PV226: Process Mining seminar

Martin Macák

macak@mail.muni.cz

September 26, 2019

LAB OF SOFTWARE ARCHITECTURES
AND INFORMATION SYSTEMS

FACULTY OF INFORMATICS
MASARYK UNIVERSITY, BRNO

Outline

1. Basic overview of Process Mining
2. Course introduction

Process mining

- Process-centric data analysis

- What really happened in the past?
- Why did it happen?
- What is likely to happen in the future?
- When and why do people deviate?
- How to redesign a process to improve it?
- ...

Process mining

- Typically working with event logs which represent processes
- These logs have to contain cases (sequences of events)

```
Martin;order_start
Martin;select_hamburger
Martin;choose_card_payment
Martin;confirm_order
Martin;order_end
```

lasaris

Process mining

- Each event has:
  - caseId
  - activity
  - timestamp (optional)
  - resource (optional)
  - other data (optional)

```
1;order_accept;Dec 2, 2017 10:30:58 AM;Peter;21
1;order_cooked;Dec 2, 2017 10:39:24 AM;Victor;24
1;order_delivered;Dec 2, 2017 11:12:37 AM;Emma;19
```

lasaris

Process mining

- Sometimes, the mapping is not clear

```
1;order_accept;Dec 2, 2017 10:30:58 AM;Peter;21
1;order_cooked;Dec 2, 2017 10:39:24 AM;Victor;24
2;order_accept;Dec 2, 2017 10:40:21 AM;Peter;21
3;order_accept;Dec 2, 2017 10:42:19 AM;Greg;34
1;order_delivered;Dec 2, 2017 11:12:37 AM;Emma;19
2;order_cooked;Dec 2, 2017 11:17:04 AM;Victor;24
2;order_delivered;Dec 2, 2017 11:24:00 AM;Peter;21
```

- For example, the name of the worker can be:
  - resource
  - activity
  - caseId

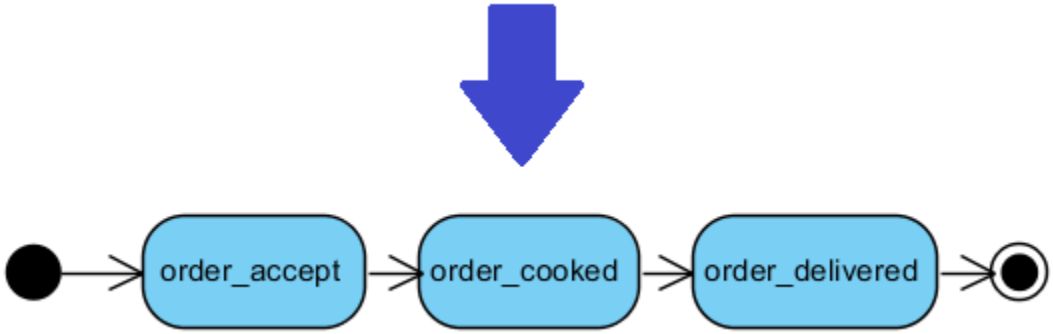
lasaris

Analysis of the past

1. Process discovery techniques
   - From the event log, we create a model that represents how the process was executed in reality
   - Model can be represented as a petri net, activity diagram, BPMN diagram, …
2. Conformance checking techniques
   - We can check the deviations from the created model in historic data

Process discovery

```
1;order_accept;Dec 2, 2017 10:30:58 AM;Peter;21
1;order_cooked;Dec 2, 2017 10:39:24 AM;Victor;24
2;order_accept;Dec 2, 2017 10:40:21 AM;Peter;21
3;order_accept;Dec 2, 2017 10:42:19 AM;Greg;34
1;order_delivered;Dec 2, 2017 11:12:37 AM;Emma;19
2;order_cooked;Dec 2, 2017 11:17:04 AM;Victor;24
2;order_delivered;Dec 2, 2017 11:24:00 AM;Peter;21
```



lasaris

Process discovery challenges

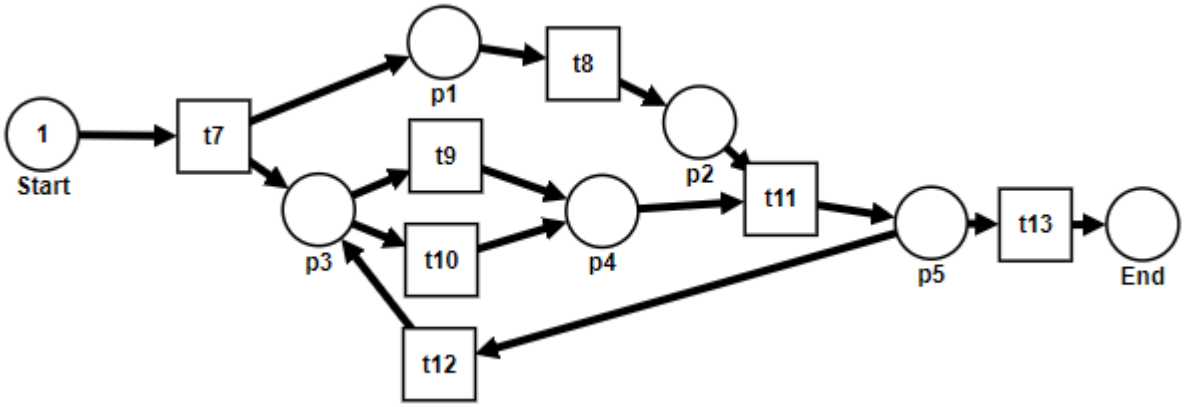
- Concurrency
- Loops
- Noisy behavior
- No negative examples in the log
- Too many allowed behaviors

Process discovery activities

- Explore processes at run-time
- Discover process models
- Compare the model of desired behavior with the model of reality
- Check the deviations in historic data
- Promote the model that shows the desired behavior

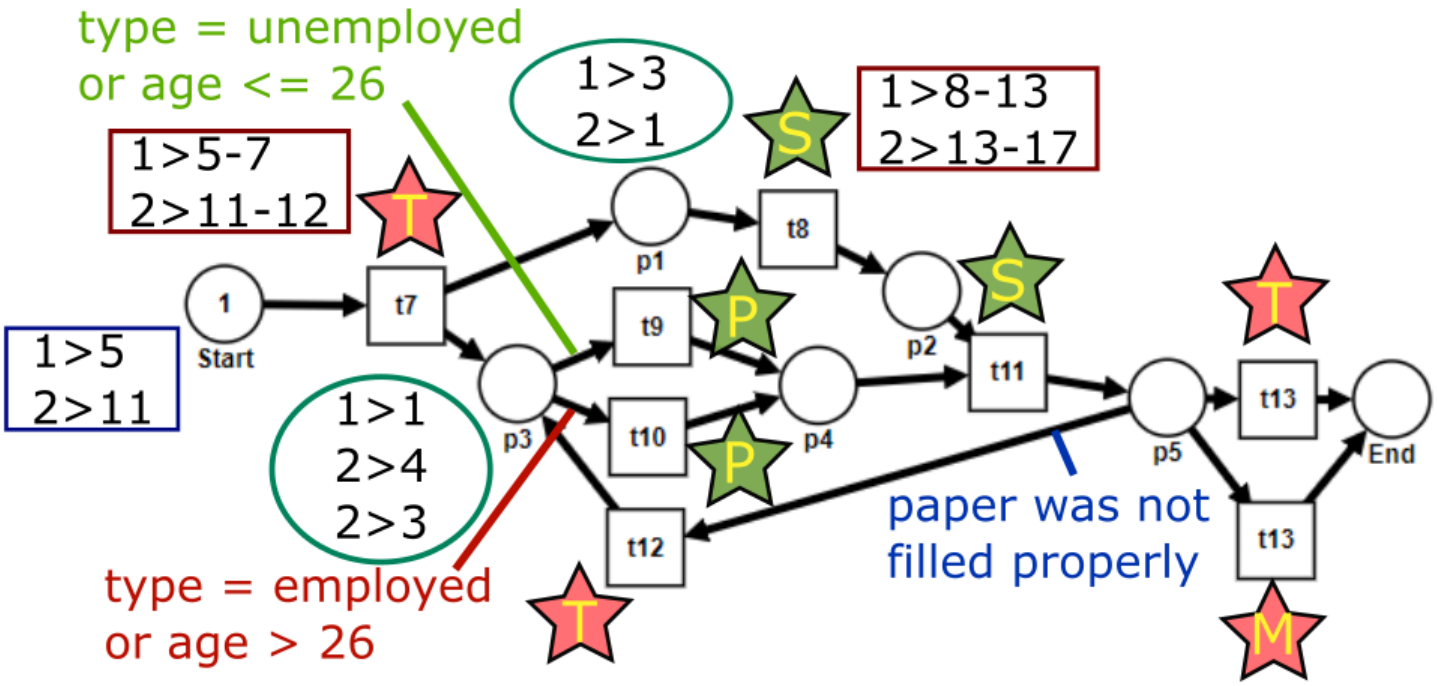
Adding additional perspectives

- Control flow is not the only perspective
- We can enhance the existing process models with:
  - Social network analysis
  - Organizational structures
  - Resource behavior analysis
  - Time perspective
  - Decision points mining
  - …



Additional perspectives

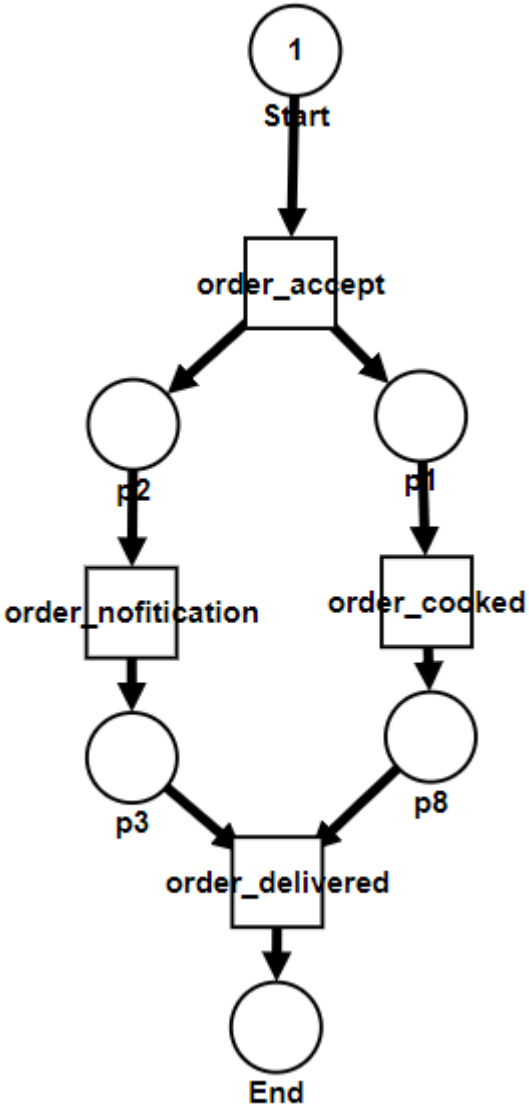
- We can add many others
- We can combine them to the integrated model
- Our model is enhanced, we might get better results



Conformance checking

- We can use the existing model to identify deviations in the behavior from logs

```
1;order_accept;
1;order_nofitication;
1;order_delivered;
1;order_cooked;
//NOK
```



Analysis of the present

- Also called operational support
- We use our model to analyze running cases
- We can:
  - Detect deviations in real-time data using the model of the desired behavior
  - Do real-time predictions (prob. of success, remaining time,…)
  - Make recommendations

Operational support: Detect deviations

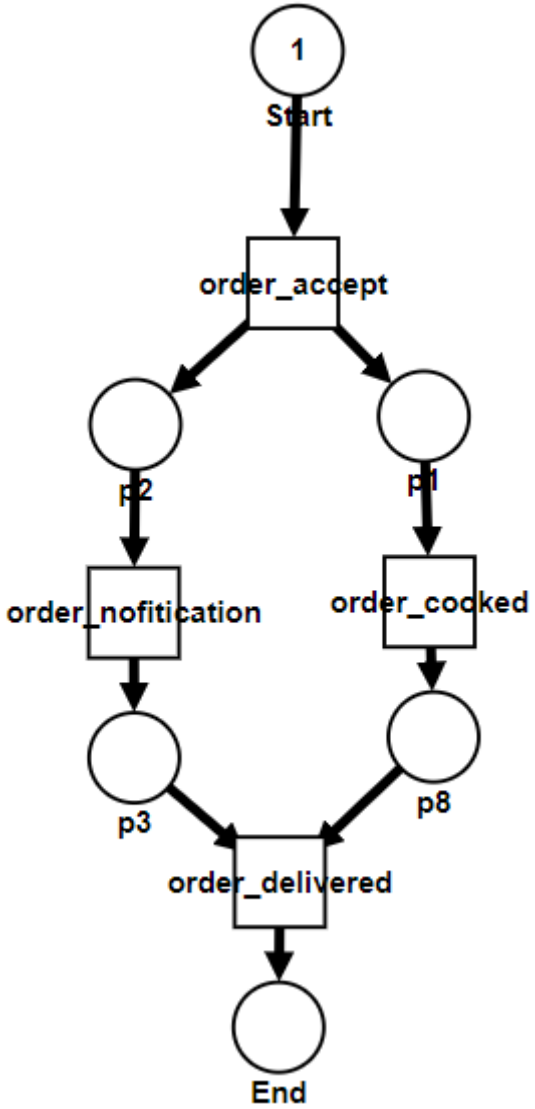
- We consider only the partial trace of a particular case
- We want immediate response when the deviation occurs

a) Token-based replay
b) Business rules

Detect deviations: Token-based replay

- Check the conformance with the model

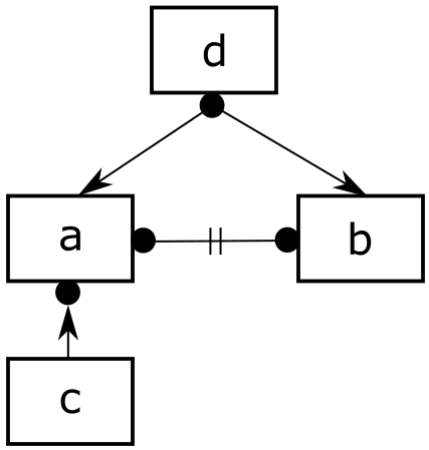
```
1;order_accept;          //OK
1;order_nofitication;    //OK
1;order_cooked;          //OK
1;order_delivered;       //OK

5;order_accept;          //OK
5;order_nofitication;    //OK
5;order_delivered;       //NOK
```



Detect deviations: Business rules

- Specific rules we want to follow
- To define them, we can use *Declare*
  - Constraint-based workflow language that uses graphical notations and semantics based on Linear Temporal Logic
- Example:
  - **a** and **b** cannot happen in the same case
  - **a** cannot happen before **c** has happened
  - every **d** have to be eventually followed by **a** or **b**

$$!((\blacklozenge a) \wedge (\blacklozenge b))$$
$$(!a)\ W\ c$$
$$\square(d \Rightarrow (\blacklozenge(a \vee b)))$$



lasaris

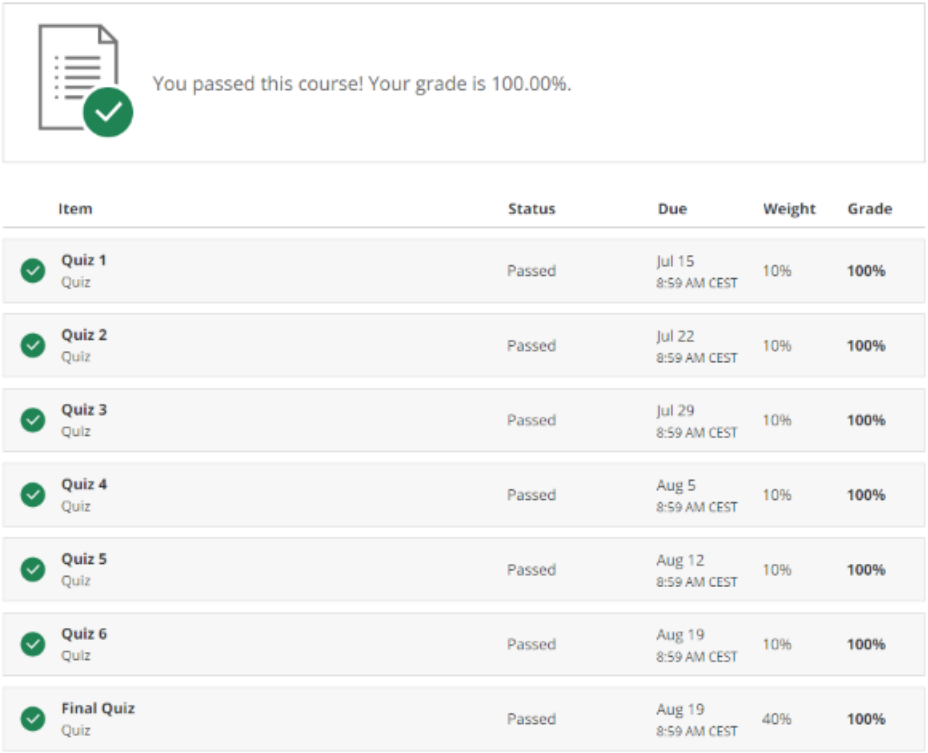
Operational support: Predict & Recommend

- We can apply data mining techniques (supervised learning, …)
- Examples of predictions:
  - Total cost of the current case
  - Total service time for the current case
  - Probability of meeting the deadline
  - Remaining flow time
- Examples of recommendations:
  - Minimize the total costs
  - Maximize the number of accepted cases
  - Minimize resource usage
  - Minimize the remaining flow time

lasaris

PV226 Course information

- e-learning (recommended: 2. – 7. week)
- https://www.coursera.org/learn/process-mining



You passed this course! Your grade is 100.00%.

Item	Status	Due	Weight	Grade
Quiz 1 Quiz	Passed	Jul 15 8:59 AM CEST	10%	100%
Quiz 2 Quiz	Passed	Jul 22 8:59 AM CEST	10%	100%
Quiz 3 Quiz	Passed	Jul 29 8:59 AM CEST	10%	100%
Quiz 4 Quiz	Passed	Aug 5 8:59 AM CEST	10%	100%
Quiz 5 Quiz	Passed	Aug 12 8:59 AM CEST	10%	100%
Quiz 6 Quiz	Passed	Aug 19 8:59 AM CEST	10%	100%
Final Quiz Quiz	Passed	Aug 19 8:59 AM CEST	40%	100%

lasaris

PV226 Course information

- Project
  - Application of Process Mining to a problem
  - You can come up with your own topic, set your own difficulty
  - You can work in pairs
  - We will have a meeting (**24.10.**?) where we will discuss your topics
  - **5.12.** – presentation of your work in Lasaris seminar (A319)
  - Optional consultations of your project / email communication ☺

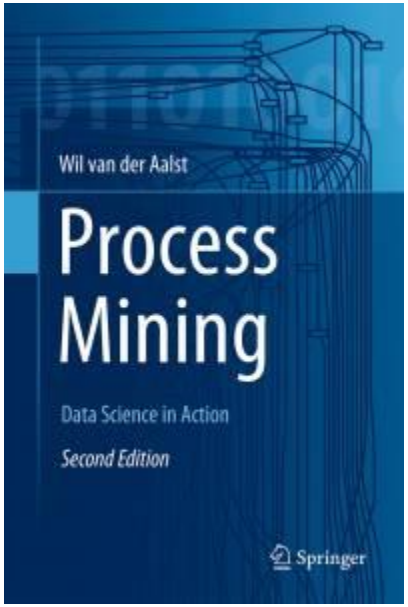
- Examples of project types:
  - Process discovery in tool Disco (https://fluxicon.com/disco/)
  - Process analysis in tool ProM (http://www.promtools.org/)
  - Process analysis using Python (https://github.com/pm4py/pm4py-source)
  - Survey research paper about the specific usage of Process Mining

Questions?

lasaris

Sources

- Presentation based on the book **Process Mining: Data Science in Action**
- https://www.springer.com/gp/book/9783662498507
- Use school VPN and you can download it! ☺



lasaris