On Behavioral Process Model Similarity Matching: A Centroid-based Approach (Enlarged Abstract of [BBJ15])

Michaela Baumann[1], Michael Heinrich Baumann[1,2], Stefan Jablonski[1]

Summary A great number of related work points out the need for similarity detection between process models, be it the management of process model repositories, compliance checking, the reuse of model parts, etc. [BL12]. Among the different aspects of process models used for measuring similarity, behavior is an important part. [BBJ15] introduces a behavioral similarity measure suitable for many-to-many correspondences (cf. [Ba14]) that is still easy to compute.

Methods The paper uses an abstract definition of process models which is transferable to many process modeling languages. In a first step, a map between the compared models is established, mapping sets of activities to sets of activities. An example mapping is given in Fig. 1 indicated with different patterns of the activities.

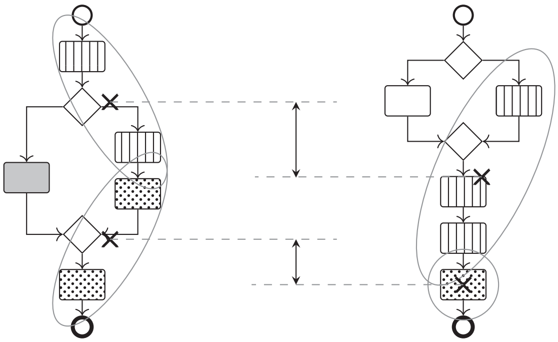

Fig. 1: Schematical representation of the comparison of (positional) centroids for mapped sets of process tasks.

In a second step, three behavioral features are determined for each activity: relative position, repeatability, and optionality. The relative position of activity a is calculated as the length of the shortest path from start to a divided by the length of the shortest path from start to end crossing a. For the grey activity in Fig. 1 this would be $3/6 = 0.5$. Repeatability and optionality are boolean. In the example, all activites are not repeatable but some are optional. For each mapped set and for each feature these values are averaged. In Fig. 1, the positions of the mapped sets, that are the averaged positions of the respective activ-

[1] University of Bayreuth, Faculty of Mathematics, Physics & Computer Science, {firstname.surname}@uni-bayreuth.de

[2] The work of M. H. Baumann is supported by a scholarship from Hanns-Seidel-Stiftung which is funded by Bundesministerium für Bildung und Foschung.

ities are indicated with an x, called centroid, each. The differences of a centroid and its image then contribute to the similarity: the lower the differences in average, the more similar the process models. An extension is given to penalize inhomogeneous activity sets via variances.

To validate the approach, a comparison with causal footprints (CF) [Di11], a widely accepted technique for measuring behavioral similarity, and "smallest" CF [BL12] is applied. All approaches show the same trend but the centroid-based (CB) approach shows greater variations: "similar" models are rated more similar and "dissimilar" models are rated more dissimilar than by CF and smallest CF. The major distinction is the computation effort of the approaches: the CB approach scales linearly with the number of activities whereas the smallest CF method scales quadratically and the CF method exponentially.

Discussion In order to evaluate the unpenalized CB approach against common sense of (behavioral) similarity, we carried out another study.[3] For five reference models we provided three alternative models, each, and asked modeling experts for the most similar and the most dissimilar one. The results are presented in Tab. 1.

		1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3	4-1	4-2	4-3	5-1	5-2	5-3
	CB	94	61	89	83	100	75	85	89	48	65	91	97	99	93	98
survey	sim	55	9	36	0	100	0	9	91	0	36	0	64	27	0	73
	dissim	0	82	18	9	0	91	18	9	73	64	36	0	27	46	27

Tab. 1: Unpenalized CB data and survey data rounded and in percent. CB values are all between 0 and 100 while the survey values sum up to 100.

For the first four reference models, the expert judgements support the CB approach. The fifth model, however, does not fit exactly. Nonetheless, the CB approach shows satisfying results in case studies. It does neither consider causal dependencies nor execution traces but position, repeatability, and optionality. The CB approach is a tradeoff between simplification and calculation effort.

References

[Ba14] Baumann, Michael Heinrich; Baumann, Michaela; Schönig, Stefan; Jablonski, Stefan: Towards Multi-perspective Process Model Similarity Matching. In: EOMAS, volume 191 of LNBIP, pp. 21–37. Springer Berlin Heidelberg, 2014.

[BBJ15] Baumann, Michaela; Baumann, Michael Heinrich; Jablonski, Stefan: On Behavioral Process Model Similarity Matching: A Centroid-based Approach. In: ICCGI 2015, volume 5, pp. 125–131. thinkmind, 2015.

[BL12] Becker, Michael; Laue, Ralf: A comparative survey of business process similarity measures. Computers in Industry, 63(2):148–67, 2012.

[Di11] Dijkman, Remco; Dumas, Marlon; van Dongen, Boudewijn; Käärik, Reina; Mendling, Jan: Similarity of business process models: Metrics and evaluation. Information Systems, 36(2):498 – 516, 2011.

[3] The corresponding material is available under `https://docs.google.com/forms/d/1VGwvJ7RrCO8gQs9rvHEqsiBdBfhG5gt86-fwmqnUE1I/viewform?c=0&w=1&usp=mail_form_link` [2016-04-11]