ProMatch.KOM: Tool Support for Process Model Analysis and Improvement

Melanie Siebenhaar, Michael Niemann, Julian Eckert, Ralf Steinmetz

Multimedia Communications Lab (KOM),
Department of Computer Science,
Technische Universität Darmstadt, Germany

Abstract. Due to changing market conditions, today's enterprises are constantly forced to improve the quality and efficiency of their business processes. The detection of weaknesses and improvement areas in business process models in comparison to reference models represents a major challenge. Partly automated supportive systems are of great benefit for process analysts in this context.

`ProMatch.KOM` is a hybrid, user-friendly approach to analyse process models with respect to conformance to reference models. Central features are the combined string-based, semantic and structural comparison of two process models described as event-driven process chains (EPC), the mutual assignment of areas comprising similar activities, the computation of process area similarity values, and the identification of process area types. `ProMatch.KOM` has been prototypically implemented as a plug-in for the process mining framework `ProM`.

1 Introduction

Business processes are at the heart of an enterprise for the creation of value and must meet several requirements. Therefore, companies often decide in favour of the partial or complete certification of their IT system as documented evidence of conformity to the requirements. Companies are required to reconcile internal processes with those part of reference frameworks (e.g., IT Governance frameworks). Due to its intensive amount of work, time and cost, this comparison is often performed by external consultants. In order to support the internal process analysis improving existing processes by reference models, partly automated systems can be valuable. In the scope of reference frameworks, a major challenge is the detection of similarities and differences with respect to activities, since the reference processes are represented by rather abstract models (i.e., they focus on the actual control flow to a lesser extent) [1][2]. ProMatch.KOM allows to compare two process models and to determine and mark areas of semantic and structural similarity in the form of grouped nodes, so-called related clusters, using color shades. In addition, also process model parts that are unique to one of the models are highlighted.

2 Features

ProMatch.KOM has been implemented as proof-of-concept and for evaluation purposes as analysis plug-in for the Process Mining Framework ProM [3]. Currently, it is in an early prototype stage. In order to support an process analysts and owners, two process models described as event-driven process chains (EPC) [4] can be loaded, if they are described using the EPC Markup Language (EPML) [5].

2.1 Task Correspondences and Structural Analysis

ProMatch.KOM can be used for a comparison of the two given processes models determining the similarities and differences with respect to syntactic and semantic task correspondences and structural counterparts. A semantic task comparison is mandatory since different process designers may use different terms for the description of tasks. To detect the task correspondences, ProMatch.KOM makes use of a combination of syntactic and semantic similarity measures to assess the similarity of the node labels within both process models. As syntactic measures, metrics provided by the SimMetrics [6] library are applied. SimMetrics os a Java-based open source library of similarity and string distance metrics (e.g., Levenshtein Distance). All metrics take two strings as input and return a real value between 0.0 and 1.0. Due to its normalised outputs, the library also allows for a composition of metrics. For a semantic comparison, the DISCO [7] tool is used, which permits to determine the semantic distributional similarity between two given terms and the set of distributionally most similar terms for a given term based on word distributions. Similar to the SimMetrics library, results are provided in terms of a numerical value between 0.0 and 1.0. Concerning structural comparison, a basic structural analysis is currently part of ProMatch.KOM, identifying simple structure types (e.g., AND-Split/Join, Iteration).

2.2 Grouping of Similarities and Differences

In addition, the similarities and differences are grouped and classified depending on their degree of similarity. The results of the comparison are then graphically outlined and presented to the process analyst as side-by-side view highlighting the similarities and differences in terms of colored clusters (cf. Figure 1). Differing clusters are generally highlighted in red color. Furthermore, two kinds of similar clusters are distinguished: clusters that are only semantically equal and clusters which are semantically and structurally equal. The former clusters are displayed in blue color, while the latter clusters are shown in green color. In both cases, the brightness of the colors indicates the degree of semantic discrepancy, i.e., the darker the coloring of the similar clusters, the more similar they are. This enables the process analyst to quickly detect areas for structural improvement. The decomposition of the process models into colored clusters also reveals global structural differences, if the alignment of the clusters differs. In order to facilitate the visual perception, unique IDs are assigned to corresponding clusters. For the

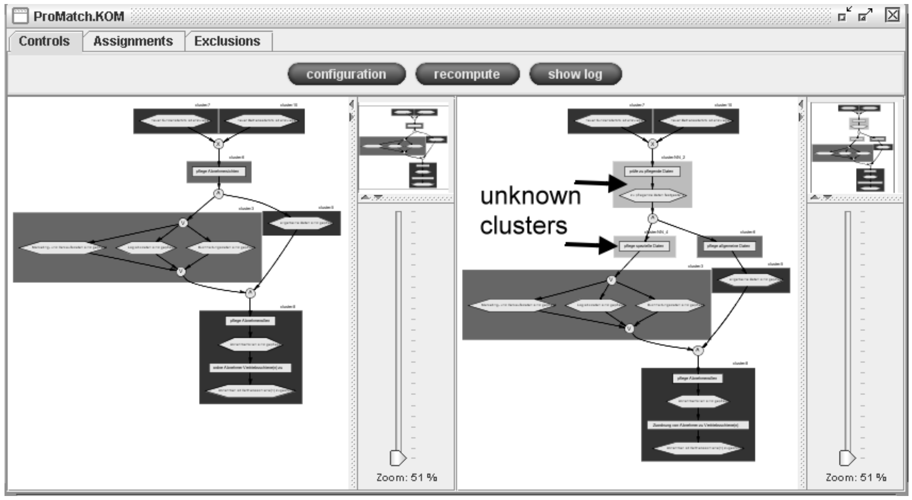

Fig. 1. The ProMatch.KOM workbench

determination of clusters, a decomposition of the EPC graphs into canonical single-entry single-exit (SESE) regions [8] is performed. In addition, to be able to retrace and reason about the computation, all results are provided to the process analyst in a log file.

2.3 User Feedback and Recomputation

Since an automated analysis and matching approach may not be aware of all required information of the processes or may not be able to determine semantic relations of process instances correctly, assignments can be incorrect or can be missing in the result. Therefore, a process analyst must be able to adjust the results. For this purpose, mandatory node assignments can be specified and nodes can be excluded with regard to a subsequent recomputation considering the new constraints. This can be done via the menu (depicted in Figure 2) opened by pressing the right mouse button on a node in the side-by-side-view of the two models. Either the respective node can be excluded, i.e. marked as "unknown" node, with respect to a subsequent recomputation, the respective node can be marked as one part of a mandatory assignment selecting the *Change corresponding node* option in the menu. Subsequently, the corresponding node as second part of the mandatory assignment is selected in the same way within the other model. After pressing the assign button in the top menu, the two selected nodes are added as mandatory assignment to the assignments list. A further comparison of the two models is performed, which is aware of the new restrictions, when pressing the recompute button.

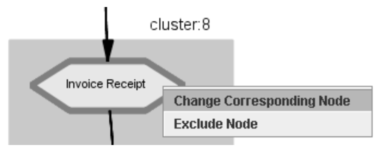

Fig. 2. The user feedback menu

2.4 Configuration and customization of similarity measures

Different process models of one process, created by different people, differ in the way tasks are described. That is, that an adequate comparison has to consider the variety of task descriptions. Thus, the comparison provided by ProMatch.KOM can be customised with regard to individual, weighted similarity measures and global syntactic and semantic weights. In addition, an optional neighbourhood search can be factored into the comparison, so that task correspondences are preferred which share a neighbourhood of similar tasks in both models. This can be useful when comparing models of the same process, where it can be assumed that similar sets of tasks are performed in a similar manner within the same range in the process.

2.5 Complexity

Concerning the computational complexity, the presented approach performs in a divide-and-conquer style based on a disjunctive decomposition of the process models into SESE regions. The computation of cycle equivalence classes required for the SESE region determination can be performed in $O(E)$ time, the Hungarian algorithm [9], which is applied to determine the best matching candidates for all corresponding nodes takes $O(n^3)$ time and the comparison of the nodes of both models can be performed in $O(n * m)$ time. Also the other utilized algorithms are polynomial in time having at most quadratic complexity. In addition, the maximum cluster size could be specified in advance, so that manageable problem sizes can always be achieved. Therefore, the approach presented within this paper yields the desired results in $O(n^3)$ time, even if the size of the models increases.

3 Demonstration

We demonstrate the application of ProMatch.KOM within two scenarios: a comparison of two small, well-structured process models and a comparison of an original (larger) ITIL [2] process with a corresponding (modified) process model. The first results of both comparisons are improved by using the feedback and recomputation mechanism provided. In addition, the influence of different weights, similarity metrics and the optional neighbourhood comparison is demonstrated. A screencast of ProMatch.KOM, which shows the initial comparison in scenario 1 is available under: `http://www.youtube.com/watch?v=KU_YSTQDMsU`.

4 Future Work

Since the current version of ProMatch.KOM only provides a rudimentary structural analysis, a more fine-grained structural analysis will be implemented in the next version to permit the proposal of adequate change operations. This can be achieved by analysing differences between global clusters with respect to their position in the model and differences concerning the nodes contained within a cluster. To enable a global structural analysis, the clusters can be further collapsed to larger heterogeneous clusters. A further enhancement are suggestions for change operations, which can be derived from the information collected during the clustering and analysis steps. These change operations can then be presented to the process analyst for assessment and realisation in order to better comply with the reference process.

References

1. K. Brand and H. Boonen, *IT Governance based on CObIT 4.1.* Van Haren Publishing, Zaltbommel, The Netherlands, 3. ed., 2007.
2. Office of Governance Commerce (OGC), *ITIL v3: Information Technology Infrastructure Library Version 3*, vol. 1-5. 2007. `http://www.itil.org`.
3. B. F. van Dongen, A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, and W. M. P. van der Aalst, "The ProM Framework: A New Era in Process Mining Tool Support," pp. 444–454, 2005.
4. A.-W. Scheer and K. Schneider, "ARIS - Architecture of Integrated Information Systems," in *Handbook on Architectures of Information Systems* (P. Bernus, K. Mertins, and G. Schmidt, eds.), pp. 605–624, Springer Berlin Heidelberg, 2006.
5. J. Mendling and M. Nüttgens, "EPC Markup Language (EPML). An XML-Based Interchange Format for Event-Driven Process Chains (EPC)," Tech. Rep. JM-2005-03-10, Vienna University of Economics and Business Administration, 2005.
6. S. Chapman, N. L. P. Group, D. of Computer Science, and U. of Sheffield, "SimMetrics. Similarity Metrics," 2006. Last accessed on: 2010-05-28.
7. P. Kolb, "DISCO: A Multilingual Database of Distributionally Similar Words," in *KONVENS 2008 - Ergänzungsband: Textressourcen und lexikalisches Wissen* (A. Storrer, A. Geyken, A. Siebert, and K.-M. Würzner, eds.), (Berlin), 2008.
8. R. Johnson, D. Pearson, and K. Pingali, "Finding Regions Fast: Single Entry Single Exit and Control Regions in Linear Time," technical report, Ithaca, NY, USA, 1993.
9. G. A. Mills-Tettey, A. Stentz, and M. B. Dias, "The Dynamic Hungarian Algorithm for the Assignment Problem with Changing Costs," Tech. Rep. CMU-RI-TR-07-27, Robotics Institute, Pittsburgh, PA, July 2007.