The ROAD from sensor data to process instances

Andreas Rogge-Solti
Colleagues

- Arik Senderovich
- Andreas Rogge-Solti
- Avigdor Gal
- Jan Mendling
- Avishai Mandelbaum

- (some slides / figures are taken with thanks from “Tech Talk @ Technion” by Arik Senderovich)
IBM Healthcare Report

- Improvement Potential
- Inefficiency (% of Spending)

Health: 42% (of 4,270 Bil. USD)

- 34% (of the 42%)
- Transport.
- Gov. and Safety
- Electricity
- Clothing
- Comm.
"if you can not measure it, you can not improve it."
- Lord Kelvin
Hospital setting

1000 patients/day
250 providers/day
Big brother’s watching you (but he doesn’t understand what’s going on...)

What is the activity?

Blood Draw → Exam → Infusion
Process Perspective

- Sensor data of people and resources keep track of their locations.
- Processes are driven by **interactions** between resources.
- **Activities** are performed within these interactions
- Process knowledge tells us what should be happening

- Related areas:
  - Automatic process discovery with **abstraction gaps**
  - Multi-dimensional **conformance checking**
  - **Activity recognition** techniques
Research Problem: 
*aggReated Observations to Activity Data (ROAD)*

![Diagram](image)

**Fig. 2: The ROAD to solution.**
Paving the ROAD
Preprocessing Sensor Data

(res₁, 9:00, receiver₁) (res₁, 9:52, receiver₂)
(res₁, 9:37, receiver₁) (res₁, 10:03, receiver₂)

(res₁, loc₁, 9:00, 9:37) (res₁, loc₂, 9:52, 10:03)
Many activities are conducted by a set of resources interacting at a certain location

Some examples:

- **Doctors** treating **patients** in an exam room
- A **technician** repairing a **printer** in a garage
- A **photographer** developing pictures in a darkroom
- **N philosophers** dining around a table arguing about forks
- ...
Assumptions

- Sensor data is already cleaned.
- Activities result in interactions.
- Interactions don’t span multiple process instances
- A change in the resource set of a location can initiate and terminate new/ongoing interactions
- Interaction requires two entities (no general requirement)
- We can observe the case entities (e.g. patients, printers, ...)
- Dimensions of process knowledge (e.g. location, duration, assignments) are independent, given the activity
Start of the ROAD
Interaction Mining

- Many combinations of resources over time
  - R1 + R2
Many combinations of resources over time

- R1 + R2
- R1 + R3
Start of the ROAD
Interaction Mining

- Many combinations of resources over time
  - R1 + R2
  - R1 + R3
  - R1 + R2 + R3

General framework for the description of interactions:
- Selection
- Grouping
- Filtering
- Construction
Middle of the ROAD
Matching problem

Interactions

Activities

Exam
Infusion
Blood Draw
Matching problem
Encoded Knowledge

Interactions
Activities

Durations

Exam
Infusion
Blood Draw
Matching problem
Encoded Knowledge

Interactions

Precedence Relations

Activities

- Exam
- Infusion
- Blood Draw
Matching problem
Solution as ILP

Interactions

Activities

Exam
Infusion
Blood Draw
Let’s deep dive into the ROAD
General formulation

- Maximize $g(x)$ subject to $B^T x \leq 0$

- where $g(x) = \sum_{i \in I} \sum_{a \in A} w_{i,a} x_{i,a}$
Encoding Probabilities as Weights

- Given probabilistic process knowledge (e.g., distribution of time, resources, locations)
- The **weight** of mapping an interaction to an activity is given as:

  \[ w_{i,a} = P(x_{i,a} \mid i = (E, l, s, c)) \]

- Applying Bayes’ theorem, we get:

  \[ w_{i,a} = \pi_a f_{D_a}(i.c - i.s) f_{E_a}(i.E) f_{L_a}(i.l) \]

Prior of an interaction coming from activity a (can be estimated by the share that a has in all activities)
Encoding Probabilities as Weights

- Given probabilistic process knowledge (e.g., distribution of time, resources, locations)
- The **weight** of mapping an interaction to an activity is given as:

\[ w_{i,a} = P(x_{i,a} \mid i = (E, l, s, c)) \]

- Applying Bayes’ theorem, we get:

\[ w_{i,a} = \pi_a f_{D_a}(i.c - i.s) f_{E_a}(i.E) f_{L_a}(i.l) \]

Value of the probability density function of the duration distribution of activity “a” at the length of the interaction
Encoding Precedence Constraints

- Relation $\prec_P \subseteq A \times A$ captures precedence amongst activities

<table>
<thead>
<tr>
<th></th>
<th>Soft encoding</th>
<th>Hard encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\forall (a, b) \in \prec_P, \forall i \in I_c :$</td>
<td>$z_{i,a,b} \leq 1 - x_{i,b} + y_{i,a}$</td>
<td>$x_{i,b} \leq y_{i,a}$</td>
</tr>
<tr>
<td></td>
<td>$z_{i,a,b} \geq 1 - x_{i,b}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$z_{i,a,b} \geq y_{i,a}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$z_{i,a,b} \in {0, 1}$</td>
<td></td>
</tr>
</tbody>
</table>
Encoding Assumptions

Assumptions

- An entity cannot be involved in two interactions at the same time

\[ \forall (i_1, i_2) \in I_{exclusive}: \sum_{a \in A} x_{i_1,a} + \sum_{a \in A} x_{i_2,a} \leq 1 \]

- An activity instance corresponds to at most a single interaction, and every interaction stems from at most a single activity instance.

\[ \forall i \in I_c: \sum_{a \in A} x_{i,a} \leq 1; \quad \forall j \in J, \forall a \in A: \sum_{i \in I_c^j} x_{i,a} \leq 1 \]
Matching problem
Solving the ILP

Interactions

Solution

Activities

Exam

Infusion

Blood Draw
Evaluation Setup

Preparation

- **experiment parameters** (e.g., number of patients, rooms)
- **Process knowledge** (e.g., precedence constraints, duration distributions)

1. Generate process model
2. Simulate
3. Convert to raw intervals
4. ROAD (algorithm with ILP solver)
5. Measure distance between event logs

Evaluation

- **Quality measures** (e.g., similarity measures)
- **Reconstructed AD log**
(a) Similarity with respect to *increasing entropy* in the mapping to activities.

(b) Similarity with respect to *increasing noise* reflecting deviations from the process knowledge.

(c) Similarity with respect to *increasing overlap* in activities per location.
entropy: \(0 + 0 + 0 = 0\)

typeface (37x0 to 663x450)

entropy: \(2 + 2 + 1 = 5\)

(c) Similarity with respect to increasing overlap in activities per location.
End of the ROAD?

- **Interactions** as intermediate knowledge layer
- Consistently encoding **different dimensions of process knowledge** (e.g., time, resources, locations, precedence constraints)

- We’re able to **reconstruct activities** of a process
  - The more **process knowledge** -> the better the results
  - Higher noise -> linear decrease in accuracy (precedence constraints)
  - Bigger rooms -> exponential runtime, due to combinations of intervals

- Helping to discover incompliancies and inefficiencies in location aware business processes
DEPARTMENT FÜR INFORMATIONSVERARBEITUNG
DEPARTMENT OF INFORMATION SYSTEMS
Welthandelsplatz 1, 1020 Vienna, Austria

DR. ANDREAS ROGGE-SOLTI
T +43-1-313 36-4753
andreas.rogge-solti@wu.ac.at
www.wu.ac.at