Mine Your Own Business

Turning (Big) Data into Real Value using Process Mining

prof.dr.ir. Wil van der Aalst

Central and Eastern European Software Engineering Conference in Russia (CEE-SECR 2013), Moscow, October 25th 2013
Scooby Doo, Where Are You!

In:

Mine Your Own Business

© 1969 Hanna-Barbera Productions, Inc.

Season 1, Episode 4 (1969)
How to get started?

Evidence-based BPM

Process Mining: The missing link

Big (Event) Data

Process discovery

Aligning reality and model

How to get started?
Why do (larger) organizations have software?
It is always about supporting or improving business processes
Business process problem or IT problem?
BPM efforts focus on models
• enormous investments in process models
• large collections of "dead" process models
• not taken seriously, unrelated to reality

Models should be:
- descriptive,
- predictive, and/or
- prescriptive
REALITY
Worst game ever.
How to get started?

Evidence-based BPM and Auditing

Process Mining: The missing link

Big (Event) Data

Process discovery

Aligning reality and model

How to get started?
Motivation: Increasing awareness of the value of (Big) Data

• "In God we trust. All others must bring data" (William Edwards Deming, statistician),
• "Data is a precious thing and will last longer than the systems themselves" (Tim Berners-Lee),
• "Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital" (Aaron Levenstein, statistician),
• "Every 2 days we create as much information as we did up to 2003" (Eric Schmidt, Google CEO, August 4, 2010).
The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

In 10 years we will have 50 times as much data! (IDC)
90% of the world's data was created in the last two years.
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT
Big Data ?
Big ... or fast and efficient?
Process-awareness is an essential but often forgotten ingredient when converting big data into real value.
How to get started?

Evidence-based BPM and Auditing

Process Mining: The missing link

Aligning reality and model

Big (Event) Data

Process discovery

How to get started?
process mining

(data mining, machine learning, business intelligence)

process model analysis
(simulation, verification, optimization, gaming, etc.)

performance-oriented questions, problems and solutions

compliance-oriented questions, problems and solutions

data-oriented analysis
(data mining, machine learning, business intelligence)
小草对您微微笑
请您把路绕一绕
KEEP OFF GRASS
let's play
Play-Out

process model

event log
Play-Out (Classical use of models)
Let’s not worry about syntax (there is difference between analysis and presentation)
Play-In

event log

process model
Play-In

A B C D  A E D  A E D
A C B D  A B C D  A C B D
A C B D  A C B D
A B C D

start
A
B
E
C
D
end

p1
p2
p3
p4
Example Process Discovery
(Vestia, Dutch housing agency, 208 cases, 5987 events)
Example Process Discovery
(ASML, test process lithography systems, 154966 events)
Example Process Discovery
(AMC, 627 gynecological oncology patients, 24331 events)
Replay

- event log
- process model
- extended model showing times, frequencies, etc.
- diagnostics
- predictions
- recommendations
Replay
Replay can detect problems

Problem! token left behind

Problem! missing token
Conformance Checking

(WOZ objections Dutch municipality, 745 objections, 9583 events, f = 0.988)
Replay can extract timing information
Performance Analysis Using Replay

(WOZ objections Dutch municipality, 745 objections, 9583 event, f = 0.988)
Disco Demo
Models are like the glasses required to see and understand event data!
How to get started?

Evidence-based BPM and Auditing

Process Mining: The missing link

Aligning reality and model

Big (Event) Data

Process discovery

How to get started?
A language is **learnable in the limit** if there exists a perfect child that generates only finitely many hypotheses.
Learning is not easy …

- Even simple language classes like regular languages are not learnable in the limit.
- Many settings: the evil or well-behaving mothers, with or without negative examples, frequencies, etc.

Classical approaches (before 1998) did not consider concurrency and definitely not end-to-end business process models.

Hence \( \cong \) trace in event log

language \( \cong \) process model
Process discovery algorithms
(small selection)

- automata-based learning
- heuristic mining
- genetic mining
- stochastic task graphs
- ETM genetic algorithm
- fuzzy mining
- mining block structures
  - α algorithm
  - α# algorithm
  - α++ algorithm

- distributed genetic mining
- language-based regions
- state-based regions
- LTL mining
- Inductive Miner (infrequent)
- neural networks
- hidden Markov models
- multi-phase mining
- conformal process graph
- partial-order based mining
- ILP mining
Mine your Map
models are like maps, their usefulness is determined by the intended use, i.e., there is not a single "perfect map".
How to get started?

Evidence-based BPM and Auditing

Process Mining: The missing link

Big (Event) Data

Process discovery

Aligning reality and model

How to get started?
Conformance Checking

- An activity that should not happen happened
- An activity that should happen did not happen
- An activity was executed by the wrong person
- An activity was executed too late
- Two activities were swapped
Alignments are essential!

- conformance checking to diagnose deviations
- squeezing reality into the model to do model-based analysis
Loops of “W_Completeren aanvraag” and “W_Nabellen offertes” are often performed

“O_DECLINED” and “W_Wijzigen contractgegevens” are often skipped

Many moves on log of “O_CANCELLED”, “O_CREATED”, “O_SELECTED”, “O_SENT” occurred with the same frequency value (i.e. 60) before parallel branch

Many moves on log of “W_Afhandelen leads” (> 2200 times) occurred in the end of traces

Work of Arya Adriansyah (Replay project)
Loops of “W_Completeren aanvraag” and “W_Nabellen offertes” are often performed in the end of traces. Loops of “W_Completeren aanvraag” and “W_Nabellen offertes” are often performed.

Many moves on log of “O_CANCELLED” and “A_CANCELLED” occurred with the same frequency value (i.e., 60) before parallel branch leads (>2200 times) occurred in the end of traces.

Synchronous moves of “Completeren aanvraag” Move on log of “Completeren aanvraag”

Many moves on log of “O_DECLINED” and “W_Wijzigen contractgegevens” are often skipped.

Move on log of “O_DECLINED” and “W_Wijzigen contractgegevens” are often skipped.

Synchronous moves of “Completeren aanvraag”

Move on model towards end of traces

Moves on model towards end of traces

Many moves on log of “O_DECLINED” and “W_Wijzigen contractgegevens” are often skipped.

Move on log of “O_DECLINED” and “W_Wijzigen contractgegevens” are often skipped.

Synchronous moves of “Completeren aanvraag”

Move on model towards end of traces

Moves on model towards end of traces

Auditor’s toolbox
“O_ACCEPTED” has average sojourn time of 27.07 minutes, while “AREGISTERED”, “A_ACTIVATED”, and “A_APPROVED” have average sojourn time of 29.56 minutes.

Activity “W_Wijzigen contractgegevens” is the bottleneck, but it occurred rarely (only 4 times).

The average waiting time for the input place of “W_Nabellen offertes+START” is very long (2.83 days) compared to the average waiting time of other places.
traffic jams are projected on map
Demand TomTom!
Do not settle for restrictive information systems and static process models

predict: when will I be home

recommend: turn right

adapt: use real-time traffic information
Evidence-based BPM and Auditing

Process Mining: The missing link

Big (Event) Data

Process discovery

Aligning reality and model

How to get started?
How to get started?
Collect data: Events are everywhere!

- **Minimal requirement:** events referring to an activity name and a process instance.
- **Good to have:** timestamps, resource information, additional data elements.
- **Challenges:** scoping and sometimes correlation.

Databases, ERP systems (SAP etc.), WFM/BPM logs, message logs, audit trails, etc.
Get at a process mining tool, e.g., ProM

600+ plug-ins available covering the whole process mining spectrum

Download from: www.processmining.org
Commercial process mining tools

- **Disco** (Fluxicon)
- **Perceptive Process Mining** (before Futura Reflect and BPM|one)
- ARIS Process Performance Manager
- QPR ProcessAnalyzer
- Celonis Discovery
- Interstage Process Discovery (Fujitsu)
- Discovery Analyst (StereoLOGIC)
- XMAnalyzer (XMPro)
- …

Example of a dedicated process mining consulting firm: **ProcessGold AG**.
Questions:

- What kind problems would you like to address (cost, time, risk, compliance, service, etc.)?
- Related to discovery, conformance, enhancement?
- Iterative process: can be “curiosity driven” initially.
Join our expedition: Mine your processes!

- **Process mining**
- **Data-oriented analysis**
  - (Data mining, machine learning, business intelligence)
- **Process model analysis**
  - (Simulation, verification, etc.)
- **Performance-oriented questions, problems and solutions**
- **Compliance-oriented questions, problems and solutions**

**Data-oriented analysis**
(data mining, machine learning, business intelligence)