Business Process Management #processscience #bpmatqut

# Process Mining For The Impatient

Adam Burke, April 2021

@AdamBurkeware https://adamburkeware.net/



the university for the real world





### **Process Mapping**







#### **Process Mining**



#### **Process Mining**



#### From System Logs to Process Models





Adapted from van der Aalst [100,p276]<sub>8</sub>

#### Sampling vs Mining



### Technical detail

A taste

### Event Logs

- Timestamp(s)
- Case ID
- Activity
- Other resource information

#### Event Logs

- Timestamp(s)
- Case ID
- Activity
- Other resource information

Eg ... just off the top of my head ...

2021-04-19 16:14:02.221 TraceIncoming NewOrderSingle 11=BOB9381-1;

### Event Logs (formally)

• The timestamp-ordered activity messages for a Case ID form a *trace* 

#### Event Logs (formally)

- The timestamp-ordered activity messages for a Case ID form a *trace*
- A *trace* is then a *sequence* of activities, eg <a,c,e,a>, or <NOS, ACK, PFILL, PFILL, FILL>

#### Event Logs (formally)

- The timestamp-ordered activity messages for a Case ID form a *trace*
- A *trace* is then a *sequence* of activities, eg <a,c,e,a>, or <NOS, ACK, PFILL, PFILL, FILL>

• A *log* is a multiset (bag) of traces, eg [[  $\langle a,c,e,a \rangle^{21}$ ,  $\langle c,b,f \rangle^4$ ,  $\langle b,b,b,e \rangle^8$  ]]

#### A Taste Of Process Discovery - Alpha Algorithm

• Alpha Algorithm - first (very) useful process mining discovery algorithm (2002)

#### A Taste Of Process Discovery - Alpha Algorithm

- Alpha Algorithm first (very) useful process mining discovery algorithm (2002)
- Since superseded, but still useful intro / teaching tool
- "The Bubble Sort of Process Model Discovery"

#### **Ordering Relations**

Sequence (>)	a followed-by b
Strict sequence (->)	a followed-by b and b never-follows a
No follow (#)	a never-follows b and b never-follows a
Parallel	a followed-by b and b followed-by a



#### Alpha Algorithm (steps 1-4)

- 1. TL = set of all traces in the log
- 2. TI = set of all first activities in any trace < b, ... >
- 3. TO = set of all final activities in any trace  $\langle a, ..., e \rangle$
- 4. XL = set of (A,B) pairs where A and B are sets of neighbour activities



#### Alpha Algorithm (steps 5-8)

- 5. YL = eliminate subset pairs from XL, leaving only maximal pairs
- 6. PL = a set of places, one input, one output, and one for every element of YL
- 7. FL = a set of flows (edges) connecting every activity in YL with an incoming and outcoming place in PL
- 8. Return Petri net (PL, TL, FL)



#### An Output Petri Net (example)



#### My Research - Stochastic Process Mining



### BUT WHY?

#### 6.2.2 Algorithm

After showing the basic idea and some examples, we describe the  $\alpha$ -algorithm [157].

**Definition 6.4** ( $\alpha$ -algorithm) Let *L* be an event log over  $T \subseteq \mathscr{A}$ .  $\alpha(L)$  is defined as follows:

- 1.  $T_L = \{t \in T \mid \exists_{\sigma \in L} t \in \sigma\},\$
- 2.  $T_I = \{t \in T \mid \exists_{\sigma \in L} t = first(\sigma)\},\$
- 3.  $T_{\mathcal{O}} = \{t \in T \mid \exists_{\sigma \in L} \ t = last(\sigma)\},\$
- 4.  $X_L = \{(A, B) \mid A \subseteq T_L \land A \neq \emptyset \land B \subseteq T_L \land B \neq \emptyset \land$  $\forall_{a \in A} \forall_{b \in B} a \to_L b \land \forall_{a_1, a_2 \in A} a_1 \#_L a_2 \land \forall_{b_1, b_2 \in B} b_1 \#_L b_2 \},$
- 5.  $Y_L = \{(A, B) \in X_L \mid \forall_{(A', B') \in X_L} A \subseteq A' \land B \subseteq B' \Longrightarrow (A, B) = (A', B')\},\$
- 6.  $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{ i_L, o_L \},\$
- 7.  $F_L = \{(a, p_{(A,B)}) \mid (A, B) \in Y_L \land a \in A\} \cup \{(p_{(A,B)}, b) \mid (A, B) \in Y_L \land b \in A\}$ B  $\cup$  { $(i_L, t) | t \in T_I$ }  $\cup$  { $(t, o_L) | t \in T_O$ }, and
- 8.  $\alpha(L) = (P_L, T_L, F_L).$

#### Desire Paths

Complement top-down managerial understanding with bottom-up data-driven understanding

Find out what people and systems are doing in your organization



User TM @ wikicommons

#### Toolkits and Introductions

- Python pm4py
  - <u>https://pm4py.fit.fraunhofer.de/</u>
  - <u>https://github.com/pm4py</u>

- Java ProM
  - <u>https://www.promtools.org/doku.php</u>

- Commercial Tools
  - Apromore, Celonis, ...





### Process Mining (BPM)

Process Mining came out of and overlaps with Business Process Management (BPM)

Process Mining: Data Science in Action (2016)

Coursera MOOC https://www.coursera.org/learn/process-mining

Gartner 2020 Report

https://www.gartner.com/en/documents/3991229/mark t-guide-for-process-mining

 $\mathrm{TU}/\mathrm{Eindhoven},\,\mathrm{RWTH}$  Aachen, QUT, ...



Data Science in Action Second Edition



Business Process Management #processscience #bpmatqut

## Questions / Discussion





the university for the real world @AdamBurkeware <u>https://adamburkeware.net/</u>