On the Similarity of Process Change Operations

Georg Kaes[(✉)] and Stefanie Rinderle-Ma

Faculty of Computer Science, University of Vienna, Vienna, Austria
{georg.kaes,stefanie.rinderle-ma}@univie.ac.at

Abstract. Process flexibility is a vital part for almost any business area. Change logs are a central asset for documenting adaptations in processes, since they capture key information about associated change operations. Comparing multiple change operations offers interesting data for many analysis questions, e.g., for analyzing previously applied change operations and for supporting users in future adaptions. In this paper, we discuss different change perspectives and present metrics for comparing change operations. Their applicability and feasibility are evaluated based on a prototypical implementation and based on real world process logs.

Keywords: Business process analysis · Process change similarity · Similarity metrics

1 Introduction

Being able to adapt processes when the situation requires to do so is vital in almost any business domain. Change operations are the key concept when adapting processes. For analyzing dependencies between change operations, change processes [1] and change trees [2] can be used. On top of that, being able to compare change operations would be beneficial in the following situations [3]:

Planning Future Adaptations: When planning adaptations in certain situations, the person responsible for planning the change could analyze the adaptations which have been made before. Imagine a nursing home, where the therapy plan of each patient is represented as a process instance. Whenever a patient shows a new symptom, his and only his therapy process has to be changed in order to deal with the new problem. If two patients have problems with their digestion, some drugs could be applied for a certain amount of time, their diet plans could be changed, or some other therapies could be applied. Which type of therapy is applied usually depends on various individual circumstances, such as their medical history, pre-existing conditions, allergies etc.

Analyzing Past Change Operations: When evaluating change operations, identifying similar process change operations can add relevant information to an analysis. In a hospital, it can be analyzed in which situations similar therapies have been applied to a patient's therapy plan. Side effects of change operations

can also be compared: Think of a patient who got ill, and received some kind of treatment. Additionally, all therapies which include activities which burden the patient's immune system have to be removed. By analyzing the similarity of such side effects, the evaluation of such situations can be improved.

Both situations benefit from assessing the similarity of the involved change operations based on metrics that measure how much a change operations is similar to another one in a given context. Though literature proposes several metrics for process similarity (e.g., [4]) and instance similarity [5], metrics for change similarity have mainly be neglected so far. This paper approaches this gap based on the following research questions:

Q1: How can process change similarity metrics be defined in general? Considering which perspectives? Based on which information?

Q2: In which scenarios can different change similarity metrics be used? What are their advantages and disadvantages?

Q1 and Q2 are tackled as follows: First, we analyze the attributes of process change operations and their effects on different process perspectives, i.e., model, data, time, and resource perspective ($\mapsto$ Q1). The limitations of comparing the similarity of two process change operations solely based on their attributes is discussed in the sequel ($\mapsto$ Q2). An alternative approach is to exploit the effects of applying change operations to processes. Hence, two change similarity metrics are provided that exploit the effects on the resource and time perspective of the underlying processes ($\mapsto$ Q1). They can be useful in change scenarios where, for example, the model perspective is not available to users deciding on the change due to privacy reasons. The feasibility and applicability of these metrics is evaluated against metrics that consider the effects of change operations on the model perspective of processes ($\mapsto$ Q2).

The rest of the paper is structured as follows: Sect. 2 introduces fundamentals, perspectives, and change effects (Q1). This is followed by a discussion about a metric which focuses solely on the attributes of a process change operation (Sect. 3). In Sect. 4, we present effect-based metrics for comparing process change operations (Q2), which are evaluated in Sect. 5. Section 6 presents related work and Sect. 7 concludes our paper.

2   Fundamentals

This section introduces basic concepts of process change operation similarity.

2.1   Basic Definitions

The process schema and the process fragment are basic components of process change operations [6]. A change operation is applied to a process schema. The fragment defines what is being inserted, removed, or in any other way affected by the change operation. A process schema/fragment is defined as follows:

Definition 1 (Process Schema, Process Fragment). *A process schema S is defined as $S := (N, E, D, DE, Res, Temp)$ where*

- *N denotes a set of nodes, i.e. tasks and gateways such as XOR and parallel splits and joins*
- *E denotes the set of control flow edges, $E \subseteq N \times N$*
- *D is a set of data elements.*
- *$DE \subseteq (N \times D) \cup (D \times N)$ is a set of data edges connecting nodes with data elements (write) and data elements with nodes (read).*
- *$Res : N \mapsto 2^{R \times \mathbb{N}}$ denotes a function that assigns each activity the required number of resources, i.e., $Res(n) = \{(r, x) | r \in R, x \in \mathbb{N}, r$ assigned to $n\}$ where R is the set of all resources.*
- *$Temp : N \mapsto \mathcal{N}$ denotes the point in time associated with a node $n \in N$.*

The resource assignment $Res(n)$ determines the number of resources that are required to perform task n, for example, 2 nurses and 1 doctor for a surgery. $Temp(n)$ assigns a point in time to task n. This expression of temporal requirements in processes is simple; more powerful definitions (e.g., [7]) exist. However, for the purpose of constructing first similarity metric proposal time points are assumed as being sufficient.

In the following definition, process change operations are defined according to literature [6].

Definition 2 (Process Change Operation). *A process change operation Δ is defined as a tupel $\Delta := (t, f, p, S)$ where*

- *t denotes the type of the change ($t \in \{INSERT, DELETE\}$).*
- *f denotes the process fragment $f := (N, E, D, DE, Res, Temp)$ which is used by the change operation.*
- *p denotes the position of the change operation.*
- *S denotes the process schema $S := (N, E, D, DE, Res, Temp)$ the change operation is applied to.*

Further attributes such as rationale or the goal of the process change (cf. [8,9]) will be considered in future work. Moreover, this work focuses on INSERT and DELETE as change types. Based on these types other types such as MOVE can be expressed.

2.2   Types of Process Change Metrics

Basically, comparing process change operations can be based on the *attributes* of a change operation Δ, such as its position, goal, type or schema (cf. Definition 2) and its *effects*. The suitability of comparing change operations based on their attributes is discussed by means of an example in Sect. 3.

For comparing change effects, at first, we have to define what change effects actually are. In literature, there are multiple definitions for effects in general, depending on the context. In *Hinge et al.* [10], effects relate to process tasks

and are notated in conjunctive normal form (CNF). In *Cossentino et al.* [11], effect logs contain the result of a programs state after steps have been executed. *Rinderle et al.* [8] uses the term *effects* related to change operations to describe the difference between a model S, and an adapted model S' where S' results from applying change Δ to S.

With respect to literature, the definition by *Rinderle et al.* [8] seems most suitable as it is related to change operations. However, the approach focuses mostly on the model and data perspective of the change, even though a schema definition such as in Definiton 1 comprises additional perspectives such as the resource and time perspectives of process schemas. In order to provide a more comprehensive picture on change effects, a short discussion on change effects in relation to process perspectives seems useful. Hereby we follow [6].

The *model perspective* captures the structural and behaviorial perspective of a process. In [12], the formal semantics of change patterns for the control flow of a process are defined. These formal semantics describe the effects of six groups of process adaptation patterns, such as insertion patterns, deletion patterns, or replace patterns.

The *data perspective* describes the data, or information flow of a process model. Related to a process change operation, the data perspective covers the data flow regarding the unchanged process schema, the process fragment, and the resulting process schema. [13] describes several data patterns for the groups *Data Visibility*, which defines in which parts of a process a data element can be accessed, *Data Interaction*, which defines the interaction with data elements in the process, *Data Transfer*, which focuses on the actual data flow, and *Data-based Routing*, which defines the influence of data elements on a processes execution. Analyzing the similarity of the effects of two process change operations on the data flow can yield interesting information, e.g. if some data element which is critical to the processes execution was affected by the process change or not.

The *resource perspective* covers the organizational aspects of a process such as actors, roles, and organizational units. Effects of process change operations on this perspective can be relevant when planning future changes, specifically which and how many resources are affected by the change operation (either because they have to do more work, less work, or something at a different point in time). In a nursing home, required resources for a long-running therapy can be planned and analyzed before adaptation. If resources are scarce, different adaptations can be compared in order to find the most efficient one. In this paper, we will describe a metric which focuses on comparing the effects of two process change operations on the resource flow.

The *time perspective* covers all temporal aspects of the process. It is also relevant for change operations, e.g., when the change operation has been conducted (cf. logging changes in [1]) or which due dates are defined for tasks in the change fragment. When comparing the effects of two change operations on a processes time perspective, one can analyze for how long the fragment in a process change will affect the affected schema.

Conclusion: In general both, metrics that are based on attributes or change effects can be used to measure the similarity of change operations. Table 1 summarizes what can specifically be compared using either the attributes or the effects of a change operation along the different perspectives.

Table 1. Comparing change operations based on their attributes or effects for different perspectives

↓ Perspective	Attribute based similarity of	Effect based similarity of
Model/Data:	control and data flow of the original schemas and change fragments	the change of control and data flow due to process change operations
Resource:	utilizing resources in the original schemas and change fragments	resource requirement change due to process change operations
Time:	temporal constraints of tasks in the original schemas and change fragments	temporal shifts in the resulting processes due to the process change operations

3 Comparing the Attributes of Process Change Operations

This section illustrates that a metric which is solely based on the attributes of a process change operation only yields satisfactory results, if it is tailored to the specific situation (Q2).

When comparing two process change operations $\Delta_1 := (t_1, f_1, p_1, S_1)$ and $\Delta_2 := (t_2, f_2, p_2, S_2)$, one could use the four basic attributes which define them for comparison, i.e. the operation type, the fragment, the position and the schema. For each of these values, different metrics already exist: For the process schema ($sim(S_1, S_2)$ and fragment ($sim(f_1, f_2)$), techniques for the similarity of process models can be used [4,14,15]. The positions of the change operations ($sim(p_1, p_2)$), which are defined by the pre- and postset of the element, can be compared by adapting the node matching similarity as defined in [4]. For comparing the change operation type ($sim(t_1, t_2)$), which is typically a string such as INSERT or DELETE, approaches which compare string similarity [16] or equivalence could be used.

Overall, for $\Delta_1 := (t_1, f_1, p_1, S_1)$ and $\Delta_2 := (t_2, f_2, p_2, S_2)$ the following attribute-based metric can be formulated:

$$sim_{attr}(\Delta_1, \Delta_2) := w_t * sim(t_1, t_2) + w_f * sim(f_1, f_2) + w_s * sim(S_1, S_2) + w_p * sim(p_1, p_2) \tag{1}$$

where the sum of the weights $w_t + w_f + w_s + w_p = 1$.

As the following example shows, the weights which should be used for each of these values highly depend on the situation, i.e., the schema, fragment, position, and change operation type.

In Fig. 1, three process schemas for a surgery in a hospital are shown: The first one displays the process for an adult, the second for an elderly person, who can still care for himself, and the third shows the process for a child, where the parents still have the right to decide. With a few minor differences, the basic process works as follows: First, general information about the patient is being collected, then, he is prepared for surgery, the surgery is being conducted, and finally, a report is delivered.
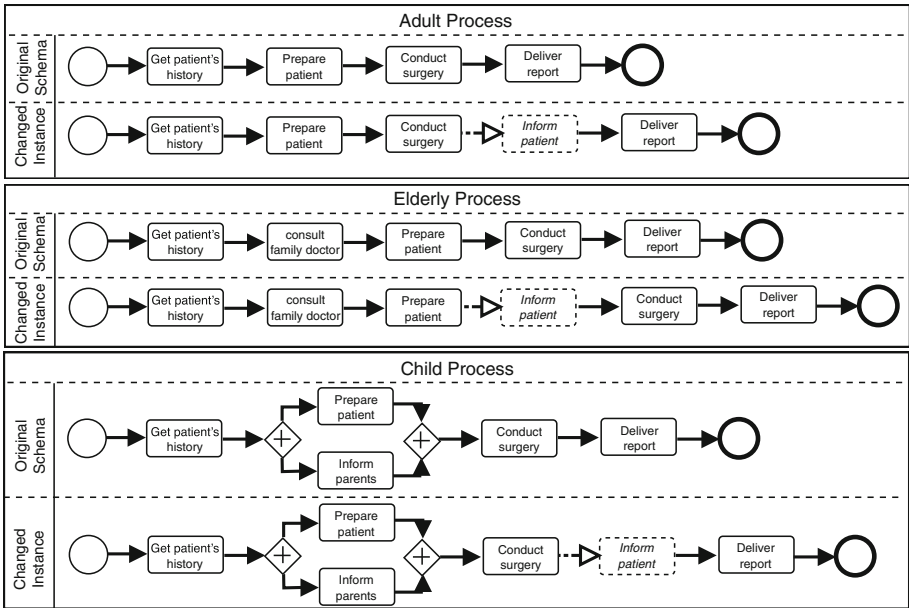

Fig. 1. Three examples of process change operations

In certain cases, adaptations to these basic processes are necessary: If a patient requires additional information *before* the surgery, an *Inform Patient* step is added to the process instance, where another doctor consults the patient. This is shown in the changed instance of the elderly process.

Sometimes, the surgery did not reach its desired goal, and another surgery will be necessary. Thus, the patient has to be informed separately *after* the surgery. For this, an *Inform Patient* step is added to the process instance *after* the surgery. This is shown in the changed instance of the adult process.

Following, the *position* of the *Inform Patient* step highly influences its semantical meaning: If it is added before the surgery, it usually just means that another doctor has to be consulted; if it is added after the surgery, it might hint that something went wrong. Thus, when comparing two change operations with the

element *Inform Patient*, one may regard the weight of the position $w_{position}$ as very high, while the other weights are quite low.

When doing surgeries on children, their parents have to be included in every step of the process. Each time the patient is informed, his or her parents also have to be informed. Besides this, the *Inform Patient* step has the same semantic meaning if added before or after the surgery as for the other two processes.

When comparing the change operations of the adult and the elderly patient, the *Inform Patient* step is added for the adult after the surgery, and for the elderly person before the surgery. Since the two process schemas do not have much effect on this particular change operation, but the position does have a very high effect, one may set the position's weight as very high, and the other weights quite low. However, when comparing the process of the elderly patient with the child's process, the child also receives the *Inform Patient* step after the surgery, thus indicating that another surgery might be required. However, since in a child's process the parents also have to be included into every step, the process schema has a higher influence. Thus, the corresponding weights would be different in these two cases.

This example supports the observation that the attribute-based metric for comparing process change operations (cf. Eq. 1) requires to carefully choose the weights depending on the particular situation changes take place. In other words, the significance of the results highly depends on the choice of the weights. Hence, in the next section, metrics are proposed that are based on change effects, i.e., metrics that abstract from the change attributes.

4   Effect Based Change Similarity Metrics

In the last section we have shown that a comparison metric which is solely based on a process change operation's attributes only yields satisfactory results if the relevance of each of the attributes is known. As an alternative, in this section we provide metrics which are based on the effects of a process change operation. A process change operation may have - depending on the design of the fragment which is inserted to or deleted from the schema - effects on all process perspectives presented in Sect. 2.2.

A common scenario in which process change operations may be compared is that of some domain expert who uses the resulting data for comparing possible adaptations, or for analyzing past adaptations, as discussed in the introduction. Depending on the question at hand, different perspectives on the process fragment may be of interest. In this section we chose to set our focus on the resource and temporal perspective for two reasons: First, these perspectives can provide answers to interesting questions for a domain expert. For a practitioner such as a nurse in a nursing home the resource and temporal perspectives provide answers to questions such as *How much staff will be required* or *How long will this therapy take?* Second, metrics which are solely based on the temporal and resource perspective can be used to compare process change operations without any knowledge of the schema after the change, or about the other perspectives

of the change operation's fragment. This can be relevant when the whole process cannot be accessed (a) due to privacy issues or (b) because the whole information is not relevant to the question. Think of a nursing home where a nurse has to plan the resources of the upcoming weeks: She may not have to know which steps the therapy processes contain exactly - all she needs is data about the required resources. By removing critical information about the patient's therapy processes (and reducing the information only to the resource and time perspective) this data can also be used with less data privacy issues.

4.1 Resource Perspective Metrics

In this section, we present a similarity metric for process change effects from the resource perspective, which captures all required organizational and other resources. When planning a process adaptation, it may be interesting to compare available adaptations based on the resources they require. Thus, one can choose the adaptation which requires the least resources if two or more adaptations are similar otherwise.

For this we present the aggregated resource view for process schemas and process fragments that are used for changes. This view builds the basis for creating a metric to compare the resource perspective of two change operations. For each task, a definition of resources which are connected to this task are required. The effects of a change operation $\Delta = (t, f, p, S)$ on the resources perspective of the underlying process schema S can be determined based on the resource assignments of S and the process fragment f. In order to determine the effects, the aggregated resource view of a process schema/fragment is defined as follows:

Definition 3 (Aggregated Resource View). *Let $S = (N, E, D, DE, Res, Temp)$ be a process schema. Then the aggregated resource assignment ρ_S for S is defined as*

$$\rho_S := \{(r,s)|\exists (r,x) \in \bigcup_{n \in N} Res(n) \wedge s = \sum_{n \in N, (r,x) \in Res(n)} x\}$$

Consider Change Fragment 1 ($CF1$) and Change Fragment 2 ($CF2$) to be inserted or deleted by change operations as depicted in Fig. 2 with $CF1 = (N1, E1, D1, DE1, Res, Temp)$ and $CF2 = (N2, E2, D2, DE2, Res, Temp)$. The required resources for each task are depicted in italic as attributes of the task (so for executing task A resource *nurse* is required two times, while for executing task B resource *doctor* is required once.) This leads to the following sets of tasks and aggregated resource views:

- $N1 = \{A, A, B, C\}$ and $N2 = \{A, B, C, D\}$[1]
- For $CF1$: $Res(A) = (\text{nurse}, 2)$, $Res(B) = (\text{doctor}, 1)$, $Res(A) = (\text{nurse}, 2)$, $Res(C) = (\text{nurse}, 1)$
- For $CF2$: $Res(A) = (\text{nurse}, 2)$, $Res(C) = (\text{nurse}, 1)$, $Res(B) = (\text{doctor}, 1)$, $Res(D) = (\text{assistant}, 1)$
- $\rho_{CF1} := \{(nurse, 5), (doctor, 1)\}$
- $\rho_{CF2} := \{(nurse, 3), (doctor, 1), (assistant, 1)\}$

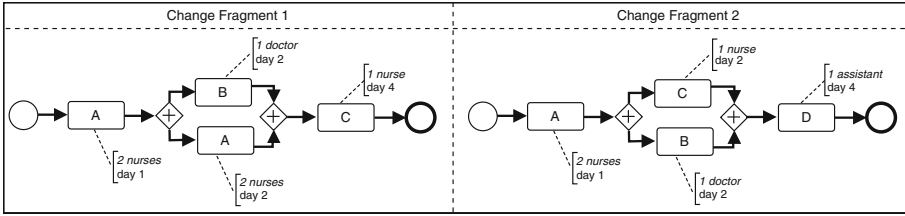[1] The sets are seen as bags due to multiple occurrence of activities.

Fig. 2. Example for aggregated resource perspective in process fragments

Definition 4 (Similarity Metrics for Resource Assignment). *Let ρ_1, ρ_2 be two aggregated resource assignments. Then the similarity between ρ_1 and ρ_2 is defined as follows:*

$$sim(\rho_1, \rho_2) := \begin{cases} \frac{\sum_{\rho_1,\rho_2}\{|y_1-y_2||\exists(r,y_1)\in\rho_1,\exists(r,y_2)\in\rho_2\}}{\sum_{\rho_1,\rho_2}\{max(z)|\exists(r,z)\in\rho_1\cup\rho_2\}} & if \ \rho_1 \cup \rho_2 \neq \emptyset \\ 0 & otherwise \end{cases}$$

$sim(\rho_1, \rho_2)$ relates the number of resources that are required for ρ_1 and ρ_2 to the maximum number of each required resource. For the example shown in Fig. 2 we have $\rho_1 := \{(nurse, 5), (doctor, 1)\}$ and $\rho_2 := \{(nurse, 3), (doctor, 1), (assistant, 1)\}$. Thus 4 resources are required for ρ_1 and ρ_2, i.e., 3 nurses and 1 doctor. The maximum number for each particular resource are 5 (nurse), 1 (doctor), and 1 assistant, summing up to 7. Overall, $sim(\rho_1, \rho_2) = \frac{4}{7} \approx 0.57$ for this example.

Metric $sim(\rho_1, \rho_2)$ can be calculated for the resource assignments of two fragments $f1$, $f2$ that are used by change operations $\Delta_1 = (t1, f1, p1, S1)$ and $\Delta_2 = (t2, f2, p2, S2)$. Doing so it becomes possible to measure the effects on the resource assignments of the underlying process schemas $S1$ and $S2$. It has only be further distinguished whether Δ_1 and Δ_2 insert or delete the fragments. For insertion, typically, resource assignments will be added, for deletion, resource assignments will be removed. Note that intentionally the resource assignments of the underlying schemata $S1$ and $S2$ are not considered as the metric is designed in an independent manner. The reason behind is to enable similarity calculation also for different schemas $S1$ and $S2$.

Definition 5 (Change Resource Similarity (CRS)). *Let $\Delta_1 = (t1, f1, p1, S1)$, $\Delta_2 = (t2, f2, p2, S2)$ be two change operations and ρ_1 (ρ_2) the aggregated resource assignment for $f1$ ($f2$). The Change Resource Similarity CRS between changes Δ_1 and Δ_2 is defined as follows:*

$$CRS(\Delta_1, \Delta_2) := \begin{cases} sim(\rho_1, \rho_2) \ if \ t1 = t2 \\ -sim(\rho_1, \rho_2) \ otherwise \end{cases}$$

4.2 Towards a Similarity Metric for the Timed Resource Perspective

The CRS metric compares the required resources of a process fragment. Sometimes, especially in long-running process settings, the time perspective is also relevant. Think about a nursing home planning the required resources for the next weeks: It may be interesting how two change operations affect the resource view only in a certain time frame. Definition 6 incorporates the time information into the aggregated resource view:

Definition 6 (Timed Aggregated Resource View). *Let $S = (N, E, D, DE, Res, Temp)$ be a process schema with aggregated resource view ρ_S. The timed aggregated resource view τ_S is defined as follows:*

$$\tau_S := \{(r, s, t1, t2)|(r, s) \in \rho_S,$$
$$t1 = min\{Temp(n)|n \in N \wedge \exists (r, x) \in Res(n)\},$$
$$t2 = max\{Temp(n)|n \in N \wedge \exists (r, x) \in Res(n)\}\}$$

Informally, the timed aggregated resource view determined for each tuple in the aggregated resource view the earliest and latest point in time the associated resource was required. For the change fragments $CF1$ and $CF2$ depicted in Fig. 2, we obtain

$\tau_{CF1} = \{(\text{nurse}, 5, 1, 4), (\text{doctor}, 1, 2, 2)\}$ and
$\tau_{CF2} = \{(\text{nurse}, 3, 1, 2), (\text{doctor}, 1, 2, 2), (\text{assistant}, 1, 4, 4)\}$.

Similarity between the timed aggregated resource views of two process schemas or fragments can be defined as follows. For comparing the temporal perspective a simple comparison between the intervals for each aggregated resource is utilized, i.e., calculating the differences between upper and lower interval limit divided by the sum of the interval lengths. This similarity value is combined with the CRS by weighing both equally. Note that if more complex temporal information is assigned to the tasks, more sophisticated similarity metrics can be used.

Definition 7 (Similarity Metrics for Timed Resource Assignment). *Let τ_1, τ_2 be two timed aggregated resource assignments for resource assignments ρ_1 and ρ_2. Then the similarity between τ_1 and τ_2 is defined as follows:*

$$sim(\tau_1, \tau_2) := \begin{cases} \frac{1}{2} * (sim(\rho_1, \rho_2) + \frac{\sum_{t1, t2} sim(t1, t2)}{max(|\tau_1|, |\tau_2|)}) & if \ \tau_1 \cup \tau_2 \neq \emptyset \\ 0 & otherwise \end{cases}$$

where $t1 = (r1, s1, l1, u1) \in \tau_1, t2 = (r2, s2, l2, u2) \in \tau_2$ and

$$sim(t1, t2) := \begin{cases} 1 & if \ r1 = r2 \wedge l1 = l2 \wedge u1 = u2 \\ 1 - \frac{|l1 - l2| + |u1 - u2|}{|u1 - l2| + |u2 - l1|)} & if \ r1 = r2 \\ 0 & otherwise \end{cases}$$

The metrics combines the similarity between the aggregated resource views with the more specific assessment of the resource requirements related to time. Each value can be also considered in a separated manner.

The corner cases for the metrics would be to have (a) highly similar or equal resource assignments that are due at the same time and (b) highly similar or equal resource assignments that are due at totally different times. For case (a), intuitively, the timed aggregated resource metrics yields a value close to or equal 1. For case (b) assume

$\rho_1 = \{(nurse, 3, 1, 3), (doctor, 1, 4, 5)\}$ and $\rho_2 = \{(nurse, 3, 4, 5), (doctor, 1, 6, 7)\}$. In this case $sim(\rho_1, \rho_2) = 1$. Then: $sim((nurse, 3, 1, 3), (nurse, 3, 4, 5)) = 1 - \frac{3+2}{1+4} = 0$ and $sim((doctor, 1, 4, 5), (doctor, 1, 6, 7)) = 1 - \frac{2+2}{1+3} = 0$. Hence, $sim(\tau_1, \tau_2) = 0.5$. This means that the high similarity between ρ_1 and ρ_2 is reduced by half because the same resources are required, but at totally different times.

For the example in Fig. 2, applying Definition 7 yields:

$sim((nurse, 3, 1, 2), (nurse, 5, 1, 4)) = 1 - \frac{2}{1+3} = 0.5$ and
$sim((doctor, 1, 2, 2), (doctor, 1, 2, 2)) = 1$.

Then: $sim(\tau_{CF1}, \tau_{CF2}) = \frac{0.57 + \frac{0.5+1}{3}}{2} = 0.54$.

This means that the deviations in the time assignments reduce the similarity of the aggregated resource assignment a bit.

For an example where $\rho_1 = \{(nurse, 3, 2, 4)\}$ and $\rho_2 = \{(nurse, 1, 2, 4)\}$ the resource assignment similarity would yield $sim(\rho_1, \rho_2) = 0.66$.

Looking at the time requirements, $sim((nurse, 3, 2, 4), (nurse, 2, 2, 4)) = 1$ and hence $sim(\tau_1, \tau_2) = \frac{0.66+1}{2} = 0.83$.

In this case the similarity increased when incorporating the time as a different number of the same resource is required at exactly the same time.

For comparing changes along their timed aggregated resource views a first proposal for a similarity metric is as follows:

Definition 8 (Timed Change Resource Similarity (TCRS)). *Let $\Delta_1 = (t1, f1, p1, S1)$, $\Delta_2 = (t2, f2, p2, S2)$ be two change operations and τ_1 (τ_2) the timed aggregated resource assignment for $f1$ ($f2$). The Timed Change Resource Similarity TCRS between changes Δ_1 and Δ_2 is defined as follows:*

$$TCRS(\Delta_1, \Delta_2) := \begin{cases} sim(\tau_1, \tau_2) \; if \; t1 = t2 \\ -sim(\tau_1, \tau_2) \; otherwise \end{cases}$$

Looking at the different examples and corner cases above, TCRS seems to make sense. However, the observations from the metrics start to become blurred if the earliest and latest point in time a resource is required span a longer time frame. The interpretation can be different. The resources could be required rather at the beginning and the end of the process or during the entire execution of the process. Both cases would be treated the same. Hence, a more fine-granule comparison becomes necessary. This aspect will be addressed in future work.

5   Implementation and Evaluation

In this section we present the implementation and evaluation of the approach. Change logs are provided by the *Apelands* [17] data set, a game-based experimentation environment for flexible and individual process settings.

5.1   Implementation

The prototype[2] developed as a basis for our evaluation provides an interface for inspecting process models, instances (Fig. 3) and the applied change operations. The metrics discussed in this paper are shown in Fig. 4.
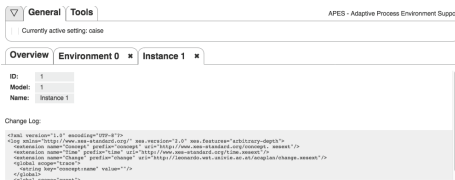
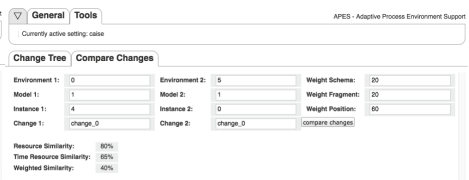**Fig. 3.** Details of a process instance in the prototype

Fig. 4. Comparing two change operations in the prototype

5.2   Discussion of Applicability

Using (T)CRS, one can compare process change operations even when there is no information about the resulting process schemas or about other perspectives of the process change operations. This can be useful for cases where (a) other information is not available, e.g. due to data privacy issues or (b) if the other data is not required to the current analysis. Thus, change operation similarity based on the temporal and resource perspective alone can be used as an alternative to structural or behavioral similarity metrics. In the next section, we will show that the results from (T)CRS actually correlate with structural similarity of the process schemas after the change operation has been executed.

5.3   Comparing (T)CRS Effect Similarity with Structural Similarity

We evaluate the effect similarity of two change operations as measured by (T)CRS against *node matching similarity* (NMS) of the resulting process schemas.

For the evaluation we have used data set 1 from the Apelands project[3]. This data set contains 136 change operations for 23 different basic models, of which

---

[2] http://cs.univie.ac.at/project/apes.

[3] The data set can be found at http://cs.univie.ac.at/project/apes.

107 are based on the same model. Apelands is a game-based experimentation and evaluation service for flexible process settings. While playing the round-based game, players adapt process instances, thus generating process change logs. These change logs contain all perspectives required for the metrics presented in this paper, i.e. a list of required resources for each activity which can be added to a process instance, and information about the game round the activity is supposed to be executed. Listing 1.1 shows a change fragment from the game. Each contains two activities which have been planned for the next two rounds (c.f. <round/> tag). In the <resources/> Tag, the required resources are shown.

Listing 1.1. Apelands change fragment

```
<fragment>
  <call>
    <name>'Fire Elemental'</name> <round>1</round>
    <resource name="Alchemy Lab" count=''2''/>
  </call>
  <call>
    <name>'Knight'</name> <round>2</round>
    <resource name=''Barracks'' count=''2''/>
  </call>
</fragment>
```

We compare the results of (T)CRS for the change operations *which are based on the same model* with the similarity of the resulting process model as calculated by NMS. In this special case, NMS similarity of the resulting process models can be seen as the effects of the change operations from the control flow perspective. We compare a randomly chosen change operation from this data set against all other change operations. As Figs. 5 and 6 show, both CRS and TCRS correlate with NMS. In most cases, the similarity as calculated by (T)CRS is higher than NMS, since (T)CRS are based on a subset of the attributes of the process models. Attributes not considered by (T)CRS which are different do not have any effect on (T)CRS scores, but lower the NMS score. This relationship can also be seen in the fact that TCRS has a closer correlation to NMS than CRS, since it also incorporates the time-related attributes, which are not used by CRS.

Conclusion: This evaluation shows that the similarity of change operation effects as calculated by (T)CRS correlates with the similarity of the resulting process schemas. Thus, such a metric can be used to compare process change operations independent of attribute-specific weights.

Threats to Validity: Our approach was evaluated with experimental data generated from a game, not with data from several different settings in which change operations occur. Also, we did not interview domain experts about the impact of the results of our metrics. These evaluations would be interesting additions to our validation which is based on a state of the art approach from process similarity, and will be addressed as future work.

6   Related Work

Soundness notions and checks for the application of change operations to process models and instances have been subject to several approaches. Structural
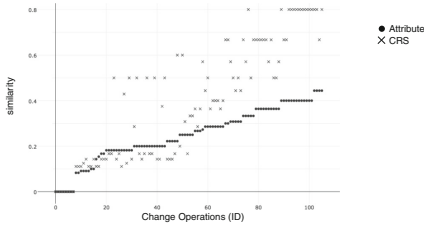
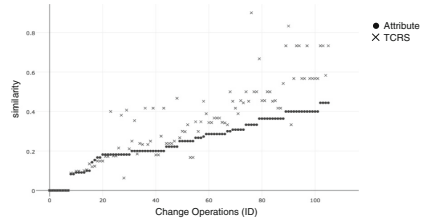**Fig. 5.** Comparing CRS with the attribute similarity

Fig. 6. Comparing TCRS with the attribute similarity

soundness with respect to control and data flow has been tackled in [18], whereas correctness criteria for the behavorial soundness of change operations are provided and compared in [19]. An overview on how to define and apply change operations on business processes is provided in [6]. Other approaches have focused on the representation of change information based on change logs [8], change processes [1], and change trees [2]. Especially, change processes and change trees aim at presenting information on past change operations to users in order to support decisions on future changes, although no assessment of the similarity of change operation is provided. The most related approach is ProCycle [3] where change operations are augmented with CBR-based techniques such that users can comment on the reasons for conducting a change. Then the changes can be compared. The approach at hand is different as it does not rely on additional comments, but only considers information that is available based on the change operations themselves.

In contrast to change operation similarity, process equivalence and similarity have been analyzed frequently in current literature. [20] discusses label equivalence, attribute equivalence, position equivalence and regional equivalence for processes which execute web services. [21] uses behavioral profiles of processes to compare their similarity. Comparing these profiles leads to a behavior based matching of processes. [14] measures the similarity of two processes based on causal footprints, which consist of a set of look-back and look-ahead links. [4] defines three metrics for measuring the similarity between process models, namely (a) node matching similarity, (b) structural similarity and (c) behavioral similarity. [22] defines the difference between two process models by a difference model that is visualized based on a difference graph. Doing so, differences between models can be visually inspected. [5] proposes similarity metrics between process instances, taking into consideration different perspective as well.

7   Conclusion and Future Work

Process change operations are usually applied when the situation requires domain experts to do so. Being able to show which change operations have similar effects on a certain perspective can be interesting for the person in charge of the adaptation: When adapting therapy processes in a nursing home, using

a resource perspective metric the nurse can see which therapy process fragments require similar resources. In conjunction with information about available resources for the upcoming weeks, such a clustering can facilitate the planning of process adaptations. When evaluating the effects of former process adaptations, comparison metrics can be used to create cluster of similar process change operations, thus improving analysis.

Effect metrics for other perspectives will be part of future work. Also comparing data structures based on change logs, e.g. change processes [1] and change trees [2] can be discussed based on the presented similarity metrics.

References

1. Günther, C.W., Rinderle, S., Reichert, M., van der Aalst, W.: Change mining in adaptive process management systems. In: Meersman, R., Tari, Z. (eds.) OTM 2006. LNCS, vol. 4275, pp. 309–326. Springer, Heidelberg (2006). doi:10.1007/11914853_19
2. Kaes, G., Rinderle-Ma, S.: Mining and querying process change information based on change trees. In: Barros, A., Grigori, D., Narendra, N.C., Dam, H.K. (eds.) ICSOC 2015. LNCS, vol. 9435, pp. 269–284. Springer, Heidelberg (2015). doi:10.1007/978-3-662-48616-0_17
3. Weber, B., Reichert, M., et al.: Providing integrated life cycle support in process-aware information systems. Int. J. Coop. Inf. Syst. **18**, 115–165 (2009)
4. Dijkman, R., Dumas, M., van Dongen, B., Krik, R., Mendling, J.: Similarity of business process models: metrics and evaluation. Inf. Syst. **36**, 498–516 (2011). Special Issue: Semantic Integration of Data, Multimedia, and Services
5. Pflug, J., Rinderle-Ma, S.: Process instance similarity: potentials, metrics, applications. In: Debruyne, C., et al. (eds.) OTM 2016. LNCS, vol. 10033. Springer, Cham (2016). doi:10.1007/978-3-319-48472-3_8
6. Reichert, M., Weber, B.: Enabling Flexibility in Process-Aware Information Systems - Challenges, Methods, Technologies. Springer, Heidelberg (2012)
7. Lanz, A., Reichert, M., Weber, B.: Process time patterns: a formal foundation. Inf. Syst. **57**, 38–68 (2016)
8. Rinderle, S., Reichert, M., Jurisch, M., Kreher, U.: On representing, purging, and utilizing change logs in process management systems. In: Dustdar, S., Fiadeiro, J.L., Sheth, A.P. (eds.) BPM 2006. LNCS, vol. 4102, pp. 241–256. Springer, Heidelberg (2006). doi:10.1007/11841760_17
9. Rinderle, S., Weber, B., Reichert, M., Wild, W.: Integrating process learning and process evolution – a semantics based approach. In: Aalst, W.M.P., Benatallah, B., Casati, F., Curbera, F. (eds.) BPM 2005. LNCS, vol. 3649, pp. 252–267. Springer, Heidelberg (2005). doi:10.1007/11538394_17
10. Hinge, K., Ghose, A., Koliadis, G.: Process SEER: a tool for semantic effect annotation of business process models. In: Enterprise Distributed Object Computing Conference, pp. 54–63 (2009)
11. Xu, H., Savarimuthu, B.T.R., Ghose, A., Morrison, E., Cao, Q., Shi, Y.: Automatic BDI plan recognition from process execution logs and effect logs. In: Cossentino, M., Fallah Seghrouchni, A., Winikoff, M. (eds.) EMAS 2013. LNCS, vol. 8245, pp. 274–291. Springer, Heidelberg (2013). doi:10.1007/978-3-642-45343-4_15

12. Rinderle-Ma, S., Reichert, M., Weber, B.: On the formal semantics of change patterns in process-aware information systems. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 279–293. Springer, Heidelberg (2008). doi:10.1007/978-3-540-87877-3_21
13. Russell, N., ter Hofstede, A.H.M., Edmond, D., van der Aalst, W.M.P.: Workflow data patterns: identification, representation and tool support. In: Delcambre, L., Kop, C., Mayr, H.C., Mylopoulos, J., Pastor, O. (eds.) ER 2005. LNCS, vol. 3716, pp. 353–368. Springer, Heidelberg (2005). doi:10.1007/11568322_23
14. 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
15. Becker, M., Laue, R.: A comparative survey of business process similarity measures. Comput. Ind. **63**, 148–167 (2012)
16. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions and reversals. Soviet physics doklady **10**, 707–710 (1966). (in Russian)
17. Kaes, G., Rinderle-Ma, S.: Generating data from highly flexible and individual process settings through a game-based experimentation service. In: Datenbanksysteme für Business, Technologie und Web, pp. 331–350 (2017)
18. Reichert, M., Dadam, P.: Adeptflex–supporting dynamic changes of workflows without losing control. J. Intell. Inf. Syst. **10**, 93–129 (2011)
19. Rinderle, S., Reichert, M., Dadam, P.: Correctness criteria for dynamic changes in workflow systems a survey. Data Knowl. Eng. **50**, 9–34 (2004)
20. Rinderle-Ma, S., Reichert, M., Jurisch, M.: On utilizing web service equivalence for supporting the composition life cycle. Int. J. Web Serv. Res. **8**, 41–67 (2011)
21. Kunze, M., Weidlich, M., Weske, M.: m3 - a behavioral similarity metric for business processes. In: Services und ihre Komposition (2011)
22. Kriglstein, S., Wallner, G., Rinderle-Ma, S.: A visualization approach for difference analysis of process models and instance traffic. In: Daniel, F., Wang, J., Weber, B. (eds.) BPM 2013. LNCS, vol. 8094, pp. 219–226. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40176-3_18