Selection of Business Process Alternatives Based on Operational Similarity for Multi-subsidiary Organizations

Avi Wasser and Maya Lincoln

University of Haifa, Israel
{awasser,mlincoln}@haifa.ac.il

Abstract. This work suggests a method for machine-assisted support for multi-subsidiary organizations in selecting business process alternatives, based on operational similarity. Operational similarity between processes can be derived from process repositories using a linguistic analysis of process descriptors. The suggested method aims to assist operation managers in multi-subsidiary organizations in identifying similar processes, that can substitute processes that cannot be carried out within a certain subsidiary. This decision making is based on knowledge that is encapsulated within existing business process repositories. The method is demonstrated using a real-life process repository from the paper manufacturing industry.

Keywords: Business process decisions, Process similarity, Multi-subsidiary organizations, Business process repositories, Semantic analysis, Process ontologies, Natural language processing.

1 Introduction

In many cases multi-subsidiary organizations perform similar processes in different geographical locations. For example, an insurance company may operate in several branches, each delivers a similar product (insurance) by conducting different processes for achieving a common goal. In some cases, such "parallel" processes can even be very different. For example, pharmaceuticals companies may manufacture different medicines in different factories- using a totally different process.

Due to various reasons, there are situations in which an organizational subsidiary cannot carry out a process. For example, due to a problem in one of its machines, a pharmaceuticals factory may not be able to proceed with its manufacturing process. In these cases the enterprise may need to perform the process in one of its other subsidiaries. Since each such subsidiary has a different modus-operandi, it is required to compare between the subsidiaries' processes and select the subsidiary that operates in the most similar way to the "blocked" process.

Research in this field has focused mainly on semantic (textual) similarity analysis (e.g. [3,4]) or structural similarity (e.g. [15]) between processes. Despite

Y.T. Demey and H. Panetto (Eds.): OTM 2013 Workshops, LNCS 8186, pp. 579–586, 2013.

the efficiency of these methods, they are not always relevant for the problem in hand: in some cases minor semantic differences may indicate major operational differences. For example, two manufacturing processes can be highly similar in textual terms, except one difference in one of their activities' tag: instead of "ink cleaning" in one process - the other process involves "ink injection." In this case it is possible that another process that contains more non-common activities, but involves the activity "ink injection," may be more similar to the blocked process n operational terms, and therefore may be more adequate to replace it.

This work aims to take state-of-the-art process similarity comparison methods a step forward by: (1) proposing a machine assisted mechanism that will take into account the operational characteristics of process models and suggest a ranked list of similar processes in operational terms; and (2) applying the suggested framework on real-life processes.

The suggested method aims to assist operation managers in identifying similar processes in case the original processes cannot be carried out. The extended framework is illustrated throughout the paper using an example based on real-life processes from the manufacturing industry.

The paper is organized as follows: we present related work in Section 2, positioning our work with respect to previous research. In Section 3 we present models and a method for evaluating process model similarity as background to this work. We describe the method for analyzing operational similarity between process models in Section 4. We conclude in Section 5.

2 Related Work

Research on standardization and analysis of the content layer of business process models mainly focuses on the analysis of linguistic components - actions and objects that describe business activities. Most existing languages for business process modeling and implementation are activity-centric, representing processes as a set of activities connected by control-flow elements indicating the order of activity execution [15]. In recent years, an alternative approach has been proposed, which is based on objects (or artifacts/entities/documents) as a central component for business process modeling and implementation. This relatively new approach focuses on the central objects along with their life-cycles. Such object-centric approaches include artifact-centric modeling [10,1], data-driven modeling [9] and proclets [11].

Although most works in the above domain are either object or activity centric, only a few works combine the two approaches in order to exploit an extended knowledge scope of the business process. The work in [5] presents an algorithm that generates an information-centric process model from an activity-centric model. The works in [8,7,6] present the concept of business process descriptor that decomposes process names into objects, actions and qualifiers, and suggest several taxonomies to express the operational knowledge encapsulated in business process repositories. In this work we take this model forward by: (a) testing it on real-life processes from the high-tech domain; (b) showing how the suggested taxonomies can assist in common usages of business process management.

Research on process similarity mainly focuses on three metrics for measuring the similarity between business process models. The label-based metric tries to match the nodes in process models by comparing their labels. Based on this matching, a similarity score is calculated taking into account the overall size of the models [3,4]. The structure based metric compares between graphs based on graph-edit distance [2]. This metric takes into account both the node labels and the topology of the process models. The third metric - the behavioral one - takes into account the causal relations between activities in a process model. These causal relations are represented as a causal footprint [14,13].

Despite the efficiency of the above described methods, they are all based on the *semantics* of the business process model. Nevertheless, a *semantic*-based similarity does not necessarily indicate an *operational* similarity. Minor semantic differences may indicate major operational gaps.

In this work we combine the notions of state-of-the-art similarity analysis methods with the operational analysis model of business process content as proposed in [7] and extended in [16]. We also use the concepts presented in [17] for segmenting the operations of enterprises based on operational characteristics. The proposed framework takes state-of-the-art similarity comparison methods forward by: (a) comparing between process models based on *operational* (and not *textual*) similarity; and (b) testing the proposed method on real-life processes.

3 Background: Models and Methods for Evaluating Process Model Similarity

3.1 The Descriptor Model

This section describes a formal model of business process decomposition and analysis as presented in [8] and further developed in [7]. We first introduce the descriptor model, and then, based on the descriptor model, we introduce four taxonomies of objects and actions. Finally, we describe the descriptor space model, which will be used in the next section. To illustrate the taxonomies we make use of the paper manufacturing repository (see Section 1).

The Activity Decomposition Model. In the Process Descriptor Catalog model ("PDC") [8] each activity is composed of one action, one object that the action acts upon, and possibly one or more action and object qualifiers. Qualifiers provide an additional description to actions and objects. In particular, a qualifier of an object is roughly related to an object state. State-of the art Natural Language Processing (NLP) systems. For example, the activity "Manually clean assembly machine" generates an activity descriptor containing the action "clean," the action qualifier "manually," the object "machine" and the object qualifier "assembly."

The Action and Object Taxonomies. The descriptor model has two basic elements, namely objects and actions, and it serves as a basis for four taxonomies, namely: the *action hierarchy model*, the *object hierarchy model*, the *action sequence model*, and the *object lifecycle model* (see [7]).

The action and object hierarchy models organize a set of activity descriptors according to the hierarchical relationships among business actions and objects, respectively. This hierarchical dimension of actions and objects is determined by their qualifiers: an addition of a qualifier to an action or an object makes them more specific, since the qualifier limits their meaning to a specific range. In the action hierarchy model, for example, the action “Manually clean” is a subclass (a more specific form) of “Clean,” since the qualifier “Manually” limits the action of “Clean” to reduced action range.

The action sequence model organizes a set of activity descriptors according to the relationships among business actions and objects in terms of execution order. In this model, each object holds a graph of ordered actions that are applied to that object. For example, the object “Assembly machine”is related to the following action sequence: “Clean” followed by “Reset,” “Start,” and finally “Monitor.”

The object lifecycle taxonomy model organizes a set of activity descriptors according to the relationships among business actions and objects in terms of execution order. In this model, each object holds a graph of ordered objects that expresses the object’s lifecycle, meaning - the possible ordering of the object’s states. For example, the object “Assembly machine” is part of the following object lifecycle: “Uncleaned assembly machine”–> “Cleaned assembly machine”–>“Working assembly machine”–> “Stopped assembly machine.”

The Quad-Dimensional Descriptor Space. The Quad-Dimensional Descriptor Space is described in [7] as follows. Based on the activity decomposition model, it is possible to visualize the operational range of a business process model as a descriptor space comprised of related objects and actions. The descriptor space is a quad-dimensional space describing a range of activities that can be carried out within a process execution flow. The coordinates represent the object dimension, the action dimension, and their qualifiers. Therefore, each space coordinate represents an activity as a quadruple $AC = \langle O, OQ, A, AQ \rangle$, where O is an object, OQ is a set of object qualifiers, A is an action, and AQ is a set of action qualifiers.

For every two coordinates in the descriptor space a *distance function* is defined, representing a linear combination of changes within each of its dimensions. This distance is used in this work for calculating the operational similarity between labels in process models. For more details on how this function is calculated - see [7].

3.2 Structure based Similarity between Graphs

The work in [2] presents a graph distance measure that is based on the structure similarity between graphs, and more specifically, on the maximal common subgraph of two graphs. In that work each graph comprises labeled nodes and edges. Labels are matched only when they contain identical texts.

We will use the above graph distance measure to calculate the similarity between process models, with one main difference: labels on vertices will be compared based on their operational distance (see Section 3.1).

4 Method for Evaluating Operational Similarity of Process Models

In this section we present a framework for supporting decisions related to the selection of alternative processes when the original process is blocked. The framework receives the name of a blocked process as input and outputs a sorted list of alternative process models - available at the process repository of a multi-subsidiary organization.

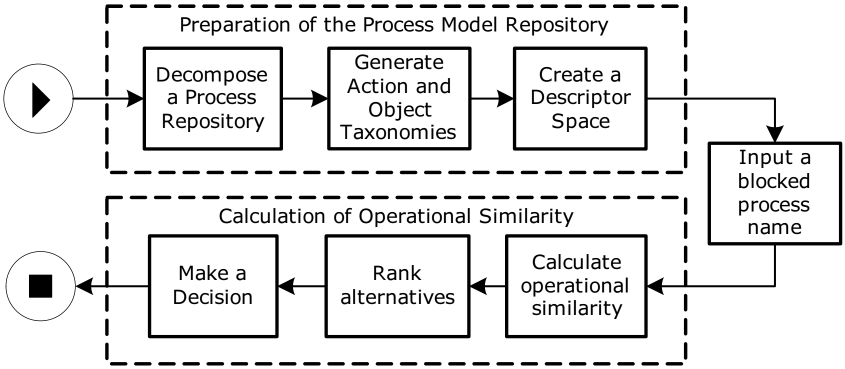

Fig. 1. A framework for evaluating operational similarity of process models

We propose a seven-step framework to support the selection of business process alternatives in multi-subsidiary organizations, as illustrated in Fig. 1 (using "Yet Another Workflow Language" (YAWL) [12]). First, we create an operationally meaningful decomposition of a process repository. We use state-of-the-art Natural Language Processing (NLP) techniques to automatically decompose the content layer (text) of the repositories into its structured linguistic components (objects and actions and their qualifiers). As part of this decomposition, each business activity is encoded automatically as a *descriptor*, using the Process Descriptor Catalog ("PDC") notation (see Section 3.1).

Second, we analyze the generated decomposition and create four action and object models, that represent operational aspects of the process repository (Section 3.1). As a third step we build a descriptor space using the four action and object taxonomies (see Section 3.1). The above first three steps are conducted automatically on a given process model repository of a multi-subsidiary organization, and are considered as "preparation" phases - since they are conducted only once (or after each time the repository changes).

As a fourth step, an operations manager inputs a blocked process name. This name can be given either as a descriptor or in natural language. If the second option is being used- the name is automatically decomposed into a descriptor format. The next three steps aim at producing a list of alternative process models that can replace the original, blocked process, as follows.

In the fifth step the blocked process model is compared to each process in the repository-in all subsidiaries, including the subsidiary of the blocked process itself (since maybe an alternative process from this repository will be found useful as well). The comparison between each process model pairs is performed using the method suggested in [2] with the following modification: labels on vertices are compared based on their operational distance in the descriptor space. In case the distance is different than "no distance" (see [7]), it means that there is some level of similarity between the nodes' functionality, and therefore the comparison algorithm proceeds until it reaches a "no distance" node. As a result, this step outputs for each process model in the repository: (1) its maximal common subgraph comparing to the blocked process model, and (2) a set of distances - calculated for each matched node pairs.

The sixth step is aimed at ranking the process models in the repository according to their operational similarity to the blocked process model. In general, we aim at a maximal size of the common subgraph, with higher similarity scores of its vertices. The second consideration is more important for our goal: in case of a trade-off, we prefer fewer matched nodes with higher similarity than the opposite - since the first case allows the subsidiary to perform its process partially, while the second only allows it to perform a larger part of its process - but differently. Due to the above considerations, the total grade for each process model is calculated as follows: (1) eliminate all options where the maximal common subgraph contains less than x% of the blocked process size (x is a parameter that should be determined according to the operations manager preferences); (2) For each of the remaining process models: sort the set of the nodes' similarity scores - where the highest score is at the top of the list; and (3) Sort the remaining process models - so that a process model is ranked higher in the list if its node score list starts with higher scores. As a result, a list of alternative process models is generated and ranked according to the user's input, so that higher ranked suggestions are believed to be closer to the user needs. Finally, at the seventh step the operations manager receives suggestions and supporting information (the common subgraph) and makes his decision accordingly.

5 Conclusions

We proposed a framework for machine-assisted support for multi-subsidiary organizations in selecting process alternatives, based on operational similarity. The proposed framework provides a starting point that can already be applied in real-life scenarios, yet several research issues remain open. We mention two such extensions here. First, extending the comparison method by referring also to labels on edges, not only on vertices. Second, adding a case study and experiments to measure the efficiency of the proposed framework that will perform actual validation in a multi-subsidiary organizational environment.

References

1. Bhattacharya, K., Gerede, C.E., Hull, R., Liu, R., Su, J.: Towards formal analysis of artifact-centric business process models. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 288–304. Springer, Heidelberg (2007)
2. Bunke, H., Shearer, K.: A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters 19(3), 255–259 (1998)
3. Dijkman, R., Dumas, M., Dongen, B.V., Käärik, R., Mendling, J.: Similarity of business process models: Metrics and evaluation. Information Systems 36(2), 498–516 (2011)
4. Ehrig, M., Koschmider, A., Oberweis, A.: Measuring similarity between semantic business process models. In: Proceedings of the Fourth Asia-Pacific Conference on Comceptual Modelling, vol. 67, pp. 71–80. Australian Computer Society, Inc. (2007)
5. Kumaran, S., Liu, R., Wu, F.Y.: On the duality of information-centric and activity-centric models of business processes. In: Bellahsène, Z., Léonard, M. (eds.) CAiSE 2008. LNCS, vol. 5074, pp. 32–47. Springer, Heidelberg (2008)
6. Lincoln, M., Gal, A.: Searching business process repositories using operational similarity. In: Meersman, R., Dillon, T., Herrero, P., Kumar, A., Reichert, M., Qing, L., Ooi, B.-C., Damiani, E., Schmidt, D.C., White, J., Hauswirth, M., Hitzler, P., Mohania, M. (eds.) OTM 2011, Part I. LNCS, vol. 7044, pp. 2–19. Springer, Heidelberg (2011)
7. Lincoln, M., Golani, M., Gal, A.: Machine-assisted design of business process models using descriptor space analysis. In: Hull, R., Mendling, J., Tai, S. (eds.) BPM 2010. LNCS, vol. 6336, pp. 128–144. Springer, Heidelberg (2010)
8. Lincoln, M., Karni, R., Wasser, A.: A Framework for Ontological Standardization of Business Process Content. In: International Conference on Enterprise Information Systems, pp. 257–263 (2007)
9. Müller, D., Reichert, M., Herbst, J.: Data-driven modeling and coordination of large process structures. In: Meersman, R., Tari, Z. (eds.) OTM 2007, Part I. LNCS, vol. 4803, pp. 131–149. Springer, Heidelberg (2007)
10. Nigam, A., Caswell, N.S.: Business artifacts: An approach to operational specification. IBM Systems Journal 42(3), 428–445 (2003)
11. Van der Aalst, W.M.P., Barthelmess, P., Eliis, C.A., Wainer, J.: Proclets: A framework for lightweight interacting workflow processes. International Journal of Cooperative Information Systems 10(4), 443–482 (2001)
12. van der Aalst, W.M.P., Ter Hofstede, A.H.M.: YAWL: yet another workflow language. Information Systems 30(4), 245–275 (2005)
13. van Dongen, B.F., Dijkman, R., Mendling, J.: Measuring similarity between business process models. In: Bellahsène, Z., Léonard, M. (eds.) CAiSE 2008. LNCS, vol. 5074, pp. 450–464. Springer, Heidelberg (2008)
14. van Dongen, B.F., Mendling, J., van der Aalst, W.M.P.: Structural patterns for soundness of business process models. In: 10th IEEE International Enterprise Distributed Object Computing Conference, EDOC 2006, pp. 116–128. IEEE (2006)

15. Wahler, K., Kuster, J.M.: Predicting Coupling of Object-Centric Business Process Implementations. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 148–163. Springer, Heidelberg (2008)
16. Wasser, A., Lincoln, M.: Semantic machine learning for business process content generation. In: Meersman, R., Panetto, H., Dillon, T., Rinderle-Ma, S., Dadam, P., Zhou, X., Pearson, S., Ferscha, A., Bergamaschi, S., Cruz, I.F. (eds.) OTM 2012, Part I. LNCS, vol. 7565, pp. 74–91. Springer, Heidelberg (2012)
17. Wasser, A., Lincoln, M., Karni, R.: Accelerated enterprise process modeling through a formalized functional typology. In: van der Aalst, W.M.P., Benatallah, B., Casati, F., Curbera, F. (eds.) BPM 2005. LNCS, vol. 3649, pp. 446–451. Springer, Heidelberg (2005)