

Manifesto de Process Mining

Um manifesto é uma "declaração pública de princípios e intenções" de um grupo de pessoas. Este manifesto é escrito por membros e apoiantes do grupo de trabalho IEEE em Process Mining. O objectivo deste grupo de trabalho é a promoção da investigação, desenvolvimento, ensino, implementação, evolução,

e compreensão do Process Mining.

Process Mining é uma área de investigação relativamente recente que se posiciona não só entre Computational Intelligence e Data Mining, mas também entre a modelação e análise de processos. A ideia base do Process Mining é descobrir, monitorizar e melhorar processos reais (i.e., processos tal como são executados na realidade e não como desenhados em fluxogramas ou outro suporte) através da extracção de conhecimento a partir de registos de eventos usualmente disponíveis nos sistemas (de informação) existentes.

Process Mining inclui a descoberta automatizada de processos (i.e., obtenção de processos de negócio a

Conteúdo:

| | |
|---------------------------------|----|
| Process Mining – Estado da Arte | 3 |
| Normas de Orientação | 7 |
| Desafios | 11 |
| Epílogo | 16 |
| Glossário | 16 |

As técnicas de Process Mining estão orientadas para extrair conhecimento a partir de registos de eventos (event logs), frequentemente disponíveis nos sistemas de informação actualmente existentes. Estas técnicas oferecem novos meios para descobrir, monitorizar e melhorar processos num vasto domínio de aplicações. Existem dois factores principais para o crescente interesse em Process Mining. Por um lado, cada vez mais eventos são registados, facilitando assim a criação de históricos de processos. Por outro lado, existe a necessidade de melhorar processos de negócio em ambientes competitivos e de rápida evolução. Este manifesto é escrito pelo grupo de trabalho IEEE em Process Mining e tem como objectivo a promoção do Process Mining. Adicionalmente, através da definição de um conjunto de normas de orientação e da identificação de desafios importantes, este manifesto pretende servir de guia para engenheiros de software, investigadores, consultores, gestores de empresas e utilizadores finais. O objectivo é aumentar a maturidade do Process Mining como uma nova ferramenta para melhoria, controlo e suporte de processos de negócio operacionais.

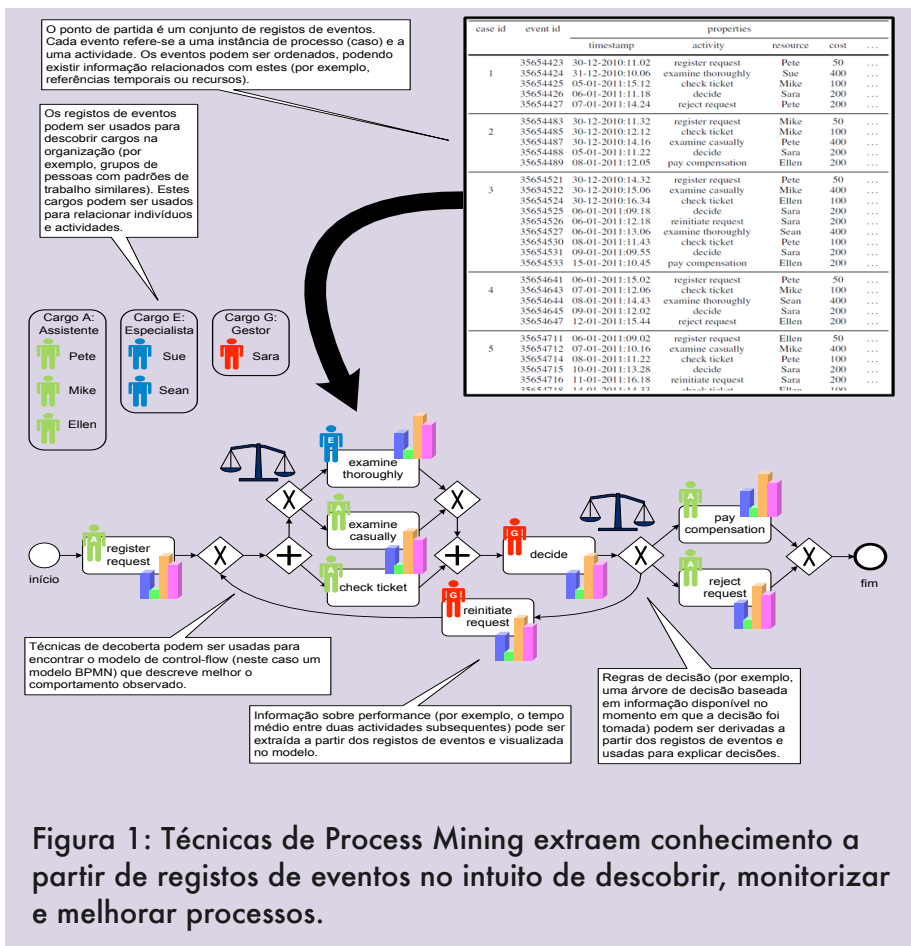


Figura 1: Técnicas de Process Mining extraem conhecimento a partir de registos de eventos no intuito de descobrir, monitorizar e melhorar processos.

Sigma estão também relacionadas. Estas abordagens caracterizam-se por "observar os processos ao microscópio" no intuito de se verificar se são possíveis eventuais aperfeiçoamentos. O Process Mining é uma tecnologia que facilita CPM, BPI, TQM, Six Sigma, e similares.

Enquanto ferramentas de BI e práticas de gestão tais como Six Sigma e TQM focam-se na optimização da performance operacional, como por exemplo a redução de defeitos e do tempo de execução, organizações estão também a orientar esforços em gestão corporativa, riscos, e conformidade. Legislação tal como Sarbanes-Oxley Act (SOX) e Basel II Accord evidencia o foco em questões de conformidade. Técnicas de Process Mining disponibilizam meios mais rigorosos para verificação de conformidade e averiguação da validade e fiabilidade da informação acerca dos processos centrais das organizações.

Ao longo da última década, dados de eventos têm-se tornado cada vez mais acessíveis e as técnicas de Process Mining têm evoluído. Adicionalmente, tal como já mencionado, as tendências de gestão relacionadas com melhoria de processos (por exemplo, Six Sigma, TQM, CPI, e CPM) e conformidade (SOX, BAM, etc.) podem beneficiar com Process Mining. Felizmente, algoritmos de Process Mining têm sido implementados em diversos sistemas de informação comerciais e académicos. Actualmente, existe um grupo activo de investigadores a trabalhar em Process Mining, tendo este tornado-se numa das áreas mais em foco na área de Business Process Management (BPM). Adicionalmente, a indústria tem demonstrado grande interesse em Process Mining. Cada vez mais empresas de software têm disponibilizado funcionalidades de Process Mining nos seus produtos. Exemplos de produtos de software com estas funcionalidades são: ARIS Process 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), e Reflect\$|sone (Pallas Athena). O

partir de um conjunto de registos de eventos), verificação de conformidade (i.e., monitorização de desvios através da comparação de modelos com registos de eventos), análise

organizacional ou de redes sociais, construção automatizada de modelos de simulação, extensão de modelos, reparação de modelos, previsão de casos, e recomendação baseada em históricos.

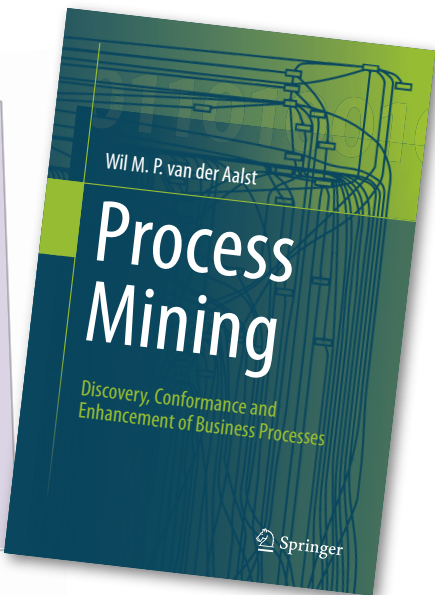
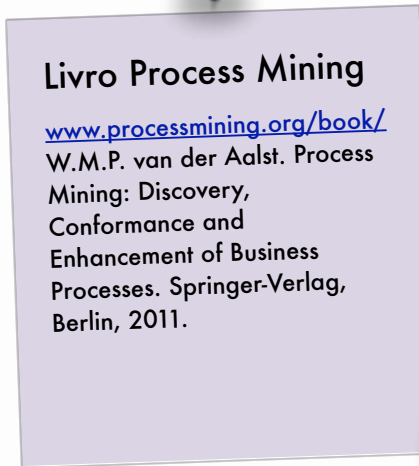
O Process Mining proporciona uma importante interligação entre Data Mining e modelação e análise de processos de negócio. Sob o domínio de Business Intelligence (BI) muitas palavras-chave têm sido mencionadas para identificar as - relativamente simples - ferramentas de monitorização e criação de relatórios. Business Activity Monitoring (BAM) abrange as tecnologias de monitorização em tempo real de processos de negócio. Complex Event Processing (CEP) engloba as tecnologias de processamento de largas quantidades de eventos, utilizando-as para monitorizar, orientar e optimizar o negócio em tempo real. Corporate Performance Management (CPM) são palavras-chave para identificar a avaliação de performance de um processo ou organização. Abordagens de gestão tais como Continuous Process Improvement (CPI), Business Process Improvement (BPI), Total Quality Management (TQM), e Six

Os objectivos do grupo de trabalho são:

- 1) A divulgação do estado da arte do Process Mining a investigadores, gestores de negócio, consultores, engenheiros de software, e utilizadores finais.
- 2) A promoção de técnicas e ferramentas de Process Mining e a estimulação de novas aplicações.
- 3) A participação no processo de normalização da operação de recolha de dados de eventos.
- 4) A organização de tutoriais, sessões especiais, workshops, fóruns de discussão.
- 5) A publicação de artigos, livros, vídeos, e edições especiais de revistas científicas.

crescente interesse em análise de processos baseada em registos motivou a formação deste grupo de trabalho.

O grupo de trabalho foi formado em 2009 no âmbito do Data Mining Technical Committee (DMTC) da Computational Intelligence Society (CIS) do Institute of Electrical and Electronic Engineers (IEEE). O actual grupo de trabalho é composto por representantes de *empresas de software* (por exemplo, Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Stereologic), *empresas de consultadoria/utilizadores finais* (por exemplo, ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chili, BWI Systeme GmbH, Excellentia BPM, Rabobank), e *institutos de investigação* (por exemplo, TU/e, University of Padua, Universitat Politècnica de Catalunya, New Mexico State University, IST - Universidade Técnica de Lisboa, University of Calabria, Penn State University, University of Bari, Humboldt-Universität zu Berlin, Queensland University of Technology, Vienna University of Economics and Business, 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).

Desde a sua formação em 2009, têm vindo a decorrer diversas actividades relacionadas com os objectivos atrás descritos. Por exemplo, vários workshops e sessões especiais foram (co-)organizados pelo grupo de trabalho, nomeadamente, os workshops em Business Process Intelligence (BPI'09, BPI'10, e BPI'11) e sessões especiais em conferências principais IEEE (por exemplo, CIDM'11). Conhecimento foi disseminado em tutoriais (por exemplo, WCCI'10 e PMPM'09), cursos de verão (ESSCaSS'09, ACPN'10, CICH'10, etc.), vídeos (cf.

www.processmining.org), e diversas publicações incluindo o primeiro livro acerca Process Mining recentemente publicado pela Springer. O grupo de trabalho também (co-)organizou o primeiro Business Process Intelligence Challenge (BPIC'11): uma competição na qual os participantes têm como objectivo extrair conhecimento relevante a partir de um grande e complexo conjunto de registos de eventos. Em 2010, o grupo de trabalho também adoptou o XES (www.xes-standard.org), um formato para registos de eventos que é extensível e suportado pela biblioteca OpenXES (www.openxes.org) e por ferramentas tais como ProM, XESame, Nitro, etc.

Para obter mais informação acerca das actividades do grupo de trabalho o leitor é convidado a visitar www.win.tue.nl/ieeetfpm/.

2. Process Mining: Estado da Arte

As capacidades em expansão dos sistemas de informação e outros sistemas computacionais estão bem caracterizados pela lei de Moore. Gordon Moore, o co-fundador da Intel, previu em 1965 que o número de componentes dos circuitos integrados iria duplicar a cada ano. Durante os últimos cinquenta anos esse crescimento exponencial verificou-se de facto, apesar de ter sido a um ritmo ligeiramente inferior. Estes avanços resultaram num crescimento espectacular do "universo digital" (i.e., todos os dados armazenados e/ou compartilhados electronicamente). Adicionalmente, o universo real e o digital continuam a aproximar-se.

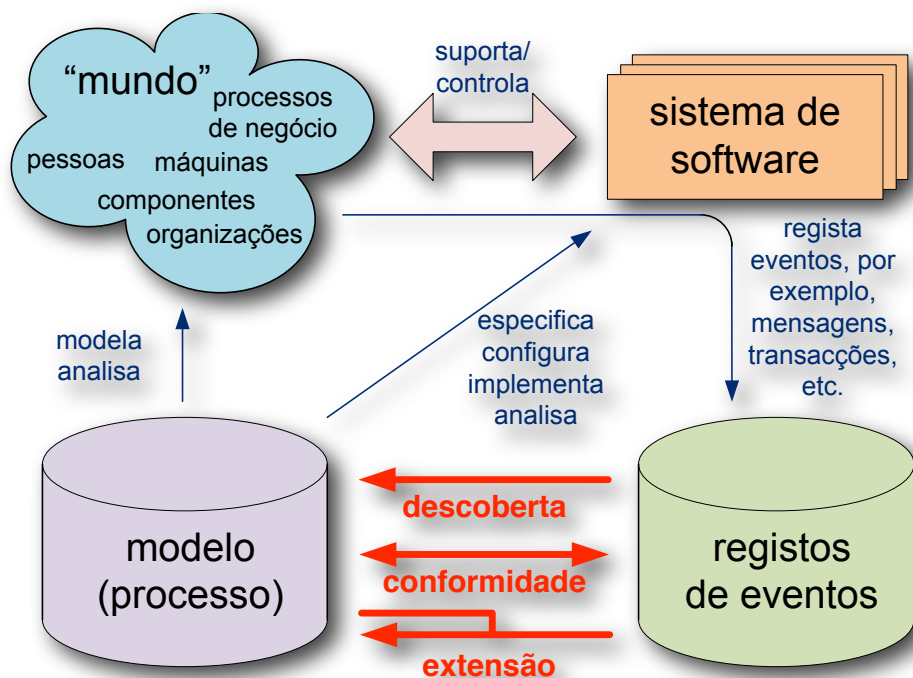


Figura 2: O posicionamento dos três tipos principais de Process Mining: (a) descoberta, (b) verificação de conformidade e (c) extensão.

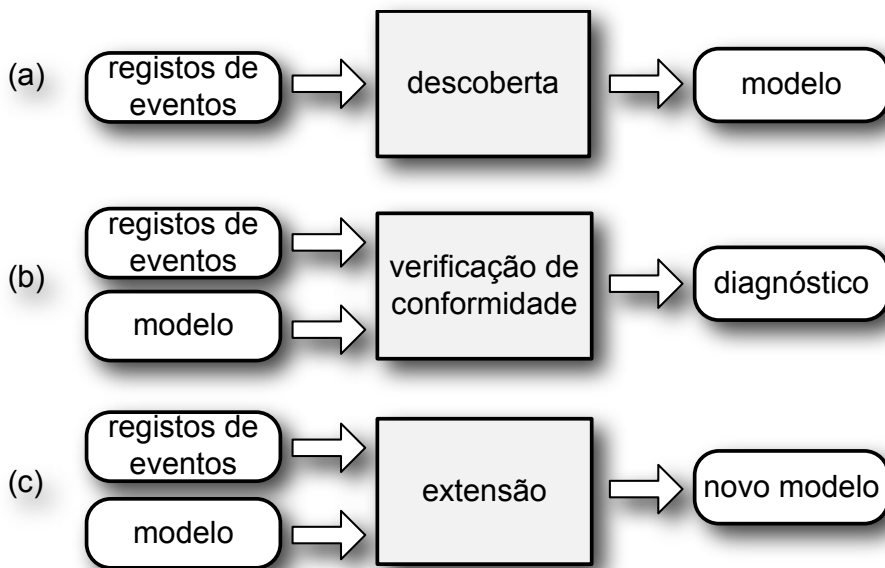


Figura 3: Os três tipos básicos de Process Mining em termos de inputs e outputs: (a) descoberta, (b) verificação de conformidade, e (c) extensão.

O crescimento do universo digital directamente relacionado com processos em organizações possibilita o registo e análise de eventos. Eventos podem ser vistos como o levantamento de dinheiro numa caixa multibanco, a afinação de uma máquina de raios X por um especialista, o pedido de renovação da carta de condução, o envio da declaração de impostos, e a recepção de um bilhete electrónico por um viajante. O desafio é explorar os dados de eventos de uma forma

proveitosa, como por exemplo, compreender o processo, identificar pontos de estrangulamento, antecipar problemas, registar violações de regras, recomendar medidas, e agilizar processos. O objectivo de Process Mining é exactamente este.

O ponto de partida para Process Mining é um conjunto de *registos de eventos*. Todas as técnicas de Process Mining partem do pressuposto que é possível registar sequencialmente eventos de tal forma que cada evento

representa uma *actividade* (i.e., uma etapa bem definida num processo) e está relacionado com um caso particular (i.e., uma instância de processo). Informação adicional sobre eventos poderá também ser mantida nos registos de eventos. Na verdade, quando disponível, técnicas de Process Mining usam informação complementar tal como os *recursos* (i.e., pessoa ou dispositivo) que executaram ou iniciaram a actividade, as *referências temporais* do evento, ou *elementos de dados* registados com o evento (por exemplo, a quantidade de produtos de uma ordem de compra).

Tal como ilustrado na Fig. 2, registos de eventos podem ser usados para executar três tipos de Process Mining. O primeiro tipo de Process Mining é *descoberta*. Uma técnica deste tipo usa exclusivamente um conjunto de registos de eventos para produzir um modelo (i.e., não requer qualquer tipo de informação à priori). Para muitas organizações é surpreendente o facto de ser possível descobrir processos reais apenas com base em históricos de processos. O segundo tipo de Process Mining é *conformidade*. Técnicas deste tipo comparam modelos de processo existentes com um conjunto de registos de eventos do mesmo processo. Verificação de conformidade pode ser usada para verificar se o processo tal

Características do Process Mining:

1. O Process Mining não está limitado a descoberta de control-flow.

A descoberta de modelos de processo a partir de registos de eventos dá asas à imaginação de praticantes e académicos. Desta forma, a descoberta de control-flow é frequentemente vista como a aplicação mais interessante de Process Mining. Contudo, Process Mining não se limita a descoberta de control-flow. Descoberta de processos é simplesmente uma das três formas básicas de Process Mining (descoberta, conformidade e extensão). Assim sendo, o âmbito não está limitado a control-flow; as perspectivas da organização, caso e tempo também desempenham um papel importante.

2. O Process Mining não é simplesmente um tipo específico de Data Mining.

Process Mining pode ser visto como o "elo em falta" entre Data Mining e BPM tradicional. A maioria das técnicas de Data Mining não são orientadas a processos. Podendo potencialmente descrever concorrência, modelos de processo não são comparáveis às estruturas simples de Data Mining tais como árvores de decisão e regras de associação. Assim sendo, são necessários tipos de representação e algoritmos completamente novos.

3. O Process Mining não está limitado a análises restritas.

Técnicas de Process Mining extraem conhecimento a partir de dados de eventos históricos. Apesar de serem usados dados referentes a acontecimentos passados, os resultados podem ser aplicados a casos em execução. Por exemplo, o tempo de conclusão de uma ordem de compra em execução pode ser previsto usando um modelo de processo.

como descrito nos registos de eventos está em conformidade com os modelos e vice-versa. É importante referir que diferentes tipos de modelos podem ser considerados: verificação de conformidade pode ser aplicada em modelos processuais, modelos organizacionais, modelos de processo declarativos, regras de negócio, leis, etc. O terceiro tipo de Process Mining é *extensão*. Técnicas deste tipo complementam ou aperfeiçoam um modelo de processo existente usando informação acerca de um processo descrito num conjunto de registos de eventos. Enquanto a verificação de conformidade determina o posicionamento entre o modelo e a realidade, este terceiro tipo de Process Mining tem como objectivo complementar ou aperfeiçoar o modelo fornecido à priori. Por exemplo, usando as referências temporais dos registos de eventos, é possível complementar-se o modelo de modo a que informação sobre níveis de serviço, tempos de execução e frequências seja também considerada para análise.

A Fig. 3 descreve os três tipos de Process Mining em termos de inputs e outputs. Técnicas de descoberta usam um conjunto de registos de eventos para produzir um modelo. O modelo resultante é tipicamente um modelo de processo (por exemplo, rede de Petri, BPMN, EPC, ou diagrama de actividade UML), contudo, o modelo pode também descrever outras perspectivas (por exemplo, rede social). Técnicas de verificação de conformidade usam um conjunto de registos de eventos e um modelo como

inputs. O output consiste em informação de diagnóstico identificando diferenças ou semelhanças entre modelos e registos. Técnicas de extensão de modelos (reparação ou aperfeiçoamento) também requerem um conjunto de registos de eventos e um modelo como inputs. O output será um modelo estendido.

O Process Mining poderá abranger diferentes perspectivas. A *perspectiva de control-flow* foca-se na ordenação de actividades. O objectivo desta perspectiva é encontrar uma boa caracterização de todas as sequências possíveis. O resultado é expressado tipicamente em termos de uma rede de Petri ou alguma outra notação de processo (por exemplo, EPCs, BPMN, ou diagramas de actividade UML). A *perspectiva organizacional* foca-se na informação acerca de recursos implícita nos registos, i.e., quais os actores (por exemplo, pessoas, sistemas, funções, ou departamentos) envolvidos e como estes se relacionam. O objectivo é tanto descrever a rede social como estruturar a organização através da classificação de pessoas em termos de funções e unidades organizacionais. A *perspectiva de caso* foca-se em aspectos de casos de processo. Um caso pode obviamente ser caracterizado pelo seu percurso no processo ou pelos actores que o executam. Contudo, casos podem ser também caracterizados por valores descritos nos dados. Por exemplo, se um caso representa uma ordem de reposição de stock, poderá ser interessante saber qual é o fornecedor ou o número de produtos requisitados.

A *perspectiva temporal* diz respeito ao timing e à frequência de eventos. Sendo que os eventos são anotados com referências temporais, é possível a descoberta de pontos de estrangulamento, medição dos níveis de serviço, monitorização da utilização de recursos, e previsão do tempo de processamento necessário para conclusão dos casos existentes.

Existem alguns equívocos comuns relacionados com Process Mining. Alguns vendedores, analistas e investigadores limitam o âmbito de Process Mining a uma técnica especial de Data Mining para descoberta de processos que pode ser somente usada para análises restritas. Não sendo este o caso, salienta-se portanto as três características na caixa de texto da página anterior.

O ciclo de vida de BPM ilustrado na Fig. 4 é usado para posicionar o Process Mining. O ciclo de vida de BPM apresenta as sete fases de um processo de negócio e correspondente(s) sistema(s) de informação. Na *fase de (re)desenho* é criado um novo modelo de processo ou adaptado um modelo existente. Na *fase de análise* são analisados vários modelos candidatos. Depois da fase de (re)desenho, o modelo é implementado (*fase de implementação*) ou um sistema existente é (re)configurado (*fase de reconfiguração*). Na *fase de execução* o modelo escolhido é posto em prática. Durante esta fase o processo é *monitorizado*. Adicionalmente, pequenos ajustes poderão ser feitos sem grandes alterações no processo (*fase de ajuste*). Na *fase de diagnóstico* o processo posto em prática é analisado, sendo que os resultados desta fase poderão dar origem à repetição da fase de redesenho. Process Mining é uma ferramenta de valor para a maior parte das fases apresentadas na Fig. 4. Obviamente, a fase de diagnóstico pode beneficiar com Process Mining. Contudo, Process Mining não se limita somente a esta fase. Por exemplo, na fase de execução, técnicas de Process Mining podem ser usadas para *suporte operacional*. Previsões e recomendações baseadas em modelos originários de dados históricos podem ser usadas para auxiliar casos em execução. Formas similares de suporte à decisão podem ser usadas para ajustar processos e orientar o processo de (re)configuração.

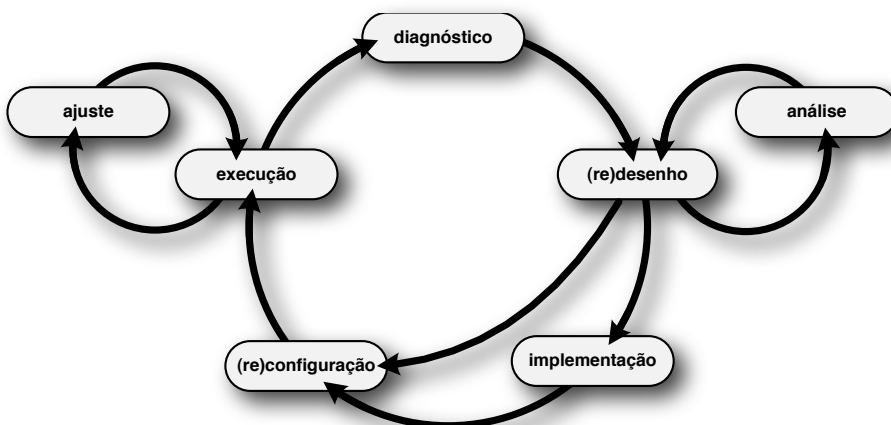


Figura 4: O ciclo de vida de BPM identificando as várias fases de um processo de negócio e correspondente(s) sistema(s) de informação; Process Mining está (potencialmente) presente em todas as fases (excepto na fase de implementação).

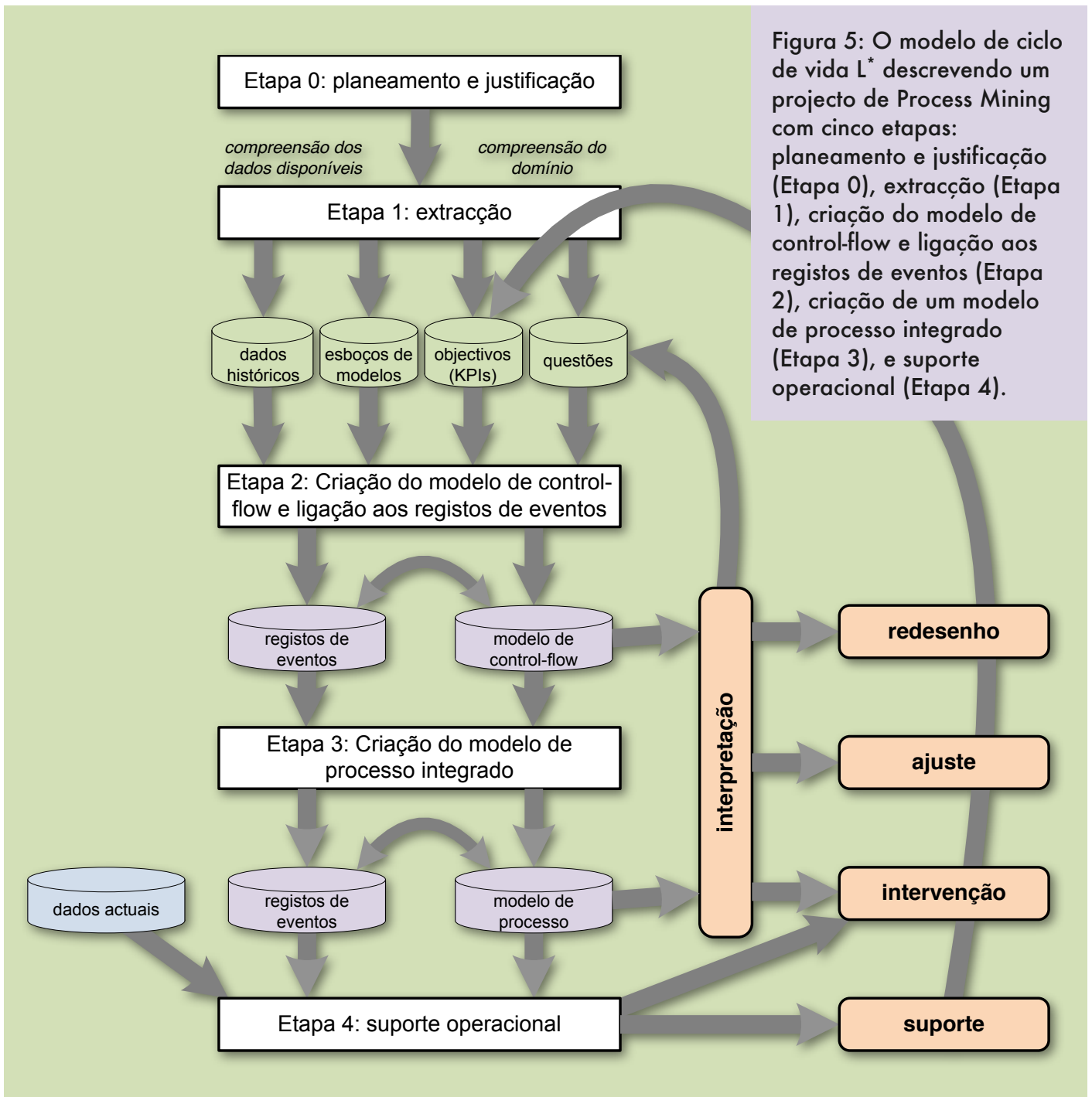


Figura 5: O modelo de ciclo de vida L* descrevendo um projecto de Process Mining com cinco etapas: planeamento e justificação (Etapa 0), extração (Etapa 1), criação do modelo de control-flow e ligação aos registos de eventos (Etapa 2), criação de um modelo de processo integrado (Etapa 3), e suporte operacional (Etapa 4).

Enquanto Fig. 4 apresenta o ciclo de vida de BPM, Fig. 5 foca-se em actividades e artefactos concretos de Process Mining. Fig. 5 descreve as possíveis etapas de um projecto de Process Mining. Qualquer projecto de Process Mining começa com um planeamento e uma correspondente justificação (Etapa 0). Depois do processo ser inicializado, dados referentes a eventos, modelos, objectivos, e questões precisam de ser obtidos a partir de sistemas, especialistas de domínio, e gestão (Etapa 1). Isto requer compreensão dos dados disponíveis ("O que é que pode ser usado para análise?") e

compreensão do domínio ("Quais são as questões importantes?"), e resulta em artefactos tal como os apresentados na Fig. 5 (i.e., dados históricos, esboços de modelos, objectivos e questões). Na Etapa 2 o modelo de control-flow é contruído e ligado aos registos de eventos. Nesta etapa, técnicas de Process Mining automatizadas podem ser aplicadas. O modelo de processo descoberto já poderá responder a algumas das questões e desencadear acções de redesenho e ajuste. Adicionalmente, os registos de eventos poderão ser filtrados e adaptados usando o modelo (por exemplo, remover casos raros ou

outliers, e inserir eventos em falta). Por vezes esforços consideráveis são necessários para correlacionar eventos pertencentes à mesma instância de processo. Os eventos restantes são relacionados com entidades do modelo de processo. Quando o processo estiver relativamente estruturado, o modelo de control-flow poderá ser estendido com outras perspectivas (por exemplo, dados, tempo, e recursos) durante a Etapa 3. A relação entre os registos de eventos e o modelo resultante da Etapa 2 é usada para estender o modelo (por exemplo, referências temporais dos registos associados são usadas para estimar

tempos de espera de actividades). Isto pode ser usado para encontrar resposta a questões adicionais e poderá desencadear novas acções. Por fim, os modelos resultantes da Etapa 3 poderão ser usados para suporte operacional (Etapa 4). Conhecimento extraído a partir de dados históricos de eventos é combinado com informação acerca dos casos em execução. Isto pode ser usado para intervir, prever, e fazer recomendações. Etapas 3 e 4 somente podem ser alcançadas se o processo for suficientemente estável e estruturado.

Actualmente, existem técnicas e ferramentas que suportam todas as etapas apresentadas na Fig. 5. Contudo, Process Mining é um paradigma relativamente recente e grande parte das ferramentas actualmente disponíveis ainda não atingiram um bom estado de maturidade. Para além disso, é frequente que os utilizadores prospectivos não tenham consciência do potencial e das limitações de Process Mining. Por isto, este manifesto cataloga algumas normas de orientação (cf. secção seguinte) e desafios (cf. página 11) para utilizadores de técnicas de Process Mining e também investigadores e contribuintes que estejam interessados nos avanços do estado da arte.

3. Normas de Orientação

Tal como em qualquer nova tecnologia, existem erros obvios que podem ser cometidos aquando da aplicação de Process Mining em casos reais. Assim sendo, seis normas de orientação são descritas para prevenir utilizadores/analistas de cometer tais erros.

GP1: Dados de eventos devem ser tratados como passageiros de primeira classe

Eventos registados são o ponto de partida para qualquer actividade de Process Mining. Apesar de registos de eventos serem referidos como colecções de eventos, isto não implica que eventos tenham de ser mantidos em arquivos de registos dedicados. Eventos podem estar armazenados em

tabelas de bases de dados, registos de mensagens, arquivos de correio electrónico, registos de transacções, e outras fontes de dados. Mais importante que o formato de arquivo, é a *qualidade* dos registos de eventos. A qualidade dos resultados de Process Mining depende largamente do input usado. Assim sendo, registos de eventos devem ser tratados como *passageiros de primeira classe* nos sistemas de informação que suportam os processos em análise. Infelizmente, registos de eventos são muitas vezes tidos como "subprodutos" usados para depuração (debug) ou caracterização (profile). Por exemplo, os dispositivos médicos da Philips Healthcare registam eventos simplesmente porque os engenheiros de software introduziram no código "comandos de impressão no ecrã". Apesar de existirem algumas normas informais para introdução de tais comandos no código, é necessária uma abordagem mais sistemática para melhorar a qualidade de registos de eventos. Dados de eventos devem ser vistos como passageiros de primeira classe (em vez de passageiros de segunda).

Existem vários critérios para avaliar a qualidade de dados de eventos. Eventos devem ser *fidedignos*, i.e., deverá ser seguro assumir que os eventos registados foram realmente executados e que a informação destes eventos é verdadeira. Registos de eventos devem ser *completos*, i.e., dada uma determinada aplicação, não existem eventos em falta. Qualquer evento registado deve ter uma *semântica* bem definida. Adicionalmente, os dados de eventos devem ser *seguros* no sentido de que privacidade e segurança são tidos em conta na operação de registo dos eventos. Por exemplo, os agentes devem estar cientes dos tipos de eventos que estão a ser registados e a forma como serão usados.

A Tabela 1 descreve os cinco níveis de maturidade de registos de eventos, desde excelente qualidade (★★★★★) até fraca qualidade (★). Por exemplo, os registos de eventos da Philips Healthcare situam-se no nível ★★★, i.e., eventos são registados automaticamente e o comportamento registado corresponde à realidade. Contudo, nenhuma abordagem sistemática é aplicada para definir a semântica de eventos e assegurar o alcance de um nível particular. Técnicas

Normas de Orientação:

GP1: Dados de eventos devem ser tratados como passageiros de primeira classe

GP2: Extracção de registos deve ser impulsionada por questões

GP3: Concorrência, escolha e outras construções básicas de control-flow devem ser suportados

GP4: Os eventos devem estar relacionados com os elementos do modelo de processo

GP5: Os modelos devem ser tratados como abstrações intencionais da realidade

GP6: Process Mining deve ser um processo contínuo

de Process Mining podem ser aplicadas em registos que se situam nos níveis ★★★★★, ★★★★★, e ★★★. Em princípio, a aplicação do Process Mining em registos de ★★ ou ★ é também possível. Contudo, a análise de tais registos é tipicamente problemática e os resultados podem não ser fidedignos. Na verdade, não faz muito sentido aplicar Process Mining em registos de nível ★.

Para tirar o máximo partido do Process Mining, organizações devem concentrar-se em registos de eventos com máximo nível de qualidade possível.

GP2: Extracção de registos deve ser impulsionada por questões

Tal como ilustrado na Fig. 5, actividades de Process Mining precisam de ser impulsionadas por questões. Sem que hajam questões concretas é bastante difícil extrair dados relevantes de eventos. Considere-se, por exemplo, as milhares de tabelas de uma base de dados de um sistema ERP (por exemplo, SAP). A selecção de tabelas para extracção de dados não será possível se não existirem questões

| Nível | Caracterização | Exemplos |
|-------|--|--|
| ★★★★★ | Nível mais alto: os registos de eventos são de excelente qualidade (i.e., fidedignos e completos) e eventos são bem definidos. Eventos são registados de forma automática, sistemática, consistente, e segura. Questões de privacidade e segurança são tratadas de forma adequada. Adicionalmente, os eventos registados (e todos os seus atributos) têm uma semântica clara. Isto implica a existência de uma ou mais ontologias nas quais eventos e correspondentes atributos são explicados. | Registos de sistemas de BPM anotados com semântica. |
| ★★★★ | Eventos são registados automaticamente e de uma forma sistemática e consistente, i.e., registos são fidedignos e completos. Ao contrário de sistemas que operam ao nível ★★★, conceitos como instâncias (casos) de processos e actividade são suportados de uma forma explícita. | Os registos de eventos de sistemas tradicionais de BPM/workflow. |
| ★★★ | Eventos são registados automaticamente, apesar de não existir qualquer abordagem sistemática no registo de eventos. Contudo, ao contrário dos registos de nível★★, existem garantias que os eventos registados correspondem à realidade (i.e., os registos de eventos são fidedignos mas não necessariamente completos). Considere-se, por exemplo, os eventos registados por um sistema ERP. Apesar de ser preciso extrair os eventos de várias tabelas, a informação pode ser considerada correcta (por exemplo, é seguro assumir que um pagamento registado pelo sistema realmente ocorreu e vice versa). | Tabelas em sistemas ERP, registos de eventos em sistemas CRM, registos de transacções de sistemas de mensagens, registos de eventos de sistemas de alta tecnologia, etc. |
| ★★ | Eventos são registados automaticamente, i.e., como um subproduto de um sistema de informação. O alcance varia, i.e., nenhuma abordagem sistemática é seguida para decidir que eventos devem ser registados. Para além disso, o sistema de informação pode ser facilmente contornável. Assim sendo, eventos poderão não ser registados ou ser inapropriadamente. | Registos de eventos de sistemas de gestão de produtos e documentos, registos de erros de sistemas embutidos, folhas de cálculo de engenheiros, etc. |
| ★ | Nível mais baixo: os registos de eventos são de fraca qualidade. Eventos registados poderão não corresponder à realidade e poderão haver eventos em falta. Registos de eventos elaborados manualmente apresentam tipicamente estas características. | Indícios deixados em documentos de papel ao longo do seu percurso pela organização ("post-its"), registos médicos em papel, etc. |

Tabela 1: Níveis de maturação de registos de eventos.

concretas.

Um modelo de processo tal como o ilustrado na Fig. 1 descreve o ciclo de vida de casos (i.e., instâncias de processos) de um tipo específico. Assim, antes de se aplicar qualquer técnica de Process Mining, é necessário escolher o tipo de caso a ser analisado. Esta escolha, podendo não ser fácil, deve ser impulsionada pelas questões a ser respondidas. Considere-se, por exemplo, a execução de ordens de compra de clientes. Cada ordem poderá consistir em múltiplos itens caso o cliente pretenda múltiplos

productos numa só ordem, e resultar em múltiplas entregas. Cada entrega poderá agregar itens de múltiplas ordens. Assim sendo, pode-se concluir que existe uma relação muitos-para-muitos entre ordens e entregas, e uma relação um-para-muitos entre ordens e itens. Usando uma base de dados que suporte as operações descritas é possível descobrir diferentes modelos de processo. Por um lado, é possível extrair dados com o objectivo de descrever o ciclo de vida de ordens de compra específicas. Por outro lado, é também possível extrair dados com o

objectivo de descobrir os ciclos de vida de certos itens ou entregas.

GP3: Concorrência, escolha e outras construções básicas de control-flow devem ser suportados

Existe actualmente um grande número de linguagens de modelação de processos (por exemplo, BPMN, EPCs, redes de Petri, BPEL, e diagramas de actividade UML). Algumas destas

linguagens oferecem um vasto número de elementos de modelação (por exemplo, BPMN disponibiliza mais de 50 elementos gráficos distintos) enquanto outras primam pela simplicidade (por exemplo, redes de Petri são simplesmente compostas por três elementos diferentes: posições, transições, e arcos). A descrição de control-flow é a espinha dorsal de qualquer modelo de processo. Construções básicas de workflow (também conhecidas como *padrões*) suportadas pela generalidade das linguagens são sequência, encaminhamento paralelo (AND-splits/joins), escolha (XOR-splits/joins), e ciclos. Obviamente, as técnicas de Process Mining devem ser capazes de suportar estes padrões. Contudo, existem casos em que não é possível lidar com concorrência, somente sistemas de transição/cadeias de Markov têm essa capacidade.

A Fig. 6 ilustra o efeito da aplicação de técnicas de Process Mining incapazes de lidar com concorrência (não existem quaisquer AND-split/joins). Considere-se um conjunto de registos de eventos $L = \{ \langle A, B, C, D, E \rangle, \langle A, B, D, C, E \rangle, \langle A, C, B, D, E \rangle, \langle A, C, D, B, E \rangle, \langle A, D, B, C, E \rangle, \langle A, D, C, B, E \rangle \}$. L contém casos que começam com A e terminam com E. As actividades B, C, e D ocorrem aleatoriamente entre A e E. O modelo BPMN da Fig. 6(a) apresenta uma representação compacta do processo subjacente usando dois gateways AND. Neste caso, os outros dois modelos BPMN da Fig. 6 são candidatos óbvios. O modelo BPMN da Fig. 6(b) é compacto mas permite demasiado comportamento (por exemplo, segundo o modelo, são permitidos casos como $\langle A, B, B, B, E \rangle$ que não têm grande suporte nos registos de eventos). O modelo BPMN da Fig. 6(c) suporta todos os casos de L, mas pelo o facto de os representar explicitamente, a representação do processo subjacente não é compacta. O exemplo mostra que, para modelos de casos reais em que podem existir dezenas de actividades potencialmente concorrentes, os modelos resultantes podem ser excessivamente abstractos (i.e., permitem demasiado comportamento) e/ou extremamente complexos se comportamentos

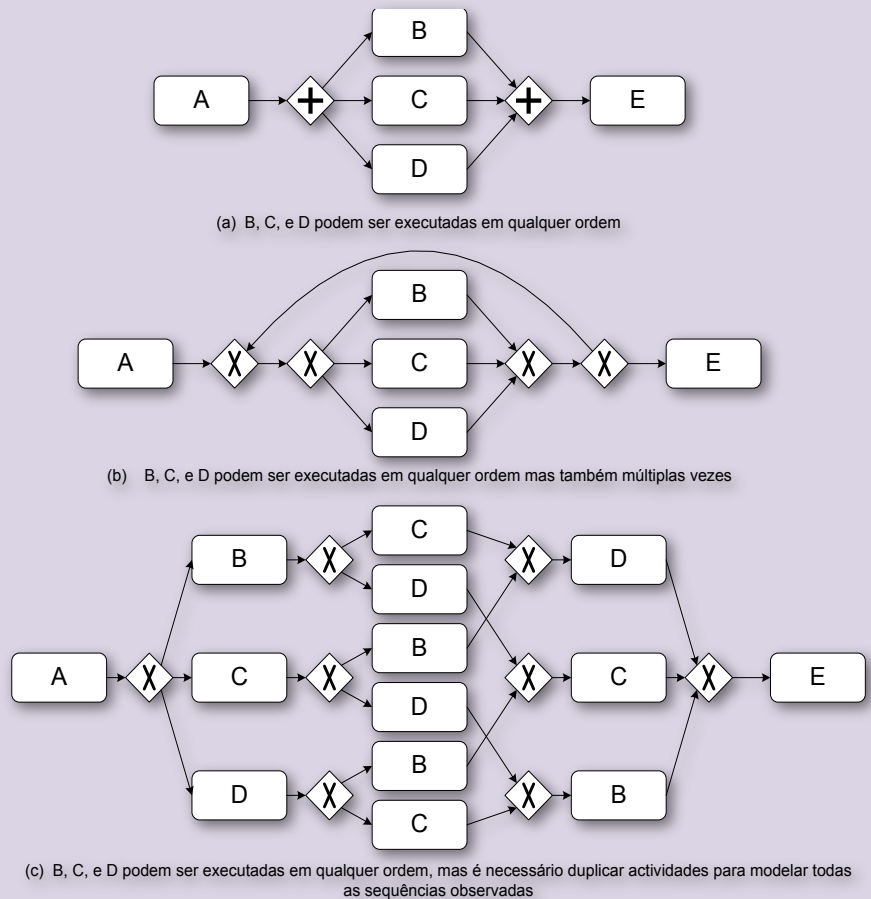


Figura 6: Exemplificação de problemas em que não é possível representar directamente concorrência (i.e., AND-splits/joins). No exemplo somente três actividades são concorrentes (B, C, e D). Caso existissem, por exemplo, 10 actividades concorrentes, seria necessário construir um modelo com $2^{10}=1024$ estados e $10! = 3.628.800$ sequências de execução.

concorrentes não forem suportados.

Tal como ilustrado na Fig. 6, é importante que, pelo menos, os padrões básicos de workflow sejam suportados. Para além dos padrões básicos já mencionados, é também recomendável o suporte a OR-splits/joins porque estes oferecem uma representação compacta de decisões e sincronizações parciais.

GP4: Os eventos devem estar relacionados com os elementos do modelo de processo

Como indicado na Secção 2, é um equívoco assumir que o Process Mining se limita apenas à descoberta de

modelos de processo. Como mostrado na Fig. 1, o modelo de processo descoberto pode abranger várias perspectivas (organizacional, tempo, dados, etc.) Além disso, a descoberta é apenas um dos três tipos de Process Mining representados na Fig. 3. Os outros dois tipos (verificação de conformidade e expansão do modelo) dependem fortemente da relação entre os elementos que constituem o modelo e os eventos descritos nos registos. Essa relação pode ser utilizada para a "reconstituição" do conjunto de registos de eventos no modelo. A reconstituição poderá ser utilizada para revelar discrepâncias entre os registos de eventos, a realidade e o modelo. Por exemplo, alguns eventos nos registos não são passíveis de ocorrer de acordo

com o que se encontra definido no modelo. Para esse efeito, podem ser usadas técnicas de verificação de conformidade para quantificar e diagnosticar tais discrepâncias. As referências temporais nos registos de eventos podem ser usadas para analisar o comportamento temporal durante a reconstituição do processo. Diferenças temporais entre actividades que têm uma relação de causalidade podem ser usadas para anotar o modelo com tempos de espera esperados. Estes exemplos mostram que a relação entre os eventos descritos nos registos e os elementos no modelo serve como ponto de partida para diferentes tipos de análise.

Em alguns casos, pode não ser trivial estabelecer essa relação. Por exemplo, um evento pode referir-se a duas actividades diferentes ou não ser claro a qual actividade se refere. Tais ambiguidades necessitam de ser removidas, a fim de interpretar os resultados de Process Mining correctamente. Além do problema de relacionar eventos a actividades, há ainda o problema de relacionar eventos com instâncias do processo. Isto é comumente referido como *correlação de eventos*.

GP5: Os modelos devem ser tratados como abstrações intencionais da realidade

Os modelos derivados a partir de dados dos eventos fornecem *pontos de vista sobre a realidade*. Tal ponto de vista deve proporcionar uma abstracção intencional de determinado comportamento capturado nos registos de eventos. Para um determinado conjunto de registos de eventos, pode haver múltiplos pontos de vista que são úteis. Além disso, as diversas partes interessadas podem pretender analisar diferentes pontos de vista. Na verdade, modelos descobertos a partir de registos de eventos devem ser vistos como "mapas" (como mapas geográficos). Esta norma de orientação fornece informação importante, alguma da qual encontra-se abaixo descrita.

Em primeiro lugar, é importante notar que não existe o conceito de um mapa "para determinada área geográfica". Dependendo do uso pretendido, existem mapas para

diversos fins: mapas de estradas, mapas para caminhada, mapas para ciclismo, etc. Todos estes mapas mostram uma visão sobre a mesma realidade e seria absurdo supor que haveria uma coisa como o "mapa perfeito". O mesmo acontece para os mapas de processos: o modelo deve enfatizar as coisas relevantes para um determinado tipo de utilizador. Os modelos de processo descobertos podem-se concentrar em perspectivas diferentes (control-flow, dados, tempo, recursos, custos, etc) e mostrar os diferentes níveis de granularidade e precisão, por exemplo, um gestor pode querer ver de forma grosseira um modelo de processo com foco nos custos e um analista de processos pode querer ver um modelo de processo detalhado com foco nos desvios do fluxo. Por outro lado diferentes partes interessadas podem querer ver um processo em diferentes níveis de análise: *estratégico* (decisões a este nível têm impacto a longo prazo e são baseadas em dados de eventos agregados ao longo de um período mais longo), *tático* (decisões neste nível têm impacto a médio prazo e são na sua maioria baseados em dados recentes) e *operacional* (decisões a este nível têm impacto imediato e são baseados em dados de eventos relacionados com casos que se encontram em execução).

Em segundo lugar, é útil adoptar ideias de cartografia, quando se trata de produzir modelos de processo que sejam compreensíveis. Por exemplo, mapas de estradas nacionais não costumam ter representados caminhos agrícolas e aldeias. Objectos menos significantes são deixados de fora ou dinamicamente agrupados (por exemplo, as ruas das cidades são fundidas dentro da localização geográfica da cidade). Os cartógrafos não só eliminam detalhes irrelevantes, mas também usam cores para realçar características importantes. Além disso, elementos gráficos têm um tamanho especial para indicar sua significância (por exemplo, o tamanho das linhas e pontos podem variar, quando representam auto-estradas ou capitais de distrito). Mapas geográficos também têm uma clara interpretação do eixo do x e do eixo do y, ou seja, o layout de um mapa não é arbitrário assim como as coordenadas de elementos têm um significado. Esta abordagem cartográfica está em

contraste com os modelos de processos convencionais que geralmente não usam cor, tamanho, localização e características que o tornem mais compreensíveis. No entanto, as ideias da cartografia podem facilmente ser incorporadas na descoberta de mapas de processos. Por exemplo, o tamanho de uma actividade pode ser usado para reflectir a frequência de execução ou outra propriedade, indicando a sua significância estatística (por exemplo, custos, ou utilização de recursos). A largura de um arco pode reflectir a importância da dependência causal, a coloração de arcos podem ser usada para destacar pontos de estrangulamento na execução do processo.

As considerações acima mostram que é importante escolher a representação certa e ajusta-la para determinado público-alvo. Isto é importante para a visualização de resultados para os utilizadores finais e para os guiar nos algoritmos de descoberta de modelos adequados (ver também Desafio C5).

GP6: Process Mining deve ser um processo contínuo

O Process Mining pode ajudar a fornecer de forma clara "mapas" de processo que estão directamente ligados aos dados dos eventos ocorridos. Tanto os dados de eventos históricos e os dados actuais podem ser projectados em tais modelos. Não nos podemos esquecer que os processos vão mudando à medida que vão sendo analisados. Dada a natureza dinâmica dos processos, não é aconselhável olhar para o Process Mining como um processo que se executa uma única vez. O objectivo não deve ser o de criar um modelo rígido, mas sim permitir aos utilizadores e analistas a possibilidade de olharem para os mapas de processo de forma contínua e diária.

Compare isso com o uso de mashups utilizando etiquetas georreferenciadas. Existem milhares de mashups usando o Google Maps (por exemplo, aplicações para disponibilizar informações sobre as condições de trânsito, imobiliário, restaurantes de fast food, ou horários de exibição de um filme num determinado mapa seleccionado). As pessoas podem facilmente aumentar ou diminuir o zoom e utilizar esses mapas

para interagir com eles (por exemplo, os congestionamentos de tráfego são disponibilizados no mapa e o utilizador pode seleccionar um problema particular para ver os detalhes). Também deve ser possível conduzir o Process Mining utilizando dados em tempo real. Usando a metáfora dos "mapas cartográficos" podemos pensar em eventos com coordenadas GPS que podem ser disponibilizados em tempo real. De forma análoga aos sistemas de navegação automóvel, as ferramentas de Process Mining podem ajudar os utilizadores (i) facilitando a navegação nos processos, (ii) mostrando informação dinâmica em mapas de processo (por exemplo, mostrando "os congestionamentos de tráfego" em processos de negócio), e (iii) fornecendo previsões sobre os casos que se encontram em execução (por exemplo, estimar o tempo de conclusão de um caso que está atrasado). Estes exemplos demonstram que é uma pena não usar os modelos de processo de forma mais activa. Portanto, o Process Mining deve ser visto como um processo contínuo que disponibiliza informação accionável em diversas escalas de tempo (minutos, horas, dias, semanas e meses).

4. Desafios

O Process Mining é uma ferramenta importante para as organizações modernas que precisam de gerir processos operacionais. Por um lado, há um crescimento incrível de dados sobre os eventos processados. Por outro lado, processos e informação precisam ser perfeitamente alinhados, a fim de atender aos requisitos relacionados com a conformidade, eficiência e serviço ao cliente. Apesar das capacidades de utilização do Process Mining, ainda subsistem desafios importantes que precisam ser abordados. Estes ilustram que o Process Mining é uma técnica emergente como disciplina de gestão dos processos de negócio. Abaixo, listamos alguns destes desafios. Esta lista não tem como objectivo ser completa. Ao longo do tempo, devido aos avanços que sejam alcançados com a utilização do Process Mining, novos desafios podem surgir ou desafios existentes podem desaparecer.

C1: Procurar, integrar e melhorar a qualidade dos dados de eventos

A extracção de dados de eventos ainda requer esforços consideráveis. Normalmente, vários obstáculos precisam ser superados:

- Os dados podem estar *distribuídos* por uma variedade de sistemas. Esses dados precisam ser integrados. Isto tende a ser problemático quando diferentes identificadores, chaves ou apontadores, são usados nas diversas fontes de dados. Por exemplo, um sistema usa nome e data de nascimento para identificar uma pessoa enquanto outro sistema usa o número da segurança social.
- Os dados dos eventos estão muitas vezes em torno dos objectos em vez de estarem em torno do processo. Por exemplo, produtos alimentares, paletes e contentores podem ter etiquetas RFID e os eventos registados referem-se a essas etiquetas. No entanto, para monitorizar uma encomenda de um cliente esses dados centrados nos objectos com os quais estão relacionados, precisam ser integrados e pré-processados.
- Os dados relativos aos eventos podem estar *incompletos*. Um problema comum é que os eventos não apontam explicitamente para as instâncias do processo. Muitas vezes é possível obter esta informação, mas isso pode levar a um esforço considerável. Pode também ocorrer que informação temporal sobre os eventos pode estar em falta. Pode ser necessário efectuar interpolação das referências temporais, a fim de continuar a utilizar a dimensão tempo nas análises que se pretendem efectuar.
- Um conjunto de registos de eventos pode conter *valores atípicos*, excepcionais, também conhecidos como *ruído*. Como definir o que é ruído nos registos? Como detectar o ruído? Estas perguntas precisam ser respondidas para assegurar a qualidade dos dados.
- Os registos de eventos podem conter informação com *níveis diferentes de granularidade*. Nos registos de eventos de um sistema de informação

Desafios:

C1: Procurar, integrar e melhorar a qualidade dos dados de eventos

C2: Lidar com registos de eventos complexos e com diversas características

C3: Criar valores de referência para estabelecer pontos de comparação

C4: Lidar com o desvio do conceito do processo

C5: Melhorar a representação de resultados para a descoberta de processos

C6: Equilíbrio entre a qualidade dos critérios como Robustez, Simplicidade, Precisão e Generalização

C7: Process Mining inter-organizacional

C8: Fornecer suporte operacional

C9: Utilizar Process Mining conjuntamente com outro tipo de análises

C10: Melhorar a usabilidade para os que não são especialistas

C11: Melhorar a compreensão para os que não são especialistas

hospitalar, eventos registados no sistema podem referir-se a exames sanguíneos simples ou a procedimentos cirúrgicos complexos. As referências temporais também podem ter níveis diferentes de granularidade que variam entre milissegundos (28-9-2011: h11m28s32ms342) e a data em que ocorreram (28-9-2011).

- Os eventos ocorrem num *contexto* particular (intervalo de tempo, condições específicas do contexto de negócio, sazonalidade, etc.). Esse contexto pode explicar certos fenómenos, por exemplo, o tempo

de resposta é superior à média devido ao aumento da carga de trabalho existente ou devido a feriados. Para o analisar, é desejável que o contexto em que as acções foram executadas seja tido em conta. Isso implica a integração de dados dos eventos com os dados do contexto, o que poderá facilmente tornar a análise extremamente complexa caso demasiadas variáveis sejam adicionadas. Este problema é conhecido como "maldição da dimensionalidade" (curse of dimensionality).

São necessárias melhores ferramentas e métodos para resolver os problemas atrás descritos. Além disso, como indicado anteriormente, as organizações precisam de tratar os registos de eventos assegurando a qualidade dos dados ao invés de os utilizarem como sendo um subproduto qualquer. O objectivo é obter registos de eventos ★★★★★ (ver Tabela 1). As lições aprendidas no contexto de data warehousing são úteis para garantir registos de eventos de elevada qualidade. Por exemplo, uma simples verificação quando os dados são registados pode ajudar a reduzir significativamente a proporção de dados incorrectos registados.

C2: Lidar com registos de eventos complexos e com diversas características

Os registos de eventos podem ter características muito diferentes. Alguns conjuntos de registos de eventos podem tornar-se extremamente grandes, enquanto outros são tão pequenos que não há dados suficientes disponíveis para tirar conclusões confiáveis.

Em alguns domínios, são registados enormes quantidades de eventos. Por isso, é necessário um esforço adicional para melhorar o desempenho e a escalabilidade. Por exemplo, o sistema ASML utilizado na indústria dos semi-condutores, está a monitorizar continuamente todos os scanners dos produtos que se encontram em produção. Estes scanners são usados por várias organizações (por exemplo, Samsung e Texas Instruments) para produzir chips (aproximadamente 70% dos chips são produzidos utilizando os scanners da ASML). As ferramentas existentes têm dificuldades em lidar

com os petabytes de dados obtidos neste tipo de domínios. Além do número de eventos registados, há outras características tais como: a média de eventos por caso, a semelhança entre os casos ocorridos, o número de eventos únicos e o número de caminhos únicos. Considere um conjunto de registos de eventos L1 com as seguintes características: 1000 casos, em média 10 eventos por caso e pequena variação (por exemplo, vários casos seguem os mesmos caminhos ou seguem caminhos muito semelhantes). O conjunto de registos de eventos L2 contém apenas 100 casos, mas em média há 100 eventos por caso e todos os casos seguem um caminho único. Claramente, L2 é muito mais difícil de analisar do que L1, embora os dois registos tenham tamanhos semelhantes (cerca de 10.000 eventos).

Como os registos de eventos contêm apenas o comportamento da amostra, não devem ser assumidos como completos. As técnicas de Process Mining precisam lidar com factos incompletos, usando uma suposição: o facto de que algo não aconteceu, não significa que não pode acontecer. Isto representa um desafio para conjuntos de registos de eventos com grande variabilidade.

Como mencionado anteriormente, alguns registos contêm eventos num nível de abstracção muito baixo. Esses conjuntos de registos de eventos tendem a ser extremamente grandes e são de pouco interesse para as partes interessadas. Portanto, gostaríamos de agregar eventos de nível inferior em eventos de alto nível. Por exemplo, quando se analisa os processos de diagnóstico e tratamento de um determinado grupo de pacientes pode não existir interesse nos testes individuais registados no sistema de informação do laboratório do hospital.

Neste momento, as organizações precisam de usar uma abordagem de tentativa e erro para ver se um conjunto de registos de eventos é adequado para o Process Mining. Portanto, as ferramentas devem permitir um teste de avaliação de viabilidade rápido dado um conjunto de dados em particular. Tal teste deve indicar possíveis problemas de desempenho e advertir para os registos de eventos que estão longe de ser completos ou, pelo contrário, serem muito detalhados.

C3: Criar valores de referência para estabelecer pontos de comparação

O Process Mining é uma tecnologia emergente. Isso explica porque não existem ainda valores de referência contra os quais possam ser comparados. Por exemplo, dezenas de técnicas de descoberta de processo estão disponíveis a partir de diferentes fornecedores de sistemas de informação, mas não há consenso sobre a qualidade destas técnicas. Embora haja grandes diferenças em termos de funcionalidade e desempenho, é difícil comparar as diversas técnicas e ferramentas. Por isso, a existência de valores de referência, consistindo em padrões de dados e critérios de avaliação, necessitam de ser desenvolvidos.

Para as técnicas clássicas de Data Mining, existem e estão disponíveis valores de referência em que se pode confiar. Esses valores de referência têm estimulado os fornecedores de sistemas de informação e investigadores a melhorar o desempenho das suas técnicas. No caso do Process Mining este processo constitui outro tipo de desafio. Por exemplo, o modelo relacional introduzido por Codd em 1969 é simples e amplamente adoptado. Como resultado, é necessário pouco esforço para converter dados de uma base de dados para outra e não existem problemas de interpretação. Para a análise de processos, não existe nenhum modelo dessa natureza. As normas de modelação de processos são muito mais complicadas e poucos fornecedores de sistemas suportam exactamente o mesmo conjunto de conceitos. Os processos são mais complexos do que simplesmente um conjunto de dados armazenados numa base de dados.

Mesmo assim, é importante criar valores de referência para o Process Mining. Algum trabalho inicial já está disponível. Por exemplo, existem várias métricas para medir a qualidade dos resultados de Process Mining (robustez, simplicidade, precisão e generalização). Além disso, vários conjuntos de registos de eventos estão publicamente disponíveis (cf. www.processmining.org). Veja por exemplo os registos de eventos usados

para o Business Process Intelligence Challenge (BPIC'11) organizado pelo grupo de trabalho (cf. doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54).

Por um lado, devem existir valores de referência com base em conjuntos de dados reais. Por outro lado, existe a necessidade de criar conjuntos de dados fictícios para capturar determinadas características. Esses conjuntos de dados fictícios ajudam a desenvolver técnicas de Process Mining que são adaptados para registos com eventos incompletos, ruído, ou conjuntos de dados específicos que foram registados para o efeito.

Além da criação de valores de referência que sejam representativos, também é necessário mais consenso sobre os critérios utilizados para avaliar a qualidade dos resultados do Process Mining (ver também o Desafio C6). Além disso, a *validação cruzada* de técnicas de Data Mining pode ser usada para julgar o resultado. Considere por exemplo a verificação k-fold. Pode-se dividir o conjunto de registos de eventos em k partes. k-1 partes podem ser usadas para construir o modelo de processo, enquanto a parte restante pode ser usada para avaliar o modelo resultante através de técnicas de verificação de conformidade. Isto pode ser repetido k vezes, proporcionando assim alguma informação sobre a qualidade do modelo.

C4: Lidar com o desvio do conceito do processo

O *desvio do conceito* refere-se à situação em que o processo muda enquanto está a ser analisado. Por exemplo, no início do conjunto de registos de eventos duas actividades podem ser simultâneas enquanto mais tarde, observa-se que nos registos essas actividades tornam-se sequenciais. Os processos podem mudar devido a mudanças sazonais (por exemplo, "em Dezembro há mais procura" ou "na tarde de sexta-feira há menos funcionários disponíveis") ou devido a alterações no contexto empresarial (por exemplo, "o mercado está cada vez mais competitivo"). Tais mudanças têm impacto na análise efectuada e é vital detectá-las e analisá-las. O desvio no conceito do processo pode ser descoberto dividindo o conjunto de registos de eventos em conjuntos menores

analisando o rasto da divisão efectuada. Tal análise de "segunda ordem" requer muitos mais dados de eventos. Mesmo assim, dificilmente os processos de negócio são estáticos e os desvios de conceito são importantes para a gestão de processos. Portanto, investigação adicional e novas ferramentas são necessárias para lidar com o desvio de conceito.

C5: Melhorar a representação de resultados para a descoberta de processos

Uma técnica utilizada para a descoberta do processo produz um modelo de representação do processo usando uma notação particular (por exemplo, BPMN ou redes de Petri). No entanto, é importante separar a visualização do resultado da representação utilizada durante o processo de descoberta. A selecção de uma notação pré-determinada abrange frequentemente vários pressupostos implícitos. Essa escolha limita o domínio de análise; processos que não podem ser representados por uma determinada notação não podem ser descobertos. Essa "polarização representativa" utilizada durante o processo de descoberta deve ser uma escolha consciente e não deve ser (apenas) impulsionada pela preferência de representação gráfica do processo.

Considere o exemplo da Fig. 6: a capacidade de uma notação representar concorrência pode ter impacto tanto na visualização do modelo descoberto como no tipo de modelos considerados pelo algoritmo. Se a notação escolhida não permitir concorrência (Fig. 6(a) não é possível) nem múltiplas actividades com a mesma designação (Fig. 6(c) não é possível), então apenas os modelos problemáticos, como o mostrado na Fig. 6(b), são possíveis. Este exemplo mostra que a escolha mais cuidadosa da notação de representação é um factor importante.

C6: Equilíbrio entre a qualidade dos critérios como Robustez, Simplicidade, Precisão e Generalização

Os registos de eventos estão muitas vezes longe de estarem completos, ou seja, apenas um exemplo de comportamento é fornecido. Os modelos de processo geralmente permitem um número exponencial ou até mesmo infinito de variantes e recursividade (designados comumente como ciclos). Além disso, algumas instâncias de processo podem ter uma probabilidade muito menor do que outras. Portanto, não é realista supor que estão disponíveis todas as evidências nos registos de eventos. Para ilustrar que é impraticável assumir como adquirido que os registos de eventos estão completos e contêm todas as informações para proceder à análise, considere um processo constituído por 10 actividades que podem ser executadas em paralelo e um conjunto de registos de eventos correspondente, que contém informações sobre 10.000 casos. O número total de intercalações possíveis no modelo com 10 actividades simultâneas é $10! = 3.628.800$. Por isso, é impossível que cada intercalação esteja presente nos registos, pois há menos casos (10.000) do que variantes potenciais (3.628.800). Mesmo se houver milhões de casos nos registos de eventos, é extremamente improvável que todas as variações possíveis estejam presentes. Um problema adicional é que algumas variantes são menos frequentes do que outras. Estas podem ser consideradas como "ruído". É impossível construir um modelo robusto com tal interferência ruidosa. O modelo descoberto precisa abstrair isso; é melhor investigar o comportamento de variantes menos executadas, contrastando com a verificação de conformidade.

Ruído e dados incompletos transformam a descoberta do modelo do processo num desafio. Na realidade, há quatro dimensões da qualidade que concorrem entre si: (a) Robustez, (b) Simplicidade, (c) Precisão e (d) Generalização. Um modelo com boa *robustez* permite observar a maior parte dos eventos nos registos. Um modelo tem uma *robustez* perfeita se todas as variantes descritas nos

registos de eventos podem ser reproduzidas desde o início do modelo do processo até ao fim. O modelo mais *simples* que pode explicar o comportamento descrito nos registos de eventos é o melhor modelo. Este princípio é conhecido como Occam's Razor. Robustez e simplicidade por si só não são suficientes para julgar a qualidade de um modelo de processo que acaba de ser descoberto. Por exemplo, é muito fácil construir uma rede de Petri extremamente simples ("flower model") que é capaz de reproduzir todas as variantes de um conjunto de registos de eventos (mas também outro qualquer conjunto que se refere ao mesmo conjunto de actividades). Da mesma forma, não é desejável ter um modelo descoberto que apenas permite observar um determinado comportamento descrito nos registos de eventos. Lembre-se que os registos de eventos contêm apenas um determinado comportamento e que muitas outras possibilidades de execução podem não ter sido ainda registadas. Um modelo é *preciso* se ele não permite ter demasiadas variantes. Claramente, o "flower model" carece de precisão. Um modelo que não é preciso não tem robustez suficiente porque oculta informação sobre a forma como o processo é executado. A falta de robustez é um problema que tende a generalizar em demasia o comportamento do processo comparativamente ao que se encontra descrito nos registos de eventos (i.e., distorce a realidade que se encontra registada). Um modelo deve generalizar e não restringir o comportamento apenas com os dados registados. Um modelo que não é capaz de *generalizar* cria distorção na análise. A distorção é criada porque o modelo de processo gerado apenas mostra um determinado comportamento descrito nos registos de eventos (i.e., o modelo do processo está de acordo com os dados registados, mas uma outra amostra de dados do processo pode criar um modelo de processo completamente diferente).

Obter um equilíbrio entre Robustez, Simplicidade, Precisão e Generalização é o grande desafio. Esta é a razão por que a maioria das técnicas mais poderosas de descoberta de modelos de processo fornece vários

parâmetros. É preciso desenvolver melhores algoritmos para equilibrar melhor as quatro dimensões de qualidade que concorrem entre si. Além disso, quaisquer parâmetros utilizados devem ser compreensíveis pelos utilizadores finais.

C7: Process Mining inter-organizacional

Tradicionalmente, o Process Mining é aplicado dentro do universo de uma única organização. No entanto, à medida que os Web Services, a integração da cadeia logística e o cloud computing tornam-se mais difundidos, existem cenários onde os registos de eventos de várias organizações estão disponíveis para análise. Em princípio, existem duas definições para *Process Mining inter-organizacional*.

Em primeiro lugar, deveremos considerar que hoje em dia as empresas colaboram entre si, fazendo parceria na execução de instâncias do processo. Pode-se assumir que um processo inter-organizacional é como um "quebra-cabeças", ou seja, o processo é cortado em peças sendo distribuídas por organizações que têm necessidade de cooperar para executar determinadas actividades. Analisar os registos de eventos relativos apenas a uma das organizações envolvidas é insuficiente. Para se descobrir o processo numa perspectiva "end-to-end", os conjuntos de registos de eventos de diferentes organizações precisam ser integrados. Esta é uma tarefa não trivial porque os eventos precisam ser correlacionados através das fronteiras organizacionais.

Em segundo lugar, também podemos considerar o cenário onde as diferentes organizações executam o mesmo processo, partilhando experiências, conhecimento, ou uma infra-estrutura comum. Considere, por exemplo o sistema de informação da Salesforce.com. Os processos de vendas e de gestão das oportunidades de negócio de muitas organizações são geridos e suportados pela Salesforce. Por um lado, essas organizações compartilham uma infra-estrutura comum (processos, bases de dados, etc.). Por outro lado, as empresas não são obrigadas a seguir um modelo de processo rígido porque o sistema pode

ser configurado para apoiar as diversas variantes de execução do mesmo processo. Como outro exemplo, considere a processos básicos executados em qualquer município (por exemplo, a emissão de licenças de construção). Apesar de todos os municípios num determinado país terem necessidade de apoiar o mesmo conjunto básico de processos, pode haver diferenças. Obviamente, é interessante analisar tais variações entre as diferentes organizações. Estas organizações podem aprender umas com as outras e os fornecedores de sistemas de informação podem melhorar os seus serviços com base nos resultados de Process Mining inter-organizacional.

Novas técnicas de análise precisam ser desenvolvidas para ambos os tipos de Process Mining inter-organizacional. Estas técnicas devem também considerar a privacidade e segurança. As organizações podem não desejar partilhar informações por razões de confidencialidade ou falta de confiança. Portanto, é importante desenvolver técnicas de Process Mining capazes de preservar privacidade.

C8: Fornecer suporte operacional

Inicialmente, o foco do Process Mining debruçava-se sobre a análise de dados históricos. Hoje, muitas fontes de dados são actualizadas em (quase) tempo real, que em conjunto com o poder computacional disponível tornam possível a análise de eventos ainda em execução. Portanto, o Process Mining não deve ser limitado a uma perspectiva restrita de análise, podendo também ser usado para apoio operacional em tempo real. Podem ser identificadas três actividades de apoio operacional: *detectar*, *prever* e *recomendar*. No momento em que uma instância se desvia do modelo de referência, isso pode ser detectado e o sistema pode gerar um alerta. Muitas vezes gostaríamos que essas notificações fossem geradas imediatamente (para ainda sermos capazes de influenciar os resultados) e não de uma forma indirecta. Os dados históricos podem ser usados para construir modelos preditivos. Estes podem ser utilizados para orientar o

processo na execução de determinada instância. Por exemplo, é possível prever o tempo de processamento restante. Baseado em tais previsões, pode-se construir sistemas de recomendação que propõe acções específicas para reduzir os custos ou diminuir o tempo. Aplicando técnicas de Process Mining em tempo real, é possível criar desafios adicionais em termos de poder computacional e qualidade de dados.

C9: Utilizar Process Mining conjuntamente com outro tipo de análises

A gestão de operações, e particularmente investigação operacional, são ramos da ciência de gestão bastante dependentes da modelação. Para esse efeito, existe um conjunto variado de modelos matemáticos que são usados e que vão desde a programação linear e planeamento de projectos até filas de espera, cadeias de Markov, e simulação. O Data Mining pode ser definido como "a análise de conjuntos de dados (tipicamente grandes conjuntos) para encontrar relacionamentos que à partida assumimos que não existem, e para resumir e apresentar os dados de forma mais compreensível e útil". Uma grande variedade de técnicas têm vindo a ser desenvolvidas: classificação (por exemplo, a aprendizagem de árvores de decisão), regressão, clustering (por exemplo, k-means) e descoberta de padrões (por exemplo, a aprendizagem de regras de associação).

Ambos os domínios (gestão de operações e Data Mining) fornecem técnicas de análise de valor. O desafio é combinar as técnicas destes domínios com o Process Mining. Considere, por exemplo a simulação. As técnicas de Process Mining podem ser usadas para entender um modelo de simulação com base em dados históricos. Posteriormente, o modelo de simulação pode ser usado para fornecer suporte operacional. Por causa da estreita ligação entre os registos de eventos e o modelo de processo, o modelo pode ser usado para reproduzir as instâncias de processo que ocorreram e pode-se começar a executar simulações do estado actual proporcionando assim uma forma rápida de prever o futuro

com base em dados reais.

Da mesma forma, é desejável combinar Process Mining com métodos de *análise visual*. A análise visual combina a análise automatizada com visualizações interactivas para uma melhor compreensão de grandes e complexos conjuntos de dados. A análise visual explora as incríveis capacidades dos seres humanos de ver padrões em dados não estruturados. Ao combinar técnicas automatizadas de Process Mining com capacidades de análise visuais e interactivas, é possível extrair informação importante a partir dos dados dos eventos descritos nos registos.

C10: Melhorar a usabilidade para os que não são especialistas

Um dos objectivos do Process Mining é criar "modelos de processos vivos", ou seja, modelos que são usados numa base diária, em vez dos modelos estáticos que estão estruturados num qualquer documento. Novos dados de eventos podem ser usados para descobrir determinado comportamento emergente. A ligação entre os dados de eventos e os modelos de processo permite prever o estado actual do processo e as actividades que estão a ser executadas em tempo real. Assim os utilizadores finais podem interagir com os resultados do Process Mining

diariamente. Tais interacções são muito valiosas, mas também exigem interfaces intuitivos com os utilizadores finais. O desafio é esconder os sofisticados algoritmos de Process Mining por detrás de interfaces amigáveis que, automaticamente, definem os parâmetros e sugerem métodos adequados de análise.

C11: Melhorar a compreensão para os que não são especialistas

Mesmo que seja fácil de gerar resultados usando Process Mining, isso não significa que os resultados sejam realmente úteis. O utilizador pode ter problemas em entender os resultados ou ser tentado a inferir conclusões incorrectas. Para evitar tais problemas, os resultados devem ser apresentados através de uma representação adequada (ver também GP5). Além disso, o nível de confiança dos resultados deve ser sempre claramente indicado. Podem existir poucos dados para justificar conclusões particulares. Na verdade, as técnicas de descoberta de processos existentes, não alertam sobre o estado de robustez do processo. Um modelo é sempre apresentado, mesmo quando é evidente que não há dados suficientes para justificar qualquer conclusão.

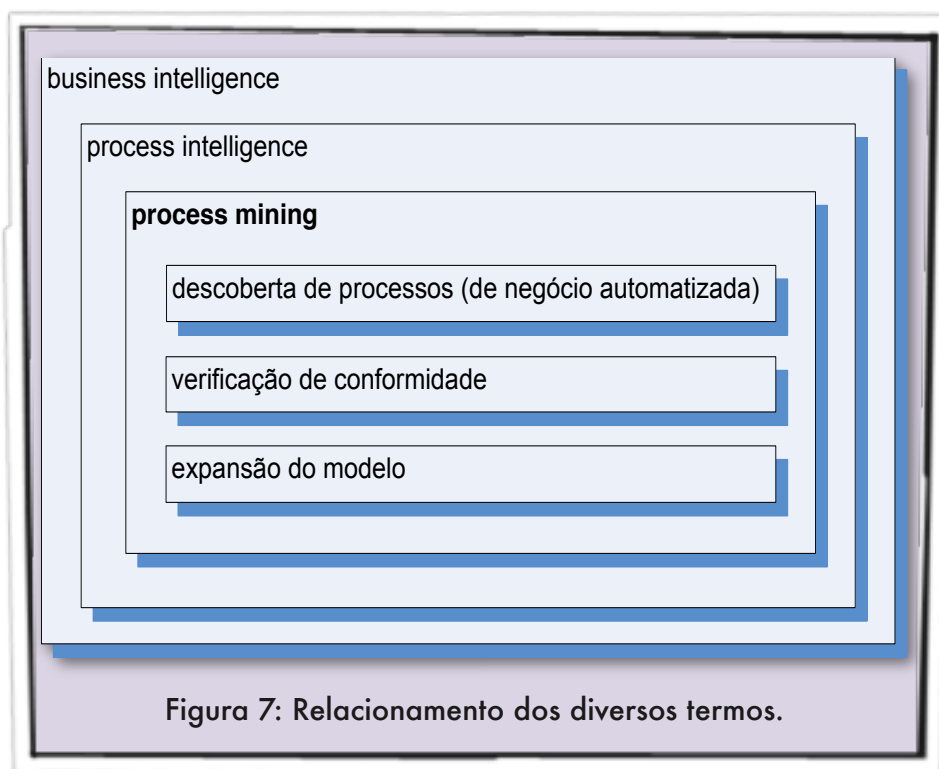


Figura 7: Relacionamento dos diversos termos.

Epílogo

O grupo de trabalho IEEE em Process Mining tem como objectivo (a) promover a aplicação do Process Mining, (b) orientar as pessoas responsáveis pelo desenvolvimento de software, consultores, gestores e utilizadores finais para a utilização das melhores técnicas de análise, e (c) estimular a pesquisa sobre o Process Mining. Este manifesto afirma os princípios e intenções do grupo de trabalho. Depois de introduzir o tema do Process Mining, o manifesto cataloga algumas normas de orientação (Secção 3) e desafios (Secção 4). As normas de orientação podem ser usadas para evitar erros óbvios. A lista de desafios destina-se a direccionar esforços entre investigação e desenvolvimento. Ambos visam aumentar o nível de maturidade do Process Mining.

Para concluir, algumas palavras sobre a terminologia. Os seguintes termos são usados no domínio de Process Mining: Workflow Mining, Process Mining (de processos de negócio), Automated Business Process Discovery e Process Intelligence. Diferentes organizações parecem estar a usar diferentes termos para conceitos que se sobrepõem. Por exemplo, o Gartner está a promover o termo "Automated Business Process Discovery" (ABPD) e a Software AG está a usar "Process Intelligence" para se referir ao seu sistema de informação. O termo Workflow Mining parece menos adequado tal como a criação de modelos de workflow (fluxo de processo) é apenas uma das aplicações possíveis do Process Mining. Da mesma forma que a utilização do termo "negócio" diminui o âmbito de certas aplicações de Process Mining. Existem inúmeras aplicações de Process Mining (por exemplo, análise de sistemas de alta tecnologia ou análise de sítios na Internet) onde este tipo de inclusão parece ser inapropriado. Apesar da descoberta de processos ser uma parte importante do espectro do Process Mining, isto é apenas uma das possibilidades de utilização. A verificação de conformidade, previsão, análise organizacional, análise de redes sociais, e outras, vão muito mais além da descoberta de processos.

A Fig. 7 relaciona alguns dos termos mencionados. Todas as tecnologias e métodos que procuram

fornecer informação adicional que pode ser usada para tomar decisões podem ser posicionadas sob a égide do Business Intelligence (BI). O Process Intelligence (de negócio) pode ser visto como a combinação de BI e BPM, ou seja, técnicas que são usadas para analisar e melhorar os processos de negócio. O Process Mining pode ser visto como uma concretização de inteligência de processo a partir dos registos de eventos. A descoberta automática de modelos de processo é apenas um dos três tipos básicos de Process Mining. A Fig. 7 pode ser um pouco enganadora dado que a maioria das ferramentas de BI não suporta funcionalidades de Process Mining, conforme descrito neste documento. O acrónimo BI é frequentemente desviado de forma conveniente para uma determinada ferramenta ou método que cobre apenas uma pequena parte do amplo espectro de BI.

Podem existir razões comerciais para se usar termos alternativos. Alguns fornecedores de sistemas podem também estar interessados em enfatizar aspectos particulares (por exemplo, descoberta ou inteligência). Contudo, para evitar confusão, é melhor usar o termo "Process Mining" para os aspectos analisados neste manifesto.

Glossário

Actividade: um passo bem definido num processo. Os eventos podem referir-se ao início, conclusão, cancelamento, etc., de uma actividade para uma determinada instância de um processo.

Business Intelligence (BI): conjunto alargado de ferramentas e métodos que usam dados para suportar a tomada de decisão.

Business Process Intelligence: ver Process Intelligence.

Business Process Management (BPM): disciplina que combina conhecimento de tecnologias de informação e conhecimento de ciências de gestão, e aplica ambos a processos de negócios operacionais.

Caso: ver Instância de Processo.

Data Mining: análise de conjuntos de dados (tipicamente grandes conjuntos) para encontrar relacionamentos inesperados e sintetizar os dados de forma a proporcionar novo conhecimento.

Este manifesto foi publicado originalmente em "Business Process Management Workshops 2011", Lecture Notes em Business Information Processing, Vol. 99, Springer-Verlag, 2011, e tem sido traduzido em vários idiomas. Mais informação acerca do grupo de trabalho IEEE em Process Mining pode ser encontrada em: <http://www.win.tue.nl/ieeetfpm/>

Descoberta automatizada de um processo de negócio: ver Descoberta de Processos.

Desvio de conceito do processo: fenómeno que faz alterar a forma como o processo é executado ao longo do tempo. As instâncias do processo poderão mudar gradualmente (ou repentinamente) devido a factores sazonais ou aumento de competição, tornando assim a análise mais complicada.

Descoberta de Processos: um dos três tipos básicos de Process Mining. Baseado num conjunto de registos de eventos, a estrutura de um modelo de processo é compreendida. Por exemplo, o algoritmo α é capaz de descobrir uma rede de Petri por identificação de padrões de processo em conjuntos de eventos.

Evento: uma acção registada nos registos de eventos, por exemplo, o início, a conclusão, ou o cancelamento de uma actividade de uma instância particular de um processo.

Expansão do modelo: um dos três tipos básicos de Process Mining. O modelo de Processo é expandido ou melhorado usando a informação extraída dos registos de eventos. Por exemplo, os pontos de estrangulamento podem ser identificados usando referências temporais, através da repetição de um conjunto de registos de eventos num dado modelo de processo.

Generalização: uma medida que determina de que forma o modelo permite analisar comportamento que nunca foi identificado. Um modelo "extremamente robusto" pode não ser capaz de generalizar adequadamente.

Instância de Processo: entidade que está a ser processada pelo processo em análise. Eventos referem-se às instâncias de processo. Exemplos de

instâncias de processo são ordens de compra, pedidos de empréstimo, etc.

MXML: formato XML para trocar registos de eventos. O formato XES substitui o MXML como o novo formato de Process Mining independentemente do sistema de informação usado.

Precisão: medida para determinar se o modelo impede um comportamento muito diferente do comportamento observado nos registos de eventos. Um modelo com baixa precisão é pouco robusto.

Process Intelligence: um ramo do Business Intelligence com foco em Business Process Management.

Process Mining: técnicas, ferramentas e métodos para detectar, monitorizar e melhorar os processos que são executados (i.e., não são processos que se assume que existem e estão documentados por exemplo em diagramas de arquitecturas de processo) extraindo conhecimento a partir de registos de eventos comumente disponíveis nos sistemas de informação.

Process Mining inter-organizacional: utilização de técnicas de Process Mining sobre registos de eventos de organizações diferentes.

Registos de eventos: conjunto de registos de eventos utilizados como input para o Process Mining. Os eventos não necessitam de estar registados num conjunto de dados único (por exemplo, os eventos podem

estar dispersos em diversas bases de dados).

Representational Bias: a linguagem seleccionada para apresentar e construir resultados de Process Mining.

Robustez: uma medida que determina de que forma um determinado modelo de processo permite que o seu comportamento seja compreendido a partir dos registos de eventos. O modelo tem total robustez se todas as instâncias de processo nos registos de eventos poderem ser reproduzidas desde o início até ao fim.

Simplicidade: uma medida para implementação do princípio de Occam's Razor, ou seja, o modelo mais simples que pode explicar o comportamento visto nos registos de eventos, é o melhor modelo. A simplicidade pode ser quantificada de várias formas, por exemplo, número de nós e arcos no modelo.

Suporte Operacional: análise em tempo real de dados de eventos com o objectivo de monitorizar e influenciar instâncias de processo. Três actividades de apoio operacional podem ser identificadas: *detectar* (gerar um alerta, se o comportamento observado se desvia do comportamento modelado), *prever* (prever um comportamento futuro com base no comportamento passado, por exemplo, prever o tempo de processamento restante), e *recomendar* (sugerir acções apropriadas para realizar um

objectivo particular, por exemplo, para minimizar os custos).

Verificação de Conformidade: analisar se uma actividade nos registos de eventos está de acordo com o modelo e vice-versa. O objectivo é detectar discrepâncias e medir a sua severidade. A verificação de conformidade é um dos três tipos básicos de Process Mining.

XES: norma baseada em XML para conjuntos de registos de eventos. A norma foi aprovada pelo grupo de trabalho IEEE em Process Mining como o formato de intercâmbio padrão para conjuntos de registos de eventos (cf. www.xes-standard.org).

Autores

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