Manifiesto sobre Minería de Procesos

(versión final)

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, and Moe Wynn

IEEE Task Force on Process Mining[*]
http://www.win.tue.nl/ieeetfpm

Resumen Las técnicas de minería de procesos son capaces de *extraer conocimiento de los registros de eventos* comúnmente disponibles en los sistemas de información actuales. Estas técnicas proveen nuevos medios para *descubrir, monitorear y mejorar los procesos* en una variedad de dominios de aplicación. Hay dos razones principales para el creciente interés en minería de procesos. Por un lado, se registran más y más eventos, proporcionando información detallada acerca de la historia de los procesos. Por otro lado, hay una necesidad de mejorar y apoyar los procesos de negocio en ambientes competitivos y que cambian rápidamente. Este manifiesto es creado por la *IEEE Task Force on Process Mining* (Fuerza de Trabajo de la IEEE sobre Minería de Procesos) y está dirigido a promover el tópico de minería de procesos. Además, al definir un conjunto de principios rectores y listar importantes desafíos, este manifiesto espera servir como una *guía para desarrolladores de software*, *científicos*, *consultores*, *gerentes de negocio*, y *usuarios finales*. El objetivo es incrementar la madurez de la minería de procesos como una nueva herramienta para mejorar el (re)diseño, control, y apoyo a los procesos de negocio operacionales.

1. IEEE Task Force on Process Mining

Un *manifiesto* es una "declaración pública de principios e intenciones" por un grupo de personas. Este manifiesto es escrito por los miembros y personas que respaldan la *IEEE Task Force on Process Mining* (Fuerza de Trabajo de la IEEE sobre Minería de Procesos). El objetivo de esta fuerza de trabajo es promover la investigación, desarrollo, educación, implementación, evolución, y entendimiento acerca de la minería de procesos.

La minería de procesos es una disciplina de investigación relativamente jóven que se ubica entre la inteligencia computacional y la minería de datos, por una parte, y la modelación y análisis de

[*] La versión original apareció en los *BPM 2011 Workshops proceedings*, Lecture Notes in Business Information Processing, Springer-Verlag, 2011.

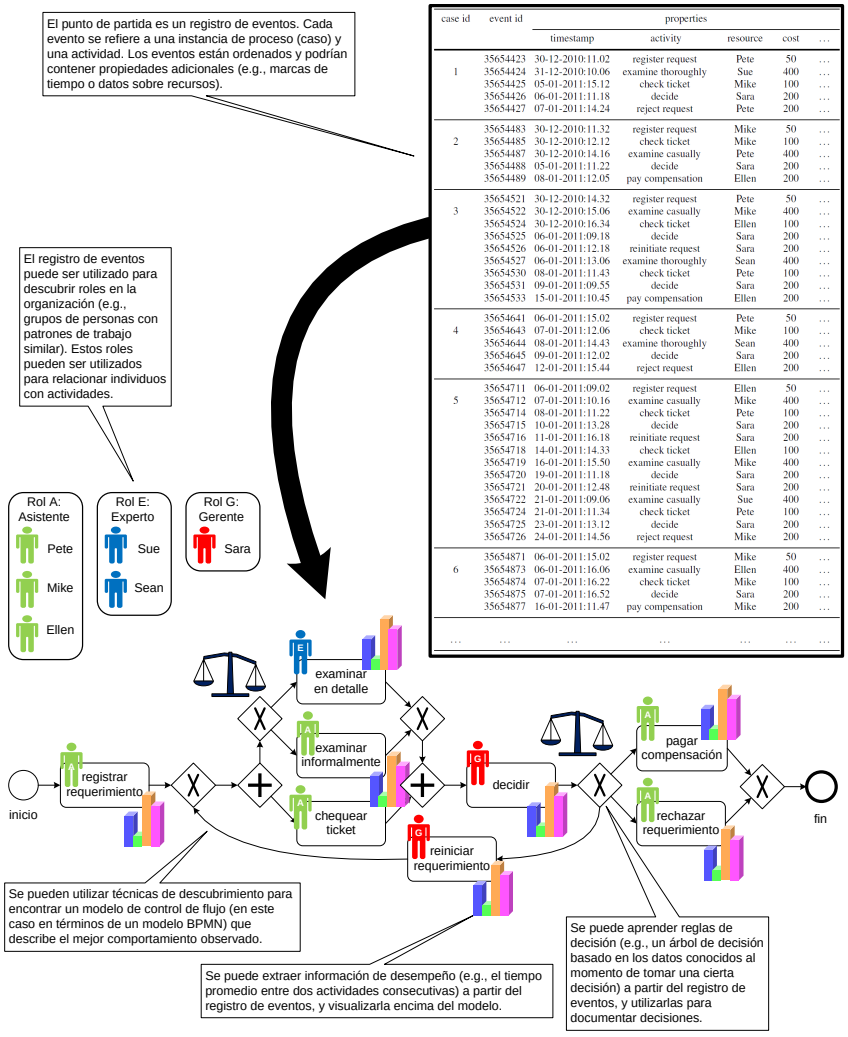
procesos, por otra. La idea de la minería de procesos es *descubrir, monitorear y mejorar los procesos reales* (i.e., no los procesos supuestos) *a través de la extracción de conocimiento de los registros de eventos* ampliamente disponibles en los actuales sistemas (de información) (ver Fig. 1). La minería de procesos incluye el descubrimiento (automático) de procesos (i.e., extraer modelos de procesos a partir de un registro de eventos), la verificación de conformidad (i.e., monitorear desviaciones al comparar el modelo y el registro de eventos), la minería de redes sociales/organizacionales, la construcción automática de modelos de simulación, la extensión de modelos, la reparación de modelos, la predicción de casos, y las recomendaciones basadas en historia.

La minería de procesos proveee un puente importante entre la minería de datos y la modelación y análisis de procesos de negocio. Bajo el paragua de la *Inteligencia de Negocios* (*Business Intelligence*, BI), se han introducido muchas palabras de moda para referirse a herramientas más bien simples para hacer reportería y paneles de control. El *Monitoreo de Actividades de Negocio* (*Business Activity Monitoring*, BAM) se refiere a las tecnologías que facilitan el monitoreo en tiempo real de los procesos de negocio. El *Procesamiento de Eventos Complejos* (*Complex Event Processing*, CEP) se refiere a las tecnologías que permiten procesar grandes cantidades de eventos, utilizándolos para monitorear, dirigir y optimizar el negocio en tiempo real. La *Gestión del Desempeño Corporativo* (*Corporate Performance Management*, CPM) es otra palabra de moda para medir el desempeño de un proceso u organización. También se relaciona con enfoques de gestión, tales como el *Mejoramiento Continuo de Procesos* (*Continuous Process Improvement*, CPI), el *Mejoramiento de Procesos de Negocio* (*Business Process Improvement*, BPI), la *Gestión de Calidad Total* (*Total Quality Management*, TQM), y *Six Sigma*. Estos enfoques tienen en común que los procesos son "puestos bajo el microscopio" para ver si son posibles mejoras adicionales. La minería de procesos es una tecnología que facilita CPM, BPI, TQM, Six Sigma, y similares.

Mientras BI y los enfoques de gestión tales como Six Sigma y TQM buscan mejorar el desempeño operacional, e.g., reducir el tiempo de flujo y los defectos, las organizaciones también están poniendo más énfasis en el *gobierno corporativo*, los *riesgos*, y el *cumplimiento de normativas*. Legislaciones como la ley Sarbanes-Oxley (SOX) y el Acuerdo de Basilea II ilustran el foco en tópicos de cumplimiento de normativas. Las técnicas de minería de procesos ofrecen un medio para chequear de manera más rigurosa el cumplimiento de normativas y establecer la validez y confiabilidad de la información acerca de los procesos críticos de una organización.

Durante la última década, los datos sobre los eventos han comenzando a estar disponibles y las técnicas de minería de procesos han madurado. Además, como ya se mencionó, las tendencias de gestión relacionadas al mejoramiento de procesos (e.g., Six Sigma, TQM, CPI, y CPM) y cumplimiento de normativas (SOX, BAM, etc.) se pueden beneficiar de la minería de procesos. Afortunadamente, los algoritmos de minería de procesos han sido implementados en diversos sistemas académicos y comerciales. Hoy en día, hay un grupo activo de investigadores trabajando en minería de procesos y ha llegado a ser uno de los "tópicos de moda" en la investigación en Gestión de Procesos de Negocio (*Business Process Management*, BPM). Además, hay un enorme interés de la industria por la minería de procesos. Más y más proveedores de software están agregando funcionalidades de minería de procesos en sus herramientas. Ejemplos de productos de software con capacidades de minería de procesos son: 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), y Reflect|one (Pallas Athena). El creciente interés en el análisis de procesos basado en registros de eventos motivó la fundación de una Fuerza de Trabajo en Minería de Procesos.

La fuerza de trabajo se estableció en 2009 en el contexto del Comité Técnico de Minería de Datos (*Data Mining Technical Committee*, DMTC) de la Sociedad de Inteligencia Computacional (*Computational Intelligence Society*, CIS) del Instituto de Ingenieros Eléctricos y Electrónicos (*Institute of Electrical and Electronic Engineers*, IEEE). La fuerza de trabajo actual tiene miembros que representan a *proveedores de software* (e.g., Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Stereologic), *empresas consultoras/usuarios finales* (e.g.,



case id	event id	properties				
		timestamp	activity	resource	cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	register request	Pete	50	...
	35654522	30-12-2010:15.06	examine casually	Mike	400	...
	35654524	30-12-2010:16.34	check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	decide	Sara	200	...
	35654526	06-01-2011:12.18	reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	check ticket	Pete	100	...
	35654531	09-01-2011:09.55	decide	Sara	200	...
	35654533	15-01-2011:10.45	pay compensation	Ellen	200	...
4	35654641	06-01-2011:15.02	register request	Pete	50	...
	35654643	07-01-2011:12.06	check ticket	Mike	100	...
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400	...
	35654645	09-01-2011:12.02	decide	Sara	200	...
	35654647	12-01-2011:15.44	reject request	Ellen	200	...
5	35654711	06-01-2011:09.02	register request	Ellen	50	...
	35654712	07-01-2011:10.16	examine casually	Mike	400	...
	35654714	08-01-2011:11.22	check ticket	Pete	100	...
	35654715	10-01-2011:13.28	decide	Sara	200	...
	35654716	11-01-2011:16.18	reinitiate request	Sara	200	...
	35654718	14-01-2011:14.33	check ticket	Ellen	100	...
	35654719	16-01-2011:15.50	examine casually	Mike	400	...
	35654720	19-01-2011:11.18	decide	Sara	200	...
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	...
	35654722	21-01-2011:09.06	examine casually	Sue	400	...
	35654724	21-01-2011:11.34	check ticket	Pete	100	...
	35654725	23-01-2011:13.12	decide	Sara	200	...
	35654726	24-01-2011:14.56	reject request	Mike	200	...
6	35654871	06-01-2011:15.02	register request	Mike	50	...
	35654873	06-01-2011:16.06	examine casually	Ellen	400	...
	35654874	07-01-2011:16.22	check ticket	Mike	100	...
	35654875	07-01-2011:16.52	decide	Sara	200	...
	35654877	16-01-2011:11.47	pay compensation	Mike	200	...
...	...	...	...	...	...	...

Figura 1. Las técnicas de minería de procesos extraen conocimiento de los registros de eventos con el fin de descubrir, monitorear, y mejorar los procesos

ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chile, BWI Systeme GmbH, Excellentia BPM, Rabobank), e *institutos de investigación* (e.g., TU/e, University of Padua, Universitat Politècnica 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, 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, Pontificia Universidad Católica de Chile).

Los objetivos concretos de la fuerza de trabajo son:

- generar conciencia en los usuarios finales, desarrolladores, consultores, gerentes de negocio, e investigadores acerca del estado del arte en minería de procesos,
- promover el uso de técnicas y herramientas de minería de procesos y estimular nuevas aplicaciones,
- tener un rol en los esfuerzos de estandarización para el registro de datos de eventos,
- organizar tutoriales, sesiones especiales, talleres (workshops), paneles, y
- publicar artículos, libros, videos, y ediciones especiales de revistas científicas.

Desde su fundación en 2009, ha habido varias actividades relacionadas a los objetivos anteriores. Por ejemplo, varios workshops y sesiones especiales fueron (co-)organizados por la fuerza de trabajo, e.g., los workshops sobre Inteligencia de Procesos de Negocio (BPI'09, BPI'10, y BPI'11) y las sesiones especiales en las principales conferencias de la IEEE (e.g. CIDM'11). El conocimiento fue diseminado vía tutoriales (e.g. WCCI'10 y PMPM'09), escuelas de verano (ESSCaSS'09, ACPN'10, CICH'10, etc.), videos (cf. www.processmining.org), y varias publicaciones, incluyendo el primer libro sobre minería de procesos recientemente publicado por Springer.[1] La fuerza de trabajo también (co-)organizó el primer Desafío de Inteligencia de Procesos de Negocio (*Business Process Intelligence Challenge*, BPIC'11): una competencia donde los participantes tuvieron que extraer conocimiento relevante de un registro de eventos grande y complejo. En 2010, la fuerza de trabajo también estandarizó *XES* (www.xes-standard.org), un formato de registro estándar que es extensible y está respaldado por la *OpenXES library* (www.openxes.org) y por herramientas tales como ProM, XESame, Nitro, etc.

Se invita al lector a visitar http://www.win.tue.nl/ieeetfpm para más información acerca de las actividades de la fuerza de trabajo.

2. Minería de Procesos: Estado del Arte

Las capacidades en expansión de los sistemas de información y otros sistemas que dependen de la computación, están bien caracterizadas por la ley de Moore. Gordon Moore, el co-fundador de Intel, vaticinó en 1965 que el número de componentes en los circuitos integrados se duplicaría todos los años. Durante los últimos 50 años el crecimiento ha sido de hecho exponencial, si bien es cierto que a un paso ligeramente más lento. Estos avances resultaron en un crecimiento espectacular del "universo digital" (i.e., todos los datos almacenados y/o intercambiados electrónicamente). Además, el universo digital y el real continúan acercándose a estar más y más alineados.

El crecimiento de un universo digital que está bien alineado con los procesos en las organzaciones hace posible registrar y analizar *eventos*. Los eventos podrían variar desde el retiro de dinero en efectivo desde un ATM, un doctor ajustando una máquina de rayos-X, un ciudadano solicitando una licencia de conducir, el envío de una declaración de impuestos, y la recepción de un número de boleto electrónico por un viajero. El desafío es aprovechar los datos de eventos en una forma significativa, por ejemplo, para proveer un mejor entendimiento, identificar cuellos de botella, anticipar problemas, registrar violaciones de políticas, recomendar contramedidas, y simplificar procesos. La minería de procesos apunta a hacer exactamente eso.

El punto de partida de la minería de procesos es un *registro de eventos*. Todas las técnicas de minería de procesos asumen que es posible registrar eventos *secuencialmente* tal que cada evento se refiera a una *actividad* (i.e., un paso bien definido en algún proceso) y se relacione a un *caso* particular (i.e., una instancia de proceso). Los registros de eventos podrían almacenar información adicional acerca de los eventos. De hecho, siempre que sea posible, las técnicas de minería de procesos usan información extra, tales como el *recurso* (i.e., persona o dispositivo) que ejecuta o inicia la actividad, la marca de tiempo del evento, o *elementos de datos* registrados con el evento (e.g., el tamaño de un pedido).

Como se muestra en la Fig. 2, los registros de eventos pueden ser utilizados para realizar tres tipos de minería de procesos. El primer tipo de minería de procesos es el *descubrimiento*. Una técnica de descubrimiento toma un registro de eventos y produce un modelo sin usar ninguna información a-priori. El descubrimiento de procesos es la técnica de minería de procesos más destacada. Para muchas organizaciones es sorprendente ver que las técnicas existentes son realmente capaces de descubrir los procesos reales meramente basado en las muestras de ejecución en los registros de eventos. El segundo tipo de minería de procesos es la *conformidad*. Aquí, se compara un modelo de proceso existente con un registro de eventos del mismo proceso. La verificación de conformidad puede ser usada para chequear si la realidad, tal como está almacenada en el registro de eventos, es equivalente al modelo y viceversa. Note que distintos tipos de modelos pueden ser

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

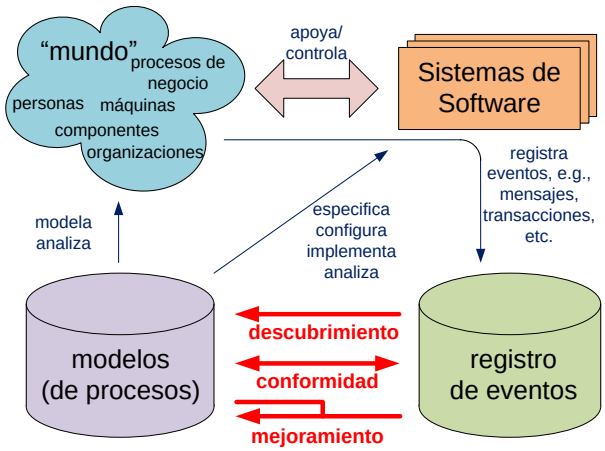


Figura 2. Posicionamiento de los tres tipos principales de minería de procesos: (a) *descubrimiento*, (b) verificación de *conformidad*, y (c) *mejoramiento*.

considerados: la verificación de conformidad puede ser aplicada a modelos procedurales, modelos organizacionales, modelos de procesos declarativos, políticas/reglas de negocio, regulaciones, etc. El tercer tipo de minería de procesos es el *mejoramiento*. Aquí, la idea es extender o mejorar un modelo de proceso existente usando la información acerca del proceso real almacenada en algún registro de eventos. Mientras la verificación de conformidad mide el alineamiento entre el modelo y la realidad, este tercer tipo de minería de procesos busca cambiar o extender el modelo a-priori. Por ejemplo, al usuar marcas de tiempo en el registro de eventos, uno puede extender el modelo para mostrar cuellos de botella, niveles de servicio, tiempos de procesamiento, y frecuencias.

La Fig. 3 describe los tres tipos de minería de procesos en términos de entradas y salidas. Las técnicas para descubrimiento toman un registro de eventos y producen un modelo. El modelo descubierto es típicamente un modelo de proceso (e.g., una red de Petri, un BPMN, un EPC, o un diagrama de actividad UML), sin embargo, el modelo podría también describir otras perspectivas (e.g., una red social). Las técnicas de verificación de conformidad necesitan un registro de eventos y un modelo como entrada. La salida consiste en información de diagnóstico mostrando las diferencias y elementos en común entre el modelo y el registro de eventos. Las técnicas para mejoramiento de modelos (reparar o extender) también necesitan un registro de eventos y un modelo como entrada. La salida es un modelo mejorado o extendido.

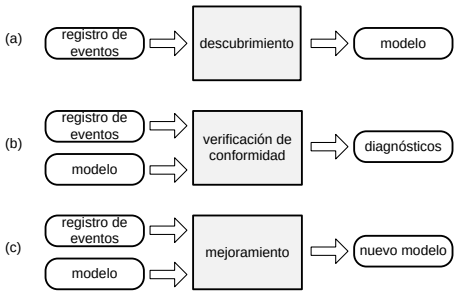


Figura 3. Los tres tipos básicos de minería de procesos explicados en términos de entradas y salidas: (a) descubrimiento, (b) verificación de conformidad, y (c) mejoramiento.

La minería de procesos podría cubrir diferentes perspectivas. La *perspectiva de control de flujo* se enfoca en el control de flujo, i.e., el orden de ejecución de las actividades. El objetivo de

explorar esta perspectiva es encontrar una buena caracterización de todos los caminos posibles. El resultado se expresa típicamente en términos de una red de Petri o alguna otra notación de procesos (e.g., EPCs, BPMN, o diagramas de actividad UML). actores (e.g., personas, sistemas, o departamentos) están involucrados y cómo se relacionan. El objetivo es ya sea estructurar la organización clasificando a las personas en términos de roles y unidades organizacionales, o mostrar la red social. La *perspectiva de casos* se enfoca en las propiedades de los casos. Obviamente, un caso puede ser caracterizado por su ruta en el proceso o por los actores que trabajan en él. Sin embargo, los casos también pueden ser caracterizados por los valores de los correspondientes elementos de datos. Por ejemplo, si un caso representa un pedido de reposición, podría ser interesante conocer el proveedor o la cantidad de productos solicitados. La *perspectiva de tiempo* se relaciona con la ocurrencia y frecuencia de los eventos. Cuando los eventos tienen asociados marcas de tiempo, es posible descubrir cuellos de botella, medir niveles de servicio, monitorear la utilización de recursos, y predecir el tiempo de procesamiento restante de casos en ejecución.

Hay algunas ideas erradas en relación a minería de procesos. Algunos proveedores, analistas, e investigadores limitan el alcance de la minería de procesos a una técnica especial de minería de datos para el descubrimiento de procesos que puede solo ser usada para análisis *offline*. Esto *no* es así, por lo tanto, enfatizamos las siguientes tres características.

- *La minería de procesos no está limitada al descubrimiento del control de flujo.* El descubrimiento de modelos de procesos desde los registros de eventos llena la imaginación tanto de profesionales como de académicos. Por lo tanto, el descubrimiento del control de flujo es a menudo visto como la parte más emocionante de la minería de procesos. Sin embargo, la minería de procesos no se limita al descubrimiento del control de flujo. Por una parte, el descubrimiento es sólo una de las tres formas básicas de minería de procesos (descubrimiento, conformidad, y mejoramiento). Por otra parte, el alcance no está limitado al control de flujo; las perspectivas organizacional, de casos y de tiempo también cumplen un rol importante.
- *La minería de procesos no es sólo un tipo específico de minería de datos.* La minería de procesos se puede ver como el "eslabón perdido" entre la minería de datos y el BPM tradicional basado en modelos. La mayoría de las técnicas de minería de datos no están en absoluto centradas en procesos. Los modelos de proceso que potencialmente exhiben concurrencia son incomparables a las estructuras de minería de datos simples, tales como los árboles de decisión y las reglas de asociación. Por lo tanto, se necesitan nuevos tipos de representación y de algoritmos.
- *La minería de procesos no está limitada al análisis offline.* Las técnicas de minería de procesos extraen conocimiento de los datos de eventos históricos. Aunque se utilizan datos "post mortem", los resultados pueden ser aplicados a casos en ejecución. Por ejemplo, se puede predecir el tiempo de finalización de un pedido del cliente parcialmente realizado, usando un modelo de proceso descubierto.

Para posicionar la minería de procesos, usamos el ciclo de vida de BPM mostrado en la Fig. 4. El ciclo de vida de BPM muestra las siete fases de un proceso de negocio y sus correspondientes sistemas de información. En la *fase de (re)diseño* se crea un nuevo modelo de proceso o se adapta un modelo de proceso existente. En la *fase de análisis* se analiza un modelo candidato y sus alternativas. Después de la fase de (re)diseño, se implementa el modelo (*fase de implementación*) o se (re)configura un sistema existente (*fase de (re)configuración*). En la *fase de ejecución* se ejecuta el modelo diseñado. Durante la fase de ejecución el proceso es *monitoreado*. Además, se podrían realizar pequeños ajustes sin rediseñar el proceso (*fase de ajuste*). En la *fase de diagnóstico* se analiza el proceso ejecutado y la salida de esta fase podría gatillar una nueva fase de rediseño del proceso. La minería de procesos es una herramienta valiosa para la mayoría de las fases mostradas en la Fig. 4. Obviamente, la fase de diagnóstico puede beneficiarse de la minería de procesos. Sin embargo, la minería de procesos no está limitada a la fase de diagnóstico. Por ejemplo, en la fase de ejecución, las técnicas de minería de procesos se pueden usar para el *soporte operacional*. Se pueden utilizar predicciones y recomendaciones basadas en modelos aprendidos usando información histórica para influenciar los casos en ejecución. Se pueden utilizar formas similiares de apoyo a la toma de decisiones para ajustar los procesos y guiar la (re)configuración de procesos.

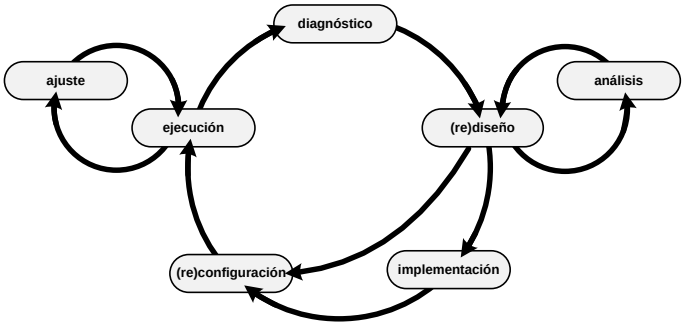


Figura 4. El ciclo de vida de BPM identificando las diferentes fases de un proceso de negocio y sus correspondientes sistemas de información; la minería de procesos cumple (potencialmente) un rol en todas las fases (excepto la fase de implementación).

Mientras la Fig. 4 muestra el ciclo de vida BPM como un todo, la Fig. 5 se enfoca en las actividades y artefactos concretos de minería de procesos. La Fig. 5 describe las etapas posibles en un proyecto de minería de procesos. Cualquier proyecto de minería de procesos comienza con una planificación y una justificación para esta planificación (Etapa 0). Después de iniciado el proyecto, se necesita extraer los datos de eventos, modelos, objetivos, y preguntas a partir de los sistemas, expertos del dominio, y la gestión (Etapa 1). Esto requiere un entendimiento de los datos disponibles ("¿Qué puede ser usado para el análisis?") y un entendimiento del dominio ("¿Cuáles son las preguntas importantes?") y tiene como resultado los artefactos mostrados en la Fig. 5 (i.e., datos históricos, modelos hechos a mano, objetivos, y preguntas). En la Etapa 2, se construye el modelo de control de flujo y se le relaciona con el registro de eventos. Aquí, se pueden utilizar técnicas automáticas de descubrimiento de procesos. El modelo de procesos descubierto ya podría proveer respuestas a algunas de las preguntas y gatillar acciones de rediseño o ajuste. Además, se podría filtrar o adaptar el registro de eventos usando el modelo (e.g., eliminando actividades poco frecuentes o casos atípicos, e insertando eventos faltantes). Algunas veces se necesitan significativos esfuerzos para correlacionar eventos que pertenecen a la misma instancia de un proceso. Los eventos restantes están relacionados con entidades del modelo de proceso. Cuando el proceso es relativamente estructurado, el modelo de control de flujo podría ser extendido con otras perspectivas (e.g., datos, tiempo, y recursos) durante la Etapa 3. La relación entre el registro de eventos y el modelo establecido en la Etapa 2 se utiliza para extender el modelo (e.g., se utilizan las marcas de tiempo de los eventos asociados para estimar los tiempos de espera para las actividades). Esto podría utilizarse para responder preguntas adicionales y podría gatillar acciones adicionales. En última instancia, los modelos construidos en la Etapa 3 podrían ser utilizado para apoyar las operaciones (Etapa 4). El conocimiento extraído de los datos de eventos históricos se combina con la información acerca de los casos en ejecución. Esto podría utilizarse para intervenir, predecir, y recomendar. Las Etapas 3 y 4 sólo se pueden alcanzar si el proceso es suficientemente estable y estructurado.

Actualmente, hay técnicas y herramientas que pueden apoyar todas las etapas mostradas en la Fig. 5. Sin embargo, la minería de procesos es un paradigma relativamente nuevo y la mayoría de las herramientas actuales son todavía algo inmaduras. Además, los usuarios potenciales a menudo no son conscientes del potencial y las limitaciones de la minería de procesos. Por lo tanto, este manifiesto cataloga algunos principios rectores (cf. Sección 3) y desafíos (cf. Sección 4) para los usuarios de las técnicas de minería de procesos, así como también para los investigadores y desarrolladores que están interesados en el avance del estado del arte.

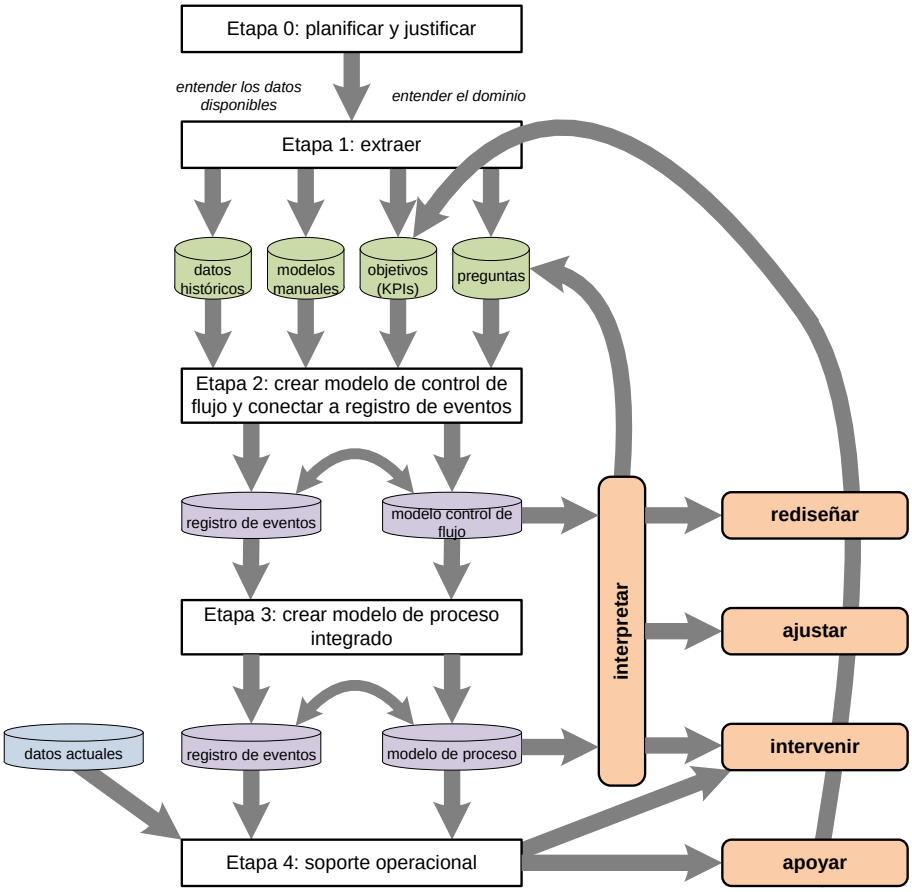


Figura 5. El modelo de ciclo de vida L^* describe un proyecto de minería de procesos consistente de cinco etapas: planificar y justificar (Etapa 0), extraer (Etapa 1), crear un modelo de control de flujo y conectarlo con el registro de eventos (Etapa 2), crear un modelo de proceso integrado (Etapa 3), y proveer soporte operacional (Etapa 4).

3. Principios Rectores

Como con cualquier nueva tecnología, hay errores obvios que pueden cometerse cuando se aplica minería de procesos en entornos de la vida real. Por lo tanto, listamos seis *principios rectores* para evitar que los usuarios/analistas comentan dichos errores.

3.1. PR1: Los Datos de Eventos Deberían Ser Tratados como Ciudadanos de Primera Clase

El punto de partida para cualquier actividad de minería de procesos son los eventos registrados. Nos referimos a colecciones de eventos como *registros de eventos*, sin embargo, esto no implica que los eventos deban estar almacenados en archivos de registro dedicados. Los eventos podrían estar almacenados en tablas de bases de datos, registros de mensajes, archivos de correo, registros de transacciones, y otras fuentes de datos. Más importante que el formato de almacenamiento, es la *calidad* de tales registros de eventos. La calidad de un resultado de minería de procesos en gran medida depende de la entrada. Por lo tanto, los registros de eventos deberían ser tratados como *ciudadanos de primera clase* en los sistemas de información que apoyan los procesos a ser analizados. Desafortunadamente, los registros de eventos son a menudo meramente un "sub-producto" utilizado para depurar o medir el rendimiento del software. Por ejemplo, los dispositivos

médicos de Philips Healthcare registran eventos simplemente porque los desarrolladores de software han insertado "instrucciones de impresión" en el código. Aunque hay algunas directrices informales para agregar dichas instrucciones al código, se necesita un enfoque más sistemático para mejorar la calidad de los registros de eventos. Los datos de eventos deberían ser vistos como ciudadanos de primera clase (más que ciudadanos de segunda clase).

Hay varios criterios para juzgar la calidad de los datos de eventos. Los eventos deben ser *confiables*, i.e., debería ser seguro asumir que los eventos registrados realmente ocurrieron y que los atributos de los eventos son correctos. Los registros de eventos deberían ser *completos*, i.e., dado un determinado contexto, no puede faltar ningún evento. Cualquier evento registrado debe tener una *semántica* bien definida. Además, los datos de eventos deben ser *seguros* en el sentido que se tengan en cuentas consideraciones de privacidad y seguridad al registrar los eventos. Por ejemplo, los actores deben ser conscientes del tipo de eventos que se registra y la forma en que se utilizan.

La Tabla 1 define cinco niveles de madurez de un registro de eventos que van desde excelente calidad ($\star\star\star\star\star$) a mala calidad ($\star$). Por ejemplo, los registros de eventos de Philips Healthcare residen en el nivel $\star\star\star$, i.e., los eventos se registran automáticamente y el comportamiento registrado calza con la realidad, pero no existe un enfoque sistemático para asignar semántica a los eventos y para garantizar cobertura en un nivel particular. Las técnicas de minería de procesos pueden ser aplicadas a registros de eventos en niveles $\star\star\star\star\star$, $\star\star\star\star$ y $\star\star\star$. En principio, también es posible aplicar minería de procesos utilizando registros de eventos en niveles $\star\star$ o $\star$. Sin embargo, el análisis de dichos registros de eventos es generalmente problemático y los resultados no son confiables. De hecho, no tiene mucho sentido aplicar minería procesos a registros de eventos en el nivel $\star$.

Para obtener beneficio de la minería de procesos, las organizaciones deben apuntar a registros de eventos en el nivel de calidad más alto posible.

3.2. PR2: La Extracción de Registros de Eventos Debería Ser Impulsada por Preguntas

Como se muestra en la Fig. 5, las actividades de minería de procesos necesitan ser impulsadas por preguntas. Sin preguntas concretas es muy difícil extraer datos de eventos significativos. Considere, por ejemplo, los miles de tablas en la base de datos de un sistema ERP como SAP. Sin preguntas concretas es imposible seleccionar las tablas relevantes para la extracción de datos.

Un modelo de proceso como el mostrado en la Fig. 1 describe el ciclo de vida de los casos (i.e., instancias de proceso) de un tipo particular. Por lo tanto, antes de aplicar cualquier técnica de minería de procesos hay que elegir el tipo de casos a ser analizado. Esta elección debe ser impulsada por las preguntas que se necesita contestar, y esto puede no ser trivial. Considere, por ejemplo, el manejo de pedidos de los clientes. Cada pedido de un cliente podría consistir de múltiples líneas de pedido, dado que el cliente podría solicitar varios productos en un solo pedido. Un pedido del cliente podría resultar en varias entregas. Una entrega puede referirse a líneas de pedido de múltiples pedidos. Por lo tanto, existe una relación de muchos a muchos entre los pedidos y las entregas, y una relación de uno a muchos entre los pedidos y las líneas de pedido. Dada una base de datos con datos de eventos relacionados con los pedidos, las líneas de pedido, y las entregas, hay diferentes modelos de proceso que pueden ser descubierto. Se puede extraer datos con el objetivo de describir el ciclo de vida de cada pedido. Sin embargo, también es posible extraer datos con el objetivo de descubrir el ciclo de vida de cada línea de pedido o el ciclo de vida de cada entrega.

3.3. PR3: Se Debería Dar Soporte a Concurrencia, Elección y Otros Conceptos Básicos de Control de Flujo

Existe una gran cantidad de lenguajes de modelación de procesos (e.g., BPMN, EPC, redes de Petri, BPEL, y los diagramas de actividad UML). Algunos de estos lenguajes proporcionan muchos elementos de modelación (e.g., BPMN ofrece más de 50 elementos gráficos distintos), mientras que otros son muy básicos (e.g., las redes de Petri se componen de sólo tres elementos

Cuadro 1. Niveles de madurez para los registros de eventos.

Nivel	Caracterización
⋆⋆⋆⋆⋆	Nivel más alto: el registro de eventos es de excelente calidad (i.e., confiable y completo) y los eventos están bien definidos. Los eventos se registran de manera automática, sistemática, confiable, y segura. Se toman en cuenta adecuadamente consideraciones acerca de la privacidad y la seguridad. Además, los eventos registrados (y todos sus atributos) tienen una semántica clara. Esto implica la existencia de una o más ontologías. Los eventos y sus atributos se refieren a esta ontología. *Ejemplo:* registros de eventos anotados semáticamente de los sistemas BPM.
⋆⋆⋆⋆	Los eventos se registran automáticamente y de manera sistemática y confiable, i.e., los registros de eventos son confiables y completos. A diferencia de los sistemas operando a nivel ⋆⋆⋆, se da soporte de manera explícita a nociones tales como instancia de proceso (caso) y actividad. *Ejemplo:* los registros de eventos de los sistemas tradicionales de BPM/workflow.
⋆⋆⋆	Los eventos se registran automáticamente, pero no se sigue un enfoque sistemático para registrar los eventos. Sin embargo, a diferencia de los registros de eventos en el nivel ⋆⋆, hay algún nivel de garantía que los eventos registrados calzan con la realidad (i.e., el registro de eventos es confiable pero no necesariamente completo). Considere, por ejemplo, los eventos registrados por un sistema ERP. Aunque se necesita extraer los eventos de una variedad de tablas, se puede asumir que la información es correcta (e.g., es razonable asumir que un pago registrado por el ERP efectivamente existe, y viceversa). *Ejemplo:* las tablas en un sistema ERP, los registros de eventos de sistemas CRM, registros de transacciones de sistemas de mensajería, registros de eventos de sistemas de alta tecnología, etc.
⋆⋆	Los eventos se registran automáticamente, i.e., como un subproducto de algún sistema de información. La cobertura varía, i.e., no se sigue un enfoque sistemático para decidir qué eventos se registran. Además, es posible pasar por alto el sistema de información. Por lo tanto, podrían faltar eventos o éstos podrían no registrarse correctamente. *Ejemplo:* los registros de eventos de sistemas de gestión de documentos y productos, registros de errores de sistemas embebidos, planillas de ingenieros de servicios, etc.
⋆	Nivel más bajo: los registros de eventos son de mala calidad. Los eventos registrados podrían no corresponder a la realidad y podrían faltar eventos. Los registros de eventos en los cuales los eventos se registran manualmente suelen tener dichas características. *Ejemplo:* trazas dejadas en documentos en papel que se trasladan a través de la organización (notas tipo "Post-it"), expedientes médicos en papel, etc.

diferentes: sitios, transiciones y arcos). La descripción del control de flujo es la columna vertebral de cualquier modelo de proceso. Los conceptos basicos de control de flujo (también conocidos como *patrones*) a los cuales todos los lenguajes principales dan soporte son secuencia, paralelismo (AND-splits/joins), elección (XOR-splits/joins), y ciclos. Obviamente, las técnicas de minería de procesos deberían dar soporte a estos patrones. Sin embargo, algunas técnicas no son capaces de manejar la concurrencia y sólo permiten cadenas de Markov/sistemas de transicción.

La Fig. 6 muestra el efecto de usar técnicas de minería de procesos que no son capaces de descubrir la concurrencia (sin AND-split/joins). Consider un registro 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 contiene casos que parten con A y finalizan con E. Las actividades B, C, y D ocurren en cualquier orden entre

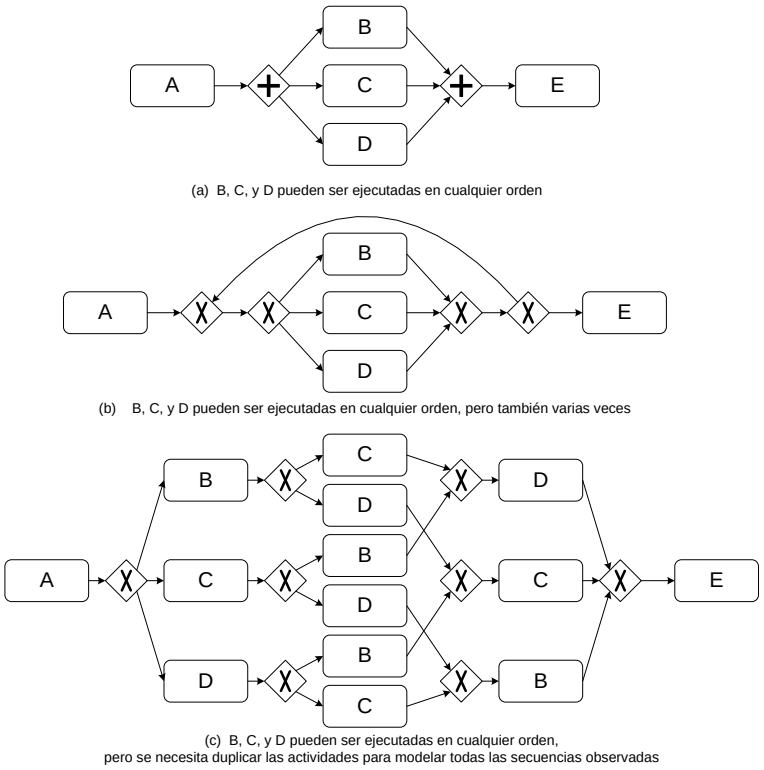


Figura 6. Ejemplo que ilustra los problemas que ocurren cuando no se puede expresar directamente la concurrencia (i.e., AND-splits/joins). En el ejemplo sólo tres actividades (B, C, y D) son concurrentes. Imagine los modelos de proceso resultantes cuando hay 10 actividades concurrentes ($2^{10} = 1{,}024$ estados y $10! = 3{,}628{,}800$ posibles secuencias de ejecución).

A y E. El modelo BPMN en la Fig. 6(a) muestra una representación compacta del proceso subyacente usando dos compuertas AND. Suponga que la técnica de minería de procesos no soporta compuertas AND. En este caso, los otros dos modelos BPMN en la Fig. 6 son candidatos obvios. El modelo BPMN en la Fig. 6(b) es compacto, pero permite demasiados comportamientos (e.g., casos tales como $\langle A, B, B, B, E\rangle$ son posibles de acuerdo al modelo, pero no son probables de acuerdo al registro de eventos). El modelo BPMN en la Fig. 6(c) permite los casos en L, pero considera todas las secuencias explícitamente, por lo que no es una representación compacta del registro de eventos. El ejemplo muestra que para modelos de la vida real con docenas de actividades potencialmente concurrentes, los modelos resultantes son sub-ajustados (i.e., permiten demasiados comportamientos) y/o son extremadamente complejos si no se soporta concurrencia.

Como se ilustra en la Fig. 6, es importante dar soporte al menos a los patrones básicos de *workflow*. Además de los patrones básicos mencionados, es también desesable dar soporte a OR-splits/joins, ya que estos proporcionan una representación compacta de decisiones inclusivas y sincronizaciones parciales.

3.4. PR4: Los Eventos Deberían Estar Relacionados a Elementos del Modelo

Como se indicó en la Sección 2, es un error pensar que la minería de procesos se limita al descubrimiento de control de flujo. Como se muestra en la Fig. 1, el modelo de proceso descubierto podría cubrir diversas perspectivas (perspectiva organizacional, perspectiva temporal, perspectiva de datos, etc.). Además, el descubrimiento es sólo uno de los tres tipos de minería de procesos mostrados en la Fig. 3. Los otros dos tipos de minería de procesos (verificación de conformidad y mejoramiento) dependen fuertemente de la relación entre los *elementos en el modelo* y los

eventos en el registro de eventos. Esta relación podría ser usada para "repetir la ejecución" del registro de eventos sobre el modelo. La repetición de la ejecución podría ser utilizada para revelar discrepancias entre un registro de eventos y un modelo, e.g., algunos eventos en el registro de eventos no son posibles de acuerdo al modelo. Las técnicas para la verificación de conformidad, cuantifican y diagnostican dichas discrepancias. Las marcas de tiempo en el registro de eventos se pueden utilizar para analizar el comportamiento temporal durante la repetición de la ejecución. Las diferencias de tiempo entre las actividades relacionadas causalmente se pueden utilizar para agregar tiempos de espera estimados en el modelo. Estos ejemplos muestran que la relación entre los eventos en el registro de eventos y los elementos en el modelo sirve como punto de partida para diferentes tipos de análisis.

En algunos casos puede no ser trivial establecer dicha relación. Por ejemplo, un evento podría referirse a dos actividades diferentes o no estar claro a qué actividad se refiere. Tales ambigüedades deben ser eliminadas a fin de interpretar correctamente los resultados de la minería de procesos. Además del problema de relacionar los eventos con actividades, existe el problema de relacionar los eventos con instancias de proceso. Esto comúnmente se conoce como *correlación de eventos*.

3.5. PR5: Se Debería Tatar a los Modelos como Abstracciones Útiles de la Realidad

Los modelos derivados de datos de eventos proporcionan *puntos de vista sobre la realidad*. Dichos puntos de vista deberían proporcionar una abstracción útil del comportamiento capturado en el registro de eventos. Dado un registro de eventos, podría haber múltiples puntos de vista que son útiles. Además, las diversas partes interesadas pueden requerir diferentes puntos de vista. De hecho, los modelos descubiertos a partir de los registros de eventos deberían ser visto como "mapas" (como los mapas geográficos). Este principio rector proporciona importantes intuiciones, dos de los cuales se describen a continuación.

En primer lugar, es importante tener en cuenta que no hay tal cosa como "el mapa" de un área geográfica particular. Dependiendo del uso previsto existen diferentes mapas: mapas de carreteras, mapas de senderismo, mapas de ciclismo, etc. Todos estos mapas muestran un punto de vista sobre una misma realidad, y sería absurdo suponer que habría tal cosa como "el mapa perfecto". Lo mismo vale para los modelos de proceso: el modelo debería enfatizar las cosas relevantes para un determinado tipo de usuario. Los modelos descubiertos podrían focalizarse en diferentes perspectivas (control de flujo, flujo de datos, tiempo, recursos, costos, etc.) y mostrarlas en diferentes niveles de granularidad y precisión, e.g., un gerente podría querer ver un modelo de proceso informal grueso que se focaliza en los costos, mientras que un analista de procesos podría querer ver un modelo de proceso detallado que se focaliza en las desviaciones del flujo normal. También tenga en cuenta que las diferentes partes interesadas podrían desear ver un proceso en diferentes niveles: *nivel estratégico* (las decisiones en este nivel tienen efectos a largo plazo y se basan en datos de eventos agregados a través de un período más largo), *nivel táctico* (las decisiones en este nivel tienen efectos a mediano plazo y se basan principalmente en datos recientes), *nivel operacional* (las decisiones en este nivel tienen efectos inmediatos y se basan en datos de eventos relacionados con casos en ejecución).

En segundo lugar, es útil adoptar ideas de la cartografía a la hora de producir mapas comprensibles. Por ejemplo, los mapas de carreteras pasan por alto las carreteras y ciudades menos importantes. Las cosas menos importantes son dejadas de lado o dinámicamente agrupadas en formas agregadas (e.g., calles y suburbios se agrupan en ciudades). Los cartógrafos no sólo eliminan los detalles irrelevantes, sino que también utilizan colores para destacar características importantes. Además, los elementos gráficos tienen un tamaño especial para indicar su relevancia (e.g., el tamaño de las líneas y puntos puede variar). Los mapas geográficos también tienen una interpretación clara de los eje-x y eje-y, i.e., el diseño de un mapa no es arbitrario, dado que las coordenadas de los elementos tienen un significado. Todo esto está en marcado contraste con los principales modelos de procesos que no utilizan típicamente los atributos de color, tamaño, y ubicación para hacer los modelos más comprensible. Sin embargo, las ideas de la cartografía se pueden incorporar fácilmente en la construcción de los mapas de proceso descubiertos. Por ejemplo, el tamaño de una actividad puede ser utilizado para reflejar su frecuencia o alguna otra propiedad que indica

su relevancia (por ejemplo, costos o uso de recursos). El ancho de un arco puede reflejar la importancia de la dependencia causal correspondiente, y el color de los arcos se puede utilizar para destacar los cuellos de botella.

Las observaciones anteriores muestran que es importante seleccionar la representación correcta y sintonizarla para la audiencia objetivo. Esto es importante para la visualización de los resultados para los usuarios finales y para guiar a los algoritmos de descubrimiento hacia modelos adecuados (véase también el Desafío D5).

3.6. PR6: La Minería de Procesos Debería Ser un Proceso Continuo

La minería de procesos puede ayudar a proveer "mapas" significativos que están conectados directamente a los datos de eventos. Tanto los datos de eventos históricos como los datos actuales se pueden proyectar en estos modelos. Además, los procesos cambian mientras están siendo analizados. Dada la naturaleza dinámica de los procesos, no es recomendable ver a la minería de procesos como una actividad puntual. El objetivo no debería ser la creación de un modelo fijo, sino que dar vida a los modelos de procesos, de manera que se incentive a los usuarios y analistas a mirarlos diariamente.

Compare esto con el uso de mashups utilizando geo-tagging. Hay miles de mashups que utilizan Google Maps (e.g., las aplicaciones que proyectan en un mapa seleccionado información sobre las condiciones del tráfico, bienes raíces, restaurantes de comida rápida, o las carteleras de cine). La gente puede sin problemas ampliar y reducir dichos mapas e interactuar con ellos (e.g., los atascos de tráfico se proyectan en el mapa y el usuario puede seleccionar un problema particular para ver los detalles). También debería ser posible realizar minería de procesos basada en datos de eventos en tiempo real. Utilizando la "metáfora del mapa", podemos pensar que los eventos tienen coordenadas GPS que pueden ser proyectadas en los mapas en tiempo real. En forma análoga a los sistemas de navegación para automóviles, las herramientas de minería de procesos pueden ayudar a los usuarios finales a (a) navegar a través de los procesos, (b) proyectar información dinámica sobre los mapas de procesos (e.g., mostrando "atascos de tráfico" en los procesos de negocio), y (c) proporcionar predicciones respecto a los casos en ejecución (e.g., estimando el "tiempo de arribo" de un caso que se retrasó). Estos ejemplos demuestran que es una pena que no se utilicen modelos de procesos más activamente. Por lo tanto, la minería de procesos debería ser vista como un proceso continuo, proporcionando información útil en diversas escalas de tiempo (minutos, horas, días, semanas y meses).

4. Desafíos

La minería de procesos es una herramienta importante para las organizaciones modernas que necesitan gestionar procesos operacionales no triviales. Por un lado, hay un increíble crecimiento en la cantidad de datos de eventos. Por otro lado, los procesos y la información necesitan estar perfectamente alineados para cumplir requerimientos relacionados con el cumplimiento de normas, eficiencia y servicio al cliente. A pesar de la aplicabilidad de la minería de procesos aún hay desafíos importantes que necesitan ser abordados; estos ilustran que la minería de procesos es una disciplina emergente. A continuación, entregamos una lista de algunos de estos desafíos. No se pretende que esta lista sea completa y, en el tiempo, podrían aparecer nuevos desafíos o podrían desaparecer desafíos existentes debido a los avances en minería de procesos.

4.1. D1: Encontrar, Fusionar y Limpiar Datos de Eventos

Todavía toma esfuerzos considerables extraer datos de eventos apropiados para la minería de procesos. Típicamente, se necesita superar varios obstáculos:

- Los datos pueden estar *distribuídos* en varias fuentes. Esta información necesita ser fusionada. Esto tiende a ser problemático cuando se utilizan distintos identificadores en las diferentes fuentes de datos. Por ejemplo, un sistema usa el nombre y la fecha de nacimiento para identificar una persona mientras que otro sistema utiliza el número de seguridad social de la persona.

- Los datos de eventos están a menudo "centrados en objetos" más que "centrados en procesos". Por ejemplo, productos individuales, pallets, y contenedores pueden tener etiquetas RFID y eventos registrados referidos a esas etiquetas. Sin embargo, para monitorear una orden de un cliente en particular, tales eventos centrados en objetos necesitan ser fusionados y preprocesados.
- Los datos de eventos pueden estar *incompletos*. Un problema común es que los eventos no apuntan explícitamente a intancias del proceso. A menudo es posible derivar esta información, pero esto puede tomar esfuerzos considerables. También la información de tiempo puede estar perdida para algunos eventos. Uno puede necesitar interpolar las marcas de tiempo para aún utilizar la información de tiempo disponible.
- Un registro de eventos puede contener datos atípicos (*outliers*), i.e., comportamiento excepcional también referido como *ruido*. ¿Cómo definir outliers? ¿Cómo detectar tales outliers? Se necesita responder estas preguntas para limpiar los datos de eventos.
- Los registros de eventos pueden contener eventos en *diferentes niveles de granularidad*. En el registro de eventos de un sistema de información de un hospital los eventos se pueden referir a examenes de sangre simples o a procedimientos quirúrgicos complejos. También las marcas de tiempos pueden tener diferentes niveles de granularidad que van desde una precisión de milisegundos (28-9-2011:h11m28s32ms342) a información de tiempo gruesa (28-9-2011).
- Los eventos ocurren en un *contexto* particular (clima, carga de trabajo, día de la semana, etc.). Este contexto puede explicar ciertos fenómenos, por ejemplo, el tiempo de respuesta es más largo que el usual debido a trabajos-en-desarrollo o por vacaciones. Para el análisis, es deseable incorporar este contexto. Esto implica la fusión de datos de eventos con datos de contexto. Aquí la "la maldición de la dimensionalidad" aparece cuando el análisis se convierte en intratable al agregar demasiadas variables.

Se necesitan mejores herramientas y metodologías para abordar los problemas anteriores. Además, como indicamos antes, las organizaciones necesitan tratar los registros de eventos como ciudadanos de primera clase más que un sub-producto. El objetivo es obtener registros de eventos ⋆⋆⋆⋆⋆ (ver Tabla 1). Aquí, las lecciones aprendidas en el contexto de almacenamiento de datos (datawarehousing) son muy útiles para asegurar una alta calidad en los registros de eventos. Por ejemplo, revisiones simples durante el ingreso de datos pueden ayudar a reducir significativamente la proporción de datos de eventos incorrectos.

4.2. D2: Lidiar con Registros de Eventos Complejos con Diversas Características

Los registros de eventos pueden tener diferentes características. Algunos registros de eventos pueden ser extremadamente grandes lo cual hace difícil manipularlos mientras otros registros de eventos son tan pequeños que no tienen suficientes datos para obtener conclusiones confiables.

En algunos dominios, se registran cantidades alucinantes de eventos. Por lo tanto, se necesitan esfuerzos adicionales para mejorar el desempeño y la escalabilidad. Por ejemplo, ASML está continuamente monitoreando todos sus escáner de obleas (*wafer scanners*). Estos escáner de obleas son utilizados por varias organizaciones (e.g., Samsung y Texas Instruments) para producir chips (aprox. 70 % de los chips son producidos utilizando los escáner de obleas de ASML). Las herramientas existentes tienen dificultades para lidiar con los petabytes de datos recolectados en tales dominios. Además del número de eventos registrados hay otras características tales como el número promedio de eventos por caso, similaridades entre casos, el número de eventos únicos, y el número de caminos únicos. Considere un registro de eventos $L1$ con las siguientes características: 1000 casos, un promedio de 10 eventos por caso, y poca variación (e.g., varios casos siguen el mismo camino o caminos muy similares). El registro de eventos $L2$ contiene sólo 100 casos, pero en promedio hay 100 eventos por caso y todos los casos siguen un camino propio. Claramente, $L2$ es mucho más difícil de analizar que $L1$ aún cuando los dos registros de eventos tienen tamaños similares (aproximadamente 10.000 eventos).

Dado que los registros de eventos contienen sólo muestras de comportamiento, no se puede asumir que ellos están completos. Las técnicas de minería de procesos necesitan lidiar con la

incompletitud de datos utilizando un "supuesto de mundo abierto": el hecho que algo no sucedió no significa que no pueda suceder. Esto hace difícil lidiar con registros de eventos pequeños con mucha variabilidad.

Como mencionamos antes, algunos registros contienen eventos en niveles de abstracción muy detallados. Estos registros tienden a ser extremadamente grandes y los eventos individuales de bajo nivel son de poco interés para las partes interesadas. Por lo tanto, uno desearía agregar los eventos de bajo nivel en eventos de alto nivel. Por ejemplo, cuando se analiza los procesos de diagnóstico y tratamiento de un grupo particular de pacientes, uno puede no estar interesado en los examenes individuales registrados en el sistema de información del laboratorio del hospital.

Actualmente, las organizaciones necesitan utilizar un enfoque de prueba y error para ver si un registro de eventos es apropiado para minería de procesos. Por lo tanto, las herramientas deberían permitir realizar un examen rápido de factibilidad dado un conjunto de datos particular. Dicho examen debería indicar potenciales problemas de desempeño y advertir sobre registros que están lejos de ser completos o que son muy detallados.

4.3. D3: Crear Puntos de Referencia Representativos

La minería de procesos es una tecnología emergente. Esto explica por qué aún faltan buenos puntos de referencia (*benchmarks*). Por ejemplo, docenas de técnicas de descubrimiento de procesos están disponibles y diferentes proveedores ofrecen distintos productos, pero no hay consenso sobre la calidad de esas técnicas. Aunque hay diferencias gigantescas en funcionalidad y desempeño, es difícil comparar las diferentes técnicas y herramientas. Por lo tanto, se necesita desarrollar buenos puntos de referencias que consistan de conjuntos de datos de ejemplo y criterios de calidad representativos.

Para las técnicas clásicas de minería de datos, hay muchos y buenos puntos de referencias disponibles. Estos puntos de referencias han estimulado a los proveedores e investigadores a mejorar el desempeño de sus técnicas. En el caso de la minería de procesos esto es más desafiante. Por ejemplo, el modelo relacional introducido por Codd en 1969 es simple y ampliamente soportado. Como resultado, toma poco esfuerzo convertir datos desde una base de datos a otra, y no hay problemas de interpretación. Para los procesos hace falta un modelo así de simple. Los estándares propuestos para la modelación de procesos son mucho más complicados, y pocos proveedores soportan exactamente el mismo conjunto de conceptos. Los procesos son simplemente más complejos que los datos tabulares.

No obstante, es importante crear puntos de referencia para minería de procesos. Ya están disponibles algunos trabajos iniciales. Por ejemplo, hay varias métricas para medir la calidad de los resultados de la minería de procesos (ajuste, simplicidad, precisión y generalización). Además, varios registros de eventos están disponibles públicamente (cf. www.processmining.org). Vea por ejemplo el registro de eventos utilizado para el primer Desafío de Inteligencia de Procesos de Negocio (*Business Process Intelligence Challenge*, BPIC'11) organizado por la Fuerza de Trabajo (cf. doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54).

Por un lado, deberían haber puntos de referencia basados en bases de datos de la vida real. Por otro lado, está la necesidad de crear conjuntos de datos sintéticos capturando características particulares. Tales conjuntos de datos sintéticos ayudan a desarrollar técnicas de minería de procesos que son hechas a la medida para registros de eventos incompletos, registros de eventos con ruido, o para poblaciones específicas de procesos.

Además de la creación de puntos de referencia representativos, hay también necesidad de mayor consenso sobre los criterios utilizados para juzgar la calidad de los resultados de la minería de procesos (ver también Desafío D6). Además, se puede adaptar las técnicas de *validación cruzada* de la minería de datos para juzgar el resultado. Considere por ejemplo la validación cruzada de k-iteraciones. Uno puede dividir el registro de eventos en k partes. $k-1$ partes pueden ser utilizadas para aprender un modelo de proceso, y las técnicas de chequeo de conformidad pueden ser utilizadas para juzgar el resultado con respecto a la parte restante. Esto puede ser repetido k veces, para así proveer alguna idea sobre la calidad del modelo.

4.4. D4: Lidiar con el Cambio de Tendencia

El término *cambio de tendencia* (*concept drift*) se refiere a la situación en la cual el proceso está cambiando mientras está siendo analizado. Por ejemplo, en el comienzo del registro de eventos dos actividades pueden ser concurrentes mientras que más tarde en el registro esas actividades se convierten en secuenciales. Los procesos pueden cambiar debido a cambios periódicos/estacionales (e.g., "en Diciembre hay más demanda" o "en la tarde del Viernes hay menos empleados disponibles") o debido a condiciones cambiantes (e.g., "el mercado se está volviendo más competitivo"). Tales cambios impactan los procesos y es vital detectarlos y analizarlos. El cambio de tendencia en un proceso puede ser descubierto al dividir el registro de eventos en registros más pequeños y analizar las "huellas" de los registros más pequeños. Tal análisis de "segundo orden" require muchos más datos de eventos. No obstante, pocos procesos están en un estado estable, y entender el cambio de tendencia es de suma importancia para la gestión de los procesos. Por lo tanto, se necesita más investigación y soporte en las herramientas para analizar adecuadamente el cambio de tendencia.

4.5. D5: Mejorar el Sesgo Representacional Utilizado para el Descubrimiento de Procesos

Una técnica de descubrimiento de procesos produce un modelo utilizando un lenguaje particular (e.g., BPMN o Redes de Petri). Sin embargo, es importante separar la visualización del resultado de la representación utilizada durante el proceso de descubrimiento propiamente tal. La selección de un lenguaje objetivo a menudo abarca varios supuestos implícitos. Esto limita el espacio de búsqueda; los procesos que no pueden ser representados por el lenguaje elegido no pueden ser descubierto. Este así llamado "sesgo representacional" utilizado durante el proceso de descubrimiento debería ser una elección consciente y no debería (sólo) estar impulsada por la representación gráfica preferida.

Considere por ejemplo la Fig. 6: si el lenguaje objetivo permite o no concurrencia puede tener un efecto en la visualización del modelo descubierto y la clase de modelos considerada por el algoritmo. Si el sesgo representacional no permite concurrencia (Fig. 6(a) no es posible) y no permite que múltiples actividades tengan la misma etiqueta (Fig. 6(c) no es posible), entonces sólo modelos problemáticos tales como los de la Fig. 6(b) son posibles. Este ejemplo muestra que se necesita una selección más cuidadosa y refinada del sesgo representacional.

4.6. D6: Balancear Criterios de Calidad tales como Ajuste, Simplicidad, Precisión y Generalización

Los registros de eventos están a menudo lejos de ser completos, es decir, sólo se cuenta con un comportamiento de ejemplo. Los modelos de procesos típicamente permiten un número exponencial o aún infinito de trazas diferentes (en caso de iteraciones). Además, algunas trazas pueden tener una probabilidad mucho más baja que otras. Por lo tanto, no es realista asumir que toda traza posible está presente en el registro de eventos. Para ilustrar que es poco práctico asumir que los registros de eventos están completos, considere un proceso consitente de 10 actividades que pueden ser ejecutadas en paralelo y un registro de eventos correspondiente que contiene información acerca de 10.000 casos. El número total de posibles caminos entrelazados (*interleavings*) en el modelo con 10 actividades concurrentes es $10! = 3.628.800$. Así, es imposible que cada camino entrelazado esté presente en el registro de eventos si hay menos casos (10.000) que las trazas potenciales (3.628.800). Aún si hay millones de casos en el registro, es extremadamente improbable que todas las posibles varaciones estén presente. Una complicación adicional es que algunas alternativas son menos frecuentes que otras. Estas pueden ser consideradas como "ruido". Es imposible construir un modelo razonable para tales comportamientos ruidosos. El modelo descubierto necesita abstraerse de esto; es mejor investigar comportamientos de baja frecuencia utilizando verificación de conformidad.

El ruido y la incompletitud hacen que el descubrimiento del proceso sea un problema desafiante. De hecho, hay cuatro dimensiones de calidad que compiten: (a) ajuste, (b) simplicidad, (c) precisión y (d) generalización. Un modelo con buen *ajuste* permite la mayor parte del comportamiento visto en el registro de eventos. Un modelo tiene un ajuste perfecto si todas las trazas del registro de eventos pueden ser repetidas por el modelo de comienzo a fin. El modelo *más simple* que puede explicar el comportamiento visto en el registro es el mejor modelo. Este principio es conocido como la Navaja de Occam. El ajuste y la simplicidad por sí solos no son suficiente para juzgar la calidad de un modelo de proceso descubierto. Por ejemplo, es muy fácil construir una red de Petri extremadamente simple ("modelo de flor") que es capaz de repetir todas las trazas en el registro de eventos (pero también cualquier otro registro de eventos referente al mismo conjunto de actividades). Similarmente, no es deseable tener un modelo que sólo permita el comportamiento exacto del registro de eventos. Recuerde que el registro contiene sólo comportamiento de ejemplo y que muchas trazas que son posibles pueden no haber sido vistas aún. Un modelo es *preciso* si no permite muchos comportamientos. Claramente, el "modelo de flor" carece de precisión. Un modelo que no es preciso está "subajustado". El subajuste es el problema en que el modelo sobre generaliza el comportamiento de ejemplo en el registro de eventos (i.e., el modelo permite comportamientos muy diferentes de aquellos vistos en el registro). Un modelo debería generalizar y no restringir comportamientos sólo a los ejemplos vistos en el registro. Un modelo que no *generaliza* está "sobreajustado". El sobreajuste es el problema que un modelo muy específico es generado mientras que es obvio que el registro sólo provee comportamientos de ejemplo (i.e., el modelo explica la muestra particular del registro, pero una siguiente muestra del registro del mismo proceso puede producir un modelo de proceso completamente diferente).

Balancear ajuste, simplicidad, precisión y generalización es desafiante. Esta es la razón por la que la mayoría de las técnicas de descubrimiento de procesos más potentes proveen varios parámetros. Se necesita desarrollar algoritmos mejorados para un mejor balance entre las cuatro dimensiones de calidad que compiten. Además, cualquier parámetro utilizado debería ser entendible por los usuarios finales.

4.7. D7: Minería Inter-Organizacional

Tradicionalmente, la minería de procesos se aplica dentro de una sola organización. Sin embargo, como la tecnología de servicios, la integración de cadenas de abatecimiento y la computación en la nube se extienden cada vez más, hay escenarios donde los registros de eventos de múltiples organizaciones están disponibles para análisis. En principio, hay dos escenarios para la *minería de procesos inter-organizacional*.

Primero, podemos considerar el escenario colaborativo donde diferentes organizaciones trabajan en conjunto para manejar instancias de procesos. Uno puede imaginar dicho proceso inter-organizacional como un "rompecabezas", i.e., el proceso completo se descompone en partes y se distribuye en varias organizaciones que necesitan cooperar para completar los casos exitosamente. Analizar el registro de eventos dentro de una de estas organizaciones involucradas es insuficiente. Para descubrir los procesos de punta-a-punta, se necesita fusionar los registros de eventos de diferentes organizaciones. Esto no es una tarea trivial, ya que los eventos necesitan ser correlacionados a través de las fronteras organizacionales.

Segundo, también podemos considerar el escenario donde diferentes organizaciones están ejecutando esencialmente el mismo proceso mientras comparten experiencias, conocimiento, o una infraestructura común. Considere por ejemplo Salesforce.com. Los procesos de ventas de varias organizaciones son gestionados y soportados por Salesforce. Por un lado, estas organizaciones comparten un infraestructura (procesos, bases de datos, etc.). Por otro lado, ellos no están obligados a seguir un modelo de proceso estricto ya que el sistema puede ser configurado para soportar variantes del mismo proceso. Como otro ejemplo, considere los procesos básicos ejecutados en cualquier municipalidad (e.g., entregar permisos de construcción). Aunque todas las municipalidades en un país necesitan proveer el mismo conjunto básico de procesos, puede también haber diferencias. Obviamente, es interesante analizar tales variaciones entre las diferentes organizaciones. Estas organizaciones pueden aprender la una de la otra, y los proveedores de servicios pueden

mejorar sus servicios y ofrecer servicios de valor agregado basados en los resultados de la minería de procesos inter-organizacional.

Se necesita desarrollar nuevas técnicas de análisis para ambos tipos de minería de procesos inter-organizacional. Estas técnicas también deberían considerar problemas de privacidad y seguridad. Las organizaciones podrían no querer compartir información por razones de competencia o debido a falta de confianza. Por lo tanto, es importante desarrollar técnicas de minería de procesos que preserven la privacidad.

4.8. D8: Proporcionar Soporte Operacional

Inicialmente, el foco de la minería de procesos estuvo en el análisis de datos históricos. Hoy, sin embargo, muchas fuentes de datos se actualizan (casi) en tiempo real, y hay suficiente capacidad computacional para analizar los eventos cuando ellos ocurren. Por lo tanto, la minería de procesos no debería estar restringida al análisis fuera de línea (off-line) y puede también ser utilizada para el soporte operacional en línea (on-line). Se pueden identificar tres actividades de soporte operacional: *detectar*, *predecir* y *recomendar*. En el momento en que un caso se desvía del proceso predefinido, esto puede ser detectado y el sistema puede generar una alerta. A menudo, uno quisiera generar tales notificaciones inmediatamente (para aún ser capaces de influir en las cosas), y no en una modalidad fuera de línea. Los datos históricos pueden ser utilizados para construir modelos predictivos. Estos pueden ser utilizados para guiar las instancias de proceso en ejecución. Por ejemplo, es posible predecir el tiempo de procesamiento restante de un caso. Basado en tales predicciones, uno puede también construir sistemas de recomendación que propongan acciones particulares para reducir costos o acortar el tiempo de flujo. Aplicar las técnicas de minería de procesos en tales escenarios en línea crea desafíos adicionales, en términos de capacidad de cómputo y calidad de datos.

4.9. D9: Combinar Minería de Procesos con Otros Tipos de Análisis

La gestión de operaciones, y en particular la investigación de operaciones, es una rama de la ciencia de gestión que depende fuertemente del modelamiento. Se utiliza una variedad de modelos matemáticos que van desde programación lineal y planificación de proyectos hasta modelos de colas, cadenas de Markov, y simulación. La minería de datos puede ser definida como "el análisis de conjuntos de datos (a menudo grandes) para encontrar relaciones insospechadas y para resumir los datos en formas novedosas que sean al mismo tiempo entendibles y útiles para el dueño de los datos". Se ha desarrollado una amplia variedad de técnicas: clasificación (e.g., árboles de decisión), regresión, segmentación (e.g., k-means) y descubrimiento de patrones (e.g., aprendizaje de reglas de asociación).

Ambos campos (gestión de operaciones y minería de datos) proveen técnicas de análisis valiosas. El desafío es combinar las técnicas en estos campos con la minería de procesos. Considere por ejemplo la simulación. Las técnicas de minería de procesos se pueden utilizar para aprender un modelo de simulación basado en datos históricos. Posteriormente, se puede utilizar el modelo de simulación para proveer soporte operacional. Debido a la cercana conexión entre el registro de eventos y el modelo, se puede utilizar el modelo para repetir la historia, y uno podría comenzar simulaciones desde el estado actual proporcionando un "botón de avance rápido" hacia el futuro basado en datos en tiempo real.

Similarmente, es deseable combinar la minería de procesos con la *analítica visual*. La anlítica visual combina el análisis automático con visualizaciones interactivas para un mejor entendimiento de conjuntos de datos grandes y complejos. La analítica visual explota las sorprendentes capacidades de los humanos para ver patrones en datos no estructurados. Al combinar las técnicas automáticas de minería de procesos con la analítica visual interactiva, es posible extraer más ideas a partir de los datos de eventos.

4.10. D10: Mejorar la Usabilidad para los No Expertos

Uno de los objetivos de la minería de procesos es crear "modelos de procesos reales", i.e., modelos de procesos que se utilizan a diario, más que modelos estáticos que terminan en algún archivo. Se puede usar los nuevos datos de eventos para descubrir comportamientos emergentes. La relación entre los datos de eventos y los modelos de procesos permiten la proyección del estado actual y las actividades recientes en modelos actualizados (al día). Por lo tanto, los usuarios finales pueden interactuar con los resultados de la minería de procesos de forma diaria. Tales interacciones son muy valiosas, pero también requieren interfaces de usuario intuitivas. El desafío es esconder los sofisticados algoritmos de minería de procesos detrás de interfaces de usuario amigables que automáticamente definan parámetros y sugieran tipos de análisis apropiados.

4.11. D11: Mejorar el Entendimiento para los No Expertos

Aún cuando es fácil generar los resultados de la minería de procesos, esto no significa que los resultados sean realmente útiles. El usuario puede tener problemas para entender la salida, o es tentado a inferir conclusiones incorrectas. Para evitar tales problemas, los resultados deberían ser presentados utilizando una representación apropiada (ver también PR5). Además, la fiabilidad de los resultados debería estar siempre claramente indicada. Podría haber muy pocos datos para justificar conclusiones particulares. De hecho, las técnicas existentes de descubrimiento de procesos típicamente no alertan acerca de un bajo ajuste o un sobreajuste. Siempre muestran un modelo, aún cuando es claro que hay muy pocos datos para justificar cualquier conclusión.

5. Epílogo

La Fuerza de Trabajo de la IEEE sobre Minería de Procesos (*IEEE Task Force on Process Mining*) tiene como objetivos (a) promover la aplicación de minería de procesos, (b) guiar a desarrolladores de software, consultores, gerentes y usuarios finales en el uso de técnicas en el estado del arte, y (c) estimular la investigación en minería de procesos. Este manifiesto declara los principales principios e intenciones de la fuerza de trabajo. Después de introducir el tópico de minería de procesos, el manifiesto hace un catálogo de algunos principios rectores (Sección 3) y desafíos (Sección 4). Los principios rectores pueden ser utilizado para evitar errores obvios. La lista de desafíos tiene como objetivo dirigir los esfuerzos en investigación y desarrollo. Ambos apuntan a aumentar el nivel de madurez de la minería de procesos.

Para concluir, unas pocas palabras sobre terminología. Los siguientes términos se utilizan en el espacio de la minería de procesos: minería de flujo de tareas (*workflow mining*), minería de procesos (de negocio), descubrimiento automático de procesos (de negocio), e inteligencia de procesos (de negocio). Diferentes organizaciones parecen utilizar diferentes términos para conceptos que se traslapan. Por ejemplo, Gartner está promoviendo el término "Descubrimiento Automático de Procesos de Negocio" (*"Automated Business Process Discovery"*, ABPD), y Software AG está usando "Inteligencia de Procesos"(*"Process Intelligence"*) para referirse a su plataforma de control. El término "minería de flujo de tareas" (*workflow mining*) parece menos apropiado ya que la creación de modelos de flujos de tareas es sólo una de las muchas aplicaciones posibles de minería de procesos. Similarmente, la agregación del término "de negocio" reduce el alcance a ciertas aplicaciones de minería de procesos. Hay numerosas aplicaciones de minería de procesos (e.g., analizar el uso de sistemas de alta tecnología o analizar sitios web) donde esta agregación parece ser inapropiada. Aunque el descubrimiento de procesos es una parte importante del espectro de la minería de procesos, éste es sólo uno de los muchos casos de uso. La verificación de conformidad, la predicción, la minería organizacional, el análisis de redes sociales, etc., son otros casos de uso que se extienden más allá del descubrimiento de procesos.

La Figura 7 relaciona algunos de los términos recién mencionados. Todas las tecnologías y métodos que apuntan a proveer información útil que puede ser utilizada para apoyar la toma de decisiones pueden ser posicionadas bajo el paraguas de la Inteligencia de Negocios (*Business*

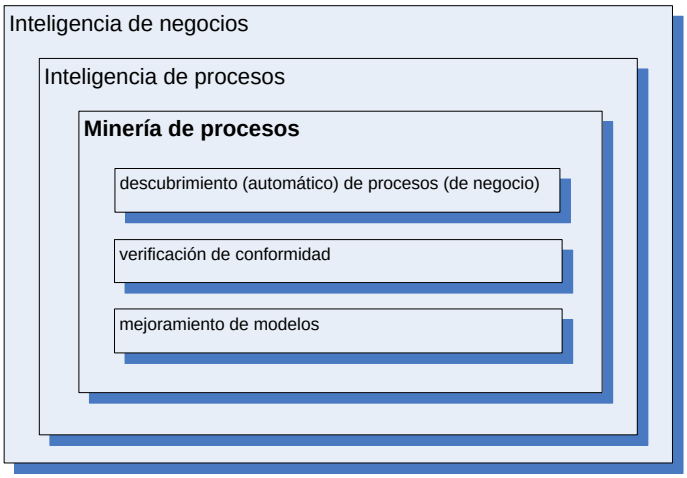


Figura 7. Relacionando los diferentes términos.

Intelligence, BI). La inteligencia de procesos (de negocio) puede ser vista como la combinación de BI y BPM, i.e., se utiliza las técnicas de inteligencia de negocios para analizar y mejorar los procesos y su gestión. La minería de procesos puede ser vista como una concretización de la inteligencia de procesos que toma los registros de eventos como punto de inicio. El descubrimiento (automático) de procesos (de negocio) es sólo uno de los tres tipos básicos de minería de procesos. La Figura 7 puede ser un poco engañosa en el sentido que la mayoría de las herramientas de BI no proveen funcionalidades de minería de procesos, tal como se describe en este documento. El término BI es a menudo convenientemente sesgado hacia una herramienta o método en particular que cubre sólo una pequeña parte del amplio espectro de BI.

Puede haber razones comerciales para usar términos alternativos. Algunos proveedores pueden también querer enfatizar un aspecto en particular (e.g., descubrimiento o inteligencia). Sin embargo, para evitar confusión, es mejor utilizar el témino "minería de procesos" para la disciplina cubierta por este manifiesto.

Glosario

- **Actividad**: es un paso bien definido en el proceso. Los eventos pueden referirse al inicio, conclusión, cancelación, etc., de una actividad para una instancia específica del proceso.
- **Ajuste (*Fitness*)**: es una medida para determinar cuán bien un modelo dado se ajusta al comportamiento observado en el registro de eventos. Un modelo tiene un ajuste perfecto si todas las trazas en el registro de eventos pueden ser reproducidas por el modelo de principio a fin.
- **Caso**: véase **Instancia de un Proceso**.
- **Cambio de Tendencia (*Concept Drift*)**: es el fenómeno en que los procesos suelen cambiar con el tiempo. El proceso observado podría cambiar gradualmente (o de imprevisto) debido a cambios estacionales o al aumento de la competencia, complicando así el análisis.
- **Minería de Datos**: análisis de conjuntos de datos (a menudo grandes) para encontrar relaciones inesperadas y para resumir los datos de manera que proporcionen nuevos entendimientos.
- **Descubrimiento Automático de Procesos de Negocios**: véase **Descubrimiento de Procesos**.
- **Descubrimiento de Procesos**: es uno de los tres tipos básicos de minería de procesos. Basado en un registro de eventos, se crea un modelo de proceso. Por ejemplo, el algoritmo α es capaz de descubrir una red de Petri mediante la identificación de patrones de procesos en colecciones de eventos.

- **Evento**: es una acción almacenada en el registro, por ejemplo, el inicio, conclusión o cancelación de una actividad para una instancia particular de un proceso.
- **Generalización**: es una medida para determinar cuán bien el modelo es capaz de describir comportamiento desconocido. Un modelo con "sobreajuste" no es capaz de generalizar lo suficiente.
- **Gestión de Procesos de Negocio (*Business Process Management*, BPM)**: es la disciplina que combina conocimiento sobre tecnología de información y conocimiento sobre las ciencias de gestión y lo aplica en conjunto a los procesos de negocio operacionales.
- **Inteligencia de Negocios (*Business Intelligence*, BI)**: es una amplia colección de herramientas y métodos que utilizan datos para apoyar la toma de decisiones.
- **Instancia de un Proceso**: es la entidad siendo ejecutada por el proceso que es analizado. Los eventos se refieren a instancias del proceso. Ejemplos de instancias de un proceso son: pedidos de los clientes, reclamos de seguros, solicitudes de préstamos, etc.
- **Inteligencia de Procesos**: es una rama de la Inteligencia de Negocios centrada en la Gestión de Procesos de Negocio.
- **Inteligencia de Procesos de Negocio**: véase **Inteligencia de Procesos**.
- **Minería de Procesos**: son técnicas, herramientas y métodos para descubrir, monitorear y mejorar los procesos reales (es decir, no los procesos supuestos) a través de la extracción de conocimiento de los registros de eventos, ampliamente disponibles en los actuales sistemas de información.
- **Mejoramiento de Modelos**: es uno de los tres tipos básicos de minería de procesos. Un modelo de proceso se extiende o mejora con la información extraída de un registro de eventos. Por ejemplo, se pueden identificar cuellos de botella reproduciendo un registro de eventos en un modelo de proceso, mientras se examinan las marcas de tiempo.
- **MXML**: es un formato basado en XML para el intercambio de registros de eventos. XES reemplaza a MXML como el nuevo formato para minería de procesos no dependiente de la herramienta.
- **Minería de Procesos Inter-Organizacional**: la aplicación de las técnicas de minería de procesos sobre registros de eventos procedentes de diferentes organizaciones.
- **Precisión**: es una medida para determinar si el modelo prohíbe un comportamiento muy diferente al comportamiento observado en el registro de eventos. Un modelo con baja precisión es "subajustado".
- **Registro de Eventos**: es la colección de eventos utilizados como entrada para la minería de procesos. Los eventos no necesitan ser almacenados en un archivo de registro por separado (por ejemplo, los eventos pueden estar dispersos en diferentes tablas de bases de datos).
- **Sesgo Representacional**: es el lenguaje seleccionado para la presentación y construcción de los resultados de la minería de procesos.
- **Simplicidad**: es una medida que pone en práctica el concepto de la Navaja de Occam, i.e ., el modelo más simple que pueda explicar el comportamiento observado en el registro de eventos, es el mejor modelo. La simplicidad se puede cuantificar de distintas maneras, por ejemplo, la cantidad de nodos y arcos en el modelo.
- **Soporte Operacional**: es un análisis en línea de los datos de eventos con el objetivo de supervisar e influir en las instancias del proceso en ejecución. Se pueden identificar tres actividades de soporte operacional: detectar (generar una alerta si el comportamiento observado se desvía del comportamiento modelado), predecir (predecir el comportamiento futuro basado en el comportamiento pasado, e.g., predecir el tiempo de procesamiento restante), y recomendar (sugerir las medidas adecuadas para alcanzar un objetivo concreto, e.g., minimizar costos).
- **Verificación de Conformidad**: analiza si la realidad, según consta en un registro de eventos, se ajusta al modelo y viceversa. El objetivo es detectar las discrepancias y medir su gravedad. La verificación de conformidad es uno de los tres tipos básicos de minería de procesos.
- **XES**: es un estándar XML para los registros de eventos. El estándar ha sido adoptado por la *IEEE Task Force on Process Mining* como el formato de intercambio de registros de eventos por defecto. (cf. www.xes-standard.org).

