

Anti-Alignments in Conformance Checking – The Dark Side of Process Models

Thomas Chatain Josep Carmona Boudewijn van Dongen

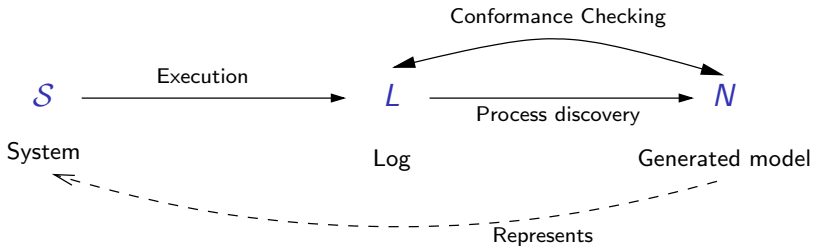
LSV, ENS Paris-Saclay, Cachan, France
chatain@lsv.ens-cachan.fr

Universitat Politècnica de Catalunya, Barcelona, Spain
jcarmona@cs.upc.edu

Eindhoven University of Technology, The Netherlands
b.f.v.dongen@tue.nl

Advanced Tutorial on Process Mining
Zaragoza
June 27, 2017

Process Mining

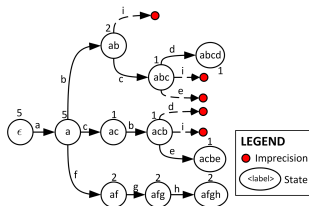
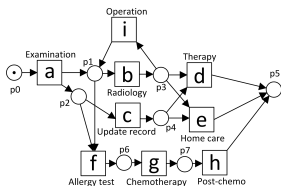


Conformance Checking:

- ▶ N fits L if $L \subseteq \mathcal{L}(N)$
- ▶ N is precise if $\mathcal{L}(N) \setminus L$ is small
- ▶ N generalizes L with respect to S if $\mathcal{L}(N)$ contains some unobserved behavior in $\mathcal{L}(S) \setminus L$

Measuring Precision – State of the Art

Log:
 $\langle a, b, c, d \rangle$
 $\langle a, c, b, e \rangle$
 $\langle a, f, g, h \rangle$

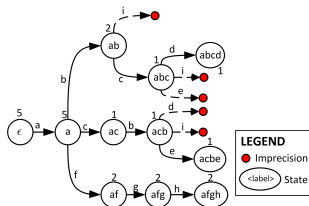
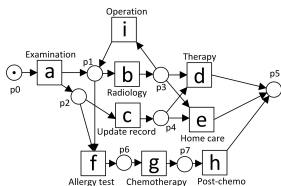


Alignment-based precision metrics [Adriansyah *et al.*]

- ▶ Build a representation $\mathcal{A}_{\Gamma(N,L)}$ of the part of the behaviour of the model which is covered by the log
- ▶ Count escaping points in $\mathcal{A}_{\Gamma(N,L)}$

Measuring Precision – State of the Art

Log:
 $\langle a, b, c, d \rangle$
 $\langle a, c, b, e \rangle$
 $\langle a, f, g, h \rangle$



Alignment-based precision metrics [Adriansyah *et al.*]

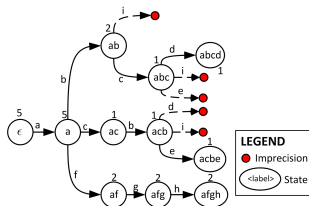
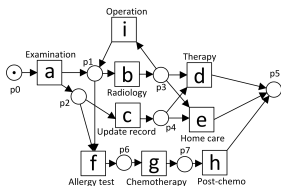
- ▶ Build a representation $\mathcal{A}_{\Gamma(N,L)}$ of the part of the behaviour of the model which is covered by the log
- ▶ Count escaping points in $\mathcal{A}_{\Gamma(N,L)}$

Drawbacks of a_p :

- ▶ Short sighted: only a step ahead of log behavior is considered

Measuring Precision – State of the Art

Log:
 $\langle a, b, c, d \rangle$
 $\langle a, c, b, e \rangle$
 $\langle a, f, g, h \rangle$



Alignment-based precision metrics [Adriansyah *et al.*]

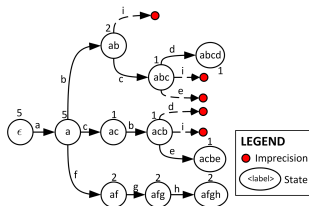
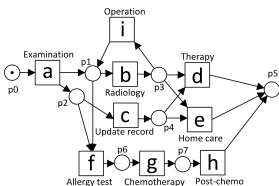
- ▶ Build a representation $\mathcal{A}_{\Gamma(N,L)}$ of the part of the behaviour of the model which is covered by the log
- ▶ Count escaping points in $\mathcal{A}_{\Gamma(N,L)}$

Drawbacks of a_p :

- ▶ Short sighted: only a step ahead of log behavior is considered
- ▶ Non-monotonic: observing a new trace may unveil new imprecisions

Measuring Precision – State of the Art

Log:
 $\langle a, b, c, d \rangle$
 $\langle a, c, b, e \rangle$
 $\langle a, f, g, h \rangle$
 $\langle a, b, i, b, c, d \rangle$



Alignment-based precision metrics [Adriansyah *et al.*]

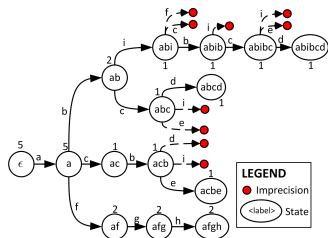
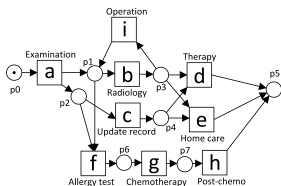
- ▶ Build a representation $\mathcal{A}_{\Gamma(N,L)}$ of the part of the behaviour of the model which is covered by the log
- ▶ Count escaping points in $\mathcal{A}_{\Gamma(N,L)}$

Drawbacks of a_p :

- ▶ Short sighted: only a step ahead of log behavior is considered
- ▶ Non-monotonic: observing a new trace may unveil new imprecisions

Measuring Precision – State of the Art

Log:
 $\langle a, b, c, d \rangle$
 $\langle a, c, b, e \rangle$
 $\langle a, f, g, h \rangle$
 $\langle a, b, i, b, c, d \rangle$



Alignment-based precision metrics [Adriansyah *et al.*]

- ▶ Build a representation $\mathcal{A}_{\Gamma(N,L)}$ of the part of the behaviour of the model which is covered by the log
- ▶ Count escaping points in $\mathcal{A}_{\Gamma(N,L)}$

Drawbacks of a_p :

- ▶ Short sighted: only a step ahead of log behavior is considered
- ▶ Non-monotonic: observing a new trace may unveil new imprecisions

Anti-alignments – Motivation

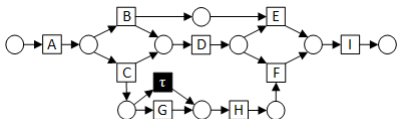
Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$



Motivation

In order to measure precision, find the run of N which is most **misaligned** with the log L .

Anti-alignments – Motivation

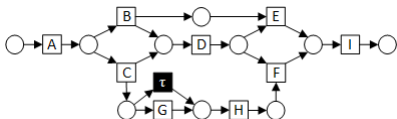
Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$



Motivation

In order to measure precision, find the run of N which is most **misaligned** with the log L .

Here: $\langle A, B, D, E, I \rangle$

Anti-alignments

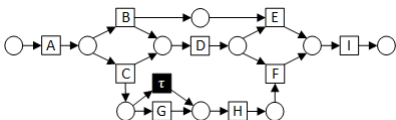
Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$



- ▶ $L \subset \Sigma^*$: a log (set of traces) of an observed system
- ▶ N : a (labeled) Petri net model (constructed by process discovery)

Definition (Anti-alignment)

An (n, m) -anti-alignment of a model N w.r.t. a log L is a run $\gamma \in \mathcal{L}(N)$ such that

- ▶ $|\gamma| \leq n$ and
- ▶ for every $\sigma \in L$, $dist(\gamma, \sigma) \geq m$.

Anti-alignments

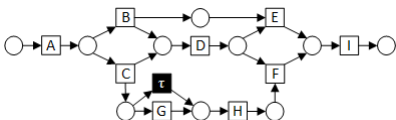
Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$



- ▶ $L \subset \Sigma^*$: a log (set of traces) of an observed system
- ▶ N : a (labeled) Petri net model (constructed by process discovery)

Definition (Anti-alignment)

An (n, m) -anti-alignment of a model N w.r.t. a log L is a run $\gamma \in \mathcal{L}(N)$ such that

- ▶ $|\gamma| \leq n$ and
- ▶ for every $\sigma \in L$, $dist(\gamma, \sigma) \geq m$.

Definition (Edit distance $dist(\gamma, \sigma)$)

Number of edits/deletions/insertions needed to edit γ to σ .

Anti-alignments: Example

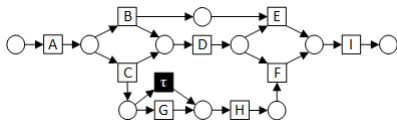
Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$



$(3, 5)$ anti-alignment $\langle A, B, D, E, I \rangle$

Best Anti-alignment for Given Length n

For a given length n :

- ▶ Assuming that the model N has a (full) run γ of length $\leq n$, this run is a $(n, 0)$ -anti-alignment
- ▶ Every $(n, m + 1)$ -anti-alignment is also a (n, m) -anti-alignment
- ▶ There cannot exist any $(n, n + 1)$ -anti-alignment

$\max^n(N, L)$

Define $\max^n(N, L)$ as the largest m for which there exists a (n, m) -anti-alignment.

I.e. find the run of length n of the model which deviates most from all the observed traces.

Anti-alignments to Measure Precision

- ▶ $L \subset \Sigma^*$: a log (set of traces) of an observed system
- ▶ N : a (labeled) Petri net model (constructed by process discovery)

Anti-alignment-based precision metrics

$$P_I(N, L) = 1 - \frac{\max^n(N, L)}{n}$$

with

- ▶ n : (in the order of) the maximal length for a trace in the log
- ▶ $\max^n(N, L)$: the largest m for which there exists a (n, m) -anti-alignment

Clearly, $\max^n(N, L) \in [0 \dots n]$ which implies $P_I(N, L) \in [0 \dots 1]$.

Anti-alignments to Measure Precision – Exercise

Sort the models by decreasing precision.

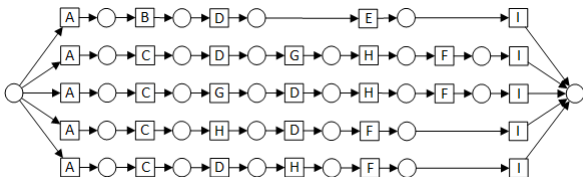
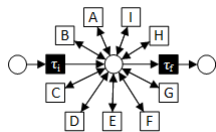
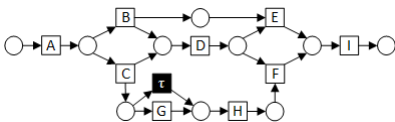
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

Log: $\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$



Anti-alignments to Measure Precision – Exercise

Sort the models by decreasing precision.

For each model, find the best anti-alignment of length ≤ 7 .

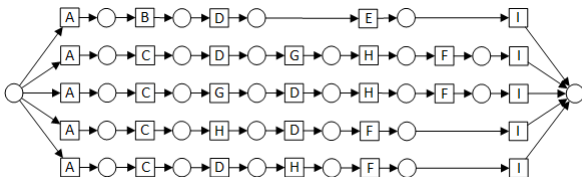
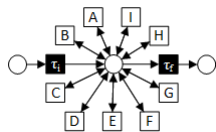
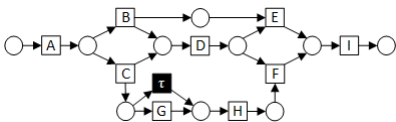
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

Log: $\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$



Anti-alignments to Measure Precision – Exercise

Sort the models by decreasing precision.

For each model, find the best anti-alignment of length ≤ 7 .

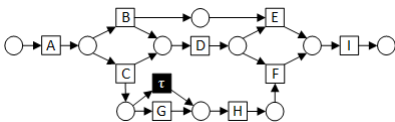
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

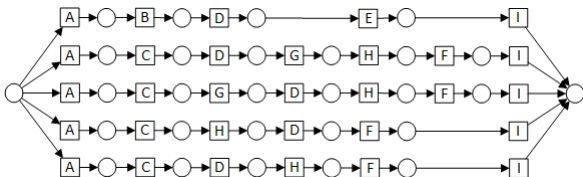
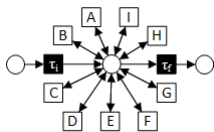
Log: $\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$



Anti-alignment $\langle A, C, G, H, D, F, I \rangle$
 $P_I(N_1, L) = 0.857$



Anti-alignments to Measure Precision – Exercise

Sort the models by decreasing precision.

For each model, find the best anti-alignment of length ≤ 7 .

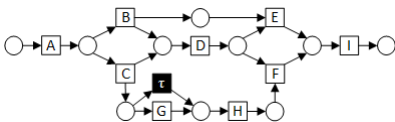
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

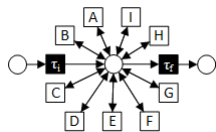
Log: $\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

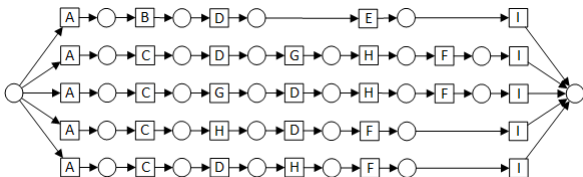
$\langle A, C, D, H, F, I \rangle$



Anti-alignment $\langle A, C, G, H, D, F, I \rangle$
 $P_I(N_1, L) = 0.857$



Anti-alignment
 $\langle I, I, I, A, A, A, A \rangle$
 $P_I(N_2, L) = 0$



Anti-alignments to Measure Precision – Exercise

Sort the models by decreasing precision.

For each model, find the best anti-alignment of length ≤ 7 .

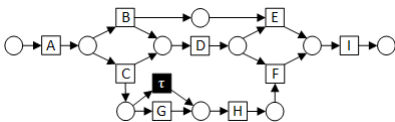
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

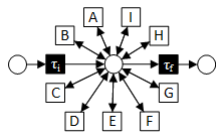
Log: $\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

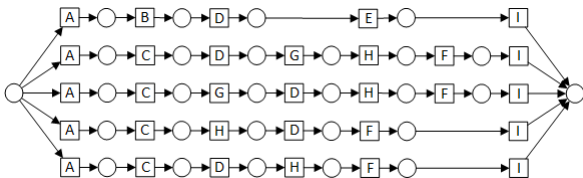
$\langle A, C, D, H, F, I \rangle$



Anti-alignment $\langle A, C, G, H, D, F, I \rangle$
 $P_I(N_1, L) = 0.857$



Anti-alignment
 $\langle I, I, I, A, A, A, A \rangle$
 $P_I(N_2, L) = 0$



No $(7, 1)$ -anti-alignment
 $P_I(N_3, L) = 1$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

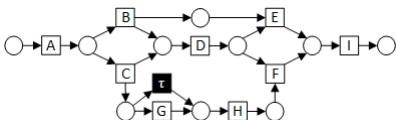
Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

Example

Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$



Best anti-alignment

$\langle A, B, D, E, I \rangle$

$\max^n(N, L)$

4

$P_I(N, L)$

$\frac{3}{7}$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

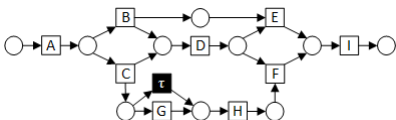
Example

Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, B, D, E, I \rangle$



Best anti-alignment $\max^n(N, L) \quad P_I(N, L)$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

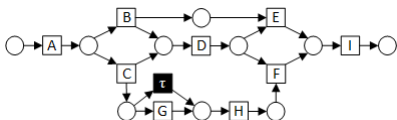
Example

Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, B, D, E, I \rangle$



Best anti-alignment

$\langle A, C, H, D, F, I \rangle$

$\max^n(N, L) \quad P_I(N, L)$

2

$\frac{5}{7}$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

Example

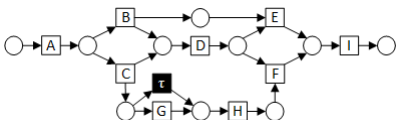
Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, B, D, E, I \rangle$

$\langle A, C, D, H, F, I \rangle$



Best anti-alignment $\max^n(N, L) \quad P_I(N, L)$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

Example

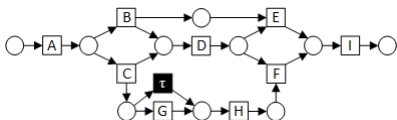
Log L :

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, B, D, E, I \rangle$

$\langle A, C, D, H, F, I \rangle$



Best anti-alignment

$\langle A, C, H, D, F, I \rangle$

$\max^n(N, L) \quad P_I(N, L)$

2

$\frac{5}{7}$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

Example

Log L :

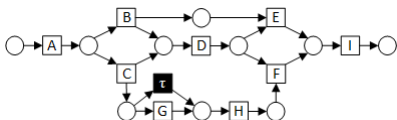
$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, B, D, E, I \rangle$

$\langle A, C, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$



Best anti-alignment $\max^n(N, L) \quad P_I(N, L)$

Monotonicity w.r.t. New Observations

Observing a new trace which happens to be already a run of the model, can only increase the precision measure.

Theorem

For every N, L and for every $\sigma \in \mathcal{L}(N)$,

$$P_I(N, L \cup \{\sigma\}) \geq P_I(N, L)$$

Hint: every (n, m) -anti-alignment for $(N, L \cup \{\sigma\})$ is also a (n, m) -anti-alignment for (N, L) .

Example

Log L :

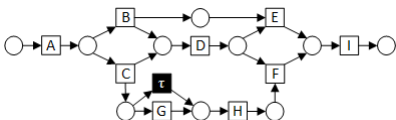
$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, B, D, E, I \rangle$

$\langle A, C, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$



Best anti-alignment

$\langle A, C, G, H, D, F, I \rangle$

$\max^n(N, L) \quad P_I(N, L)$

1

$\frac{6}{7}$

Monotonicity w.r.t. Language

Theorem

Given two models N_1 and N_2 , if $\mathcal{L}(N_1) \subseteq \mathcal{L}(N_2)$, then N_1 is more precise than N_2 .

$$\mathcal{L}(N_1) \subseteq \mathcal{L}(N_2) \implies P_I(N_1, L) \geq P_I(N_2, L)$$

Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
→ Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
 → Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace:

σ

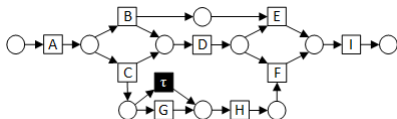
$\langle A, B, D, E, I \rangle$
 $\langle A, C, D, G, H, F, I \rangle$
 $\langle A, C, G, D, H, F, I \rangle$
 $\langle A, C, H, D, F, I \rangle$
 $\langle A, C, D, H, F, I \rangle$

Anti-alignment

γ

Distance

$dist(\sigma, \gamma)$



Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
 → Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace:

σ

$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

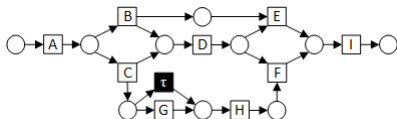
$\langle A, C, D, H, F, I \rangle$

Anti-alignment

γ

Distance

$dist(\sigma, \gamma)$



Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
 → Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace:

σ

$\langle A, B, D, E, I \rangle$
 $\langle A, C, D, G, H, F, I \rangle$
 $\langle A, C, G, D, H, F, I \rangle$
 $\langle A, C, H, D, F, I \rangle$
 $\langle A, C, D, H, F, I \rangle$

Anti-alignment

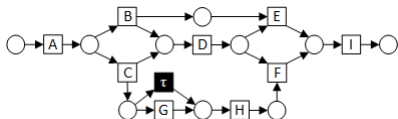
γ

$\langle A, B, D, E, I \rangle$

Distance

$dist(\sigma, \gamma)$

0



Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
 → Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace:

σ

$\langle A, B, D, E, I \rangle$
 $\langle A, C, D, G, H, F, I \rangle$
 $\langle A, C, G, D, H, F, I \rangle$
 $\langle A, C, H, D, F, I \rangle$
 $\langle A, C, D, H, F, I \rangle$

Anti-alignment

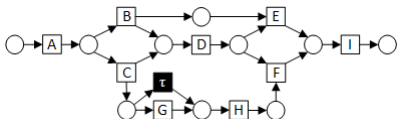
γ

$\langle A, B, D, E, I \rangle$

Distance

$dist(\sigma, \gamma)$

0



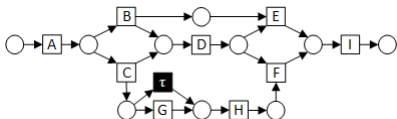
Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
 → Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace: σ	Anti-alignment γ	Distance $dist(\sigma, \gamma)$
$\langle A, B, D, E, I \rangle$	$\langle A, B, D, E, I \rangle$	0
$\langle A, C, D, G, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, G, D, H, F, I \rangle$		
$\langle A, C, H, D, F, I \rangle$		
$\langle A, C, D, H, F, I \rangle$		



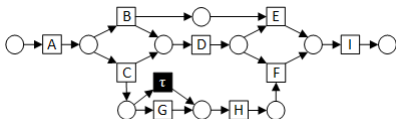
Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
 → Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace: σ	Anti-alignment γ	Distance $dist(\sigma, \gamma)$
$\langle A, B, D, E, I \rangle$	$\langle A, B, D, E, I \rangle$	0
$\langle A, C, D, G, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, G, D, H, F, I \rangle$		
$\langle A, C, H, D, F, I \rangle$		
$\langle A, C, D, H, F, I \rangle$		



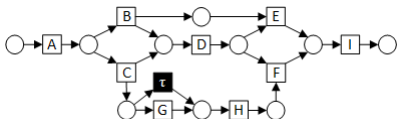
Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
→ Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace: σ	Anti-alignment γ	Distance $dist(\sigma, \gamma)$
$\langle A, B, D, E, I \rangle$	$\langle A, B, D, E, I \rangle$	0
$\langle A, C, D, G, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, G, D, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, H, D, F, I \rangle$		
$\langle A, C, D, H, F, I \rangle$		



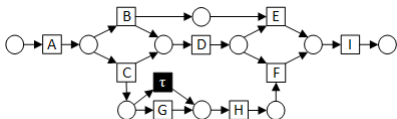
Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
→ Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace: σ	Anti-alignment γ	Distance $dist(\sigma, \gamma)$
$\langle A, B, D, E, I \rangle$	$\langle A, B, D, E, I \rangle$	0
$\langle A, C, D, G, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, G, D, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, H, D, F, I \rangle$		
$\langle A, C, D, H, F, I \rangle$		



Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
→ Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace:

σ

$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

$\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$

Anti-alignment

γ

$\langle A, B, D, E, I \rangle$

$\langle A, C, G, H, D, F, I \rangle$

$\langle A, C, G, H, D, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

Distance

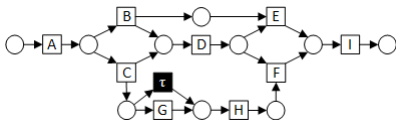
$dist(\sigma, \gamma)$

0

2

2

0



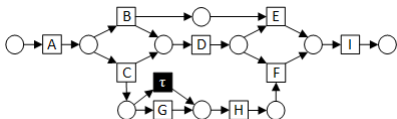
Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
→ Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace: σ	Anti-alignment γ	Distance $dist(\sigma, \gamma)$
$\langle A, B, D, E, I \rangle$	$\langle A, B, D, E, I \rangle$	0
$\langle A, C, D, G, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, G, D, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, H, D, F, I \rangle$	$\langle A, C, H, D, F, I \rangle$	0
$\langle A, C, D, H, F, I \rangle$		



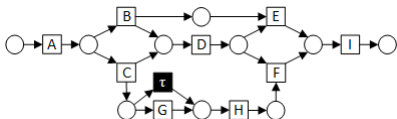
Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
→ Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace: σ	Anti-alignment γ	Distance $dist(\sigma, \gamma)$
$\langle A, B, D, E, I \rangle$	$\langle A, B, D, E, I \rangle$	0
$\langle A, C, D, G, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, G, D, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, H, D, F, I \rangle$	$\langle A, C, H, D, F, I \rangle$	0
$\langle A, C, D, H, F, I \rangle$	$\langle A, C, D, H, F, I \rangle$	0



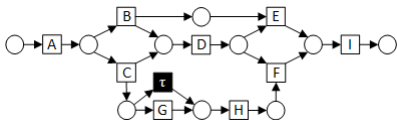
Trace-based precision

Trace-based precision

- ▶ For every trace σ in the log L , compute a best anti-alignment γ of N w.r.t. $L \setminus \{\sigma\}$.
→ Intuition: γ close to σ tends to indicate that the model is precise.
- ▶ Evaluate the distance between γ and σ .
- ▶ Average over all the σ (possibly weighted by frequencies).

Example

Log trace: σ	Anti-alignment γ	Distance $dist(\sigma, \gamma)$
$\langle A, B, D, E, I \rangle$	$\langle A, B, D, E, I \rangle$	0
$\langle A, C, D, G, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, G, D, H, F, I \rangle$	$\langle A, C, G, H, D, F, I \rangle$	2
$\langle A, C, H, D, F, I \rangle$	$\langle A, C, H, D, F, I \rangle$	0
$\langle A, C, D, H, F, I \rangle$	$\langle A, C, D, H, F, I \rangle$	0



Trace-based precision: $P_t(N, L) = \frac{1}{5} \cdot (1 + \frac{5}{7} + \frac{5}{7} + 1 + 1) = 0.886$

Summary on Precision

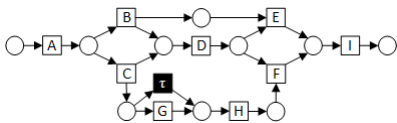
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

Log: $\langle A, C, G, D, H, F, I \rangle$

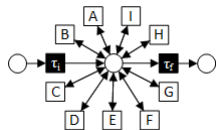
$\langle A, C, H, D, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$



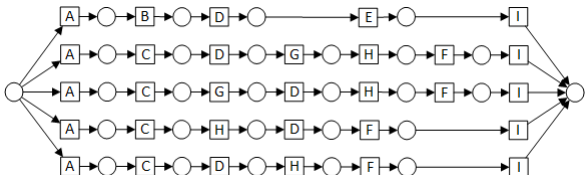
$$P_I(N_1, L) = 0.857$$

$$P_t(N_1, L) = 0.886$$



$$P_I(N_2, L) = 0$$

$$P_t(N_2, L) = 0$$



$$P_I(N_3, L) = 1$$

$$P_t(N_3, L) = 1$$

Generalization

Generalization of a model N with respect to a log L is a measure for the ability of N to predict previously unseen, *but correct*, behavior.

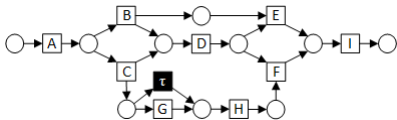
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

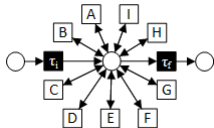
Log: $\langle A, C, G, D, H, F, I \rangle$

$\langle A, C, H, D, F, I \rangle$

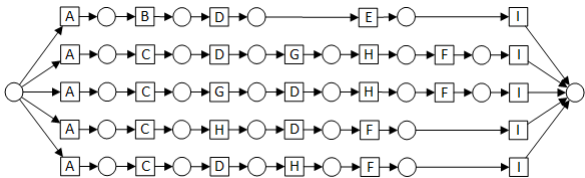
$\langle A, C, D, H, F, I \rangle$



Fairly precise
Fairly generalizing



Very imprecise
Generalizing

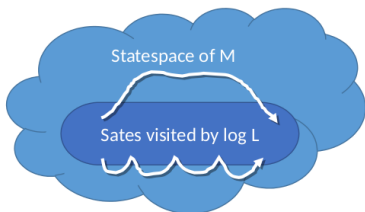


Very precise
Not generalizing

Generalization

Idea: a model generalizes if it allows for new sequences while not introducing too many new states.

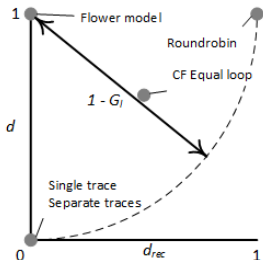
Again, we use anti-alignments!



Recovery distance

For a run γ of model N ,

- ▶ S_γ set of states of N visited by γ
- ▶ S_L : set of states of N visited by L (assuming fitting model: $L \subseteq \mathcal{L}(N)$)
- ▶ Recovery distance = $\max_{s \in S_\gamma} \min_{s' \in S_L} |\sigma|$



Generalization – Results

Like for precision, we provide log-based and trace-based metrics.

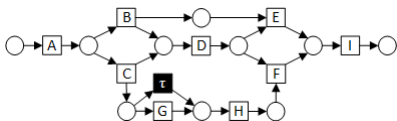
$\langle A, B, D, E, I \rangle$

$\langle A, C, D, G, H, F, I \rangle$

Log: $\langle A, C, G, D, H, F, I \rangle$

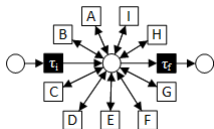
$\langle A, C, H, D, F, I \rangle$

$\langle A, C, D, H, F, I \rangle$



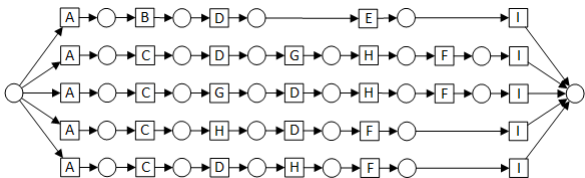
$$G_l(N_1, L) = 0.143$$

$$G_t(N_1, L) = 0.270$$



$$G_l(N_2, L) = 1$$

$$G_t(N_2, L) = 1$$

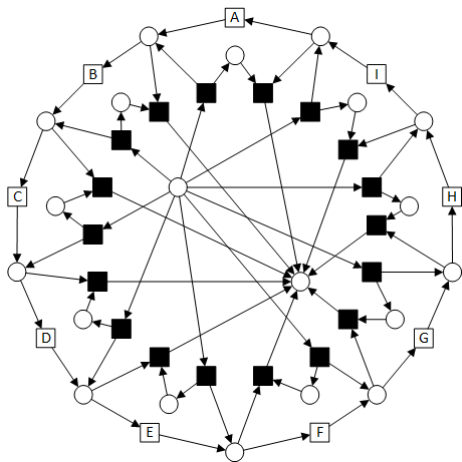


$$G_l(N_3, L) = 0$$

$$G_t(N_3, L) = 0$$

Unfair Generalization

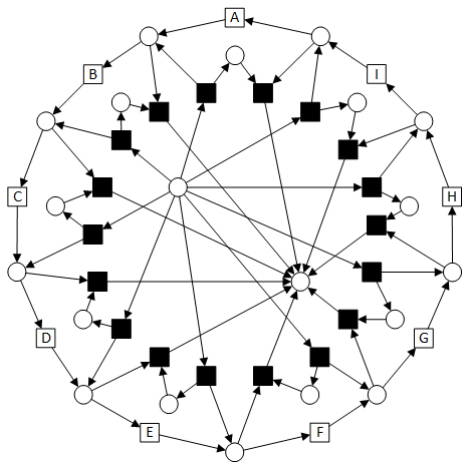
$\langle A, B, D, E, I \rangle$
 $\langle A, C, D, G, H, F, I \rangle$
 Log: $\langle A, C, G, D, H, F, I \rangle$
 $\langle A, C, H, D, F, I \rangle$
 $\langle A, C, D, H, F, I \rangle$



Unfair Generalization

$\langle A, B, D, E, I \rangle$
 $\langle A, C, D, G, H, F, I \rangle$
 Log: $\langle A, C, G, D, H, F, I \rangle$
 $\langle A, C, H, D, F, I \rangle$
 $\langle A, C, D, H, F, I \rangle$

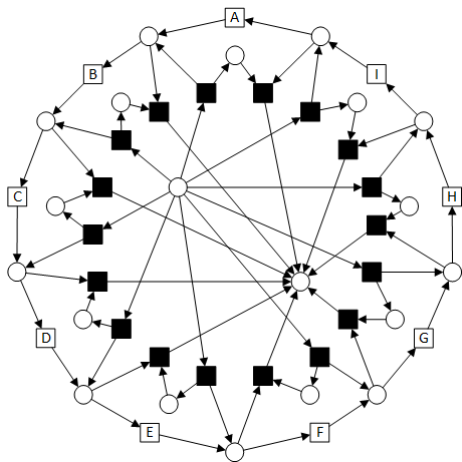
- ▶ many new behaviours
- ▶ ... but many new states



Unfair Generalization

Log: $\langle A, B, D, E, I \rangle$
 $\langle A, C, D, G, H, F, I \rangle$
 $\langle A, C, G, D, H, F, I \rangle$
 $\langle A, C, H, D, F, I \rangle$
 $\langle A, C, D, H, F, I \rangle$

- ▶ many new behaviours
- ▶ ... but many new states



$$G_l(N_4, L) = 0$$

$$G_t(N_4, L) = 0$$

Conclusion

- ▶ **Anti-alignments**: run of the model which maximizes its distance to the observed traces
- ▶ New metric for **precision** in process mining
 - ▶ monotonic w.r.t. new observations
- ▶ Also used for measuring generalization

- ▶ Implementations
 - ▶ DARKSIDER (using SAT encoding)
`www.lsv.ens-cachan.fr/~chatain/darksider`
 - ▶ uses Hamming distance or variations of it
 - ▶ Also available in ProM
`www.promtools.org/`

- ▶ Wanted!
 - ▶ Efficient way to compute anti-alignments for edit distance