Detection of Similarity Between Business Process Models with the Integration of Semantics in Similarity Measures

Wiem Kbaier and Sonia Ayachi Ghannouchi

November 20, 2022

Detection of Similarity between Business Process Models With the integration of Semantics in Similarity Measures

Wiem Kbaier[1] and Sonia Ayachi Ghannouchi[2]

[1] Higher Institute of Computer Science and Communication Techniques Hammam Sousse (ISITCOM)
University of Sousse, Sousse, Tunisia
wiem.kbaier@isitc.u-sousse.tn
[2] RIADI Laboratory
University of Manouba, Manouba, Tunis
s.ayachi@coselearn.org

Abstract. Business process models play an important role in today's organizations and they are stored in models repositories. Organizations need to handle hundreds or even thousands of process models within their model repositories, which serve as a knowledge base for business process management. Similarity measures can detect similarities between Business process models and consequently they play an important role in the management of business processes. Existing researches are mostly based on the syntactic similarities based on labels of activities and deal with mapping of type 1:1. To address the problem, semantic similarities remain difficult to detect and this problem is accentuated when dealing with mapping of type n:m and considering large models. In this paper, we will present a solution for detecting similarities between business process models by taking into account the semantics. We will use a genetic algorithm, which is a well-known metaheuristic, to find a good enough mapping between two process models.

Keyword: Business Process models, Similarity measures, Semantics, Genetic Algorithm, Matching.

1 Introduction

In organizations, thousands of business processes (BP) are modeled and stored due to the diversity of needs and operations associated with these processes. In fact, process documentation generates a large number of process models in a repository. So, a repository helps to improve business model development and resolves flaws in process modeling [2]. But, with the rapidly changing environment, organizations must be able to quickly and flexibly adjust their business processes to meet new demands. However, it is extremely complicated to create business processes from the begining. Hence the need to manage repositories of business process models so that organizations continually improve their operations [2]. For this, the detection of similarity within the repository between business process models is mandatory but the problem is how to be able to recover models that are similar or containing similar fragments.

Besides, similarity measures are frequently used in text similarity analysis and Clustering. Any measure of similarity or distance usually measures the degree of proximity between two entities, which can be in any format of text, such as documents, sentences or even terms. These similarity measures can be useful in identifying similar entities and distinguishing clearly different entities from one another. Similarity measures are very effective, and sometimes choosing the right measure can make a big difference in decision making.

In this context, similarity measures are used to detect the similarity between business process models. Therefore, calculating the similarity between BPs is a task performed in a wide variety of business process management applications. So, similarity measures [5] can be useful in many cases such as merging BPs and managing repositories to check if similar models are stored in the repository. Also, similarity measures can facilitate the reuse of BP models because they reduce time and cost. Therefore, it is important to find existing BP models and reuse them.

Within a business, customer requirements can change. So it is necessary to have a similarity measure that simplifies changes by determining the processes that meet these needs. This simplifies the management and facilitates the reuse of these processes. Besides, it is necessary to measure the degree of conformity between a reference model and a given model using similarity measures. Moreover, during execution, services are called. But, these services can fail for example due to a computer failure. So you have to find identical or similar services to automate execution [5].

A basic technique required for many approaches of process model similarity is matching. More precisely, PMM (Process Model Matching) is composed of techniques that allow the automatic identification of corresponding activities between two business process models. Several correspondence techniques have been developed.

The remainder of the paper is organized as follows. Section 2 presents the measures and the matching and their problems are explained in section 3. Section 4 illustrates the related work found. Section 5 presents our approach. Section 6 concludes the paper.

2 Similarity measures

Mainly, there are four measures which are most used by different authors. These are syntactic, semantic, structural and behavioral measures. In addition, according to a similarity search carried out by Dumas et al. [9] the similarity measures are classified according to three criteria which are: the labels (label), the graphical structure of the model, and the execution semantics. In fact, the grouping of similarity measures varies slightly according to the different proposals of the authors as we will show in the related work section.

2.1 Syntactic measures

Syntactic measures relate to simple comparisons of strings and do not take into account the meaning or context of words. The most used syntactic measures are the Levenshtein distance which counts the number of edit operations (add / remove / substitute) and the Jaro-Winkler distance, which works similarly, but produces a value included between 0 and 1. The Jaccard and Dice coefficient measures both calculate the similarity between two activity labels as a function of the number of shared and unshared words [13]. Also, the authors consider cosine similarity, Jensen-Shannon distance measure and the substring measure that allows taking into account substring relationships between the activities.

The limitation of these measures is not only their inability to recognize synonymous terms, but also their tendency to view unrelated words as similar. Take, for example, the unrelated words “contract” and “contact”. The Levenshtein distance between these words is only 1, indicating a strong similarity between the terms [13].

2.2 Semantic measures

Semantic measures aim to take into account the meaning of words. A very common strategy for doing this is identifying synonyms using the lexical database WordNet [17].

The most prominent semantic measure is the Lin similarity. Lin Similarity is a method for calculating the semantic relatedness of words based on their information content according to WordNet taxonomy [13]. To use the Lin measure for measuring the similarity between two activities (which mostly contain more than a single word), we have to combine Lin's similarity with the bag-of-words model. The bag-of-words model turns an activity into a set of words, ignoring grammar and word order. The Lin similarity can then be obtained by identifying the word pairs from the two bags with the highest Lin score and calculating their average.

Other measures based on the WordNet dictionary include Wu & Palmer and Lesk [13]. The first calculates the similarity between two words by considering the path length between these words in WordNet taxonomy. The latter compares the WordNet dictionary definitions of the two words. Some approaches also directly verify hypernym relationships (a hypernym is a more common word). For example, some authors consider “car” and “vehicle” to be identical words since “vehicle” is a hypernym of “car” [13].

The semantic measures used are very basic and based on the WordNet dictionary. This is a significant problem because any measure based on WordNet returns a similarity score of zero if a term is not in the WordNet dictionary. Although the WordNet dictionary is quite comprehensive, it does not cover the complex compound words that we often find in process models. For example, “problem report” or “budget plan” activities are measured by determining synonyms for each word and not considering compound words [13].

2.3 Structural measures

The structural similarity mainly reflects the similarity of process model topology which expresses the logical relationship between the activities. It depends on the relationship between the relevant data and the control flow. Therefore, the structure is one of the important static attributes of the process model. The relevant aspects of this category come from graph theory and the general graph-structured based similarity between models can be quantified by the graph edit distance [22]. Dijkman et al. [7] define the graph edit distance between different process models and design the basic edit operations and the similarity formula. The graph edit distance between two graphs is the minimum number of graph edit operations required to switch from one graph to another. Different graph edit operations can be taken into account such as: deleting or inserting nodes, substituting nodes, and deleting or inserting edges.

2.4 Behavioral measures

Behavioral similarity emphasizes the execution semantics of business process models. This is usually expressed by a set of allowed execution traces of the process model. These traces can be generated by simulation or during the actual execution of a process and are usually stored in a process log. Currently, most of the behavioral similarities of processes are obtained by measuring the similarity of simulated traces of process models [22].

3 Problem illustration

In this section, the limitation of similarity measures and the choice of dimensions are discussed. Then, we illustrate the cardinality problem and a presentation of genetic algorithm.

3.1 Similarity measures

Becker and Laue [5] provide a detailed survey on the exact calculations used by process model similarity measures. So, the results show that different similarity measures rank the similarity between BP very differently. For example, when we take two business process models and then we calculate the similarity with two different measures, the first measure might indicate that they are similar, while the second measure indicates that they are not similar. Therefore, Becker and Laue [5] conclude that there is not a single "perfect'' similarity measure. At the same time, it is unclear which measures can be meaningfully applied in a specific context.

In addition, Szmeja et al. [18] proposed a classification of similarity measures related to semantics into different dimensions. Each dimension represents a different type of similarity.

For these reasons, we take as starting point the similarity measures classified by dimension of semantics because they allow to add extra information to the similarity score, and to highlight differences and similarities between results.

3.2 Dimensions of semantic similarity

Understanding the meaning of a dimension is absolutely necessary in order to decide which dimensions to utilize. For this, the semantics of each dimension is explained in detail.

- Lexical : Entities are lexically similar, when the words used to label them are similar according to a dictionary.
- Co-occurrence : Objects are co-occurrently similar, when they often appear together.
- Taxonomic : Objects are taxonomicaly similar, when they are of similar classes, kinds or types
- Descriptive : Objects are descriptively similar, when they have similar properties, attributes or characteristics
- Physical : Objects are physically similar, when their physical characteristics and appearance are similar
- Compositional : Objects are compositionaly similar, when they have a similar set of parts or ingredients
- Membership : Objects are membership similar, when they have similar sets of representatives, instances or members

We precisely focus on lexical dimensions because our objective is to detect similarity by introducing semantics.

3.3 Cardinality Problem

Few works have addressed the problem of cardinality when detecting similarities between business process models. It is possible to determine the activity of the first model which corresponds to another activity of the second model. This corresponds to the case of simple cardinalities [1: 1]. On the other hand, we may find problems in detecting the corresponding activities because one activity can be similar to a set of activities of the other model. We are talking in that case about complex cardinalities which are either [1: N] or [M: N]. Recent works in the area of business process model mapping show an interest in complex mappings.

A considerable number of complex matches cannot be successfully identified because two models can differ in terms of the terms they use (synonyms) as well as their level of detail. Given these differences, the correct recognition of correspondences between two process models can become a complex and difficult task. To solve this problem, we can detect the mapping with the Genetic Algorithm (GA). The GA is a population-based approach which means that it starts by an initial solution with many points. In our case, for detecting the optimal mapping, it starts by one activity of the first model and browses the other activities of the second model until it finds the best correspondence. We consider a set of solutions for a problem and select the set of best ones out of them.

3.4 Genetic Algorithm

The GA is a search algorithm which was based on survival of the fittest of natural evolution. This metaheuristic is based on the method of natural choice of the best one for producing the children of the next generation maintaining the law of Darwin's theory. GAs are based on the random search technique but in a structured manner. These heuristics depend on the method of natural selection and natural genetics. So, in each generation, a new set of offspring is generated, utilizing portions of the fittest individuals of the previous generations. The GA can be applied with a coding of the parameter set and they start its execution from a group of points not from a single point [14].

Since we are interested in a solution for our problem based on metaheuristic algorithms, it is important to note that these algorithms are used to solve real-life complex problems arising in different fields. These metaheuristic algorithms are broadly classified into two categories namely single solution and population based metaheuristic algorithm. We have chosen to continue with population based metaheuristic algorithm because the first category such as simulated annealing, tabu search (TS), microcanonical annealing (MA), and guided local search (GLS) utilize single candidate solution and improve this solution by using local search. However, the solution obtained from single-solution based metaheuristics may stuck in local optima.

Population-based metaheuristics utilize multiple candidate solutions during the search process. These metaheuristics maintain the diversity in population and avoid the solutions which stuck in local optima. The well-known population-based metaheuristic algorithm is the genetic algorithm [14].

For instantiating a genetic algorithm to solve the problem at hand, five phases are considered in a genetic algorithm. The first one concerns the encoding of a solution (i.e., the representation of an initial population), the second is the fitness function to be used to evaluate the quality of a candidate solution and the third one concerns the crossover and mutation operators.

Genetic Algorithm

Randomly initialize population (t)
Evaluate Fitness of population (t)
Repeat
1. Select parents from population (t)
2. Perform crossover on parents creation population (t+1)
3. Perform Mutation on population (t+1)
4. Evaluate fitness of population (t+1)
Until best individual is good enough

4 Related work

According to the literature, several papers [1,7-10,11,16,19,20] provided an overview of similarity measures such as the definition and calculation of certain measures.
Dijkman et al. [7] were interested in determining the similarity between business process models that are documented in an organization's repository. Indeed, Dijkman et al. [7] presented three similarity measures to solve this problem which are evaluated experimentally in terms of "precision" and "recall". Thus, according to this evaluation, the authors deduced that the results of the measurements are comparable but the structural measure is more efficient than the others. The measures found in [7] are:

- Node matching similarity which compares the labels linked to the elements of models.
- Structural similarity which compares element labels according to the model topology.
- Behavioral similarity which compares the labels of the elements as well as the causal relationships captured in the process model.

The detection of similarities between business process models is a very interesting subject and it is adopted by several authors because it allows solving some problems such as the measurement of the conformity between the reference models and the current models and finding associated models in a repository.

So, Becker and Laue [5] gave a detailed overview of the exact calculations used by similarity measures of process models. According to this paper, the comparison of two business process models should be established in two steps. In the first step, the activities of one model that match the activities of the other model are identified. This step is known by matching i.e. if the models have been created in different organizations or if they describe a business process at different levels of details, the corresponding activities must be identified either by using one of the existing algorithms or on the basis of expert judgment. In the second step, the similarity measures between models are applied.

Moreover, Zhou et al. [22] have shown that measuring process similarity plays an important role in business process management. Additionally, to establish efficient use of processes, the paper [22] specializes in detecting similar patterns by applying process models which are predefined descriptions of business processes and process logs that can be viewed as an objective observation of the actual execution behavior of the process. In other words, they specialize in structural measures and behavioral measures.

Another paper [4] proposed an automated approach to query a repository of business process models through queries for structurally and semantically relevant models. Specifically, a user establishes a request “BPMN_Q” and therefore he must receive a list of models ordered by relevance according to the request. The objective of this paper is to find models through queries then apply similarity measures only to models with the same scope.

This idea is further confirmed by Jabeen et al. [13]. Therefore, the search for process models favors the detection of correspondences between model activities.

In addition, there are authors [4,11,12,15] who work using Semantic Business Process Models (SBPM). Thus, a business process model is considered as a semantic model if it contains descriptions based on the ontology of the process models, this helps to resolve ambiguity issues caused by the use of synonyms and homonyms. On the other hand, the authors of [10] aim to solve the ambiguity problem of PM by using an ontology-based description of process models. Indeed, the detection of labels of similar elements automatically helps ensure the interconnectivity and interoperability of business processes. For this reason, an OWL_DL description is applied for models with the Petri net notation.

Several works have dealt with the matching between business process models, hence the development of different techniques for the matching of process models (PMM: Process Model Matching).

Furthermore, the evaluations described by Cayoglu et al. [6] and Antunes et al. [3] can be regarded as related works in which process model matching techniques are summarized and compared to each other.

However, these techniques are heuristic and, therefore, their results are uncertain and should be evaluated on a common basis. So due to the lack of datasets and frameworks for the evaluation of techniques, the idea of Process Model Matching Contest 2013 (PMMC'13) was created by Cayoglu et al. [6] which aimed to address the need for effective evaluation by defining mapping issues between process models in published data sets.

The domain experts developed two sets of benchmark data called PMMC'13 to allow the researcher validating their work. These databases are validated by expert opinion and also by three performance measures (precision, recall, F score). Likewise, Antunes et al. [3] launched a second competition in 2015 which is the same concept as the first competition in 2013 but with an improved database and more techniques to evaluate.

In addition, [13] aims to improve the performance of PM matching techniques by first establishing an analysis of a literature review on all existing process model matching techniques and secondly takes place by performing an analysis of PMMC 2015 to determine the most suitable technique. In fact, PMM techniques consist of automatically identifying the corresponding activities between two models that represent similar or identical behavior. In fact, the paper shows that the techniques presented in PMMC'13 and PMMC'15 are mainly based on syntactic measures and semantic measures based on Wordnet.

According to [21] business process models can be compared, for example, to determine their consistency. [21] reported that any comparison between process models relies on a match that identifies which activity in one model matches which activity in another. This paper introduces the matching tools which are called matchers. Weidlich et al. [21] present the ICoP framework (Identification of Correspondences between Process Models), which can be used to develop such matchers. The framework allows the creation of matchers from reusable components or newly developed components. It focuses on detecting complex matches between activity groups.

5 OUR APPROACH

Our approach makes it possible to detect the similarity between business process models and more precisely the detection of similar fragments in large models, while better taking into consideration the semantics at the level of the different similarity measures and not only for the semantic similarity measure.

Indeed, according to the work found, certain measures apply the semantics by accessing the WordNet dictionary. But the problem is that these metrics only apply at the activity label level and do not take into account other elements of the models such as gateways and events.

For this reason, we propose to begin with the first step of process model matching which is the first line matcher. In this step, we apply the genetic algorithm that helps to identify the best mapping because we apply the Jiang similarity measure and resolve the problem of complex cardinality. After obtaining the best mapping, we apply the structural measures for this mapping. So, we deal with both semantic and structural measures.

5.1 Steps of Genetic Algorithm

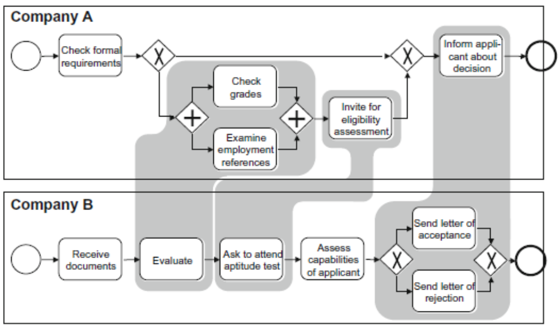

Fig. 1. Two business processes and their correspondences[13]

To create this representation of a Candidate solution presented in figure 1, we propose to use a string representation. Each chromosome (solution) contains the labels of the activities of each model. Thus each activity in the two BPs has the chance to be matched. The size of the chromosome is equal to max (|BP1|, |BP2|). For example a candidate solution (a possible mapping) between the two processes in figure 1 is represented by an array of strings. The size of the array presented in table 1 is 6.

TABLE I. AN EXAMPLE OF CANDIDATE SOLUTION

Check formal requirements	Examine employement references	Check grades	Inform applicant about decisions	Invite for eligibility assessment	

In this example, the model of company B is the source BP.

The first dimension indicates that an activity in the BP of company B « Receive documents » is mapped to an activity in the second BP of company A « Check formal requirements ».

The second activity in the BP of company B « Evaluate » is mapped to an activity in the second one «Examine employement references », the activity « Ask to attend aptitude test » is mapped to « Check grade » in the BP of company A. The activity « Assess capabilities of applicant » is mapped to the activity « Inform applicant about decisions ». The activity « Send letter of acceptance » is mapped to the activity « Invite for eligibility assessment ». The activity « Send letter of rejection » is not mapped to any activity. Thus, the candidate solution represented by this chromosome is the matching M= {(Receive documents,

Check formal requirements),(Evaluate, Examine employement references), (Ask to attend aptitude test, Check grade),(Assess capabilities of applicant, Inform applicant about decisions),(Send letter of acceptance, Invite for eligibility assessment)),(Send letter of rejection),∅)}. Then, we need to apply crossover and mutation to obtain the optimal mapping.
The process established for recovering models which are similar or which contain a similar fragment is in three stages which are represented in figure 2.

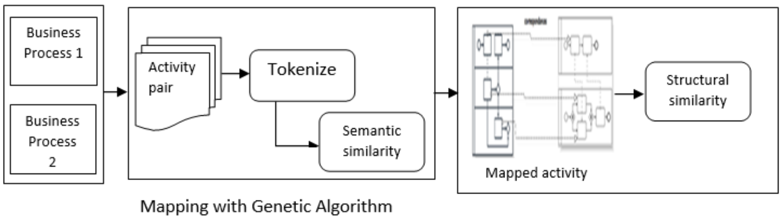

Fig. 2. Overview of the Proposed Approach.

The first step is to identify the activities of one model that match the activities of the other model. The Genetic Algorithm shows this step of matching i.e. describing which activity in the first model corresponds to an activity in the second model using an algorithm that uses activity pairs to give you all possible mappings then with the fitness function, we can choose the best mapping with complex cardinality.

The second step is the application of the similarity measures. In fact, these measures are stored in a database and classified by dimensions.

So, for calculating the semantic similarity we need activity from BP1 and activity from BP2. First, each label of these activities are splited into a set of tokens. Then, each token is lemmatized to be able to find the score of similarity between it and its most similar token in the second label using the Jiang similarity measure.

We calculate the semantic similarity by using the Jiang measure. This method has used a corpus in addition to a hierarchic ontology (taxonomic links). The distance between two concepts C1 and C2, formulated in this work is the difference between the sum of the information content of the two concepts and the information content of their most informative subsumer:

$$SimJnc(c1, c2) = \frac{1}{IC(C1) + IC(C2) - 2 * IC(LCS)}$$

This measure is insightful to the shortest path length between C1 and C2 and the density of concepts along this same path. Where, IC stands for information contest and LCS is the lowest common subsumer of C1 and C2 defined as the common parent of them with minimum node distance.

The last step is the recommendation that is to say after knowing if the two models are similar, it is necessary to act and choose an action to be done as illustrated in figure 3. In this context, when the similarity calculation is applied in the case of companies merging, the recommendation is to merge two models and therefore adopt a newly established model. If we are managing the company's repository, the possible recommendation is to reuse the reference model by making the necessary changes. If the two models are not similar then the recommendation is to add a new model to the repository.

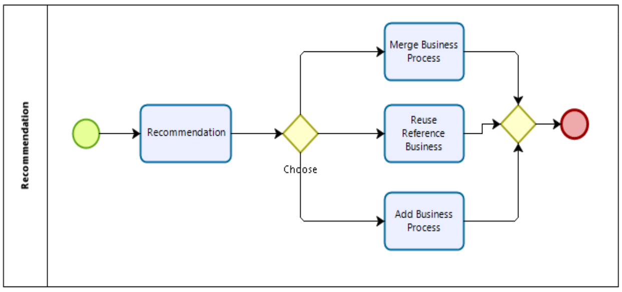

Fig. 3. Proposed recommendation

To justify our proposal, we will apply our results obtained by the genetic algorithm with the databases presented in Process Model Matching Contest 2015. For the evaluation of the proposed approach, we have selected three benchmark datasets, which are specific to three different domains. The three datasets were developed for the PMMC'15. We have chosen to use these three datasets because PMMC'15 content has been used in the evaluation of many process model matching techniques and they are available online. Also, the involvement of leading experts of the domain and the frequent use of the datasets for the evaluation of matching techniques are reasons for our choice.

This evaluation is done by applying our approach to these databases. we apply our approach on PMMC'15 models. Then, we calculate the three performance measures (precision, recall, F score) of these datasets. Once we obtain the performance values of our prototype then we compare them with the performance measures of PMMC'15 models established by the other techniques.

6 Conclusion

This paper studies three classes of similarity measures (syntactic, semantic and structural) designed to answer process model similarity queries. So, our objective is to detect the similarity between business process models and more precisely the detection of similar fragments while considering the semantics at the genetic algorithm. Our approach was to apply semantic similarity in genetic algorithm to obtain the best mapping with complex cardinality using Genetic Algorithm. Then, we apply the structural similarity for the mapping obtained to discover if the two business processes are similar or not and act with recommendations. For future research, first we will evaluate our approach by experimental validation. Then, we will develop a prototype to be able to apply the genetic algorithm and we will evaluate the results obtained by the databases published in PMMC 2015. We will also study the benefits of our approach in various case studies. Furthermore, we will investigate the detection of similarity in the case of large models.

References

1. Aiolli, F., Burattin, A., & Sperduti, A. (2011). A Metric for Clustering Business Processes Based on Alpha Algorithm Relations. Department of Pure and Applied Mathematics, University of Padua, Italy, 1-17.
2. Ali, M., & Shahzad, K. (2018, June). Enhanced benchmark datasets for a comprehensive evaluation of process model matching techniques. In Workshop on Enterprise and Organizational Modeling and Simulation (pp. 107-122). Springer, Cham.
3. Antunes, G., Bakhshandeh, M., Borbinha, J., Cardoso, J., Dadashnia, S., Di Francescomarino, C., ... & Weidlich, M. (2015). The process model matching contest 2015 (Vol. 248, pp. 127-155). Geellschaft für Informatik.
4. Awad, A., Polyvyanyy, A., & Weske, M. (2008, September). Semantic querying of business process models. In 2008 12th International IEEE Enterprise Distributed Object Computing Conference (pp. 85-94). IEEE.
5. Becker, M., & Laue, R. (2012). A comparative survey of business process similarity measures. Computers in Industry, 63(2), 148-167.
6. Cayoglu, U., Dijkman, R., Dumas, M., Fettke, P., García-Bañuelos, L., Hake, P., ... & Weidlich, M. (2013, August). Report: The process model matching contest 2013. In International Conference on Business Process Management (pp. 442-463). Springer, Cham.
7. Dijkman, R., Dumas, M., Van Dongen, B., Käärik, R., & Mendling, J. (2011). Similarity of business process models: Metrics and evaluation. Information Systems, 36(2), 498-516.
8. Dijkman, R. M., van Dongen, B. F., Dumas, M., García-Bañuelos, L., Kunze, M., Leopold, H., ... & Yan, Z. (2013). A short survey on process model similarity. In Seminal contributions to information systems engineering (pp. 421-427). Springer, Berlin, Heidelberg.
9. Dumas, M., García-Bañuelos, L., & Dijkman, R. M. (2009). Similarity search of business process models. IEEE Data Eng. Bull., 32(3), 23-28.
10. Ehrig, M., Koschmider, A., & Oberweis, A. (2007, January). Measuring similarity between semantic business process models. In Proceedings of the fourth Asia-Pacific conference on Comceptual modelling-Volume 67 (pp. 71-80).
11. Gerth, C., Luckey, M., Küster, J. M., & Engels, G. (2010, July). Detection of semantically equivalent fragments for business process model change management. In 2010 IEEE International Conference on Services Computing (pp. 57-64). IEEE.
12. Humm, B. G., & Fengel, J. (2012, May). Semantics-based business process model similarity. In International Conference on Business Information Systems (pp. 36-47). Springer, Berlin, Heidelberg.
13. Jabeen, F., Leopold, H., & Reijers, H. A. (2017, June). How to make process model matching work better? An analysis of current similarity measures. In International Conference on Business Information Systems (pp. 181-193). Springer, Cham.
14. Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. Multimedia Tools and Applications, 80(5), 8091-8126.
15. Koschmider, A., & Oberweis, A. (2007, March). How to detect semantic business process model variants?. In Proceedings of the 2007 ACM symposium on Applied computing (pp. 1263-1264).
16. Schoknecht, A., Thaler, T., Fettke, P., Oberweis, A., & Laue, R. (2017). Similarity of business process models—a state-of-the-art analysis. ACM Computing Surveys (CSUR), 50(4), 1-33.
17. Shahzad, K., Pervaz, I., & Nawab, A. (2018). WordNet based Semantic Similarity Measures for Process Model Matching. In BIR Workshops (pp. 33-44).
18. Szmeja, P., Ganzha, M., Paprzycki, M., & Pawłowski, W. (2018). Dimensions of semantic similarity. In Advances in Data Analysis with Computational Intelligence Methods (pp. 87-125). Springer, Cham.
19. Thaler, T., Schoknecht, A., Fettke, P., Oberweis, A., & Laue, R. (2016, September). A comparative analysis of business process model similarity measures. In International Conference on Business Process Management (pp. 310-322). Springer, Cham.
20. Van Dongen, B., Dijkman, R., & Mendling, J. (2013). Measuring similarity between business process models. In Seminal contributions to information systems engineering (pp. 405-419). Springer, Berlin, Heidelberg.
21. Weidlich, M., Dijkman, R., & Mendling, J. (2010, June). The ICoP framework: Identification of correspondences between process models. In International Conference on Advanced Information Systems Engineering (pp. 483-498). Springer, Berlin, Heidelberg.

22. Zhou, C., Liu, C., Zeng, Q., Lin, Z., & Duan, H. (2019). A comprehensive process similarity measure based on models and logs. IEEE Access, 7, 69257-6927