Process Mining: Case Study

Sareh Sadeghian
Somayeh HosseinZadeh
Contents

• Process Mining for Analyzing Inventory Processes

• Process Mining in Healthcare
  – Case Study No. 1
  – Case Study No. 2
  – Case Study No. 3

• Process Mining for Analyzing Software Processes

• Process Mining for Legacy Systems
Process Mining for Analyzing Inventory Processes

- Dendro: a mattress production company in Poland
- Goal: conformance checking
- Data log recorded during 5 months
Process Discovery Results
Distribution of Activities
Process Mining in Healthcare
Case Study No. 1

- Academic hospital in Netherlands
- 627 oncology patients
Challenges

• for each activity it was only known on which day the service had been delivered

• 376 different event names for activities
  • Filtering log by only keeping the initial lab event
  • Aggregating log by summarizing events that happened within one department
    • ‘echo abdomen’, ‘thorax’ and ‘CT brain’ were just represented as ‘radiology’
  • The result was an event log with less than 60 different activities
Process Discovery Result
Clustering Algorithm
Process of the Largest Cluster
Case Study No. 2

- Sao Sebastiao hospital in Portugal
- 27930 patients treated in 2009
Process Discovery Result
Regular Process
Infrequent process
Case Study No. 3
• The General hospital of Chania is a 400 bed hospital providing all types of health services.

• The management committee of the hospital asked the research team to analyze their emergency department process.
• The research team drew a flow chart based on this description, which was confirmed as the expected process flow for the emergency department.
Sample of data

• A sample of data covering 250 patients who visited the emergency department from the hospital information system to compare the actual process with the expected process using process mining techniques.
As a first step, the actual process was discovered using process mining (see pictures of the simplified process below).
Differences could be found by comparing the real process with the assumed process:

- There are **more entry points as well as exit points** to the process than the ones illustrated by the flow chart. While the additional start activities are not so important, the additional exit points require administrative actions.

- There were cases, where **patients exit right after Diagnosis** — a path which did not exist in the flow chart. Of course, a doctor can diagnose that the sickness is not severe, but letting patients leave without any examinations or lab tests is still an action that could be challenged.

- People may take some lab tests (especially biochemical and enzyme test which take considerable time), **leave the hospital and return later**. This path is not illustrated in the flow chart.

- There were several cases where **prescriptions were made without any examinations**. Again, this is possible according to medical guidelines, but it could be investigated.
Focus of the analysis

• The focus of the analysis was **not check for conformance** but to **understand the actual process**. Recall that the process in is **not a rigid prescription** of the process but a description of the process by a domain expert.

• What became apparent, however, was that the assumed process was not suitable as a basis for simulation, because — due to a **lack of alignment with the real process** — it would produce faulty results.
• Due to their complexity, these spaghetti diagrams were of little usefulness for the management committee. So, the challenge was to create simpler process maps that could be used to derive insights about the process.

• A common technique for more structured processes is to concentrate on the most frequent variants in the process.
Derive 3 diagrams
Difference can be detected by comparing the process flows for

• In Cluster (a)
  – the ‘Registration’ step is often skipped.
  – **Green ~34%**, Yellow ~57%, and Red ~8%.
    • This can be attributed to the economic crisis

• In Cluster (b)
  – ‘Registration’ is also skipped, but this can be attributed to a higher emergency of cases (much more Red cases than normal).
  – this cluster has a **higher frequency of lab tests** (all patients have blood / biochemical).
  – this cluster has a **high percentage of patients that enter a clinic** rather than just leaving the hospital.

• Cluster (c) is closest to the expected flow
  – People get **registered**, have **some tests**, and are **forwarded towards the exit** via the expected way (after a prescription or a treatment in the ED room).
Why is this important?

• In retrospect, we were able to correlate the patients’ clusters with their triage.

• to predict more accurately the workload per activity, create a better resource allocation plan, etc.

• to provide interesting insights, such as that some patients with not severe illnesses came to the hospital at night to avoid the registration costs of 5 Euro, which can be attributed to the economic crisis.
Process Mining for Analyzing Software Processes
Introduction

• Both user interaction and system’s internal behavior can be analyzed with the help of process mining.

• The results of this analysis can significantly influence the architecture, design, testing, and development of the software system.
1. The team wanted to analyze the productive behavior of the users in order to see the system failures, bottlenecks, and to gather statistics.

2. Several critical performance challenges appeared with an increasing number of users, they had to be identified and solved.
Case 1: User positive behavior (Frequency View)
User negative behavior (Performance View)
Case 2: System Performance Analysis
Detailed statistics of calls

<table>
<thead>
<tr>
<th>Value</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(RQ)_tnetnk_c.xcSuppTkPrkductService.letDivisiknDxbFkrPrkduct</td>
<td>4654</td>
</tr>
<tr>
<td>P(RS)_tnetnk_c.xcSuppTkPrkductService.letDivisiknDxbFkrPrkduct</td>
<td>4654</td>
</tr>
<tr>
<td>D(RQ)_ckrpuq.bexdQueueEntry.bexdQueueEntries</td>
<td>3648</td>
</tr>
<tr>
<td>D(RS)_ckrpuq.bexdQueueEntry.bexdQueueEntries</td>
<td>3648</td>
</tr>
<tr>
<td>D(RQ)_tsuccx.nixtnixncikvnireggit.evseckvnireggirt</td>
<td>2637</td>
</tr>
<tr>
<td>D(RS)_tsuccx.nixtnixncikvnireggit.evseckvnireggirt</td>
<td>2637</td>
</tr>
<tr>
<td>D(RQ)_tsuccx.nixtnixnxVxtCknfigurxtkn.findxll</td>
<td>2329</td>
</tr>
<tr>
<td>D(RS)_tsuccx.nixtnixnxVxtCknfigurxtkn.findxll</td>
<td>2329</td>
</tr>
<tr>
<td>D(RQ)_tsuccx.nixtnixxTkurNunberDefinitkn.letTkurNunberDefinitknByCriterix</td>
<td>2329</td>
</tr>
<tr>
<td>D(RS)_tsuccx.nixtnixxTkurNunberDefinitkn.letTkurNunberDefinitknByCriterix</td>
<td>2329</td>
</tr>
<tr>
<td>D(RQ)_iek.iekDxbQuery.letVxLkkxtiknTypelekZkne</td>
<td>2328</td>
</tr>
<tr>
<td>D(RS)_iek.iekDxbQuery.letVxLkkxtiknTypelekZkne</td>
<td>2328</td>
</tr>
<tr>
<td>D(RQ)_drpckd.Txnsldtkn.letTxnsldtkn</td>
<td>2328</td>
</tr>
<tr>
<td>D(RS)_drpckd.Txnsldtkn.letTxnsldtkn</td>
<td>2328</td>
</tr>
<tr>
<td>P(RQ)_tncnk.berxvttkn.letBrkchubeDtx</td>
<td>2328</td>
</tr>
<tr>
<td>P(RS)_tncnk.berxvttkn.letBrkchubeDtx</td>
<td>2328</td>
</tr>
</tbody>
</table>
System Performance Analysis
Summary

• showing two successful applications of process mining in a concrete enterprise software project.

• From our point of view, this is a very fruitful application domain, because productive software systems provide a big amount of data in form of logs and traces. This data can and should be analyzed in order to improve the software quality.
Process Mining for Legacy Systems
Legacy systems
One of the most interesting use cases for process mining is the understanding of legacy systems.

In many cases the developers are long gone when changes to these systems must be made and it becomes a huge burden just to maintain these often mission-critical systems.
Legacy systems

- Two texts on process mining for legacy systems by Steve Kilner:
  - ‘Reinvent Your Business With Business Process Mining’ about the opportunities of understanding software systems from a business perspective.
  - ‘Process Mining: A Living View of Systems’ even provides open source example code that you can use to extract the data.
why is the so-called greenfield development, where you make a fresh start, often not possible and people have to **put up with all these old systems** that nobody understands anymore?

- *Replacing legacy systems is costly, risky and disruptive to organizations.*
• What does process mining add compared to traditional approaches such as static code analysis techniques?

  – Anyone who has been a programmer working with existing code knows that is impossible to look at a large program, let alone an entire system, and grasp everything that could happen within it.

  – By creating or obtaining event logs of executing programs, possibly through program instrumentation if necessary, it is possible to observe the paths that are actually used, along with their frequency. By examining individual cases it is possible to then correlate data inputs with resulting path variances.
Thanks for Your Attention