

Aktuelles Schlagwort: Process Mining

Rafael Accorsi, Meike Ullrich
Abteilung Telematik
Albert-Ludwigs-Universität Freiburg, Deutschland
accorsi@iig.uni-freiburg.de

Wil M. P. van der Aalst
Department of Industrial Engineering and Innovation Sciences
Eindhoven University of Technology, The Netherlands
w.m.p.v.d.aalst@tue.nl

1 Geschäftsprozessmanagement

Getrieben durch technologischen Fortschritt und digitale Vernetzung nimmt der Einsatz von Informationssystemen zur computergestützten und automatisierten Verarbeitung von betrieblichen Prozessabläufen zu. Informationssysteme wie beispielsweise Workflow-Management Systeme (WfMS), Enterprise-Resource-Planning Systeme (ERP) oder Customer-Relationship-Management Systeme (CRM) bilden das Rückgrat der meisten Organisationen. Sogenannte prozessgestützte Informationssysteme (engl. *process-aware information systems*, PAIS) steuern und führen operationale Prozesse auf der Basis von Prozessmodellen aus und tragen damit zu einer flexiblen und effizienten Prozessausführung bei. Während der Abwicklung von Prozessen durch Informationssysteme wird die Prozessausführung kontinuierlich in Logs festgehalten und protokolliert. Sie hinterlässt damit Spuren in den explosionsartig wachsenden Datenmengen des digitalen Universums.

Gleichzeitig unterliegt die Prozessausführung von verschiedenen Seiten einer Reihe nicht-funktioneller Anforderungen. Dies erfordert die Analyse von Prozessen hinsichtlich der Einhaltung solcher Anforderungen, was allgemein unter dem Begriff „Geschäftsprozessmanagement“ (engl. *Business Process Management*, BPM) subsumiert wird. Aus wirtschaftlicher Sicht steht die stetige Anpassung und Optimierung von Prozessen zur Reduzierung von Kosten im Vordergrund. Dazu kommen gesetzliche, datenschutzrechtliche, organisatorische und regulatorische Vorgaben, die im Zusammenhang mit der Prozessausführung eingehalten werden müssen (*Compliance*), beispielsweise interne Kontrollen nach Basel II oder dem Sarbanes-Oxley-Act (SOX). Notwendige Kontrollmaßnahmen in Bezug auf Compliance lassen sich bei gleichzeitig auftretendem Wunsch nach Flexibilität bei der Prozessausführung jedoch nicht vollständig und zuverlässig zur Design- und Laufzeit durchsetzen. Somit ist eine ex-post, also nachträglich stattfindende Analyse von großer Bedeutung, um beispielsweise ineffektive Kontrollen oder vorsätzliche Umgehungen aufdecken und zurechenbar machen zu können.

Aus diesen Anforderungen ergibt sich für Unternehmen zweifelsfrei ein großes Interes-

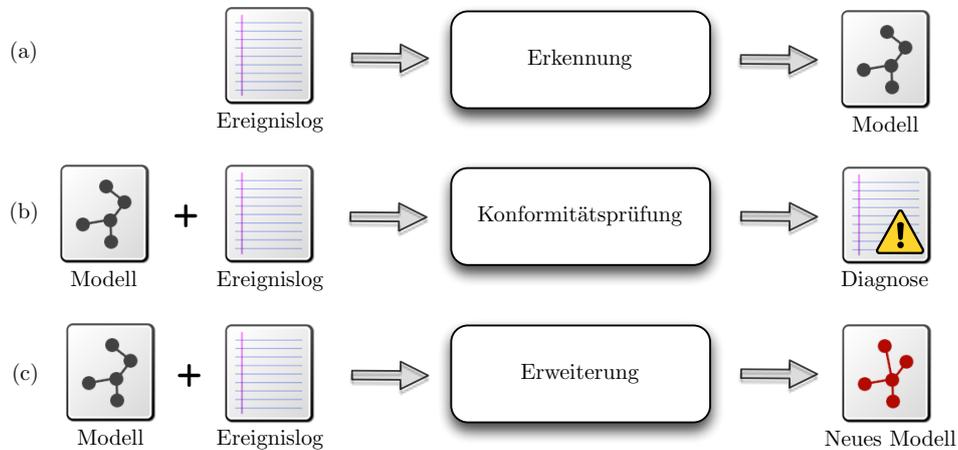


Abbildung 1: Drei Typen von Verfahren des Process Mining mit ihren Ein- und Ausgaben: (a) Erkennung, (b) Konformitätsprüfung und (c) Erweiterung [oPM11].

se daran, Einblick in die komplexen und damit auch wenig transparenten Prozessabläufe zu erhalten und zu erkennen, wie in Modellen idealisierte Prozesse („*to-be*“) tatsächlich ausgeführt werden („*as-is*“). Verfahren des Process Minings bieten diese Möglichkeit. Sie stützen sich dabei für die automatisierte Spurensuche bezüglich der Prozessausführung auf konkrete Fakten aus den zahlreich vorhandenen Logdaten und ersetzen dadurch traditionelle Methoden wie Fragebögen, Interviews oder Workshops mit ihren limitierten und vorwiegend idealisierten Angaben als Grundlage für die Analyse.

2 Process Mining

In Abgrenzung zum klassischen Data Mining fokussiert das Process Mining statt der Daten- auf die Prozessebene, um Einblicke aus einer prozessorientierten Sicht zu ermöglichen. Dazu sind die zu analysierenden Logs grundlegend aus *Ereignissen* aufgebaut, die designierten Aktivitäten im Prozess entsprechen. Gleichzeitig ist jedes Ereignis einem konkreten *Fall* (einer Prozessinstanz) zugeordnet. Die chronologische Abfolge der Ereignisse zu einem Fall stellt somit einen Pfad von Aktivitäten, d.h. einen *Ablauf* (engl. *Trace*) des Prozesses dar. Zusätzliche Informationen zu Ereignissen umfassen Angaben zu den Subjekten, die eine Aktivität ausführen, ihre Rollenzugehörigkeit, den Zeitstempel der Ausführung und die Charakteristika der im Prozess verarbeiteten Objekte, beispielsweise den Wert einer Kreditanfrage.

Auf der Grundlage solcherart gestalteter Ereignislogs bietet das Process Mining drei verschiedene Typen von Verfahren, wie in Abbildung 1 dargestellt. Ihr Ziel ist es, (a) dem Prozess zugrundeliegende Prozessmodelle aus den Logdaten zu extrahieren (*Erkennung*),

(b) die Übereinstimmung von Prozessmodellen oder Vorgaben mit dem durch die Logdaten dokumentierten Verhalten zu prüfen (*Konformitätsprüfung*) und (c) bestehende Prozessmodelle um zusätzliche Informationen zu erweitern und zu verbessern (*Erweiterung*).

Unter Anwendung der Verfahren des Process Minings können unterschiedliche Aspekte des Prozesses berücksichtigt werden. Beispiele sind der Kontrollfluss, d.h. die Abfolge von Aktivitäten, die Organisationsstruktur der Prozessteilnehmer, die verarbeiteten Datenobjekte oder auch die zeitliche Performanz der Prozessausführung.

Das theoretische Fundament für die konkrete Implementation einzelner Verfahren bilden Konzepte aus verschiedenen Teilgebieten der Informatik. Im Zentrum stehen Petri-Netze, welche eine ausdrucksstarke formale Grundlage zur Repräsentation von Prozessen bieten. Darauf aufbauend können „Token-Games“ verwendet werden, um die Konformität von Modell und Ereignislog festzustellen; LTL-basierte Verifikation, um nicht-konforme Ausführungen aufzuspüren; und Algorithmen aus dem Gebiet des maschinellen Lernens (z.B. *Genetic Mining*) zur Erkennung von Prozessen.

Erkennung. Process Mining-Verfahren zur Erkennung (engl. *Process Discovery*) konstruieren einzig und allein aus gegebenen Ereignislogs ohne Hinzunahme weiterer a-priori-Informationen ein Prozessmodell und ermöglichen somit einen unvoreingenommenen Einblick in die Struktur des Prozesses.

Ein Beispiel für ein Prozessmodell aus der Perspektive des Kontrollflusses stellt das in Abbildung 2 enthaltene Prozessmodell (a) dar. Dieses wurde mit Hilfe der Anwendung Disco (siehe Abschnitt 3) direkt aus einem Ereignislog für ein Szenario des Gesundheitswesens (siehe Tabelle 1) extrahiert. Anhand der so erzeugten Darstellung in Form eines Fuzzy-Modells kann die Abfolge der Prozessaktivitäten während der Prozessausführung sichtbar gemacht werden. Dabei wird über die verwendete Linienstärke einzelner Verbindungen deren Häufigkeit visualisiert, ebenso kann eine Animation des Prozessablaufs einzelner Fälle ausgeführt werden. Um Einsicht in die Organisationsstruktur des Prozesses zu erhalten, können Beziehungen als Kanten zwischen Prozessteilnehmern (Knoten) durch ein soziales Netzwerk dargestellt werden. Beispielsweise geben die Verbindungen in dem mit Hilfe der Anwendung ProM (siehe Abschnitt 3) erstellten Prozessmodell (b) aus Abbildung 2 an, dass Teilnehmer ähnliche Prozessaktivitäten ausführen. Dadurch können die Prozessteilnehmer gruppiert und z.B. Rückschlüsse auf die Rollenverteilung in der Organisation gezogen werden.

Bei der Anwendung dieser Art von Process Mining-Verfahren entfällt nicht nur der Aufwand für die manuelle Erstellung eines Prozessmodells, die extrahierten („de facto“) Prozessmodelle sind zudem wirklichkeitsgetreu und reflektieren den tatsächlichen Prozessablauf, da sie ausschliesslich auf Fakten und nicht auf Annahmen basieren. In den Punkten Geschwindigkeit und Genauigkeit bietet die Anwendung von Process Mining-Verfahren zur Erkennung demnach wesentliche Vorteile gegenüber traditionellen Ansätzen zur Erstellung von Prozessmodellen.

Konformitätsprüfung. Ein Prozessmodell kann *deskriptiven* oder *normativen* Charakter aufweisen. Im ersten Fall beschreibt das Modell, was tatsächlich passiert, während es

Instanz	Aktivität	Teilnehmer	Patient	Zeitstempel
452	<i>Anmeldung</i>	2	81	2011-12-24, 11:10:21
452	<i>Untersuchung</i>	9	81	2011-12-24, 11:15:21
452	<i>Beratung</i>	5	81	2011-12-24, 12:17:10
452	<i>Entlassung</i>	3	81	2011-12-24, 12:47:11
453	<i>Anmeldung</i>	2	42	2011-12-24, 11:16:35
453	<i>Untersuchung</i>	10	42	2011-12-24, 11:27:12
453	<i>Beratung</i>	7	42	2011-12-24, 11:52:37
453	<i>Entlassung</i>	2	42	2011-12-24, 11:59:54
454	<i>Anmeldung</i>	2	101	2011-12-24, 11:11:21
454	<i>Untersuchung</i>	10	101	2011-12-24, 11:15:21
454	<i>Aufnahme</i>	7	101	2011-12-24, 12:17:10
454	<i>Operation</i>	9	101	2011-12-24, 12:47:11
454	<i>Check</i>	7	101	2011-12-24, 12:47:11
454	<i>Entlassung</i>	2	101	2011-12-24, 13:29:11
453	<i>Anmeldung</i>	1	107	2011-12-24, 12:16:35
453	<i>Untersuchung</i>	9	107	2011-12-24, 13:09:10
453	<i>Beratung</i>	8	107	2011-12-24, 13:41:07
453	<i>Entlassung</i>	2	107	2011-12-24, 14:05:51

Tabelle 1: Auszug aus dem Ereignislog für ein Szenario des Gesundheitswesens. Die Ereignisse (Zeilen) sind nach a) Instanzen (also Fällen) und b) Zeitstempeln sortiert.

im zweiten Fall definiert, was passieren soll. Sowohl für von Hand entworfene Prozessmodelle, als auch für solche, die mit Verfahren zur Erkennung gewonnen werden, kann es wünschenswert sein, einen Abgleich mit der in den Ereignislogs dokumentierten Realität vorzunehmen.

Verfahren der Konformitätsprüfung ermöglichen es, Diskrepanzen zwischen dem modellierten und dem protokollierten Verhalten aufzuzeigen, wie in Abbildung 3 skizziert. Dabei können nicht übereinstimmende Elemente in Log und Modell lokalisiert und visuell hervorgehoben werden. Zudem ist eine Bestimmung des globalen Grads der Übereinstimmung (engl. *fitness*) mit Hilfe spezieller Metriken und entsprechender Verfahren möglich.

Werden Prozessvorgaben wie beispielsweise Sicherheitsanforderungen in Form von Regeln über Prozesseigenschaften modelliert, kann ihre Einhaltung automatisiert überprüft und von den Vorgaben abweichende Fälle aufgezeigt werden [AS12]. Zur Formalisierung und anschließenden Verifikation von Anforderungen kommen dabei Sprachen auf Grundlage von z.B. linearer temporaler Logik oder Prädikatenlogik erster Stufe zum Einsatz. Das Process Mining verspricht aus diesem Gesichtspunkt großes Potential für Unternehmen durch eine verbesserte Unterstützung im Bereich interner Revision und Audit, sowie der Compliance-Prüfung.

Erweiterung. Das wesentliche Ziel von Process Mining-Verfahren des Typs Erweiterung (engl. *Enhancement*) ist es, ein gegebenes Prozessmodell auf Basis der Informationen aus den Ereignislogs entweder (a) zu erweitern oder (b) zu korrigieren, um dadurch ein aussagekräftigeres oder optimiertes Prozessmodell zu erhalten.

Ein auf diese Art erzeugtes und erweitertes Prozessmodell kann Unternehmen wertvolle Informationen liefern. Ein Beispiel dafür bietet die Analyse von Entscheidungspunkten (engl. *decision point analysis*). Hierbei wird eine Grundidee aus dem Bereich des maschinellen Lernens erfolgreich umgesetzt. Sind im Prozessmodell Entscheidungspunkte, d.h. Stellen mit exklusiven Verzweigungen in verschiedene Ausführungspfade enthalten, können für diese Punkte Geschäftsregeln ermittelt werden, die angeben, unter welchen Voraussetzungen für einen Fall ein bestimmter Pfad im Prozessmodell eingeschlagen wird. Ein weiteres typisches Beispiel berührt das Gebiet der Performanzanalyse. Mit dieser werden für die Ausführung der Aktivitäten des gegebenen Prozessmodells sogenannte Key Performance Indicators (KPIs) ermittelt. Durch die Angabe von minimalen, maximalen und durchschnittlichen Ausführungszeiten können z.B. Wartezeiten oder Engstellen in Prozessen identifiziert und die Prozessausführung optimiert werden.

3 Tool Support

Ein Großteil der bisher entwickelten Process Mining-Verfahren ist im Open-Source Rahmenwerk ProM¹ implementiert und somit frei verfügbar. ProM wird an der Technischen Universität Eindhoven (TU/e) verwaltet. Die ProM-Architektur basiert auf der Einbindung von Plugins, so dass Entwickler neuer Verfahren auf eine einheitliche Umgebung und gemeinsame Funktionen zurückgreifen können. ProM hat sich dadurch als de facto Standard für Process Mining vor allem in der Wissenschaftsgemeinde etabliert. Die Anwendung Disco² der Firma Fluxicon ermöglicht den professionellen Einsatz von Process Mining-Verfahren durch Endanwender, wobei die Nutzung dieser Werkzeuge auch im akademischen Umfeld erfolgen kann.

Zunehmend werden auch kommerzielle Anwendungen aus dem Bereich des Prozessmanagements um Analyse- und Diagnosefunktionen erweitert, die in enger Beziehung zum Process Mining stehen. Beispiele dafür liefern unter anderem die Anwendungen ARIS Process Performance Manager (Software AG), Enterprise Visualization Suite (Businesscape), Interstage BPME (Fujitsu), Reflectone (Pallas Athena/Preceptive) und Reflect (Futura Process Intelligence/Preceptive). In diesem Kontext kann das Process Mining auch als eine Reihe von Methoden zur Realisierung von Business Intelligence (BI) oder auch Business Process Intelligence (BPI) verstanden werden.

¹<http://www.processmining.org/prom/start>

²<http://www.fluxicon.com/disco/>

4 Der Weg zur Einsatzfähigkeit in einem „BPM-in-the-large“-Anwendungskontext

Gegenwärtig gehen Verfahren des Process Minings von einem sogenannten „BPM-in-the-small“-Anwendungskontext aus. Ein solcher sieht beispielsweise vor, dass Prozesse in einer einheitlichen Modellierungssprache erfasst werden, eine unveränderliche Struktur aufweisen und nur über geringfügige Abhängigkeiten zu anderen Prozessen verfügen. Dem gegenüber steht der sogenannte „BPM-in-the-large“-Anwendungskontext [HFL⁺11], bei dem diese Annahmen aufgehoben werden. „BPM-in-the-large“ baut u.a. auf kollaborative und heterogene Modellierung sowie konfigurierbare und flexible Prozesse auf, was den tatsächlichen Gegebenheiten in der Praxis entspricht. Aufgrund der exorbitanten Menge an Daten, die in einem solchen Anwendungskontext generiert werden, ist er gleichzeitig charakteristisch für das viel beachtete „Big Data“-Szenario [MCB⁺11].

Um in diesem Zusammenhang aus Ereignislogs entscheidende Verbesserungen mittels Process Mining erzielen zu können, müssen die derzeitigen Ansätze weiterentwickelt und dabei einsatzfähig gestaltet werden. Die gegenwärtigen Herausforderungen umfassen beispielsweise die Vorbereitung des Ereignislogs (im Sinne einer Konsolidierung und Bereinigung), die Erstellung aussagekräftiger Benchmarks zur Evaluation der Verfahren und die Erkennung von sogenannten „Konzeptverschiebungen“, d.h. laufenden Änderungen an der Prozessstruktur. Überdies muss der Fokus derzeitiger Techniken verstärkt auf Datenobjekte im Prozess gerichtet werden, z.B. um den Fluss sensitiver Daten und Informationen während der Prozessausführung darstellbar zu machen.

Fazit. Process Mining ist ein mächtiges Werkzeug für moderne Unternehmen zum Management nicht-trivialer operativer Prozesse. Das Einsatzgebiet ist dabei nicht auf Geschäftsprozesse begrenzt – Process Mining lässt sich auf jeden Prozess anwenden, der Spuren in Logs hinterlässt. Neue Algorithmen und Verfahren machen es möglich, immer komplexer werdende Ereignisdaten zu analysieren und auf dieser Grundlage Prozesse effizienter und sicherer zu gestalten.

Weitere Informationen zu Process Mining – einschließlich relevanter Publikationen, Anwendungen und beispielhafter Ereignislogs für Versuchszwecke – sind unter <http://www.processmining.org> erhältlich.

Literatur

- [AS12] Rafael Accorsi und Thomas Stocker. On the Exploitation of Process Mining for Security Audits: The Conformance Checking Case. *ACM Symposium on Applied Computing*, 2012.
- [HFL⁺11] Constantin Houy, Peter Fettke, Peter Loos, Wil M. P. van der Aalst und John Krogstie. Business Process Management in the Large. *Business & Information Systems Engineering*, 3(6):385–388, 2011.

- [MCB⁺11] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh und Angela H. Byers. Big data: The next frontier for innovation, competition, and productivity, 2011.
- [oPM11] IEEE Task Force on Process Mining. Process Mining Manifesto. In Florian Daniel, Kamel Barkaoui und Schahram Dustdar, Hrsg., *Business Process Management Workshops (1)*, Jgg. 99 of *Lecture Notes in Business Information Processing*, Seiten 169–194. Springer, 2011.
- [vdA11] Wil M. P. van der Aalst. *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

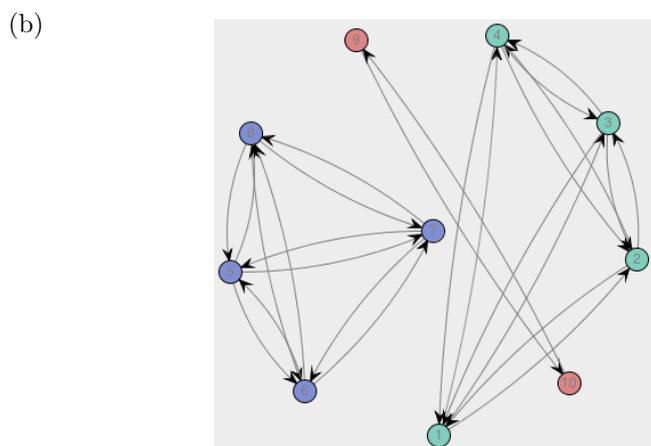
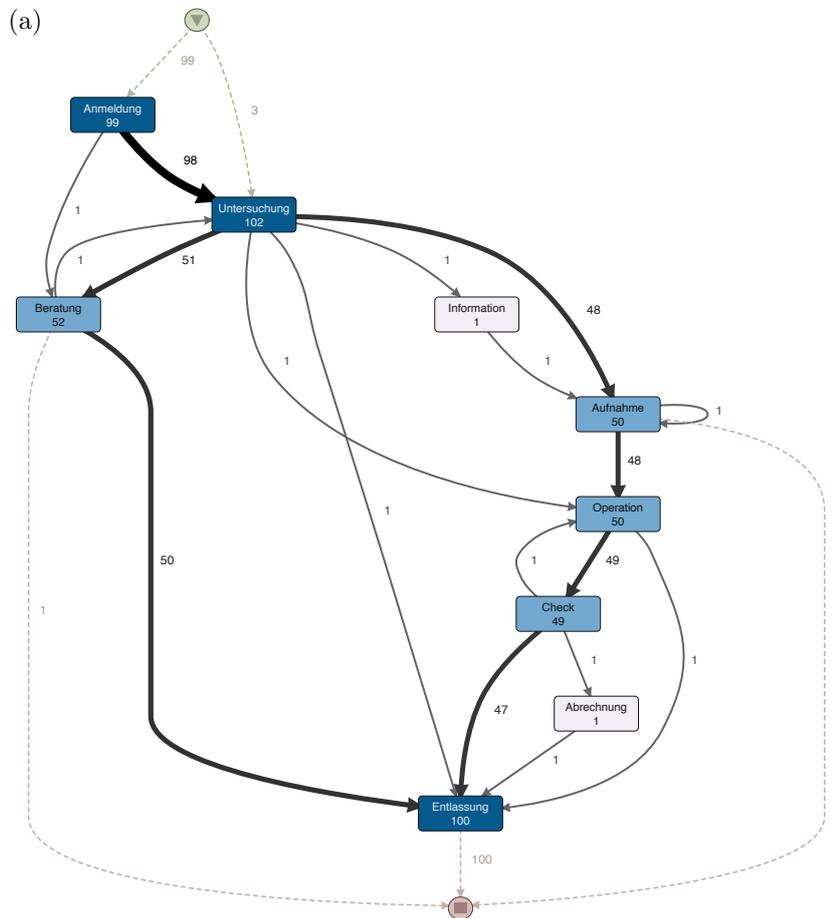


Abbildung 2: Aus einem Ereignislog eines Szenarios des Gesundheitswesens extrahierte Prozessmodelle: der Kontrollfluss wird durch das Fuzzy-Modell in (a) dargestellt, das soziale Netzwerk in (b) reflektiert Prozessteilnehmer die ähnliche Aktivitäten durchführen.

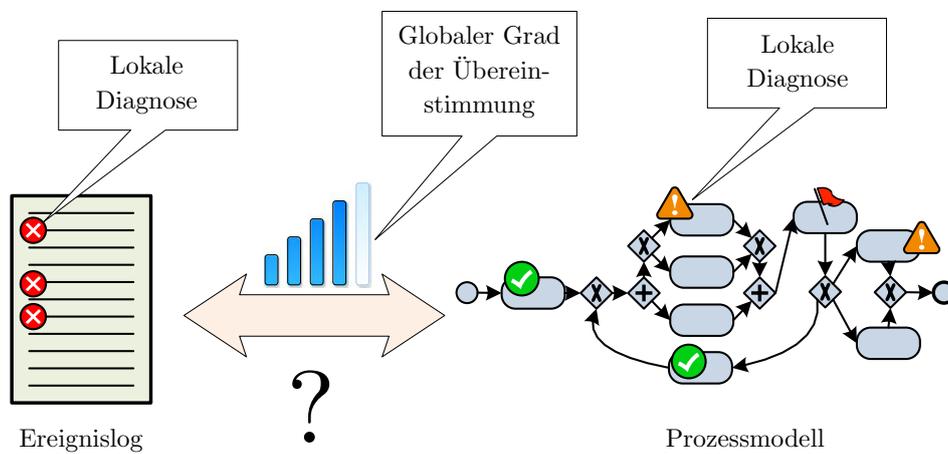


Abbildung 3: Konformitätsprüfung: Abgleich von Ereignislog und Prozessmodell. Dabei kann der Grad der Übereinstimmung global bestimmt werden, während die lokale Diagnose die nicht übereinstimmenden Elemente in Log und Modell hervorhebt [vdA11, S. 192].