

Process Mining Manifest

Ein *Manifest* ist eine "öffentliche Erklärung von Zielen und Absichten", getragen von einer Gruppe von Personen. Das vorliegende Manifest wurde von Mitgliedern und Unterstützern der *IEEE Task Force on Process Mining* verfasst. Das Ziel dieser Task Force ist die Förderung der Forschung, der Entwicklung, der Ausbildung, der Umsetzung, der Fortentwicklung und des Verständnisses rund um Process Mining.

Process Mining ist eine vergleichsweise junge Wissenschaftsdisziplin, angesiedelt zwischen Computational Intelligence und Data Mining auf der einen Seite und Prozessmodellierung und Analyse auf der anderen. Die Grundidee von Process Mining ist es, *reale Prozesse* (im Gegensatz zu vermuteten oder angenommenen Prozessen) *durch Extrahieren von Wissen aus Ereignislogs* heutiger (Informations-)systeme zu erkennen, zu überwachen und zu verbessern. Abbildung 1 skizziert diesen Anspruch. Process Mining umfasst das (automatische) Erkennen von Prozessen (d.h. die Konstruktion von Prozessmodellen für ein Ereignislog), die Übereinstimmungsprüfung (d.h. die Überwachung von Abweichungen durch den Vergleich von Modell und

Log), das Erkennen von sozialen bzw. organisatorischen Netzwerken, die automatische Erstellung von Simulationsmodellen, die Erweiterung und Verbesserung von Modellen, das Treffen von Vorhersagen für Fälle, sowie das Ableiten von Empfehlungen auf Basis der Fallhistorie.

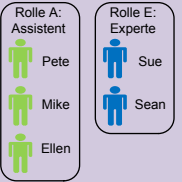
Inhalt:

Process Mining: Stand der Forschung	3
Leitsätze	7
Herausforderungen	11
Epilog	15
Glossar	16

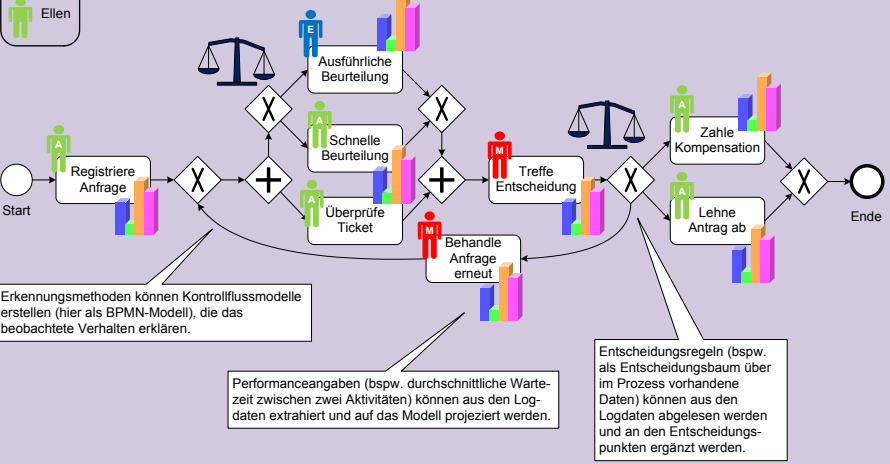
Mit Hilfe von Methoden des Process Mining können Einsichten aus Ereignisdaten gewonnen werden, welche heutzutage in vielen Informationssystemen anfallen. Diese Methoden eröffnen neue Möglichkeiten, um Prozesse in einer Vielzahl von Anwendungsszenarien zu erkennen, zu überwachen und zu verbessern. Es gibt zwei wesentliche Gründe für das wachsende Interesse an Process Mining. Einerseits werden immer mehr Ereignisse informationstechnisch erfasst, und somit wird die Historie eines Prozesses im Detail aufgezeichnet. Andererseits besteht das Bedürfnis, Prozesse bei steigender Wettbewerbsdynamik zu verbessern und zu unterstützen. Dieses Manifest wurde von der *IEEE Task Force on Process Mining* verfasst. Es zielt darauf ab, den Bekanntheitsgrad von Process Mining zu erhöhen. Zudem sollen die aufgeführten Leitprinzipien und Herausforderungen als Anregung für Forscher, Entwickler, Berater, Manager und Endanwender dienen. Das Ziel ist dabei die Anwendbarkeit von Process Mining als Werkzeug der Prozessverbesserung, -überwachung und -unterstützung weiter zu verbessern.

Ausgangspunkt sind Ereignislogdaten. Jedes Ereignis bezieht sich auf eine Prozessinstanz (Fall) und eine Aktivität. Ereignisse sind geordnet. Weitere Eigenschaften (Zeitstempel oder Rollen) können vorhanden sein.

Die Ereignislogdaten können benutzt werden, um Rollen in der Organisation zu identifizieren (bspw. Gruppen von Personen, die ähnliche Arbeit verrichten). Diese Rollen drücken die Beziehung zwischen Personen und Aktivitäten aus.



Fall ID	Ereignis ID	Eigenschaften			
		Zeitstempel	Aktivität	Rolle	Kosten
1	35654423	30-12-2010:11.02	Registrierte Anfrage	Pete	50
	35654424	31-12-2010:10.06	Ausführliche Beurteilung	Sue	400
	35654425	05-01-2011:15.12	Überprüfte Ticket	Mike	100
	35654426	06-01-2011:11.18	Treffe Entscheidung	Sara	200
2	35654427	07-01-2011:14.24	Lehne Antrag ab	Pete	200
	35654483	30-12-2010:11.32	Registrierte Anfrage	Mike	50
	35654485	30-12-2010:12.12	Überprüfte Ticket	Mike	100
	35654487	30-12-2010:14.16	Schnelle Beurteilung	Pete	400
	35654488	05-01-2011:11.22	Treffe Entscheidung	Sara	200
	35654489	08-01-2011:12.05	Zahle Kompensation	Ellen	200
3	35654521	30-12-2010:14.32	Registrierte Anfrage	Pete	50
	35654522	30-12-2010:15.06	Schnelle Beurteilung	Mike	100
	35654524	30-12-2010:16.34	Überprüfte Ticket	Ellen	400
	35654525	06-01-2011:09.18	Treffe Entscheidung	Sara	200
	35654526	06-01-2011:12.18	Behandle Anfrage erneut	Sara	200
	35654527	06-01-2011:13.06	Ausführliche Beurteilung	Sean	400
	35654530	08-01-2011:11.43	Überprüfte Ticket	Pete	100
4	35654531	09-01-2011:09.55	Treffe Entscheidung	Sara	200
	35654533	15-01-2011:10.45	Zahle Kompensation	Ellen	200
	35654641	06-01-2011:15.02	Registrierte Anfrage	Pete	50
5	35654643	07-01-2011:12.06	Überprüfte Ticket	Mike	100
	35654644	08-01-2011:14.43	Ausführliche Beurteilung	Sean	400
	35654645	09-01-2011:12.02	Treffe Entscheidung	Sara	200
	35654647	12-01-2011:15.44	Lehne Antrag ab	Ellen	200
6	35654711	06-01-2011:09.02	Registrierte Anfrage	Ellen	50
	35654712	07-01-2011:10.16	Schnelle Beurteilung	Mike	400
	35654714	08-01-2011:11.22	Überprüfte Ticket	Pete	100
	35654715	10-01-2011:13.28	Treffe Entscheidung	Sara	200
	35654716	11-01-2011:16.18	Behandle Anfrage erneut	Sara	200
	35654718	14-01-2011:14.33	Überprüfte Ticket	Ellen	100
	35654719	16-01-2011:15.50	Schnelle Beurteilung	Mike	400
	35654720	19-01-2011:11.18	Treffe Entscheidung	Sara	200
	35654721	20-01-2011:12.48	Behandle Anfrage erneut	Sara	200
	35654722	21-01-2011:09.06	Schnelle Beurteilung	Sue	400
7	35654724	21-01-2011:11.34	Überprüfte Ticket	Pete	100
	35654725	23-01-2011:13.12	Treffe Entscheidung	Sara	200
	35654726	24-01-2011:14.56	Lehne Antrag ab	Mike	200
	35654871	06-01-2011:15.02	Registrierte Anfrage	Mike	50
	35654873	06-01-2011:16.06	Schnelle Beurteilung	Ellen	400
	35654874	07-01-2011:12.22	Überprüfte Ticket	Mike	100
8	35654875	07-01-2011:14.52	Treffe Entscheidung	Sara	200
	35654877	16-01-2011:11.47	Zahle Kompensation	Mike	200



Erkennungsmethoden können Kontrollflussmodelle erstellen (hier als BPMN-Modell), die das beobachtete Verhalten erklären.

Performanceangaben (bspw. durchschnittliche Wartezeit zwischen zwei Aktivitäten) können aus den Logdaten extrahiert und auf das Modell projiziert werden.

Entscheidungsregeln (bspw. als Entscheidungsbaum über im Prozess vorhandene Daten) können aus den Logdaten abgelesen werden und an den Entscheidungspunkten ergänzt werden.

Abbildung 1: Process Mining Techniken extrahieren Informationen aus Ereignislogdaten mit denen Prozesse erkannt, überwacht und verbessert werden können.

Die Absichten der Task Force sind insbesondere:

- 1) Endnutzer, Entwickler, Berater, Manager und Wissenschaftler über den neuesten Stand der Technik im Bereich des Process Mining zu informieren,
- 2) die Nutzung von Process-Mining-Techniken und - Werkzeugen voranzutreiben und neue Anwendungsgebiete zu fördern,
- 3) das Engagement in Standardisierungsinitiativen für das Aufzeichnen von Ereignisdaten,
- 4) die Organisation von Tutorien, Tagungen, Workshops, Diskussionsrunden, und
- 5) das Veröffentlichen von Artikeln, Büchern, Videos und Special Issues von Zeitschriften.

den Fall des Compliance Managements. Process-Mining-Techniken erlauben eine strenge Überprüfung von Compliance, sowie die Beurteilung der Aussagekraft und Zuverlässigkeit von Angaben über die Kernprozesse einer Organisation.

Im Laufe der letzten Jahre sind Ereignisdaten jederzeit verfügbar geworden und die Process-Mining-Techniken wurden erheblich weiter entwickelt. Wie bereits angesprochen können Ansätze zur Prozessverbesserung (zum Beispiel Six Sigma, TQM, CPI und CPM) und zum Compliance Management (SOX, BAM, etc.) von Process Mining profitieren. Erfreulicherweise sind Process-Mining-Algorithmen bereits in einer Vielzahl von akademischen und kommerziellen Systemen implementiert worden. Etliche Wissenschaftler widmen sich aktiv dem Process Mining, welches eines der "heißen Themen" der Forschung im Bereich des Geschäftsprozessmanagements geworden ist. Auch von Seiten der Industrie gibt es ein großes Interesse an Process Mining. Mehr und mehr Softwarehersteller erweitern ihre Werkzeuge mit Process-Mining-Funktionalität. Beispiele für Softwareprodukte mit Process-Mining-Fähigkeiten sind: ARIS Process

Process Mining stellt eine wichtige Brücke zwischen Data Mining und der Geschäftsprozessmodellierung und Analyse dar. Unter dem Slogan Business Intelligence (BI) wurde eine Vielzahl von Modewörtern eingeführt um einfache Berichts- und Dashboard-Werkzeuge zu vermarkten. Business Activity Monitoring (BAM) ist eine Technik, welche das Überwachen von Geschäftsprozessen in Echtzeit erlaubt. Complex Event Processing (CEP) bezeichnet Techniken, welche die Verarbeitung von großen Mengen an Ereignissen erlauben, um einen Geschäftsprozess in Echtzeit zu überwachen, zu steuern und zu verbessern. Corporate Performance Management (CPM) ist ein weiteres Modewort für das Messen der Leistung eines Prozesses oder einer Organisation. Darüberhinaus gibt es einen Bezug zu Managementansätzen,

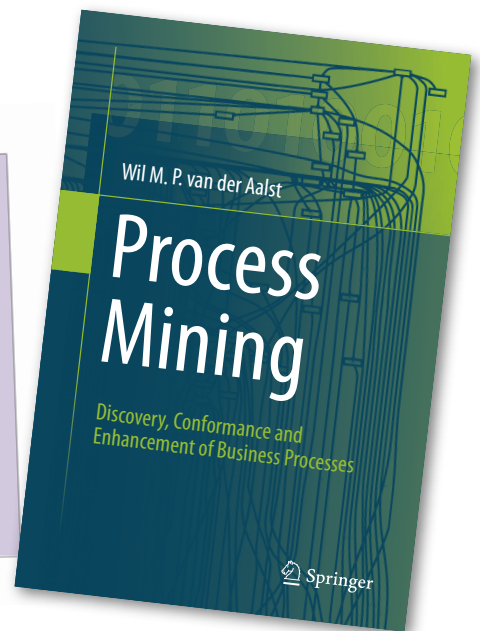
wie dem Continuous Process Improvement (CPI), Business Process Improvement (BPI), Total Quality Management (TQM) und Six Sigma. Diese Ansätze zeichnen sich dadurch aus, dass ein Prozess "unter dem Mikroskop begutachtet wird" um weitere Verbesserungen zu identifizieren. Process Mining ist somit eine Technik, welche den Einsatz von CPM, BPI, TQM, Six Sigma und verwandten Ansätzen ermöglicht. Während BI-Werkzeuge und Managementansätze wie Six Sigma und TQM auf die Verbesserung der operationellen Leistung abzielen, zum Beispiel die Verringerung von Flusszeiten und Defekten, rücken auch Corporate Governance, Risiken und Compliance verstärkt in den Fokus von Organisationen. Gesetzesinitiativen wie der Sarbanes-Oxley-Act (SOX) und die Basel-II-Vorschriften illustrieren dies für

Performance Manager (Software AG), Comprehend (Open Connect), Discovery Analyst (Stereologic), Flow (Fourspark), Futura Reflect (Futura Process Intelligence), Interstage Automated Process Discovery (Fujitsu), OKT Process Mining suite (Exeura), Process Discovery Focus (Iontas/Verint), ProcessAnalyzer (QPR), ProM (TU/e), Rbminer/Dbminer (UPC) und Reflec|one (Pallas Athena). Dieses wachsende Interesse an Log-basierter Prozessanalyse führte zu der Gründung einer Task Force zum Thema Process Mining.

Die Task Force wurde im Jahr 2009 im Rahmen des Data Mining Technical Committee (DMTC) der Computational Intelligence Society (CIS) des Institute of Electrical and Electronic Engineers (IEEE) ins Leben gerufen. Die aktuellen Mitglieder der Task Force repräsentieren Softwarehersteller (zum Beispiel Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Stereologic), Beratungshäuser/Endnutzer (zum Beispiel ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chili, BWI Systeme GmbH, Excellentia BPM, Rabobank) und wissenschaftliche Einrichtungen (zum Beispiel TU/e, University of Padua, Universität

Process Mining Buch

www.processmining.org/book/
W.M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.



Politécnica de Catalunya, New Mexico State University, IST - Technical University of Lisbon, University of Calabria, Penn State University, University of Bari, Humboldt-Universität zu Berlin, Queensland University of Technology, Wirtschaftsuniversität Wien, Stevens Institute of Technology, University of Haifa, University of Bologna, Ulsan National Institute of Science and Technology, Cranfield University, K.U. Leuven, Tsinghua University, University of Innsbruck, University of Tartu).

Seit der Gründung im Jahr 2009 wurden zahlreiche Aktivitäten hinsichtlich dieser Ziele unternommen. Workshops und spezielle Tracks auf Konferenzen wurden von der Task

Force mit organisiert, zum Beispiel die Business Process Intelligence Workshop Serie (BPI'09, BPI'10 und BPI'11) sowie Tracks auf IEEE Konferenzen (z.B. CIDM'11). Process-Mining-Kenntnisse wurden durch Tutorien (z.B. WCCI'10 und PMPM'09), Sommerschulen (ESSCaSS'09, ACPN'10, CICH'10, etc.), Videos (siehe auch www.processmining.org) und etliche Publikationen, wie das erste Buch über das Thema, welches kürzlich im Springer Verlag erschien^[1], verbreitet. Die Task Force hat auch die erste Business Process Intelligence Challenge (BPIC'11) mitorganisiert: ein Wettbewerb, bei dem die Teilnehmer aussagekräftige Informationen aus einem großen, komplexen Ereignislog gewinnen sollten. Weiterhin hat die Task Force im Jahr 2010 XES (www.xes-standard.org) standardisiert, ein Standardformat für das Aufzeichnen von Ereignissen. Es zeichnet sich durch Erweiterbarkeit aus und wird von der OpenXES library (www.openxes.org) und Werkzeugen wie ProM, XESame, Nitro, etc. unterstützt.

Weitere Informationen zur Task Force sind unter <http://www.win.tue.nl/ieeefpm/> zu finden.

2. Process Mining: Stand der Forschung

Die wachsenden Möglichkeiten, die Informationssysteme und anderen computergestützte Anwendungen bieten, werden treffend von Moores Gesetz charakterisiert. Gordon Moore, Mitgründer von Intel, sagte 1965 voraus, dass die Anzahl der Komponenten in integrierten Schaltkreisen sich jedes Jahr verdoppelt

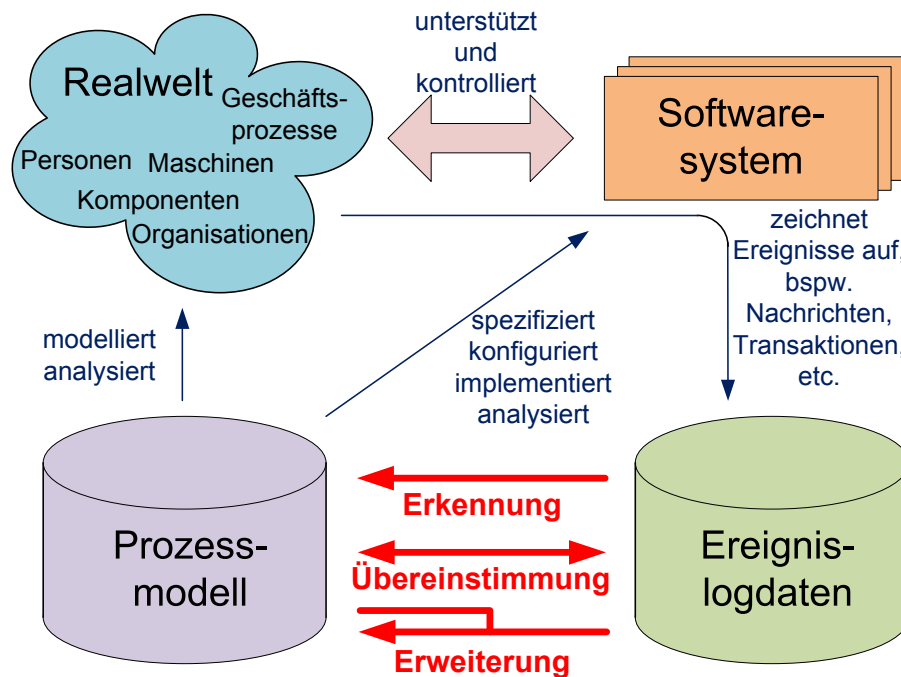


Abbildung 2: Positionierung der drei wichtigsten Typen des Process Mining: (a) Erkennung, (b) Übereinstimmungsprüfung und (c) Erweiterung.

würden. In den vergangenen 50 Jahren war das Wachstum tatsächlich exponentiell, allerdings etwas weniger schnell als vorhergesagt. Dieser Fortschritt ebnete den Weg für ein spektakuläres Wachstum der digitalen Welt und der elektronisch gespeicherten Daten. Zudem sind die digitale und die reale Welt mehr und mehr miteinander verwoben.

Das Wachstum dieser digitalen Welt, welches mit den Prozessen in modernen Organisationen eng verbunden ist, ermöglicht es Ereignisse aufzuzeichnen und zu analysieren. Ereignisse beziehen sich auf eine Vielzahl von Vorgängen wie das Abheben von Geld an einem Geldautomat, das Ausrichten eines Röntgengeräts durch einen Arzt, das Beantragen eines neuen Führerscheins, die Einreichung einer Steuererklärung oder der Empfang eines elektronischen Tickets für einen Flugreisenden. Die Herausforderung besteht darin, diese Ereignisdaten in einer geeigneten Form nutzbar zu machen, zum Beispiel um Vorgänge besser zu verstehen, Engpässe zu identifizieren, Probleme zu antizipieren, Vertragsverletzungen aufzuzeigen, Gegenmaßnahmen vorzuschlagen und Prozesse zu verschlanken. Process Mining zielt genau auf diese Art von Fragen ab.

Ausgangspunkt für Process Mining sind Ereignislogdaten. Alle Methoden des Process Mining gehen davon aus, dass diese Daten die zeitliche Reihenfolge der Ereignisse wiedergeben, und dabei jedes Ereignis einer Aktivität (Prozessschritt) und einem spezifischen Fall (Prozessinstanz) zugeordnet ist. Ereignislogdaten können weitere Informationen über Ereignisse enthalten. In der Tat nutzen Process-Mining-Methoden, wo immer möglich, solche Extrainformationen zu Ressourcen (etwa Personen oder Geräte), die die Aktivität durchführen, den Zeitstempel des Ereignisses oder zu dem Ereignis gehörende Datenelemente (bspw. Auftragsvolumen).

Abbildung 2 zeigt, dass Ereignislogdaten für drei Arten des Process Mining genutzt werden können. Die erste Art ist *Erkennung*. Eine Erkennungsmethode nimmt Ereignislogdaten als Eingabe und erzeugt ein Modell ohne dafür weitere Informationen zu benötigen. Prozesserkennung ist die bekannteste

Process-Mining-Methode. Für viele Organisationen ist es überraschend zu sehen, dass bestehende Methoden in der Tat echte Prozessmodelle ableiten können, obwohl nur beispielhafte Ausführungen in Form von Logdaten vorliegen. Die zweite Form von Process Mining ist die *Übereinstimmungsprüfung*. Hierbei wird ein bestehendes Prozessmodell mit zugehörigen Ereignislogdaten verglichen. Die Übereinstimmungsprüfung kann verwendet werden, um zu bestimmen, inwiefern die in den Logdaten dokumentierte Realität tatsächlich mit dem Modell übereinstimmt. Es können hierfür verschiedenen Modelle genutzt werden: Die Übereinstimmungsprüfung kann sich auf Ablaufmodelle, Organigramme, deklarative Modelle, Geschäftsregeln, Richtlinien etc. beziehen. Der dritte Typ des Process Mining ist die *Erweiterung*. Dabei ist die Idee, ein Prozessmodell auf Basis der Informationen aus den Ereignislogdaten zu erweitern und zu verbessern. Während die Übereinstimmungsprüfung den Unterschied zwischen Realität und Modell misst, zielt diese dritte Art des Process Mining darauf ab das vorliegende Modell zu verändern. Zum Beispiel können anhand der Zeitstempel Engpässe, Durchlaufzeiten und Häufigkeiten im Modell angezeigt werden.

Abbildung 3 zeigt die drei Typen des Process Mining mit Blick auf ihre Eingabe und Ausgabe. Methoden der Erkennung nutzen Ereignislogdaten und

liefern ein Modell. Das erzeugte Modell ist typischerweise ein Prozessmodell (bspw. ein Petri-Netz, UML-Aktivitätsdiagramm, BPMN- oder EPK-Modell), kann aber auch andere Perspektiven beschreiben (zum Beispiel ein soziales Netzwerk). Methoden der Übereinstimmungsprüfung benötigen Ereignislogdaten und ein Modell als Eingabe. Die Ausgabe umfasst diagnostische Informationen mit Blick auf Unterschiede und Übereinstimmungen zwischen Modell und Logdaten. Methoden der Erweiterung brauchen ebenfalls Modell und Logdaten als Eingabe. Die Ausgabe ist ein erweitertes Modell.

Process Mining kann weiterhin verschiedene Perspektiven abdecken. Die *Kontrollflussperspektive* betrachtet den Kontrollfluss, sprich die Abfolge der Aktivitäten. Das Ziel beim Mining dieser Perspektive besteht darin, eine treffende Beschreibung aller Ausführungspfade zu finden. Das Ergebnis wird typischerweise als Petri-Netz oder in Form eines ähnlichen Prozessmodells ausgedrückt (bspw. EPK, BPMN oder UML-Aktivitätsdiagramm). Die *Organisationsperspektive* stellt Informationen über Ressourcen in Logdaten heraus. Dies können verschiedene Arten von Akteuren sein, bspw. Personen, Rollen oder Organisationseinheiten, und deren Beziehungen zueinander. Das Ziel ist es hierbei entweder die Struktur der Organisation aufzudecken, in dem man Personen den zugehörigen Rollen und

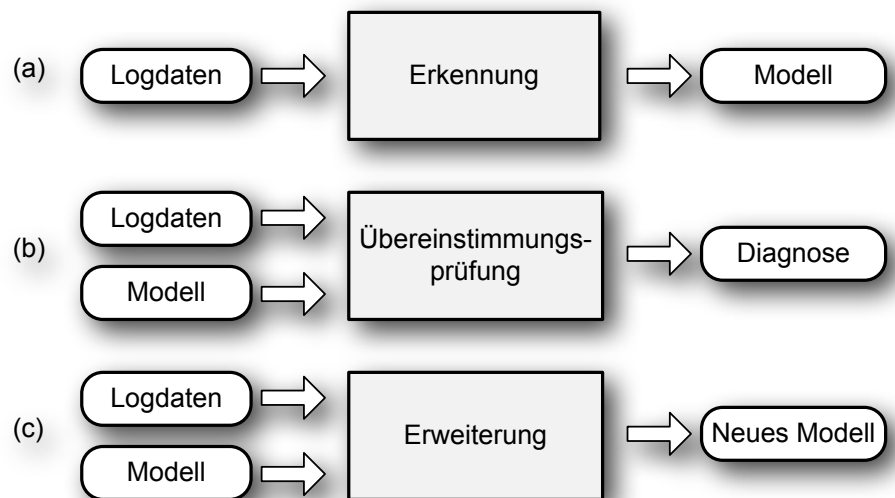


Abbildung 3: Die drei Arten des Process Mining dargestellt in Bezug auf ihre Eingabe und Ausgabe: (a) Erkennung, (b) Übereinstimmungsprüfung und (c) Erweiterung.

Eigenschaften von Process Mining:

1. Process Mining ist nicht auf die Kontrollflusserkennung beschränkt. Die Erkennung von Prozessmodellen anhand von Ereignislogdaten regt Praktiker und Akademiker gleichermaßen an. Daher wird die Erkennung oft als die spannendste Anwendung von Process Mining gesehen. Allerdings ist Process Mining nicht auf dieses Szenario beschränkt. Einerseits ist die Erkennung lediglich eine der drei Typen des Process Minings (Erkennung, Übereinstimmungsprüfung und Erweiterung). Andererseits ist der Rahmen nicht auf die Kontrollflussperspektive beschränkt. Die Organisations-, die Fall- und die Zeitperspektive spielen ebenfalls eine wichtige Rolle.

2. Process Mining ist nicht bloß eine spezielle Art des Data Mining. Process Mining kann als das fehlende Bindeglied zwischen Data Mining und dem traditionell modellgetriebenen Geschäftsprozessmanagement betrachtet werden. Die meisten Methoden des Data Mining sind in keiner Weise prozessorientiert. Prozessmodelle zeichnen sich durch Nebenläufigkeit aus, und können insofern nicht mit einfachen Strukturen des Data Mining wie Entscheidungsbäumen und Assoziationsregeln betrachtet werden. Daher sind gänzlich neue Repräsentationen und Algorithmen erforderlich.

3. Process Mining ist nicht auf Ex-Post-Analysen begrenzt. Methoden des Process Mining extrahieren Wissen aus historischen Ereignisdaten. Obwohl Daten "post mortem" verwendet werden, können die Ergebnisse auf laufende Fälle angewandt werden. Zum Beispiel kann die Endzeit einer in Bearbeitung befindlichen Bestellung auf Basis eines hergeleiteten Prozessmodells vorhergesagt werden.

Organisationseinheiten zuordnet, oder deren Beziehungen als soziales Netzwerk anzuzeigen. Die *Fallperspektive* konzentriert sich auf Eigenschaften von bestimmten Fällen. Offenkundig kann ein Fall anhand seines Ausführungspfades und der eingebundenen Akteure beschrieben werden. Ebenfalls kann er mithilfe der Werte der genutzten Datenelemente charakterisiert werden. Im Falle einer Nachbestellung mag es zum Beispiel interessant sein, den Lieferanten zu kennen und die Anzahl der bestellten Produkte. Die *Zeitperspektive* blickt auf den Zeitpunkt und die Häufigkeit von Ereignissen. Wenn Ereignislogdaten Zeitstempel enthalten, kann man anhand dessen Engpässe entdecken, Service-Levels messen, Ressourcenauslastung bestimmen und die verbleibende Ausführungszeit von laufenden Fällen vorhersagen.

Es kursieren einige falsche Annahmen zu Process Mining. Manche Hersteller, Analysten und Forscher wähen die Anwendung von Process Mining auf die Prozesserkennung im Sinne einer speziellen Data-Mining-Technik begrenzt. Dies ist allerdings *nicht* der Fall. Deshalb sollen im Folgenden drei Eigenschaften von Process Mining herausgestellt werden.

Process Mining lässt sich mithilfe des Business Process Management (BPM) Lebenszyklus einordnen (siehe Abbildung 4). Der BPM-Lebenszyklus definiert sieben Phasen eines Geschäftsprozesses und seiner zugehörigen Informationssysteme. In der Phase des *(Neu)entwurfs* wird ein neues Prozessmodell erstellt oder ein existierendes Prozessmodell angepasst.

In der *Analysephase* werden ein entworfenes Modell und seine Alternativen analysiert. Nach der *(Neu)entwurfphase* wird das Modell in der *Implementierungsphase* umgesetzt oder ein existierendes System wird neu konfiguriert (*Phase der (Neu)konfiguration*). In der *Ausführungsphase* werden Fälle anhand des entworfenen Modells abgearbeitet. In dieser Phase wird der Prozess auch *überwacht*. Zudem können einfache Änderungen vorgenommen werden ohne den Prozess im Ganzen neuzugestalten (*Anpassungsphase*). In der *Diagnosephase* wird die Ausführung des Prozesses analysiert. Die Ergebnisse dieser Phase können einen weiteren *(Neu)entwurf* des Prozesses auslösen. Process Mining ist ein nützliches Werkzeug für die meisten dieser Phasen aus Abbildung 4. Offenkundig kann die Diagnosephase von Process Mining profitieren. Allerdings ist Process Mining nicht auf diese Phase beschränkt. Zum Beispiel können Methoden des Process Mining in der Ausführungsphase zur Unterstützung genutzt werden. Auf Basis von aus historischen Daten abgeleiteten Modellen können für laufende Fälle Prognosen und Empfehlungen abgegeben werden. Ähnliche Formen der Entscheidungsunterstützung können zur Anpassung und zur *(Neu)konfiguration* der Prozesse beitragen.

Während Abbildung 4 den gesamten BPM-Lebenszyklus betrachtet, konzentriert sich Abbildung 5 auf die konkreten Aktivitäten des Process Mining mit seinen Artefakten.

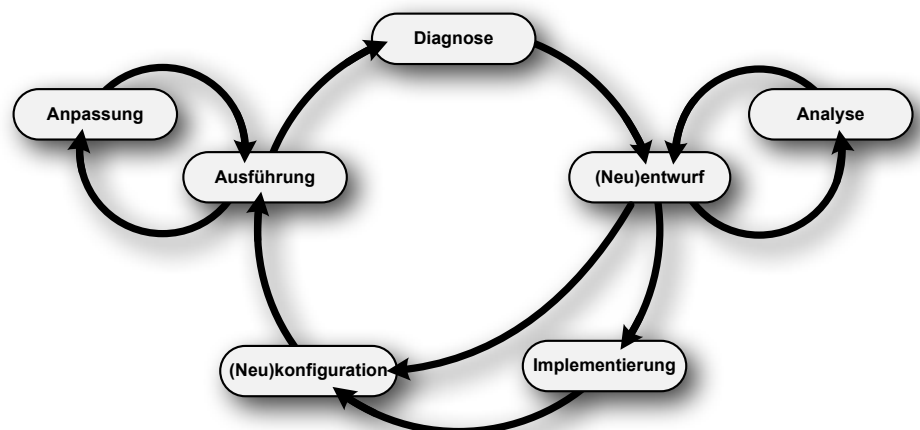
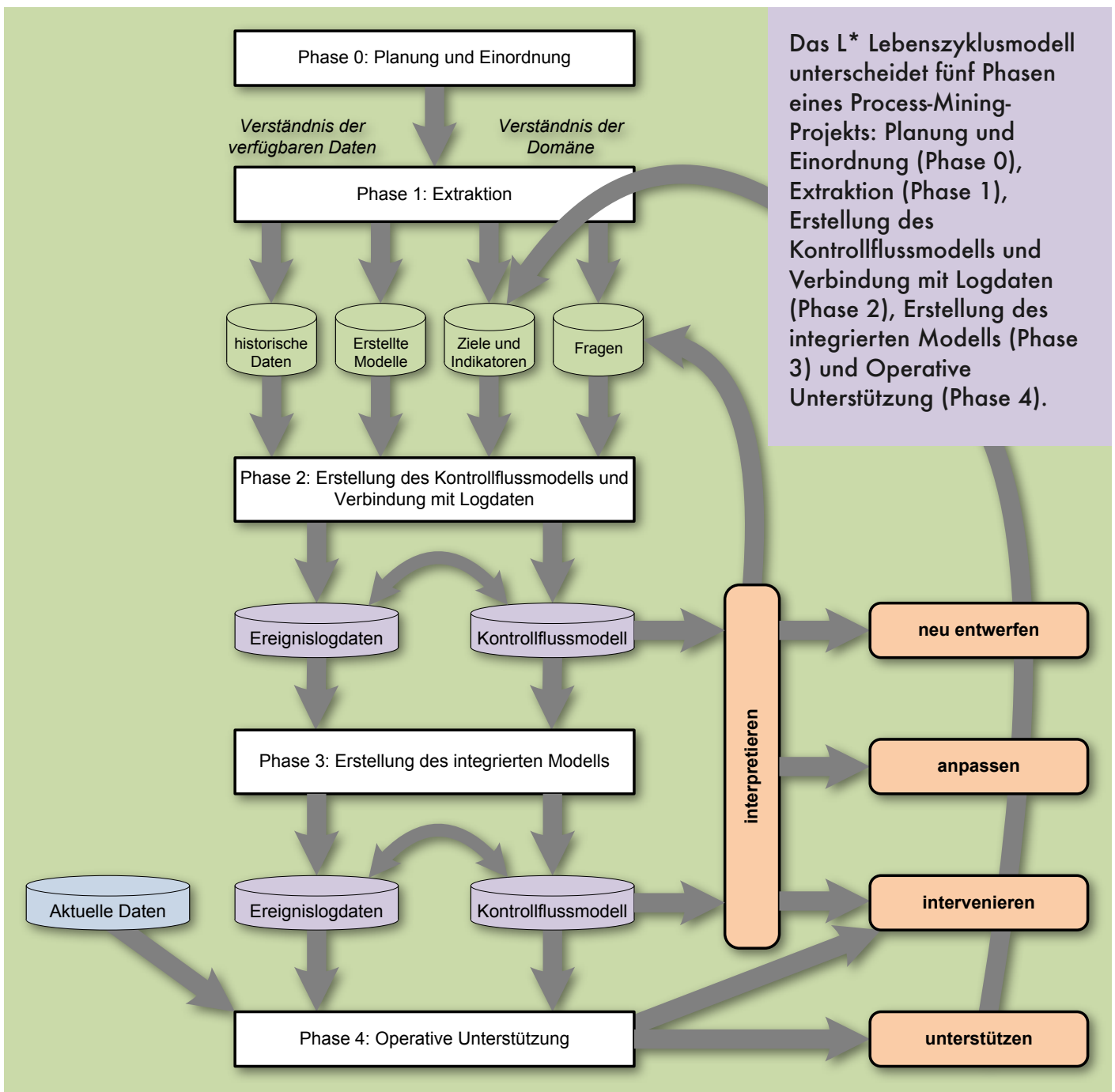


Abbildung 4: Der BPM-Lebenszyklus unterscheidet verschiedene Phasen eines Prozesses und seiner zugehörigen Informationssysteme. Process Mining kann potentiell in all diesen Phasen eingesetzt werden (mit Ausnahme der Implementierungsphase)



Das L* Lebenszyklusmodell unterscheidet fünf Phasen eines Process-Mining-Projekts: Planung und Einordnung (Phase 0), Extraktion (Phase 1), Erstellung des Kontrollflussmodells und Verbindung mit Logdaten (Phase 2), Erstellung des integrierten Modells (Phase 3) und Operative Unterstützung (Phase 4).

Abbildung 5 unterscheidet fünf Phasen eines Process-Mining-Projekts. Jedes Process-Mining-Projekt startet mit einer Planung und der Einordnung dieses Plans (Phase 0). Nach der Einrichtung des Projekts müssen die Ereignisdaten, Modelle, Ziele und Analysefragen mithilfe der Systeme, Domänenexperten und des Managements extrahiert werden (Phase 1). Dies erfordert ein Verständnis der verfügbaren Daten ("Was kann zur Analyse genutzt werden?") und ein Verständnis der Domäne ("Was sind die wichtigen Fragestellungen?") sowie der vorhandenen Artefakte (historische Daten, erstellte Modelle, Ziele und Indikatoren sowie Fragen). In der

zweiten Phase wird das Kontrollflussmodell abgeleitet und mit den Ereignisdaten verbunden. Hier können automatische Methoden der Modellerkennung genutzt werden. Die abgeleiteten Modelle können mitunter bereits Antworten auf einige der Fragen liefern, und somit einen (Neu)entwurf oder eine Anpassung mit sich bringen. Zudem können die Logdaten anhand des Modells gefiltert und adaptiert werden (bspw. wenn seltene Aktivitäten ausgeblendet oder fehlende Ereignisse eingefügt werden). Ereignisse einer eindeutigen Prozessinstanz zuzuordnen, kann einen beträchtlichen Aufwand mit sich bringen. Die verbleibenden Ereignisse werden dann Elementen des

Prozessmodells zugewiesen. Falls der Prozess ausreichend strukturiert ist, kann das Prozessmodell in der dritten Phase um weitere Perspektiven ergänzt werden (bspw. Daten, Zeit und Ressourcen). Die Verbindung zwischen dem Ereignislogdaten und dem Modell, wie es in der Phase 2 hergestellt wurde, hilft dabei, das Modell zu erweitern (bspw. können Zeitstempel der zugeordneten Ereignisse genutzt werden, um Wartezeiten für Aktivitäten abzuschätzen). Dieses kann dann eingesetzt werden, um weitergehende Fragen zu beantworten. Letztendlich kann das in Phase 3 konstruierte Modell für eine operative Unterstützung genutzt werden (Phase 4). Aus den historischen Ereignisdaten extrahiertes

Grad	Beschreibung	Beispiel
★★★★★	Höchster Grad: Die Qualität des Ereignislogs ist exzellent (d.h., belastbar und vollständig) und Ereignisse sind wohldefiniert. Ereignisse werden automatisch, systematisch, zuverlässig und sicher aufgezeichnet. Datenschutz und Sicherheitsaspekte werden angemessen berücksichtigt. Weiterhin haben die aufgezeichneten Ereignisse (und ihre Attribute) eine klare Semantik. Dies impliziert die Existenz ein oder mehrerer Ontologien, die von den Ereignissen und ihren Attribute referenziert werden.	Semantisch annotierte Logs von BPM-Systemen.
★★★★	Ereignisse werden automatisch, systematisch und zuverlässig aufgezeichnet, so dass die Ereignislogs belastbar und vollständig sind. Im Gegensatz zu Systemen des ***-Grades werden Konzepte wie Prozessinstanzen (Fälle) und Aktivitäten explizit unterstützt.	Ereignislogs klassischer BPM/ Workflow-Systeme.
★★★	Ereignisse werden automatisch aufgezeichnet, es fehlt jedoch ein systematischer Ansatz für die Aufzeichnung. Im Gegensatz zu Logs des Grades ** gibt es ein gewisses Maß an Sicherheit, dass die aufgezeichneten Ereignisse der Realität entsprechen, d.h. das Ereignislog ist belastbar aber nicht zwangsläufig vollständig. Ein Beispiel hierfür wären Ereignisse, welche von einem ERP-System aufgezeichnet werden. Auch wenn die Ereignisse aus einer Vielzahl von Datenbanktabellen extrahiert werden müssen, so kann man annehmen, dass die Ereignisse korrekt sind. So wird eine Buchung, welche vom ERP-System aufgezeichnet wurde, auch wirklich existieren, während eine fehlende Aufzeichnung als Nichtvorhandensein interpretiert werden kann.	Datenbanktabellen in ERP-Systemen, Ereignislogs von CRM-Systemen, Transaktionslogs von Nachrichtensystemen, Ereignislogs von High-Tech-Systemen, usw.
★★	Ereignisse werden automatisch als Nebenprodukt eines Informationssystems aufgezeichnet. Die Abdeckung der Aufzeichnung variiert, es gibt keinen systematischen Ansatz, um zu entscheiden, welche Ereignisse aufgezeichnet werden. Darüberhinaus ist es möglich das Informationssystem zu umgehen, wodurch Ereignisse eventuell im Log fehlen oder nicht korrekt aufgezeichnet werden.	Ereignislogs von Dokumentenverwaltungen oder Produktionssystemen, Fehlerlogs von eingebetteten Systemen, Arbeitsübersichten von Ingenieuren, usw.
★	Niedrigster Grad: die Ereignislogs sind von schlechter Qualität. Aufgezeichnete Ereignisse entsprechen mitunter nicht der Realität, während andere Ereignisse fehlen. Händisch erstellte Ereignislogs zeigen typischerweise diese Charakteristika.	Bearbeitungsvermerke in Dokumenten, die innerhalb einer Organisation zirkulieren ("Gelbe Zettel"), schriftliche Krankenakten, usw.

Tabelle 1: Reifegrade für Ereignislogs.

Wissen wird dabei mit Informationen zu in Verarbeitung befindlichen Fällen zusammengeführt. Dies kann genutzt werden, um zu intervenieren, vorherzusagen und Empfehlungen auszusprechen. Die Phasen 3 und 4 können nur erreicht werden, wenn der Prozess ausreichend stabil und strukturiert ist.

Mittlerweile gibt es Methoden und Werkzeuge, um alle Phasen aus Abbildung 5 zu unterstützen. Dennoch ist Process Mining ein neues Paradigma und viele der derzeit verfügbaren Werkzeuge sind recht unausgereift. Zudem sind potentielle Nutzer oft nicht mit den Potentialen und Grenzen des

Process Mining vertraut. Daher benennt dieses Manifest eine Reihe von Leitsätzen (Abschnitt 3) und Herausforderungen (Abschnitt 4), die sowohl für Anwender als auch für Forscher und Entwickler gedacht sind, die die bestehenden Methoden weiterentwickeln möchten.

3. Leitsätze

Wie für jede neue Technik, gibt es auch für das Process Mining offensichtliche Fehler, welche im praktischen Einsatz gemacht werden können. Aus diesem Grund stellen wir sechs Leitsätze vor,

welche Anwender und Analysten vor diesen Fehlern bewahren sollen.

LS1: Ereignisse sind fundamentale Informationsträger

Ausgangspunkt für alle Aktivitäten des Process Mining sind aufgezeichnete Ereignisse. Sammlungen solcher Ereignisse werden *Ereignislogs* genannt. Dies heißt jedoch nicht zwangsläufig, dass Ereignisse in Form dedizierter Logdateien vorliegen. Ereignisse können in Datenbanktabellen, Nachrichtenlogs, Mailarchiven, Transaktionslogs, und

Leitsätze:

LS1: Ereignisse sind fundamentale Informationsträger.

LS2: Extraktion von Ereignisdaten basiert auf konkreten Fragestellungen.

LS3: Parallelität, Entscheidung und andere elementare Kontrollflusskonzepte werden unterstützt.

LS4: Ereignisse beziehen sich auf Modellelemente.

LS5: Modelle sind zweckmässige Abstraktionen der Realität.

LS6: Process Mining ist ein kontinuierlicher Prozess.

anderen Datenquellen gespeichert sein. Weit wichtiger als das Speicherformat ist die *Qualität* des Ereignislogs. Die Qualität des Process Mining hängt in großem Umfang von den Eingabedaten ab. Ereignisse sollen daher von den prozessunterstützenden Informationssystemen als *fundamentale Informationsträger* gehandhabt werden. Leider sind Ereignislogs oft nur als ein Nebenprodukt verfügbar, genutzt für das Debuggen oder Profiling eines Informationssystems. Die Ereignisse der medizinischen Systeme von Philips Healthcare sind ein Beispiel: Die Ereignislogs werden durch Ausgabeanweisungen erzeugt, die von Softwareentwickler in den Programmcode eingefügt wurden. Auch wenn es informale Richtlinien für das Einbetten solcher Ausgabeanweisungen gibt, so braucht es einen systematischen Ansatz um Ereignislogs hoher Qualität zu erstellen. Dies gelingt nur, wenn Ereignisse als fundamentale Informationsträger aufgefasst werden.

Es gibt eine Reihe von Kriterien nach welchen die Qualität von Ereignisdaten beurteilt wird. Ereignisse sollten *belastbar* sein. Es sollte sicher sein, dass die aufgezeichneten Ereignisse tatsächlich stattfanden und dass die Attributwerte der Ereignisse korrekt sind. Ereignisse sollten *vollständig* sein. In einem bestimmten Rahmen sollten keine Ereignisse in der

Aufzeichnung fehlen. Jedes aufgezeichnete Ereignis sollte eine wohldefinierte *Semantik* haben. Außerdem sollten Ereignisdaten *sicher* sein, d.h. Datenschutz und Sicherheitsaspekte sind im Rahmen der Aufzeichnung berücksichtigt worden. Alle Akteure sollten wissen, welche Ereignisse aufgezeichnet werden und wie sie verwendet werden.

Tabelle 1 definiert fünf Reifegrade für Ereignislogs, von höchster Qualität (*****) zu niedrigster Qualität (*). Die Ereignislogs von Philips Healthcare sind ein Beispiel für den Grad ***, d.h. die Ereignisse werden automatisch aufgezeichnet und entsprechen der Realität, es fehlt jedoch ein systematischer Ansatz, um die Semantik der Ereignisse zu definieren und eine entsprechende Vollständigkeit sicherzustellen. Process-Mining-Techniken sind auf Ereignislogs der jeweiligen Reifegrade *****, **** und *** anwendbar. Generell kann Process Mining auch für Ereignislogs der Reifegrade ** oder * zur Anwendung kommen. Die Analyse solcher Ereignislogs ist jedoch oft problematisch und die Resultate sind nicht belastbar. Tatsächlich ist es nicht sinnvoll Process Mining auf Logs des Grades * anzuwenden.

Um einen großen Nutzen aus Process Mining zu ziehen, sollten Organisationen die Aufzeichnung von Ereignislogs des höchsten Reifegrades anstreben.

LS2: Extraktion von Ereignisdaten basiert auf konkreten Fragestellungen

Wie in Abb. 5 gezeigt sollten alle Aktivitäten des Process Mining durch konkrete Fragestellungen getrieben sein. Ohne solche Fragestellungen ist es sehr schwierig, aussagekräftige Ereignisdaten zu extrahieren. Ein Beispiel ist hier die Vielzahl der Tabellen einer Datenbank eines ERP Systems wie SAP. Ohne konkrete Fragestellungen ist es unmöglich, die für die Datenextraktion relevanten Tabellen ausfindig zu machen.

Ein Prozessmodell wie jenes in Abb. 1 beschreibt den Lebenszyklus von Fällen (d.h. Prozessinstanzen) eines bestimmten Typs. Aus diesem Grund ist es notwendig, den Typ der zu untersuchenden Fälle zu bestimmen, bevor Process-Mining-Techniken

angewendet werden. Diese Entscheidung sollte auf Basis der zu beantwortenden Fragen getroffen werden, was unter Umständen nicht trivial ist. Man stelle sich das Beispiel einer Bearbeitung von Kundenbestellungen vor. Jede Bestellung besteht mitunter aus mehreren Bestellpositionen, da ein Kunde mit einer Bestellung potentiell mehrere Produkte bestellt. Eine Kundenbestellung kann wiederum zu mehreren Lieferungen führen. Eine Lieferung kann Bestellpositionen unterschiedlicher Bestellungen beinhalten. Somit gibt es eine N-zu-M Beziehung zwischen Bestellungen und Lieferungen, und eine 1-zu-N Beziehung zwischen Bestellungen und Bestellpositionen. Sofern eine Datenbank Ereignisdaten mit Bezug zu Bestellungen, Bestellpositionen und Lieferungen beinhaltet, gibt es unterschiedliche Prozessmodelle welche erkannt werden können. Man kann Daten mit dem Ziel erheben, den Lebenszyklus einzelner Bestellungen zu beschreiben. Allerdings können die Daten auch mit dem Ziel extrahiert werden, den Lebenszyklus einzelner Bestellpositionen oder einzelner Lieferungen zu beschreiben.

LS3: Parallelität, Entscheidung und andere elementare Kontrollflusskonzepte werden unterstützt

Es gibt eine Vielzahl von Prozessmodellierungssprachen, zum Beispiel BPMN, EPKs, Petri-Netze, BPEL und UML-Aktivitätsdiagramme. Einige dieser Sprachen umfassen eine große Anzahl von Modellierungselementen (z.B. definiert die BPMN mehr als 50 verschiedene graphische Elemente), während sich andere Sprachen auf wenige Basiselemente beschränken (z.B. Petri-Netze, welche aus nur drei unterschiedlichen Elementen gebildet werden: Stellen, Transitionen und Pfeilen). Die Kontrollflussspezifikation ist das Rückrad jedes Prozessmodells. Allen etablierten Sprachen unterstützten folgende elementare Workflow-Konstrukte (auch bekannt als *Pattern*): Sequenzen, parallele Ausführungen (AND-Splits/Joins), Entscheidungen (XOR-Splits/Joins) und Schleifen. Offensichtlich sollten diese

Pattern auch von Process-Mining-Techniken unterstützt werden. Dennoch gibt es Techniken, die nicht mit Parallelität umgehen können und nur Markow-Ketten oder Transitionssysteme unterstützen.

Abbildung 6 verdeutlicht, welche Probleme entstehen, wenn Process-Mining-Techniken keine Parallelität erkennen können (keine AND-Split/Joins). Sei folgendes Ereignislog gegeben: $L = \{ \langle A, B, C, D, E \rangle, \langle A, B, D, C, E \rangle, \langle A, C, B, D, E \rangle, \langle A, C, D, B, E \rangle, \langle A, D, B, C, E \rangle, \langle A, D, C, B, E \rangle \}$. L enthält Fälle, welche mit A starten und mit E enden. Die Aktivitäten B, C und D treten in beliebiger Reihenfolge zwischen A und E auf. Das BPMN-Modell in Abb. 6(a) ist eine kompakte Darstellung des zugrundeliegenden Prozesses, welche zwei AND-Gateways verwendet. Nimmt man an, dass die Process Mining-Technik keine AND-Gateways unterstützt, so können die beiden anderen BPMN-Modelle in Abb. 6 erstellt werden. Das BPMN-Modell in Abb. 6(b) ist kompakt, erlaubt aber zusätzlich Verhalten (Fälle wie zum Beispiel $\langle A, B, B, B, E \rangle$ werden durch Modell erlaubt, sind aber unwahrscheinlich vor dem Hintergrund des gegebenen Ereignislogs). Das BPMN-Modell in Abb. 6(c) spiegelt alle Fälle in L wieder. Es kodiert jedoch alle Sequenzen explizit und bietet somit keine kompakte Darstellung des entsprechenden Verhaltens. Das Beispiel verdeutlicht, dass für Modelle mit dutzenden potentiell parallelen Aktivitäten, die erzeugten Modelle stark "unterformt" sind bzw. sehr komplex werden, sofern Parallelität nicht unterstützt wird.

Wie durch Abb. 6 verdeutlicht, ist es wichtig zumindest die elementaren Kontrollflusskonzepte zu unterstützen. Neben den bereits erwähnten Pattern ist es wünschenswert auch OR-Splits/Joins zu berücksichtigen, da diese Konzepte kompakte Modelle mit inklusiven Entscheidungen und partieller Synchronisation erlauben.

Wie bereits in Abschnitt 2 angedeutet, ist die Vorstellung falsch, dass Process Mining auf Kontrollflusserkennung

LS4: Ereignisse beziehen sich auf Modellelemente

Wie bereits in Abschnitt 2 angedeutet, ist die Vorstellung falsch, dass Process Mining auf Kontrollflusserkennung

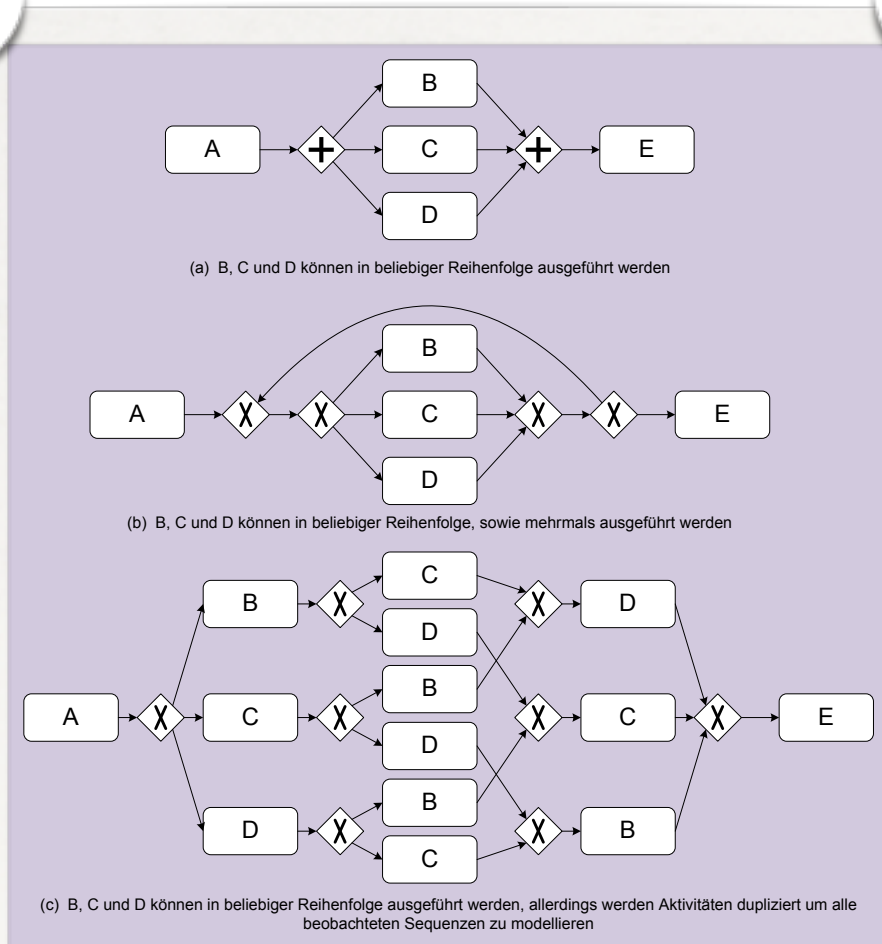


Abbildung 6: Ein Beispiel, das die Probleme verdeutlicht, die entstehen wenn Parallelität (d.h. AND-Splits/Joins) nicht direkt ausgedrückt werden kann. Das Beispiel beinhaltet nur drei parallele Aktivitäten (B, C und D). Um die Tragweite des Problems zu verstehen, kann man sich die resultierenden Prozessmodelle für 10 parallelen Aktivitäten ($2^{10} = 1024$ Zustände und $10! = 3,628,800$ mögliche Ausführungsfolgen) vorstellen.

beschränkt wäre. Abbildung 1 illustriert, dass die erkannten Prozessmodelle verschiedene Modellierungsperspektiven abdecken (Organisationsperspektive, Zeit-Perspektive, Datenperspektive, usw.). Weiterhin ist die Erkennung von Prozessmodellen nur eine Art des Process Mining, siehe auch Abb. 3. Die beiden anderen Arten des Process Mining, Übereinstimmungsprüfung und Erweiterung, hängen stark von der Beziehung zwischen Elementen des Modells und Ereignissen im Log ab. Eine solche Beziehung ist notwendig, um ein Ereignislog auf einem Modell "abspielen" zu können. Das Abspielen eines Logs erlaubt es Abweichungen zwischen dem Modell und dem Log offen zu legen, z. B. dass einige Ereignisse im Log nicht vom Modell

erlaubt sind. Techniken der Übereinstimmungsprüfung bestimmen und quantifizieren solche Abweichungen. Zeitstempel in dem Ereignislog können genutzt werden, um das zeitliche Verhalten während des Abspielens zu untersuchen. Zeitliche Abstände zwischen kausal bedingten Aktivitäten erlauben die Bestimmung von Wartezeiten, welche in das Prozessmodell übernommen werden können. Diese Beispiele verdeutlichen, dass die Beziehung zwischen Ereignissen eines Logs und Elementen eines Prozessmodells der Ausgangspunkt für eine Reihe unterschiedlicher Analysen ist. In einigen Fällen kann es jedoch schwierig sein, eine solche Beziehung herzustellen. Ein Ereignis kann sich zum Beispiel auf zwei verschiedene

Aktivitäten beziehen, so dass es nicht eindeutig zugeordnet werden kann. Diese Uneindeutigkeit muss beseitigt werden, um die Resultate des Process Mining richtig interpretieren zu können. Neben dem Problem der Beziehung zwischen Ereignissen und Aktivitäten stellt die Zuordnung der Ereignisse zu Prozessinstanzen, auch bezeichnet als *Ereigniskorrelation*, eine Herausforderung dar.

LS5: Modelle sind zweckmässige Abstraktionen der Realität

Modelle welche aus Ereignisdaten gewonnen werden stellen *Sichten auf die Realität* dar. Eine Sicht sollte eine zweckmässige Abstraktion des Verhaltens eines Ereignislogs darstellen. Für ein Ereignislog kann es durchaus mehrere nützliche Sichten geben. Verschiedene Akteure benötigen eventuell auch unterschiedliche Sichten. Aus Ereignislogs erstellte Modelle sollten als "Landkarten" gesehen werden. Dieser Leitsatz liefert wichtige Einblicke, von denen zwei im Folgenden beschrieben sind.

Erstens ist es wichtig zu sehen, dass es nicht "die Landkarte" für eine bestimmte Region gibt. Abhängig von dem jeweiligen Einsatzzweck gibt es unterschiedliche Landkarten: Straßenkarten, Wanderkarten, Radwanderkarten, usw. All diese Karten sind eine Sicht auf die Realität und es wäre absurd anzunehmen, dass so etwas wie "die perfekte Landkarte" existiert. Gleiches gilt für Prozessmodelle. Ein Modell sollte die Dinge herausstellen, die für einen bestimmten Akteur relevant sind. Die erstellten Modelle können den Fokus auf unterschiedliche Perspektiven legen (Kontrollfluss, Datenfluss, Zeit, Ressourcen, Kosten, usw.) und die in unterschiedlicher Granularität und Präzision darstellen. So wird ein Manager eventuell nach einem grobgranularen, informalen Modell verlangen, welches den Fokus auf die Kosten legt. Ein Prozessanalyst hingegen ist vielleicht eher an einem detaillierten Prozessmodell interessiert, welches die Abweichungen vom normalen Ablauf hervorhebt. Darüber hinaus verlangen verschiedene Akteure potentiell nach Prozessmodellen auf verschiedenen Ebenen: *strategische Ebene* (Entscheidungen auf dieser

Ebene haben langfristige Auswirkungen und basieren auf aggregierten Ereignisdaten, welche über einen langen Zeitraum gewonnen wurden), *taktische Ebene* (Entscheidungen auf dieser Ebene haben mittelfristige Auswirkungen und basieren zumeist auf aktuellen Daten) und *operationale Ebene* (Entscheidungen auf dieser Ebene haben unmittelbare Auswirkungen und basieren auf Ereignisdaten laufender Fälle).

Zweitens ist es nützlich auf Ideen aus der Kartographie zurückzugreifen um verständliche Landkarten zu erzeugen. Ein Beispiel sind Straßenkarten, welche von weniger wichtigen Straßen und Städten abstrahieren. Weniger wichtige Dinge werden entweder ausgelassen oder zusammengefasst, z.B. verschmelzen Straßen und Bezirke zu einer Stadt. Kartographen lassen nicht nur irrelevante Details aus, sie nutzen auch Farben, um besondere Merkmale hervorzuheben. Weiterhin wird die Größe von graphischen Elementen genutzt, um ihre Signifikanz zu kodieren (z.B. werden unterschiedliche Linienstärken oder Punktgrößen verwendet). Landkarten zeichnen sich durch eine eindeutige Interpretation ihrer x- und y-Achse aus, da das Layout nicht beliebig ist, sondern die Koordinaten der Elemente eine Bedeutung haben. Dies steht in großem Kontrast zu den meisten Prozessmodellen. Farben, unterschiedliche Elementgrößen, und bestimmte Anordnungen werden typischerweise nicht verwendet, um eine bessere Verständlichkeit zu erreichen. Die Ideen aus der Kartographie können jedoch einfach in die Erstellung von Prozesslandkarten integriert werden. So kann die Größe einer Aktivität genutzt werden, um ihre Ausführungshäufigkeit anzuzeigen, oder eine andere Eigenschaft welche auf ihre Relevanz hinweist (Kosten oder Nutzung von Ressourcen). Die Linienstärke eines Pfeils kann die Wichtigkeit der entsprechenden kausalen Abhängigkeit visualisieren, während eine Einfärbung von Pfeilen auf Engpässe hinweisen kann.

Diese Beobachtungen legen nahe, dass es wichtig ist, eine gute Darstellung für bestimmte Adressaten zu wählen und abzustimmen. Dies gilt besonders mit Hinblick auf die Visualisierung von Resultaten für

Endnutzer und für die Abstimmung von Algorithmen, um zweckmäßige Modelle zu erstellen (siehe auch Herausforderung H5).

LS6: Process Mining ist ein kontinuierlicher Prozess

Process Mining erlaubt es aussagekräftige "Landkarten" zu erstellen, welche direkt mit Ereignisdaten verbunden sind. Sowohl historische also auch aktuelle Ereignisdaten können auf diese Modelle projiziert werden. Darüberhinaus verändern sich Prozesse während sie analysiert werden. Aufgrund dieser Dynamik, ist es nicht ratsam Process Mining als einen einmaligen Vorgang anzusehen. Das Ziel sollte somit nicht ein feststehendes Modell sein. Vielmehr ist es erstrebenswert Prozessmodelle mit Leben zu füllen, so dass Nutzer und Analysten ermutigt werden die Modelle täglich zu betrachten.

An dieser Stelle ist ein Vergleich mit Mashup-Applikationen, die Geo-Tagging verwenden, hilfreich. Es gibt Tausende von Mashup-Anwendungen, die Google Maps nutzen, um Informationen über die Verkehrslage, Immobilien, Restaurants oder Kinos auf eine ausgewählte Karte zu projizieren. Man kann in diese Karten stufenlos hineinzoomen und mit ihnen interagieren (sind z.B. Stauwarnungen auf eine Karte projiziert, so können Nutzer eine Warnung auswählen, um weitere Informationen zu erhalten). Ebenso sollte es möglich sein Process Mining auf Grundlage von Echtzeitereignisdaten durchzuführen. Im Sinne der "Landkartenmetapher" kann man sich Ereignisse vorstellen, welche in Echtzeit mittels GPS Koordinaten auf eine Landkarte projiziert werden. Analog zu einem Navigationssystem eines Autos können Process-Mining-Programme Endnutzern helfen, da sie es erlauben (a) durch Prozesse zu navigieren, (b) dynamische Inhalte auf Prozesslandkarten zu projizieren (z. B. können "Verkehrsstaus" eines Geschäftsprozesses angezeigt werden) und (c) Vorhersagen für laufende Prozessinstanzen zu treffen (z. B. wird die "Ankunftszeit" eines verzögerten Falls geschätzt). Diese Beispiele zeigen wieviel Potential dadurch verschont wird, dass Prozessmodelle nicht aktiver

Herausforderungen:

H1: Ereignisdaten finden, zusammenführen und reinigen.

H2: Mit komplexen Ereignislogs mit vielfältigen Charakteristika umgehen.

H3: Repräsentative Benchmarks anlegen.

H4: Mit Konzeptverschiebung umgehen.

H5: Darstellungsausrichtung für Prozessentdeckung verbessern.

H6: Qualitätskriterien wie Fitness, Einfachheit, Präzision und Generalisierung in Einklang bringen.

H7: Organisationsübergreifendes Mining.

H8: Operationelle Unterstützung anbieten.

H9: Process Mining mit anderen Analysearten kombinieren.

H10: Nutzbarkeit für Nicht-Experten verbessern.

H11: Verständlichkeit für Nicht-Experten verbessern.

genutzt werden. Aus diesem Grund sollte Process Mining als ein kontinuierlicher Prozess angesehen werden, welcher entscheidungsrelevante Informationen für verschiedene Zeiträume (Minuten, Stunden, Tage, Wochen und Monate) bereitstellt.

4. Herausforderungen

Process Mining ist ein wichtiges Werkzeug für moderne Organisationen, die nicht-triviale operationelle Prozesse ausführen. Einerseits ist ein nahezu unglaubliches Wachstum an Ereignisdaten zu beobachten. Andererseits müssen Prozesse und Informationen perfekt aufeinander abgestimmt sein, um alle Anforderungen hinsichtlich Compliance, Effizienz und Kundenorientierung zu befriedigen. Obwohl Process Mining bereits in der Praxis einsetzbar ist, gibt es noch einige wichtige Herausforderungen, die einer Lösung bedürfen. Dies zeigt gleichzeitig, dass Process Mining eine sich stetig weiterentwickelnde Disziplin ist. Im Folgenden erläutern wir einige dieser Herausforderungen. Unsere Aufstellung erhebt keinen Anspruch auf Vollständigkeit; mit der Zeit werden neue Herausforderungen auftreten und bestehende werden aufgrund der Fortschritte im Process Mining ihre Lösung finden.

H1: Ereignisdaten finden, zusammenführen und reinigen

Es müssen nach wie vor erhebliche Anstrengungen unternommen werden,

um Ereignisdaten zu erheben, die für Process Mining geeignet sind. Folgende typische Hürden sind dabei zu nehmen:

- Ereignisdaten können über eine Vielzahl von Quellen verteilt sein. Diese Daten müssen zusammengeführt werden. Hier können Probleme auftreten, wenn verschiedene Schlüsselfelder in verschiedenen Datenquellen genutzt werden. Beispielsweise könnte ein System Personen anhand von Name und Geburtsdatum identifizieren, während ein anderes System hierfür die Sozialversicherungsnummer der Person nutzt.
- Ereignisdaten sind oft "objekt-zentriert" statt "prozess-zentriert" gespeichert. So können beispielsweise einzelne Produkte, Paletten und Container mit RFID-Tags versehen sein und Ereignisse beziehen sich auf diese RFID-Tags. Um jedoch die Aufträge eines bestimmten Kunden nachverfolgen zu können, müssen diese "objekt-zentrierten" Ereignisse zusammengeführt und vorverarbeitet werden.
- Ereignisdaten können *unvollständig* sein. Ein typisches Problem ist, dass Ereignisse nicht explizit auf Prozessinstanzen verweisen. Oftmals kann diese Information aus den vorhandenen Daten abgeleitet werden, bisweilen jedoch nur unter erheblichem Aufwand. Auch kann Zeit-Information einzelner Ereignisse fehlen. In diesem Fall müssen die Zeit-Stempel anderer Ereignisse interpoliert werden, um die verfügbare Zeit-Information noch nutzen zu können.

- Ein Ereignislog kann *Ausreißer* enthalten, d.h. Ausnahmeverhalten, das auch als *Rauschen* bezeichnet wird. Was definiert einen Ausreißer? Wie erkennt man solche Ausreißer? Diese Fragen müssen beantwortet werden, um Ereignisdaten zu reinigen.
- Logs können Ereignisdaten verschiedener *Granularität* enthalten. Das Ereignislog eines Krankenhaus-Informationssystems kann sowohl einfache Ereignisse einfacher Bluttests enthalten als auch Ereignisse komplexer chirurgischer Eingriffe. Auch Zeitstempel können in unterschiedlicher Granularität vorliegen, von präzisen Daten im Millisekundenbereich (28-9-2011:h11m28s32ms342) bis hin zu groben Datumsinformationen (28-9-2011).
- Ereignisse können in einem bestimmten *Kontext* wie *Wetter*, *Arbeitslast*, *Tag der Woche*, etc. auftreten. Dieser Kontext kann helfen bestimmte Phänomene zu erklären, wie beispielsweise dass Antwortzeiten steigen, wenn viele Aufträge vorliegen oder die Anfrage in der Ferienzeit gestellt wird. Für solche Erklärungen müssen Ereignisdaten und Kontextdaten zusammengeführt werden. Hier macht sich der "Fluch der Vieldimensionalität" bemerkbar, da eine Analyse praktisch undurchführbar wird, wenn zu viele Variablen aufgenommen werden. Bessere Methoden und Werkzeuge sind nötig, um die genannten Probleme zu lösen. Darüberhinaus müssen Organisationen, wie bereits erläutert, lernen, Ereignislogs als wertvolle

Datenquelle anzusehen und nicht nur als irgendein Nebenprodukt. Das Ziel ist es, *****-Ereignislogs zu gewinnen (siehe Tabelle 1). Hier können sich die Erkenntnisse aus dem Umfeld des *Data Warehousing* als nützlich erweisen, um qualitativ hochwertige Ereignislogs zu erhalten. So können beispielsweise schon einfache Überprüfungen während der Dateneingabe helfen, den Anteil an inkorrekten Ereignisdaten signifikant zu senken.

H2: Mit komplexen Ereignislogs mit vielfältigen Charakteristika umgehen

Ereignislogs können sehr vielfältige Charakteristika aufweisen. Manche Ereignislogs sind extrem groß, was ihr Verarbeitung erschwert. Andere Logs hingegen sind so klein, dass nicht genug Daten für belastbare Schlussfolgerungen vorliegen.

In manchen Domänen werden Ereignisse in schier unglaublicher Zahl aufgezeichnet. Dementsprechend sind weitere Anstrengungen nötig, die Leistung und Skalierbarkeit der Process-Mining-Techniken zu verbessern. Beispielsweise überwacht ASML kontinuierlich alle seine Wafer-Scanner. Diese Wafer-Scanner werden in verschiedenen Organisationen (z.B. Samsung und Texas Instruments) eingesetzt, um Microchips herzustellen (ca. 70% aller Chips werden auf Wafer-Scannern von ASML hergestellt). Verfügbare Techniken haben Probleme, mit Datenmengen in der Größenordnung einiger Petabytes umzugehen, die in diesen Domänen anfallen. Neben der schlichten Anzahl aufgezeichneter Ereignisse gibt es weitere Charakteristika wie Ereignisse pro Fall, Ähnlichkeit von Fällen, die Anzahl spezifischer Ereignisse und die Anzahl spezifischer Pfade. Nehmen wir an, ein Ereignislog L1 enthalte 1000 Fälle mit jeweils 10 Ereignissen (d.h. viele Fälle folgen demselben oder sehr ähnlichen Pfaden). Ein Ereignislog L2 könnte hingegen nur 100 Fälle zu je 100 Ereignissen enthalten, die alle einem anderen, spezifischen Pfad folgen. Offensichtlich ist L2 sehr viel schwerer zu analysieren als L1 obwohl beide Logs mit ca. 10000 Ereignissen ähnlich groß sind.

Da Ereignislogs nur Beispielverhalten enthalten, können wir nicht annehmen sie wären vollständig.

Process-Mining-Techniken müssen in der Lage sein mit Unvollständigkeit umzugehen, in dem sie eine "offene Welt" annehmen ("open world assumption"): Die Beobachtung, dass etwas nicht eingetreten ist bedeutet nicht, dass es gar nicht eintreten kann. Entsprechend herausfordernd ist die Analyse kleiner Eventlogs mit großer Variabilität.

Wie bereits erläutert enthalten einige Logs Ereignisse auf einem sehr niedrigen Abstraktionsgrad. Diese Logs tendieren dazu extrem groß zu sein, während gleichzeitig individuelle Ereignisse auf der untersten Ebene von geringem Interesse für die am Prozess beteiligten Personen und Organisationen sind. Aus diesem Grund ist es sinnvoll Ereignisse der untersten Ebene in höherwertige Ereignisse zu aggregieren. Beispielsweise könnten bei der Analyse von Diagnose- und Behandlungsprozessen einer bestimmten Patientengruppe die individuellen Testergebnisse, die im Informationssystem des Krankenhauslabors aufgezeichnet wurden, uninteressant sein.

Zum gegenwärtigen Zeitpunkt sind Organisationen auf Versuch-und-Irrtum-Ansätze angewiesen, um zu erkennen, ob ein Ereignislog für Process Mining geeignet ist. Aus diesem Grund sollten Werkzeuge einen schnellen Brauchbarkeitstest für einen gegebenen Datensatz anbieten. Solch ein Test sollte potentielle Leistungsprobleme in der Analyse anzeigen und den Nutzer warnen, wenn ein Log vorliegt, das nicht vollständig genug ist oder zu detailliert.

H3: Repräsentative Benchmarks anlegen

Process Mining ist eine Technik, die sich noch in der Entwicklung befindet. Dies erklärt, weshalb gute Benchmarks bislang fehlen. So sind beispielsweise dutzende Prozessentdeckungstechniken verfügbar und verschiedene Firmen bieten unterschiedliche Produkte an, es gibt jedoch keinerlei Konsens über die Qualität dieser Techniken. Obwohl diese riesige Unterschiede in Funktionalität und Leistung aufweisen, ist es schwierig Techniken und Werkzeuge miteinander zu vergleichen. Aus diesem Grund müssen gute Benchmarks, bestehend aus

Beispieldatensätzen und repräsentativen Qualitätskriterien, entwickelt werden.

Für klassische Data-Mining-Techniken sind viele gute Benchmarks verfügbar. Diese Benchmarks haben Werkzeuganbieter und Forscher dazu angeregt die Leistungsfähigkeit ihrer Techniken zu verbessern. Process Mining ist hier herausfordernder. Beispielweise ist das relationale Modell von Codd (1969) einfach und allseits unterstützt. Daher gelingt eine Konvertierung von Daten einer Datenbank in eine andere Datenbank mit geringem Aufwand und es entstehen keine Interpretationsprobleme. Für Prozesse fehlt ein vergleichbares Modell. Die bislang vorgeschlagenen Standards zur Prozessmodellierung sind erheblich komplizierter und nur wenige Anbieter unterstützen exakt dieselben Konzepte. Prozesse sind schlicht komplexer als Tabellendaten.

Gleichwohl ist es wichtig, repräsentative Benchmarks für Process Mining zu erstellen. Erste Resultate sind bereits verfügbar. So gibt es diverse Maße, um die Qualität eines Results im Process Mining zu messen (Fitness, Einfachheit, Präzision und Generalisierung). Darüberhinaus sind mehrere Ereignislogs frei verfügbar (siehe www.processmining.org); ein repräsentatives Beispiel ist das Ereignislog, das in der ersten, von der Taskforce organisierten, Business Process Intelligence Challenge (BPIC'11) verwendet wurde (siehe doi: 10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54).

Benchmarks sollten einerseits auf Datensätzen aus der Praxis basieren. Andererseits werden auch synthetische Benchmarks benötigt, die bestimmte Charakteristika abbilden. Solche synthetischen Benchmarks können beispielsweise helfen, Process-Mining-Techniken für unvollständige Ereignislogs, Ereignislogs mit Rauschen, oder bestimmte Klassen von Prozessen zu entwickeln.

Neben dem Anlegen repräsentativer Benchmarks muss auch stärkerer Konsens über die Kriterien gefunden werden, anhand derer die Qualität eines Process-Mining-Resultats bewertet wird (siehe auch Herausforderung H6). Zusätzlich lassen sich *Kreuz-Validierungstechniken* aus

dem Bereich des Data Mining adaptieren, um ein Ergebnis zu bewerten. Nehmen wir das k-fache Prüfen als Beispiel. Hierbei wird das Ereignislog in k Teile aufgespalten. k-1 Teile werden benutzt, um ein Prozessmodell zu entdecken. Techniken der Übereinstimmungsprüfung können dann das Resultat bezüglich des letzten Logteils überprüfen. Dies kann k mal wiederholt werden, was einige Einsichten in die Qualität des Modells und der verwendeten Techniken liefert.

H4: Mit Konzeptverschiebung umgehen

Der Begriff *Konzeptverschiebung* beschreibt eine Situation, in der ein Prozess sich verändert während er analysiert wird. Beispielsweise können am Anfang der Analyse zwei Ereignisse nebenläufig auftreten und später in der Analyse nur noch sequentiell geordnet auftreten. Prozesse können sich periodisch oder saisonal verändern (z.B. "im Dezember gibt es eine höhere Nachfrage" oder "Freitag Nachmittag sind weniger Mitarbeiter verfügbar") oder sich ändernden Bedingungen unterliegen ("der Markt unterliegt stärkerer Konkurrenz"). Solche Änderungen wirken sich auf die Prozesse aus und es ist wichtig sie zu erkennen und zu analysieren. Konzeptverschiebung kann in einem Ereignislog entdeckt werden, in dem das Log in kleinere Logs aufgespalten wird und deren "Fußabdrücke" analysiert und verglichen werden. Diese Analyse "zweiter Stufe" setzt viele Ereignisdaten voraus. Gleichwohl sind nur wenige Prozesse unveränderlich, und Konzeptverschiebung zu verstehen ist eine Primäraufgabe bei der Verwaltung von Prozessen. Daher sind weitere Forschung und Werkzeugunterstützung nötig, um Konzeptverschiebung adäquat analysieren zu können.

H5: Darstellungsausrichtung für Prozessentdeckung verbessern

Eine Technik zur Prozessentdeckung erzeugt Modelle in einer bestimmte Sprache (z.B. BPMN oder Petrinetze). Allerdings ist es wichtig zwischen der Darstellung des Ergebnisses und der

Darstellung während der eigentlichen Entdeckung zu unterscheiden. Die Wahl einer bestimmten Zielsprache bringt häufig eine Reihe impliziter Annahmen mit sich. Diese schränkt den Suchraum ein: Prozesse, die in der Zielsprache nicht darstellbar sind, können auch nicht entdeckt werden. Diese sogenannte

"Darstellungsausrichtung" (*representational bias*), die während der Prozessentdeckung genutzt wird, sollte eine bewusste Wahl sein und nicht (nur) durch die bevorzugt graphische Darstellung getrieben sein.

Betrachten wir das Beispiel in Abb. 6: Ob die Zielsprache Nebenläufigkeit darstellen kann oder nicht hat sowohl Auswirkungen auf die Darstellung des entdeckten Modells als auch auf die Klasse der Modelle, die vom Algorithmus in Betracht gezogen werden. Falls die Darstellungsausrichtung keine Nebenläufigkeit gestattet (Abb. 6(a) ist nicht darstellbar) und auch mehrere Aktivitäten mit demselben Namen verbietet (Abb. 6(c) ist nicht darstellbar), dann sind nur problematische Modelle, wie das in Abb. 6(b) gezeigte, möglich. Dieses Beispiel illustriert, dass eine vorsichtigeren und feinere Wahl der Darstellungsausrichtung nötig ist.

H6: Qualitätskriterien wie Fitness, Einfachheit, Präzision und Generalisierung in Einklang bringen

Ereignislogs sind oftmals alles andere als vollständig. Das heißt es wurde nur Beispielverhalten aufgezeichnet. Prozessmodelle erlauben typischerweise exponentiell viele oder sogar unendlich viele verschiedene Abläufe (im Falle von Schleifen). Weiterhin treten einige Abläufe mit deutlich geringerer Wahrscheinlichkeit als andere auf. Es ist daher unrealistisch anzunehmen, jeder mögliche Ablauf wäre auch im Log vorhanden. Folgendes Beispiel illustriert die unplausible Annahme eines vollständigen Logs. Betrachten wir einen Prozess der 10 Aktivitäten nebenläufig ausführt und ein zugehöriges Log, das Informationen über 10.000 Fälle enthält. Die Gesamtzahl aller möglichen Ausführungen des Modells mit 10

nebenläufigen Aktivitäten ist $10! = 3.628.800$. Es ist daher unmöglich, dass jede mögliche Ausführung im Ereignislog (mit nur 10.000 Fällen) enthalten ist. Selbst wenn das Log Millionen von Fällen enthält ist es extrem unwahrscheinlich, dass jede mögliche Variation auch im Log vorkommt. Eine weitere Komplikation ist, dass manche Alternativen seltener auftreten als andere. Diese können als "Rauschen" aufgefasst werden. Es ist nahezu unmöglich, ein vernünftiges Prozessmodell für solch verrauschtes Verhalten zu konstruieren. Das entdeckte Prozessmodell muss davon abstrahieren; selten auftretendes Verhalten (Rauschen) lässt sich besser mit Techniken der Übereinstimmungsprüfung untersuchen.

Rauschen und Unvollständigkeit sind eine Herausforderung für die Prozessentdeckung. Es gibt vier Qualitätsmaße: (a) Fitness, (b) Einfachheit, (c) Präzision, und (d) Generalisierung. Ein Modell mit guter *Fitness* beschreibt einen Großteil des Verhaltens, das auch im Ereignislog aufgezeichnet ist. Ein Modell mit perfekter Fitness kann alle Abläufe des Ereignislog wieder abspielen - von Anfang bis Ende. Das *einfachste* Modell, das das aufgezeichnete Verhalten abspielen kann, gilt als das beste Modell. Dieses Prinzip ist auch als Occams Rasiermesser bekannt. Fitness und Einfachheit allein reichen nicht aus, um die Qualität eines entdeckten Prozessmodells zu bewerten. Beispielsweise ist es sehr leicht ein extrem einfaches Prozessmodell ("Blumenmodell" oder "flower model") zu konstruieren, das alle Abläufe des Logs abspielen kann (aber auch jedes andere Log, das sich auf dieselbe Menge von Aktivitäten bezieht). Es ist gleichermaßen unerwünscht ein Modell zu entdecken, das genau das im Log aufgezeichnete Verhalten beschreibt. Erinnern wir uns, dass ein Log nur Beispielverhalten beschreibt und dass viele mögliche Abläufe möglicherweise noch nicht beobachtet wurden. Ein Modell ist *präzise*, wenn es nicht "zu viel" Verhalten erlaubt. Das Blumenmodell weist nur eine geringe Präzision auf. Ein unpräzises Modell ist "unterformt" (*underfitting*). Unterformung tritt auf, wenn das Modell das Beispielverhalten im Log zu stark generalisiert (d.h. das Modell beschreibt auch Verhalten, das sich

grundlegend vom im Log aufgezeichneten Verhalten unterscheidet). Ein Modell, das gar nicht *generalisiert*, ist "überformt" (*overfitting*). Überformung tritt auf, wenn ein sehr spezifisches Modell erzeugt wird, obwohl klar ist, dass das Log nur Beispielverhalten beinhaltet (d.h. das Modell erklärt eine spezielle Probe des aufgezeichneten Verhaltens, aber bereits die nächste Probe desselben Prozesses kann ein ganz anderes Prozessmodell ergeben).

Fitness, Einfachheit, Präzision und Generalisierung miteinander abzuwägen ist eine Herausforderung. Speziell aus diesem Grund bieten viele leistungsfähige Prozessentdeckungstechniken verschiedene Parameter. Bessere Algorithmen müssen entwickelt werden, um die vier Qualitätskriterien besser in Einklang zu bringen. Darüberhinaus sollten die Parameter für Nutzer verständlich sein.

H7: Organisationsübergreifend es Mining

Klassisch wird Process Mining innerhalb einer einzelnen Organisation angewandt. Jedoch in dem Maße, wie sich dienstbasierte Techniken, Wertschöpfungskettenintegration und Cloud-Computing verbreiten, werden Szenarien denkbar, in denen Ereignislogs mehrerer Organisationen zur Analyse verfügbar sind. Es gibt zwei Arten von *organisationsübergreifendem Process Mining*.

Zunächst können wir die kollaborative Situation betrachten, in der verschiedene Organisationen gemeinsam eine Prozessinstanz vorantreiben. Der organisationsübergreifende Prozess ähnelt hier einem Puzzle, d.h. der Gesamtprozess ist in Teile geschnitten und über mehrere Organisationen verteilt, die kooperieren müssen, um den Fall abzuschließen. Das Ereignislog nur innerhalb einer dieser Organisationen zu analysieren reicht nicht aus. Um den gesamten Prozess von Anfang bis Ende zu entdecken, müssen die Ereignislogs der verschiedenen Organisationen zusammengeführt werden. Dies ist eine nicht-triviale Aufgabe, da Ereignisse über Organisationsgrenzen hinweg zueinander in Beziehung gesetzt werden müssen.

In der zweiten Situation für organisationsübergreifendes Mining führen verschiedene Organisationen im Grund denselben Prozess aus und teilen Erfahrungen, Wissen oder eine gemeinsame Infrastruktur. Betrachten wir beispielsweise *Salesforce.com*. Die Verkaufsprozesse vieler Organisationen werden von *Salesforce* verwaltet und unterstützt. Einerseits teilen diese Organisationen dieselbe Infrastruktur (Prozesse, Datenbanken, etc.). Andererseits sind sie nicht gezwungen einem strengen Prozessmodell zu folgen, da das System konfiguriert werden kann, um Varianten desselben Prozesses zu unterstützen. Offensichtlich wäre es interessant diese Variationen zwischen den einzelnen Organisationen zu untersuchen. Die Organisationen könnten voneinander lernen und Anbieter könnten ihr Dienste verbessern und Mehrwertdienste basierend auf den Ergebnissen des organisationsübergreifendem Mining anbieten.

Neue Analysetechniken müssen für beide Arten von organisationsübergreifendem Mining entwickelt werden. Diese Techniken sollten auch Datenschutz und Sicherheitsaspekte betrachten. Organisationen möchten bestimmte Informationen nicht mit anderen teilen, sei es aus Konkurrenzgründen oder aus Mangel an Vertrauen. Es ist daher wichtig datenschutz-bewahrende Process-Mining-Techniken zu entwickeln.

H8: Operationelle Unterstützung anbieten

Anfangs lag der Fokus im Process Mining auf der Analyse historischer Daten. Heutzutage werden viele Datenquellen jedoch in (nahezu) Echtzeit aktualisiert und ausreichend Rechenleistung ist verfügbar, um Ereignisse zu analysieren während sie passieren. Daher sollten Process-Mining-Techniken nicht allein auf Offline-Analyse beschränkt werden, sondern können auch für operationelle Unterstützung zur Laufzeit eingesetzt werden. Drei Aktivitäten im operationellen Support können identifiziert werden: *erkennen*, *vorhersagen*, und *empfehlen*. Der Moment, in dem ein Fall vom vorgegebenen Prozess abweicht, kann erkannt werden und das System kann eine Warnmeldung erzeugen. Oftmals möchte man solche Meldungen

unverzüglich erzeugen (um noch Einfluss nehmen zu können) und nicht erst im Offline-Fall. Historische Daten können genutzt werden, um Vorhersagemodelle zu erzeugen. Diese können genutzt werden, um laufende Prozessinstanzen zu leiten. Beispielsweise ist es möglich die noch verbleibende Bearbeitungszeit eines Falls vorherzusagen. Basierend auf solchen Vorhersagen lassen sich auch Empfehlungssysteme bauen, die bestimmte Schritte vorschlagen, um Kosten oder Bearbeitungszeiten zu reduzieren. Process-Mining-Techniken in solchen Echtzeitsituationen einzusetzen stellt weitere Herausforderungen an die Rechenleistung und die Datenqualität.

H9: Process Mining mit anderen Analysearten kombinieren

Produktionsmanagement, und insbesondere *Operations Research*, ist ein Zweig der Wirtschaftswissenschaften, der stark auf Modellierung basiert. Hier werden eine Vielzahl mathematischer Modelle von Linearer Programmierung und Projektplanung über Warteschlangenmodelle und Markov-Ketten bis hin zu Simulation eingesetzt. Data Mining kann definiert werden als "die Analyse (oftmals großer) Datensätze, um unerwartete Beziehungen zu finden und Daten auf neuartige Weise zusammenzufassen, die sowohl verständlich als auch nützlich für den Eigner der Daten sind". Eine Vielzahl von Techniken wurden entwickelt: Klassifikation (z.B. Erlernen von Entscheidungsbäumen), Regression, Clustering (z.B. k-Means-Clustering) und Muster-Erkennung (z.B. Erlernen von Assoziationsregeln).

Beide Bereiche (Produktionsmanagement und Data Mining) bieten wertvolle Analysetechniken. Die Herausforderung ist hier, Techniken aus diesen Gebieten mit Process Mining zu kombinieren. Nehmen wir Simulation als Beispiel. Process-Mining-Techniken können genutzt werden, um ein Simulationsmodell auf Basis historischer Daten zu erlernen. Anschließend kann das Simulationsmodell für operationelle Unterstützung genutzt werden. Aufgrund der engen Verbindung zwischen Ereignislog und Modell kann die Historie auf dem Modell abgespielt werden und der Prozessfortgang kann

ab dem gegenwärtig erreichten Zustand simuliert werden, wodurch ein "Schnell-Vorlauf-Knopf" in die Zukunft basierend auf aktuellen Daten bereitgestellt wird.

Ähnlich ließen sich Process Mining und *Visuelle Analyse* (*visual analytics*) kombinieren. Visuelle Analyse vereint automatische Analyse mit interaktiver Visualisierung, um ein besseres Verständnis großer und komplexer Datenmengen zu erreichen. Sie nutzt dabei die verblüffende Fähigkeit des Menschen, Muster in unstrukturierten Daten zu erkennen. In dem automatisierte Process-Mining-Techniken mit interaktiver, visueller Analyse kombiniert wird, ist es möglich tiefere Einsichten in Ereignisdaten zu gewinnen.

H10: Nutzbarkeit für Nicht-Experten verbessern

Ein Ziel von Process Mining ist es, "lebendige Prozessmodelle" zu erzeugen, d.h. Prozessmodelle, die täglich genutzt werden, anstatt statischer Modelle, die in irgendeinem Archiv landen. Neue Ereignisdaten können genutzt werden, um sich neu entwickelndes Verhalten zu entdecken. Der Zusammenhang zwischen Ereignisdaten und Prozessmodellen erlaubt es, den aktuellen Zustand und die zuletzt ausgeführten Aktivitäten auf aktuelle Prozessmodelle zu projizieren. Auf diese Art und Weise können Endanwender mit den Ergebnissen des Process Mining täglich arbeiten. Solche Interaktionen sind sehr wertvoll, erfordern allerdings eine intuitiv benutzbare Programmoberfläche. Die Herausforderung ist hier die ausgefeilten Process-Mining-Techniken mit benutzerfreundlichen Schnittstellen und Oberflächen zu versehen, die automatisch Parameter wählen und passende Analyseaufgaben vorschlagen.

H11: Verständlichkeit für Nicht-Experten verbessern

Selbst wenn es einfach ist, Prozess-Mining-Resultate zu erzeugen, bedeutet dies noch lange nicht, dass die Resultate auch brauchbar sind. Ein Nutzer könnte Probleme haben die Ausgabe zu verstehen oder könnte geneigt sein falsche Schlussfolgerungen zu ziehen. Um solche Probleme zu vermeiden, sollten die Ergebnisse in

einer brauchbaren Darstellung ausgegeben werden (siehe auch LS5). Darüberhinaus sollte die Belastbarkeit der Resultate stets klar angezeigt werden. Es könnten beispielsweise zu wenig Daten vorliegen, um eine bestimmte Schlussfolgerung zu ziehen. Tatsächlich warnen bestehende Entdeckungstechniken nicht vor einem niedrigen Fitnesswert oder vor Überformung. Sie zeigen stets ein Modell, selbst wenn klar ist, dass zu wenig Daten für jegliche Schlussfolgerung vorliegen.

Epilog

Ziel der IEEE Task Force on Process Mining ist es (a) die Anwendung von Process Mining zu fördern und zu verbreiten, (b) Softwareentwickler, Berater, Manager und Endnutzer bei der Nutzung modernster Techniken zu unterstützen, und (c) die Forschung im Bereich Process Mining weiter voranzubringen und zu stimulieren. Dieses Manifest beschreibt die Grundprinzipien und Intentionen der Task Force. Nach der Einführung von Process Mining, werden in diesem Manifest einige Leitsätze (Kapitel 3) und Herausforderungen aufgezeigt. Die dargestellten Prinzipien sollen dazu dienen häufige Fehler zu verhindern. Die Liste der Herausforderungen hingegen, soll als Wegweiser für Forschung und Entwicklung dienen. Beides soll dazu beitragen, den Reifegrad von Process-Mining-Techniken zu erhöhen.

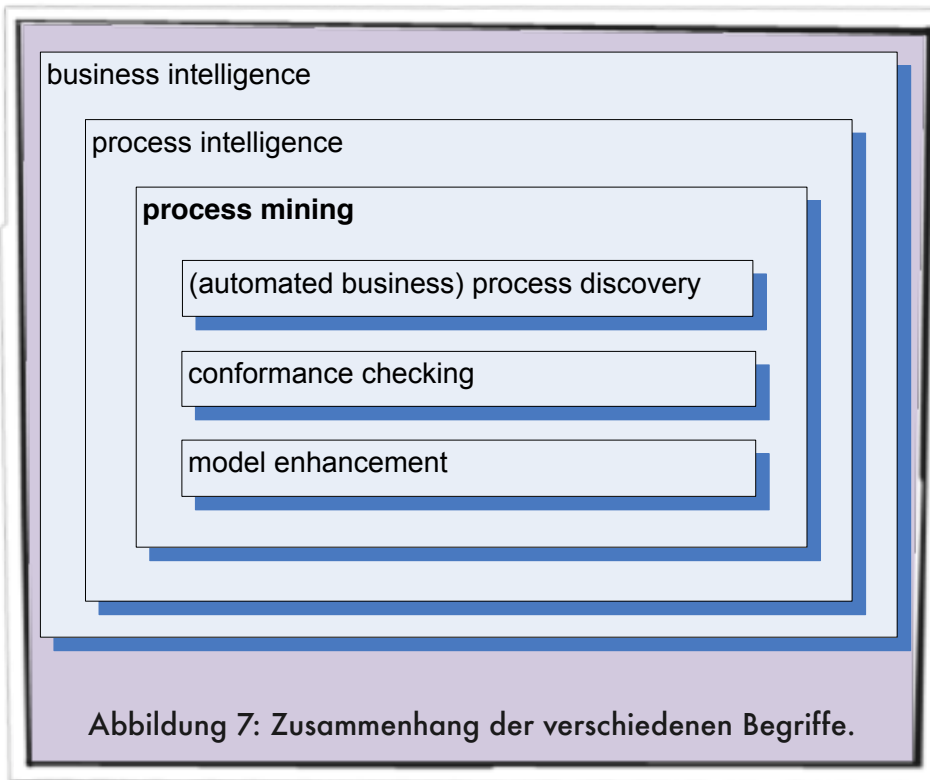
Abschließend sollen noch einige Worte zur Terminologie folgen. Die im Folgenden aufgeführten Begriffe werden im Bereich von Process Mining verwendet: Workflow Mining, (Business) Process Mining, Automatische (Geschäfts-) Prozesserkennung (Automated (Business) Process Discovery) und (Business) Process Intelligence. Verschiedene Organisationen verwenden in diesem Zusammenhang oft unterschiedliche Begriffe für übergreifende Konzepte. Gartner prägt beispielsweise den Begriff "Automated Business Process Discovery" (ABPD) während die Software AG den Term "Process Intelligence" für Ihre Controlling-Plattform nutzt. Des Weiteren erscheint der Term "Workflow Mining" weniger passend, da die Erstellung von Workflow-Modellen nur

eine von vielen Process Mining Anwendungen ist. Gleichmaßen beschränkt das Hinzufügen des Terms "Business" den Begriff auf bestimmte Anwendungen. Es gibt zahlreiche Process-Mining-Anwendungen, wie z.B. die Analyse der Verwendung von High-Tech-Systemen oder die Analyse von Webseiten, bei denen diese Ergänzung nicht passend erscheint. Auch wenn die Prozesserkennung ein wichtiger Teil des Process-Mining-Spektrums ist, ist sie dennoch nur einer von vielen Anwendungsfällen.

Übereinstimmungsprüfung, Vorhersage, Organisationserkennung, die Analyse sozialer Netzwerke, etc. sind weitere Anwendungsgebiete, die über die Prozesserkennung hinaus gehen.

Abbildung 7 zeigt die Zusammenhänge einiger der zuvor genannten Begriffe. Alle Technologien und Methoden, die der Gewinnung von gesicherten Informationen zur Entscheidungsfindung dienen, können dem Oberbegriff Business Intelligence (BI) untergeordnet werden. (Business) Process Intelligence kann als Kombination von BI und BPM angesehen werden. D.h. BI-Techniken werden hierbei genutzt, um Prozesse bzw. das Prozessmanagement zu verbessern. Process Mining kann als Konkretisierung von Process Intelligence angesehen werden, bei der Ereignislogs als Ausgangspunkt verwendet werden. (Automatische Geschäfts-) Prozesserkennung ist nur eine der drei Arten von Process Mining. Abbildung 7 ist eventuell etwas irreführend, da die meisten BI-Softwareapplikationen keine der in diesem Dokument beschriebenen Process-Mining-Funktionalitäten unterstützen. Der Term BI wird oft in geeigneter Weise in Hinblick auf ein

Dieses Manifest wurde zuerst im Band "Business Process Management Workshops 2011," Lecture Notes in Business Information Processing, Vol. 99, Springer-Verlag, 2011 veröffentlicht und inzwischen in mehrere Sprachen übersetzt. Besuchen Sie die Webseite der IEEE Taskforce on Process Mining unter <http://www.win.tue.nl/ieeetfpm/> für weitere Informationen.



bestimmtes Tool oder eine bestimmte Methode angepasst und umfasst dann nur noch einen kleinen Teil des breiten BI Spektrums.

Es mag kommerzielle Gründe geben alternative Begriffe zu verwenden. Einige Anbieter wollen dabei z.B. einen bestimmten Aspekt herausheben (z.B. Erkennung oder Intelligence). Nichtsdestotrotz, hilft es Verwirrung zu vermeiden, wenn der Begriff "Process Mining" ausschließlich für die in diesem Manifest beschriebene Disziplin verwendet wird.

Glossar

Aktivität (Activity): ein wohl-definierter Schritt in einem Prozess. Ereignisse können auf den Start, Abschluss, Abbruch, etc. einer Aktivität für eine bestimmte Prozess-Instanz verweisen. Automatische Geschäftsprozess-Entdeckung (Automated Business Process Discovery): siehe Prozess-Entdeckung

Business Intelligence (BI): vielfältige Sammlung von Werkzeugen und

Methoden, die Daten nutzen, um Entscheidungen zu unterstützen
Business Process Intelligence: see Process Intelligence.

Darstellungsausrichtung: die gewählte Zielsprache, in der Ergebnisse des Process Mining konstruiert und präsentiert werden.

Data Mining: die Analyse (oftmals großer) Datenmengen, um unerwartete Beziehungen zu finden und Daten auf neue Art und Weise zusammenzufassen, mit dem Ziel neue Einsichten zu gewinnen.

Einfachheit (Simplicity): ein Maß, das Occams Rasiermesser operationalisiert: das einfachste Prozessmodell, das das im Log aufgezeichnete Verhalten erklären kann, ist das beste Modell. Einfachheit kann auf verschiedene Art quantifiziert werden, z.B. als Anzahl von Knoten und Kanten im Modell.

Ereignis (Event): eine Aktion, die in einem Log aufgezeichnet wurde, z.B. der Anfang, Abschluss oder Abbruch einer Aktivität in einer bestimmte Prozessinstanz.

Ereignislog (Event Log): Sammlung von Ereignissen, die als Eingabe für Process Mining verwendet wird. Ereignisse müssen nicht in einer separate Logdatei gespeichert werden (z.B. können Ereignisse über verschiedene Tabellen verteilt sein).

Erweiterung (Model Enhancement): eine der drei grundlegenden Arten des Process Mining. Ein Prozessmodell wird anhand von Informationen aus einem Ereignislog erweitert oder verbessert.

Autoren

Wil van der Aalst
Arya Adriansyah
Ana Karla Alves de Medeiros
Franco Arcieri
Thomas Baier
Tobias Blickle
Jagadeesh Chandra Bose
Peter van den Brand
Ronald Brandtjen
Joos Buijs
Andrea Burattin
Josep Carmona
Malu Castellanos
Jan Claes
Jonathan Cook
Nicola Costantini
Francisco Curbera
Ernesto Damiani
Massimiliano de Leoni

Pavlos Delias
Boudewijn van Dongen
Marlon Dumas
Schahram Dustdar
Dirk Fahland
Diogo R. Ferreira
Walid Gaaloul
Frank van Geffen
Sukriti Goel
Christian Günther
Antonella Guzzo
Paul Harmon
Arthur ter Hofstede
John Hoogland
Jon Espen Ingvaldsen
Koki Kato
Rudolf Kuhn
Akhil Kumar
Marcello La Rosa
Fabrizio Maggi

Donato Malerba
Ronny Mans
Alberto Manuel
Martin McCreesh
Paola Mello
Jan Mendling
Marco Montali
Hamid Motahari
Nezhad
Michael zur Muehlen
Jorge Munoz-Gama
Luigi Pontieri
Joel Ribeiro
Anne Rozinat
Hugo Seguel Pérez
Ricardo Seguel Pérez
Marcos Sepúlveda
Jim Sinur
Pnina Soffer
Minseok Song
Alessandro Sperduti

Giovanni Stilo
Casper Stoel
Keith Swenson
Maurizio Talamo
Wei Tan
Chris Turner
Jan Vanthienen
George Varvaressos
Eric Verbeek
Marc Verdonk
Roberto Vigo
Jianmin Wang
Barbara Weber
Matthias Weidlich
Ton Weijters
Lijie Wen
Michael Westergaard
Moe Wynn

Beispielsweise können Engpässe erkannt werden, wenn das Ereignislog auf einen Prozessmodell abgespielt wird und dabei die Zeitstempel der Ereignisse analysiert werden.

Fall (Case): siehe Prozessinstanz.

Fitness: ein Maß, das bestimmt wie gut ein gegebenes Modell das Verhalten, das in einem Ereignislog aufgezeichnet wurde, beschreibt. Ein Modell hat perfekte Fitness, wenn alle Abläufe des Logs von Anfangs bis Ende auf dem Modell abgespielt werden können.

Generalisierung (Generalization): ein Maß, das bestimmt wie gut ein Modell Verhalten beschreibt, das nicht im Ereignislog aufgezeichnet wurde. Ein "überformtes" Modell generalisiert nicht genug.

Geschäftsprozessmanagement (GPM): die Disziplin, die Wissen aus Informationssystemen und aus den Wirtschaftswissenschaften verknüpft und auf operationelle Geschäftsprozesse anwendet.

Konzeptverschiebung (Concept Drift): das Phänomen, dass Prozesse sich mit der Zeit verändern. Der beobachtete Prozess kann sich allmählich oder plötzlich aufgrund von saisonalen Änderungen oder erhöhtem Wettbewerb verändern, was die Analyse verkompliziert.

MXML: ein XML-basiertes Format zum Austausch von Ereignislogs. XES ersetzt MXML als neues werkzeug-unabhängiges Ereignislog-Format.

Operationelle Unterstützung (Operational Support): Analyse von Ereignisdaten zur Laufzeit mit dem Ziel laufende Prozessinstanzen zu überwachen und zu beeinflussen. Drei Aktivitäten können in der operationellen Unterstützung identifiziert werden: *erkennen* (erzeuge eine Warnung, wenn beobachtetes Verhalten vom modellierten Verhalten abweicht), *vorhersagen* (sage zukünftiges Verhalten auf Basis von früherem Verhalten voraus, z.B. sage die verbleibende Bearbeitungszeit voraus), und *empfehlen* (schlage passende Aktionen vor, um ein bestimmtes Ziel zu erreichen, z.B. Kosten minimieren).

Organisationsübergreifendes Process Mining (Cross-Organizational Process Mining): die Anwendung von Process-Mining-Techniken auf Ereignislogs verschiedener Organisationen.

Präzision (Precision): ein Maß, das feststellt ob das Modell Verhalten erlaubt, das grundlegend verschieden vom im Log aufgezeichneten Verhalten ist. Ein Modell mit geringer Präzision ist "unterformt".

Process Intelligence: ein Zweig der Business Intelligence, der sich auf Geschäftsprozessmanagement konzentriert.

Process Mining: Techniken, Werkzeuge und Methoden, um reale Prozesse (d.h. nicht bloß vermutete Prozesse), zu entdecken, zu beobachten und zu

verbessern, in dem Wissen von Ereignislogs, die in heutigen (Informations-)Systemen verfügbar sind, extrahiert wird.

Prozessentdeckung (Process Discovery): eine der drei grundlegenden Arten von Process Mining. Auf Basis eines Ereignislogs wird ein Prozessmodell erlernt. Beispielsweise kann der alpha-Algorithmus ein Petrinetz entdecken, in dem er Prozessmuster in einer Sammlung von Ereignissen identifiziert.

Prozessinstanz (Process Instance): die Entität, die vom Prozess, der analysiert wird, verarbeitet wird. Ereignisse beziehen sich auf Prozessinstanzen. Beispiele von Prozessinstanzen sind Kundenanträge, Schadensfälle einer Versicherung, Kreditanträge, etc.

Übereinstimmungsprüfung (Conformance Checking): Prüfung, ob die Realität, wie sie Ereignislog aufgezeichnet wurde, mit dem Modell übereinstimmt. Das Ziel ist es, Abweichungen zu erkennen und ihre Schwere zu messen.

Übereinstimmungsprüfung ist eine von drei grundlegenden Arten des Process Mining.

XES: ist ein XML-basierter Standard für Ereignislogs. Der Standard wurde von der IEEE Task Force on Process Mining als Standard-Austauschformat für Ereignislogs angenommen (s. www.xes-standard.org).