Alignment-based Quality Metrics in Conformance Checking

Boudewijn van Dongen, Josep Carmona, Thomas Chatain

Alignment-based Quality Metrics in Conformance Checking

B.F. van Dongen[1] and J. Carmona[2] and T. Chatain[3]

Abstract: The holy grail in process mining is a process discovery algorithm that, given an event log, produces fitting, precise, properly generalizing and simple process models. Within the field of process mining, conformance checking is considered to be anything where observed behaviour, e.g., in the form of event logs or event streams, needs to be related to already modelled behaviour.

In the conformance checking domain, the relation between an event log and a model is typically quantified using fitness, precision and generalization. In this paper, we present metrics for fitness, precision and generalization, based on alignments and the newer concept named anti-alignments.

Keywords: alignments, process mining, quality metrics, fitness, precision, generalization

1 Quality Metrics in Process Mining

The holy grail in process mining is a process discovery algorithm that, given an event log, produces fitting, precise, properly generalizing and simple process models. Event logs are generally considered to be accurate representations of the behaviour of a system in such as way that each event refers to an *activity* that was executed in the context of a *case*. By deriving a process model from such an event log, process discovery algorithms give insights into the underlying system. There has been always a discussion on how to interpret process discovery results, i.e. how does the produced model relate to the actual, but unknown, system in four quality dimensions [BvDvdA14]:

Fitness quantifies how much of the observed behaviour is captured by the model,
Generalization quantifies how well the model explains unobserved system behaviour,
Precision quantifies how much behavior exists in the model that was not observed, and
Simplicity quantifies the complexity of the model and is not considered in this paper.

Consider an example event log in Table 1, together with several process models as depicted in Figure 1 to 4. The model in Figure 1 shows the "ideal" process discovery result, i.e. the model that is fitting, fairly precise and properly generalizing. Models 2 and 3 are chosen such that they score poorly on at least one of the dimensions precision or generalization, while model 4 scores around 0.5 in both dimensions.

Tab. 1: An example event log

Trace	Frequency
$\langle A,B,D,E,I\rangle$	1207
$\langle A,C,D,G,H,F,I\rangle$	145
$\langle A,C,G,D,H,F,I\rangle$	56
$\langle A,C,H,D,F,I\rangle$	23
$\langle A,C,D,H,F,I\rangle$	28

[1] b.f.v.dongen@tue.nl, Eindhoven University of Technology, Eindhoven (The Netherlands)
[2] jcarmona@cs.upc.edu, Universitat Politècnica de Catalunya, Barcelona (Spain)
[3] chatain@lsv.ens-cachan.fr, LSV, ENS Cachan, CNRS, INRIA, Universit Paris-Saclay, Cachan (France)

Table 2 contains our conformance metrics for fitness (F), proposed in [Ad14, vdAAvD12] and for precision (P) and generalization (G) proposed in [vDCC16] for the models in Figure 1 to Figure 4. In this paper, we explain the intuition behind these three metrics and we refer to literature for further details.

Tab. 2: Precision and Generalization for all models (excerpt from [vDCC16])

Model		F	P	G
Figure 1	Generating model	1.000	0.871	0.206
Figure 2	Single trace	0.915	1.000	0.000
Figure 3	Flower model	1.000	0.000	1.000
Figure 4	G,H as self-loops	1.000	0.588	0.466

Alignment based Fitness. A model *fits* a log if all traces in the log can be replayed by the model. A fitness metric quantifies the fit of a log in a model. Several different measures exist for this quality dimension [BvDvdA14, Ad14, vdAAvD12, and references therein]. The most recent and robust technique uses a cost-based alignment between the traces in the event log and the most optimal execution of the process model [Ad14].

An ***alignment*** is a sequence of pairs that refer to an event from a trace and a transition in the model, or $\gg$ elements indicating deviations. The projection of these pairs onto the first element yields the trace from the log and the projection onto the second elements yields a firing sequence in the model. Each pair in an alignment is called a move. If both parts of the pair are equally labelled, we call such a pair a *synchronous move*. A *model move* is a pair $\frac{\gg}{t}$, i.e. a transition is fired, but no corresponding event appeared in the log and a *log move* is a pair $\frac{a}{\gg}$, i.e. an event appears in the log, but there is no corresponding transition to be fired in the model.

Consider for example the trace $\langle A,C,H,D,F,I\rangle$ and model 2. The optimal alignment for this trace in that model is $\langle \frac{A}{A}, \frac{\gg}{B}, \frac{C}{\gg}, \frac{H}{\gg}, \frac{D}{D}, \frac{\gg}{E}, \frac{F}{\gg}, \frac{I}{I}\rangle$ with 5 deviations and a fitness of $1-\frac{5}{11}=0.55$.

Typically, a cost function is used to compute so-called optimal alignments, such that the number of model moves, i.e. pairs $(a,\gg)$, and log moves, i.e. pairs $(\gg,t)$ in the alignment is minimized. Then, using an optimal alignment, i.e. an alignment with minimal cost, for each trace fitness is calculated by adding all penalties for log and model moves and

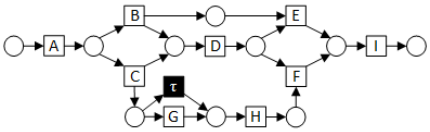

Fig. 1: The ideal model. Fitting, fairly precise and properly generalizing.

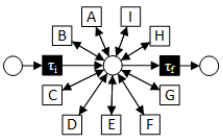

Fig. 3: The flower model. Fitting and generalizing, but very imprecise.

Fig. 2: Most frequent trace. Precise, but not fitting or generalizing.

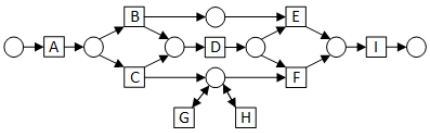

Fig. 4: A model with G and H in self-loops

dividing that by the worst-case costs, i.e. the costs of an alignment with only log and model moves. For more details on fitness, we refer to [BvDvdA14, Ad14, vdAAvD12, and references therein].

Anti-alignment based Precision. A model is considered *precise* if the behaviour the model allows for that was not observed in the event log is small. In this section, we present the precision metric proposed in [vDCC16], which uses anti-alignments as introduced in [CC16]. An ***anti-alignment*** is a firing sequence of a model which differs (according to some distance metric) sufficiently from all the observed traces in a log[4].

Consider the log in Table 1 and model 4. A maximal anti-alignment of length 10 is the firing sequence $\langle A,C,G,G,D,G,G,G,F,I\rangle$ with minimal edit distance $\frac{4}{10} = 0.4$. In contrast, model 1 has only one anti-alignment, namely $\langle A,C,G,H,D,F,I\rangle$ with length 7 and distance $\frac{1}{7}$) and model 2 and 3 do not have anti-alignments.

In [vDCC16], we suggest a fresh view on precision, using anti-alignments[5]. The intuition behind our metric is as follows. A very precise process model allows for exactly the observed traces to be executed and not more. Hence, if one trace is removed from the log, this trace becomes the anti-alignment for the remaining log as it is the only execution of the model that is not in the log. We use this property to estimate precision.

For each trace in the log, we compute a maximal anti-alignment for the model and the log without that trace. This anti-alignment is guaranteed to reach the final marking. Then, we compute the distance between the removed trace and the anti-alignment found which we average over the log, *not* taking into account the relative frequencies of the traces in the log. If the language of the model equals the log, then the anti-alignments will be equal to each removed trace, hence the precision is 1. If for every trace, an anti-alignment can be produced which has maximal distance from that trace, the precision is 0.

Frequencies of traces are not considered as the comparison is between the language of the model and the observed traces. Observing one trace more frequently than another should not influence the precision of the model as the unobserved behaviour does not change.

In trace-based precision, the length of the anti-alignment considered is bounded by the length of the removed trace. This guarantees that an anti-alignment exists in the log without that trace, but also limits the possibility to see imprecise executions of the model that are much longer than the lengths of the observed traces. Therefore, we also defined a log-based precision metric, which uses an anti-alignment of the model with respect to the entire log of a much greater length than the longest trace observed in the log. The log-based precision metric uses a single anti-alignment of considerable maximum length to determine the amount of behaviour allowed by the model, but not observed in the event log. Our final precision metric is a weighted sum of log- and trace-based precision.

Anti-alignment based Generalization. In contrast to precision, which relates the log and the model, *generalization* relates some unknown system to the log and the model as it rea-

[4] We refer to `http://www.lsv.ens-cachan.fr/~chatain/darksider/` for the anti-alignment tool.
[5] This approach is implemented in the ProM package "anti-alignments" `http://www.promtools.org/`.

sons over expected future behaviour. Generalization aims to estimate the extent to which unobserved, but likely possible behaviour, is explained by the model.

In order to quantify generalizations, we consider not only the sequential behaviour that is actually allowed by the model, but we also quantify how different this behaviour is when considering the state space of the model. (Structured) loops and parallel structures, which are most commonly used to achieve generalization when modelling a system, have introduce new traces while introducing fewer new states. Therefore, in our generalization metric, we consider the notion of a *recovery distance* for an anti-alignment.

Like for precision, we first consider trace-based generalization following the same leave-one-out procedure. This way, the model is guaranteed to contain an anti-alignment of some distance (i.e. the removed trace). Not using trace-based generalization would lead us to consider all models non-generalizing if the log equals the language of the model. In contrast to precision, generalization depends on the frequency of the traces and once again, using the average over trace-based and log-based generalization we get one metric.

2 Conclusions and Future Challenges

In this paper we presented three metrics for fitness [vdAAvD12], precision [vDCC16] and generalization [vDCC16] based on alignments and anti-alignments. The future challenges in the domain of conformance checking are in *scalability*, i.e. dealing with volume and velocity and variability of input data; *approximation*, i.e. balancing computation time with accuracy of alignments; *understandability*, i.e. providing explanations of the root causes between observed and modelled behaviour; *multi-perspective conformance*, i.e. considering data, resources and time; *measurability*, i.e. providing a better understanding of the different conformance dimensions and solid metrics having a formal interpretation.

References

[Ad14] Adriansyah, A.: Aligning observed and modeled behavior. PhD thesis, Eindhoven University of Technology, 2014.

[BvDvdA14] Buijs, J.C.A.M.; van Dongen, B.F.; van der Aalst, W.M.P.: Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization and Simplicity. Int. J. Cooperative Inf. Syst., 23(1), 2014.

[CC16] Chatain, T.; Carmona, J.: Anti-alignments in Conformance Checking - The Dark Side of Process Models. In 37th International Conference, PETRI NETS 2016. volume 9698 of Lecture Notes in Computer Science. Springer, pp. 240–258, 2016.

[vdAAvD12] van der Aalst, W.M.P.; Adriansyah, A.; van Dongen, B.F.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery, 2(2):182–192, 2012.

[vDCC16] van Dongen, B.F.; Carmona, J.; Chatain, T.: A Unified Approach for Measuring Precision and Generalization Based on Anti-Alignments. In: Business Process Management - 14th International Conference, BPM 2016, Proceedings to appear.