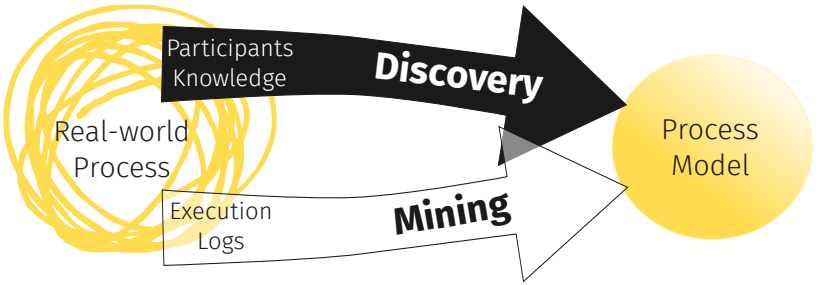
Università della Svizzera italiana | Faculty of Informatics

Business Process Modeling, Management and Mining

Business Process Mining

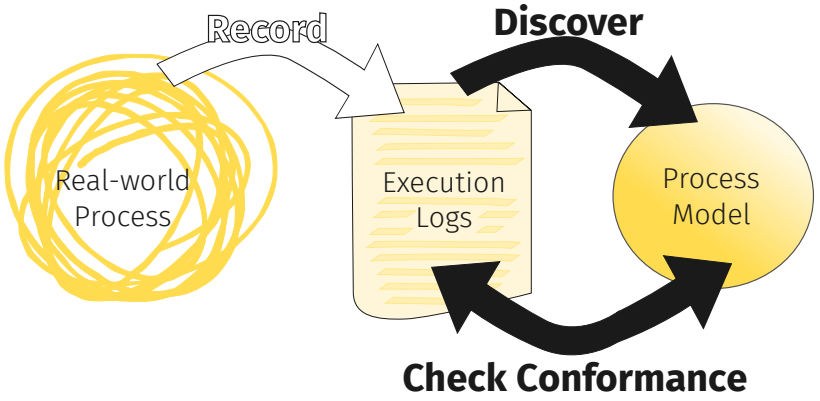
Prof. Cesare Pautasso
http://www.pautasso.info
cesare.pautasso@usi.ch
@pautasso



Discovery Gap

- Reality vs. idealized process models
- Participants provide only incomplete and inconsistent knowledge
- Focus on happy path and planned exceptions
- Both "as-is" and "to-be" models are disconnected from what actually happens in the organization

Process Mining Steps



Logging

- **Time**: When did an event happen?
- **Resource**: Who was involved in the event?
- **Activity**: What was the event about?
- **Case**: Which process instance?

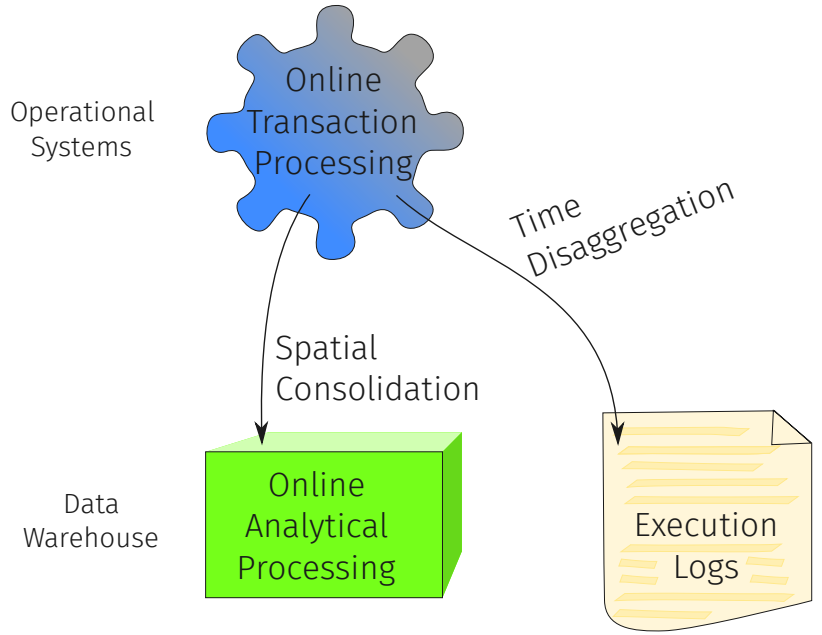
Capture enough information that can be trusted to be fed into the mining algorithm

Log Granularity

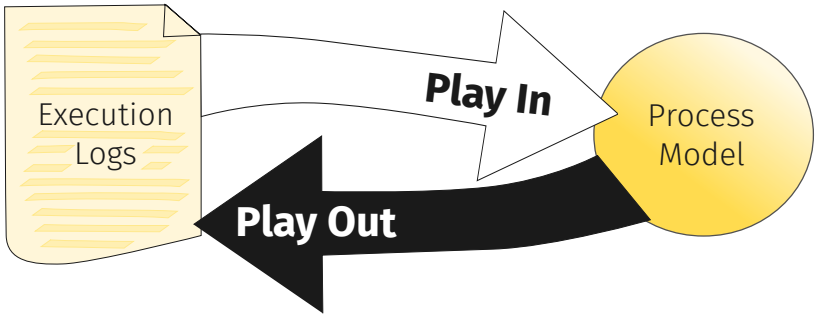
- **Coarse-grained**: one event per activity (instantaneous tasks)
- **Fine-grained**: one event for each state transition of each activity (ready, waiting, started, completed)

Where to find events?

- As they happen:
  - Intercept and record every action and interaction
  - Force people to report task completions
- After the fact:
  - Reconstruct events from historical databases
  - Look for timestamps in schemas and map them to an event



Extract the process model by mining the logs



Monitor the process execution

Mining Outcome

1. Process Model:
control flow with partial ordering of events
2. Social Network:
based on the frequency of handovers
3. Decision rules:
branch probability based on data/known state
4. Performance:
activity and process duration statistics, resource utilization

Process Mining Algorithm

Assumptions

- Every event belongs to **one** process **instance**
(how to find the corresponding instance?)
- Every logged event carries a **timestamp**
(the time of the event or the time when the event was logged?)
- The log contains only **completed** processes
(how to filter incomplete ones?)
- The log contains **all** possible events and event pairs
(behavioral completeness)

Basic Mining Algorithm

1. Abstract event log: identify event sequences
2. Determine partial order of event pairs
3. Classify causal, parallel and non-succession relationships
4. Build the footprint matrix
5. Reconstruct the control flow

Order Relationship

$a > b$

task a is directly followed by b

Causality Relationship

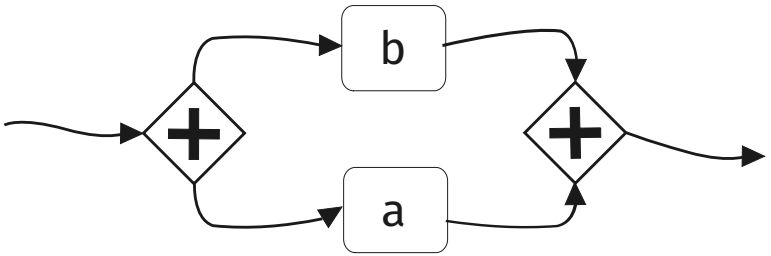
$a \rightarrow b := a > b \wedge !(b > a)$

task a is directly followed by task b,
and task a never directly follows task b

Parallelism Relationship

$a \parallel b \;:=\; a>b \;\wedge\; b>a$

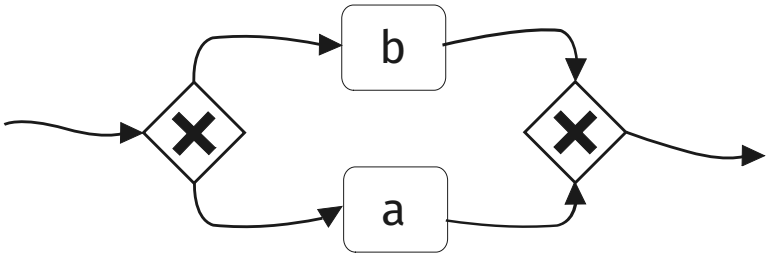
task a both directly follows and directly precedes task b



No-Direct-Succession Relationship

$a \# b \;:=\; !(a>b) \;\wedge\; !(b>a)$

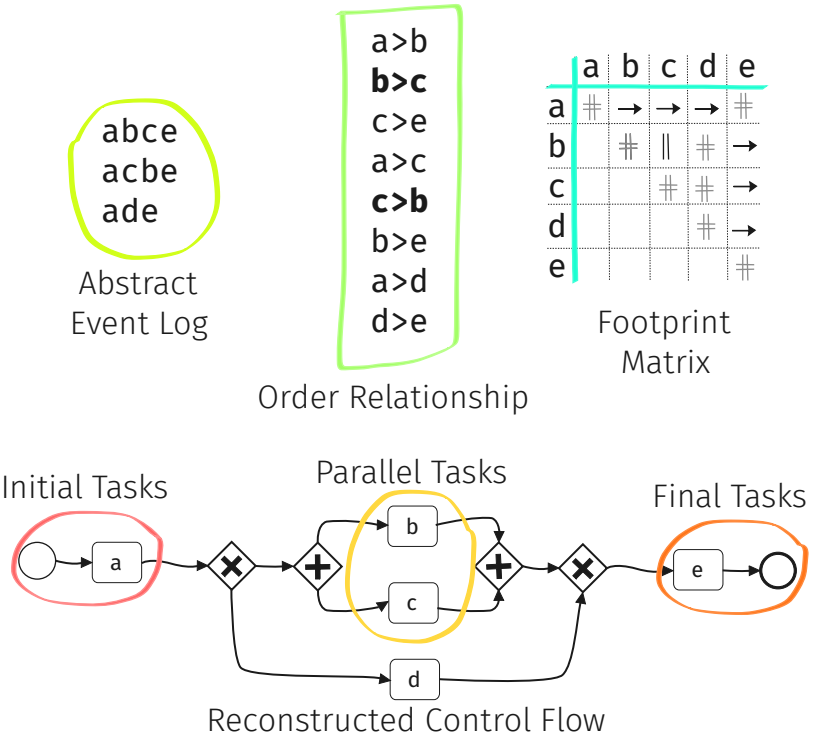
neither task a directly follows b, nor task b directly follows task a



Footprint Matrix

Complete classification of all possible event pairs according to the causality, parallelism and no-direct-succession relationships

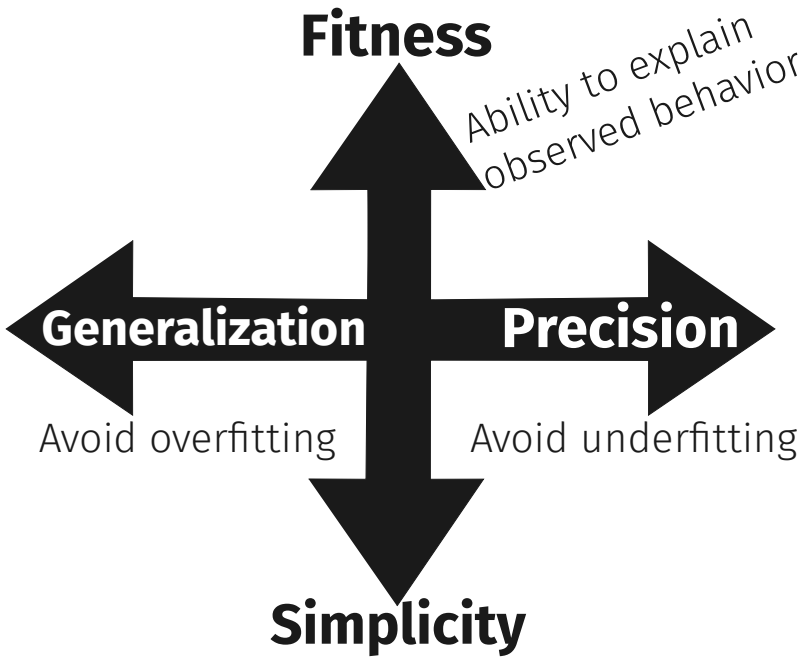
Input for the control flow reconstruction algorithm

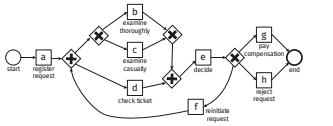


Challenges

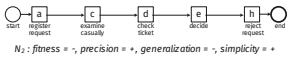
How to deal with:

- Incomplete input event log (only positive examples)
- Unclear separation between regular and exceptional behavior
- Large amounts of noisy data
- Underfitting, fitting or overfitting?

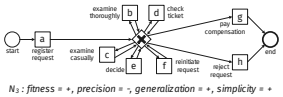




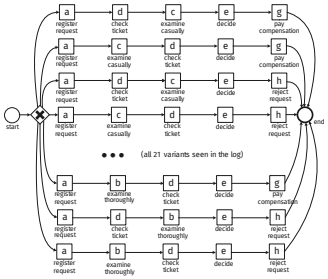
N_1 : *fitness* = +, *precision* = +, *generalization* = +, *simplicity* = +



N_2 : *fitness* = -, *precision* = +, *generalization* = -, *simplicity* = +



N_3 : *fitness* = +, *precision* = -, *generalization* = +, *simplicity* = +



N_4 : *fitness* = +, *precision* = +, *generalization* = -, *simplicity* = -

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

a	register request
b	examine thoroughly
c	examine casually
d	check ticket
e	decide
f	reinitiate request
g	pay compensation
h	reject request

References

- Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo Reijers, **Fundamentals of Business Process Management**, Chapter 10, Springer, 2013, ISBN 978-3-642-33142-8
- Wil M. P. van der Aalst, **Process Mining**: Discovery, Conformance and Enhancement of Business Processes, Springer 2011
- Jan Dirk Van-der-Burg, Desire Lines/Olifantenpaadjes (http://www.olifantenpaadjes.nl)