Adaptive Process Model Matching – Improving the Effectiveness of Label-based Matching through Automated Configuration and Expert Feedback

Christopher Klinkmüller

Data61 CSIRO, Eveleigh NSW 2015, Australia
christopher.klinkmuller@data61.csiro.au

Abstract. Organizations often store information about their process landscape in large process model collections. To utilize those knowledge bases, process model matchers assist experts in keeping track of model relationships by automatically identifying corresponding activities. Yet, state-of-the-art matchers achieve an overall low effectiveness. That is, their results contain only a few existing, but many non-existing correspondences. Hence, practical application is far from being realized. In light of this, the thesis introduces ADBOT, an interactive matching technique that analyzes manually corrected matcher results to adapt itself to the characteristics of the respective model collection. At heart, ADBOT relies on BOT and OPBOT. BOT is a configurable matcher that solely examines activity labels and OPBOT analyzes model collection characteristics to automatically configure BOT without human intervention. While the three matchers pose different requirements on model and feedback availability, they all achieve a high effectiveness in relation to the state-of-the-art, as shown in a comparative evaluation. ADBOT as the most effective technique even yields improvements of up to 70%.

Keywords: Process model matching · process similarity · process model collection management

1 Introduction

Process models are an essential tool to communicate, analyze, design and continuously improve business processes. In organizations that use process models large process model collections are not uncommon. To leverage the knowledge that is hidden in such extensive collections, research has devised approaches for model collection management [4] such as process similarity search and querying, process model merging, compliance checking, model comparison and clone detection. Many of those approaches require the identification of correspondences, i.e., activities that represent similar functionality in different models. Yet, this step is generally associated with huge manual efforts and research has thus developed *process model matchers* to automate it. But, comparative evaluations in terms

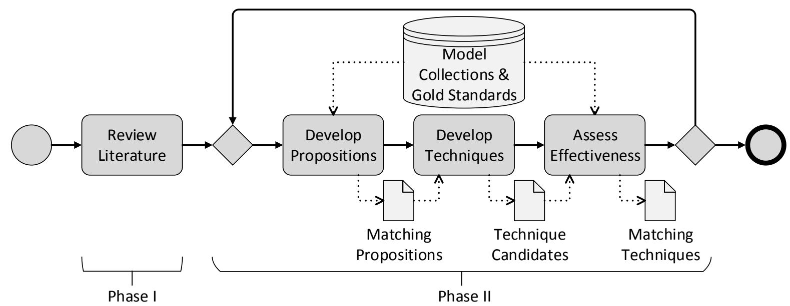

Fig. 1. Research design

of the process model matching contests in 2013 and 2015 [1, 3] demonstrate that state-of-the-art matchers achieve a low *effectiveness* when applied to real-world model collections. That is, the matchers detect only a few of the existing correspondences, while making many irrelevant suggestions. In this context, the thesis studies how the effectiveness can be improved to achieve practical applicability.

2 Research Methodology

The research methodology that the thesis is based on is oriented towards the *information systems research* framework [5]. That is, the research design (shown in Figure 1) is driven by the need for effective matchers in real-world scenarios and aims to contribute new findings to the scientific knowledge base.

The first phase deals with a *literature review* which follows guidelines for literature reviews in information systems research [2]. The goal is to examine prior work and to identify the shortcomings that the thesis must address.

The second phase attends to the development of matching techniques under the consideration of the shortcomings. The activities in this phase are carried out iteratively. Each iteration starts with the *development of matching propositions* where design decisions are examined with regard to their validity. Valid decisions are then considered as matching propositions. In the *development of technique candidates* the propositions are integrated into matching algorithms in different ways. Lastly, *the effectiveness is assessed* and those candidates that achieve a high effectiveness are proposed as matching techniques. It is important to note that due to this research approach, matching techniques are not treated as black boxes, i.e., as the sum of their parts. Instead, the design decisions are individually examined. Thus, the thesis does not only deliver matching techniques, but a catalogue of (in-)validated design decisions that future work can build upon.

The respective experiments rely on empirical datasets that contain model collections, i.e., sets of model pairs, and the respective gold standards which comprise manually identified correspondences. There are two types of datasets. *Development* datasets are used in each iteration to evaluate the design decisions and the effectiveness. By contrast, *evaluation datasets* are exclusively applied in a final iteration to examine the generalizability of the findings. In total, there are four datasets, each containing 36 model pairs. While the *birth registration*

(BR) and the *university admission* (UA) dataset from the matching contest in 2013 [3] are used for development, the *SAP reference model* (SR) and the *alma web* (AW) dataset are the evaluation datasets. SR and AW were created in the context of this thesis following the procedure that was applied to establish BR and UA. SR was made available to the second matching contest [1].

In the experiments a variety of research methods is employed to assess the design decisions and the effectiveness of the matchers. Statistical measures such as the Kolmogorov-Smirnov test [7], the information gain [9], or the correlation coefficient are applied to quantitatively examine cause-effect relationship associated with the design decisions. If a deeper understanding of the relationships is required, categorizing qualitative analyses [8] are carried out. Additionally, well-established measures from information retrieval [6] and the matching contests are used to evaluate the effectiveness based on a comparison of the gold standards and the matchers' suggestions. These measures are (i) the *precision* (P) as the percentage of correct suggestions; (ii) the *recall* (R) as the percentage of identified correspondences; and (iii) the *f-measure* as their harmonic mean.

3 Hypotheses and Evidence

The results of the thesis are summarized in four hypotheses. In the following, the hypotheses are outlined along with the respective evidence from the thesis.

H1: The identification of correspondences between business process models is a challenge for organizations which is not sufficiently supported by existing approaches.

This hypothesis addresses the practical relevance which was already outlined in Section 1 where various problems related to model collection management are listed. Moreover, the scientific demand for the development of more effective matchers was verified through the literature review. A total of 17 publications that introduced matchers were identified and examined. In nine publications the applicability of matchers was restricted by assumptions that e.g., limit the supported modeling languages, that impose requirements on the labeling vocabulary, or that exclude situations where sets of activities correspond. Further, the matchers' effectiveness is low and varies across datasets. Finally, prior work focused on illustrating the matchers' usage or on black-box evaluations that examined the effectiveness, but not the validity of design decisions. These observations confirm the need for further research aimed at improving the effectiveness.

The remaining hypotheses address different sources of information for matching techniques. Each perspective is associated with its own matcher. Table 1 shows the effectiveness of these matchers and of the top-performing matchers from the matching contests [1, 3]. Note that matchers from this thesis were excluded and that for each dataset a different matcher achieved the best results.

H2: Label-based matching techniques yield a varying and generally insufficient effectiveness.

Various natural language analysis techniques related to (i) the decomposition of labels into words, (ii) the syntactic and semantic comparison of words, and (iii) the reduction of differences in label specificity were discussed and analyzed. The results were then used to devise BOT, a label-based matcher that can be adjusted to the model collection via five parameters. BOT comprises a default configuration that performs nearly as good as the three best performing matchers from the matching contests. Finally, a qualitative analysis of BOT's results reveals that many falsely suggested correspondences are due to only small variations in the labels, e.g., when two labels only differ with regard to one word. Further, labels of correspondences that are not detected contain synonyms and homonyms or involve very generic descriptions. This analysis shows that to achieve a high effectiveness, domain-specific knowledge sources are required.

H3: The maximization of the effectiveness of label-based matching techniques is enabled by the analysis of control flow information.

There are three scenarios for the use of control-flow information. First, activities are compared with regard to control-flow properties such as their position in the models. Second, candidates for complex correspondences can be established by extracting connected sub-graphs from the models. Third, there is the consistency analysis which assesses the degree to which control-flow dependencies between activities in one model resemble those of their corresponding counterparts in a second model. While experiments show that the first two options are not generally valid, the third option is identified as a promising design decision. Consequently, OPBOT automatically executes many BOT configurations and suggests a combination of the most consistent results. OPBOT generally detects configurations that perform better or close to BOT's default configuration.

H4: The effectiveness of matching techniques is improved by the utilization of expert feedback.

The process of feedback collection is iterative: a pair of models is selected, then automatically matched, and experts finally correct the suggestions. That way, a ground truth is established which can be used to adapt matchers and to prepare them for yet unmatched model pairs. In particular, ADBOT builds upon OPBOT and contains three strategies to learn from the feedback. First, the assessment of semantic word relations is adapted to better reflect the domain-specific relationships. Second, it relies on transitivity which is a reliable indicator, i.e., two activities are matched, if they correspond to the same activity. Third, OPBOT's BOT configurations are constantly re-adjusted. ADBOT outperforms

Table 1. BOT, OPBOT, and ADBOT vs. the state-of-the-art (SotA) matchers

	BR			UA			SR			AW		
Approach	P	R	F	P	R	F	P	R	F	P	R	F
SotA	.50	.42	.46	.37	.39	.38	.79	.60	**.68**	-	-	-
BOT	.66	.34	.45	.46	.39	.38	.77	.57	.66	.96	.25	.40
OPBOT	.61	.45	.52	.58	.36	.41	.60	**.65**	.63	.73	.34	.46
ADBOT	**.74**	**.81**	**.78**	**.70**	**.73**	**.68**	**.90**	.55	**.68**	**.93**	**.86**	**.89**

the automated matchers from the contests by up to 70%. To reduce the expert workload, ADBOT also sorts the model pairs and feedback collection can be turned after 30 to 50% of the pairs, as the improvements level off at this point.

4 Conclusion

The results of the thesis are summarized in the main research hypothesis which was validated through the verification of the four sub-hypotheses H1 to H4:

> **H0:** The adaptation of business process model matching techniques to model collections is necessary to ensure a high effectiveness and the analysis of the control flow as well as of expert feedback provides means to implement this adaptation.

An important aspect in future work is the extension of the datasets to cover further matching scenarios and achieve a broader evaluation. Moreover, more emphasis should be put on the expert involvement. As experts operate in different contexts, matchers should be adaptive to different views. To reduce human efforts, modeling tools must provide interfaces that allow for easily interpreting and correcting correspondences. This also includes strategies to derive expert feedback from other model collection tasks like model merging or querying.

References

1. Antunes, G., Bakhshandeh, M., Borbinha, J., Cardoso, J., Dadashnia, S., Di Francescomarino, C., Dragoni, M., Fettke, P., Gal, A., Ghidini, C., Hake, P., Khiat, A., Klinkmüller, C., Kuss, E., Leopold, H., Loos, P., Meilicke, C., Niesen, T., Pesquita, C., Peus, T., Schoknecht, A., Sheetrit, E., Sonntag, A., Stuckenschmidt, H., Thaler, T., Weber, I., Weidlich, M.: The process model matching contest 2015. In: EMISA 2015 (2015)
2. vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: ECIS 2009 (2009)
3. Cayoglu, U., Dijkman, R., Dumas, M., Fettke, P., García-Bañuelos, L., Hake, P., Klinkmüller, C., Leopold, H., Ludwig, A., Loos, P., Mendling, J., Oberweis, A., Schoknecht, A., Sheetrit, E., Thaler, T., Ullrich, M., Weber, I., Weidlich, M.: The process model matching contest 2013. In: PMC-MR 2013 (2013)
4. Dijkman, R.M., La Rosa, M., Reijers, H.A.: Managing large collections of business process models – current techniques and challenges. Comput. Ind. **63**(2), 91–97 (2012)
5. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Quart. **28**(1), 75–105 (2004)
6. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge, England (2008)
7. Massey Jr., F.J.: The kolmogorov-smirnov test for goodness of fit. J. Am. Stat. Assoc. **46**(253), 68–78 (1951)
8. Mayring, P.: Qualitative content analysis. FQS **1**(2) (2000)
9. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Pearson Education Limited, Harlow (2014)