Advanced Process Mining

Bruno M. Pacheco

April 15, 2020

Contents

Lecture 1: Introduction to Process Mining

13.04.2020

Overview

Definitions

About Process Mining.

Definition 1. A Process is a discrete collection of activities executed to achieve a goal.

Definition 2. Mining is "Gaining knowledge of, and, insights in (business) processes by analysing the event data stored during execution of the process".

Basic process data example:

Table 1: Basic example of event data

package	task	timestamp
1eg081hr	load	12:36
1eg081hr	dispatch	12:38
41yp39he	load	12:37

The 'package' column is the *Case identifier*, it identifies the instance of the process. 'task' is the *Activity* column, it explains what is being performed at a data record (row). 'timestamp' is the *Timestamp* column, a time reference to the data record. The last one is not necessary and can come in several different ways, even sometimes having record for beginning and end.

We can also have different data in a log, but these are the bare basics.

Definition 3. An *Event* captures the (partial) execution of an activity within a process instance.

Events are the atoms of our data, the rows in the dataset.

Definition 4. A *Trace* is a collection of events related to a process instance.

We can make traces by the *Case identifier* reference.

Process Mining - Anaylsis

There are three branches of Process Mining:

Definition 5. *Process Discovery* is deriving process models from process logs.

Definition 6. *Conformance Checking* is assessing if reality in the data is conforming to specification (from a process model).

Definition 7. *Process Enhancement* is providing performance information to a process model (as an overlay).

Process discovery acts from event logs, in the model. Conformance checking acts between the model and the event logs, acting on both. Enhancement acts from logs on the models, with a feedback from the model.

Lecture 2: Introduction to Process Mining

13.04.2020

Mathematical Background

Mathematical Preliminaries

Here we will set the grounds for the further lectures.

definition A **Set** is a group of objects without duplicates. definition

Sets allow for union, intersection, difference and complement operations.

Power of a set

$$P(X) = \{X' | X' \subset X\}.$$

remark The empty set $\emptyset$ is always a subset of every set. remark

Cartesian product

Given $X_1, \ldots, X_n$ sets,

$$X_1 \times \ldots \times X_n = \{(x_1, \ldots, x_n) \, | x_i \in X_i, \forall 1 \leq i \leq n\}.$$

Functions

Functions are necessary for dealing with sets.

Definition 8. Given sets X and Y, a set $f \in X \times Y$ is a **function** $\Leftrightarrow \forall x \in X, \exists y \in Y\,((x,y) \in f) \wedge \nexists! y' \in Y, y' \neq y\,((x,y') \in f)$

I.e., in a function $f : X \to Y$, there is one, and only one mapping for every element of X. Basic function definition.

Definition 9. $f : X \nrightarrow Y$ is a **partial function** $\Leftrightarrow \exists X' \in X \mid f : X' \to Y$ is a function.

I.e., a partial function map elements from a subset of the domain.

Remark. A function $f : X \to Y$ is said **injective** $\Leftrightarrow f^{-1} : Y \to X$ is a partial function.

Remark. A function $f : X \to Y$ is said **surjective** $\Leftrightarrow \forall y \in Y, \exists x \in X \mid (x,y) \in f$.

I.e., surjectiveness is coverage of Y.

Remark. A function is said **bijective** if it is both *injective* and *surjective*.

Multisets

Multisets are sets that allow for multiple elements. Usually written as $B = \left\{a^2, b^3, c, d\right\}$, given $X = \{a, b, c, d\}$. They are represented as a function.

Definition 10. A **Multiset** is a function $B : X \to \mathbb{N}$ that maps the elements of a set X in their amounts.

Multisets allow for several operations such as extension and union (across different base sets as well).

Sequences

Sequences are enumerated collections of objects. Given a set $X = \{a, b, c\}$ a sequence is usually written as $< a, b, c, d >$.

Definition 11. Given a set X, a **sequence** σ is a function $\sigma : \{1, \ldots, n\} \to X, n \in \mathbb{N}$.

Sequences are a subset of $\mathbb{N} \times X$.

Definition 12. Given a set X, X^* is the set of all possible sequences over X.

Sequences can be concatenated: $\sigma_1 =< a, b >, \sigma_2 =< c, d >, \sigma_1 \times \sigma_2 =< a, b, c, d >$.

Petri Nets

[ADD IMAGEM?]

Definition 13. A **Bipartite graph** is a graph $G = \{(v, e) \mid v \in V, e \in E \subset V \times V\}$, V represents the vertices (or nodes) and E represents the edges, where $\exists V_1, V_2 \subset V, V_1 \cap V_2 = \emptyset \wedge V_1 \cup V_2 = V$, such that $\forall e = (v_1, v_2) \in E, v_1 \in V_1, v_2 \in V_2$.

I.e., there is no edge between the elements of V_1 or V_2.

Definition 14. Given P, T sets of *places* and *transitions*, respectively, and $F \subset (P \times T) \cup (T \times P)$ the set of arcs, a tuple $N \in P \times T \times F$ is a **Petri Net** $\Leftrightarrow$ P and T are partitions of the graph $G = \{(v, e) \mid v \in P \cup T, e \in F\}$.

Let $N = (P, T, F)$ be a Petri Net, we can define

- $\cdot x = \{y \in P \cup T \mid (y, x) \in F\}$ as the pre-set of a vertex x.
- $x \cdot = \{y \in P \cup T \mid (x, y) \in F\}$ as the post-set of a vertex x.
- $M \in B(P)$ is called a marking of N
  - Given $p \in P$, we write $M(p)$ to denote the number of tokens in p.

Definition 15. A transition $t \in T$ is **enabled**, written $M[t\rangle$, $\Leftrightarrow$ $\forall p \in \cdot t, M(p) > 0$.

If a transition $t \in T$ is **fired**, we obtain a new marking $M' \in B(P)$ in the following way:

$$M'(p) = \begin{cases} M(p) + 1, & p \in t \cdot \setminus \cdot t \\ M(p) - 1, & p \in \cdot t \setminus \cdot t \\ M(p), & \end{cases}.$$