Leveraging Frequencies in Event Data

a Pledge for Stochastic Process Mining

Sander Leemans

Algorithms & Theories for the Analysis of Event Data 2022

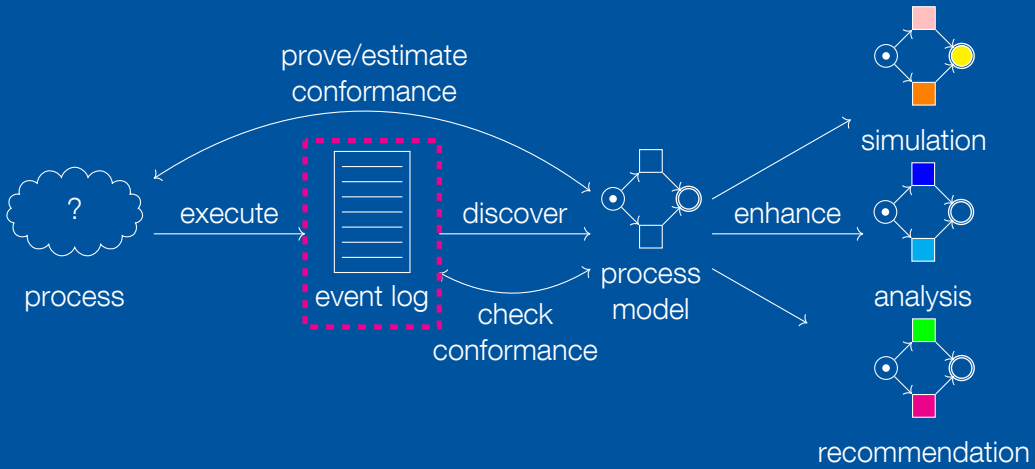
About me

- Bachelor & master: computer science and engineering
- Master: science education and communication
- PhD: robust process mining with guarantees
- Postdoc/lecturer/senior lecturer QUT, Australia
- Professor of BPM, RWTH Aachen, Germany

Processes

process

Process Mining



Event Log

case ID	activity	timestamp	resource	amount	vehicleClass	…
135	create fine	09:30	A	$39	A	
135	send fine	09:39	B	$39	A	
135	insert notification	09:40	A	$39	A	
136	create fine	10:45	A	$185	C	
136	payment	10:50	C	$185	C	
⋮						

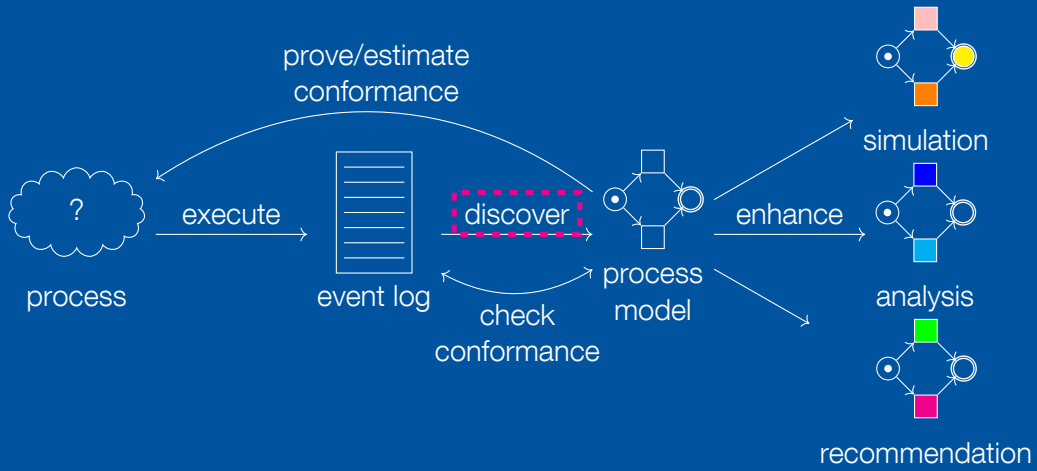
↓

⟨create fine, send fine, insert notification⟩

⟨create fine, payment⟩

…

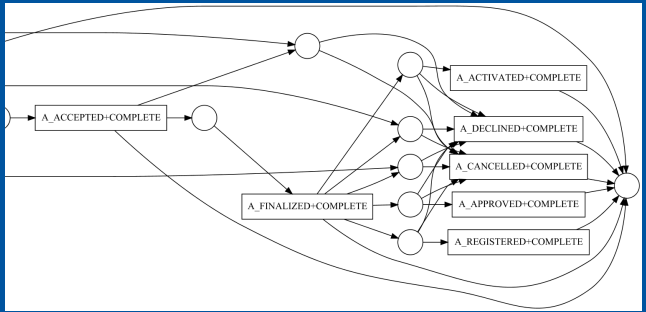
Process mining



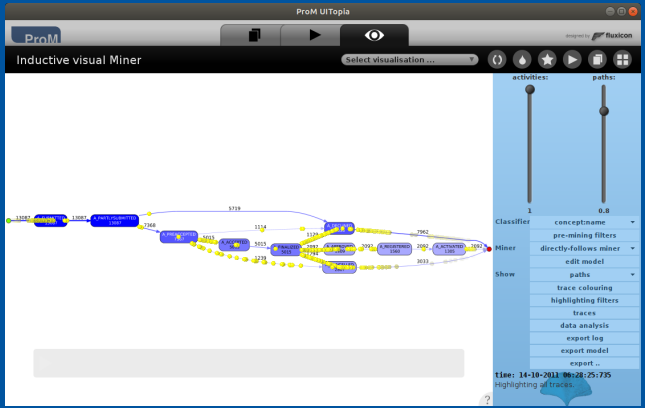
Process discovery

- α, α^{++}, $\alpha^{\$}$, …
- ILP
- Heuristic Miner
- Flexible Heuristics Miner
- Genetic Miner
- Split Miner
- Evolutionary Tree Miner
- Directly Follows Model Miner
- Indulpet Miner
- Inductive Miner - infrequent
- Inductive Miner - all operators
- Inductive Miner - …

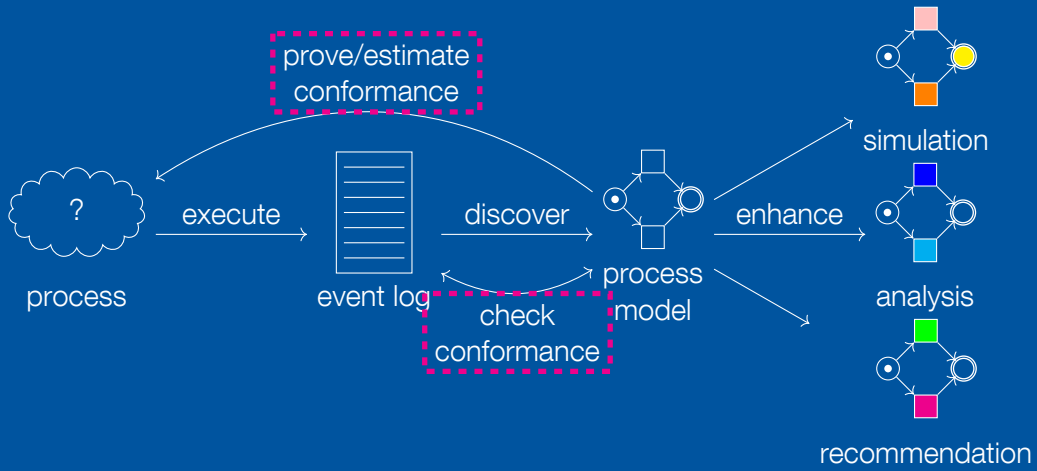
sound



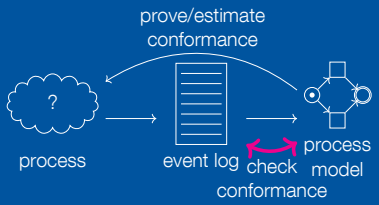
visual Miner

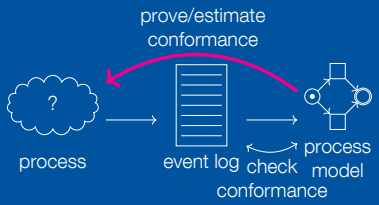


Process mining



Model quality





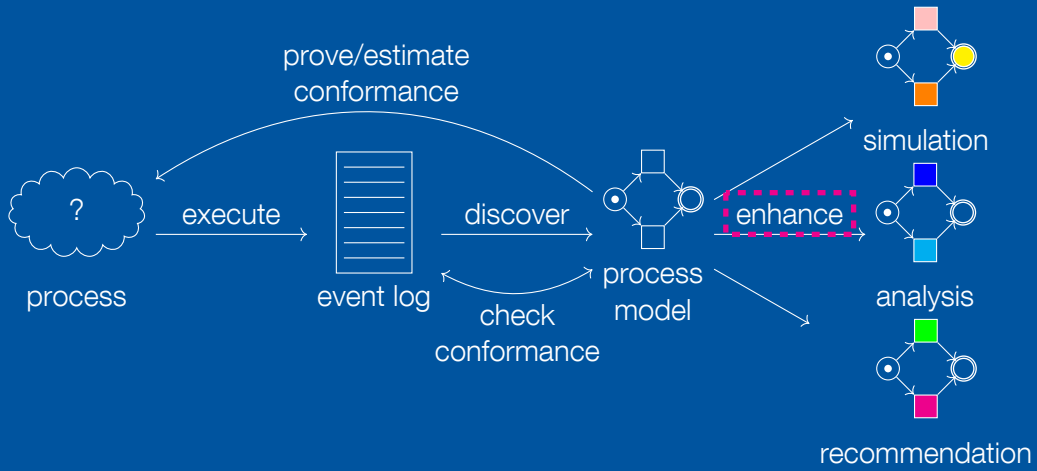
Alignments

demo

Discovered model and process language equivalent if:

- process $\in$ class
- event log without noise
- event log complete

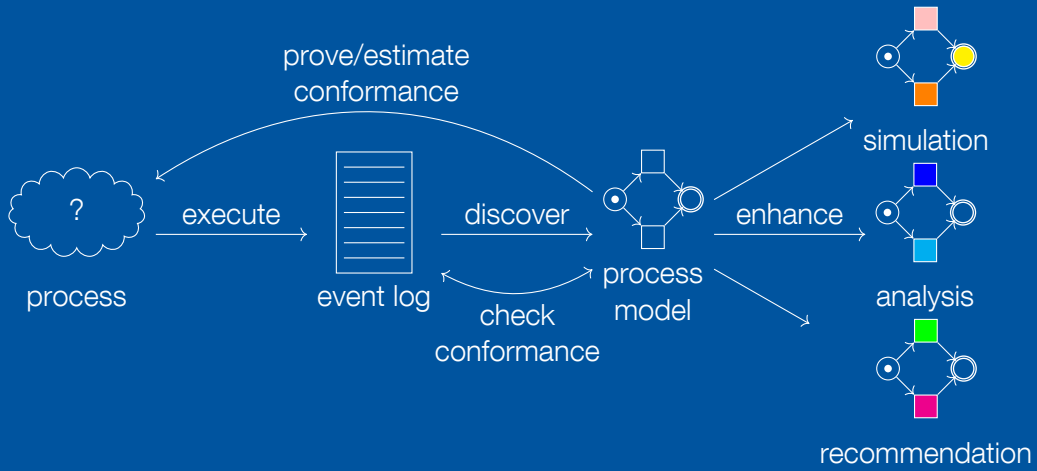
Process mining



Enhancement

demo

Process mining

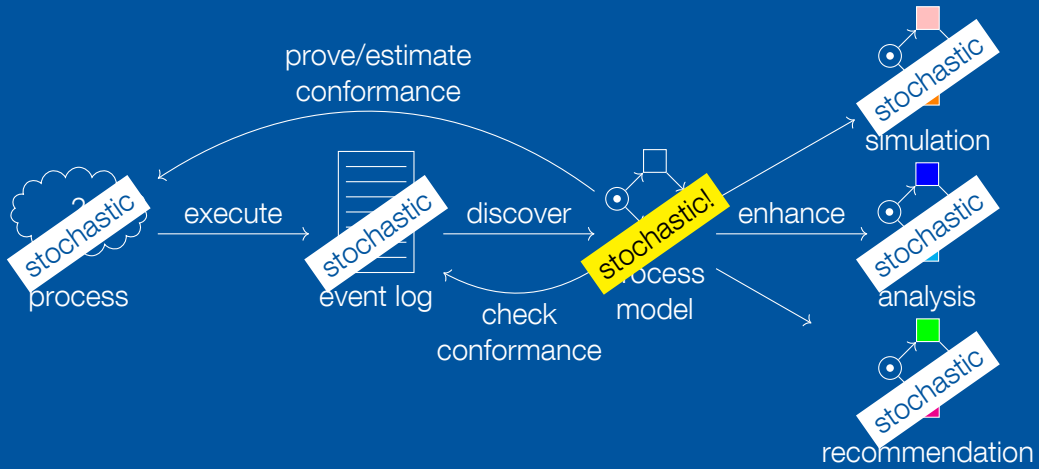


Frequencies

$$\begin{aligned} L_1 = [&\langle \text{register}, \text{check}, \text{accept}\rangle^{10000}, \\ &\langle \text{register}, \text{check}, \text{reject}\rangle^{10000}, \\ &\langle \text{register}, \text{accept}\rangle^{1}, \\ &\langle \text{accept}, \text{register}\rangle^{1}] \end{aligned}$$

$$\begin{aligned} L_2 = [&\langle \text{register}, \text{check}, \text{accept}\rangle^{9500}, \\ &\langle \text{register}, \text{check}, \text{reject}\rangle^{9500} \\ &\langle \text{register}, \text{accept}\rangle^{1002}] \end{aligned}$$

Frequencies in process mining

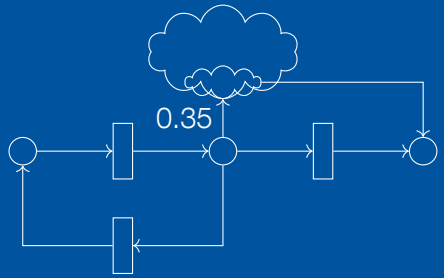


Simulation & recommendation

ad-hoc

- Derived from log
- Testing on training logs
- No idea of quality
- No reasoning on quality
- Adjust stochastic perspective
- Verify stochastic perspective

Precision

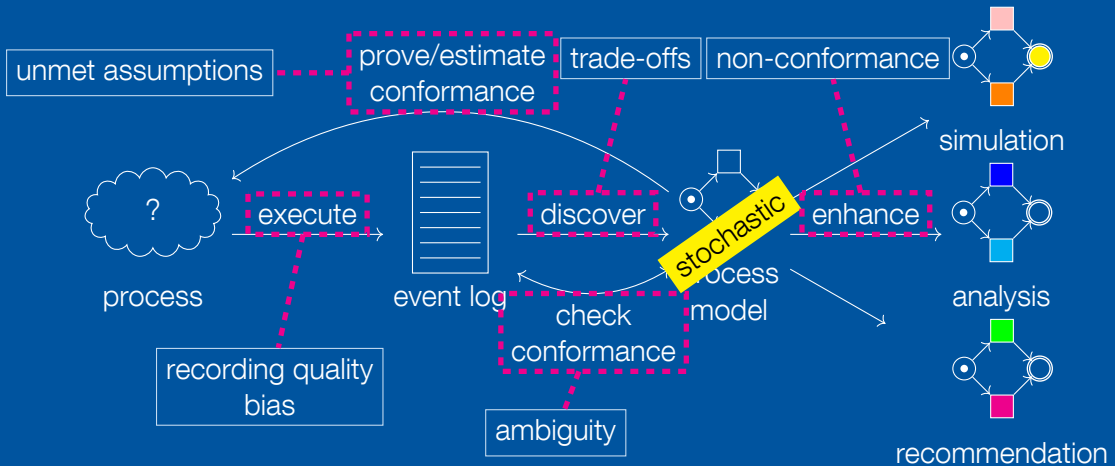


Part of the model's behaviour captured in the log

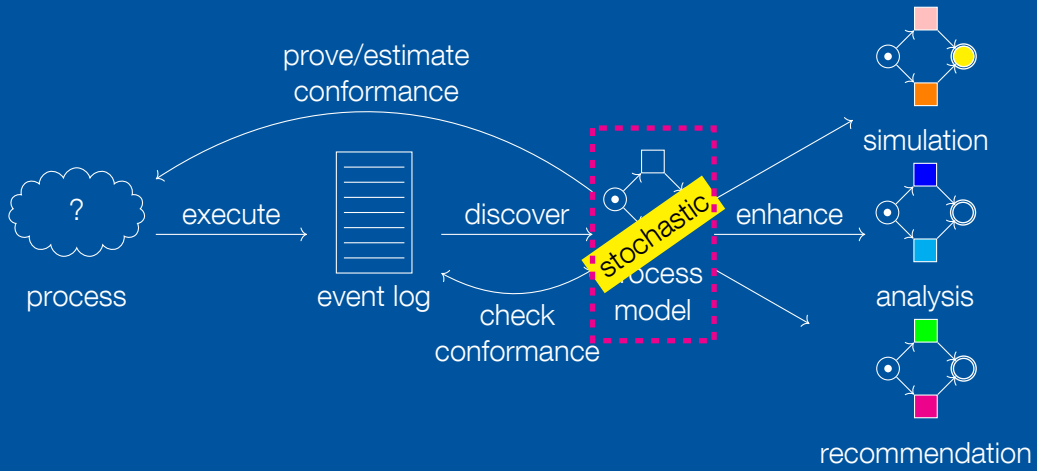
$$\frac{\text{model} \cap \text{log}}{\text{model}}$$

- Token replay
- Visited edges in model
- Visited edges in state space
- Visited edges in history-based state space
- Entropy
- Earth-movers' distance

Reliability of conclusions

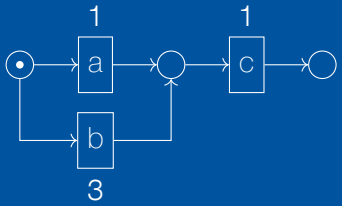


Stochastic process mining



Stochastic process models

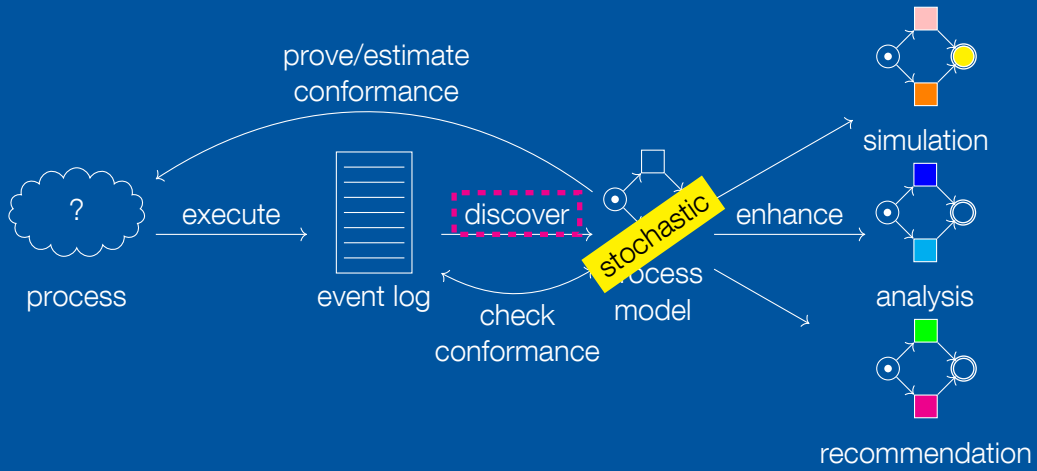
Stochastic Petri net:



Stochastic language:

$$[\langle a, c\rangle^{0.25}$$
$$\langle b, c\rangle^{0.75}]$$

Stochastic process mining



Stochastic process discovery

- GDT_SPN Miner
- Weight estimators
- Toothpaste Miner
- Stochastic Declare

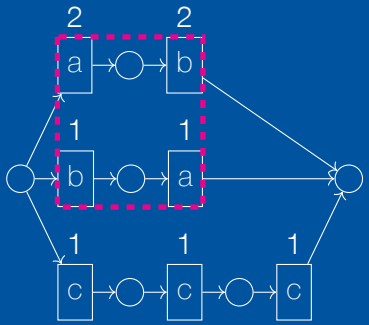
Toothpaste Miner

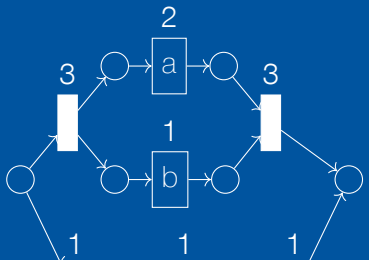
Step 1: construct trace model
Step 2: apply reduction rules

- e.g.: concurrent reduction
- e.g.: probabilistic loop roll

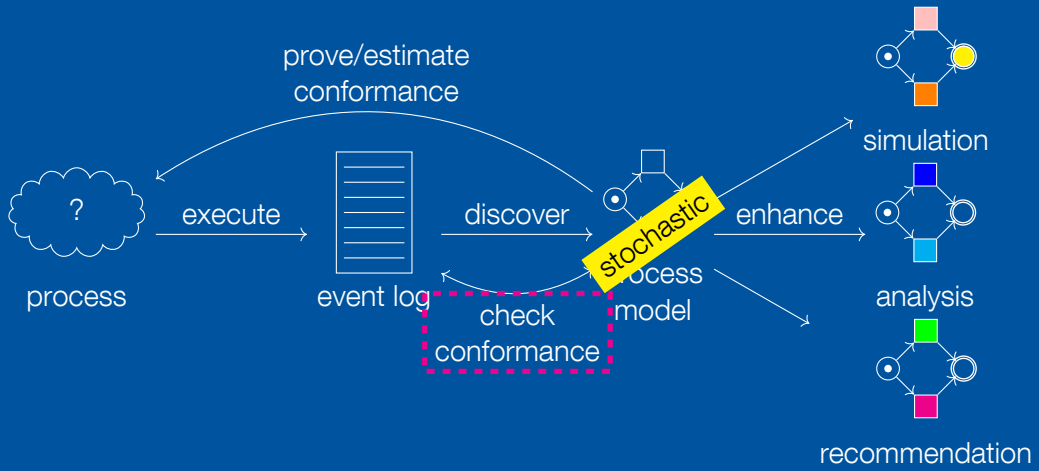
$$[\langle a, b\rangle^2, \langle b, a\rangle, \langle c, c, c\rangle]$$

Internally: probabilistic process trees





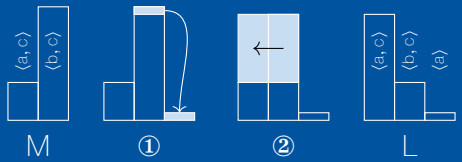
Stochastic conformance checking



Earth Movers' Stochastic Conformance

$$M = [\langle a, c\rangle^{0.25}, \langle b, c\rangle^{0.75}]$$

$$L = [\langle a, c\rangle^{0.70}, \langle b, c\rangle^{0.25}, \langle a\rangle^{0.05}]$$



Earth Movers' Stochastic Conformance:
1 - minimum cost
$1 - 0.275 = 0.725$

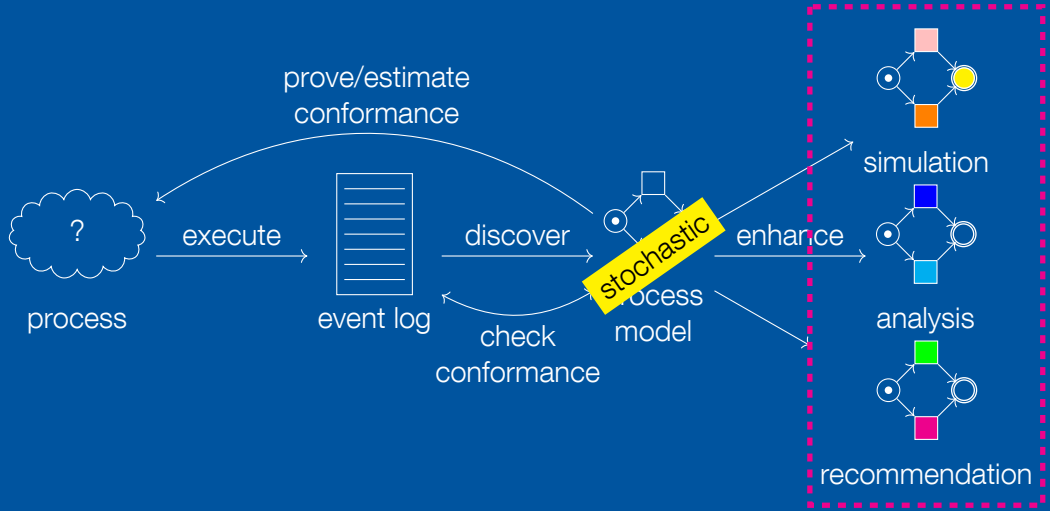
Cost:
①: 0.05 * $\delta(\langle b, c\rangle, \langle a\rangle)$
②: 0.45 * $\delta(\langle b, c\rangle, \langle a, c\rangle)$

δ: normalised Levenshtein trace distance:
①: 0.05 * 1
②: 0.45 * 0.5

Complexity: polynomial
however
future work: loop unfolding & concurrency

[1] Leemans, van der Aalst, Brockhoff, Polyvyanyy. Stochastic Process Mining: Earth Movers' Stochastic Conformance. Information Systems

Stochastic process mining



Cohort Analysis: trace attributes

case ID	activity	timestamp	resource	amount	vehicleClass	…
135	create fine	09:30	A	$39	A	
135	send fine	09:39	B	$39	A	
135	insert notification	09:40	A	$39	A	
136	create fine	10:45	A	$185	C	
136	payment	10:50	C	$185	C	
⋮						

↓

event log
with trace attributes

⟨create fine, send fine, insert notification⟩ amount $39 vehicleClass A

⟨create fine, payment⟩ amount $185 vehicleClass C

…

Cohort Analysis

1: elicit features	2: split log	3: compute EMSC
vehicleClass = A		0.523
vehicleClass = C		0.647
vehicleClass missing		0.985
amount ≥ $150		0.125
…	…	…

event log
with trace attributes

Which trace attributes influence the process?
Which filter should I apply?
Which cohorts are interesting?

The trace attribute vehicleClass coincides with large differences in processing.
The amount is weakly associated with a difference in processing.

[1]Leemans, Shabaninejad, Goel, Khosravi, Sadiq, Wynn. Identifying Cohorts: Recommending Drill-Downs Based on Differences in Behaviour for Process Mining. ER 2020

Causal analysis

Ice cream causes drowning: it has been shown that the more ice creams are sold, the more people drown.

Correlation $\neq$ causation

The choice between a and b causally influences the choice between x and y.

We can increase x by increasing a.

[1]Leemans, Tax. Causal Reasoning over Control-Flow Decisions in Process Models. CAiSE 2022

Even more goodies

$$L_1 = [\langle \text{register}, \text{check}, \text{accept}\rangle^{10000}, \langle \text{register}, \text{check}, \text{reject}\rangle^{10000}, \langle \text{register}, \text{accept}\rangle^{1}, \langle \text{accept}, \text{register}\rangle^{1}]$$

$$L_2 = [\langle \text{register}, \text{check}, \text{accept}\rangle^{9500}, \langle \text{register}, \text{check}, \text{reject}\rangle^{9500} \langle \text{register}, \text{accept}\rangle^{1002}]$$

Relation process - outcomes
Stochastic process drift
Anomaly detection
Stochastic + data models

You have been watching…

A pledge for stochastic process mining

- Simply there
- Reliability of conclusions

Goodies

- Cohort analysis
- Causal analysis

Stochastic process mining techniques

- Stochastic process discovery (Toothpaste Miner)
- Stochastic conformance checking (Earth Movers' Stochastic Conformance checking)

Sander Leemans
s.leemans@bpm.rwth-aachen.de
http://leemans.ch

RWTH AACHEN UNIVERSITY