparison among constraints. Imagine a regulatory document containing the following constraints i)"A data owner has 5 days to delete this information.", ii)"This information has to be deleted within 5 days by the data owner." which are spread across the document. A user reading this document in a chronological order might not be aware of such redundancies. In the best case this causes additional implementation effort, in the worst case implementation errors. The thesis therefore provides *definitions for determining redundant, subsuming and contradicting constraints* which are depicted in a graph-based representation, a *constraint network map*.

Derivation of mixed graphs [7]   In order to bridge the gap between constraints and business processes, the thesis employs the concept of *mixed graphs*. A mixed graph mirrors a paragraph within a regulatory document containing at least one constraint and establishes control flow connections if possible. This enables a representation of regulatory documents heading towards formal process models.

Compliance assessment of regulatory documents and process models [8]   Another key contribution of the thesis constitutes an approach for assessing the compliance between regulatory documents and process models and is the first work of its kind. We developed a *fitness score* quantifying the likelihood that a paragraph pertains to a model as well as a

cost score quantifying the distance between the obligations expressed in paragraphs and their implementation by a model. Within the latter we covered three essential types of compliance violations, missed obligatory activities, strict order and resource responsibility violations.

3. Business Process Compliance – Discovery Phase

Within the discovery phase constraints are distinguished into *intra instance constraints*, i.e., constraints affecting only one process instance and so-called *instance spanning constraints (ISC)*, i.e., constraints spanning multiple process instances of one or multiple process types [9]. A process type is hereby described by a process model and a process instance represents one execution of a process type. Process execution logs constitute the data source within the discovery phase of business process compliance.

3.1. Research Gaps

Current approaches for extracting knowledge from process execution logs are related to process discovery, e.g., decision mining [10] or declarative process mining, e.g., [11]. Yet, these approaches are only capable of discovering intra instance constraints but do not envision the discovery of ISC. Therefore, this thesis poses the question *How to extract and represent ISC from process execution logs?* Since not just the discovery of ISC but also a formalization of ISC was lacking, we further addressed *How to support integration, transparency, and usage of ISC?*

3.2. Contributions

Categorization of instance spanning constraints [12]   For addressing ISC discovery from process execution logs, we established a *categorization of ISC* based on an extensive set of real-world ISC examples. We identified five distinct categories. Four of them, i.e., simultaneous execution, constrained execution, ordered execution and non-concurrent execution, trace back to synchronization of instances. The fifth category relates to exception handling of ISC.

Algorithms for discovering instance spanning constraints [12, 13]   Within the thesis, we developed *two ISC discovery approaches*. One semi-automatic, following the idea of the decision miner, i.e., first preparing process execution logs such that so-called ISC decision points are revealed and second applying decision tree algorithms for determining the underlying decision rules. The second approach bases on the ISC categorization, is automatic, yet, due to well-selected parameters, adjustable to various settings.

Formalization of instance spanning constraints in terms of patterns [14]   In order to support the integration, transparency, and usage of ISC we developed *ISC patterns* based on timed colored workflow nets and Proclets. This fosters ISC comparison and consequently an assessment whether processes obeyed to imposed ISC during process execution.

4. Overarching Contributions

Business process compliance life cycle   The outlined contributions are linked to but not strictly limited to one phase of business process compliance. The compliance assessment approach can, e.g., be employed to check whether executed processes obeyed to the constraints imposed on them by mining a process model from a process execution log. ISC patterns can also be employed during the modeling phase and the selected formalism has a well-defined formal execution semantics enabling the smooth transformation into executable process code.

Relevance and implication to practice   Ensuring that the outlined approaches constitute solutions to relevant practical problems was always considered. On the one hand we evaluated the concepts based on manifold user stories for three different expert groups, on the other hand prototypes such as RegMiner [15] were further improved by interviews with domain experts.

Evaluation and implementation   Within the evaluation, several evaluation methods ranging from expert interviews, to case studies and common measures like precision and recall were employed. The artificial as well as real-world data sets stemmed from several domains, e.g., security, health care, manufacturing or the financial domain, demonstrating the domain independent applicability of the presented approaches. Moreover, six command-line as well as two web-based prototypes [15, 16] were developed. The latter provide a low-threshold entry for non-technical users ensuring again the transfer of research results to practice.

5. Conclusion

Within this thesis novel concepts and algorithms for the (semi-)automatic discovery and contextualization of compliance constraints from regulatory documents as well as process execution logs were presented constituting a fundamental step towards digitalized compliance management and to significantly relieve compliance and domain experts. Further research directions can target exception handling for ISC or compliance assessment of regulatory documents with further textual documents such as policies or handbooks.

Acknowledgments

References

[1] M. Dumas, M. L. Rosa, J. Mendling, H. A. Reijers, Fundamentals of Business Process Management, Second Edition, Springer, 2018. doi:10.1007/978-3-662-56509-4.

[2] H. van der Aa, C. D. Ciccio, H. Leopold, H. A. Reijers, Extracting declarative process models from natural language, in: Advanced Information Systems Engineering, CAiSE, 2019, pp. 365–382. doi:10.1007/978-3-030-21290-2_23.

[3] F. Friedrich, J. Mendling, F. Puhlmann, Process model generation from natural language text, in: Advanced Information Systems Engineering, CAiSE, 2011, pp. 482–496. doi:10.1007/978-3-642-21640-4_36.

[4] L. T. Ly, S. Rinderle-Ma, D. Knuplesch, P. Dadam, Monitoring business process compliance using compliance rule graphs, in: On the Move to Meaningful Internet Systems: OTM, 2011, pp. 82–99. doi:10.1007/978-3-642-25109-2_7.

[5] K. Winter, S. Rinderle-Ma, W. Grossmann, I. Feinerer, Z. Ma, Characterizing regulatory documents and guidelines based on text mining, in: On the Move to Meaningful Internet Systems. OTM, 2017, pp. 3–20. doi:10.1007/978-3-319-69462-7_1.

[6] K. Winter, S. Rinderle-Ma, Detecting constraints and their relations from regulatory documents using NLP techniques, in: On the Move to Meaningful Internet Systems. OTM, 2018, pp. 261–278. doi:10.1007/978-3-030-02610-3_15.

[7] K. Winter, S. Rinderle-Ma, Deriving and combining mixed graphs from regulatory documents based on constraint relations, in: Advanced Information Systems Engineering, CAiSE, 2019, pp. 430–445. doi:10.1007/978-3-030-21290-2_27.

[8] K. Winter, H. van der Aa, S. Rinderle-Ma, M. Weidlich, Assessing the compliance of business process models with regulatory documents, in: Conceptual Modeling, ER, 2020, pp. 189–203. doi:10.1007/978-3-030-62522-1_14.

[9] M. Leitner, J. Mangler, S. Rinderle-Ma, Definition and enactment of instance-spanning process constraints, in: Web Information Systems Engineering - WISE, 2012, pp. 652–658. doi:10.1007/978-3-642-35063-4_49.

[10] A. Rozinat, W. M. P. van der Aalst, Decision mining in prom, in: Business Process Management, BPM, 2006, pp. 420–425. doi:10.1007/11841760_33.

[11] F. M. Maggi, A. J. Mooij, W. M. P. van der Aalst, User-guided discovery of declarative process models, in: Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, CIDM, 2011, pp. 192–199. doi:10.1109/CIDM.2011.5949297.

[12] K. Winter, F. Stertz, S. Rinderle-Ma, Discovering instance and process spanning constraints from process execution logs, Inf. Syst. 89 (2020) 101484. doi:10.1016/j.is.2019.101484.

[13] K. Winter, S. Rinderle-Ma, Discovering instance-spanning constraints from process execution logs based on classification techniques, in: IEEE International Enterprise Distributed Object Computing Conference, EDOC, 2017, pp. 79–88. doi:10.1109/EDOC.2017.20.

[14] K. Winter, S. Rinderle-Ma, Defining instance spanning constraint patterns for business processes based on proclets, in: Conceptual Modeling, ER, 2020, pp. 149–163. doi:10.1007/978-3-030-62522-1_11.

[15] K. Winter, M. Gall, S. Rinderle-Ma, Regminer: Taming the complexity of regulatory documents for digitalized compliance management, in: BPM-D, 2020, pp. 112–116.

[16] F. Stertz, K. Winter, S. Rinderle-Ma, SVIPEX: A web service for discovering and visualizing instance spanning constraints based on process execution logs, in: BPM-D, 2020, pp. 117–121.