Karlsruher Institut für Technologie

Intelligente iSAS Sensor-Aktor-Systeme

Interaktive Echtzeitsysteme IES

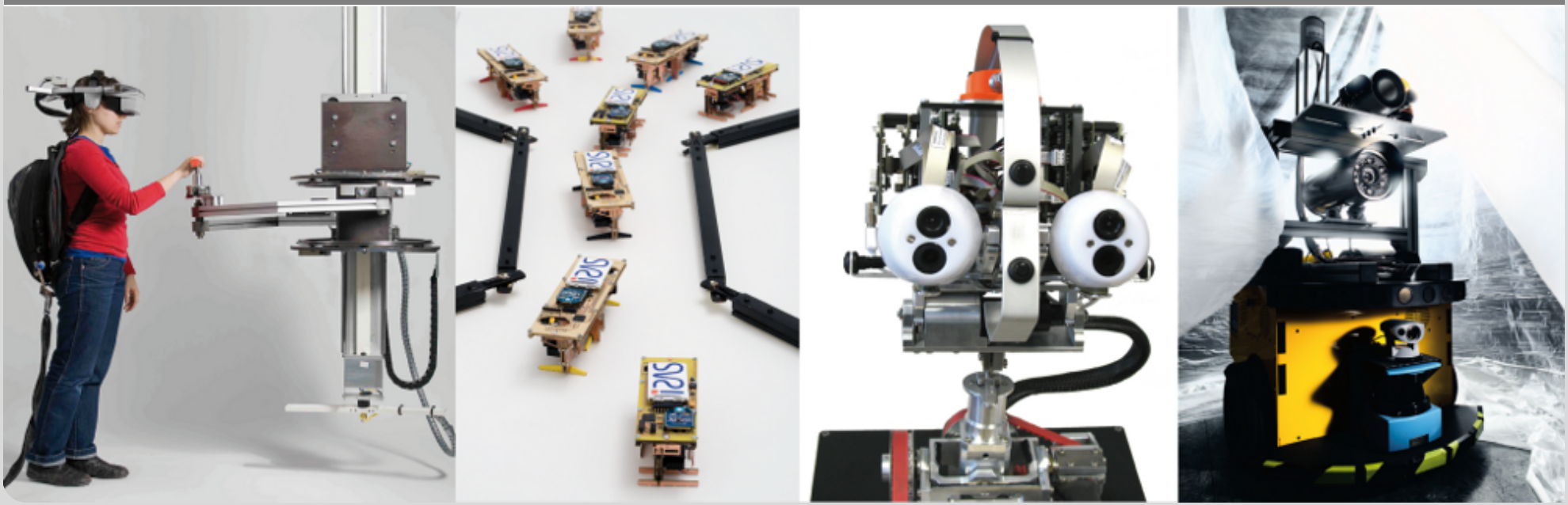
Fraunhofer IOSB

Scalable Inductive Process Mining

Mining process trees from large event logs with guarantees

Maximilian L. Franz | 24. Dezember 2016

PROSEMINAR ANTHROPOMATIK: VON DER THEORIE ZUR ANWENDUNG



Agenda

1. Motivation
2. Basic Notation
3. Inductive Mining
4. Demo

Why mine processes?

- Validate exisiting process models
- Gather new knowledge

Problem Statement

- Mine Process Model from a log of empirical data
- Balance between
  - *fitness*
  - *simplicity*
  - *generality*
  - *precision*

Why mine processes?

- Validate exisiting process models
- Gather new knowledge

Problem Statement

- Mine Process Model from a log of empirical data
- Balance between
  - *fitness*
  - *simplicity*
  - *generality*
  - *precision*

Why mine processes?

- Validate exisiting process models
- Gather new knowledge

Problem Statement

- Mine Process Model from a log of empirical data
- Balance between
  - *fitness*
  - *simplicity*
  - *generality*
  - *precision*

Why mine processes?

- Validate exisiting process models
- Gather new knowledge

Problem Statement

- Mine Process Model from a log of empirical data
- Balance between
  - *fitness*
  - *simplicity*
  - *generality*
  - *precision*

Logs

Definition (Log)

- A log L is a multi-set of traces σ_i
- A trace σ_i is a finite sequence of activities in $\mathcal{A}$: $\sigma_i \in \mathcal{A}^*$
- For simplicity let $\mathcal{A} := \{a, b, c, \dots\}$

Operators

- We consider a set of operators $\bigoplus = \{\times, \rightarrow, \wedge, \circlearrowleft\}$
- They define relation between logs (like regular expression on languages)
  - $\times$: Exclusive choice
  - $\rightarrow$: Sequence
  - $\wedge$: Parallel
  - $\circlearrowleft$: Loop

Process Trees

- Abstract Representation of a process model
- Represent Regular Expression

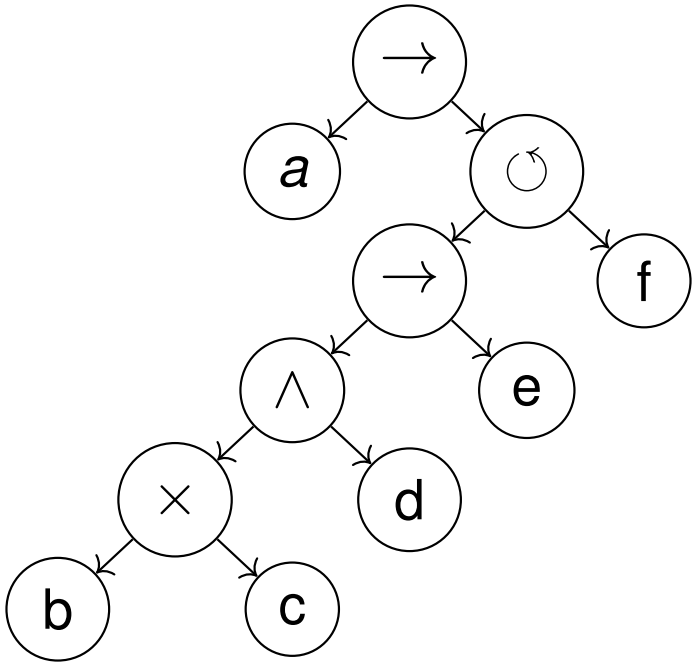


Abbildung: Process tree to log $L = \{\langle a, c, d, e, f\rangle, \langle a, c, d, e, f, d, b, e, f\rangle, \ldots\}$

Inductive Mining

Inductive Mining

Idea

Log $\rightarrow$ Directly-follows-graph (DFG) $\rightarrow$ Cuts $\rightarrow$ Sub-logs

- Consider the log $L_2 = \{\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle, \langle a, e, d\rangle^2\}$

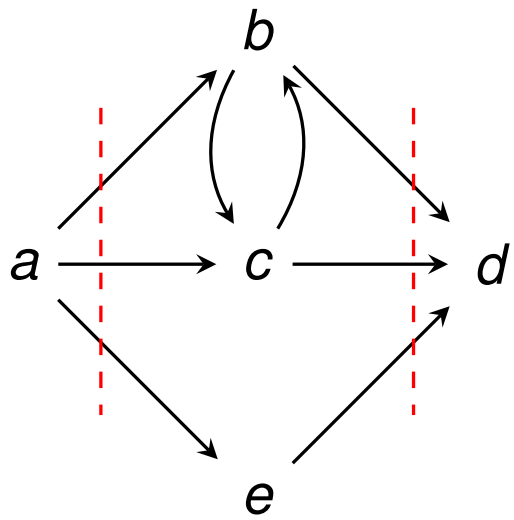


Abbildung: DFG $G_L = G(L_2)$ constructed from L_2

- Found: Sequence Cut: $\rightarrow (a, (bce), d)$

Inductive Mining

Idea

Log $\rightarrow$ Directly-follows-graph (DFG) $\rightarrow$ Cuts $\rightarrow$ Sub-logs

- Consider the log $L_2 = \{\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle, \langle a, e, d\rangle^2\}$
- Reminder: $\bigoplus = \{\times, \rightarrow, \wedge, \circlearrowleft\}$

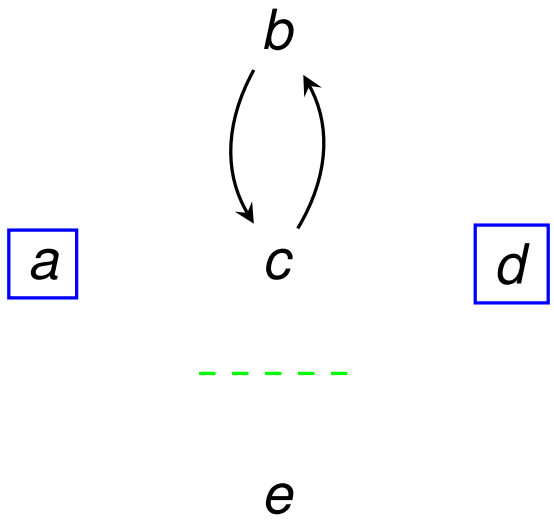


Abbildung: DFG $G_L = G(L_2)$ constructed from L_2

Inductive Mining - Result

Idea

Log $\rightarrow$ Directly-follows-graph (DFG) $\rightarrow$ Cuts $\rightarrow$ Sub-logs

- Consider the log $L_2 = \{\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle, \langle a, e, d\rangle^2\}$
- Reminder: $\bigoplus = \{\times, \rightarrow, \wedge, \circlearrowleft\}$
- Resulting Tree

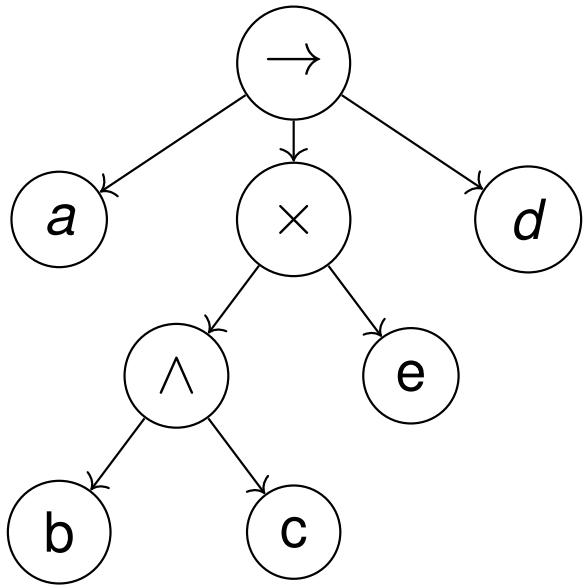


Abbildung: Process tree Q_L mined from L_2 with inductive mining after [**?**]

DEMO

Thank you for your attention