Objectivity in Process Descriptions

Christopher Klinkmüller[1], Henrik Leopold[2], Jan Mendling[3], and Ingo Weber[4]

[1] CSIRO Data61, Sydney, Australia christopher.klinkmueller@data61.csiro.au
[2] Kühne Logistics University, Hamburg, Germany henrik.leopold@the-klu.org
[3] Humboldt-Universität zu Berlin, Berlin, Germany jan.mendling@hu-berlin.de
[4] Chair of Software and Business Engineering,
Technische Universitaet Berlin, Germany, <firstname>.<lastname>@tu-berlin.de

Abstract. Process models are central artifacts for many business process management activities. They are often manually crafted, which means that modelers capture many details in the way they consider appropriate – but the problem also applies to discovered models. We, therefore, argue that we need objectivity of granularity level, objectivity of perspective, and objectivity of terminology to enable broader use of models, like comparing processes. This is currently not available, which is a roadblock for automatic analysis, empirical research, and generally use for purposes that differ from the initial model creation purpose.

1 Introduction

Process models are central artefacts for many *business process management* (BPM) activities and provide a foundation for the design, documentation, analysis, automation, and optimization of business processes [1]. Traditionally, process models have been manually created and kept up to date by modelers. Nowadays, increasingly process discovery techniques from the field of process mining are used to automatically derive models from discrete event data. Depending on the degree of BPM adoption, organizations might establish collections consisting of thousands of process models.

In essence, process models provide *concise* and *selective* representations of business processes, as they abstract from many details and express specific aspects, whose relevance depends on the models' purpose, through a few elements for which short labels provide brief natural language descriptions. Moreover, independent of whether a process model is created manually or through discovery, it provides a *selective* view. In the case of manual creation, this selectivity stems from the fact that modelers express their own perception in a way they deem appropriate. Although discovery algorithms follow precise rules to transform data into process models, selectivity arises when information needs are translated into operations that extract, preprocess, and analyze the data [2].

To create process models, modelers can rely on notations such as BPMN, EPCs, Petri Nets, etc., which define the types of elements that can be used to describe processes. They can also resort to guidelines that outline how to apply those notations so that the resulting models are of a high quality, e.g., those

CEUR-WS.org/Vol-2938/paper-PROBLEMS-18.pdf

in [3, 4]. However, the creation of models is more art than science. That is because available notations, methods, and tools abstract from the model content and do not provide guidance for how to handle selectivity when capturing processes. This freedom during modeling exacerbates the effective utilization of models within the BPM lifecycle, especially when model usage and interpretation are negatively impacted by the absence or ambiguous description of important aspects. For example, the authors in [5] attempted to consolidate a set of process models, but were challenged by the versatile labeling of similar activities. Similarly, the empirical study in [6] demonstrated that modelers tend to express aspects via natural language although appropriate modeling elements for those aspects are available. In this regard, it is important to stress that discovered models are not necessarily easier to understand than manually created models [7].

Next, we summarize fundamental challenges surrounding this problem and discuss their impact on existing work. We then outline possible future directions.

2 Existing Work and Challenges

The core of the problem can be traced back to models being concise, selective and arguably subjective process representations, or in other words to a lack of objectivity in the following senses:

Objectivity of levels of granularity: So far, there are no objective levels of granularity for describing a business process. If we accept that a process is something that can be decomposed into subprocesses [8], which may also be referred to as activities, tasks, steps, phases, stages, etc., then we observe that processes have been described and analyzed at the macro level (developments of companies over decades [9] or careers of famous musicians [10]), meso level (order-to-cash processes [11] or healthcare pathways [12]), and micro level (keystroke sequences [13] or scrolling of a computer user [14]). Modelers can choose different levels of granularity depending on what they deem appropriate for a given modeling purpose. With this observation, we do not mean to suggest that all models should be created on the same level of granularity; how to possibly react to the situation observed will be described in the next section.

Objectivity of perspectives: So far, there are no objective perspectives for describing business processes. One specific instance of this problem is the discussion of local and global views of business processes [15] and the usage of pools (blackbox or whitebox) versus lanes in BPMN [11]. Modelers construct views and system boundaries around passages of a process that they deem relevant for a given task at hand.

Objectivity of terminology: So far, there are no objectively defined terms available for describing business processes and the elements in process models. This so-called vocabulary problem is fundamental and not specific to business processes [16, 17]. Even if we refer to the same matter, we can use homonyms and synonyms [18] and describe activities from the perspective of what they aim towards, how they are done, or what they achieve [19]. Modelers are free to choose terminology in process models based on what they deem appropriate in a specific context.

These challenges have implications for various semantic application scenarios of business process models [20], e.g., for *process model matching* where algorithms are designed that automatically identify correspondences between models, i.e, activities that represent similar functionality. Process model matching has turned out to be a fundamentally hard problem and provides a perfect example for illustrating the consequences of the lack of objectivity in process modeling.

Despite the substantial attention that process model matching received, the solutions approaches that were developed have not yet yielded satisfactory and practically usable performance, as prominently demonstrated in the process model matching contests in 2013 [21] and 2015 [22]. Here, matching techniques were compared in a competitive setting and overall achieved a moderate effectiveness. This performance is a result of a low recall, i.e., matchers only identify a small portion of the existing correspondences. Generally, the most plausible strategy to lift recall is by sacrificing precision, i.e., by allowing matchers to propose a substantial amount of incorrect correspondences. This performance is a direct result of the lack of objectivity with which models are created. That is, when implementing matchers, developers can only resort to general-purpose, off-the-shelf knowledge bases and techniques, but the matchers themselves need to interpret less objective process models with heterogeneous labeling styles, domain terminology, etc. [23]. Moreover, the same control flow can be expressed in various ways. This means that the control flow relationships have a limited explanatory power for correspondences, as confirmed by empirical evidence [24].

A promising direction for improving the effectiveness is to learn from user feedback [25]. However, such a setup in the end means that instead of algorithms, it is the model creators and users who have to make sense of the models. In this regard, several studies, e.g., in [26, 27], demonstrated that humans also face challenges when interpreting models, often arriving at diverging views regarding the existing correspondences between the same pair of process models.

3 Future Directions

The creation and interpretation of models in general and of process models in particular has been an active research area for decades, resulting in a broad range of notations, practices, (anti-)patterns, and tools. Contrasting those efforts and outcomes with the severity of the problems around objectivity, it is hard to devise specific ideas for advancing the body of knowledge in this direction. Part of the problem is that selectivity is not only a bug, but to a degree also a feature: each model is created for a purpose, like documentation or performance analysis. What should be part of the model and what can be abstracted from depends on this very purpose, and impacts granularity level, perspective, and vocabulary. While vocabulary for a given context could be objectified through use of ontologies, dictionaries, or glossaries, this is not the case for the perspective and granularity dimensions, given their dependence on the purpose. A first step to addressing the problem could be the generation of taxonomies for these dimensions, and mapping of process models to taxonomy elements.

In general, research into this topic could benefit from more publicly available data in terms of large process model collections, protocols of how individuals translate processes into models, or records of how process model collections evolve over time. This is not to say that there have not been attempts to establish collections of real-world data, an endeavour that is often hampered by contractual obligations. For example, the SAP reference model has been studied in many publications; the process model matching contests [22, 21] provided process model collections along with gold standards that define the correspondence relationships in the models; Signavio's BPM Academic Initiative is providing access to models that users of the platform contributed to the initiative [28]; the annual Business Process Intelligence Challenge provides real-world event logs and publishes the contestants' analysis reports which contain protocols and interpretations for process discovery results; and the BPM conference is encouraging researchers to adopt open science principles and to submit resource papers.

The availability of extensive data collections could then be used to study similarities between process models and, in general, how they can be systematically made more comparable. For example, based on manually identified correspondence relationships, qualitative content analysis and data mining could help to better understand the different ways in which concrete aspects can be expressed and to derive objective ways for modeling those aspects, potentially using new paradigms. In this regard, it would be beneficial to forgo the common practice of relying on binary correspondence relationships. Instead, more insights might be derived when diverging views of multiple analysts are considered, and with more detailed information regarding the nature of correspondence relationships, e.g., in terms of similarity scores, classifications, or open-ended descriptions.

References

1. Malinova, M., Mendling, J.: Identifying do's and don'ts using the integrated business process management framework. Business Process Management J. (2018)
2. Klinkmüller, C., Seeliger, A., Müller, R., Pufahl, L., Weber, I.: A method for debugging process discovery pipelines to analyze the consistency of model properties. In: International Conference on Business Process Management. (2021)
3. Becker, J., Rosemann, M., von Uthmann, C.: Guidelines of business process modeling. In: Business Process Management: Models, Techniques, and Empirical Studies. (2000) 30–49
4. Mendling, J., Reijers, H.A., van der Aalst, W.M.P.: Seven process modeling guidelines (7pmg). **52**(2) (2010) 127–136
5. Gottschalk, F., Wagemakers, T.A., Jansen-Vullers, M.H., van der Aalst, W.M., La Rosa, M.: Configurable process models: Experiences from a municipality case study. In: Intl. Conf. Advanced Information Systems Engineering. (2009) 486–500
6. Pittke, F., Leopold, H., Mendling, J.: When language meets language: Anti patterns resulting from mixing natural and modeling language. In: Intl. Workshop on Process Model Collections: Management and Reuse. (2014) 118–129

7. Fahland, D., van der Aalst, W.M.: Simplifying discovered process models in a controlled manner. Information Systems **38**(4) (2013) 585–605
8. Malone, T.W., Crowston, K., Herman, G.A.: Organizing business knowledge: The MIT process handbook. MIT press (2003)
9. Pettigrew, A.M.: The character and significance of strategy process research. Strategic management journal **13**(S2) (1992) 5–16
10. Abbott, A., Hrycak, A.: Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. Am. J. of Sociology **96**(1) (1990) 144–185
11. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of business process management. Second Edition. Springer (2018)
12. Pentland, B.T., Recker, J., Wyner, G.: Rediscovering handoffs. Academy of Management Discoveries **3**(3) (2017) 284–301
13. Card, S.K., Moran, T.P., Newell, A.: The keystroke-level model for user performance time with interactive systems. Comm. of the ACM **23**(7) (1980) 396–410
14. Altmann, E.M., John, B.E.: Episodic indexing: A model of memory for attention events. Cognitive Science **23**(2) (1999) 117–156
15. Zaha, J.M., Dumas, M., Ter Hofstede, A., Barros, A., Decker, G.: Service interaction modeling: Bridging global and local views. In: IEEE EDOC. (2006) 45–55
16. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Comm. of the ACM **30**(11) (1987) 964–971
17. Gassen, J.B., Mendling, J., Bouzeghoub, A., Thom, L.H., de Oliveira, J.P.M.: An experiment on an ontology-based support approach for process modeling. Information and Software Technology **83** (2017) 94–115
18. Pittke, F., Leopold, H., Mendling, J.: Automatic detection and resolution of lexical ambiguity in process models. IEEE Trans. Software Eng. **41**(6) (2015) 526–544
19. Leopold, H., Mendling, J., Reijers, H.A., La Rosa, M.: Simplifying process model abstraction: Techniques for generating model names. Information Systems **39** (2014) 134–151
20. Mendling, J., Leopold, H., Pittke, F.: 25 challenges of semantic process modeling. International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC) **1**(1) (2015) 78–94
21. Cayoglu, U., Dijkman, R., Dumas, M., et al.: The process model matching contest 2013. In: Business Process Management Workshops, Beijing, China (2013) 442–463
22. Antunes, G., Bakhshandeh, M., Borbinha, J., et al.: The process model matching contest 2015. In: EMISA. (2015) 127–155
23. Klinkmüller, C., Weber, I., Mendling, J., Leopold, H., Ludwig, A.: Increasing recall of process model matching by improved activity label matching. In: International Conference on Business Process Management, Beijing, China (2013) 211–218
24. Klinkmüller, C., Weber, I.: Analyzing control flow information to improve the effectiveness of process model matching techniques. Decision Support Systems **100** (2017) 6–14
25. Klinkmüller, C., Weber, I.: Every apprentice needs a master: Feedback-based effectiveness improvements for process model matching. Information Systems **95** (2021) 101612
26. Rodríguez, C., Klinkmüller, C., Weber, I., Daniel, F., Casati, F.: Activity matching with human intelligence. In: BPM Forum 2016. (2016) 124–140
27. Kuss, E., Leopold, H., van der Aa, H., Stuckenschmidt, H., Reijers, H.A.: A probabilistic evaluation procedure for process model matching techniques. Data & Knowledge Engineering **117** (2018) 393–406
28. Weske, M., Decker, G., Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Model collection of the business process management academic initiative (2020)