PV226: Process Mining Seminar

Martin Macák

13. 9. 2021

Faculty of Informatics, Masaryk University
Brno

lasaris LAB OF SOFTWARE ARCHITECTURES AND INFORMATION SYSTEMS

Outline

- Basic overview of Process Mining
- Course information

What is process mining?

- Discipline that aims to understand and analyze processes.
- It uses a structured event log.
- We can use it to discover process maps.

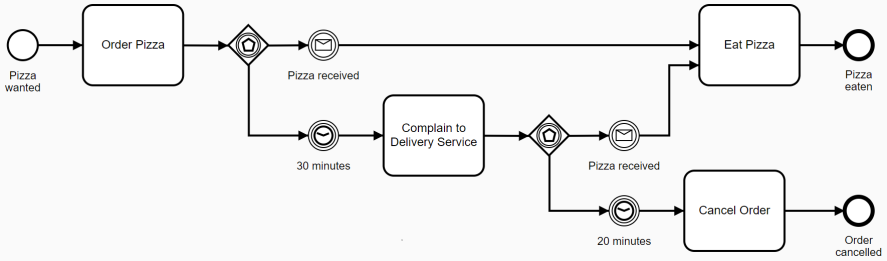


Image taken from: https://camunda.com/bpmn/examples/

Why use process mining?

- We assume how the process is performed.
- However, how the real process looks like?
- There might be differences between the real execution and the assumption:
  - special situations,
  - "shortcuts",
  - subjectivity,
  - malicious activity.

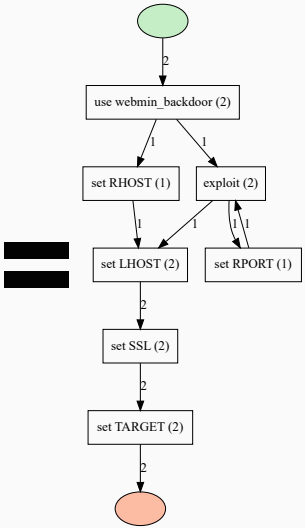
How does process mining work?

- Typically works 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
```

How does process mining work?

```
userId;timestamp;event
1;2.08.2020 10:31:43;use webmin_backdoor
1;2.08.2020 10:32:44;set RHOST
1;2.08.2020 10:34:19;set LHOST
1;2.08.2020 10:34:27;set SSL
1;2.08.2020 10:34:35;set TARGET
2;2.08.2020 10:52:55;use webmin_backdoor
2;2.08.2020 10:53:22;exploit
2;2.08.2020 10:56:24;set RPORT
2;2.08.2020 10:56:57;exploit
2;2.08.2020 10:59:51;set LHOST
2;2.08.2020 11:00:02;set SSL
2;2.08.2020 11:00:14;set TARGET
```



How does process mining work?

- Each event has:
  - $[Required]$ caseId,
  - $[Required]$ activity,
  - timestamp,
  - resource,
  - other data.

```
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
```

How does process mining work?

- 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.

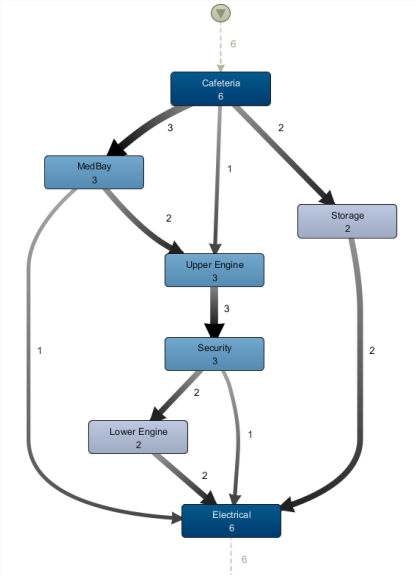
What is the difference between process mining and data mining?

What is the difference between process mining and data mining?

```
OK;1;Cafeteria;10:00
OK;1;MedBay;10:05
OK;1;Upper Engine;10:10
OK;1;Security;10:12
OK;1;Lower Engine;10:15
OK;1;Electrical;10:20

OK;3;Cafeteria;10:00
OK;3;Storage;10:04
OK;3;Electrical:10:10

malicious;5;Cafeteria;10:00
malicious;5;MedBay;10:03
malicious;5;Electrical;10:06
```

```
OK;2;Cafeteria;10:00
OK;2;MedBay;10:04
OK;2;Upper Engine;10:08
OK;2;Security;10:14
OK;2;Lower Engine;10:16
OK;2;Electrical;10:18

OK;4;Cafeteria;10:00
OK;4;Storage;10:06
OK;4;Electrical;10:12

malicious;6;Cafeteria;10:00
malicious;6;Upper Engine;10:06
malicious;6;Security;10:14
malicious;6;Electrical;10:19
```

What is the difference between process mining and data mining?



Where is process mining used?

- Healthcare
- Manufacturing
- Finance
- Public sector
- Usability
- Robotics, industry 4.0
- Utility
- Advisory, audits
- Biology
- Agriculture
- ICT
- Education
- Logistics
- Security
- Call center
- Entertainment
- Garment
- Retail
- Hotel

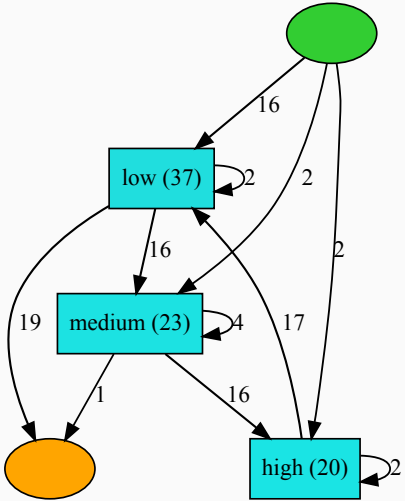
More details in [1]

Let’s start to mine!

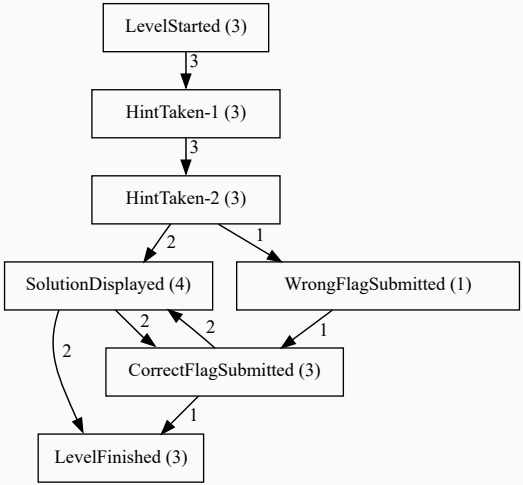
Analysis of the past

- Process discovery techniques
- From the event log, we create a model that represents how the process was executed
- Model can be represented as a petri net, activity diagram, BPMN diagram, heuristic net, ...
- Now we focus only on control flow

Analysis of the past: Employees' productivity



Analysis of the past: Cybersecurity training session



Process discovery activities

- We can:
  - explore processes,
  - 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,
  - ...

Detecting deviations in processes

- We can check the conformance with the model:

Detecting deviations in processes

- We can check the conformance with the model:

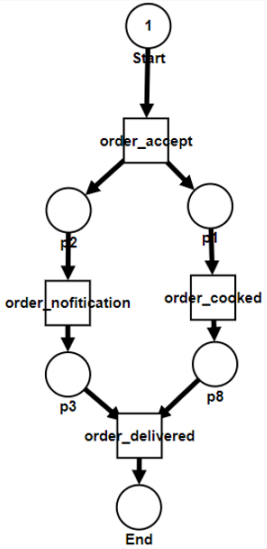
Detecting deviations in processes

- We can check the conformance with the model:

  - token-based replay,
  - business rules,
  - ...

Token-based replay

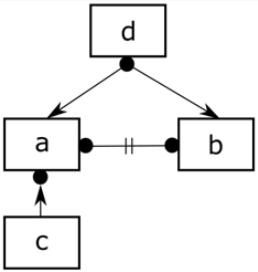
```
1;order_accept;          //OK
1;order_nofitication;    //OK
1;order_cooked;          //OK
1;order_delivered;       //OK

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



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**

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



Analysis of the present

- We analyze running cases.
- We can:
  - detect deviations in real-time data using the model of the desired behavior,
  - do real-time predictions (probability of success, remaining time, ...),
  - make recommendations.

Deviation detection: past vs. present

Deviations in past

vs

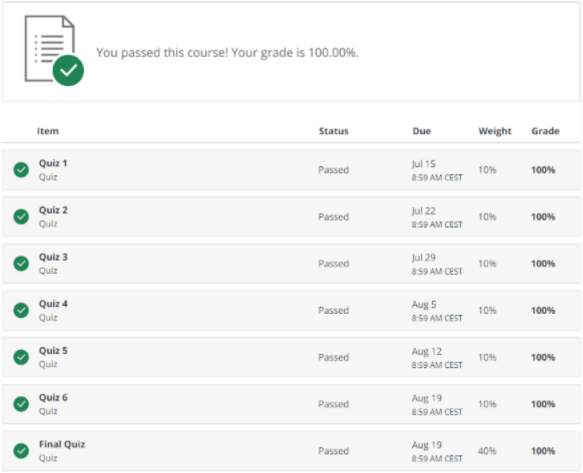
Deviations in present

Summary

- Process-centric data analysis.
- Process discovery, enhancement, and conformance checking.
- Past *vs.* present.

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%

PV226 Course information

- Project
  - You can come up with your own topic, set your own difficulty.
  - You can work in groups.
  - We will have a meeting (Oct/Nov) where we will discuss your topics.
  - The last week of semester — presentation of your work.
  - Optional consultations of your project through the whole semester.

PV226 Course information

- Project
  - You can come up with your own topic, set your own difficulty.
  - You can work in groups.
  - We will have a meeting (Oct/Nov) where we will discuss your topics.
  - The last week of semester — presentation of your work.
  - Optional consultations of your project through the whole semester.
- Examples of project types:
  - process discovery in tool Disco (https://fluxicon.com/disco/),
  - process analysis in tool ProM (http://www.promtools.org/),
  - process analysis in tool RapidMiner (https://rapidminer.com/),
  - process analysis using Python or .NET
    (https://github.com/pm4py/pm4py-source)
    (https://github.com/lasaris/ProcessM.NET),
  - survey research paper about the specific usage of Process Mining.

Project results opportunities

- conferences
  - Information Technology for Practice (http://www.cssi-morava.cz/new/index.php?id=155)
  - International Conference on Process Mining (https://icpmconference.org/2021/)
  - International Conference on Internet of Things, Big Data and Security (https://iotbds.scitevents.org/)
  - Conference on Computer Science and Intelligence Systems (https://fedcsis.org/2021/)
  - ...
- journals

Project results examples

- Towards Process Mining Utilization in Insider Threat Detection from Audit Logs (https://ieeexplore.ieee.org/abstract/document/9336573)
- Game Achievement Analysis: Process Mining Approach – result in 2 days (https://is.muni.cz/auth/th/jsq9f/)
- Using process mining for Git log analysis of projects in a software development course (https://link.springer.com/article/10.1007/s10639-021-10564-6)

Additional sources

- Process Mining book [2]
- https://www.springer.com/gp/book/9783662498507
- Use school VPN and you can download it! :)

Additional sources

- Process Mining book [2]
- https://www.springer.com/gp/book/9783662498507
- Use school VPN and you can download it! :)
- Our Discord server: https://discord.gg/CyykPVN
- Discuss anything with your colleagues and me :)

Resources

[1] C. dos Santos Garcia, A. Meincheim, E. R. F. Junior, M. R. Dallagassa,D. M. V. Sato, D. R. Carvalho, et al., Process mining techniques and applications - a systematic mapping study, Expert Systems with Applications, vol. 133, pp. 260 – 295, 2019.doi: https://doi.org/10.1016/j.eswa.2019.05.003. [Online].

[2] W. van der Aalst, Process Mining: Data Science in Action, 2nd Edition, Springer Publishing Company, Incorporated, 2016.