A Comparative Analysis of Business Process Model Similarity Measures

Tom Thaler[1(✉)], Andreas Schoknecht[2], Peter Fettke[1], Andreas Oberweis[2], and Ralf Laue[3]

[1] German Research Center for Artificial Intelligence (DFKI),
Saarland University, Saarbrücken, Germany
{tom.thaler,peter.fettke}@dfki.de
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
{andreas.schoknecht,andreas.oberweis}@kit.edu
[3] University of Applied Sciences Zwickau, Zwickau, Germany
ralf.laue@fh-zwickau.de

Abstract. To work efficiently with and unlock the potentials of business process models, measuring their similarity is a basic requirement. Thus, many automatic similarity measurement approaches have been developed during the last years, which utilize very different aspects of a model. At the same time, it is unclear which measures can be meaningfully applied in which context and how they behave in general. Hence, this paper analyzes how the values of existing similarity measures correlate and how corresponding implementations perform with respect to their resource consumption. The results of our analysis show that the similarity values of most measures highly correlate while their performance prohibits the usage of more than 50% of the measures in practice.

Keywords: Process model similarity · Process model matching

1 Introduction

Regarding a practical usage of business process models, there are many scenarios, wherefore measuring the similarity between particular models is even a basic requirement. This includes checking the conformance of process models to legal regulations or to reference models, enabling the reusability of process fragments or merging process models. At the same time, a manual measurement of similarity values and differences between hundreds or even thousands of models would take an enormous effort leading to high costs [3].

Thus, many business process model similarity measures have been developed during the last years. However, the interpretation of similarity is quite different, since several dimensions of similarity can be considered. While these are not dimensions in a strong mathematical sense, they focus on different aspects of process models (cf. Fig. 1). Since the activites in *seller p. 1* and *seller p. 2* are identically labelled, the similarity value would be 1 when only considering the natural language dimension. Yet, their graph structure and behavior are slightly
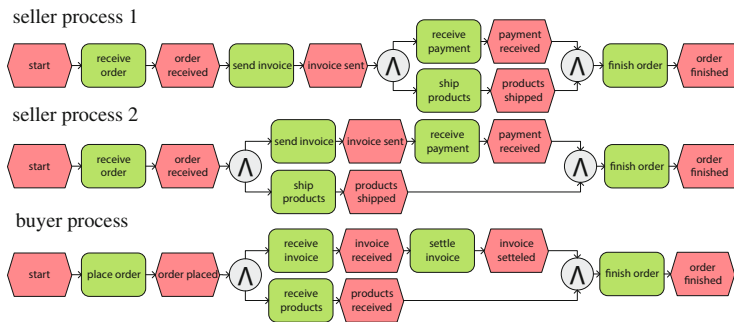
Fig. 1. Illustration of different dimensions used for similarity measurement (cf. [11]).

different. E.g., in *seller p. 1* the activity "send invoice" is always executed before "ship products" while this is not necessarily true for *seller p. 2*. Thus, the final similarity value might be less than 1. When looking at the process models *seller p. 2* and *buyer p.*, they are identical with respect to the graph structure, which would lead to a similarity value of 1 when only considering this dimension.

Considering this variety of dimensions, which are possibly addressed by existing similarity measures, it is unclear which measures can be meaningfully applied in which usage scenario and how they behave in general. Therefore, we make a first step in this direction by addressing the following questions: (1) How do the values of existing similarity measures correlate? (2) How do existing implementations perform and what does that imply for their practical usage?

The rest of the paper is organized as follows: In Sect. 2, the relevant core terms are introduced and explained in detail. Section 3 outlines related work, followed by the applied research approach in Sect. 4. Afterwards, Sect. 5 describes the results of the comparative analysis, which covers a correlation analysis of similarity values as well as a run time analysis of eight publicly available measures. The results and limitations of the paper are discussed in Sect. 6. Finally, a conclusion is given in Sect. 7.

2 Fundamentals of Business Process Model Similarity Measurement

Business process model similarity measures try to quantify the similarity between business process models in general. A similarity value is mostly expressed either on an interval or on a ratio scale, which provide the frame for the typical operationalization of business process model similarity in a metric space. A metric generally fulfills the properties non-negativity ($\forall x, y \in \mathcal{D}, d(x,y) \geq 0$), symmetry ($\forall x, y \in \mathcal{D}, d(x,y) = d(y,x)$), identity ($\forall x, y \in \mathcal{D}, x = y \Leftrightarrow d(x,y) = 0$) and triangle inequality ($\forall x, y, z \in \mathcal{D}, d(x,z) \leq d(x,y) + d(y,z)$) [24], with $\mathcal{D}$ being a domain of objects and $d : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ being a function measuring the distance between two objects. However, as shown in [14], many existing process model similarity measures do not fulfill the above mentioned properties. Only three of eleven analyzed measures met all metric properties.

The quantification of process model similarity is based on process model matches in most cases [4]. Such matches explicate correspondences between single nodes or sets of nodes of two models based on criteria such as similarity, equality or analogy [20] and are expressed using different cardinalities (1:1, 1:N, M:N). As shown in [4], at that time, most similarity measures used matching approaches creating solely 1:1 matches, while recent works tend to M:N matches [2,6]. As mentioned above, there are several dimensions of similarity in business process models, which are now briefly introduced.

Natural language: Natural language is, e.g., used for labeling the elements of a model, for naming a model or for tagging it. Such labels serve as one of the most important knowledge bases for process model similarity measurement and are analyzed with regard to syntactical and semantic aspects. While the syntactical analysis focuses on the characters of labels, the semantic analysis aims at understanding their meaning based on the used words and the grammar.

Graph structure: The relevant aspects of this dimension arise from graph theory and can be divided into general graph structure-based and business process aware control flow similarity measures. The graph structure-based similarity between models can be quantified by, e.g., quantitative metrics related to common subgraphs. Since general graph-based algorithms do not consider the meaning of any control flow connectors, they are either ignored or the existing measures are extended in order to handle them.

Behavior: This dimension focuses on the execution of business processes. Corresponding execution traces can be generated through simulation runs or during the actual execution of a process. An example for such a similarity measurement is to count the number of identical execution sequences in a trace log. Thereby, also characteristics of possible execution sequences are considered.

Human estimation: Another dimension is the human judgment on how similar process models are. One can differentiate three types of human estimation based on the knowledge of the involved people: Process experts have a grounded knowledge on the process landscape of a company, while process participants are specialists for particular processes or parts thereof. Thus, it can be assumed that process experts quantify the similarity on a general and more high level point of view, while process participants take up a detailed perspective. In contrast to that, the crowd gains its knowledge solely based on the process description (the models). Hence, the crowd quantifies the similarity with its own interpretation at the back of their mind.

Other aspects: Other aspects that are described in the literature are collected in this group, as they are infrequently used or are specific to a certain similarity measurement approach. Examples are the usage of ontology alignment techniques [5] or resources as, e.g., input and output objects [13].

3 Related Work

Currently, two structured literature surveys on process model similarity have been published [4,18]. Besides these, a few other articles give an overview on published similarity measures as well as matching techniques. One highlights open questions regarding similarity measurement [8] and another three compare different approaches in evaluation settings [2,6,9]. Becker and Laue [4] provide a detailed overview on the exact calculations used by process model similarity measures without considering their behavior in realistic scenarios. The survey conducted by Niesen and Houy [18] focuses on Natural Language Processing (NLP) techniques used in process model similarity measures. Again, the survey of Dijkman et al. [8] describes categories of problems related to similarity measurement as well as future research directions. While these surveys provide a theoretical overview on process model similarity, our article compares different similarity measures regarding their run time and similarity values. Hence, we focus on an empirical assessment point of view.

Furthermore, the evaluations described in [2,6] can be regarded as related work as process model matching techniques are summarized and compared to each other. Thus, these papers also constitute practical, empirical works. They present results of two process model matching contests, wherein several matching approaches were compared regarding their performance of finding similar activities in several process model data sets. The closest work to ours is the comparison of three process model similarity measures for process model retrieval, which are evaluated with one data set described in [9]. In contrast to this work, we additionally analyze the correlation between the similarity values of eight publicly available measures on three different data sets.

4 Analysis Objectives and Methodology

4.1 Selection of Model Data

Although a representative data set is not achievable from a statistical point of view (the overall population of process models is unknown), an experimental analysis of similarity measures is necessary to characterize their behavior in concrete application scenarios. For that purpose, one can distinguish laboratory and field investigations. In laboratory investigations the process models are (possibly synthetically) generated in a controlled environment, while in field investigations, they are generated by modelers in the real world. However, results from a laboratory investigation cannot always easily be transferred to the field. Against that background, the field should be considered as well. Hence, we use three different groups of samples with different characteristics, which are taken from a large process model corpus [19]. The data sets are described below, several corresponding model metrics are presented in [19].

1. **Field models:** No restrictions regarding the labeling of model elements are given to the modeler(s). Thus, equal or similar aspects might be modeled

in a different manner and expressed with different words. An adequate data set, containing models from the domains university admission (9 models) and birth registration (9 models), is provided in [6].

2. **Models from controlled modeling environments:** Models are created in a controlled environment, wherein different modelers independently model the same process based on a natural language text description. Since, in this way, a terminology is given by the textual description, it is assumed that it is used by the modelers. Student exercises[1] (8 models) serve as an adequate data set. An analysis based on this data set covers a laboratory investigation.

3. **Mined models:** The process models are derived using process mining techniques. Thus, the node labels are linguistically harmonized and therefore (1) unambiguous and (2) consistent over the whole collection (matching problem is faded out). The models from Dutch governance presented in [21] fulfill this requirement (80 models). However, one can discuss whether they are synthetically created in a laboratory sense or, since the processes are executed in the real world, whether they are derived from field.

4.2   Selection and Setup of Similarity Measures

In order to identify relevant similarity measurement approaches we conducted a structured literature search [22], which led to 120 relevant papers.[2] Based on this selection, all papers with an existing or known implementation, were selected. If available, the tools of the remaining 16 candidates were checked, whether either Petri Net or EPC is supported, since the models used in our study are available in these notations. Finally, the selected similarity measures address the dimensions mentioned in Sect. 2 as described in Table 1 and were set up as follows:

Similarity Score based on Common Activity Names [1]:[3] The similarity is calculated based on the number of identically labeled activities.

Graph Edit Distance Similarity [7]**(see Footnote 3):** The concept of edit distance is applied to both, node labels (string edit distance) and the graph structure (graph edit distance). The Greedy algorithm is used for the (approximate) optimization of the similarity matrix. The three mentioned quotients are equally weighted.

Causal Footprints [10]**(see Footnote 3):** The similarity measure was implemented in the research prototype RefMod-Miner. Although there is a proposal of a semantic node similarity measure, in the context at hand, two nodes are considered as equal (matched), iff they have the same label.

Percentage of Common Nodes and Edges [17]**(see Footnote 3):** This similarity measure enhances [1] by regarding all nodes and edges of process models

---

[1] Model set "Exams" is available in the model repository at http://rmm.dfki.de.

[2] A complete list of investigated publications can be retrieved from http://rmm.dfki.de/docs/BPM2016_ProcessModelSimAnalysis_LiteratureList.pdf.

[3] http://rmm.dfki.de.

Table 1. Dimensions addressed by selected measures.

Dimension/Ref	[1]	[7]	[10]	[17]	[23]	[15]	[16]	[12]
Natural language	syn	syn	syn	syn	syn	syn	syn+sem	syn
Graph structure		x	x	x	x	x	x	
Behavior								x

(instead of activities only). Thus, the structure of a process model is considered as well. The control flow connectors are ignored such that their preceding and succeeding nodes are interpreted as directly connected through edges.

Feature-based Similarity Estimation [23] **(see Footnote 3):** The approach consists of syntactical and graph-structural components. Firstly, for each node pair, the Levenshtein similarity for their labels is calculated (syntactical component). As a graph-structural component, the authors define five roles characterizing a node. The thresholds are set as proposed in the original paper. The resulting similarity matrix is optimized using the Greedy algorithm.

Activity Matching and Graph Edit Distance [15] **(see Footnote 3):** This measure extends [7] by also considering control flow connectors. The Greedy algorithm is again used for optimizing the similarity matrix. All other settings are equal to the above mentioned graph edit distance similarity.

La Rosa Similarity [16]:[4] This similarity measure extends [15] by calculating a node matching not only based on the Levenshtein similarity but also using a linguistic similarity measure based on a lexical database. The original implementation is used with the standard parameters.

Longest Common Sets of Traces [12] **(see Footnote 3):** The authors propose two components expressing in how far the traces of one model are reflected by the traces of the other model. To make the similarity values comparable, the average of both components is calculated and interpreted as the similarity value. The node mapping (which is not explicated in the original paper) is calculated using the Levenshtein similarity with a minimum threshold of 0.9.

5   Analysis Results

Overall, we analyzed 3,371 model pairs with 8 similarity measures and thereby proceeded 26,968 similarity calculations plus the corresponding node matchings. Since the practical applicability is one important aspect, the analysis was executed on a standard PC with 4 cores (3 GHz each) and 4 GB main memory.

Correlation between Similarity Values: Since all analyzed techniques use node matchings as a calculation basis, they highly depend on the corresponding

---

[4] http://www.processconfiguration.com/download.html.

node matching approach. Against this background, it is assumed that the correlation between the measures' values differ based on the data sets. While mined models might reliably be matched by all approaches, quite diverging matchings between non-linguistically harmonized models are expected. Thus, the similarity values are evaluated separately on the above mentioned model categories (1) field models, (2) models of controlled modeling environments and (3) mined models. Since the underlying node matching is of major importance, they were additionally analyzed with regard to known reference matchings considering the activities only. Table 2 includes only six mapping approaches instead of eight, since [1,10,17] use the same approach. Moreover, it looks like [7,15,16] produce exactly the same matchings in all data sets. This is in fact not the case, since only the activities are considered in the reference matchings, while the named approaches also match events or connectors, which are not covered by the analysis. Furthermore, we chose the Pearson correlation coefficient since we assume a normal distribution of the similarity values. In fact, it is not possible to determine the distribution in a methodical correct way, since, as mentioned above, it is not possible to randomly select process models from the overall population of process models. Nevertheless, the Pearson correlation coefficient is suitable for interpreting the correlation behavior for the selected data sets.

Table 2. Analysis of underlying matchings with regard to reference activity matchings.

Approach	Field						Controlled						Mined					
	TP	FP	FN	P	R	F	TP	FP	FN	P	R	F	TP	FP	FN	P	R	F
[1,10,17]	152	17	962	0.9	0.14	0.24	6	0	284	1.00	0.02	0.04	9,767	0	0	1.00	1.00	1.00
[7]	289	205	825	0.59	0.26	0.36	96	33	194	0.74	0.33	0.46	9,767	9,187	0	0.52	1.00	0.68
[23]	315	906	799	0.27	0.28	0.27	125	228	165	0.35	0.43	0.39	9,554	50,354	213	0.16	0.98	0.27
[15]	289	205	825	0.59	0.26	0.36	96	33	194	0.74	0.33	0.46	9,767	9,187	0	0.52	1.00	0.68
[16]	289	205	825	0.59	0.26	0.36	96	33	194	0.74	0.33	0.46	9,767	9,187	0	0.52	1.00	0.68
[12]	175	20	939	0.90	0.16	.027	19	1	271	0.95	0.07	0.12	9,767	1,257	0	0.89	1.00	0.94

sim = similarity measure, TP = true positives, FP = false positives, FN = false negatives, P = μ-average of precision, R = μ-average of recall, F = μ-average of f-measure.

As one can see in Table 3 and contrary to the expectations, a higher correlation of the similarity values for mined models in comparison to the other two data sets cannot be identified. As one can easily see in Table 2, the mapping quality in case of the mined models is, as expected, much higher than for the other two cases. At the same time, a higher mapping quality of the controlled models to the field models cannot generally be attested. However, that does not influence the correlation between the similarity values, since the quality of the different approaches within the different scenarios are comparable in most cases. Especially the analysis of [1,10,17] in case of the mined models is very meaningful, since they have a perfect mapping (a linguistic harmonization corresponds to a label-identical mapping). Although the first approach solely considers the equally labeled activities, while the others also consider the control flow, there is a high correlation in the resulting similarity values.

Furthermore the presented heat map shows a very high correlation between the similarity values of all measures except of [23]. As one can see in Table 2, the matching approach produces at least four times more false positive matches than all other approaches, which is also the reason for the missing significance of the correspondent correlation values. That underlines the thesis of the high dependency of a similarity measure to the underlying node mapping. Particularly, a cluster consisting of [1, 7, 15–17] can be identified. Therein, all similarity measure pairs correlate to more than 0.95 on average. This is surprising since [1] solely considers node labels for the similarity quantification, while the other four approaches also take the structure into account.

Considering the measures [10, 23], it is conspicuous that both show a comparable low correlation to most other similarity measures, but also between each other. As that is founded in the different matching approach of [10, 23] seems to measure a specific aspect of similarity. In fact, [10] bases on the causal footprints, and thus considers the causal dependencies between nodes, while [23] considers the correspondences between nodes similar to [1].

However, looking at the very low mapping quality of [1, 10, 12, 17] in the controlled scenario, a general effect on their correlation cannot be identified. As a further intermediate result, except of [10, 23], there is a very high correlation between all analyzed measures. The expected lower correlation between measures focusing on the behavior of possible process executions in comparison to those focusing solely on the labels could not be experimentally verified. Hence, except of the [10, 23], the analyzed measures are scarcely distinguishable based on their values and are therefore exchangeable in the demonstrated cases. Hence, except of [23], the similarity measures seem to be exchangeable as their values correlate to a very high degree. Other aspects such as the run time of similarity calculation could therefore be more important when choosing a measure for a specific application.

Computing Performance: The second analysis aspect of the comparative analysis is the computing performance in terms of run time. The importance of this aspect is founded in the practical applicability, for which a calculation time of several minutes or hours would generally not be desirable. As expected, based on the number of models within the data sets *University admission*, *Birth registration*, and *Student exercises*, they are suitable for fast calculations. Nearly all measures returned the similarity values in less than one minute, 50% of the calculations under five seconds.

Considering real model databases, perhaps containing thousands of individual models, it seems that more than 50% of the analyzed similarity measure implementations [10, 12, 15, 16, 23] are not suitable for an application in real contexts. On the other hand, there are three approaches [1, 7, 17] which generally provide results in short time. Especially [1, 17] have high potential for an application in real contexts, since next to the short calculation time, they also constitute the highest correlation to all other measures (Table 4).

Table 3. Pearson correlation coefficients for the analyzed data sets.

		[1]	[7]	[10]	[17]	[23]	[15]	[16]	[12]
[1]	F	1.00	0.93	0.77	0.97	0.70	0.96	0.94	0.94
	C	1.00	0.94	0.80	0.98	0.43*	0.97	0.96	0.92
	M	1.00	0.96	0.85	0.98	0.60	0.95	0.93	#
[7]	F	0.93	1.00	0.91	0.98	0.62	0.97	0.98	0.98
	C	0.94	1.00	0.91	0.95	0.49*	0.98	0.94	0.97
	M	0.96	1.00	0.87	0.99	0.55	0.94	0.98	#
[10]	F	0.77	0.91	1.00	0.86	0.55	0.85	0.87	0.85
	C	0.80	0.91	1.00	0.95	0.49	0.98	0.89	0.96
	M	0.85	0.87	1.00	0.85	0.49	0.82	0.84	#
[17]	F	0.97	0.98	0.86	1.00	0.65	0.98	0.98	0.97
	C	0.98	0.95	0.95	1.00	0.56*	0.99	0.97	0.91
	M	0.98	0.99	0.85	1.00	0.57	0.94	0.97	#
[23]	F	0.70	0.62	0.55	0.65	1.00	0.75	0.63	0.63
	C	0.43*	0.49*	0.49	0.56*	1.00	0.55	0.65	0.55*
	M	0.60	0.55	0.49	0.57	1.00	0.76	0.57	#
[15]	F	0.96	0.97	0.85	0.98	0.75	1.00	0.97	0.96
	C	0.97	0.98	0.98	0.99	0.55	1.00	0.97	0.95
	M	0.95	0.94	0.82	0.94	0.76	1.00	0.93	#
[16]	F	0.94	0.98	0.87	0.98	0.63	0.97	1.00	0.98
	C	0.96	0.94	0.89	0.97	0.65	0.97	1.00	0.96
	M	0.93	0.98	0.84	0.97	0.57	0.93	1.00	#
[12]	F	0.94	0.98	0.85	0.97	0.63	0.96	0.98	1.00
	C	0.92	0.97	0.96	0.91	0.55*	0.95	0.96	1.00
	M	#	#	#	#	#	#	#	#

p-value ≤1%, F = field models, C = controlled models, M = mined models, # = calculation aborted because of memory overflow, * = p-value >1%.

In contrast to that, it was not possible to calculate any similarity values for the *Dutch Governance* data set with the approach of [12]. Since there are no real log traces available, referring to [4], all possible traces were calculated. Depending on the size and on the complexity of the input models, this approach produced a mass of data leading to a memory overflow on the used hardware. Thus, the similarity values of [12] could not be calculated for this data set. All other run times for the data sets could be calculated and for two of them even in suitable time (*Student exercises* and *Birth registration*). At the same time, the usage of real execution data might improve this approach to a high degree. Nevertheless, such an analysis is not part of the work at hand.

Table 4. Run time of similarity calculations for different data sets.

Measure	Dutch Governance	Student exercises	Birth registration	University admission
[1]	3:28 min	0:00 min	0:02 min	0:02 min
[7]	8:40 min	0:01 min	0:04 min	0:05 min
[10]	n/a[a]	0:37 min	9:32 min	26:30 min
[17]	8:40 min	0:02 min	0:04 min	0:05 min
[23]	45:37 min	0:03 min	0:23 min	0:56 min
[15]	40:21 min	0:03 min	0:15 min	0:36 min
[16]	39:22 min	0:14 min	0:20 min	0:22 min
[12]	memory overflow	0:03 min	0:07 min	4:52 min

[a]For [10] the Dutch Governance processing had to be split because of a memory overflow. Since summing up the partial run times might have led to a corruption in comparison to the other calculations, it was decided to state it as not available.

6   Discussion and Limitations

Unfortunately, the availability of similarity measure implementations is quite limited. Only 22 implementations were mentioned in the publications, whereof only 8 were accessible and executable in the context of the analysis. Thereby, although the behavior-oriented measure originally works on process instances from real executions, here, all possible traces were derived from the models. This covers a slightly different case, since we implicitly considered the state space of the models instead of the observed behavior. Using real traces may lead to much lower calculation times and will most certainly lead to much lower memory consumption. On the other hand, using process logs would not measure the process model similarity but the process instance similarity. Against that background, the applied variant makes sense in the context of the work at hand.

We identified high differences in the intensity in terms of memory and time consumption. Both effects lead to trouble in the context of a practical application right up to a non-applicability. In turn, other approaches are able to calculate a similarity value within short time and with only little resources. In spite of these differences, the correlation between the similarity values is high in most cases.

The high correlation values are caused by the fact that all approaches work with an underlying node mapping, which is finally responsible for the similarity values. This is founded in the fact that the similarity measures are functions on the matchings. This leads to the result that it is necessary to separate process model similarity measurement into two components: (1) the node matching and (2) the calculation of a similarity value. As shown in the analysis at hand, this makes it possible to analyze the effects of addressing different dimensions of similarity in more detail. In fact, it might be meaningful to repeat the analysis using consistent matchings. However, the underlying scientific papers propose particular matching approaches, which is the reason for the design at hand. Moreover,

the proposed procedure would lead to new challenges since the cardinality of node matchings influences the applicability of a measure. Most of them need a 1:1 matching (as, e.g., in the context of behavior based similarity measures) while complex matchings cannot easily be interpreted. At the same time, 1:1 matchings would implicitly formulate the requirement that the models need to be on the same level of detail, which is generally not given in a realistic scenario.

7 Conclusion

Based on the practical empirical evaluation, it can be stated that the computational behavior of similarity measures is diverging in concrete contexts. First of all, the conceptualization of the measures has a high impact on the execution time, which ranges from 3 to 45 min up to non-computability for the similarity measurement of a set of 80 models. Besides, it was shown that the values of most measures highly correlate to each other. However, there were also two measures showing differences to the others. Thus, there are different types of similarity measures, which might reasonably be applied in different contexts.

One special scenario might be the similarity analysis of process models which are derived through process mining. Since the data basis is automatically generated, the contained information are linguistically harmonized. Hence, the analysis of node labels with NLP techniques is of minor importance, while the usage of further information like system handbooks might be meaningful. However, because of the generally high correlations, it is recommendable to apply one of the easier and faster measures like [1] in order to get a first impression of similarity between particular process models. Only if one is interested in details, and if a reliable matching is available, it is meaningful to apply a similarity measure addressing specific dimensions. For that, the *first impression* might be seen as a preselection of relevant models.

Yet, it is still an open question whether two similarity measures measure the same pragmatic aspects as, e.g., similarity of content, of the equivalence of action, or the equivalence of objective (in contrast to the above mentioned dimensions) and how that can be determined. It is also unclear how far the automatically calculated similarity values correspond to human estimations, and thus, how valid the similarity values are. In fact, the results of the investigated similarity measures are valid with regard to their technical implementation, but how far that matches specific measurement objectives, perhaps in different application scenarios, is not analyzed so far. Amongst others, one reason is, that the requirements of different application scenarios to a similarity measure are unclear and not precisely defined. Especially concerning the underlying node mapping, it should be clarified, what a correspondence is. E.g. in case of the University admission processes (field models), some universities interview the applicants, while others prefer aptitude tests. There are good arguments for and against a match [20]. Hence, it is necessary to obtain a deeper understanding on what should be understood as a correspondence and what types of correspondences do exist. If such an understanding is reached, an application of established methods

for the evaluation of process model similarity measures, e.g., in terms of validity, reliability, and objectivity, might be possible. This would considerably improve the appreciation of the capabilities of automatic similarity measurements.

References

1. Akkiraju, R., Ivan, A.: Discovering business process similarities: an empirical study with SAP best practice business processes. In: Maglio, P.P., Weske, M., Yang, J., Fantinato, M. (eds.) ICSOC 2010. LNCS, vol. 6470, pp. 515–526. Springer, Heidelberg (2010). doi:10.1007/978-3-642-17358-5_35
2. Antunes, G., et al.: The process model matching contest 2015. In: 6th International Workshop on Enterprise Modeling and Information Systems Architectures, pp. 127–155 (2015)
3. Becker, J., Delfmann, P., Dietrich, H.A., Steinhorst, M., Eggert, M.: Business process compliance checking - applying and evaluating a generic pattern matching approach for conceptual models in the financial sector. Inf. Syst. Front. **18**, 1–47 (2014)
4. Becker, M., Laue, R.: A comparative survey of business process similarity measures. Comput. Ind. **63**(2), 148–167 (2012)
5. Brockmans, S., Ehrig, M., Koschmider, A., Oberweis, A., Studer, R.: Semantic alignment of business processes. In: 8th International Conference on Enterprise Information Systems, pp. 191–196 (2006)
6. Cayoglu, U., et al.: Report: the process model matching contest 2013. In: Lohmann, N., Song, M., Wohed, P. (eds.) BPM 2013. LNBIP, vol. 171, pp. 442–463. Springer, Cham (2014). doi:10.1007/978-3-319-06257-0_35
7. Dijkman, R., Dumas, M., García-Bañuelos, L.: Graph matching algorithms for business process model similarity search. In: Dayal, U., Eder, J., Koehler, J., Reijers, H.A. (eds.) BPM 2009. LNCS, vol. 5701, pp. 48–63. Springer, Heidelberg (2009). doi:10.1007/978-3-642-03848-8_5
8. Dijkman, R.M., van Dongen, B.F., Dumas, M., García-Bañuelos, L., Kunze, M., Leopold, H., Mendling, J., Uba, R., Weidlich, M., Weske, M., Yan, Z.: A short survey on process model similarity. In: Bubenko, J., Krogstie, J., Pastor, O., Pernici, B., Rolland, C., Sølvberg, A. (eds.) Seminal Contributions to Information Systems Engineering, pp. 421–427. Springer, Heidelberg (2013)
9. Dijkman, R.M., Dumas, M., van Dongen, B.F., Käärik, R., Mendling, J.: Similarity of business process models: metrics and evaluation. Inf. Syst. **36**(2), 498–516 (2011)
10. Dongen, B., 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). doi:10.1007/978-3-540-69534-9_34
11. Fettke, P., Vella, A., Loos, P.: From measuring the quality of labels in process models to a discourse on process model quality: a case study. In: 45th Hawaii International International Conference on Systems Science, pp. 197–206 (2012)
12. Gerke, K., Cardoso, J., Claus, A.: Measuring the compliance of processes with reference models. In: Meersman, R., Dillon, T., Herrero, P. (eds.) OTM 2009. LNCS, vol. 5870, pp. 76–93. Springer, Heidelberg (2009). doi:10.1007/978-3-642-05148-7_8
13. Juntao, G., Li, Z., Wei, J.: Procuring requirements for ERP software based on semantic similarity. In: International Conference on Semantic Computing, pp. 61–70 (2007)

14. Kunze, M., Weidlich, M., Weske, M.: Behavioral similarity – a proper metric. In: Rinderle-Ma, S., Toumani, F., Wolf, K. (eds.) BPM 2011. LNCS, vol. 6896, pp. 166–181. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23059-2_15

15. La Rosa, M., Dumas, M., Uba, R., Dijkman, R.M.: Merging business process models. In: On the Move to Meaningful Internet Systems, pp. 96–113 (2010)

16. La Rosa, M., Dumas, M., Uba, R., Dijkman, R.M.: Business process model merging: an approach to business process consolidation. ACM Trans. Softw. Eng. Methodol. **22**(2), 11:1–11:42 (2013)

17. Minor, M., Tartakovski, A., Bergmann, R.: Representation and structure-based similarity assessment for agile workflows. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 224–238. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74141-1_16

18. Niesen, T., Houy, C.: Zur Nutzung von Techniken der Natürlichen Sprachverarbeitung für die Bestimmung von Prozessmodellähnlichkeiten – Review und Konzeptentwicklung. In: 12 Internationale Tagung Wirtschaftsinformatik, pp. 1829–1843 (2015)

19. Thaler, T., Dadashnia, S., Sonntag, A., Fettke, P., Loos, P.: The IWi process model corpus. Technical report, Institute for Information Systems (IWi) at the German Research Center for Artificial Intelligence (DFKI) (2015)

20. Thaler, T., Hake, P., Fettke, P., Loos, P.: Evaluating the evaluation of process matching techniques. In: Multikonferenz Wirtschaftsinformatik, pp. 1600–1612 (2014)

21. Vogelaar, J.J.C.L., Verbeek, H.M.W., Luka, B., Aalst, W.M.P.: Comparing business processes to determine the feasibility of configurable models: a case study. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM 2011. LNBIP, vol. 100, pp. 50–61. Springer, Heidelberg (2012). doi:10.1007/978-3-642-28115-0_6

22. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. **26**(2), xiii–xxiii (2002)

23. Yan, Z., Dijkman, R., Grefen, P.: Fast business process similarity search with feature-based similarity estimation. In: Meersman, R., Dillon, T., Herrero, P. (eds.) OTM 2010. LNCS, vol. 6426, pp. 60–77. Springer, Heidelberg (2010). doi:10.1007/978-3-642-16934-2_8

24. Zezula, P., Amato, G., Dohnal, V., Batko, M.: Similarity Search - The Metric Space Approach. Advances in Database Systems, vol. 32. Springer, New York (2006)