

Alignment-based conformance checking over probabilistic events

Jiawei Zheng
School of Informatics
University of Edinburgh
jw.zheng@ed.ac.uk

Petros Papapanagiotou
School of Informatics
University of Edinburgh
pe.p@ed.ac.uk

Jacques D. Fleuriot
School of Informatics
University of Edinburgh
jdf@ed.ac.uk

Abstract

Conformance checking techniques allow us to evaluate how well some exhibited behaviour, represented by a trace of monitored events, conforms to a specified process model. Modern monitoring and activity recognition technologies, such as those relying on sensors, the IoT, statistics and AI, can produce a wealth of relevant event data. However, this data is typically characterised by noise and uncertainty, in contrast to the assumption of a deterministic event log required by conformance checking algorithms. In this paper, we extend alignment-based conformance checking to function under a probabilistic event log. We introduce a weighted trace model and weighted alignment cost function, and a custom threshold parameter that controls the level of confidence on the event data vs. the process model. The resulting algorithm considers activities of lower but sufficiently high probability that better align with the process model. We explain the algorithm and its motivation both from formal and intuitive perspectives, and demonstrate its functionality in comparison with deterministic alignment using real-life datasets.

Keywords: Conformance checking, Probabilistic events, Uncertainty, Probabilistic cost function

1. Introduction

Conformance checking is the task of comparing the behaviour captured by event logs recorded during the execution of a process to the intended behaviour of a corresponding process model (Carmona et al., 2018). It evaluates how well the execution matches the modelled process and allows us to detect when things are diverging from a desired or expected workflow. A

widely used conformance checking approach is based on the *alignment* of the events in the log with the activities in the process model, where any misalignments are potential deviations (Adriansyah et al., 2011).

With the advances of sensor technologies, the Internet of Things (IoT) and modern Artificial Intelligence, we are now generating an abundance of data that can be used to detect processes as they are being performed in the physical world. However, this data is typically noisy, with errors caused by various factors such as the environment, sensor accuracy, faults, malfunctions, and the inherent uncertainty of statistical and machine learning techniques used to analyse the data. The standard *alignment* conformance checking approach is only able to fit deterministic events that are assumed to reflect reality with certainty (Cohen and Gal, 2021). This means that probabilistic event data has to be reduced to a deterministic event log in order to be used. This removes information from the event log, thus lowering the confidence on any detected deviations.

We present an extension of alignment-based conformance checking under the assumption of a probabilistic event log with a categorical distribution over a set of activities. We introduce a cost function that takes activity probabilities into consideration, and a custom parameter that allows the algorithm to align activities of lower but sufficiently high probability that better agree with the process model, as opposed to always assuming the most-probable activities occurred. In effect, this leverages the knowledge captured in the process model to address levels of uncertainty in the event data, with the aim of reducing the number of false positives and false negatives in deviation detection. We further motivate our work with an example from recent research in Activities of Daily Living (Section 2). We give a formal presentation of our algorithm (Section 5),

explain the operation and intuition behind its parameter (Section 6), and evaluate the performance on real-life datasets (Section 7). We implement our algorithm as a Python package based on the PM4Py framework (Berti et al., 2019) and open-sourced on GitHub¹.

2. Motivating Example

Smart home monitoring of Activities of Daily Living (ADLs) has seen increased attention in recent research, particularly in the context of home care (Zheng and Papapanagiotou, 2022). For instance, ADL data can be used to analyse the daily routine of an older adult (as a personal process model) and detect when they deviate from that healthy behaviour using conformance checking (Di Federico and Burattin, 2023), such as skipping medication, etc.

ADL data is typically based on smart home and wearable sensors. A Human Activity Recognition (HAR) algorithm, often based on supervised learning, helps classify that data into distinct daily activities, such as sleeping, eating, reading, etc. Subsequent analysis, including for conformance, usually assumes the most probable activity is the one that took place, and the alternatives, no matter how likely, are never considered (Cohen and Gal, 2021; Szytler et al., 2016). However, the result of HAR is uncertain due to inter-class similarity, usually taking the form of a categorical probability distribution over the possible classes of activities. For example, a HAR algorithm may classify some sensor input as the *drinking from cup* activity with 33% probability, *answering phone* with 34% probability, or *spraying from a bottle* with 33% probability, all three of which options have a similar pattern in the sensor data (Gupta and Davis, 2007).

In the context of ADL monitoring for care, a misclassification of *drinking from cup* as *answering phone*, which is only 1% more likely in our example, may lead to an alert of a deviation, for instance related to the levels of hydration of the person. This would be an unnecessary, false positive alert, caused by the fact that the conformance analysis only considers the most probable activities. Instead, we want an algorithm that operates under an assumption of uncertainty, such that it considers the process model as an additional source of information of what may have occurred in reality, beyond the noisy sensor and HAR data.

In contrast, if the probability of an activity is too low (e.g. a 3% probability of *drinking from cup* compared to a 64% probability of *answering phone*), we may want to consider it as non-conforming, even if it is expected by the process model. Therefore, incorporating the level of

confidence in the pattern of the process model compared to the uncertainty of the event log is a key requirement.

The same challenges are emerging in other domains that feature sensors and AI prominently, such as manufacturing (*Industry 4.0*), autonomous vehicles and healthcare, requiring conformance to production flows, driving regulations and care pathways respectively.

3. Related work

Alignment conformance checking algorithms transform the problem of computing alignments between event logs and process models into an optimal search problem using the A* algorithm, i.e., minimising the total alignment cost (Adriansyah et al., 2011).

Recent work on conformance checking under uncertainty has three focal points: missing data (Felli et al., 2022; Pegoraro et al., 2021), uncertain event activity mappings (van der Aa et al., 2020), and uncertain process models (Bergami et al., 2021). Koorneef et al. present a probability-based alignment cost function, based in frequencies of events in historical event logs, to obtain the most *probable* alignment. We adapt this in the context of uncertain events.

Even more recently, Cohen and Gal (2021) presented a review of the challenges related to probabilistic event data in conformance checking techniques. They highlighted the demand for new conformance checking techniques that can adapt multiple types of uncertain event data, which further motivates our work.

Based on the challenges identified by Cohen et al., Bogdanov et al. (2022a) presented a conformance checking algorithm over stochastic logs, which takes the probability of event occurrences into account explicitly, similarly to our algorithm. They apply their algorithm on *trace recovery*, i.e., recovering the original trace from a probabilistic one based on a maximal alignment to a reference process model (Bogdanov et al., 2022b). However, in their approach an event is aligned to a corresponding activity in the process model regardless of the probability that activity actually occurred, as long as it is non-zero. This assumes the occurrence of activities with very low or even close to zero probability, as long as they align with the process model, therefore discarding potential true positive deviations.

Overall, the need for conformance checking under uncertainty has been recognised by researchers in the area, but the related research is still in its infancy.

4. Preliminaries

Our approach is built upon well-established literature on alignment-based conformance checking

¹ <https://github.com/jia-wei-zheng/ProbCost>

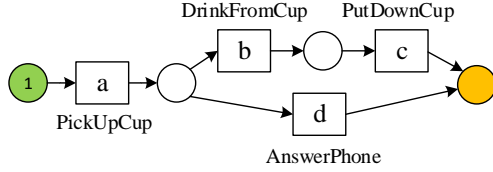


Figure 1: Example of a process model.

(Carmona et al., 2018). We provide brief definitions of relevant concepts so that this paper is self-contained.

Definition 1 (Petri net). Given a set of activities A , a *Petri net* is defined as a tuple $N = (P, T, F, \alpha)$, where P and T are sets of places and transitions respectively. F is a set of arcs representing flow relations between transitions and places, so that $F \subseteq (P \times T) \cup (T \times P)$. A labelling function $\alpha : T \rightarrow A \cup \{\tau\}$ assigns either an activity from A or τ (immediate transition not associated with any activity) to each transition in T .

Petri Nets offer a standard representation for *process models*, such as the example shown in Figure 1.

An *event log* captures information about the execution of multiple cases (or instances) of a process model. Each case is represented by a trace of events $\sigma = \langle e_0, e_1, \dots, e_n \rangle$ that correspond to *observations* of the activities that occurred during the process. We assume each event corresponds to a single activity in the process. An event is *deterministic* if it is associated with an activity in the process model in a deterministic way.

When aligning a trace of events to a model, the trace is transformed into a *Petri net* called a *trace model*.

Definition 2 (Trace model). Given a sequence of events $\sigma = \langle e_0, e_1, \dots, e_n \rangle$ over a set of activities A can be converted into a *Petri net* $N_t = (P^t, T^t, F^t, \alpha^t)$, called a *trace model*. In this, each event e_i is mapped to a transition in T^t . The transitions are then interleaved with places in P^t to form a linear sequence.

We combine the trace model with a given process model in a single Petri Net called a *synchronous product net*. In this, the set of transitions T^s consists of 3 subsets, namely *synchronous moves* T^{SM} , *model moves* T^{MM} , and *log moves* T^{LM} . Combining a transition for the same activity a from the 2 models results in a *synchronous move* $(a, a) \in T^{SM}$, meaning that an event corresponding to a is *aligned* to activity a in the process model. Transitions in the original process model that do not have a corresponding transition in the trace model are represented as *model moves* $(\gg, a) \in T^{MM}$ or *model moves on τ* (\gg, τ) and transitions in the original trace model that do not correspond to the process model as *log moves* $(a, \gg) \in T^{LM}$ (Carmona et al., 2018).

An alignment is a *sequence of transition firings*

in the synchronous product net. There exists at least one alignment, such that contains only model moves, followed by log moves. However, the goal is to find the *optimal* alignment between the process and trace given some cost function $c(t)$ for each transition $t \in T^s$.

Definition 3 (Cost function, Standard cost function). A cost function $c : T^s \rightarrow \mathbb{R}^+$ associates a non-negative cost to the firing of each transition of the synchronous product net (Bloemen et al., 2018), such that:

$$c(t) = \begin{cases} 0, & t = (\gg, \tau) \text{ (model move on } \tau) \\ 0, & t \in T^{SM} \text{ (synchronous move, e.g. } \\ & \quad (a, a)) \\ 1, & t \in T^{LM} \text{ (log move, e.g. } (a, \gg)) \\ 1, & t \in T^{MM} \text{ (model move, e.g. } (\gg, a)) \end{cases}$$

Given this, the problem of finding an *optimal alignment* is reduced to searching an execution sequence of the synchronous product with the minimum cost. The standard cost function ensures an optimal number of synchronous moves, i.e. events and process transitions that align together at the same time. A *fitness* score will be derived from the alignment.

5. Conformance checking over probabilistic events

This section presents ProbCost, our alignment algorithm. Traditional alignment-based conformance checking only considers traces of *deterministic* events. In our work, we instead consider a trace of m events $\langle e_0, \dots, e_{m-1} \rangle$ each of which can correspond to one of n possible activities $\{a_0, \dots, a_{n-1}\}$ with the *categorical probability* $p_{i,j} = p(a_j | e_i)$ that event e_i is an observation of activity a_j . This can be modelled by a probability matrix as follows:

$$\begin{matrix} & e_0 & e_1 & \dots & e_{m-2} & e_{m-1} \\ \begin{matrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{matrix} & \left\| \begin{matrix} p_{0,0} & p_{1,0} & \dots & p_{m-2,0} & p_{m-1,0} \\ p_{0,1} & p_{1,1} & \dots & p_{m-2,1} & p_{m-1,1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ p_{0,n-1} & p_{1,n-1} & \dots & p_{m-2,n-1} & p_{m-1,n-1} \end{matrix} \right\| \end{matrix} \quad (1)$$

where entries $p_{i,j} \in [0, 1]$ and $0 \leq i \leq m-1, 0 \leq j \leq n-1$ represent the probability that event e_i corresponds to activity a_j . The sum of the probabilities for each event should be 1, i.e. $\sum_{j=0}^{n-1} p_{i,j} = 1, \forall 0 \leq i \leq m-1$.

An example trace of 3 events $\langle e_0, e_1, e_2 \rangle$ each of which may correspond to 3 activity categories a, b, c can be modelled by the following probability matrix:

$$\mathbf{P} = \begin{matrix} & e_0 & e_1 & e_2 \\ \begin{matrix} a \\ b \\ c \end{matrix} & \left\| \begin{matrix} 0.3 & 0 & 0 \\ 0.7 & 0.7 & 0.3 \\ 0 & 0.3 & 0.7 \end{matrix} \right\| \end{matrix} \quad (2)$$

where the first event e_0 has a 0.3 probability to be associated with activity a and 0.7 probability to be associated with b .

This probability matrix can be modelled as a *weighted trace model* as follows:

Definition 4 (Weighted trace model). Let A be a set of n activities and σ a trace of m events, where each $e_i \in \sigma$ corresponds to each $a_j \in A$ with probability $p_{i,j}$. A weighted trace model is a Petri net with weights, i.e., $(P^{wt}, T^{wt}, F^{wt}, \alpha^{wt}, w^{wt})$, where:

- $P^{wt} = \{P_0, \dots, P_{m-1}\}$ a set of m places,
- $T^{wt} = \bigcup_{i,j} \{t_{i,j} \mid p_{i,j} > 0\}$, where $t_{i,j}$ is a transition for event e_i if activity a_j corresponds to it with probability $p_{i,j} > 0$,
- $F^{wt} = \bigcup_{i,j} \{(P_{j-1}, t_{i,j}), (t_{i,j}, P_j) \mid t_{i,j} \in T^{wt}\}$
- $\alpha^{wt}(t_{i,j}) = a_j$, for each transition $t_{i,j} \in T^{wt}$,
- $w^{wt} : T^{wt} \rightarrow [0, 1]$, where $w^{wt}(t_{i,j}) = p_{i,j}$ is a function assigning a weight to each transition based on the probabilistic matrix of the event log,

Note that, in contrast to deterministic trace models, a weighted trace model has multiple transitions for every event, linked to possible activities with different weights. The weighted trace model corresponding to probabilistic matrix (2) is shown in Figure 2.

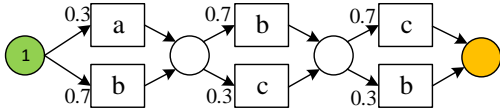


Figure 2: Weighted trace model of matrix (2).

Definition 5 (Weighted synchronous product net). Let A be a set of n activities, $N_p = (P^p, T^p, F^p, \alpha^p)$ a process model and σ an captured probabilistic trace of m events with categorical probability $p_{i,j}$ over the activities, with the corresponding weighted trace model $N_{wt} = (P^{wt}, T^{wt}, F^{wt}, \alpha^{wt}, w^{wt})$. A synchronous product net is a Petri net with weights, i.e., $N_{ws} = (P^{ws}, T^{ws}, F^{ws}, \alpha^{ws}, w^{ws})$, such that:

- $P^{ws} = P^p \cup P^{wt}$,
- $T^{ws} = (T^{MM} \cup T^{LM} \cup T^{SM})$, where $T^{MM} = \{\gg\} \times T^p$ denotes moves on model, $T^{LM} = T^{wt} \times \{\gg\}$ denotes moves on log, $T^{SM} = \{(t_1, t_2) \in T^p \times T^{wt} \mid \alpha^p(t_1) = \alpha^{wt}(t_2)\}$ denotes synchronous moves.
- $F^{ws} = \{(p, (t_1, t_2)) \in P^{ws} \times T^{ws} \mid (p, t_1) \in F^p \vee (p, t_2) \in F^{wt}\} \cup \{((t_1, t_2), p) \in T^{ws} \times P^{ws} \mid (t_1, p) \in F^p \vee (t_2, p) \in F^{wt}\}$,

- $\alpha^{ws}((t_1, t_2)) = (l_1, l_2)$ for all transitions $(t_1, t_2) \in T^{ws}$, where $l_1 = \alpha^p$ if $t_1 \in T^p$, otherwise $l_1 = \gg$; and $l_2 = \alpha^{wt}$ if $t_2 \in T^{wt}$, otherwise $l_2 = \gg$,
- $w^{ws} : T^{ws} \rightarrow [0, 1]$ a weight function for each transition, where $w^{ws}((t_1, t_2)) = w^{wt}(t_1)$ if $(t_1, t_2) \in (T^{LM} \cup T^{SM})$, otherwise $w^{ws}((t_1, t_2)) = 1$,

Note that the worst-case of the number of transitions T^{ws} is exponential to the number of events, which occurs when (i) the number of events in the trace is equal to the number of activities in the process model, (ii) each event has multiple transitions, and (iii) all transitions match to a corresponding labelled activity in the process model. This means that all events can be matched to every activity in the model. However, such high level of uncertainty and generality is unrealistic in the context of the applications that have motivated this work.

Figure 3 shows the weighted synchronous product net of our example process model (Figure 1) and weighted trace model (Figure 2).

Knowing the probabilities of each activity given an observed event allows us to calculate the cost of each transition based on the likelihood that the corresponding activity actually occurred. This likelihood is reflected in the weight w^{ws} of each transition. Activities with higher probability should have lower cost so that the algorithm is more likely to select them in the optimal alignment. We use a *log* transformation to transform probability products to sums of log probabilities. Based on this, we define a new, weighted cost function $c(t)$:

Definition 6 (Weighted cost function). Given a weighted synchronous product net $N_{ws} = (P^{ws}, T^{ws}, F^{ws}, \alpha^{ws}, w^{ws})$, we define a weighted cost function for each transition $t \in T^{ws}$ of the synchronous product net as follows:

$$c(t) = \begin{cases} 0, & t = (\gg, \tau) \\ -\log(w^{ws}(t)), & t \in T^{SM} \\ -\log(w^{ws}(t)) - \log(\epsilon), & t \in T^{LM} \\ -\log(\epsilon), & t \in T^{MM} \end{cases}$$

where $\epsilon \in (0, 1)$ is a parameter representing the level of confidence in the event log (see Section 6).

Using this cost similarly to the standard alignment algorithm (Adriansyah et al., 2011), we find an optimal alignment by using A* algorithm and derive a *fitness* score based on the alignment using the same way.

ProbCost includes the following steps: (i) build the process and weighted trace models, (ii) construct the weighted synchronous product net, (iii) use A* to find an optimal alignment based on the weighted cost function,

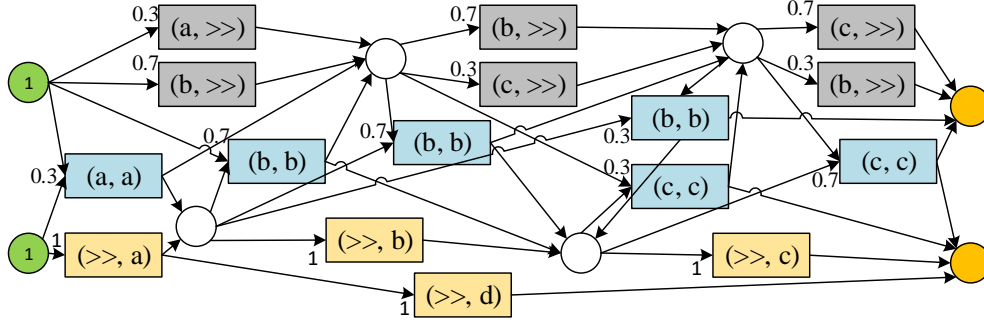


Figure 3: Weighted synchronous product net of our example models from Figure 1 and 2, annotated with the weight function w^{ws} and with different colours for model moves (yellow), log moves (grey) and synchronous moves (blue).

Table 1: Alignment results of 3 events with probability matrix (2) and the process model from Figure 1.

(a) Standard alignment					(b) Probabilistic alignment				
Event log	\gg	b	b	c	Event log	a	b	c	
Process model	a	\gg	b	c	Process model	a	b	c	

(iv) tune the parameter ϵ based on the context, (v) obtain final alignment using the tuned ϵ .

Example 1. Consider the process model (Figure 1) and the example probabilistic trace in (2), and their composed weighted synchronous product net in Figure 3, we could get the cost of the synchronous move, e.g., (a, a) is $-\log(0.3)$, the cost of the model move, e.g. (\gg, a) is $-\log(\epsilon)$, and the cost of the log move, e.g. (a, \gg) is $-\log(0.3) - \log(\epsilon)$. The optimal alignment computed by ProbCost ($\epsilon = 0.4$) is shown in Table 1b, with 100% fitness. Compared to the results of standard alignment algorithm (shown in Table 1a, with 67% fitness) which assumes the *most likely* activities for each event, i.e. the trace $\langle b, b, c \rangle$, ProbCost chooses activity a for the first event, even though it has a lower probability (0.3) than b (0.7) and forms a perfect alignment with the process model. This is because the cost of choosing the same trace $\langle b, b, c \rangle$ as the standard alignment algorithm is higher than the cost of trace $\langle a, b, c \rangle$.

This fits our intuition that even though a is less likely to have occurred given our first event observation, the process model captures the knowledge that a is *expected* to occur (e.g. based on prior observations or physical restrictions), thus increasing our perceived likelihood for a . We thus use our confidence in the modelled process to mitigate the uncertainty of the event log.

Example 2. Considering the same process model and event log with Example 1, we change ϵ to be 0.8. ProbCost gets the same optimal alignment as the

standard alignment algorithm, as shown in Table 1a.

Compared to Example 1, this shows that the ϵ parameter is able to control the acceptable probability of events. Instead of blindly accepting the activity *expected* by the process model, changing ϵ allows us to determine whether we trust the event log's indication of most probable activity or the process model's expectation. We analyse ϵ more formally in Section 6.

Complexity analysis Compared to the standard A* alignment approach, whose worst-case space complexity is linear to the number of reachable state of the synchronous product net (Carmona et al., 2018), we have the same number of reachable states and, therefore, the same space complexity. Furthermore, time complexity depends on the number of transitions between initial and final states of the synchronous product net (Carmona et al., 2018). Therefore, the worst-case time complexity of ProbCost is exponential compared to the standard approach, when the worst-case number of transitions occurs in the synchronous product net. An empirical evaluation of computation time is discussed in Section 7.2.

6. Threshold parameter ϵ

Our weighted cost function $c(t)$ includes a “threshold” parameter ϵ . As hinted in the previous section, ϵ allows the user to control the level of confidence, or *trust*, in the modelled process given an uncertain event log. An intuitive way to express this is by posing the following question:

How probable does an event need to be for it to move synchronously with a matching activity in the process model?

Let us explore a minimal example whereby ϵ allows us to control the answer to that question.

Consider a trace with a single event e_0 that

Table 2: Possible optimal alignments of a single event with 2 possible activities a , b and process model with single step a .

(a) Assuming a occurred.	(b) Assuming b occurred.										
<table border="1" style="margin: auto; border-collapse: collapse;"> <tr> <td style="padding: 2px;">Event log</td> <td style="padding: 2px;">a</td> </tr> <tr> <td style="padding: 2px;">Process model</td> <td style="padding: 2px;">a</td> </tr> </table>	Event log	a	Process model	a	<table border="1" style="margin: auto; border-collapse: collapse;"> <tr> <td style="padding: 2px;">Event log</td> <td style="padding: 2px;">b</td> <td style="padding: 2px;">\gg</td> </tr> <tr> <td style="padding: 2px;">Process model</td> <td style="padding: 2px;">\gg</td> <td style="padding: 2px;">a</td> </tr> </table>	Event log	b	\gg	Process model	\gg	a
Event log	a										
Process model	a										
Event log	b	\gg									
Process model	\gg	a									

corresponds to either of two possible activities a and b with probabilities $p_{0,a} = x$ and $p_{0,b} = 1 - x$, for some $x \in [0, 1]$. We want to check conformance of this trace with a model containing a as a single step in the process. In this case, we have 2 possible optimal alignments shown in Table 2.

The first (Table 2a) contains a synchronous move which assumes the event corresponds to activity a . The second (Table 2b) considers that the event corresponds to activity b , leading to a model move and a log move. The choice between the two should be based on the probabilities of a and b . The question is *what value does x need to be for the first alignment to be optimal?*

In the extreme cases, if $x = 1$ then a occurred with 100% probability, so the first alignment should be optimal, and symmetrically if $x = 0$ then the second alignment should be optimal. Furthermore, if we only consider the most likely activity as the one that actually occurred, which is in fact standard in activity recognition (Szt Tyler et al., 2016), then the second alignment would be optimal only when $x < 0.5$.

However, we posit that this decision may be influenced by the process model. For instance, consider the extreme case where the process model reflects domain knowledge that *only activity a is actually possible*, whereas activity b is not possible *at all*². A high probability $p_{0,b}$ might then be attributed, for example, to noise in the data, and the second alignment should *always* be sub-optimal.

More generally, we may choose to *trust* the reality (or our expectation of it) as reflected by the process model more than a noisy and uncertain event log, i.e. perform a synchronous move on a even when $x < 0.5$. Notably, Bogdanov et al. (2022a) follow an inflexible approach compared to ours, such that always considers the first alignment to be optimal even when x approaches zero.

In our approach, the parameter ϵ allows us to control how low x can be for a to be chosen by the algorithm. Given Definition 6, the total cost of the alignment in Table 2a is $-\log(x)$ (synchronous move on a), whereas the total cost of the alignment in Table 2b is $-\log(1 -$

²Of course, such an extreme model would defeat the purpose of conformance checking to begin with, since we already know what is possible, but it helps illustrate our point.

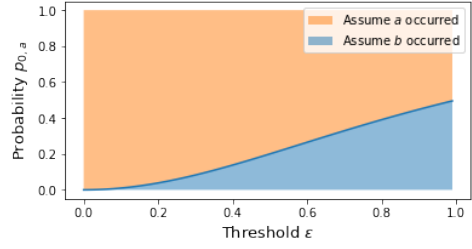


Figure 4: Relation between ϵ , $p_{0,a}$, and assumed activity.

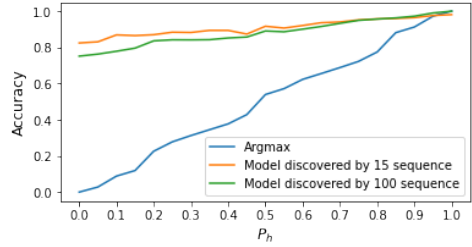


Figure 5: Trace recovery accuracy for different number of traces in model discovery.

$x) - 2\log(\epsilon)$ (move on log b and move on model). The former is optimal under the condition $\frac{x}{1-x} \geq \epsilon^2$.

For the synchronous move on the model's expected activity a to be optimal (i.e. to assume that a occurred) the ratio of $p_{0,a}$ to $p_{0,b}$ should be more than ϵ^2 .

The relationship between $p_{0,a}$, ϵ and the assumed activity is shown in Figure 4. For each combination of values for $p_{0,a}$ and ϵ above the line, the algorithm assumes a occurred and picks the first alignment as optimal. Otherwise below the line, it assumes b occurred and picks the second alignment as optimal. The lower the threshold, the lower the probability of a needs to be for a synchronous move, and therefore the more we trust the process model instead of the probability distribution of the event log. With $\epsilon \in (0, 1)$, the threshold for the probability of a can be set to be anywhere in $(0, 0.5)$.

When considering more complex process models and larger traces, with multiple possible activities per event, the interaction between ϵ , the activity probabilities and the alignment costs become harder to analyse. This makes the selection of the appropriate value of ϵ more difficult. However, a general principle applies: *when ϵ is closer to 1, the algorithm trusts the probabilities of the event log and considers the most probable activities. As the value of ϵ decreases, the algorithm puts more trust in the process model and accepts activities with lower probability in the event log if they align to those in the model.*

This selection of the appropriate value for ϵ depends on the context of each case study and our confidence

in the process model. In many of our examples when choosing between optimal alignments, we have observed ϵ to be the threshold for the *ratio of probabilities* of the activities corresponding to each alignment. This pattern is also observed in our experimental study. Pending further research on the methods to determine the appropriate value of ϵ , we select its value based on (i) the developed intuition (higher values mean higher trust on the log), (ii) the ratio of the involved probabilities and (iii) tuning ϵ on training datasets in a supervised way (see Section 7.2).

7. Experimental study

In this section, we evaluate the performance of ProbCost in *trace recovery* and *deviation detection*. These are common tasks where conformance checking can provide value in real world applications, and have been used for the evaluation of conformance checking algorithms in related literature (Bogdanov et al., 2022b). We first evaluate ProbCost’s ability to replicate the trace recovery results of Bogdanov et al. Next, we evaluate ProbCost’s performance in deviation detection compared to the standard alignment and Bogdanov et al.’s approach.

Our experiment employs three publicly available real-world datasets: (i) BPI Challenge (BPIC) 2012³ related to personal loan processes, which contains 13087 cases over 36 activities (assessing the application, calling after sent offers, etc.), (ii) BPIC 2019⁴ related to purchase order handling processes including 251734 cases over 42 activities, (submit order, payment, etc.), and (iii) ADLs daily log (Sztyley et al., 2016), which records individuals’ daily routines for 10 days, including 13 types of activities, (eat, take medicine, etc.). Both BPIC datasets have been used to evaluate algorithms in relevant literature on uncertain data settings (Bogdanov et al., 2022a; Pegoraro et al., 2020).

7.1. Evaluating trace recovery

In our first experiment, we investigate the ability of ProbCost to recover the real trace from a probabilistic event log (trace recovery) under the same assumptions as Bogdanov et al. Replicating their strategy on the same BPIC 2012 dataset, we first discover a process model from a set of randomly selected traces, then generate probabilistic events by adding noise to the original activities of a different set of 100 traces.

For each event in each trace we attach a second alternative activity to its original label, selected

randomly from the set of activities in the dataset. The original activity is assigned a randomly chosen probability p , whereas the added activity is assigned $1 - p$. A parameter $P_h \in [0, 1]$ specifies the proportion of original activities with higher probabilities than their added alternatives ($p > 1 - p$). When $P_h = 1$ all the original activities are assigned higher probability than the added alternatives, and vice versa when $P_h = 0$.

We then perform conformance checking with our algorithm, and extract the optimally aligned sequences of events as the recovered traces. We calculate the recovery accuracy by dividing the number of correctly recovered events (compared to the original ones) by the total number of events, and compare the results against the Argmax sequence (choosing the activity with highest probability for each event).

The threshold parameter ϵ is set to a low value (0.01) to simulate the intention of Bogdanov et al. to maximally align activities to the process model regardless of probability. For the model discovery we use the Inductive Miner (Adriansyah et al., 2011) as implemented in PM4Py (Berti et al., 2019), on separate sets of 15 and 100 traces. Our results compared to Argmax are shown in Figure 5,

Our results seem to match those of (Bogdanov et al., 2022b, Figure 5). Specifically, when P_h is 0, recovery accuracy based on the model discovered by 15 traces (0.81) is better than the result using 100 traces (0.75), which is same to their result. As P_h increases, the accuracy increases, until P_h reaches 1, where the same result as Argmax is obtained. This validates that their approach can be replicated by our algorithm by setting a low value of ϵ , close to 0 (e.g. 0.01) as a baseline.

7.2. Evaluating deviation detection

Experiment design In this experiment, we evaluate the performance of our algorithm when detecting deviations on all three datasets, i.e. identifying events that do not fit the process model. In the context of alignment-based conformance checking, a log move indicates a deviation (the event is inconsistent with the process model), while a synchronous move means a normal occurrence (the event fits with the activity in the process model). We compare the performance of our approach with the standard alignment algorithm over the Argmax sequences and the approach of Bogdanov et al. Note that in contrast to the assumption we made in Section 7.1, if events fit the process model but occur with low probability, we regard them as deviations.

We aim to synthesise a set of probabilistic event traces in such a way that we can control the ratio of deviations, and there exist low probability activities that

³https://data.4tu.nl/articles/_/12689204/1

⁴<https://icpmconference.org/2019/icpm-2019/contests-challenges/bpi-challenge-2019/>

are not deviations. We accomplish this as follows.

For the two BPI datasets, we first perform model discovery using 20 randomly chosen traces, and choose another 100 traces from the rest to generate probabilistic events using the same setting as previously. For the smaller ADL dataset which contains 10 traces of daily logs, we randomly chose 5 traces for model discovery and the remaining 5 for generating probabilistic events. We use $P_h = 0$ so that all the added activities have higher probabilities than the original ones ($p < 1 - p$).

Standard conformance checking over the traces against the discovered model yielded 100%, 96% and 100% fitness for the three datasets respectively. However, for the sequences of *added* activities, the average fitness and standard deviation (SD) are 23% (0.08 SD), 8% (0.11 SD), and 49% (0.15 SD), respectively. This means that the original traces conform to the process model (even though they were not used in discovery), whereas the added activities do not have a good fit. Based on this, we define deviations in a probabilistic trace by considering the original activities as *normal occurrences* and the added ones as *deviations*.

Furthermore, under our assumption of noisy data, we introduce a *deviation confidence* parameter T_d . Specifically, for each event, if the odds of the original activity happening, i.e. the ratio of the probability of the original activity p to that of the alternative is higher than T_d , i.e. $\frac{p}{1-p} \geq T_d$, then we classify this event as a normal occurrence. Otherwise, the event is a deviation.

For example, for $T_d = 0.50$, if $p < 0.33$ then the alternative occurred (true deviation). However, if $p = 0.40$, then we determine that the event is in fact an actual occurrence. Note that $p < 0.50$ always, since $P_h = 0$.

If T_d is close to 0, the original activity is a normal occurrence even when its probability is very low, as long as it is non-zero, resulting in very few deviations. As T_d increases from 0 to 1 the number of normal events decreases while the number of deviations increases.

Therefore, in this setup we control the ratio of deviations using T_d and normal occurrences have low probabilities ($P_h = 0$).

We use accuracy, F1-score, sensitivity ($\frac{TP}{TP+FN}$) and specificity ($\frac{TN}{TN+FP}$) to measure the performance of deviation detection based on the counts of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) (Arifoglu and Bouchachia, 2019). We also use Geometric mean (G-mean, i.e., $\sqrt{\text{sensitivity} \times \text{specificity}}$) metric to balance both sensitivity and specificity on imbalanced datasets (Kubat et al., 1997).

Tuning threshold ϵ Next, we set the *deviation confidence* T_d to 0.25, which means the probability of the original activity should be higher than 0.20 to be

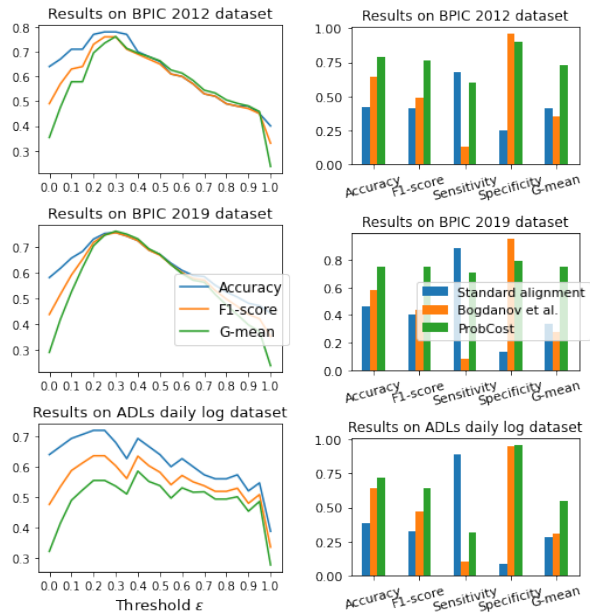


Figure 6: Deviation detection performance for different values of ϵ on different datasets ($T_d = 0.25$). Figure 7: Results for deviation detection for different values of ϵ on different datasets ($T_d = 0.25$).

considered as a normal occurrence. We evaluate the performance of ProbCost over the development set (70% of the whole data) under different values of ϵ , ranging from 0.05 to 1 with a 0.05 step. The accuracy, F1-score and G-mean are calculated for each value of ϵ to determine the optimal value that yields the best deviation detection performance, as these metrics are suitable for evaluating imbalanced datasets. The results are presented in Figure 6. The best performance was achieved for ϵ between 0.25 to 0.30 for all three datasets (variation due to classes proportions), which is basically equal to the deviation confidence T_d . This follows the ϵ selection strategy discussed in Section 6. Based on this, we set $\epsilon = T_d$ for further experiments.

Results We assess the deviation detection results using our algorithm ($\epsilon = 0.25$) compared to the standard alignment conformance checking algorithm as implemented in PM4Py and the approach by Bogdanov et al. (replicated by our algorithm with $\epsilon = 0.01$). The comparative results are shown in Figure 7.

Based on these results, ProbCost performs better than both the standard alignment algorithm and Bogdanov et al. in *Accuracy*, *F1-score* and *G-mean* metrics. The last one has the lowest *sensitivity* and mostly highest *specificity* because it always selects the activities that fit in the process model even when their probability is very low, leading to more false negatives.

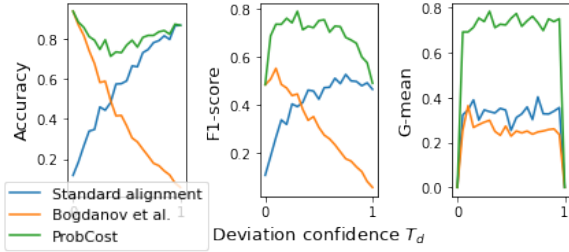


Figure 8: BPIC 2019 results for different values of T_d .

In contrast, the standard alignment algorithm has highest *sensitivity* and lowest *specificity*, because it may ignore the normally occurring activity even if its probability is close to the probability of the alternative, leading to more false positives. ProbCost achieves better *G-mean* score than the other two approaches, it suggests that it is better at correctly identifying both positive and negative cases, resulting in a better overall performance. In brief, our algorithm can better balance the confidence between the process model and the uncertain event log to obtain better results than both the others.

We further investigate the three algorithms under different deviation confidence values T_d . For this, we iterate T_d from 0 to 1 with 0.05 step to generate deviation data before running the 3 algorithms. As mentioned previously, we set $\epsilon = T_d$. The approach of Bogdanov et al. is replicated by setting $\epsilon = 0.01$.

ProbCost outperforms the others for all the values of T_d in three datasets. Results for different datasets show similar tendencies, we only show the results of BPIC 2019 in Figure 8 due to space considerations. For lower values, as expected, our performance matches that of Bogdanov et al. as no matter the probability of the original activity, it can always be aligned with the expected activity in the process model to identify a normal occurrence. However, the standard alignment algorithm gives the lowest accuracy and F1-score because it chooses the most probable activity for each event, resulting in more false positives. All three algorithms produce the lowest G-mean when $T_d = 0$, because there are no deviations, causing the sensitivity to be 0. Conversely, there are no normal occurrences when $T_d = 1$, causing specificity to be 0.

As the deviation confidence T_d increases, ProbCost achieves best performances among the approaches. The accuracy and F1-score of approach by Bogdanov et al. decrease, while they increase for the standard algorithm. This happens because as T_d increases, the original activities need to have higher probability to be considered as normal occurrences, but Bogdanov et al. accept activities expected by the

process model even with low probability, resulting in more false negatives. ProbCost can better balance the confidence on uncertain events and select the activity with sufficiently high probability to achieve best deviation detection accuracy. When T_d is increased to 1, our approach yields the same accuracy and F1-score as the standard alignment algorithm. As we put more trust on the event log (higher ϵ), we end up considering the most probable activities, similarly to the standard alignment algorithm.

Finally, we assess the execution time of ProbCost on three datasets with different sizes of events and process models. The results (shown in Table 3) indicate that the execution time of ProbCost is longer than the standard approach as expected by the complexity analysis (Section 5) because the number of transitions increases, but it is still within a reasonable range. This indicates that our approach is applicable in practice.

Overall, our approach demonstrates more resilience to noise and uncertainty in the events, and flexible control of how much the algorithm should trust the event log over the structure of the process model via ϵ .

8. Conclusion

In this paper, an extended alignment conformance checking algorithm is proposed for use with probabilistic event logs. Specifically, we propose a weighted trace model for events with a categorical probability distribution, a weighted alignment cost function, and a custom threshold parameter that controls the level of confidence on the event log vs. the process model. Furthermore, we compared the complexity results with deterministic alignment approach. Two sets of experimental studies comparing our approach to the deterministic alignment algorithm and the recent relevant approach by Bogdanov et al. show that our algorithm takes into consideration occurrence probabilities explicitly and is able to accommodate activities with lower, but sufficiently high probabilities if they fit the model. Based on this, we argue that our algorithm can better tolerate noise in the event log by leveraging the knowledge captured in the process model with the aim of reducing the false positive and false negative rates in deviation detection. As such, it is better suited to perform conformance checking under an uncertain, noisy event log, such one produced by sensors or AI-based algorithms (Cohen and Gal, 2021).

In future work, we plan to further investigate the effect of ϵ to guide the selection of an appropriate value, as we are in favour of empirical evaluation at this stage. We also plan to investigate the results with different experimental settings, such as more alternative activities

Table 3: Comparison of computation time between standard alignment and ProbCost for different datasets.

Datasets	#Cases	#Events per case				Process model		Computation time per case (in seconds)	
		min	max	avg	median	#places	#transitions	Standard	ProbCost
BPIC 2012	100	3	56	24.7	27.5	44	59	356.150	865.201
BPIC 2019	100	1	10	5	5	12	18	0.174	0.629
ADLs daily log	5	7	22	15	14	26	41	48.094	93.322

per event, etc. We also plan to improve the efficiency of algorithm and explore other types of uncertainty of event data, e.g., repeating or missing events.

Acknowledgements

This research was funded by the Legal & General Group (research grant to establish the independent Advanced Care Research Centre at University of Edinburgh). The funder had no role in conduct of the study, interpretation or the decision to submit for publication. The views expressed are those of the authors and not necessarily those of Legal & General.

References

Adriansyah, A., van Dongen, B., and van der Aalst, W. (2011). Conformance Checking Using Cost-Based Fitness Analysis. In *2011 IEEE EDOC*, pages 55–64.

Arifoglu, D. and Bouchachia, A. (2019). Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks. *Artificial Intelligence in Medicine*, 94:88–95.

Bergami, G., Maggi, F. M., Montali, M., and Peñaloza, R. (2021). Probabilistic Trace Alignment. In *ICPM 2021*, pages 9–16.

Berti, A., van Zelst, S. J., and van der Aalst, W. M. P. (2019). Process mining for python (pm4py): Bridging the gap between process- and data science. In *ICPM Demo Track 2019*, volume 2374, pages 13–16.

Bloemen, V., van Zelst, S. J., van der Aalst, W. M. P., van Dongen, B. F., and van de Pol, J. (2018). Maximizing Synchronization for Aligning Observed and Modelled Behaviour. In *BPM 2018*, pages 233–249, Cham.

Bogdanov, E., Cohen, I., and Gal, A. (2022a). Conformance Checking over Stochastically Known Logs. In *BPM Forum*, pages 105–119, Cham.

Bogdanov, E., Cohen, I., and Gal, A. (2022b). Trace Recovery from Stochastically Known Logs.

Carmona, J., van Dongen, B., Solti, A., and Weidlich, M. (2018). *Conformance Checking: Relating Processes and Models*. Springer, Cham.

Cohen, I. and Gal, A. (2021). Uncertain process data with probabilistic knowledge: Problem characterization and challenges. In *2021 BPM Workshop*, volume 2938, pages 51–56.

Di Federico, G. and Burattin, A. (2023). Do you behave always the same? In *ICPM Workshops*, pages 5–17.

Felli, P., Gianola, A., Montali, M., Rivkin, A., and Winkler, S. (2022). Conformance Checking with Uncertainty via SMT. In *BPM 2022*, volume 13420, pages 199–216, Cham.

Gupta, A. and Davis, L. S. (2007). Objects in Action: An Approach for Combining Action Understanding and Object Perception. In *CVPR 2007*, pages 1–8.

Koorneef, M., Solti, A., Leopold, H., and Reijers, H. A. (2018). Automatic Root Cause Identification Using Most Probable Alignments. In *2018 BPM Workshops*, pages 204–215, Cham.

Kubat, M., Matwin, S., et al. (1997). Addressing the curse of imbalanced training sets: one-sided selection. In *ICML*, volume 97, page 179. Citeseer.

Pegoraro, M., Uysal, M. S., and van der Aalst, W. M. P. (2020). Efficient Construction of Behavior Graphs for Uncertain Event Data. In *Business Information Systems*, pages 76–88, Cham.

Pegoraro, M., Uysal, M. S., and van der Aalst, W. M. P. (2021). Conformance checking over uncertain event data. *Information Systems*, 102:101810.

Sztyler, T., Carmona, J., Völker, J., and Stuckenschmidt, H. (2016). Self-tracking Reloaded: Applying Process Mining to Personalized Health Care from Labeled Sensor Data. *Transactions on Petri Nets and Other Models of Concurrency*, pages 160–180.

van der Aa, H., Leopold, H., and Reijers, H. A. (2020). Efficient Process Conformance Checking on the Basis of Uncertain Event-to-Activity Mappings. *IEEE TKDE*, 32(5):927–940.

Zheng, J. and Papapanagiotou, P. (2022). Predictive Behavioural Monitoring and Deviation Detection in Activities of Daily Living of Older Adults. In *HEALTHINF 2022*, pages 899–910.