

Process Mining bei Softwareprozessen

Prof. Dr. Ralf Kneuper

PVM 2016



Prof. Dr. Ralf Kneuper

- ▶ Dipl.-Mathematiker, Univ. Bonn
- ▶ PhD Computer Science, Univ. of Manchester
- ▶ Professor für Wirtschaftsinformatik, Internationale Hochschule Bad Honnef · Bonn (IUBH)
- ▶ Freiberuflicher Berater für Softwarequalitätsmanagement und Prozessverbesserung
- ▶ (Ko-)Autor mehrerer Fachbücher zu CMMI und Qualitätsmanagement

- ▶ Kontakt: r.kneuper@iubh-fernstudium.de oder ralf@kneuper.de



Agenda

Process Mining

Process Mining und Softwareprozesse

Fallstudie Process Mining von SW-Entwicklungsprozessen

Zusammenfassung und Ausblick

Agenda

Process Mining

Process Mining und Softwareprozesse

Fallstudie Process Mining von SW-Entwicklungsprozessen

Zusammenfassung und Ausblick

- ▶ Disziplin innerhalb des (Business) Process Managements, die in den letzten Jahren sehr an Bedeutung gewonnen hat
 - ▶ Ursprünglich für Softwareprozesse entwickelt, heute hauptsächlich für Geschäftsprozesse eingesetzt
- ▶ Grundidee: Nutzung von Ereignisprotokollen (Event Logs), um Informationen über die zu Grunde liegenden Prozesse und deren Bezug zu Prozessmodellen abzuleiten
- ▶ Verwandt mit Data Mining und Business Intelligence, betrachtet aber speziell Analyse von Prozessdaten
 - ▶ Dadurch Einsatz spezieller Methoden und Algorithmen möglich
 - ▶ Prozessbezogene Fragestellungen

Nutzungsmöglichkeiten von Process Mining

Drei wesentliche Nutzungsmöglichkeiten:

- ▶ Entdeckung (*Discovery*)
 - ▶ Aus den verfügbaren Ereignisdaten wird der tatsächlich durchgeführte Prozess “entdeckt”
- ▶ Konformität (*Conformance*)
 - ▶ Abgleich des tatsächlich durchgeführten Prozesses gegen den definierten Prozess
- ▶ Verbesserung (*Enhancement*)
 - ▶ Aus verfügbaren Ereignisdaten werden Verbesserungspotentiale identifiziert, z.B. Engpässe

Format von Ereignisprotokollen (Event Logs)

- ▶ Prozess besteht aus Fällen
 - ▶ z.B. Softwareentwicklungsprozess besteht aus der Bearbeitung von Anforderungen
- ▶ Ein Fall besteht aus Ereignissen
 - ▶ z.B. Analyse, Genehmigung und Implementierung von Anforderungen
- ▶ Ereignisse sind innerhalb eines Falles geordnet, z.B. über Zeitstempel
 - ▶ Zeitstempel erlauben auch Zusammenführung mehrerer Ereignisprotokolle des gleichen Prozesses, z.B. aus verschiedenen Systemen
 - ▶ Zeitstempel sind hilfreich, aber grundsätzlich nicht notwendig
- ▶ Verbreitetes Standardformat: *eXtensible Event Stream (XES)*

Agenda

Process Mining

Process Mining und Softwareprozesse

Fallstudie Process Mining von SW-Entwicklungsprozessen

Zusammenfassung und Ausblick

Process Mining und Softwareprozesse

- ▶ Erste Veröffentlichungen zu Process Mining betrachteten Softwareentwicklung (Cook/Wolf 1995, 1998)
- ▶ Auf Softwareprozesse meist schwieriger anwendbar, da weniger strukturiert, geringere Fallzahlen
 - ▶ Aber gut anwendbar auf IT-Serviceprozesse, siehe z.B. *Business Process Intelligence Challenge (BPIC)* 2013, 2014

Process Mining und Softwareprozesse

Grundsätzlich drei Ansatzpunkte für die Nutzung von Process Mining bei Softwareprozessen:

- ▶ Anwendung auf Softwareprozesse selbst (hier Schwerpunkt)
- ▶ Anwendung auf die durch IT unterstützten bzw. zu unterstützenden Anwendungsprozesse, beispielsweise zur Analyse von Anforderungen
 - ▶ Einsatz vorab zur Identifikation von Anforderungen
 - ▶ Einsatz von frühen Versionen, um Schwierigkeiten zu identifizieren (insbesondere bei agiler Vorgehensweise geeignet)
- ▶ Anwendung auf Ausführung des Programmcodes als Prozess, beispielsweise zur Analyse und Steigerung der Performanz

Process Mining von SW-Entwicklungsprozessen

- ▶ Ansätze und Werkzeuge im Wesentlichen identisch mit denen für andere Geschäftsprozesse
- ▶ Besonderheiten: Datenquellen, Vorbereitung der Daten
 - ▶ KM-Werkzeuge wie Git, Subversion
 - ▶ Werkzeuge zum Anforderungsmanagement und Bug Tracker wie BugZilla, FlySpray
 - ▶ Ein "Ereignis" ist dann z.B. ein Commit, eine Statusänderung eines Tickets im Bug Tracker oder eine Zuweisung an anderen Verantwortlichen

Agenda

Process Mining

Process Mining und Softwareprozesse

Fallstudie Process Mining von SW-Entwicklungsprozessen

Zusammenfassung und Ausblick

Process Mining von SW-Entwicklungsprozessen — Fallstudie

Mehrere kleine Entwicklungsprojekte, die mit der gleichen Instanz des Bug-Tracking-Systems *FlySpray* arbeiten

- ▶ Prozess: Bearbeitung von Tickets in FlySpray
- ▶ Fall: Bearbeitung eines bestimmten Tickets
- ▶ Ereignis: In FlySpray festgehaltener Bearbeitungsschritt

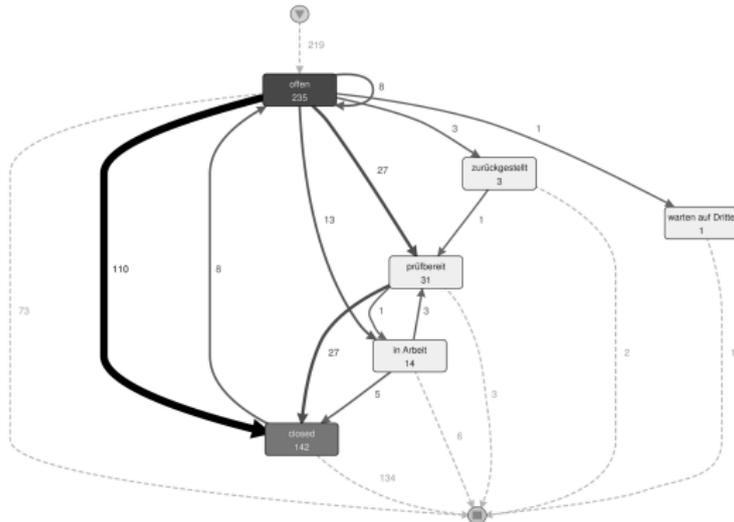
Fallstudie: Ereignisprotokoll

- ▶ Wie die meisten vergleichbaren Systeme verwendet Fly-Spray eine relationale Datenbank
- ▶ Extraktion der relevanten Daten über geeignete SQL-Abfrage
- ▶ Entstehende CSV-Datei konvertiert in XES-Format über Excel-Makro (XLS2XES)

Fallstudie: Ereignisprotokol im XES-Format

```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- XES version 1.0 -->
<!-- Created by Excel macro XLS2XES -->
<!-- Created 2016-05-24 17:10:54 -->
<!-- Columns: Case ID: b; Resource: (none);
Timestamp: c; Activity: e; Role ID: (none);
Life cycle transition: (none) -->
<log xes.version="1.0" xes.features="nested-attributes"
xmlns="http://www.xes-standard.org"
xes.creator="XLS2XES">
<extension name="Lifecycle" prefix="lifecycle"
uri="http://www.xes-standard.org/lifecycle.xesext" />
<extension name="Organizational" prefix="org"
uri="http://www.xes-standard.org/org.xesext" />
...
<string key="concept:name"
value="flyspray_history_20160524" />
<string key="lifecycle:model" value="standard" />
<trace>
  <string key="concept:name" value="1" />
  <event>
    <date key="time:timestamp"
value="2005-10-22T23:46:37" />
    <string key="concept:name" value="offen" />
  </event>
  <event>
    <date key="time:timestamp"
value="2010-12-18T20:22:31" />
    <string key="concept:name" value="closed" />
  </event>
  ...
  Process Mining
  ...
```

Fallstudie: Erzeugte Prozessdarstellung

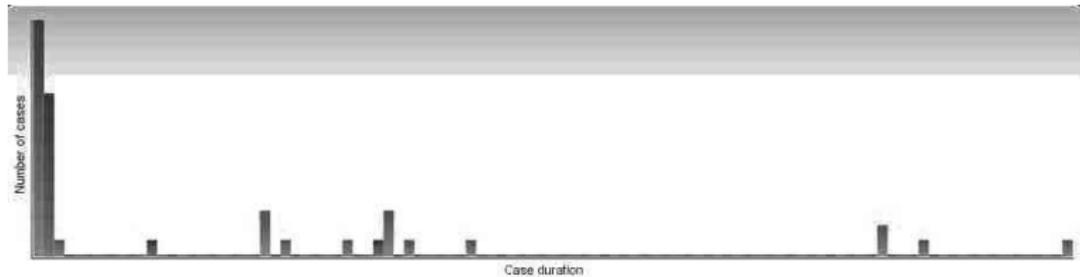


- ▶ “offen”, “prüfbereit” etc. sind hier genau genommen nicht die gleichnamigen Zustände, sondern die Ereignisse, dass ein Ticket diesen Zustand annimmt.

Fallstudie: Analyse der erzeugten Prozessdarstellung

- ▶ Tickets wurden oft nicht auf “prüfbereit” gestellt, bevor sie geschlossen wurden.
- ▶ Im Gegenteil wurde ca. die Hälfte aller Tickets direkt vom Zustand “offen” nach “closed” überführt.
- ▶ Klärung / Entscheidung über gewünschten Ablauf erforderlich

Fallstudie: Identifizierte Falldauer



Fallstudie: Analyse der identifizierten Falldauer

- ▶ Ein Teil der Tickets wurde sehr schnell erledigt, beim Rest ist die Bearbeitungsdauer sehr unterschiedlich
 - ▶ War zu erwarten, da Priorität und Aufwand der zu bearbeitenden Tickets ebenfalls sehr unterschiedlich
- ▶ Nächster Schritt: Prüfung, wie lange die Tickets in den einzelnen Zuständen verbleiben
- ▶ Ist für die meisten Zustände nicht relevant, aber Tickets dürfen nicht zu lange in einem der Zustände “in Arbeit” oder “prüfbereit” sein
- ▶ Konnte bis auf einzelne Ausnahmen bestätigt werden

Beispielsergebnisse einer anderen Fallstudie

- ▶ “Process Discovery”: Verschiedene Algorithmen liefern teilweise sehr unterschiedliche Prozesse aus den gleichen Ereignisdaten
- ▶ Konformität zu definiertem Prozess fast immer gegeben, da Ticketsystem keine anderen Prozesse zulässt
- ▶ Verbesserungspotentiale:
 - ▶ Tickets als “Projekthülse” verwendet, dementsprechend sehr lange in Bearbeitung, war aber nur durch Analyse der individuellen Tickets erkennbar. Verbesserung: neue Ticket-Kategorie, mit der man diese Tickets einfach herausfiltern kann.
 - ▶ Bestimmte Status wurden (fast) nie genutzt. a) Sonderfälle, die tatsächlich selten auftraten, aber als Status benötigt werden b) übersprungene Schritte, speziell Code Review

Agenda

Process Mining

Process Mining und Softwareprozesse

Fallstudie Process Mining von SW-Entwicklungsprozessen

Zusammenfassung und Ausblick

Zusammenfassung und Ausblick

- ▶ Process Mining ist für IT-Servicemanagementprozesse wie Incident Management problemlos anwendbar, bei Softwareentwicklungsprozessen mit Einschränkungen
 - ▶ Abhängig von Strukturierung und Häufigkeit
 - ▶ Bei iterativem Vorgehen einfacher, da kleinere Arbeitsschritte, häufigere Wiederholung
- ▶ Hilft beim Verständnis der tatsächlich durchgeführten Prozesse und der verschiedenen verwendeten Varianten