

NOVÁTICA

Revista de la Asociación de Técnicos de Informática

Nº 223, mayo-junio 2013, año XXXIX



Minería de procesos

Prácticos

los cambios y las recomendaciones son fácilmente aplicables.

Eficientes y eficaces

logramos máximos resultados con los recursos óptimos

Innovadores

con nuevas técnicas e ideas pioneras en el sector

Rigurosos

exactitud y minuciosidad en el desarrollo de nuestro trabajo.



Aplicamos conocimiento, ilusión y esfuerzo.

La mejora continua llevada a la práctica

Taller práctico de mejora de procesos, estructurado en cuatro fases mediante las cuales obtendremos:

Un **plan de acción detallado** sobre el proceso analizado, el **plan de seguimiento** de las acciones de mejora, el **caso de negocio** para la iniciativa llevada a cabo, y un **equipo de personas formadas** en las principales técnicas de mejora de procesos.

TE OFRECEMOS
UN WORKSHOP
para la Mejora de
PROCESOS

PROCES MINING

LEAN IT

info@gedos.es
<http://www.gedos.es>

G2, Gobierno y gestión de TI



Novática, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de **ATI** (Asociación de Técnicos de Informática), organización que edita también la revista **REICIS** (Revista Española de Innovación, Calidad e Ingeniería del Software).

<<http://www.ati.es/novatica/>>
<<http://www.ati.es/reicis/>>

ATI es miembro fundador de **CEPIS** (Council of European Professional Informatics Societies) y es representante de España en **IFIP** (International Federation for Information Processing); tiene un acuerdo de colaboración con **ACM** (Association for Computing Machinery), así como acuerdos de vinculación o colaboración con **AdaSpain**, **AI2**, **ASTIC**, **RITSI** e **Hispalinux**, junto a la que participa en **ProInnova**.

Consejo Editorial

Ignacio Aguiló Sousa, Guillem Alsina González, María José Escalona Cuaremas, Rafael Fernández Calvo (presidente del Consejo), Jaime Fernández Martínez, Luis Fernández Sanz, Didac Lopez Viñas, Celestino Martín Alonso, José Onofre Montes Andrés, Francesc Noguera Puig, Ignacio Pérez Martínez, Andrés Pérez Payeras, Viktu Pons i Colomer, Juan Carlos Vigo López

Coordinación Editorial

Llorenç Pagés Casas <lpages@ati.es>

Composición y autoedición

Jorge Llácer Gil de Ramales

Traducciones

Grupo de Lengua e Informática de ATI <<http://www.ati.es/gt/lengua-informatica/>>

Administración

Tomás Brunete, María José Fernández, Enric Camarero

Secciones Técnicas - Coordinadores

Acceso y recuperación de la información

José María Gómez Hidalgo (Optenet), <jmgomez@yahoo.es>

Manuel J. María López (Universidad de Huelva), <manuel.maria@diehsia.uhu.es>

Administración Pública electrónica

Francisco López Crespo (MAE), <flc@ati.es>

Sebastià Justicia Pérez (Diputación de Barcelona), <sjusticia@ati.es>

Arquitecturas

Enrique F. Torres Moreno (Universidad de Zaragoza), <enrique.torres@unizar.es>

José Filich Cardo (Universidad Politécnica de Valencia), <jfilich@disca.upv.es>

Auditoría SITIC

Marina Tourino Irolifio, <marinatourino@marinatourino.com>

Sergio Gómez-Landero Pérez (Endesa), <sergio.gomezlandero@endesa.es>

Derecho y tecnologías

Isabel Hernández Collazos (Fac. Derecho de Donostia, UPV), <isabel.hernandez@ehu.es>

Elena Davara Fernández de Marcos (Davara & Davara), <edavara@davara.com>

Enseñanza Universitaria de la Informática

Cristóbal Parja Flores (DSIP-UCM), <cparja@sip.ucm.es>

J. Angel Velázquez Iturbide (DLSI, URJC), <angel.velazquez@urjc.es>

Entorno digital personal

Andrés Marín López (Univ. Carlos III), <amarin@it.uc3m.es>

Diego Gachet Pérez (Universidad Europea de Madrid), <dgachet@uem.es>

Estándares Web

Encarna Quesada Ruiz (Virati), <encarna.quesada@virati.com>

José Carlos del Arco Prieto (TCP, Sistemas e Ingeniería), <jcarco@gmail.com>

Gestión del Conocimiento

Juan Baiget Solé (Cap Gemini Ernst & Young), <juan.baiget@ati.es>

Gobierno Cooperativo de las TI

Manuel Palao García-Suñito (ATI), <manuel@palao.com>

Miguel García-Moneda (ITI), <mgarciamoneda@ititrendsinstitute.org>

Informática y Filosofía

José Ángel Olivás Varela (Escuela Superior de Informática, UCLM), <joseangel.olivas@uclm.es>

Roberto Feliñero Oreja (UNED), <rfeliñero@gmail.com>

Informática Gráfica

Miguel Chover Sellés (Universitat Jaume I de Castellón), <mchover@lsi.uji.es>

Roberto Vivó Hernando (Eurographics, sección española), <rvido@dsic.upv.es>

Ingeniería del Software

Javier Dolado Cosin (DLSI-UPV), <adolado@lsi.uhu.es>

Daniel Rodríguez García (Universidad de Alcalá), <daniel.rodriguez@uah.es>

Inteligencia Artificial

Vicente Boti Navarro, Vicente Julián Inglada (DSIC-UPV), <[vbotti.vinglada@dsic.upv.es](mailto:(vbotti.vinglada)@dsic.upv.es)>

Interacción Persona-Computador

Pedro M. Latorre Andrés (Universidad de Zaragoza, AIPO), <platorre@unizar.es>

Francisco L. Gutierrez Vela (Universidad de Granada, AIPO), <fgutierrez@ugr.es>

Lenguaje e Informática

M. del Carmen Ugarte García (ATI), <cugarte@ati.es>

Lenguajes Informáticos

Oscar Belmonte Fernández (Univ. Jaime I de Castellón), <obelmonte@lsi.uji.es>

Inmaculada Coma Taty (Univ. de Valencia), <inmaculada.coma@uv.es>

Lingüística computacional

Xavier Gómez Guzmán (Univ. de Vigo), <xgg@uvigo.es>

Manuel Palomar (Univ. de Alicante), <mpalomar@lsi.ua.es>

Mundo estudiantil y jóvenes profesionales

Federico G. Mon Trotti (RITSI), <gmon.trotti@gmail.com>

Mikel Sáizar Páez (Asoc. de Jóvenes Profesionales, Junta de ATI Madrid), <mikelbo_uni@yahoo.es>

Profesión Informática

Rafael Fernández Calvo (ATI), <rfcalvo@ati.es>

Miguel Sarrías Gilardi (ATI), <miguel@sarrias.net>

Redes y servicios telemáticos

José Luis Marzo Lázaro (Univ. de Girona), <joseluis.marzo@udg.es>

Juan Carlos López López (UCLM), <juancarlos.lopez@uclm.es>

Robótica

José Cortés Arenas (Sopra Group), <joscorare@gmail.com>

Juan González Gómez (Universidad CARLOS III), <juang@iearobotics.com>

Seguridad

Javier Areñio Bertollin (Univ. de Deusto), <jaareñio@deusto.es>

Javier López Muñoz (ETSI Informática-UMA), <jlmm@lcc.uma.es>

Sistemas de Tiempo Real

Alejandro Alonso Muñoz, Juan Antonio de la Puente Alfaro (DIT-UPM), <caalmon@puente@dit.upm.es>

Software Libre

Jesús M. González Barahona (GSYC - URJC), <jgb@gsyc.es>

Israel Herráiz Tabernero (Universidad Politécnica de Madrid), <isra@herraz.org>

Tecnología de Objetos

Jesús García Molina (DIS-UM), <jmolina@um.es>

Gustavo Rossi (LIFIA-UNLP Argentina), <gustavo@sol.info.unlp.edu.ar>

Tecnologías para la Educación

Federico G. Mon Trotti (RITSI), <gmon.trotti@gmail.com>

César Pablo Córcoles Briónigo (UOC), <ccorcoles@uoc.edu>

Tecnologías y Empresa

Didac Lopez Viñas (Universitat de Girona), <didac.lopez@ati.es>

Alonso Álvarez García (TID), <aag@tid.es>

Tendencias tecnológicas

Gabriel Martí Fuentes (Interbits), <gabi@atinet.es>

Juan Carlos Vigo (ATI), <juancarlosvigo@atinet.es>

TIC y Turismo

Andrés Aguayo Maldonado, Antonio Guevara Plaza (Univ. de Málaga), <[aguayo.guevara@lcc.uma.es](mailto:(aguayo.guevara)@lcc.uma.es)>

Las opiniones expresadas por los autores son responsabilidad exclusiva de los mismos.

Novática permite la reproducción, sin ánimo de lucro, de todos los artículos, a menos que lo impida la modalidad de © o copyright elegida por el autor, debiéndose en todo caso citar su procedencia y enviar a **Novática** un ejemplar de la publicación.

Coordinación Editorial, Redacción Central y Redacción ATI Madrid

Plaza de España 6, 2ª planta, 28008 Madrid

Tfno. 91 4029391; fax. 91 3093685 <novatica@ati.es>

Composición, Edición y Redacción ATI Valencia

Av. del Reino de Valencia 29, 46005 Valencia

Tfno. 963740173 <novatica_prod@ati.es>

Administración y Redacción ATI Cataluña

Via Laietana 46, ppal. 1º, 08003 Barcelona

Tfno. 934125235; fax. 934127713 <secretgen@ati.es>

Redacción ATI Andalucía

Av. de la Constitución 10, 41013 Sevilla

Redacción ATI Galicia

Av. de la Constitución 10, 15100 Santiago de Compostela

Suscripción y Ventas

<novatica.subscriptions@atinet.es>

Publicidad

Plaza de España 6, 2ª planta, 28008 Madrid

Tfno. 91 4029391; fax. 91 3093685 <novatica@ati.es>

Imprenta: Derra S.A. Juan de Austria 86, 08005 Barcelona.

Depósito legal: B 15.154-1975 - ISSN: 0211-2124, CODEN NOVACD

Portada: "Mineral, vegetal, animal" - Concha Arias Pérez / © ATI

Diseño: Fernando Agresta / © ATI 2003

editorial

El "caso Snowden" y la seguridad de las redes de telecomunicación en resumen > 02

Soporte al negocio y práctica profesional: El sueño del buen editor > 03

Llorenç Pagés Casas

noticias de IFIP

IFIP TC6 Latin American Tutorials in Networking (LATIN 2013) > 03

Ramon Puigjaner Trepal

monografía

Minería de procesos

Editores invitados: Antonio Valle Salas y Anne Rozinat

Presentación. Una perspectiva sobre la minería de procesos > 05

Antonio Valle Salas, Anne Rozinat

Minería de procesos: La objetivación de la intuición en los procesos de toma de decisiones en los negocios, más transparentes gracias al análisis de los datos > 07

Anne Rozinat, Wil van der Aalst

Minería de procesos: Obtenga una radiografía de sus procesos de negocio > 11

Wil van der Aalst

El viaje del descubrimiento de procesos > 20

Josep Carmona Vargas

Posibilidades de uso de la minería de procesos en ITSM > 24

Antonio Valle Salas

Optimización dirigida por minería de procesos de un proceso de aprobación de préstamos al consumo > 31

Arjel Bautista, Lalit Wangikar, S.M. Kurnail Akbar

Mejoramiento de procesos con técnicas de minería de procesos, simulación y optimización: Caso de estudio > 41

Santiago Aguirre Mayorga, Carlos Alberto Parra Rodríguez

Detección de cambios temporales en los procesos de negocio mediante el uso de técnicas de segmentación > 49

Daniela Lorena Luengo Mundaca, Marcos Sepúlveda Fernández

secciones técnicas

Referencias autorizadas

visiones sobre Lenguajes de Programación

Cómo la metáfora de objetos llegó a ser el principal paradigma de programación > 62

Jesús J. García Molina

Elección de lenguajes de programación para la enseñanza universitaria > 67

Baltasar García Pérez-Schofield

La importancia de la labor del programador. ¿Qué se espera? ¿Cómo se prepara? > 70

Análisis desde los lenguajes de programación

Óscar Belmonte Fernández, Carlos Granell Canut

Para pensar > 79

Rafael Martínez Martínez

Programando caminos y resolviendo necesidades > 81

Alejandro Fuentes Penna

sociedad de la información

Programar es crear

El problema del CUIT (corrección del publicado en el número anterior) (Competencia UTN-FRC 2012, problema D, enunciado) > 82

Julio Javier Castillo, Diego Javier Serrano, Marina Elizabeth Cárdenas

Asuntos Interiores

Coordinación editorial / Programación de Novática / Socios Institucionales > 83

Tema del próximo número:

"Eficiencia energética en centros de proceso de datos"

Antonio Valle Salas¹, Anne Rozinat²

¹Socio Director en G2, Gobierno y Gestión de TI, SL; Director del Comité de Catalunya en itSMF España; ²Socia cofundadora de Fluxicon;

<avalle@gedos.es>, <anne@fluxicon.com>

Durante las últimas décadas, la Informática ha llegado a todos los rincones de nuestra vida. Desde los teléfonos móviles hasta los más avanzados sistemas de procesamiento de información médica, pasando por expendedoras de café o PLCs en líneas de producción. Hay componentes informatizados en casi cualquier lugar donde miremos. Y todos estos componentes generan cantidades ingentes de información que crece de forma exponencial con el paso del tiempo. Hace relativamente pocos años, el problema era conseguir información digitalizada, mientras que en estos momentos el problema se ha convertido en conseguir procesar y darle significado a toda la información que generamos.

En los últimos años hemos visto cómo desde la industria del análisis de información se aportaban diversas aproximaciones a este problema que han sido abordadas de una u otra manera en ediciones anteriores de **Novática**: se comenzó con las VLDB (*Very Large Databases*) en el número 91 de 1991, seguimos con las aproximaciones de *Datawarehouse* y el intento de descubrir patrones dentro de estos almacenes de datos con *Data Mining*, tratado en el número 138 de 1999. Lo intentamos más tarde con la *Gestión del Conocimiento* en **Novática 155** en 2002, nos dimos cuenta de cómo se podía llegar a complicar el problema con el monográfico del 209 sobre *Internet de las Cosas* en 2011 y de cómo podíamos explotar toda esta información en el número 211 sobre *Business Intelligence* en ese mismo año. Finalmente, la industria está avanzando también en una dirección que **Novática** aún no ha cubierto pero que con certeza será tratada en un futuro próximo: *Big Data*.

En esta monografía abordamos una temática especialmente interesante dentro de este amplio abanico de técnicas para el análisis de la información: la *Minería de Procesos* es una variante del *Data Mining* en la que ponemos el foco en el análisis de la información generada por los procesos que han sido informatizados y de los cuales tenemos una traza de su ejecución.

Tal y como nos explican **Anne Rozinat** y **Wil van der Aalst** en el artículo de apertura, veremos que las primeras trazas se encuentran a finales del siglo XIX, a pesar de que en términos de ciencia moderna debemos

Presentación Una perspectiva sobre la minería de procesos

Editores invitados

Antonio Valle Salas es Ingeniero Técnico en Informática de Gestión por la UPC (*Universitat Politècnica de Catalunya*) y posee tanto certificaciones metodológicas como *Certified Lean Service Professional* por la *Service Management Society*, *ITIL Service Manager* por EXIN, Auditor Certificado de Sistemas de Información (CISA) por ISACA y *COBIT Based IT Governance Foundations* por la *IT Governance Network* como certificaciones eminentemente técnicas en las herramientas de gestión de la familia HP Openview. Es Director Regional de Cataluña del itSMF España, combinando las actividades dentro del mundo de la consultoría y de la ejecución de proyectos con colaboraciones frecuentes tanto en actividades formativas en entornos universitarios como la UPC o la *Universitat Pompeu Fabra* como en el mundo editorial, donde ha colaborado en publicaciones como *IT Governance a pocket guide*, *Metrics in IT Service Organizations*, *Gestión de Servicios Una introducción a ITIL*, así como las traducciones al castellano de los libros *Soporte al Servicio* y *Provisión de Servicios ITIL V2* o en la edición en castellano de *Gestión Esencial de Servicios*.

Anne Rozinat cuenta con más de ocho años de experiencia en tecnología de minería de procesos y obtuvo el doctorado *cum laude* dentro del grupo de minería de procesos del profesor Wil van der Aalst en la Universidad Técnica de Eindhoven (Países Bajos). En la actualidad es socia cofundadora de Fluxicon y bloguera en <<http://www.fluxicon.com/blog/>>.

referirnos a los trabajos seminales de Myhill/Nerod de 1958 o a los algoritmos de Viterbi de 1978.

A finales de los años 90 había ya equipos de investigación en universidades de todo el mundo, destacando la *University of Colorado* o la *Technische Universiteit Eindhoven* (TU/e). Estos equipos desarrollaron su actividad investigadora definiendo algoritmos y métodos que permitían el tratamiento de las trazas de ejecución de procesos para realizar descubrimiento, análisis y representación de los procesos subyacentes. Pero, en aquel entonces, aún no se habían desarrollado herramientas que implementaran estos algoritmos con unos grados de usabilidad adecuados para llegar al mercado.

Así llegamos a finales del año 2003 cuando se crea la comunidad especializada *processmining.org* (un grupo de trabajo de la TU/e) y a principios de 2004 cuando aparece la primera versión de ProM, un *framework* genérico y de código abierto para la minería de procesos que se ha convertido en la principal herramienta de investigadores y analistas, y que en la actualidad está en su versión 6.3 y dispone de más de 500 *plugins* que implementan el estado del arte en este terreno.

Además, durante 2009 se creó una *Task Force* del IEEE focalizada en la minería de procesos que cuenta con miembros de más de 20 países y que representa tanto a fabricantes de software (como Software AG, HP, IBM o

Fluxicon, entre muchos otros), firmas de consultoría y análisis (como Process Sphere, Gartner o Deloitte, entre otros), y una amplia selección de instituciones educativas y de investigación (TU/e, *Universitat Politècnica de Catalunya* o *Universität zu Berlin* por citar únicamente a unas pocas). Uno de los objetivos clave de esta *task force* es la difusión de los conceptos, técnicas y beneficios que aporta la minería de procesos y para ello se publicó en 2011 el *Manifiesto por la Minería de Procesos*, documento firmado por más de 50 profesionales del sector, que ha sido traducido a 12 idiomas. No hemos podido incorporar el texto íntegro del manifiesto por motivos de espacio, pero el lector encontrará la referencia en la sección de enlaces útiles de esta monografía.

Para esta edición de **Novática** hemos tenido el privilegio de contar con un conjunto de autores que nos aportan diferentes perspectivas sobre el asunto.

Comenzamos con un artículo introductorio, en el que **Anne Rozinat** y **Wil van der Aalst** nos contextualizan los conceptos de minería de procesos y nos proponen ese mensaje tan esclarecedor de que la minería de procesos nos permite objetivar la visión que tenemos de nuestros procesos.

A continuación, **Wil van der Aalst** nos guía en el segundo artículo por las diferentes utilidades que podemos hacer de la minería de procesos: la creación de un modelo, la

comprobación del grado de cumplimiento del modelo o la mejora de un modelo ya existente. Aquí aparece otro mensaje fundamental: el uso de la minería de procesos como los Rayos X que nos permiten ver el proceso "por dentro", basándonos en los datos reales de su ejecución y analizando la totalidad de casos (en contraposición al muestreo estadístico que haríamos en una actividad de auditoría, por ejemplo).

En el siguiente artículo encontraremos la visión que nos propone **Josep Carmona Vargas** sobre la tarea de descubrimiento de un proceso a partir de las trazas. Aquí, Josep hace una aproximación entretenida a cómo podríamos utilizar la minería de procesos para descifrar el mensaje que nos da un extraterrestre para explicar su visita a la Tierra al tiempo que nos muestra la anatomía del proceso de descubrimiento.

Los artículos de introducción darán paso a un bloque de artículos centrados en casos prácticos: primero **Antonio Valle Salas** nos plantea un artículo sectorial, enfocado a los procesos propios de un Departamento de Informática y nos muestra los diferentes usos que podemos hacer de estas técnicas en el mundo de la Gestión de Servicios IT (ITSM)

A continuación, **Arjel Bautista, Lalit Wangikar** y **Syed M. Kumail Akbar** nos presentan el trabajo realizado para optimizar el proceso de aprobación de préstamos de una institución bancaria holandesa, siendo destacable que este trabajo le valió al equipo ganar el premio *BPI Challenge* 2012.

Siguiendo en la línea de los artículos que muestran casos reales de aplicación, **Santiago Aguirre Mayorga** y **Carlos**

Alberto Parra Rodríguez nos presentan un caso de uso que hace uso no sólo de la minería de procesos en el diagnóstico, sino también de técnicas de simulación para evaluar las posibles respuestas a las oportunidades de mejora encontradas en el proceso de aprobación de compra de bienes de una institución universitaria.

Finalmente, **Daniela Lorena Luengo Mundaca** y **Marcos Sepúlveda Fernández** nos dan una visión desde el punto de vista de la investigación con un artículo en el que se aborda uno de los retos planteados en el manifiesto: *lidiar con el cambio de tendencia*.

El término "cambio de tendencia" (*concept drift*) se utiliza para referirse a la situación en la que el proceso está cambiando mientras está siendo analizado; la detección de estos cambios y la inclusión de estas características en el análisis es fundamental cuando estamos trabajando en entornos que cambian rápidamente ya que, de lo contrario, podemos llegar a conclusiones erróneas en nuestros análisis.

Estos autores han contribuido con sus artículos a que tengamos una visión más clara sobre qué es, para qué sirve y hacia dónde va la minería de procesos. Es ésta una ciencia relativamente nueva pero que ya está llegando al nivel de madurez necesario para que se convierta en práctica habitual en las empresas y organizaciones. Así lo reflejan los artículos de carácter práctico que aquí publicamos.

A pesar de los retos que tenemos por delante, queda claro que hay mucho camino por recorrer: ¿Seremos capaces de superar las problemáticas que plantea el *concept drift*? ¿Podremos utilizar la minería no sólo para

conocer el pasado de un proceso sino también para predecir su futuro? ¿Conseguiremos implementar estas técnicas en los sistemas de gestión de procesos empresariales con el fin de dotarlos de sistemas predictivos o de apoyo al operador? Estamos seguro de que veremos grandes avances en este área en un futuro próximo.

Referencias útiles sobre "Minería de procesos"

Las referencias que se citan a continuación, junto con las proporcionadas en cada uno de los artículos, tienen como objetivo ayudar a los lectores a profundizar en los temas tratados en esta monografía permitiendo contrastar ideas y obtener información actualizada.

Publicaciones

- **W.M.P. van der Aalst.** *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer Verlag, 2011. ISBN 978-3-642-19344-6.
- **IEEE Task Force on Process Mining.** *Process Mining Manifesto* (en 12 idiomas). <<http://www.win.tue.nl/ieeetfpm/>>

doku.php?id=shared:process_mining_manifesto>.

Enlaces web

- **University of Wisconsin-Madison.** *Introduction to Process Mining: turning (big) data into value* (video). <http://www.youtube.com/watch?v=7oat7MatU_U>.
- **Fluxicon.** *Process Mining News*. <<http://fluxicon.com/s/newsarchive>>.
- **Grupo de Trabajo TU/e.** <<http://www.processmining.org>>.
- **Fluxicon.** *Flux Capacitor* <<http://fluxicon.com/blog/>>.
- **IEEE Task Force on Process Mining.** <<http://www.win.tue.nl/ieeetfpm/doku.php?id=start>>.

■ **LinkedIn.** *Process Mining* (comunidad) <<http://www.linkedin.com/groups/Process-Mining-1915049>>.

■ **TU/e.** *Health Analytics Using Process Mining*. <<http://www.healthcare-analytics-process-mining.org>>.

Anne Rozinat¹, Wil van der Aalst²

¹Socia cofundadora de Fluxicon; ²Universidad Técnica de Eindhoven (Países Bajos)

<anne@fluxicon.com>, <w.m.p.v.d.aalst@tue.nl>

1. Introducción

El archivo del Observatorio Naval de la Marina de los Estados Unidos guarda todos los cuadernos de bitácora de la Armada estadounidense del siglo XIX. Estos cuadernos contenían anotaciones diarias relativas a la posición, vientos, corrientes y otros detalles de miles de viajes realizados en barco. Nadie había hecho nunca nada con estos cuadernos, y se había propuesto que podrían desecharse hasta que Matthew Fontain Maury apareció en escena.

Maury (ver **figura 1**) fue un marino de la Armada estadounidense que desde 1842 dirigió el Observatorio Naval de los Estados Unidos. Evaluó sistemáticamente esos datos y creó libros ilustrados que cartografiaron los vientos y las corrientes de los océanos, mostrándolos visualmente, de tal modo que ayudaran a los capitanes de los navíos en la toma de decisiones a la hora de planificar las rutas.

En 1848 el capitán Jackson de la W. H. D. C. Wright fue el primero en utilizar los libros de Maury en un viaje de Baltimore a Río de Janeiro volviendo un mes antes de lo planeado. Tan solo siete años después de la primera edición de *Direcciones de navegación* de Maury, la industria de la navegación había ahorrado alrededor de 10 millones de dólares al año [1].

Los sistemas TI en el mundo de los negocios esconden también datos de gran valor, que a menudo no llegan a utilizarse nunca. Los procesos de negocio crean los equivalentes modernos a los "apuntes en los cuadernos de bitácora", que detallan exactamente qué actividades se llevan a cabo por quién y cómo (ver **figura 2**).

Si, por ejemplo, un proceso de compra empieza con un sistema SAP, cada paso en el proceso se indica en las correspondientes tablas SAP. De la misma manera, sistemas

¹ **Nota de Traducción:** Aunque se han realizado algunos intentos de encontrar una equivalencia en castellano para *Big Data*—macrodatos, megadatos...—ninguno de ellos consigue convencernos ni encontramos la consiguiente implantación por lo que preferimos mantener el término en inglés, tal como se emplea normalmente entre los profesionales del sector. Una definición de *Big Data* puede encontrarse en la *Wikipedia*. <http://es.wikipedia.org/wiki/Big_Data>, [consulta: 29/08/2013].

Minería de procesos: La objetivación de la intuición en los procesos de toma de decisiones en los negocios, más transparentes gracias al análisis de los datos

Traducción: María del Carmen Ugarte García (Grupo de Trabajo de Lengua e Informática de ATI)

Resumen: Big Data ya existía en el siglo XIX. Al menos esta sería la conclusión a la que podríamos llegar tras repasar la historia de Matthew Maury. Estableceremos un paralelo con las primeras evaluaciones sistemáticas de los cuadernos de bitácora y mostraremos cómo rápida y objetivamente se pueden establecer correspondencias entre los procesos basados en la evaluación de los cuadernos de bitácora y los sistemas TI.

Palabras clave: Análisis de datos, Big Data, casos de estudio, descubrimiento automático de procesos de negocio, introducción, minería de procesos.

Autores

Anne Rozinat cuenta con más de ocho años de experiencia en tecnología de minería de procesos y obtuvo el doctorado *cum laude* dentro del grupo de minería de procesos del profesor Wil van der Aalst en la Universidad Técnica de Eindhoven (Países Bajos). En la actualidad es socia cofundadora de Fluxicon y bloguera en <<http://www.fluxicon.com/blog/>>.

Wil van der Aalst es profesor en la Universidad Técnica de Eindhoven y cuenta con un índice h de más de 90 puntos entre los científicos más citados en Europa. Se le conoce sobre todo por sus trabajos en modelos de Workflow y se le considera sobradamente como el "padrino" de la minería de procesos. Su web personal se encuentra en <www.vdaalst.com>.

CRM (*Customer Relationship Management*), sistemas de etiquetado e incluso sistemas patrimoniales registran datos históricos sobre los procesos.

Estas trazas digitales son los productos secundarios del aumento de la automatización y la presencia de TI en los procesos de negocio [2].

2. Del muestreo al análisis completo

Antes de que existiera el manual de Maury sobre corrientes y mareas, los marinos se veían obligados a planificar sus rutas según su propia experiencia. Este es también el caso para la mayoría de los procesos de negocio: nadie tiene realmente una visión clara de cómo se ejecutan realmente esos procesos. En su lugar hay anécdotas, intuición y muchas opiniones subjetivas (potencialmente contradictorias) que deben reconciliarse.

El análisis sistemático de las trazas en los registros digitales a través de las llamadas técnicas de minería de procesos [3] ofrece un potencial enorme para todas las organizaciones, puesto a prueba en procesos complejos. A través del análisis de la secuencia de eventos

y sus marcas de tiempo, los procesos reales pueden ser plena y objetivamente reconstruidos y descubiertas sus debilidades. La información en los registros TI puede usarse automáticamente para generar modelos de procesos, que pueden enriquecerse más tarde

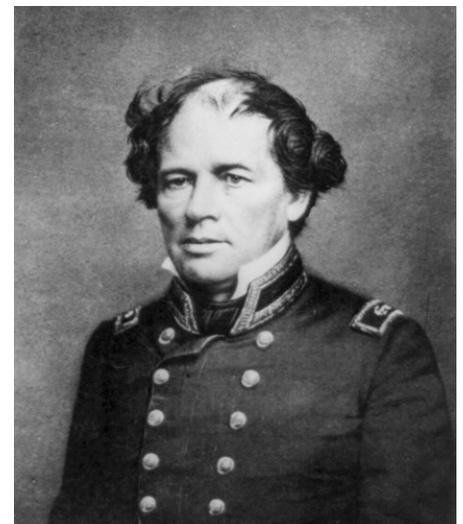
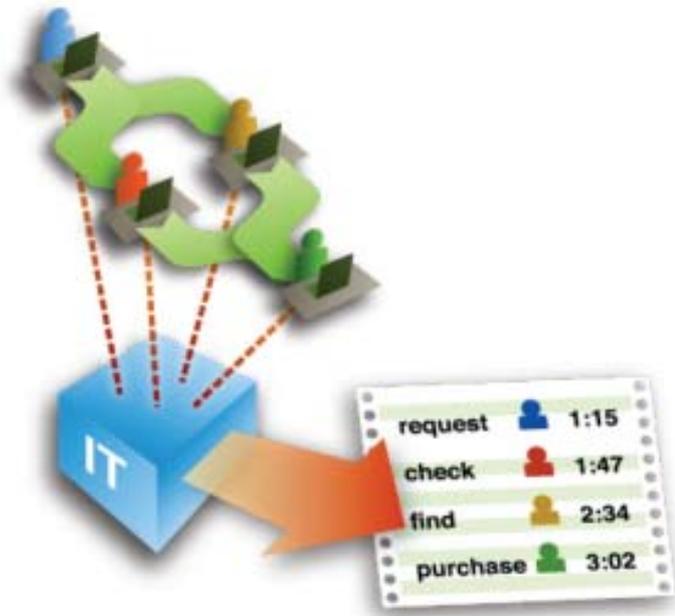


Figura 1. Matthew Fontaine Maury (Fuente: Wikipedia).

“ El descubrimiento manual del proceso a través de los clásicos talleres y entrevistas es muy costoso en tiempo y dinero, además de incompleto y subjetivo ”



mediante métricas de proceso también extraídas directamente de los datos de los registros (por ejemplo, tiempos de ejecución y tiempos de espera).

Las cuestiones típicas de la minería de procesos que pueden plantearse son:

- ¿Cómo son realmente mis procesos?
- ¿Dónde están los cuellos de botella?
- ¿Se producen desviaciones de los procesos prescritos o descritos?

A fin de optimizar un proceso, en primer lugar se debe comprender la realidad del mismo, el proceso tal cual es. Y esto, normalmente, no es una tarea sencilla porque en los procesos de negocio están implicadas numerosas personas y a menudo se distribuyen entre diferentes unidades organizativas o incluso compañías.

Cada uno ve solo una parte del proceso. El descubrimiento manual del proceso a través de los clásicos talleres y entrevistas es muy

Figura 2. En los procesos de TI se registran en detalle qué actividades se realizan, cuándo y por quién.

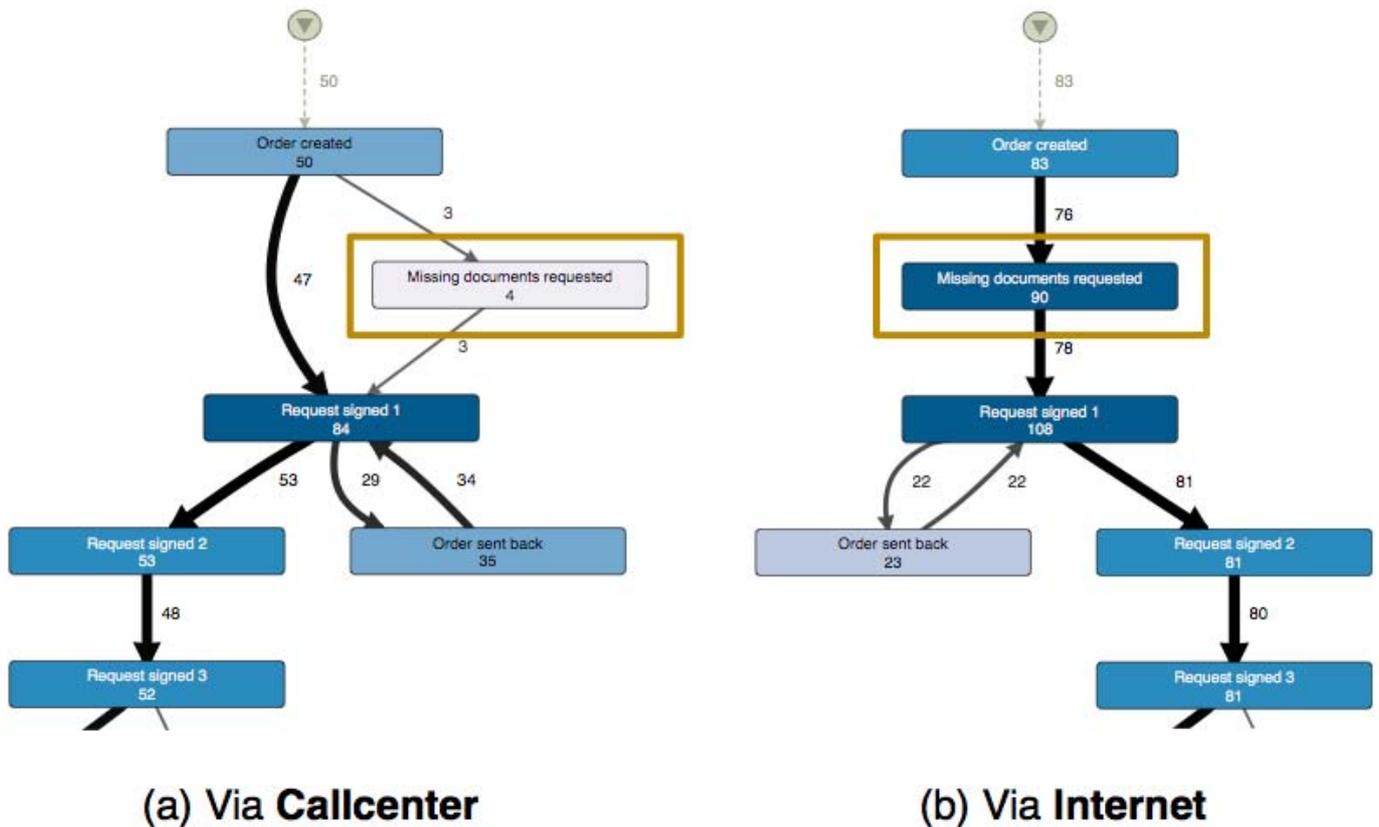


Figura 3. Visualización del proceso de devolución de casos que empiezan vía centro de llamadas (a) y por la vía del portal de Internet (b). En el caso de los casos abiertos vía Internet hay que pedir a menudo información complementaria. En los casos abiertos vía centro de llamadas ese problema no existe.

“ Como Maury hizo con los cuadernos de bitácora navales, se pueden derivar mapas de proceso objetivos que muestren cómo los procesos funcionan de verdad en el mundo real ”

costoso en tiempo y dinero, además de incompleto y subjetivo. Con las herramientas de la minería de procesos es posible hacer uso de los datos de TI existentes de los sistemas operacionales para visualizar rápida y objetivamente los procesos tal cual son y tal como están llevándose a cabo.

En talleres con depositarios de procesos, se puede, entonces, poner el foco en el análisis de las causas principales y las actividades de más valor añadido de dichos procesos.

3. Estudio de un caso concreto

En uno de nuestros proyectos hemos analizado un proceso de devolución dentro de un gran fabricante de productos electrónicos. La descripción del proceso siguiente ha sido ligeramente retocada para salvaguardar la identidad del fabricante. El punto de partida del

proyecto fue la sensación por parte del director de procesos de que este presentaba problemas graves. Las quejas de los clientes y el estudio de casos individuales indicaban que había ineficiencias y tiempos de ejecución demasiado largos en el proceso.

El proyecto fue llevado a cabo en las siguientes fases: Primero se recolectaron cuestiones y problemas concretos y se extrajeron los registros de TI de la correspondiente plataforma de servicio de todos los casos del año en curso. Se analizaron a continuación los datos de los registros junto con los directores de procesos en un taller interactivo.

Por ejemplo, en la figura 3 se puede ver un fragmento simplificado del principio del proceso de devolución. En el lado izquierdo (a) está el proceso para todos los casos que se

inician vía centro de llamadas. En el lado derecho (b) se puede ver el mismo fragmento del proceso para todos los casos que se inician a través del portal de Internet del fabricante. Ambas visualizaciones del proceso fueron construidas automáticamente utilizando el software de minería de procesos de Fluxicon, Disco, basado en los registros de TI que se han extraído.

Los números, el espesor de los arcos, y el color ilustran cuán frecuentemente cada actividad o camino se ha llevado a cabo. Por ejemplo, la visualización de los procesos iniciados a través del centro de llamadas se basa en 50 casos (ver la parte izquierda en la figura 3). Cada uno de los 50 casos empieza con la actividad de crear pedido. Después, la petición se aprueba inmediatamente en 47 casos. En 3 casos se ha de pedir más información al

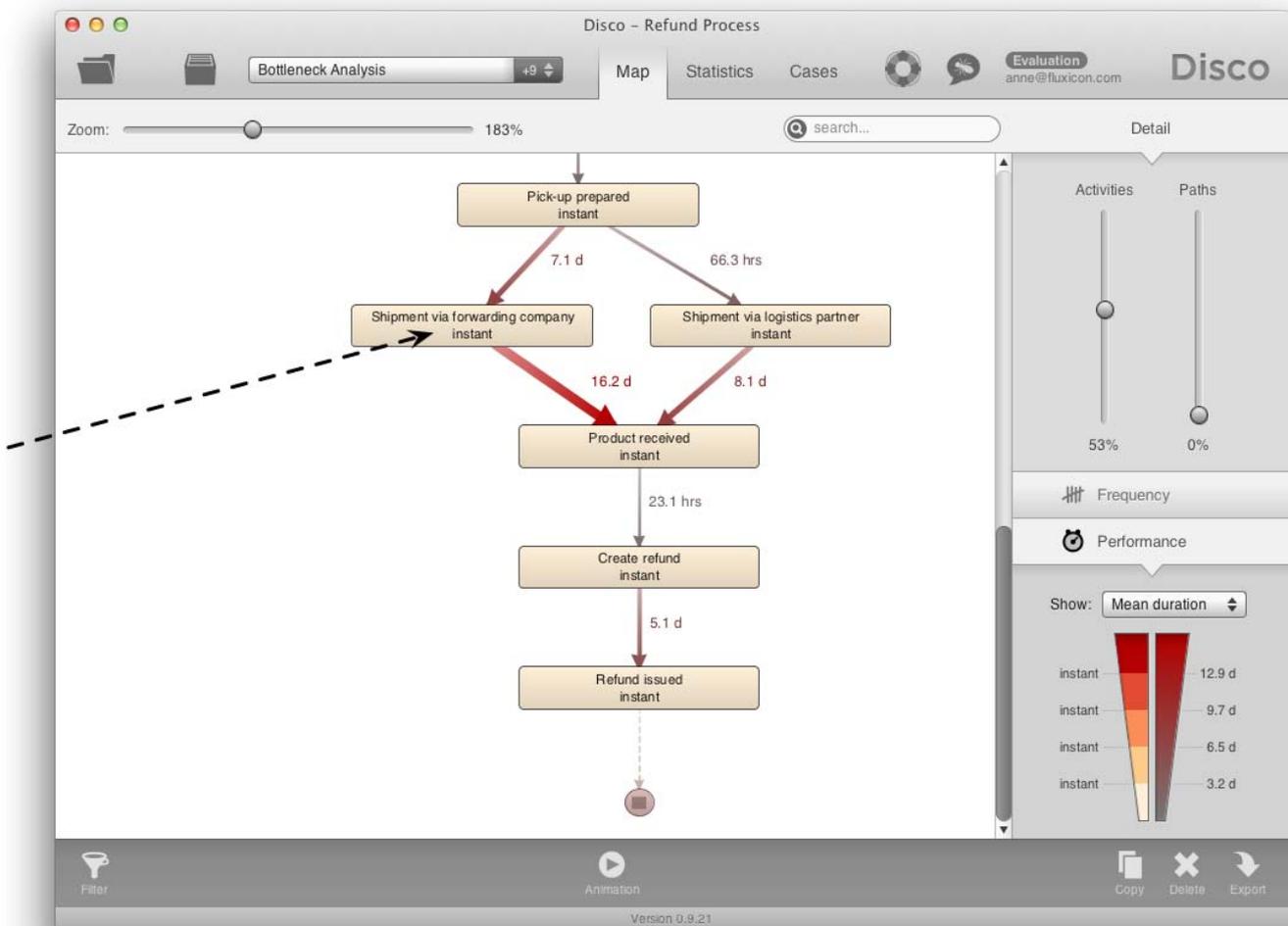


Figura 4. Pantallazo del software Disco de minería de procesos mostrando la vista del análisis del proceso. Resulta evidente que el reenvío a través de mensajería produce cuellos de botella.

cliente. Por simplicidad, solo el flujo del proceso principal se muestra aquí.

Lo que parece evidente en la **figura 3** es que, a pesar de que debe pedirse solo esporádicamente más información al cliente, esto ocurre en un montón de casos cuando los pedidos se hacen vía el portal de Internet: En el 97% de los casos analizados (77 de los 83 casos contemplados) se realiza este paso adicional en el proceso. En 12 de los 83 casos analizados (aprox. 14%) esto ocurre incluso en múltiples ocasiones (en total 90 veces para 83 casos). Este paso en el proceso es muy costoso en tiempo porque requiere una llamada o un correo electrónico por parte del proveedor de servicio. Adicionalmente, a través de la comunicación externa el proceso se retrasa para el cliente, que en un proceso de devolución ha tenido ya una mala experiencia. Por lo tanto, el problema necesita resolverse. Mediante una mejora en el portal de Internet (en lo que respecta a los datos obligatorios en el formulario que debe rellenarse para la petición de reembolso) puede prevenirse esta falta de esta información cuando se inicia el proceso.

Otro resultado del análisis es la detección de cuellos de botella detectados en las recogidas del producto llevadas a cabo por una empresa de mensajería. El fragmento de proceso de la **figura 4** muestra la media de tiempo de espera entre los pasos del proceso basándose en las marcas de tiempo de los datos históricos.

Igualmente los análisis de estos tiempos de espera se crean automáticamente mediante un software de minería de procesos. Se puede ver que antes y después del paso "Envío" por parte de la mensajería pasa un montón de tiempo. Por ejemplo, hay una media de casi 16 días entre el "Envío" y la "Recepción" del producto. La causa principal de los amplios tiempos de espera es que la empresa descubre que los productos son puestos en un palé y hasta que este no está completo no es reexpedido, lo cual lleva a retrasos, en especial para aquellos productos que se han colocado en el palé cuando este estaba casi vacío. Igualmente el proceso de devolución real en el lado de la empresa de productos electrónicos lleva también demasiado tiempo (una media de aproximadamente 5 días). Para el cliente, el proceso únicamente se ha completado en el momento de recibir el dinero.

Como resultado del análisis de la minería de procesos, se detectan las desviaciones respecto al proceso requerido. Es posible comparar los datos de los registros (y por lo tanto el proceso real) objetiva y completamente con respecto a las reglas de negocio requeridas, y aislar los casos que muestran desviaciones. Específicamente, hemos encontrado que (1) en un caso el cliente ha recibido el reembolso dos veces, (2) que en dos casos el dinero fue

reembolsado sin que la empresa manufacturera hubiera recibido el producto defectuoso, y (3) que en unos pocos casos se habían saltado el paso importante y obligatorio de la aprobación.

4. Lo último

La minería de procesos es todavía una disciplina joven y relativamente desconocida, que está accediendo al mercado a través de las primeras herramientas de software profesional, siendo respaldada por la publicación de casos particulares [4].

El Equipo de Trabajo sobre Minería de Procesos de la IEEE (*IEEE Task Force on Process Mining*) [5] se fundó en 2009 para incrementar la visibilidad de la minería de procesos. En otoño del 2011, se publicó un Manifiesto de minería de procesos (*Process Mining Manifesto*) [6], que está disponible en 13 idiomas.

Las empresas generan ya vastas cantidades de datos como un subproducto de sus procesos TI. Estos datos pueden ser analizados directamente mediante las herramientas de minería de procesos. Como Maury hizo con los cuadernos de bitácora navales, se pueden derivar mapas de proceso objetivos que muestren cómo los procesos funcionan de verdad en el mundo real [7]. Los desarrollos en el campo de Big Data¹ están ayudando al almacenamiento para posterior acceso a estos datos a fin de analizarlos de forma efectiva.

Los libros sobre corrientes y vientos de Matthew Fontaine Maury fueron tan útiles a mediados del siglo XIX, que incluso fueron utilizados de forma compulsiva por las aseguradoras [8] a fin de prevenir accidentes marítimos y garantizar la navegación sin problemas. De la misma forma, en el análisis y la optimización de los procesos de negocio llegará un momento, cuando menos lo imaginemos, en el que podremos incluso dejarlos a un lado y confiar únicamente en nuestra intuición.

Referencias

- [1] **Tim Zimmermann.** *The Race: Extreme Sailing and Its Ultimate Event: Nonstop, Round-the-World, No Holds Barred.* Mariner Books, 2004. ISBN-10: 0618382704.
- [2] **W. Brian Arthur.** *The Second Economy.* McKinsey Quarterly, 2011.
- [3] **Wil M.P. van der Aalst.** *Process Mining: Discovery, Conformance and Enhancement of Business Processes.* Springer-Verlag, 2011. ISBN-10: 3642193447.
- [4] **Alberto Manuel.** *Process Mining - Ana Aeroportos de Portugal,* 2012. BPTrends, <www.bptrends.com>.
- [5] **IEEE Task Force on Process Mining.** <<http://www.win.tue.nl/ieeetfpm/>>.
- [6] **IEEE Task Force on Process Mining.** *Process Mining Manifesto.* Business Process Management Workshops 2011, *Lecture Notes in Business Information Processing*, Vol. 99, Springer-Verlag, 2011.
- [7] **Anne Rozinat.** *How to Reduce Waste With Process Mining,* 2011. BPTrends, <www.bptrends.com>.
- [8] **Mark A. Thornton.** *General Circulation and the Southern Hemisphere,* 2005. <<http://www.lakeeriewx.com/Meteo241/ResearchTopicTwo/ProjectTwo.html>>.

Wil van der Aalst
 Universidad Técnica de Eindhoven (Países Bajos)

<w.m.p.v.d.aalst@tue.nl>

© 2013 ACM, Inc. Este artículo fue publicado previamente en inglés con el título "Process Mining" en Communications of the ACM CACM, Volume 55, Issue 8, August 2012, páginas 76-83, doi>10.1145/2240236.2240257. Se publica con los permisos de reproducción correspondientes.

1. Espectro de la minería de procesos

La minería de procesos pretende descubrir, seguir y mejorar procesos reales mediante la extracción de conocimiento de los registros (logs) de eventos disponibles en los sistemas de información actuales [1][2].

Aunque los datos de eventos son omnipresentes, las organizaciones carecen de una buena comprensión de sus procesos reales. Las decisiones de dirección tienden a basarse en diagramas de Powerpoint, en políticas locales o en paneles de gestión en lugar de en un cuidadoso análisis de datos de sucesos. El conocimiento oculto en los registros de eventos no se puede convertir en información procesable. Los avances en minería de datos hacen posible encontrar valiosos patrones en grandes conjuntos de datos y apoyar decisiones complejas basadas en tales datos. Sin embargo, cuestiones clásicas de la minería de datos como la clasificación, el agrupamiento, la regresión, el aprendizaje de reglas de asociación, y la minería de secuencia/episodio *no* se centran en procesos.

Por lo tanto, los enfoques de la gestión de procesos de negocio (*Business Process Management*, BPM) tienden a recurrir a modelos hechos a mano. La investigación sobre la minería de procesos persigue cerrar la brecha que existe entre la minería de datos y BPM. Metafóricamente, la minería de procesos se puede ver como la obtención de radiografías a través de rayos-x para diagnosticar/predecir problemas y recomendar tratamientos.

Un importante vehículo para la minería de procesos es el increíble crecimiento de los datos de eventos [4][6]. Estos datos están en todas partes – en cada sector, en cada economía, en cada organización, y en cada hogar donde uno pueda encontrar sistemas que registren eventos. Por menos de 600 dólares cualquiera puede comprar una unidad de disco duro con capacidad para almacenar toda la música del mundo. [6]. Un reciente estudio publicado en Science muestra que el espacio de almacenamiento creció de 2,6 exabytes óptimamente comprimidos (2.6 x 10

Minería de procesos: Obtenga una radiografía de sus procesos de negocio

Traducción: Josep Moya Pérez (Grupo de Trabajo de Lengua e Informática de ATI)

Resumen: Grandes avances recientes en la investigación de la minería de procesos hacen posible descubrir, analizar y mejorar procesos de negocio basados en datos de eventos. Las actividades realizadas por las personas, por las máquinas y el software dejan rastros en los así llamados event logs o registros de sucesos. Eventos como introducir un pedido de un cliente en un sistema SAP, registrarse para un vuelo, cambiar la dosis para un paciente y rechazar un permiso de obras tienen en común que todos ellos son registrados por sistemas de información. Durante la última década ha habido un espectacular crecimiento de datos. Además, el universo digital y el universo físico se están alineando cada vez más. Por lo tanto, los procesos de negocio se deberían gestionar, soportar y mejorar en base a datos de eventos en lugar de hacerlo con opiniones subjetivas y experiencias obsoletas. La aplicación de la minería de procesos en cientos de organizaciones ha mostrado que tanto gestores como usuarios tienden a sobreestimar su conocimiento de los procesos en los que están involucrados. Por consiguiente, los resultados de la minería de procesos pueden contemplarse como una radiografía que muestra lo que realmente sucede dentro de los procesos. Como las radiografías, pueden ser utilizadas para diagnosticar problemas y proponer el tratamiento adecuado. La importancia práctica de la misma y los interesantes retos científicos hacen de la minería de procesos una de las cuestiones de más actualidad en la gestión de procesos de negocio (Business Process Management, BPM). Este artículo proporciona una introducción a la minería de procesos explicando los conceptos principales y tratando de varias aplicaciones de esta tecnología emergente.

Palabras clave: Administración, dirección de proceso de negocio, inteligencia de negocio, medición, minería de datos, minería de procesos, rendimiento.

Autor

Wil van der Aalst es profesor en la Universidad Técnica de Eindhoven y cuenta con un índice h de más de 90 puntos entre los científicos más citados en Europa. Se le conoce sobre todo por sus trabajos en modelos de *Workflow* y se le considera sobradamente como el "padrino" de la minería de procesos. Su web personal se encuentra en <www.vdaalst.com>.

¹² bytes) en 1986, a 295 exabytes comprimidos en 2007. En ese año, el 94 por ciento de toda la capacidad de almacenamiento de información sobre la Tierra era digital. El otro 6 por ciento residía en los libros, en revistas y en otros formatos no digitales. Esto es en marcado contraste con 1986, cuando sólo el 0,8 por ciento de toda la capacidad de almacenamiento de información era digital. Estas cifras ilustran el crecimiento exponencial de los datos.

La mayor adopción de tecnologías como la identificación por radiofrecuencia (RFID, *Radio Frequency Identification*), los servicios basados en la localización, computación en la nube (*cloud computing*) y redes de sensores, acelerarán todavía más el crecimiento de datos de eventos. Sin embargo, las organizaciones tienen problemas para usar eficazmente esas grandes cantidades de datos de eventos. De hecho, la mayoría de corporaciones todavía diagnostican problemas basándose en ficciones (diapositivas de Powerpoint, diagramas de Visio, etc) en lugar de en hechos

(datos de eventos). Esto es ilustrativo de la mala calidad de los modelos de procesos en la práctica; por ejemplo, más del 20% de los diagramas de procesos 604 en el modelo de referencia de SAP tienen errores clarísimos y su relación con los procesos de negocio actuales soportados por SAP es confusa [7]. Por lo tanto, es vital convertir ingentes cantidades de datos de eventos en conocimiento relevante y percepciones fiables. Aquí es donde puede ayudar la minería de procesos.

La creciente madurez de la minería de procesos queda ilustrada en el Manifiesto de Minería de Procesos (*Process Mining Manifesto*) [5], publicado recientemente por la *IEEE Task Force on Process Mining*. Este manifiesto está apoyado por 53 organizaciones y 77 expertos en minería de procesos han contribuido en él. La participación activa de usuarios finales, vendedores de herramientas, consultores, analistas e investigadores ilustra la importancia de la minería de procesos como puente entre el *data mining* minería de datos y el modelado de procesos de negocio.

“ La verificación de conformidad se puede usar para comprobar si la realidad, tal y como se grabó en el registro, es conforme con el modelo y viceversa ”

El punto de partida para la minería de procesos es un *registro (log) de eventos*. Cada suceso en este registro se refiere a una *actividad* (por ejemplo, un paso bien definido en algún proceso) y está relacionado con un *caso particular* (por ejemplo, una *instancia de proceso*). Los eventos pertenecientes a un caso están ordenados y se pueden ver como una ejecución de ese proceso. Los registros de eventos pueden almacenar información adicional sobre estos sucesos. De hecho, siempre que es posible, las técnicas de minería de procesos utilizan información extra como el *recurso* (por ejemplo, una persona o un dispositivo) que ejecuta o inicia la actividad, la *marca de tiempo* del evento, o *elementos de datos* registrados con el mismo (por ejemplo, el tamaño de una orden).

Los registros de eventos se pueden usar para realizar tres tipos de minería de procesos, tal y como se muestra en la **figura 1** [1]. El primer tipo de minería de procesos es el *descubrimiento*. Una técnica de descubrimiento toma un registro de sucesos y crea un modelo sin usar a priori ninguna información. El descubrimiento de procesos es la técnica más importante en la minería de procesos. En muchas organizaciones es sorprendente ver que las técnicas existentes son realmente capaces de descubrir auténticos procesos basados simplemente en ejemplos de comportamientos grabados en los registros de eventos.

El segundo tipo de minería de procesos es la *conformidad*. Aquí, un modelo de proceso existente es comparado con un registro de eventos del mismo proceso. La verificación de conformidad se puede usar para comprobar si la realidad, tal y como se grabó en el registro, es conforme con el modelo y viceversa.

El tercer tipo es la *mejora*. Aquí la idea es ampliar o mejorar un modelo de proceso existente utilizando información sobre el proceso real registrado en el mismo fichero de eventos. Considerando que la comprobación de la conformidad mide el alineamiento entre modelo y realidad, este tercer tipo de minería de procesos pretende cambiar o extender el modelo a priori. Por ejemplo, usando marcas de tiempo en el registro de eventos se puede extender el modelo para que muestre cuellos de botella, niveles de servicio, tiempos de rendimiento y frecuencias.

2. Descubrimiento de procesos

Tal y como se muestra en la **figura 1**, el objetivo del descubrimiento de procesos es descubrir un modelo basado en un registro de eventos. Estos pueden tener toda clase de atributos (marcas de tiempo, información transaccional, uso de recursos, etc.). Todos se pueden usar en el descubrimiento de procesos. Sin embargo, por simplicidad, solemos representar los eventos sólo con nombres de actividades.

De esta manera, un caso (una instancia de proceso, por ejemplo) se puede representar por una traza que describe una secuencia de acciones. Considérese como ejemplo el registro de eventos mostrado en la **figura 1** (ejemplo tomado de [1]). Este registro contiene 1.391 casos, es decir, instancias de algún proceso de reembolso. Hay 455 instancias de proceso que siguen el rastro **acdeh**. Las actividades se representan con un único carácter: **a** = *petición de registro*, **b** = *examinar minuciosamente*, **c** = *examinar casualmente*, **d** = *comprobar resguardo*, **e** = *decidir*, **f** = *reiniciar solicitud*, **g** = *pago de indemnización* y **h** = *rechazar petición*. Por lo tanto, la traza **acdeh** modela una petición de reembolso que fue rechazada después de los pasos de registro, examen, comprobación y decisión. 455 casos siguen esta secuencia consistente en cinco etapas, es decir, la primera línea en la tabla corresponde a $455 \times 5 = 2.275$ eventos. El registro entero comprende 7.539 sucesos.

Las técnicas de descubrimiento de procesos producen modelos de procesos basados en registros de eventos como el mostrado en la **figura 2**. Por ejemplo, el algoritmo clásico α produce un modelo M_i para este registro. Este modelo se representa como una *red de Petri*. Una red de Petri consiste en *posiciones* y *transiciones*. El estado de una red de Petri, también conocido como *marcado*, se define por la distribución de *tokens* sobre posiciones. Una transición se activa si cada una de sus posiciones de entrada contiene un *token*. Por ejemplo, α se *activa* en la marca inicial de M_i debido a que la única posición de entrada de α contiene un *token* (un punto negro). La transición de **e** en M_i solamente se activa si ambas posiciones contienen un *token*. Una transición activada puede de este modo lanzar el consumo de un *token* desde cada una de sus posiciones de entrada y producir un *token* por cada una de sus posiciones de salida. Lanzar α en la marca inicial se corresponde con eliminar un *token* del inicio y con producir dos *tokens* (uno por cada posición de salida). Después de lanzar α se activan tres transiciones: **b**, **c** y **d**. El lanzamiento de **b** desactivará **c** porque el *token* está borrado de la posición de entrada compartida (y viceversa). La transición **d** es concurrente con **b** y **c**, es decir, se puede lanzar sin desactivar otra transición. La transición **e** se vuelve activa después de que **d** y **b** ó **c** hayan sucedido. Después de ejecutarse **e** tres transiciones se activan: **f**, **g** y **h**. Estas transiciones compiten por el mismo *token* modelando así una elección. Cuando **g** ó **h** se han lanzado, el proceso termina con un *token*

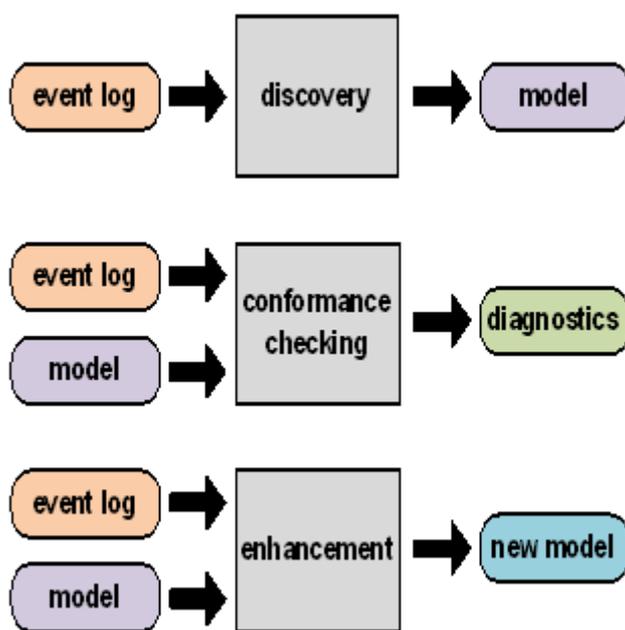


Figura 1. Los tres tipos básicos de minería de procesos explicados en términos de entrada y salida.

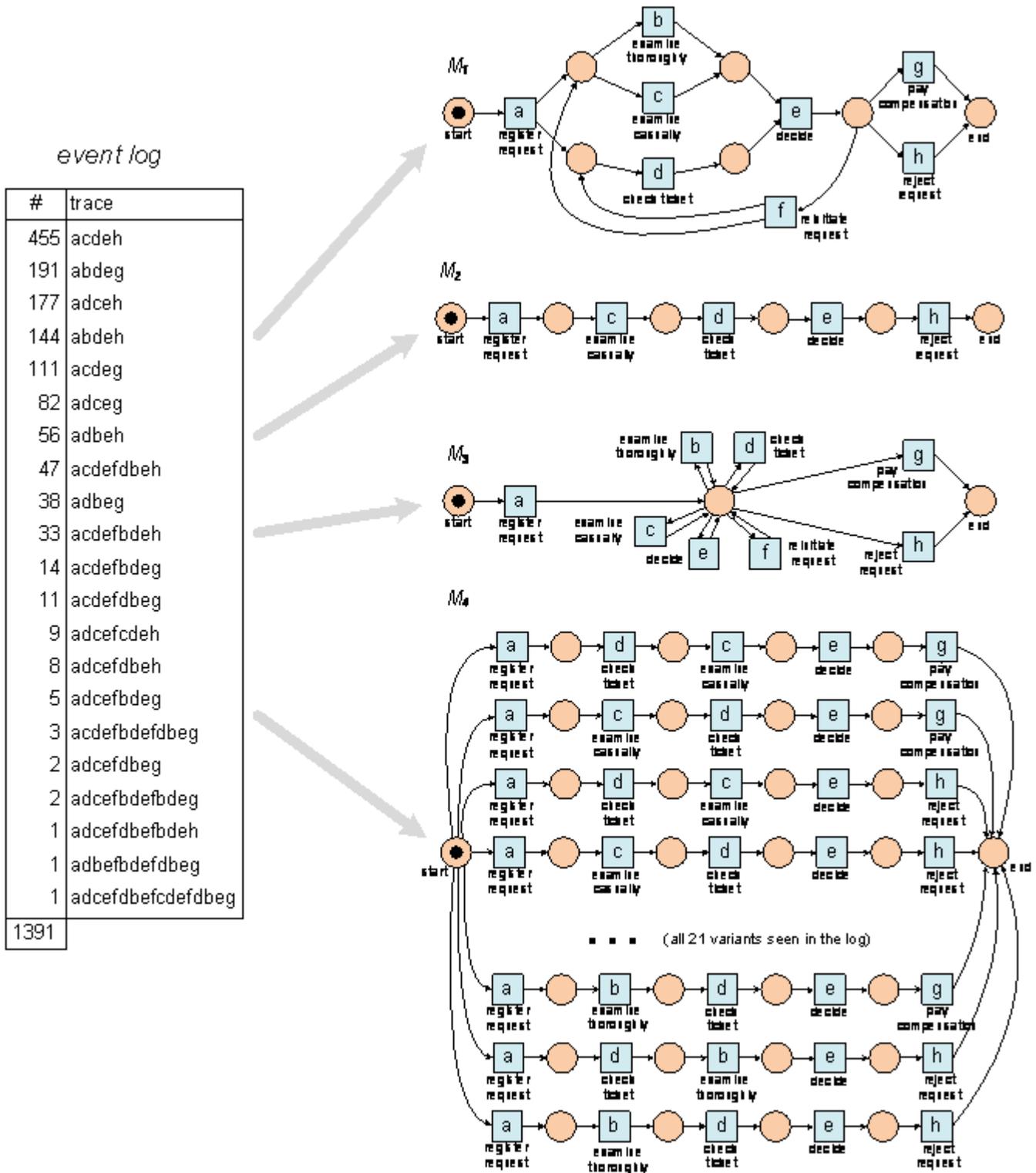


Figura 2. Un registro de eventos y cuatro modelos potenciales de procesos (M_1 , M_2 , M_3 y M_4) intentando describir el comportamiento observado.

en la posición *end*. Si **f** se lanza, el proceso regresa al estado justo después de ejecutar **a**.

Obsérvese que la transición **d** es concurrente con **b** y **c**. Las técnicas de minería de procesos necesitan poder descubrir patrones de proceso más avanzados y no se deberían restringir a simples procesos secuenciales.

Es fácil comprobar que todas las trazas en el registro de eventos pueden reproducirse por M_1 . Esto no es válido para el segundo modelo de proceso de la figura 2. M_2 sólo puede reproducir la traza más frecuente **acdch**. El modelo no *encaja* bien en el registro porque los rastros observados como **abdcg** no son posibles de acuerdo a M_2 .

El tercer modelo puede reproducir el registro de eventos completo, pero M_2 también lo permite para trazas como **ah** y **addddddg**. Por lo tanto, consideramos a M_2 como poco adecuado; se permite demasiado funcionamiento porque M_2 sobregeneraliza claramente el comportamiento observado. El modelo M_4 puede también reproducir el registro de

Una red de Petri consiste en *posiciones* y *transiciones*. El estado de una red de Petri, también conocido como *marcado*, se define por la distribución de *tokens* sobre posiciones ””

$$\gamma_1 = \begin{array}{|c|c|c|c|c|} \hline a & d & c & e & g \\ \hline a & d & c & e & g \\ \hline \end{array} \quad \text{and} \quad \gamma_2 = \begin{array}{|c|c|} \hline a & b \\ \hline a & b \\ \hline \end{array} \gg \begin{array}{|c|c|} \hline e & f \\ \hline e & f \\ \hline \end{array} \gg \begin{array}{|c|c|} \hline d & b \\ \hline d & b \\ \hline \end{array} \gg \begin{array}{|c|c|} \hline e & g \\ \hline e & g \\ \hline \end{array} \quad \text{and} \quad \gamma_3 = \begin{array}{|c|c|} \hline a & b \\ \hline a & b \\ \hline \end{array} \gg \begin{array}{|c|c|} \hline e & f \\ \hline e & f \\ \hline \end{array} \gg \begin{array}{|c|c|} \hline d & e \\ \hline d & e \\ \hline \end{array} \gg \begin{array}{|c|c|} \hline g & g \\ \hline g & g \\ \hline \end{array}$$

Tabla 1. Alineaciones de ejemplo entre las trazas y el modelo M_1 .

eventos. Sin embargo, el modelo únicamente codifica las trazas de ejemplo en el registro. Tal modelo lo denominamos "sobreadecuado" al no generalizar el comportamiento más allá de los ejemplos observados.

En años recientes se han desarrollado poderosas técnicas de minería que pueden construir automáticamente un modelo personalizado dado un registro de eventos. El objetivo de esas técnicas es crear un modelo simple que sea capaz de explicar la mayor parte de los comportamientos observados sin subvalorar o sobrevalorar el registro.

3. Comprobación de conformidad

La minería de procesos no se limita al descubrimiento de procesos. De hecho, el proceso descubierto es tan sólo el punto de partida de un análisis más profundo. Como se muestra en la figura 1, la comprobación de conformidad y la mejora relacionan modelo y registro. El modelo puede haber sido creado a mano o hallado por medio de descubrimiento de procesos. Para la comprobación de la conformidad, el comportamiento modelado y el comportamiento observado (es decir, el registro de eventos), se comparan. Cuando se comprueba la conformidad de M_2 con respecto al registro mostrado en la figura 2 es fácil observar que únicamente los 455 casos que siguieron la traza **acdeh** pueden ser repetidos de principio a fin. Si intentamos repetir la traza **acdeg** nos quedaremos parados después de la ejecución de **acde** porque **g** no está activado. Si intentamos repetir el rastro **adceh** nos atoraremos después de ejecutarse el primer paso, porque **d** no está (todavía) activado.

Existen varios enfoques para diagnosticar y cuantificar la conformidad. Un enfoque es encontrar una alineación óptima entre cada traza en el registro y el comportamiento más parecido en el modelo. Considérese por ejemplo el modelo de proceso M_1 , una traza adecuada $\sigma_1 = \text{adceg}$ una traza no adecuada $\sigma_2 = \text{abefdeg}$, y las tres alineaciones mostradas en la tabla 1.

γ_1 muestra una alineación perfecta entre σ_1

y M_1 : todos los movimientos de la traza en el registro de eventos (parte superior de la alineación) pueden ser seguidos por movimientos del modelo (parte inferior de la alineación). γ_2 muestra una alineación óptima para la traza σ_2 en el registro de eventos, y el modelo M_1 . Los dos primeros movimientos de la traza en el registro pueden ser seguidos por el modelo. Sin embargo, *e* no está activada después de ejecutar sólo *a* y *b*. En la tercera posición de la alineación γ_2 , vemos un movimiento *d* del modelo que no está sincronizado con un movimiento en el registro de eventos.

Un movimiento sólo en el modelo se denota como (\gg , *d*). En los tres movimientos siguientes el modelo y el registro concuerdan. En la séptima posición del alineamiento γ_2 sólo hay un paso del modelo y ninguno en el registro: (\gg , *b*). γ_3 muestra otra alineación óptima para la traza σ_2 . Aquí se dan dos situaciones donde el registro y el modelo no se mueven juntos: (*e*, \gg) y (*f*, \gg). Los alineamientos γ_2 and γ_3 son ambos óptimos si las penalizaciones por moverse en el registro de eventos y por moverse en el modelo son las mismas. En ambas alineaciones hay dos \gg pasos y no existen alineamientos con menos de dos \gg pasos.

La conformidad puede verse desde dos ángulos: (a) el modelo no captura el comportamiento real ("el modelo está equivocado") y (b) la realidad se desvía del modelo deseado ("el registro de sucesos es erróneo"). El primer punto de vista se toma cuando se supone que es *descriptivo*, es decir, que captura o predice la realidad. El segundo punto de vista se toma cuando el modelo es *normativo*, es decir, usado para influenciar o controlar la realidad.

Existen varios tipos de conformidad y el crear un alineamiento entre el registro y el modelo sólo es el punto de partida para la comproba-

ción de la conformidad [1]. Por ejemplo, hay varias métricas de estado físico (la habilidad para repetir). Un modelo tiene el estado 1 si todas las trazas pueden repetirse de principio a fin. Otro modelo tiene estado 0 si dicho modelo y el registro no concuerdan en ningún evento. Los modelos de procesos M_1 , M_2 y M_3 tienen un estado de 1 (o sea, en perfecto estado) con respecto al registro de sucesos mostrado en la figura 2. El modelo M_2 tiene un estado de 0,8 para el registro de eventos compuesto por 1.391 casos. Intuitivamente, esto significa que el 80% de los sucesos en el registro pueden ser explicados por el modelo. El estado físico es sólo una de varias métricas de conformidad.

Las experiencias con comprobaciones de conformidad en docenas de organizaciones revelan que los procesos en la vida real suelen desviarse de las representaciones simplificadas en Visio y PowerPoint que suelen usar los analistas de procesos.

4. Mejora de modelo

Es también posible extender o mejorar un modelo de proceso existente usando el alineamiento entre registro de eventos y modelo. Un modelo de proceso no adecuado se puede corregir utilizando los diagnósticos proporcionados por la alineación. Si el alineamiento contiene muchos (*e*, \gg) pasos, entonces podría tener sentido el salto de la actividad *e* en el modelo. Además, los registros de eventos pueden contener información sobre recursos, marcas de fecha y datos del caso. Por ejemplo, un evento refiriéndose a la acción "petición de registro" y con el caso "992564" puede tener también atributos que describan a la persona que registró la solicitud (es decir, "John"), la fecha del evento (por ejemplo, "30-11-2011:14:55"), la edad del cliente (como "45"), y el importe reclamado (por ejemplo, "650 euros"). Después del alineamiento del modelo y del registro es posible repetir el registro de sucesos sobre el modelo. Mientras se repite se pueden analizar estos atributos adicionales.

Por ejemplo, como muestra la figura 3, es posible analizar tiempos de espera entre actividades. Midiendo simplemente la diferencia

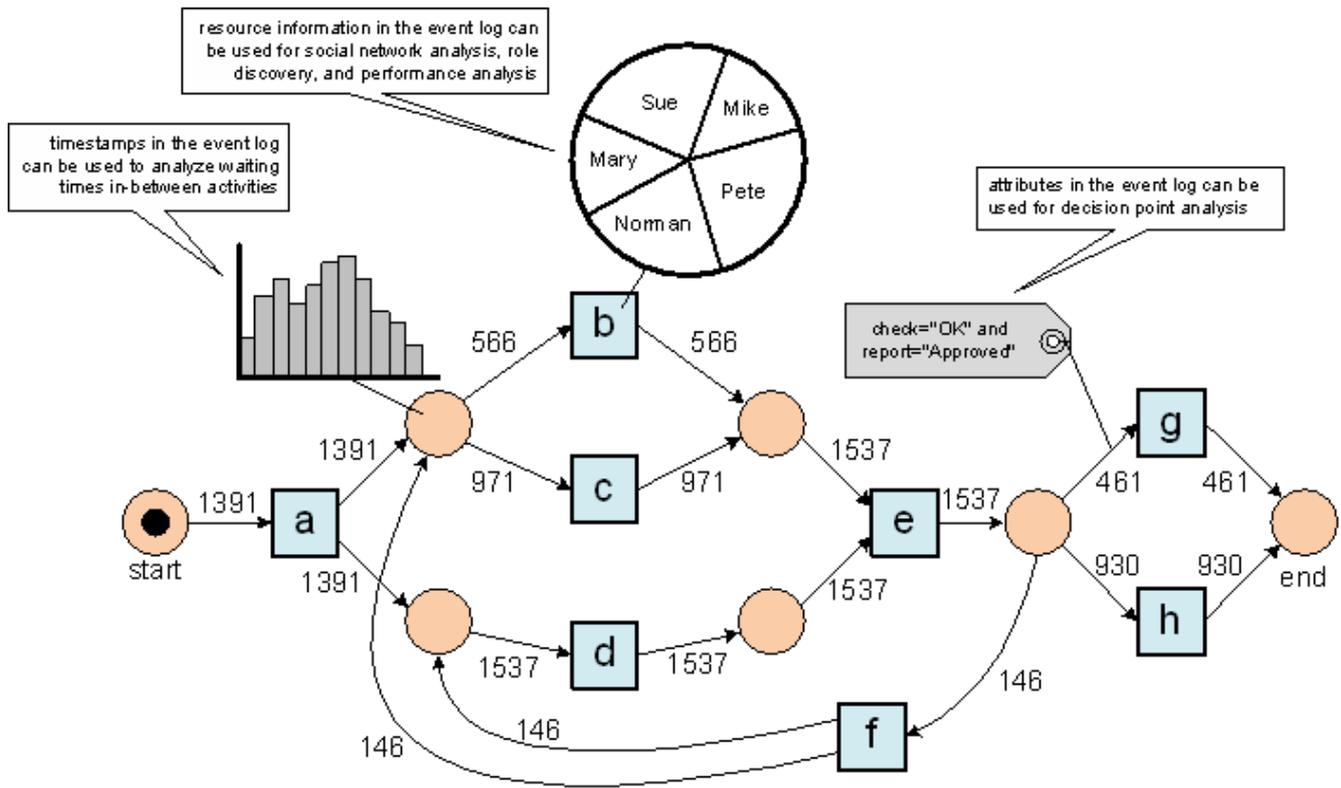


Figura 3. El modelo de proceso se puede ampliar usando atributos de eventos como las marcas de fecha, información de recursos y datos del caso. El modelo muestra también frecuencias, como por ejemplo, una decisión que se tomó 1.537 veces y en 930 casos se rechazó.

de tiempo entre eventos causalmente relacionados y calculando estadísticas básicas tales como medias, varianzas e intervalos de confianza. De esta forma es posible identificar los principales cuellos de botella.

La información sobre recursos se puede utilizar para hallar roles, como por ejemplo grupos de personas ejecutando frecuentemente actividades relacionadas. Aquí se pueden usar las técnicas de agrupamiento convencionales. También es posible construir redes sociales basadas en el flujo de trabajo y analizar las prestaciones de los recursos (como la relación entre la carga de trabajo y los tiempos de servicio).

Las técnicas de clasificación estándar pueden usarse para analizar los puntos de decisión en el modelo de proceso. Por ejemplo, la actividad e ("decidir") tiene tres posibles resultados: ("pagar", "rechazar" y "volver a hacer"). Utilizando los datos conocidos sobre el caso antes de la decisión, podemos construir un árbol de decisión explicando el comportamiento observado.

La **figura 3** muestra que la minería de procesos no se limita a hallar controles de flujo. Además, la minería de procesos no se restringe al análisis fuera de línea y se puede usar para predicciones y recomendaciones en tiempo de ejecución. Por ejemplo, el tiempo de terminación de un pedido de cliente parcial-

mente realizado puede ser predicho mediante un modelo de proceso hallado con información de tiempo.

5. La minería de procesos crea valor de varias formas

Después de introducirnos en las tres clases de minería de procesos utilizando un pequeño ejemplo, nos centraremos ahora en el valor práctico de esta disciplina. Como dijimos anteriormente, la minería de procesos se mueve por el crecimiento exponencial de los datos de eventos. Por ejemplo, de acuerdo con MGI, las empresas almacenaron más de 7 exabytes de nuevos datos en unidades de disco en el 2010 mientras que los consumidores guardaron más de 6 exabytes de nueva información en dispositivos como PC's y notebooks [6].

En lo que resta, mostraremos que la minería de procesos puede aportar valor de varias maneras. Para ilustrar esto nos referiremos a casos de estudio donde hemos usado nuestro paquete de software de código abierto ProM [1]. ProM ha sido creado y es mantenido por el grupo de minería de procesos de la Eindhoven University of Technology. No obstante, han contribuido a él grupos de investigación de todo el mundo, como la Universidad de Padua, la *Universitat Politècnica de Catalunya*, la Universidad de Calabria, la Universidad Humboldt de Berlín, la Queensland University of Technology, la Universidad Técnica de

Lisboa, la Universidad de Economía y Empresas de Viena, el Instituto Nacional de Ciencia y Tecnología de Ulsan, el K.U. Leuven, la Universidad de Tsinghua, y la Universidad de Innsbruck. Aparte de ProM existen alrededor de una decena de suministradores de software comercial que proporcionan software de minería de procesos (suelen estar embebidos en grandes herramientas), como Pallas Athena, Software AG, Futura Process Intelligence, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, y Stereologic.

5.1. Proporcionar percepciones

Durante la última década hemos aplicado nuestro software de minería de procesos ProM en más de 100 organizaciones. Algunos ejemplos son municipios (sobre 20 en total, como Alkmaar, Heusden y Harderwijk), agencias del gobierno (*Rijkswaterstaat*, *Centraal Justitiele Incasso Bureau*, y el departamento de Justicia holandés), agencias relacionadas con los seguros (como UWV), bancos (como el ING Bank), hospitales (por ejemplo, el hospital AMC y el hospital Catharina), multinacionales (como por ejemplo DSM y Deloitte), fabricantes de sistemas de alta tecnología y sus clientes (Philips Healthcare, ASML, Ricoh, y Thales), y compañías de medios de comunicación (como Winkwaves). En cada una de estas corporaciones hemos encontrado algunos de sus procesos basados en los datos de eventos que proporcionaron. En cada proceso descubierto habían partes que

“ Es también posible extender o mejorar un modelo de proceso existente usando el alineamiento entre registro de eventos y modelo ”

sorprendieron a algunos de los interesados. La variabilidad de los procesos es típicamente mucho mayor de la esperada. Tales percepciones representan un tremendo valor debido a que las sorprendentes diferencias suelen apuntar a menudo al derroche y a la mala gestión.

5.2. Mejorar el rendimiento

Como hemos explicado anteriormente, es posible reproducir los registros de eventos en modelos de proceso encontrados o hechos manualmente. Esto se puede usar para la comprobación de conformidad y en la mejora de modelos. Puesto que la mayoría de registros de sucesos contienen marcas de fecha, la repetición se puede utilizar para ampliar el modelo con información del rendimiento.

La **figura 4** ilustra algunos de los diagnósticos relacionados con el rendimiento que se pueden obtener por medio de la minería de procesos. El modelo mostrado fue hallado

basándose en 745 objeciones contra la valoración denominada WOZ ("Waardering Onroerende Zaken") en una población holandesa. Los municipios holandeses necesitan estimar el valor de las casas y apartamentos. El valor WOZ se usa como base para determinar el impuesto sobre la propiedad inmobiliaria. Cuanto más alto sea el valor WOZ, más impuesto debe pagar el propietario. Por lo tanto, muchos ciudadanos apelan contra la valoración WOZ afirmando que es demasiado alto.

Cada una de las 745 objeciones corresponden a una instancia de proceso. Juntas, estas instancias generaron 9.583 eventos teniendo todos marcas de tiempo. La **figura 4** muestra la frecuencia de las diferentes trayectorias en el modelo. Por otra parte, las diferentes etapas del modelo están coloreadas para mostrar dónde, en promedio, se emplea más tiempo. Las etapas púrpura del proceso nece-

sitan más tiempo mientras que las etapas azules usan menos tiempo. También es posible seleccionar dos actividades y medir el tiempo que transcurre entre ellas.

Como se ve en la **figura 4**, pasan de media 202,73 días entre la terminación de la actividad "OZ02 Voorverieden" (preparación) y la conclusión de "OZ16 Uitspraak" (decisión final). Esto es más largo que el promedio de tiempo de flujo total que es aproximadamente de 178 días. Alrededor de 416 de estas objeciones (aproximadamente el 56%) siguen este camino; los demás casos siguen el ramal "OZ15 Zelf uitspraak", que por término medio usa menos tiempo.

Como se puede observar en la **figura 4**, los diagnósticos se pueden utilizar para mejorar los procesos eliminando los cuellos de botella y desviando los casos. Dado que el modelo está conectado a datos de eventos, es posible

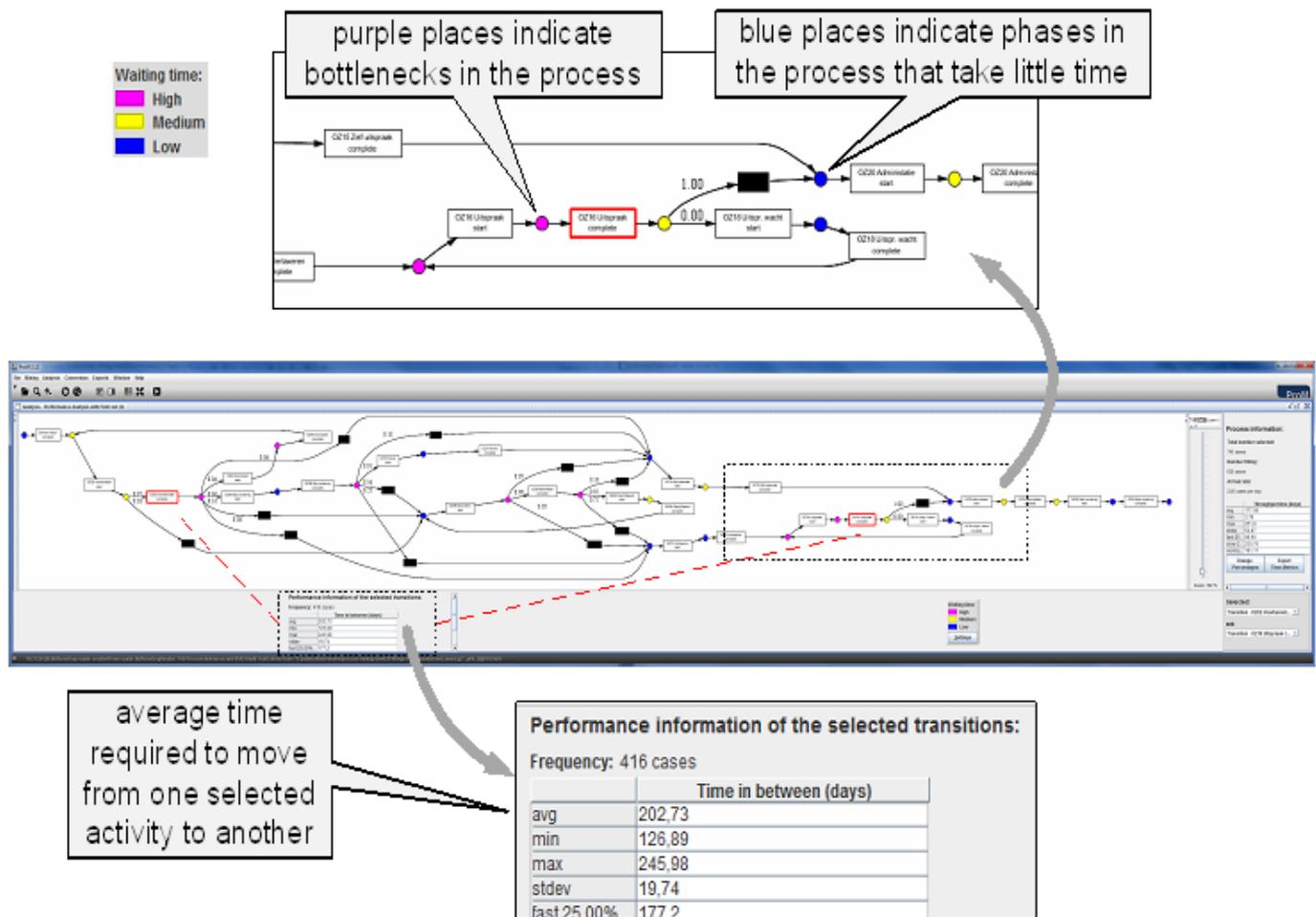


Figura 4. Análisis de rendimiento basado en 745 apelaciones contra la valoración WOZ.

“ A menudo, una "realidad PowerPoint" tiene poco en común con el proceso real, que tiene mucha más variabilidad. Sin embargo, para mejorar la conformidad y el rendimiento no debería dejar de abstraerse esta variabilidad ”

profundizar de inmediato e investigar grupos de casos que usan más tiempo que otros [1].

5.3. Asegurar la conformidad

La repetición también se puede utilizar para comprobar la conformidad tal y como se muestra en la **figura 5**. En base a las 475 apelaciones en contra de la valoración WOZ, hemos comparado también el modelo normativo y el comportamiento observado: 628 de los 475 casos se pueden repetir sin encontrar ningún problema. La adecuación del modelo y del registro es de 0,98876214 indicando que casi todos los eventos registrados se pueden explicar por el modelo. A pesar de su buena adecuación, ProM muestra claramente todas las desviaciones. Por ejemplo, "OZ12 Hertaxeren" (reevaluar propiedad) sucedió 23 veces cuando esto no estaba permitido según el modelo normativo (indicado por el "-23" en la **figura 5**). Es fácil de nuevo profundizar y ver qué tienen estos casos en común.

La conformidad del proceso de apelación que acabamos de describir es muy alta (alrededor del 99% de los eventos son posibles según el

modelo). También hemos encontrado muchos procesos con una conformidad muy baja; no es infrecuente hallar procesos donde sólo el 40% de los sucesos son posibles de acuerdo con el modelo. Por ejemplo, la minería de procesos reveló que el proceso de modelado de prueba ASML se desvió fuertemente del proceso real [9].

La creciente importancia del gobierno corporativo, riesgo y gestión de cumplimiento normativo y legislación como los de la Ley Sarbanes-Oxley (SOX), y el acuerdo Basel II, ilustran la importancia práctica de la comprobación de conformidad. La minería de procesos puede ayudar a los auditores a comprobar si los procesos se ejecutan dentro de ciertos límites establecidos por administradores, gobiernos y otras partes interesadas [3].

Las violaciones encontradas por medio de la minería pueden indicar fraudes, negligencias, riesgos e ineficiencias. Por ejemplo, en el municipio donde analizamos el proceso de apelación WOZ, hemos descubierto errores de configuración de su sistema de gestión de flujo de trabajo eIStream. Las personas tam-

bién evadieron el funcionamiento del sistema. Esto fue posible porque los administradores del sistema podrían haber cambiado manualmente el estado de los casos [8].

5.4. Mostrar variabilidad

Los modelos de proceso artesanales tienden a proporcionar una visión idealizada en el proceso de negocios que se modela. A menudo, una "realidad PowerPoint" tiene poco en común con el proceso real, que tiene mucha más variabilidad. Sin embargo, para mejorar la conformidad y el rendimiento no debería dejar de abstraerse esta variabilidad.

En el contexto de la minería de procesos solemos ver modelos tipo "spaghetti" como se puede ver en la **figura 6**. El modelo se encontró en base a un registro que contenía 24.331 eventos referidos a 376 actividades diferentes. El fichero de sucesos describe la diagnosis y el tratamiento de 627 pacientes de oncología ginecológica en el hospital AMC de Amsterdam. Las estructuras parecidas a *spaghettis* no están causadas por el algoritmo de descubrimiento, sino por la auténtica variabilidad del proceso.

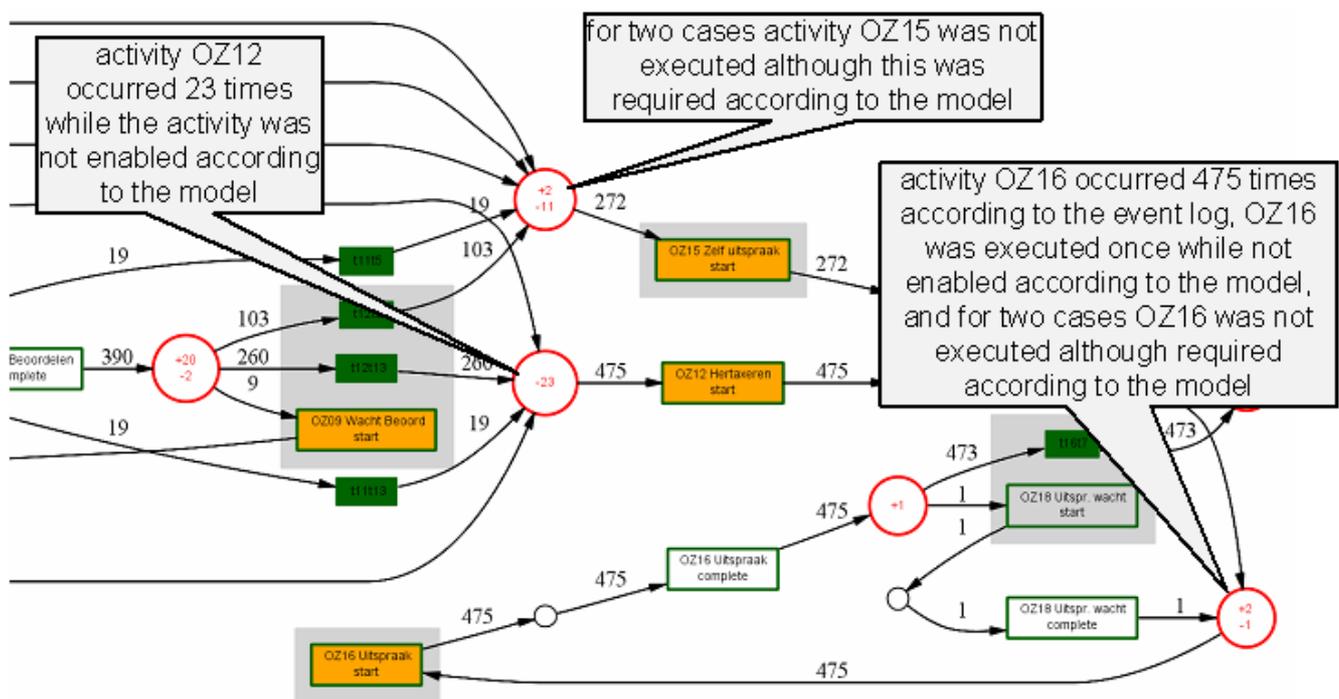


Figura 5. Análisis de la conformidad mostrando desviaciones entre el registro de eventos y el modelo de proceso.

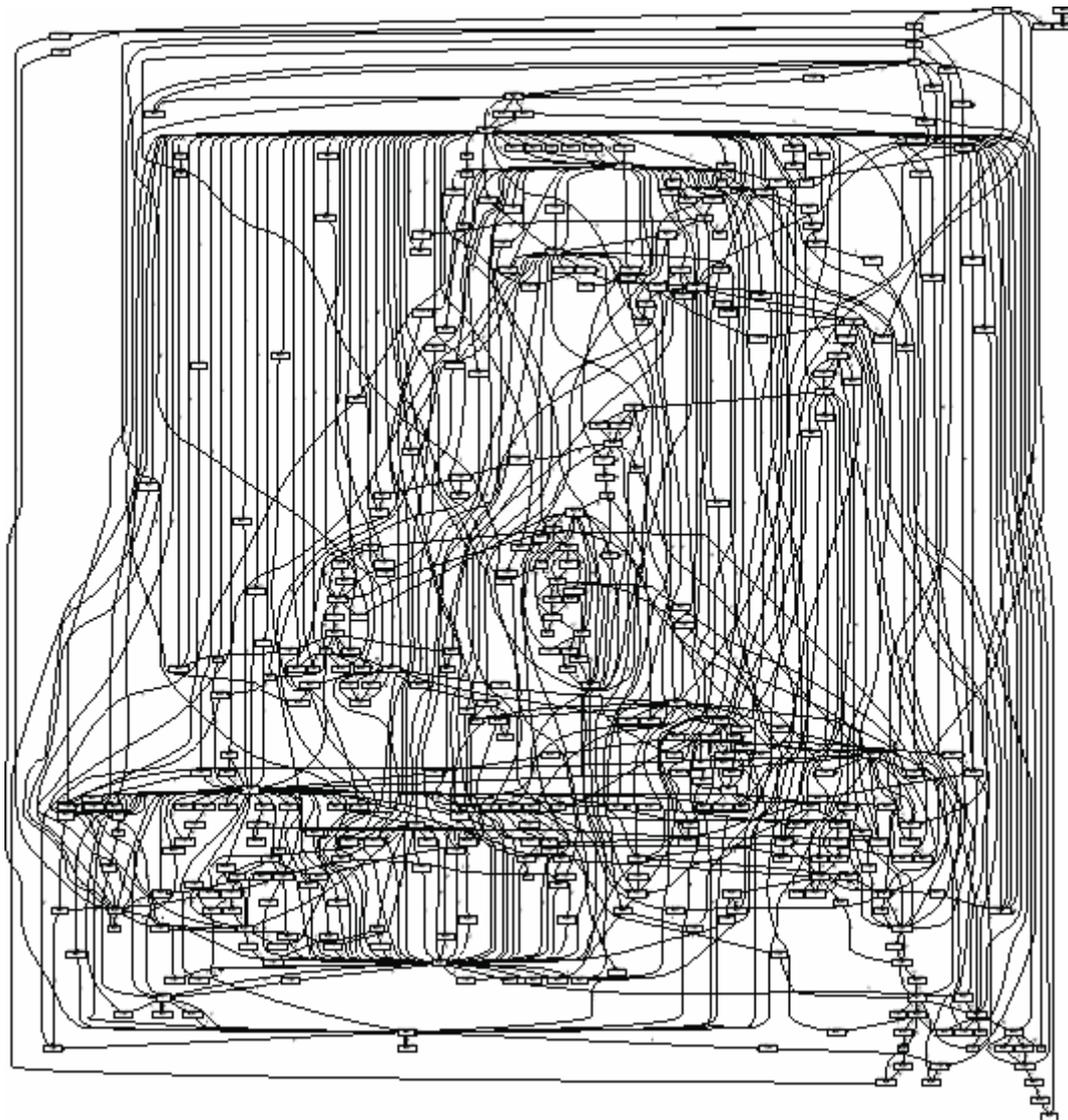


Figura 6. Modelo de proceso descubierto en un grupo de 627 pacientes de oncología ginecológica.

Aunque es importante enfrentar a las partes interesadas con la realidad, como se muestra en la figura 6, podemos también simplificar perfectamente estos modelos similares a *spaghetti*. Como cuando usamos mapas electrónicos, es también perfectamente posible ampliar y reducir [1]. Mientras reducimos, las cosas insignificantes se omiten o quedan dinámicamente agrupadas en capas añadidas, como las calles y los suburbios se amalgaman en ciudades en el Google Maps. El nivel de importancia de una actividad o conexión se puede basar en la frecuencia, en el costo, o en el tiempo.

5.5. Mejora de la fiabilidad

La minería de procesos se puede usar también para mejorar la fiabilidad de sistemas y procesos. Por ejemplo, desde 2007 hemos estado involucrados en un esfuerzo continuado para analizar los registros de eventos de los equipos de rayos-X de Philips Healthcare utilizando minería de procesos [1]. Estas máquinas registran cantidades

ingentes de eventos. En los equipamientos médicos es esencial probar que los sistemas han sido testeados bajo circunstancias realistas. Por lo tanto, el descubrimiento de procesos se usó para construir perfiles de prueba realistas. Philips Healthcare usó también minería de procesos para diagnosticar fallos. Aprendiendo de problemas anteriores, es posible hallar la causa de nuevos problemas que puedan surgir. Por ejemplo, al utilizar ProM hemos analizado bajo qué circunstancias se reemplazan ciertos componentes. Esto dio como resultado un conjunto de registros de comportamiento. Cuando un equipo de rayos-X con mal funcionamiento muestra el registro de un comportamiento particular, el ingeniero de mantenimiento sabe qué componente ha de cambiar.

5.6. Permitir predicciones

La combinación de datos históricos de eventos con otros datos en tiempo real puede ser útil para predecir problemas. Por ejemplo,

Philips Healthcare puede anticipar que un tubo de rayos-X en servicio está a punto de fallar por descubrirse patrones en el registro de eventos. Por lo tanto, se puede reemplazar el tubo antes de que la máquina empiece a fallar.

Hoy, muchas fuentes de datos se actualizan (casi) en tiempo real y existe suficiente potencia de cálculo para analizar eventos en cuanto suceden. Por lo tanto, la minería de procesos no se restringe tan sólo al análisis fuera de línea y se puede usar también para el soporte operacional *online*. Para una instancia de proceso en ejecución es posible hacer predicciones, como por ejemplo el tiempo de flujo esperado restante [1].

6. Conclusión

Las técnicas de minería de procesos permiten a las organizaciones radiografiar sus procesos de negocio, diagnosticar problemas y obtener sugerencias para solucionarlos. El descubrimiento de procesos proporciona a

menudo nuevas y sorprendentes ideas. Se pueden usar para rediseñar procesos o mejorar su gestión. La comprobación de conformidad puede utilizarse para observar dónde se desvían los procesos. Esto es muy importante, ya que las organizaciones necesitan poner más énfasis en el gobierno corporativo, en los riesgos y en el cumplimiento. Las técnicas de minería de procesos ofrecen un medio para comprobar más rigurosamente la conformidad mientras se mejora el rendimiento.

Este artículo ha introducido los conceptos básicos y mostrado que la minería de procesos puede aportar valor de varias maneras. Al lector interesado en la minería de procesos se le remite al primer libro sobre ésta [1], y al manifiesto ya mencionado [5], que está disponible en 12 idiomas. También le invitamos a visitar <www.processmining.org> para acceder a registros de eventos de ejemplo, videos, diapositivas, artículos y software.

El autor desea dar las gracias a los miembros del Grupo de Trabajo sobre Minería de Procesos del IEEE y a todos los que han contribuido al Manifiesto sobre Minería de Procesos y al *framework* ProM.

Referencias

- [1] **W. van der Aalst.** *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011. ISBN 978-3-642-19345-3.
- [2] **W. van der Aalst.** Using Process Mining to Bridge the Gap between BI and BPM. *IEEE Computer* 44, 12, pp. 77–80, 2011.
- [3] **W. van der Aalst, K. van Hee, J.M. van Werf, M. Verdonk.** Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor. *IEEE Computer* 43, 3, pp. 90–93, 2010.
- [4] **M. Hilbert, P. Lopez.** The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* 332, 6025, pp. 60–65, 2011.
- [5] **TFPM Task Force on Process Mining.** Process Mining Manifesto. *Business Process Management Workshops*, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Lecture Notes in Business Information Processing Series, vol. 99. Springer-Verlag, Berlin, pp. 169–194, 2012.
- [6] **J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers.** Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, 2011. <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>.
- [7] **J. Mendling, G. Neumann, W. van der Aalst.** Understanding the Occurrence of Errors in Process Models Based on Metrics. Proceedings of the OTM Conference on Cooperative Information Systems (CoopIS 2007). En F. Curbera, F. Leymann, and M. Weske, Eds. *Lecture Notes in Computer Science Series, vol. 4803*. Springer-Verlag, Berlin, pp. 113–130, 2007.
- [8] **A. Rozinat, W. van der Aalst.** Conformance Checking of Processes Based on Monitoring Real Behavior. *Information Systems* 33, 1, pp. 64–95, 2008.
- [8] **A. Rozinat, I. de Jong, C. Günther, W. van der Aalst.** Process Mining Applied to the Test Process of Wafer Scanners in ASML. *IEEE Transactions on Systems, Man and Cybernetics, Part C* 39, 4, pp. 474–479, 2009.



ACTUALIZACIÓN DATOS SOCIO ATI

¿Has cambiado de domicilio, de empresa, y lo has comunicado a la Secretaría General?

¿Recibes el correo postal de la asociación?

¿Te llegan los correos electrónicos enviados por las Secretarías de ATI?

Si has contestado que NO a todas estas preguntas, te agradeceríamos que enviaras un mensaje a secregen@ati.es con tus nuevos datos con el fin de tener actualizada tu ficha de socio y, de este modo, nos ayudes a mejorar la comunicación entre la asociación y sus miembros.

* Del mismo modo, si sabes de algún compañero tuyo, miembro de ATI, que no recibirá esta información, te agradeceríamos que se la hagas llegar para que se pueda poner en contacto con nosotros.

ATI Secretaría General | Vía Laietana 46, ppal. 1a. | 08003 Barcelona | 93 412 52 33 | secregen@ati.es | www.ati.es

Josep Carmona Vargas
Departament de Llenguatges i Sistemes
Informàtics, Universitat Politècnica de
Catalunya

<jcarmona@lsi.upc.edu>

El viaje del descubrimiento de procesos

Traducción: María del Carmen Ugarte García (Grupo de Trabajo de Lengua e Informàtica de ATI)

1. Introducción

La velocidad a la que crecen los datos en los sistemas de Tecnologías de la Información (TI) [1] hace crucial su automatización para permitir a empresas e instituciones gestionar sus procesos. Las técnicas automáticas abren la puerta a manejar grandes cantidades de datos, algo imposible para la capacidad humana. En este artículo hablamos sobre una de esas técnicas: modelos de descubrimiento de procesos. Para ilustrar la tarea principal detrás del descubrimiento de procesos nos serviremos de un ejemplo que esperamos les resulte simpático.

2. Un ejemplo divertido: la visita del marciano

Imaginemos que un marciano nos visita (ver figura 1) y, no importa por qué medios, quiere comunicar sus planes en lo que respecta a su visita a la Tierra. Por razones obvias no entendemos los mensajes del marciano, que tienen la apariencia mostrada en la figura 2.

A pesar de desconocer el significado de cada una de las letras en el mensaje anterior, se pueden detectar en él algunos patrones, por ejemplo una repetición de la secuencia I A C D M E (primera y últimas seis letras en la secuencia). Por lo tanto nos preguntamos: ¿cómo podemos representar el compor-



Figura 1. Imagen de nuestro marciano imaginario.

Resumen: Los modelos de proceso son un elemento de valor incalculable dentro de un sistema de Tecnologías de la Información (TI): sirven para analizar, monitorear o mejorar los procesos reales que proporcionan funcionalidad al sistema. La tecnología ha permitido a los sistemas TI almacenar en ficheros de log de eventos las marcas de las ejecuciones de los procesos, que pueden usarse para derivar los modelos de proceso que se corresponden con los procesos reales, mediante una disciplina llamada descubrimiento de procesos. En este artículo proporcionamos una visión general de esta disciplina con algunas de las alternativas existentes al día de hoy.

Palabras clave: Descubrimiento de procesos, Ingeniería del Software, métodos formales.

Autor

Josep Carmona Vargas obtuvo una Ingeniería y un Doctorado en Informática por la Universitat Politècnica de Catalunya en 1999 y en 2004, respectivamente. Es profesor asociado del Departament de Llenguatges i Sistemes Informàtics de dicha universidad. Sus intereses investigadores incluyen métodos formales, sistemas concurrentes y minería de procesos y datos. Ha sido autor o coautor de más de 50 artículos en congresos y revistas.

tamiento de los marcianos sin conocer exactamente el significado de cada unidad de información?

El descubrimiento de procesos puede ser una buena solución en este caso: el algoritmo de descubrimiento de procesos tratará de producir un modelo (formal) del comportamiento que subyace en un conjunto de secuencias. Por ejemplo, el modelo formal en notación para el Modelado de Procesos de Negocio (BPMN, Business Process Modeling Notation) [2] mostrado en la figura 3 representa muy acertadamente el comportamiento expresado en las secuencias del marciano. Para aquellos que no están familiarizados con la notación BPMN, el modelo anterior describe el siguiente proceso: cuando I ocurre, entonces (operador 'x') o el proceso B seguido de X ocurre, o bien el proceso A seguido de C y D en paralelo (operador '+') ocurre, seguidos a su vez de M. Ambos procesos activan E, que a su vez reactiva I. Es claro, que incluso sin conocer nada sobre las acciones que ha tomado el marciano, la estructura general de estas actividades resulta patente con una simple inspección del modelo BPMN.

Ahora imaginemos que en un determinado momento se descifra el significado de las letras: evalúa el total de energía en la Tierra

(I, nivel de energía alto (B), invade la Tierra (X), nivel de energía bajo (A), reúne algunos ejemplares humanos (C), infórmate sobre el sistema de reproducción humana (D), enseña a los humanos cómo incrementar sus recursos energéticos (M), informa a los marcianos del platillo volante más próximo (E). Ante esta nueva información, el valor del modelo obtenido se incrementa significativamente (a pesar de que no podamos relajarnos al descubrir la situación global que el modelo pone en evidencia).

3. Anatomía de un algoritmo simple de descubrimiento de procesos

El ejemplo anterior ilustra uno de las principales tareas de un algoritmo de descubrimiento de procesos: dado un conjunto de trazas (llamado log de eventos) correspondiente a un comportamiento particular sometido a estudio, derivar un modelo formal que represente fielmente el proceso que ha producido esas trazas. En su forma más sencilla, los algoritmos de descubrimiento de procesos ponen el foco sobre la perspectiva de control de flujo, por ejemplo se calcula la ordenación entre actividades para realizar las ejecuciones del proceso. El ejemplo anterior ha considerado esta perspectiva.

Un log de eventos debe contener suficiente información para extraer la secuencia de las

I A C D M E I B X E I A D C M E I B X E I A C D M E

Figura 2. Mensajes enviados por el marciano.

“ El núcleo de un algoritmo de descubrimiento de procesos es la capacidad de extracción de la información necesaria para llegar mediante aprendizaje a un modelo que represente el proceso ”

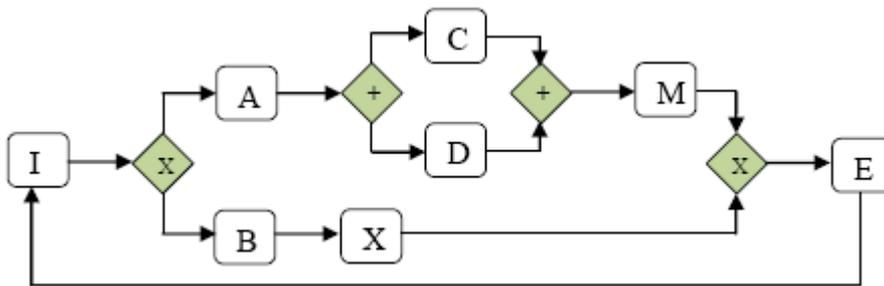


Figura 3. Modelo formal del comportamiento de las secuencias del marciano en BPMN.

actividades que están siendo monitorizadas. Típicamente se requiere un identificador de trazas, un nombre de actividad e información temporal para permitir la secuenciación correspondiente (mediante el sello de tiempo) para las actividades pertenecientes a una traza dada (determinada por un identificador de trazas). Otra información puede ser necesaria si el algoritmo de descubrimiento debe tener en cuenta información adicional como recursos (*qué cantidad se compró*), causante de la actividad (*quién llevó a cabo esa actividad?*), duración de la actividad (*cuánto duró la actividad*), entre otras. Un ejemplo del algoritmo de descubrimiento que tiene en cuenta otras dimensiones es el algoritmo para descubrir redes sociales [3], que deriva la red a través de un colaborador que realiza un proceso dado.

El núcleo de un algoritmo de descubrimiento de procesos es la capacidad de extracción de la información necesaria para llegar mediante aprendizaje a un modelo que represente el proceso.

El descubrimiento de procesos es a menudo una *tarea de aprendizaje no supervisada*, ya que el algoritmo está normalmente expuesto solo a ejemplos positivos, por ejemplo ejecuciones exitosas en el proceso en estudio: en el ejemplo de la introducción, estamos solo expuestos a lo que el marciano planea hacer, pero no sabemos lo que el marciano no piensa hacer. Esto complica la tarea de aprendizaje, ya que se espera que los algoritmos del descubrimiento del proceso produzcan modelos que sean a la vez *precisos* (el modelo resultante no debería desviarse mucho del comportamiento observado) y *generales* (el modelo debería ser capaz de generalizar los modelos observados en el *log* de eventos) [4]. Obviamente la presencia de ejemplos negati-

vos ayudaría al algoritmo de descubrimiento a mejorar estas dos métricas de calidad, pero la información negativa no está normalmente en los registros TI.

¿Cómo llegar un modelo de proceso a partir de un conjunto de trazas? Existen hoy día varios algoritmos para distintos modelos (ver **sección 4**). Sin embargo, usemos el ejemplo del marciano para razonar sobre el descubrimiento del modelo BPMN anterior. Si nos centramos en la primera letra de la secuencia (I), vemos que a veces le sigue A y a veces le sigue B, y siempre (excepto en la primera ocurrencia) E los precede. Estas observaciones pueden verse expresadas de modo gráfico en la **figura 4**.

En la notación BPMN, la relación *disjunción exclusiva* entre A o B después de I se modela usando el operador "x". La precedencia entre E e I se modela mediante una línea que conecta ambas letras. De forma simétrica, o bien M o bien X preceden a E. Igualmente, tanto C como D pueden seguir a A en cualquier orden. El bien conocido algoritmo *alfa* [5] puede encontrar la mayoría de las relaciones existentes entre pares en el *log* de eventos, y después pueden utilizarse para fabricar el modelo BPMN tal como ilustramos en la **tabla 1**.

Esta tabla puede leerse de la siguiente forma: si en el *log* de eventos A precede a B siempre, pero B es único (no hay otra letra que vaya detrás de A), entonces se crea un arco directo entre A y B. Si por el contrario hay siempre más de una letra que siga a A, entonces el operador "+" se inserta entre A y las letras que le siguen. Cuando la relación ocurre solo algunas veces la lectura es similar.

Por lo tanto se puede escanear el *log* de eventos para extraer estas relaciones (en el caso peor complejidad cuadrática respecto al *log* de eventos) y usar la tabla para crear el modelo BPMN. Sin embargo, esta es una forma de descubrimiento muy restrictiva ya que otras relaciones en la notación BPMN pueden estar escondidas en el *log* de eventos, como la relación de disjunción inclusiva, pero el algoritmo no las considera.

Los algoritmos de descubrimiento de procesos están siempre balanceando la complejidad de un algoritmo y su capacidad de modelado: el algoritmo propuesto en esta sección podría extenderse para considerar operadores o-inclusivos, pero ello lo complicaría significativamente. Más abajo analizaremos someramente este y otros problemas.

4. Algoritmos y modelos

Hay distintos modelos que pueden obtenerse a través de distintos algoritmos de descubrimiento de procesos: *redes de Petri, cadenas de eventos de proceso, redes causales, redes heurísticas, mapas de procesos de negocio*, entre otros. Hay que destacar que la mayoría de estos modelos disponen de una semántica que permite simular el modelo para así certificar si se adecúa al *log* de eventos.

La descripción de cada uno de ellos cae fuera de los límites de este artículo, pero comentaré brevemente sobre las redes de Petri, que es un modelo que a menudo producen los



Figura 4. Patrones observados en los mensajes del marciano.

“ Decidir cuál es la mejor notación de modelado para un registro de eventos es un problema arduo para el que la investigación debe proporcionar técnicas en la próxima década ”

	Always	Sometimes
A precedes B	B Unique: 	B Unique:
	General: 	General:

Tabla 1. Construcción de un modelo BPMN a partir de las relaciones descubiertas.

algoritmos de descubrimiento, debido a su semántica formal y la capacidad para representar la concurrencia. Para el modelo de nuestro ejemplo, la correspondiente red de Petri que sería puesta de manifiesto por la mayoría de los algoritmos de descubrimiento basados en Petri sería la mostrada en la figura 5.

Aquellos lectores familiarizados con las redes de Petri encontrarán un perfecto encaje entre el comportamiento subyacente en la red de Petri y la traza del marciano. Adviertan que mientras en el modelo BPMN, además de las unidades de información (en este caso las letras del alfabeto) hay otros componentes del modelo (operadores) cuya semántica define la forma en la que el modelo representa las trazas en el log de eventos.

Lo mismo ocurre con la red de Petri anterior, en la que los círculos se corresponden con el comportamiento global del modelo, que se distribuye entre la red (hay marcados algunos

círculos). Mientras el algoritmo de descubrimiento por BPMN necesita encontrar tanto las conexiones como los operadores, el algoritmo análogo de las redes de Petri debe computar tanto los círculos como las conexiones.

Hoy en día existen algunas técnicas para llevar a cabo el descubrimiento mediante redes de Petri; estas oscilan desde las relaciones extraídas de la ordenación del log de eventos mediante el algoritmo *alfa*, hasta estructuras complejas basadas en grafos que se computan sobre un autómata que representa las trazas del log de eventos.

¿Qué notación de algoritmo/modelado de descubrimiento de proceso deberemos elegir? Buena pregunta que solo tiene una respuesta parcial: no hay un modelo mejor que el resto, pero a cambio sí hay modelos mejores para un determinado tipo de comportamiento. Realmente decidir cuál es la mejor notación de modelado para un log de eventos es un pro-

blema arduo para el que la investigación debe proporcionar técnicas en la próxima década (un problema llamado *selección de sesgo representativo*). Desde un punto de vista pragmático, se deben seleccionar las notaciones de modelado de procesos con las que se esté más familiarizado, y esperar que los algoritmos de descubrimiento para esa notación sean adecuados para nuestras necesidades.

Como se ha dicho con anterioridad pueden considerarse otras perspectivas diferentes al flujo de control para los algoritmos de descubrimiento del proceso: tiempo, recursos, organización, etc.

Puede consultarse el manual de referencia [6] para profundizar en estos otros algoritmos de descubrimiento de procesos.

5. Herramientas

El descubrimiento de procesos es una disciplina bastante nueva si la comparamos con otras áreas ajenas tales como la minería de

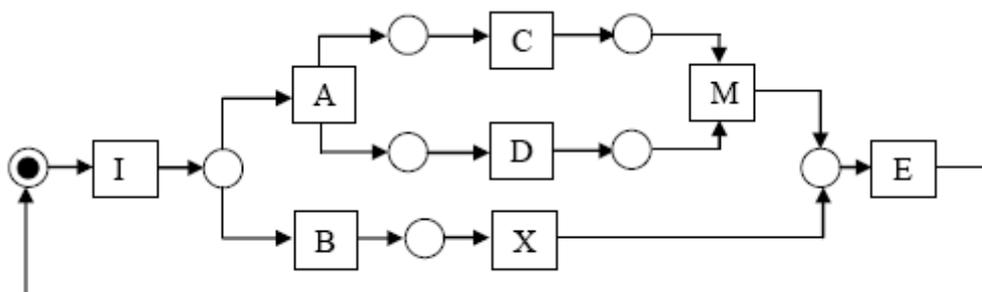


Figura 5. Red de Petri para el modelo de nuestro ejemplo.

datos o el aprendizaje automático. A pesar de ello, se pueden encontrar herramientas de minería de procesos tanto en el mundo académico (la mayor parte) como en la industria. La siguiente clasificación no es en absoluto exhaustiva, pero recoge algunas de las herramientas más relevantes que se pueden utilizar como herramientas de descubrimiento de procesos:

- **Mundo académico:** El ProM Framework, de la Universidad Técnica de Eindhoven (TU/e) es hoy día la herramienta de referencia. Es el resultado de una gran colaboración entre distintas universidades del mundo para reunir soporte algorítmico para la minería de procesos (por ejemplo no solo descubrimiento de procesos). Adicionalmente, distintos grupos han desarrollado algunas herramientas autónomas en el ámbito universitario que incorporan modernos algoritmos de descubrimiento de procesos.

- **Industria:** Algunas importantes empresas han invertido recursos en construir herramientas para el descubrimiento de procesos, por ejemplo Fujitsu (APD), pero también empresas de mediano tamaño y empresas emergentes que están más enfocadas hacia las prácticas de minería de procesos, por ejemplo Pallas Athena (ReflectOne), Fluxicon (Disco), Perspective Software (BPMOne, Futura Reflect), Software AG (ARIS Process Performance Manager), entre otras.

6. Retos

La tarea de descubrimiento de procesos puede dificultarse si aparecen alguno de los siguientes aspectos:

- **Registros de eventos incompletos:** El log de eventos contiene solo una fracción del total del comportamiento representativo del proceso. Por lo tanto, el algoritmo de descubrimiento de procesos necesita suponer parte del comportamiento que no está presente en el log de eventos, algo que en general es difícil.

- **Ruido:** El comportamiento registrado puede presentar a veces raras excepciones que no forman parte del proceso. Por lo tanto, los algoritmos de descubrimiento del proceso podrían obstaculizarse cuando aparece el ruido, p. ejp. en el descubrimiento del control de flujo algunas relaciones entre las actividades pueden resultar contradictorias. Separar el ruido de la información válida en un log de eventos es actualmente una línea de investigación.

- **Complejidad:** Debido a la magnitud de los actuales registros de eventos TI, es a menudo difícil utilizar algoritmos complejos que pueden requerir cargar el log de eventos en memoria para derivar el modelo de proceso, o aplicar técnicas cuya complejidad es no lineal al tamaño del log de eventos. En estos casos, estrategias de alto nivel (p. ejp., *divide y vencerás*) son la única posibilidad de derivar el modelo de proceso.

- **Visualización:** Incluso si el algoritmo de descubrimiento de procesos hace su trabajo y sabe cómo derivar un modelo de proceso, podría ser difícil para un humano entenderlo si hay más de un centenar de elementos (nodos, arcos). En estos casos, una descripción jerárquica, similar a la aplicación *Google Maps* donde se pueda ampliar o reducir parte del modelo, facilitará la comprensión de un modelo complejo de proceso.

Agradecimientos

Agradezco a David Antón la creación del dibujo del marciano de este artículo.

Referencias

- [1] S. Rogers. Data is Scaling BI and Analytics-Data Growth is About to Accelerate Exponentially - Get Ready. *Information and Management - Brookfield*, 21(5):p. 14, 2011.
- [2] D. Miers, S.A. White. BPMN Modeling and Reference Guide: Understanding and Using BPMN. *Future Strategies Inc.*, 2008. ISBN-10: 0977752720.
- [3] W. M. P. van der Aalst, H. Reijers, M. Song. *Discovering Social Networks from Event Logs. Computer Supported Cooperative Work*, 14(6):pp. 549-593, 2005.
- [4] A. Rozinat, W. M. P. van der Aalst. Conformance Checking of Processes Based on Monitoring Real Behavior. *Information Systems*, 33(1):pp. 64-95, 2008.
- [5] W.M.P. van der Aalst, A. Weijters, L. Maruster. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16 (9):pp. 1128-1142, 2004.
- [6] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011. ISBN-10: 3642193447.

Antonio Valle Salas

Socio Director de G2, Gobierno y Gestión de TI; Director del Comité de Catalunya en itSMF España

<avalle@gedos.es>

Posibilidades de uso de la minería de procesos en ITSM

1. Roles y responsabilidades en un modelo ITSM

Todos los modelos, estándares y marcos de referencia utilizados en el sector de la Gestión de Servicios TI (ITSM) están orientados a procesos. Esto es así porque la orientación a procesos aporta estructura y orden al trabajo realizado y permite formalizar gran cantidad de aspectos sobre las actividades que se realizan: qué actividades se deben llevar a cabo y cuándo, quién las debe realizar, qué responsabilidades hay definidas sobre las mismas, con qué herramientas se deben realizar estas actividades o cuáles son los objetivos y beneficios a largo plazo que obtiene la organización por la realización de este trabajo.

Un esquema empleado tradicionalmente para representar los diferentes componentes de un proceso es el denominado modelo ITOCO [1], en el cual se representan los elementos fundamentales de un proceso: las entradas, las salidas, las actividades realizadas y los parámetros de control (ver **figura 1**).

Este modelo nos permite diferenciar claramente tres tipos de rol necesarios para el correcto funcionamiento de cualquier proceso: los *operadores* de procesos, que son aquellos que se encargan fundamentalmente de la ejecución de las diferentes actividades, los *gestores* de procesos, que son los que se encargan de velar porque el proceso se ejecute de acuerdo a las especificaciones y de que tanto las entradas como las salidas estén de acuerdo a lo esperado (e indicado en los parámetros de control) y finalmente los *propietarios* de procesos que son los responsables de definir, con una visión más de gobernanza, cuáles son los objetivos a largo plazo, cuáles son esos parámetros de control que rigen la ejecución del proceso y de conseguir y asignar los recursos necesarios para la buena ejecución del mismo.

El trabajo del *gestor del proceso* consiste en ejecutar las actividades de control (o proceso de control) sobre el proceso gestionado, actuando sobre las desviaciones en la ejecución, sobre las variaciones en la calidad de los resultados, sobre los recursos (humanos, materiales, o de información) utilizados para la ejecución y sobre las capacidades y habilidades de las personas involucradas en la ejecución del proceso gestionado. Así, la actuación de este rol requiere una combinación de prácticas de auditoría, consultoría y sobre todo de actividades de mejora continua.

Resumen: En los sistemas de información, desde fotocopiadoras a equipos quirúrgicos pasando por los sistemas de gestión empresarial, lo habitual es que toda la información relacionada con los diferentes procesos que se llevan a cabo mediante esos sistemas quede almacenada en los logs o dietarios del sistema. El caso de los procesos empleados en la Gestión de Servicios TI (ITSM) no es diferente, y la gran mayoría de herramientas empleadas para realizar el seguimiento y control de estas actividades mantienen logs estructurados para facilitar la trazabilidad de las acciones. Sería interesante poder usar toda la información registrada en los logs para hacernos una idea exacta de cómo es realmente el proceso, poder comprobar de esta forma si el flujo real se corresponde con el diseñado en la teoría y analizarlo para mejorarlo de manera que sea más efectivo y eficiente. Esta es la principal función de la minería de procesos, en inglés process mining. Este artículo explora las diferentes capacidades que nos brinda la minería de procesos y los posibles usos que podemos darle bajo un entorno ITSM.

Palabras clave: gestión de cambios, gestión de incidencias, ITSM, minería de procesos, Service Desk, servicios.

Autor

Antonio Valle Salas es Socio Director de G2, Gobierno y Gestión de TI. Es Ingeniero Técnico en Informática de Gestión por la UPC (*Universitat Politècnica de Catalunya*) y posee tanto certificaciones metodológicas como *Certified Lean Service Professional* por la *Service Management Society*, *ITIL Service Manager* por EXIN, Auditor Certificado de Sistemas de Información (CISA) por ISACA y *COBIT Based IT Governance Foundations* por la *IT Governance Network* como certificaciones eminentemente técnicas en las herramientas de gestión de la familia *HP Openview*. Es Director Regional de Cataluña del *itSMF España*, combinando las actividades dentro del mundo de la consultoría y de la ejecución de proyectos con colaboraciones frecuentes tanto en actividades formativas en entornos universitarios como la UPC o la *Universitat Pompeu Fabra* como en el mundo editorial, donde ha colaborado en publicaciones como *IT Governance a pocket guide*, *Metrics in IT Service Organizations*, *Gestión de Servicios Una introducción a ITIL*, así como las traducciones al castellano de los libros *Soporte al Servicio y Provisión de Servicios ITIL V2* o en la edición en castellano de *Gestión Esencial de Servicios*.

2. La Gestión de Procesos en ITSM

En el sector ITSM la aproximación que se ha seguido tradicionalmente para realizar esta actividad de gestión de procesos ha sido la de utilizar una serie de herramientas metodológicas que faciliten la actividad del *gestor del proceso*:

- Definir indicadores.
- Definir cuadros de control sobre estos indicadores.
- Definir informes de seguimiento (diarios, semanales, mensuales).
- Realizar encuestas de satisfacción (a usuarios o a clientes y de forma parcial en función de los procesos gestionados).
- Realizar auditorías de cumplimiento (internas o externas).

Estas herramientas le permiten mantenerse informado al respecto del comportamiento del proceso que tiene a su cargo y le facilitan tomar decisiones que ayuden a "corregir el rumbo" de dicho proceso, pero en general adolecen de cierta rigidez a la hora de realizar un estudio más profundo de su comportamiento.

Aún así, dos aspectos clave de cualquier modelo de mejora continua son conocer cuál es la situación actual y comprender cuál es el impacto sobre él de las acciones de mejora, aspectos representados en la **figura 2**.

Es en estas etapas cuando se plantean multitud de cuestiones tales como las indicadas a continuación y para las que el *gestor* debe conseguir respuestas en su actividad diaria:

- ¿Cuál es el flujo más habitual?
- ¿Qué pasa en un determinado tipo de peticiones?
- ¿Cuánto tiempo están los diferentes casos en cada estado?
- ¿Se puede mejorar el flujo?
- ¿Dónde se atasca?
- ¿Qué actividades se repiten más?
- ¿Existe algún cuello de botella?
- Los operadores del proceso, ¿Siguen el proceso definido?
- ¿Hay segregación de funciones?

Adicionalmente, en ITSM nos encontramos con que la gran mayoría de los procesos definidos por los marcos de referencia no

“ El rol de gestor del proceso requiere una combinación de prácticas de auditoría, consultoría y sobre todo de actividades de mejora continua ”



Figura 1. Diagrama ITOCO.

acaban de encajar con las necesidades reales y del día a día: una aproximación estandarizada y rígida a los procesos no cubre las necesidades de aquellos tipos de actividad en los cuales los pasos a seguir no se conocen con antelación [2].

Un caso claro de este tipo de procesos en ITSM es el proceso de gestión de problemas, en el que para poder realizar la etapa de diagnóstico e identificación de causa(s) raíz(ces) el operador deberá decidir el siguiente paso en función de los resultados del análisis realizado en la etapa anterior. Así, nos encontramos con que el proceso de gestión de problemas es por propia naturaleza un proceso no estructurado cuyo comportamiento será totalmente diferente a un proceso estricto como pueda ser el proceso de gestión de peticiones.

3. Minería de procesos e ITSM

La primera tarea y la más delicada en la utilización de las técnicas de minería de procesos es la de obtener un *log* de buena calidad, representativo del proceso que estamos intentando analizar y con los suficientes atributos como para filtrar y orientar las tareas posteriores de análisis (ver figura 3) [3].

Afortunadamente, la gran mayoría de herramientas para la gestión de procesos ITSM disponen de logs que permiten la trazabilidad de las acciones realizadas por los diferentes

actores que intervienen, facilitando en gran medida la consecución de este tipo de trazas (ver figura 4). Estos *logs* suelen situarse en niveles de madurez IV ó V según las escalas propuestas en el Manifiesto sobre Minería de Procesos [3].

Las siguientes etapas de descubrimiento y representación son las que aportan un valor inmediato a la utilización de las técnicas de minería de procesos.

Habitualmente, los procesos diseñados están muy alejados de la ejecución real de las actividades. Esto se debe a diversos factores, entre los que influyen la generalidad en el diseño (para dar cobertura a procesos no estructurados), la flexibilidad de las herramientas (que suelen estar configuradas para permitir flujos libres en lugar de flujos cerrados) y la creatividad de los operadores, que no suelen aceptar de buena gana el sentirse encorsetados por una definición estricta de procesos.

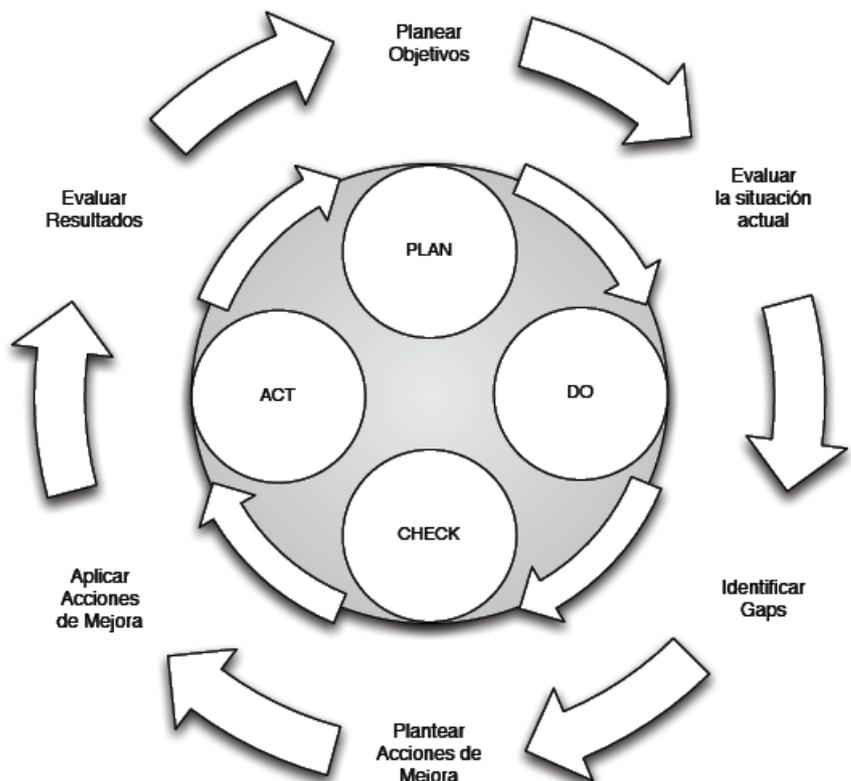


Figura 2. Ciclos de mejora continua.

“ La primera tarea y la más delicada en la utilización de las técnicas de minería de procesos es la de obtener un *log* de buena calidad, representativo del proceso que estamos intentando analizar ”

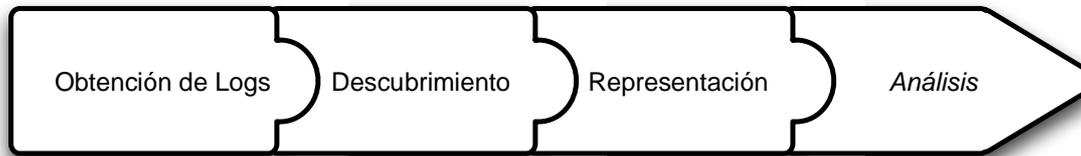


Figura 3. Fases de la minería de procesos.

Case ID	Activity	Complete Timestamp	Resource	Category	Priority
214371	Registrada	2012/04/16 13:59:51.000	Operador 136	Consulta	Media - 35
214371	Re-asignada	2012/05/02 09:25:19.000	Operador 30	Consulta	Media - 35
214371	Finalizada / Validación	2012/05/07 10:52:29.000	Operador 16	Consulta	Media - 35
214371	Cerrada	2012/05/08 09:29:39.000	Operador 136	Consulta	Media - 35
216141	Registrada	2012/04/27 13:59:16.000	Operador 136	Consulta	Media - 35
216141	En espera	2012/04/30 14:06:43.000	Operador 16	Consulta	Media - 35

Figura 4. Ejemplo de *log* de ejecución.

Por esta razón, normalmente el *propietario del proceso* y el gestor del proceso suelen tener de él una imagen idealizada, que les lleva a sorprenderse profundamente cuando tienen acceso por primera vez a una representación gráfica generada a partir del análisis de la información **real y completa** sobre su proceso.

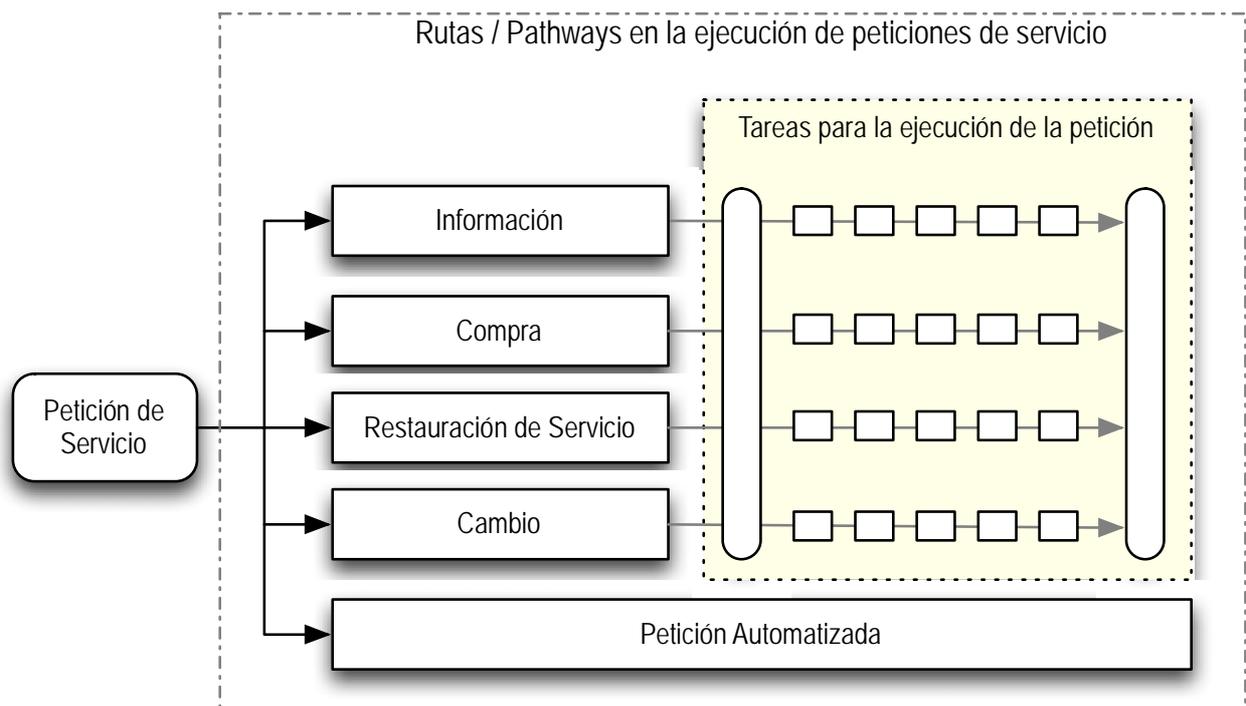
Por ejemplo, tal y como se menciona en USMBOK [4], los diferentes tipos de petición

que un usuario puede realizar a un Centro de Atención al Usuario irán a parar a un único concepto de *Service Request*, que posteriormente seguirá un flujo diferente o *Pathway*. Este flujo será "encajado" dentro de un flujo común en el módulo pertinente de la herramienta de gestión que esté utilizando el equipo de *ServiceDesk* (ver figura 5).

Para poder hacer encajar este amplio espectro de diferentes tipos de petición dentro de un

flujo relativamente general se suele huir de una definición cerrada del proceso y sus fases (en forma de autómata determinista), permitiendo un flujo abierto en el cual cada operador decide en cada momento cuál es el siguiente estado o etapa del ciclo de vida que corresponde (ver figura 6) [2].

Por eso, en el momento en que intentamos realizar un descubrimiento y representación de este tipo de actividades nos encontraremos



Fuente: USMBOK

Figura 5. El concepto *Pathway* según USMBOK.

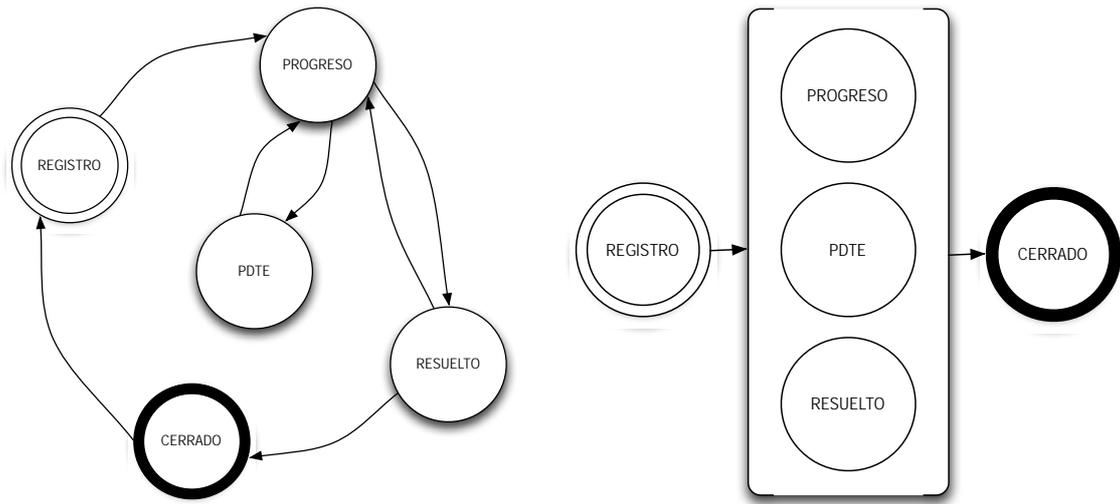


Figura 6. Flujo cerrado vs. flujo abierto.

con lo que llamamos habitualmente en la jerga de la minería de procesos un "modelo spaghetti" donde, incluso con un número reducido de casos, el volumen y la heterogeneidad de transiciones entre estados hace que el diagrama sea de escasa utilidad (ver figura 7).

Para facilitar el análisis es necesario utilizar técnicas que nos permitan dividir el problema en partes más pequeñas [5]. Podemos utilizar técnicas de *clustering* o bien simplemente filtrar el log original seleccionando el tipo de *pathway* que queremos analizar.

Previamente a las tareas de descubrimiento y representación, es conveniente enriquecer el log con toda la información que posteriormente nos permita segmentar el conjunto de datos según las diferentes dimensiones de análisis.

Por ejemplo, en este caso el disponer de un atributo que indique el tipo de petición o *Pathway* nos permitirá desglosar el modelo por peticiones, segmentando el conjunto de datos, y realizar el análisis de un tipo de petición en concreto (ver figura 8).

Por otra parte, es conveniente recordar que las técnicas de minería de procesos son independientes a la actividad realizada en el proceso: se centran en analizar cambios de estado. En este sentido, podemos ser creativos y pensar en el flujo de proceso como cualquier "cambio de estado dentro de nuestro sistema de información", de manera que podemos utilizar estas técnicas para analizar cualesquiera otras transiciones, como por ejemplo el flujo de asignación de tareas entre los diferentes actores o el flujo de escalado entre los diferentes grupos de especialidad, o aspectos más alejados del concepto de proceso como puedan ser las reclasificaciones o los cambios de prioridad de los tickets (ver figura 9).

Finalmente, en la etapa de análisis llega el momento de responder a las preguntas relativas al comportamiento del proceso. Para llevar a cabo esta labor disponemos de un amplio arsenal de herramientas:

■ **Enriquecimiento de la representación visual:** por ejemplo, en la figura 9 podemos observar que las transiciones entre operadores que más tiempo acumulan se representan con una línea más gruesa, o en la figura 8 vemos que los estados más frecuen-

tes se representan con un color más oscuro.

■ **Generación de gráficas e histogramas** que representan la información de volumetría o temporal de la que disponemos. Casos habituales de este tipo de análisis son la obtención de una gráfica que represente el número de casos abiertos a lo largo del tiempo (evolución del *backlog*) e histogramas que muestren la distribución del número de eventos que sufre cada caso o la distribución de la duración de los casos.

■ En terrenos más analíticos, podemos obtener un diagrama que represente una **Cadena de Márkov** para nuestro proceso, mostrando la probabilidad de que se produzca una determinada transición (para satisfacer preguntas del estilo "¿Cuál es la probabilidad de que un ticket cerrado se reabra?") pudiendo complementar esta información con los atributos del caso (objeto afectado, persona de contacto, tipo de petición, organización...) para enriquecer aún más el modelo de análisis (ver figura 10).

Hasta aquí hemos visto mecanismos y herramientas metodológicas que nos permiten un análisis cuantitativo y estadístico de los procesos y su comportamiento. Existe la otra

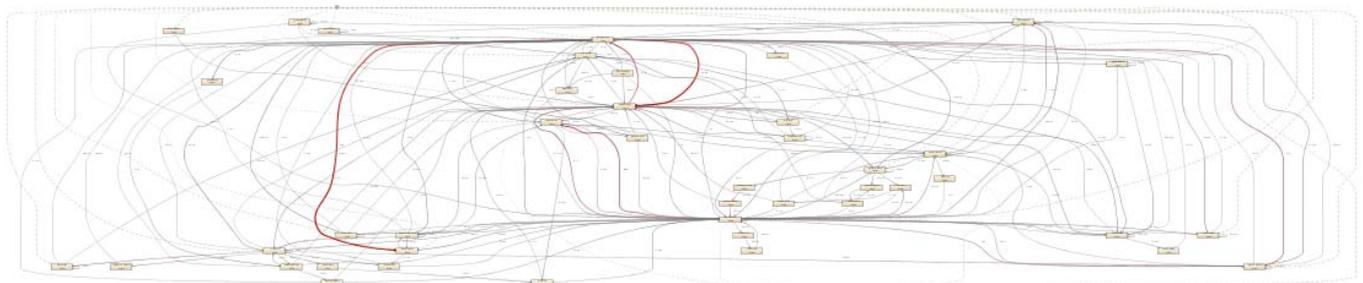


Figura 7. Modelo Spaghetti.

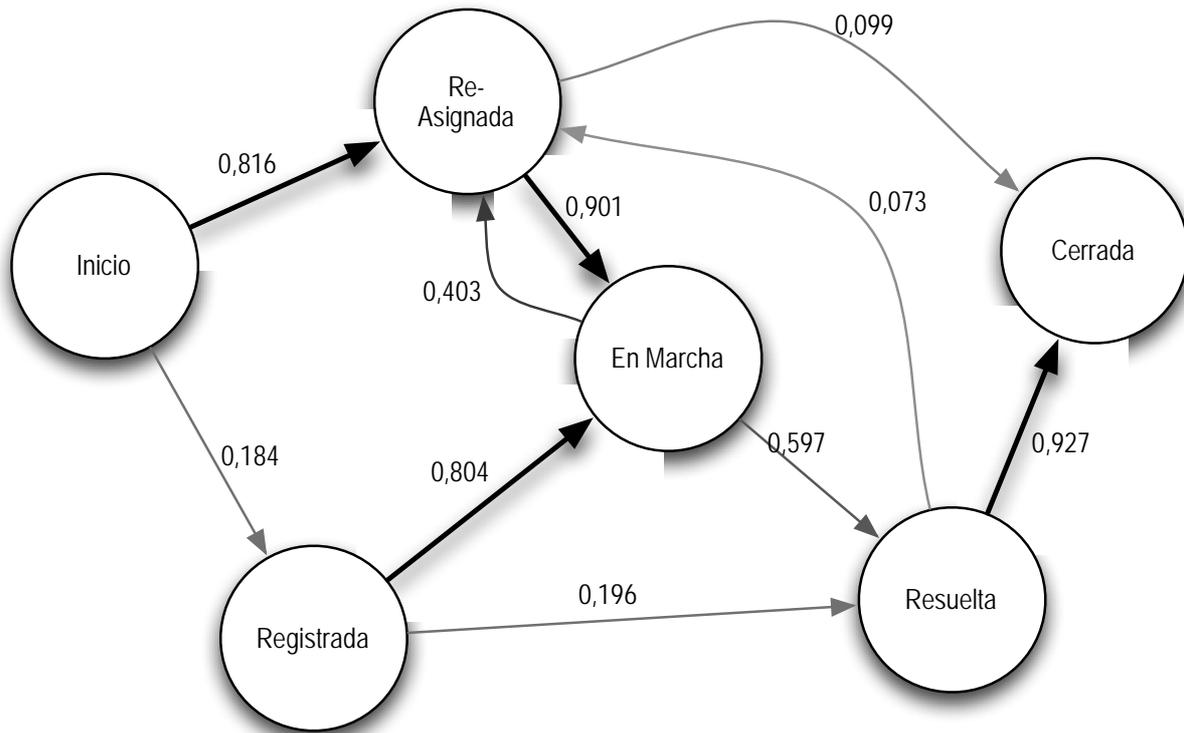


Figura 10. Cadena de Márkov simplificada.

nes al problema de la validación del cumplimiento, tal y como describe Anne Rozinat en su artículo *Conformance Checking of Processes Based on Monitoring Real Behavior* [7]:

■ Análisis del ajuste (*fitness*), que responde a la pregunta *¿El proceso observado cumple con el flujo especificado en el modelo de proceso?*

■ Análisis del grado de verosimilitud (*appropriateness*), respondiendo a la pregunta *¿El modelo de proceso describe adecuadamente el proceso observado?*

Sin embargo, calcular un índice de ajuste a un modelo concreto no será suficiente cuando estamos realizando actividades de análisis o de auditoría; en estos casos necesitaremos

disponer de mecanismos que nos permitan realizar consultas más o menos complejas sobre el log [8]. En estas situaciones, poder consultar en qué casos la actividad A se ejecutó antes que la B, o en qué casos el operador X ejecutó las actividades A y B serán de gran importancia para descubrir violaciones a las reglas de negocio o a las políticas que gobiernan la ejecución del proceso.

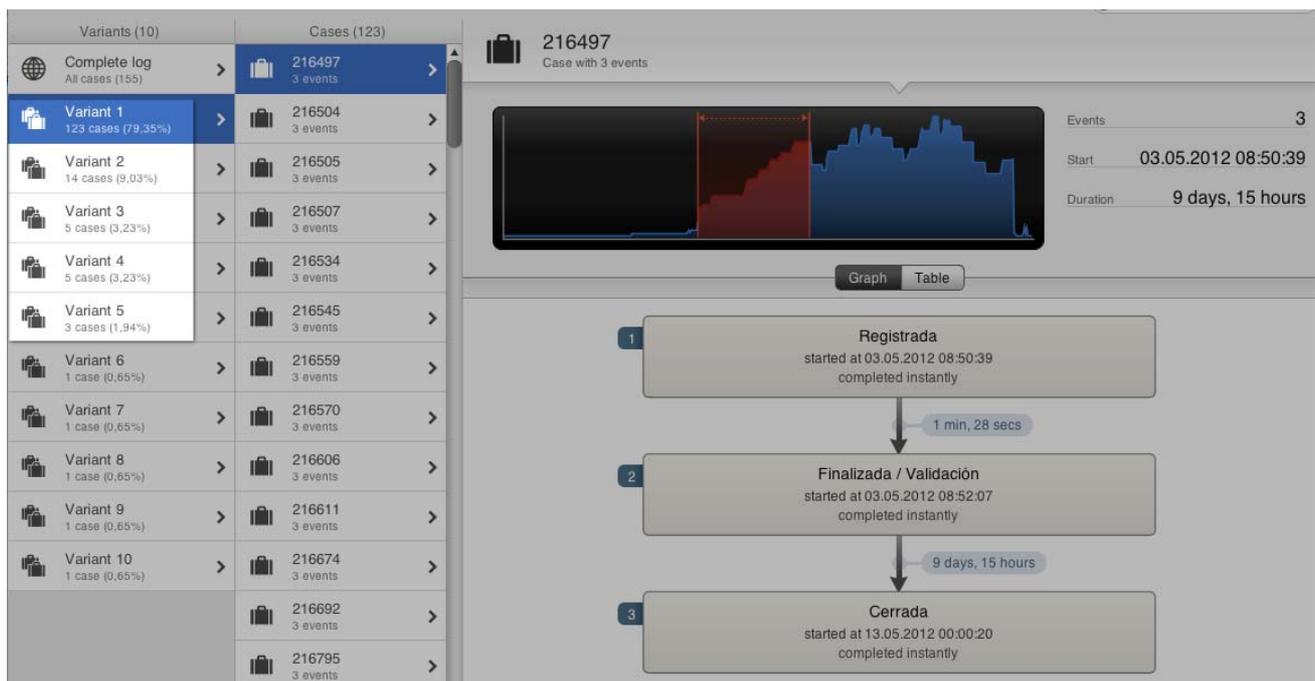


Figura 11. Variantes de un proceso.

“ En el momento en que intentamos realizar un descubrimiento y representación de este tipo de actividades nos encontraremos con lo que llamamos habitualmente en la jerga de la minería de procesos un "modelo spaghetti" ”

En nuestro caso de aplicación de estas técnicas a los procesos ITSM podemos ver aquí una aplicación interesante para asegurar la segregación de funciones en la Gestión de Cambios para aquellas organizaciones que deben cumplir con regulaciones del tipo de SOX. El siguiente paso es conseguir la monitorización de estas reglas de forma continua [9].

4. Conclusiones

La minería de procesos se presenta como un conjunto de herramientas que puede facilitar en gran medida las tareas a desempeñar por los gestores y propietarios de procesos, tanto en las labores de adquisición de la comprensión del comportamiento real del proceso como en las actividades de auditoría y mejora continua, facilitando muchas tareas de análisis que serían prácticamente imposibles o extremadamente costosas de llevar a cabo utilizando estrategias tradicionales como la generación de informes, *dashboards* e indicadores.

Aunque, en general, una de las principales dificultades que encontramos para la minería de procesos es la falta de *logs* o de información sobre la que aplicar los análisis, en el sector de la Gestión de Servicios TI esto no es un problema gracias a que las herramientas para la gestión de procesos ITSM mantienen *logs* susceptibles de ser utilizados en un análisis de minería de procesos, al tiempo que permiten definir campos auditables sobre los que generar trazas que pueden aportarnos diferentes perspectivas o dimensiones de análisis.

Por lo tanto, la minería de procesos se alza como una herramienta diferencial, totalmente adecuada y muy potente para apoyar a las prácticas ITSM en su búsqueda permanente de oportunidades de mejora tanto para los servicios TIC como para los procesos que componen el Sistema de Gestión.

Referencias

- [1] Jan van Bon. *IT Service Management Global Best Practices*, Volume 1. NL, NL: Van Haren Publishing, 2008.
- [2] Rob England. *Plus! The Standard + Case Approach*. Wellington, NZ: CreateSpace, 2013.
- [3] IEEE Task Force on Process Mining. *Manifiesto sobre Minería de Procesos* (versión final), agosto 2011. <<http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:pmm-spanish-v1.pdf>>.
- [4] Ian M. Clayton. *USMBOK - The Guide to the Universal Service Management Body of Knowledge*. CA, US: Service Management 101, 2012.
- [5] Marco Aniceto Vaz, Jano Moreira de Souza, Luciano Terres, Pedro Miguel Esposito. *A Case Study on Clustering and Mining Business Processes from a University*, 2011.
- [6] Wil M.P. van der Aalst et al. *Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor*, 2010. <<http://bpmcenter.org/wp-content/uploads/reports/2010/BPM-10-07.pdf>>.
- [7] Anne Rozinat, W.M.P. van der Aalst. *Conformance Checking of Processes Based on Monitoring Real Behavior*, 2008. <<http://www.win.tue.nl/~wvdaalst/publications/p436.pdf>>.
- [8] W.M.P. van der Aalst, H.T. de Beer, B.F. van Dongen. *Process Mining and Verification of Properties: An Approach based on Temporal Logic*, 2005.
- [9] Linh Thao Ly, Stefanie Rinderle-Ma, David Knuplesch, Peter Dadam. *Monitoring Business Process Compliance Using Compliance Rule Graphs*, 2011. <<http://dbis.eprints.uni-ulm.de/768/1/paper.pdf>>.

Arjel Bautista, Lalit Wangikar,
Syed M. Kumail Akbar
CKM Advisors, New York (EEUU)

<{abautista,lwangikar,sakbar}@ckmadvisors.com>

Optimización dirigida por minería de procesos de un proceso de aprobación de préstamos al consumo

1. Introducción

A medida que el rol del Big Data se vuelve prevalente en esta era dirigida por la información [1][2][3], los negocios de todo el mundo buscan constantemente formas de aprovechar estos recursos potencialmente valiosos. El 2012 *Business Processing Intelligence Challenge* ("Desafío de Inteligencia de Procesamiento de Negocio", BPIC 2012) es un ejercicio de análisis de uno de esos conjuntos de datos utilizando una selección de herramientas comerciales, propietarias y de código abierto, y su combinación con percepciones creativas, para entender mejor el rol de la minería de procesos en el lugar de trabajo actual.

1.1. Enfoque y alcance

La situación representada en BPIC 2012 se enfoca en los procesos de aprobación de préstamos y descubiertos de una institución financiera real en los Países Bajos. En nuestro análisis de esta información hemos buscado entender en mayor detalle y a múltiples niveles de granularidad los procesos de negocio subyacentes. También hemos buscado identificar oportunidades para mejorar la eficiencia y efectividad del proceso global.

Específicamente, intentamos investigar en detalle las siguientes áreas:

- Desarrollar una comprensión minuciosa de los datos y del proceso subyacente.
- Entender las actividades y los puntos de decisión críticos.
- Mapear el ciclo de vida de una solicitud de préstamo desde el inicio hasta su resolución definitiva.
- Identificar diferencias en el desempeño a nivel de recursos y oportunidades para intervenciones en los procesos.

Siendo como somos recién llegados a la minería de procesos, en CKM Advisors, queríamos usar esta oportunidad para poner en práctica lo aprendido en esta disciplina. También intentamos combinar las herramientas de minería de procesos con métodos analíticos tradicionales, para construir un cuadro más completo. Estamos convencidos de que al adquirir experiencia nuestro enfoque se volverá más refinado y cada vez más conducido por métodos desarrollados específicamente para la minería de procesos.

Intentamos ser tan amplios como fuera posible en nuestro análisis y profundizar donde pudiésemos. Si bien hemos realizado un aná-

Traducción: Eduardo Rodríguez Ringach (TOGAF, CGEIT, CRISC)

Resumen: Un registro de eventos (262.200 eventos; 13.087 casos) del proceso de aprobación de préstamos y descubiertos de un banco en los Países Bajos fue analizado usando diversas técnicas analíticas. Mediante una combinación de enfoques basados en hojas de cálculo, capacidades de minería de procesos y analítica exploratoria, examinamos los datos en gran detalle y a múltiples niveles de granularidad. Presentamos nuestros hallazgos sobre como desarrollamos un profundo entendimiento del proceso, evaluamos áreas potenciales de mejora de eficiencia e identificamos oportunidades para realizar predicciones basadas en conocimiento sobre el resultado final de una solicitud de préstamo. También discutimos desafíos peculiares del trabajo con tales datos, y oportunidades para mejorar el impacto de dichos análisis mediante la incorporación de elementos de datos adicionales.

Palabras clave: Analítica de datos, Big Data, inteligencia de procesos de negocio, minería de procesos.

Autores

Arjel Bautista es consultor de CKM Advisors, involucrado en el desarrollo de técnicas innovadoras de reingeniería de procesos y de investigación analítica en la empresa. En sus proyectos, ha desplegado una combinación de herramientas de minería de datos de última generación y el análisis estratégico tradicional para resolver diversos problemas relacionados con los procesos de negocio. También ha desarrollado estrategias para el análisis de texto no estructurado y otras fuentes de datos no tradicionales. Arjel ha obtenido su maestría y doctorado en química por la Universidad de Yale y un título de grado en Bioquímica por la UC San Diego.

Lalit Wangikar es socio de CKM Advisors. Como consultor, ha asesorado a clientes en especial en el sector de servicios financieros y las industrias de seguros y servicios de pago. Su área principal de especialización es el uso de *Big Data* y Analítica para conducir el impacto de negocio a través de todas las áreas empresariales clave como marketing, riesgo, operaciones y cumplimiento normativo. Ha trabajado con clientes en Norteamérica, Reino Unido, Singapur e India. Antes de asociarse a CKM Advisors, Lalit fue responsable de la práctica de Analítica de Decisiones para EXL Service / Inductis. Con anterioridad trabajó como consultor para Deloitte Consulting y Mitchell Madison Group, donde asesoró a clientes en los sectores verticales de banca y mercados de capitales.

Syed M. Kumail Akbar es consultor de CKM Advisors donde forma parte del equipo de Analítica, y colabora en minería de datos, mapeo de procesos y analítica predictiva. Previamente ha trabajado en proyectos de estrategia y operaciones en la industria de servicios financieros. Antes de unirse a CKM, trabajó como asistente de investigación en el Centro de Análisis Cuantitativo así como en el Departamento de Física de la Universidad Wesleyan. También fue cofundador de Possibilities Pakistan, una Organización No-Gubernamental dedicada a proveer acceso a orientación vocacional universitaria para los estudiantes de instituto en Pakistán. Syed se ha graduado en Física y Matemática-Economía por la Universidad de Wesleyan.

lisis detallado en unas pocas áreas, no hemos cubierto en nuestro análisis todas las posibles áreas de la minería de procesos. La determinación de las áreas que no hemos cubierto (por ejemplo, el análisis de redes sociales) se debe exclusivamente a nuestra propia comodidad y familiaridad con la materia en cuestión y no necesariamente por una limitación de los datos.

2. Materiales y métodos

2.1. Entendiendo los datos

Los datos capturan eventos de procesos para

13.087 solicitudes de préstamo/descubierto en un período de seis meses, entre octubre de 2011 y marzo de 2012. El registro de eventos está compuesto por un total de 262.200 eventos para estos casos, comenzando con un cliente que envía una solicitud y que culmina con la finalización de esa solicitud en un evento de aprobación, cancelación o rechazo. Cada solicitud tiene un único atributo, AMOUNT_REQ, que indica la cantidad pedida por el solicitante. Para cada evento el extracto muestra el tipo de evento, la etapa del ciclo de vida (Programado, Iniciado, Finali-

Tipo	Descripción
“A_” Eventos de solicitud	Se refiere a estados de la propia solicitud. Cuando un cliente inicia una solicitud, personal del banco realiza el seguimiento para completar la solicitud si es necesario y para facilitar la toma de decisiones sobre las solicitudes.
“O_” Eventos de oferta	Se refiere a estados de una oferta comunicada al cliente.
“W_” Eventos de trabajo	Se refiere a estados de elementos de trabajo que ocurren durante el proceso de aprobación. Estos eventos capturan la mayoría del esfuerzo manual llevado a cabo por los recursos del banco durante el proceso de aprobación de la solicitud. Los eventos describen esfuerzos realizados durante diversas etapas del proceso de solicitud: <ul style="list-style-type: none"> – <i>W_Afhandelen leads</i>: Seguimiento de presentaciones iniciales incompletas. – <i>W_Completeren aanvraag</i>: Compleción de solicitudes pre-aprobadas – <i>W_Nabellen offertes</i>: Seguimiento tras la transmisión de ofertas a solicitantes cualificados – <i>W_Valideren aanvraag</i>: Evaluación de la solicitud – <i>W_Nabellen incomplete dossiers</i>: Búsqueda de información adicional durante la fase de evaluación – <i>W_Beoordelen fraude</i>: Investigación de casos sospechosos de fraude – <i>W_Wijzigen contractgegevens</i>: Modificación de contratos aprobados

Tabla 1. Nombres y descripciones de eventos.

zado), un indicador de recurso y el momento de finalización.

Los eventos describen pasos a lo largo del proceso de aprobación y son clasificados en tres tipos principales. La **tabla 1** muestra los tipos de eventos y nuestra comprensión de su significado

En sí mismo, el registro de eventos es una masa complicada de información de la cual resulta complicado inferir conclusiones lógicas. Por consiguiente, como otros investigadores han señalado [4][5], es menester someter al registro a cierto grado de pre-procesamiento para reducir su complejidad global, realizar conexiones visuales entre los pasos incluidos y auxiliar manualmente al análisis y la optimización de los conceptos de negocio. Aunque recibimos un registro de eventos rigurosamente pre-procesado que podía ser analizado sin inconvenientes usando herramientas de minería de procesos, procesamos todavía más los datos para construir extractos personalizados para diversos propósitos analíticos.

2.2. Herramientas usadas para el análisis

■ **Disco**: Adquirimos una versión de evaluación de Disco 1.0.0 (Fluxicon) y la utilizamos para exportar los datos a formatos adecuados para el análisis de hojas de cálculo. Disco fue especialmente útil para facilitar la visualización de flujos de proceso y excepciones típicas.

■ **Microsoft Excel**: Usamos Excel 2010 (Microsoft) para una exploración posterior más profunda en los datos pre-procesados. Excel ayudó especialmente para realizar funciones matemáticas básicas y avanzadas y para ordenar los datos, dos capacidades que destacan por su ausencia en la aplicación Disco.

■ **CART**: Usamos una versión de evaluación de la implementación de CART (Salford Systems) para llevar a cabo el análisis de segmentación preliminar de las solicitudes de préstamo con el fin de evaluar oportunidades de priorización del esfuerzo de trabajo.

3. Entendiendo el proceso en detalle

3.1. Simplificación del registro de eventos

Tras obtener el registro de eventos BPIC 2012, intentamos en primer lugar la reducción de su complejidad global identificando y removiendo eventos *redundantes*. Para los propósitos de este análisis, se considera que un evento es redundante si ocurre en forma concurrente con otro evento o a continuación de otro evento, de forma tal que el intervalo entre ambos sea mínimo (no más de dos segundos) respecto del marco temporal del caso completo.

El análisis inicial de los datos en crudo de *Disco* reveló un total de 4.366 variantes de orden de eventos entre los 13.087 casos representados. Conjeturamos que la remoción de hasta una sola secuencia de eventos

redundantes podía resultar en una reducción significativa del número de variantes. Esta simplificación se acentúa cuando el número de variantes removidas se multiplica por otras que ocurren más allá del evento inicial.

Además eliminamos dos tipos de eventos O-type (O_CANCELLED and O_DECLINED) que ocurren simultáneamente con A_CANCELLED y A_DECLINED, respectivamente.

No consideramos para su remoción eventos de tipo W-type ya que las fases de transición son cruciales para el cálculo de tiempo de trabajo dedicado a cada caso. Una vez removidos los eventos redundantes del registro de eventos, el número de variantes se redujo a 3.346, una mejora de casi el 25% respecto del conjunto de datos no filtrado.

Dicha consolidación puede ayudar en la simplificación de datos de proceso y facilitar un análisis más rápido. La complejidad de las variantes podría reducirse más si se entrevistase a los expertos de proceso del banco como ayuda para consolidar eventos que ocurren en forma conjunta y variantes secuenciales que no son críticas para el análisis del negocio.

3.2. Determinación del flujo de caso estándar

A continuación buscamos determinar el flujo de caso estándar para una solicitud exitosa contra el cual pudieran compararse todos los otros casos. Lo hicimos descargando el proyecto simplificado en Disco y filtrando todos los casos para el atributo A_APPROVED. Fijamos a continuación al nivel más riguroso (0%) y los umbrales tanto de actividades como de caminos, lo que resultó en una representación idealizada del camino desde la presentación hasta la aprobación (ver **figura 1**).

3.3. Entendiendo los resultados de las solicitudes

Antes de embarcarnos en una revisión más detallada de los datos, entendimos que era necesario definir los resultados finales para todas las 13.087 solicitudes. Usando el flujo de caso estandarizado (ver **figura 1**) determinamos que todas las aplicaciones están sujetas a una de cuatro *suertes* en cada etapa del proceso de aprobación:

■ **Avance a la próxima etapa**: La solicitud avanza a la siguiente etapa del proceso.

■ **Aprobada**: Las solicitudes que son aprobadas y donde el cliente ha aceptado la oferta del banco son consideradas exitosas y son etiquetadas como Aprobadas, con el punto final representado por el evento A_APPROVED.

■ **Cancelada**: La solicitud es cancelada por el banco o por pedido del cliente. Las solicitudes canceladas tienen un punto final de A_CANCELLED.

“ Tras obtener el registro de eventos BPIC 2012, intentamos en primer lugar la reducción de su complejidad global identificando y removiendo eventos redundantes ”

Eventos Redundantes	Ocurrencia
A_PARTLYSUBMITTED	Inmediatamente a continuación de A_SUBMITTED en todos los 13.087 casos
O_SELECTED O_CREATED	Ambos en rápida sucesión antes de O_SENT para los 5.015 casos que fueron seleccionados para recibir ofertas. En algunos casos, O_CANCELLED (974 instancias), A_FINALIZED (2.907 instancias) o <i>W_Nabellen offeries-SCHEDULE</i> (1 instancia) ocurren entre O_SELECTED y O_CREATED en el proceso de creación de oferta.
O_ACCEPTED A_REGISTERED A_ACTIVATED	Ocurren los tres, en orden aleatorio, con A_APPROVED para las 2.246 solicitudes exitosas. En ciertos casos, O_ACCEPTED está intercalado entre estos eventos.

Tabla 2. Redundancias potenciales del registro de eventos.

■ **Denegada:** Luego de evaluar al solicitante se considera que este no es apto para recibir el préstamo o descubierto pedido. Las aplicaciones denegadas tienen un punto final de A_DECLINED.

Nos aprovechamos del algoritmo de filtrado de Disco para definir un conjunto de posibles comportamientos de punto final. Se clasificó a 299 casos como *no* resueltos porque estaban en curso en el momento en que se capturaron los datos (es decir, no contenían puntos finales de A_DECLINED, A_CANCELLED o A_APPROVED).

La figura 2 muestra un flujo de proceso de alto nivel que marca como se disponen los casos en cada uno de los pasos de proceso clave. Este análisis nos ofrece percepciones útiles sobre el impacto general en el negocio de este proceso así como el flujo de caso global a través de los pasos de proceso críticos.

Observamos varias características de desempeño de línea base en la figura 2:

- ~26% de las solicitudes se rechazan instantáneamente (3.429 de las 13.087); lo que indica criterios estrictos de selección para mover una solicitud más allá del punto de partida.
- ~24% de las restantes (2.290 de las 9.658) se rechazan después del seguimiento inicial de

la oportunidad, indicando la existencia de un proceso continuo de selección de riesgo.

■ 754 de las 3.254 solicitudes que pasan a la etapa de validación (~23%) son rechazadas, indicando posibilidades para hacer más estricto el escrutinio directo en las etapas de solicitud u oferta.

4. Evaluación del desempeño del proceso

4.1. Análisis a nivel de caso

4.1.1. Punto final de caso vs. duración general

Para intentar evaluar cómo cambia la suerte de un caso particular respecto de la duración global, representamos gráficamente la relación entre estas dos variables y la presentamos superpuesta con el monto acumulado de tiempo de trabajo sobre la vida de estos casos. Excluimos los 3.429 casos que son rechazados instantáneamente al recibirse la presentación inicial de la solicitud porque no se gasta ningún esfuerzo en ellos. De este modo, nos propusimos visualizar el punto a partir del cual ejercer esfuerzo adicional rinde un retorno mínimo o nulo en forma de solicitudes completadas (cerradas).

La figura 3 muestra una vista del ciclo de vida de todas las solicitudes, indizada en el momento de la presentación. Como se muestra en la figura, dentro de los primeros siete días las solicitudes siguen avanzando o son

rechazadas. En el Día 7, el número de casos aprobados comienza a elevarse, sugiriendo que este es el número mínimo de días requerido para completar los pasos en el flujo de caso estándar (ver figura 1). Las aprobaciones continúan hasta ~Día 23, cuando más del 80% de todos los casos que finalmente son aprobados han sido cerrados y registrados. Hay un salto significativo en el número de solicitudes canceladas en el Día 30, ya que se cancelan los casos inactivos donde no se ha recibido respuesta del solicitante después de empantanarse en las etapas de cuello de botella *Completeren aanvraag* o *Nabellen offeries*, probablemente siguiendo las políticas del banco.

Esto plantea la pregunta interesante de cuándo debería el banco detener los esfuerzos proactivos para convertir una solicitud en un préstamo, y si el banco debiera tratar a los clientes de distinta manera según los comportamientos que indican la verosimilitud de una aprobación final. Por ejemplo, el banco realiza un esfuerzo adicional de 380+ personas-día entre los Días 23 y 31, solo para terminar cancelando la mayoría de los casos pendientes al final del periodo. Con datos adicionales sobre la rentabilidad o el valor total durante la relación comercial con el cliente, se podría determinar un punto óptimo en el proceso a partir del cual no se obtendrá un valor positivo del esfuerzo adicional en los casos que no

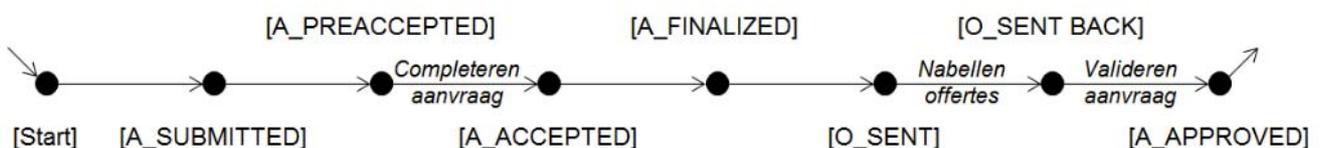


Figura 1. Flujo de Caso Estandarizado para las Solicitudes Aprobadas.

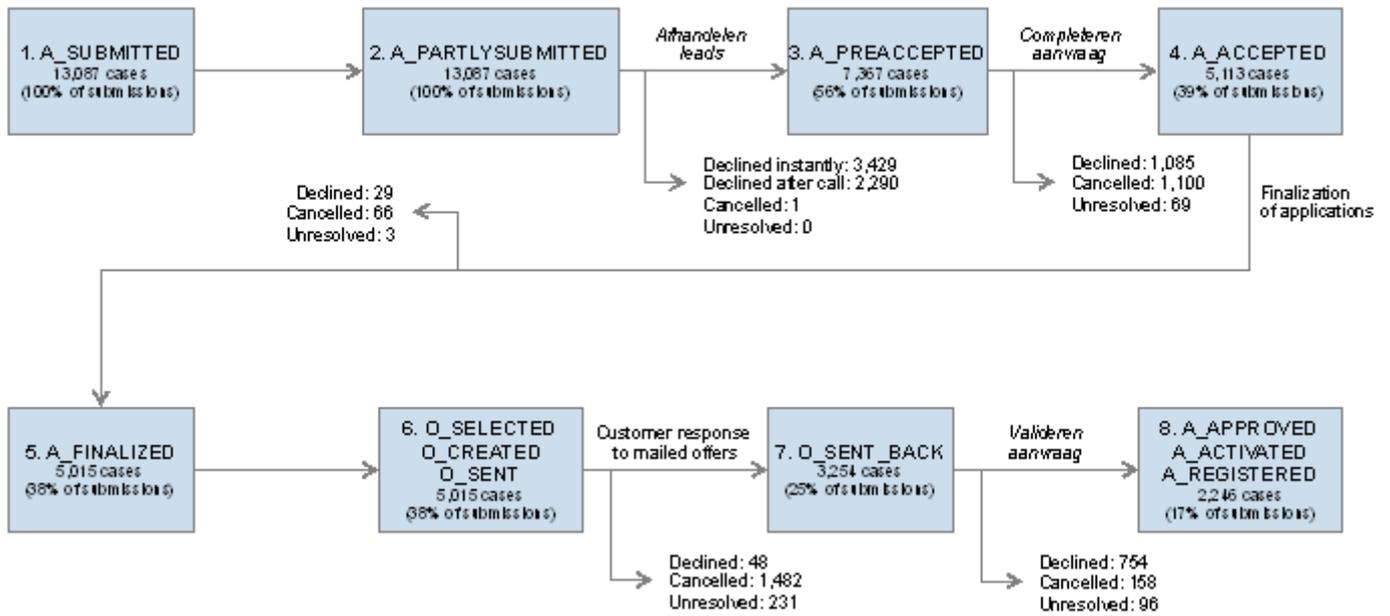


Figura 2. Pasos de proceso clave y flujo de volumen de solicitudes.

hayan alcanzado una cierta etapa.

4.1.2. Segmentación de casos por importe solicitado

Como cada caso está asociado con un im-

porte pedido por el solicitante, encontramos adecuado colocarlos en segmentos de aproximadamente el mismo número, ordenados por el valor total solicitado. En primer lugar quitamos los casos instantáneamente rechaza-

zados filtrándolos con Disco, ya que son resueltos en forma inmediata al ser presentados y no requieren ningún esfuerzo o pasos del proceso adicionales. Los 9.658 casos restantes (que incluye a los que están en curso)

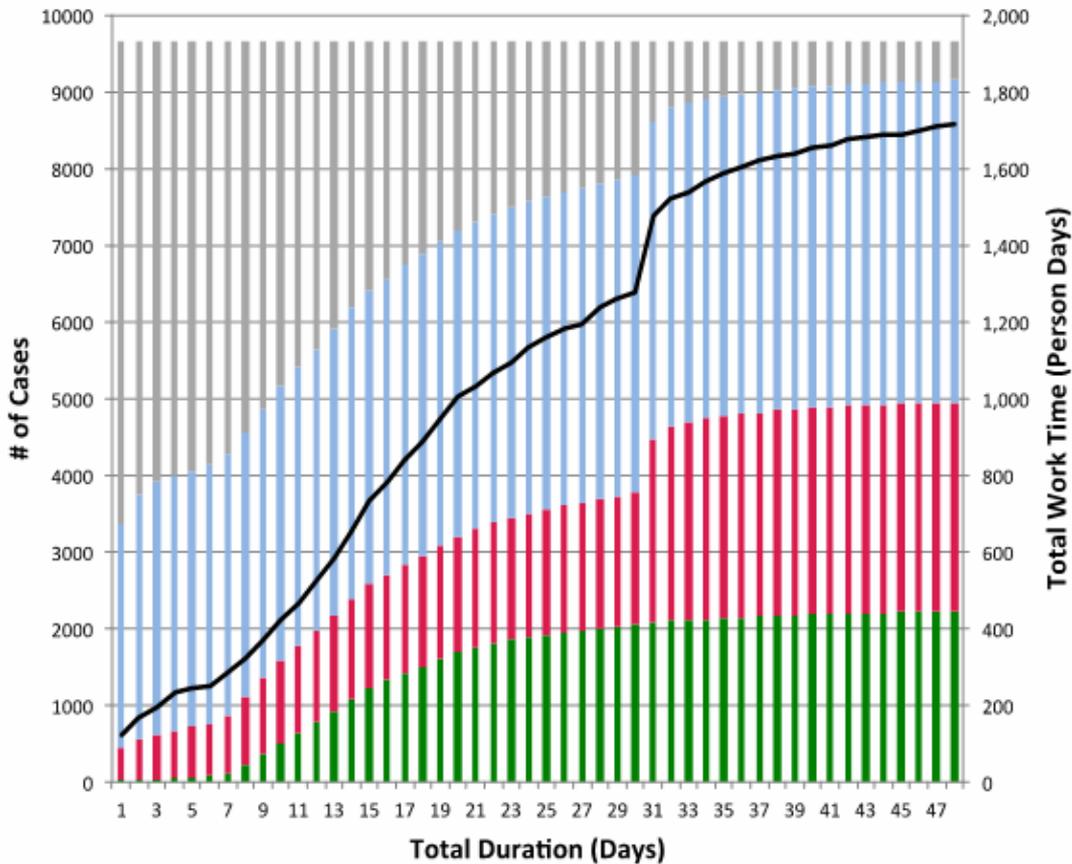


Figura 3. Distribución de casos por el resultado final y la duración, con esfuerzo de trabajo acumulado. Gris: Restantes en Curso, Azul: Rechazados Acumulados, Rojo: Cancelados Acumulados, Verde: Aprobados Acumulados. Excluye 3.472 Casos Instantáneamente Rechazados.

“ Nos propusimos visualizar el punto a partir del cual ejercer esfuerzo adicional rinde un retorno mínimo o nulo en forma de solicitudes completadas (cerradas)”

fueron luego divididos en deciles de unos 965-966 casos cada uno. Cada decil fue segmentado a su vez al clasificar los casos de acuerdo con su resultado final, y se examinaron las tendencias resultantes buscando correlaciones entre el porcentaje de aprobación con los montos pedidos (ver figura 4).

Inmediatamente observamos los porcentajes más altos de aprobación en los deciles 3 y 6, cuyos casos contenían rangos pedidos de 5.000-6.000 y 10.000-14.000, respectivamente. La razón exacta de este patrón no está clara; sin embargo especulamos que los solicitantes típicos eligen a menudo un número "redondo" en el que basar sus solicitudes (de hecho esto se refleja en los valores pedidos más frecuentes en el conjunto de datos: 5.000, 10.000 y 15.000). Quizás un cierto cambio de umbral de riesgo en el proceso de aprobación del banco cause un cambio abrupto en los porcentajes de aprobación.

4.2. Análisis a nivel de evento

4.2.1. Cálculo de la duración del evento

Buscamos conseguir una comprensión detallada de las actividades de trabajo embebidas en el proceso de aprobación, específicamente de aquellas que contribuyen con una cantidad significativa de tiempo o recursos a la resolución. El formato de los datos disponibles en este caso no se adecuaba fácilmente a este análisis. Usamos Excel para manipular los datos a nivel de evento tal como venían presentados y definimos el tiempo de trabajo (presumiblemente el esfuerzo real dedicado por recursos humanos) para cada evento como la duración de inicio a fin (transiciones START / COMPLETE, respectivamente). En cambio, el tiempo de espera fue definido como la latencia entre la programación del evento y el comienzo (SCHEDULE/START) o el tiempo transcurrido entre dos instancias de un único tipo de actividad así como entre el COMPLETE de un evento y el START de otro.

Como se muestra en la tabla 3, dos actividades *Completeren aanvraag* y *Nabellen Offertes*, contribuyen con un monto signifi-

cativo al tiempo total del caso según se lo representa en el registro de eventos. El tiempo de espera acumulado que se atribuye a cada uno de estos dos eventos puede llegar a ser tan alto como 30+ días por caso, ya que presumiblemente el banco realiza numerosos intentos de contactar al cliente antes de tener éxito. Mirando los datos con mayor detenimiento, nos dimos cuenta que el banco intenta contactar al cliente varias veces por día hasta el Día 30 para completar la solicitud, así como para el seguimiento de las ofertas que han sido realizadas pero que no han sido respondidas.

4.2.2. Actividades iniciales vs. actividades de seguimiento

El tiempo de trabajo promedio que se dedica a cada evento cambia si el banco lo realiza por primera vez o si se trata del seguimiento de un paso anterior en un caso particular (ver figura 5). Algunas diferencias en instancias iniciales y de seguimiento son mínimas (*Valideren aanvraag*), mientras que otras son más pronunciadas (*Beoordelen fraude*). En

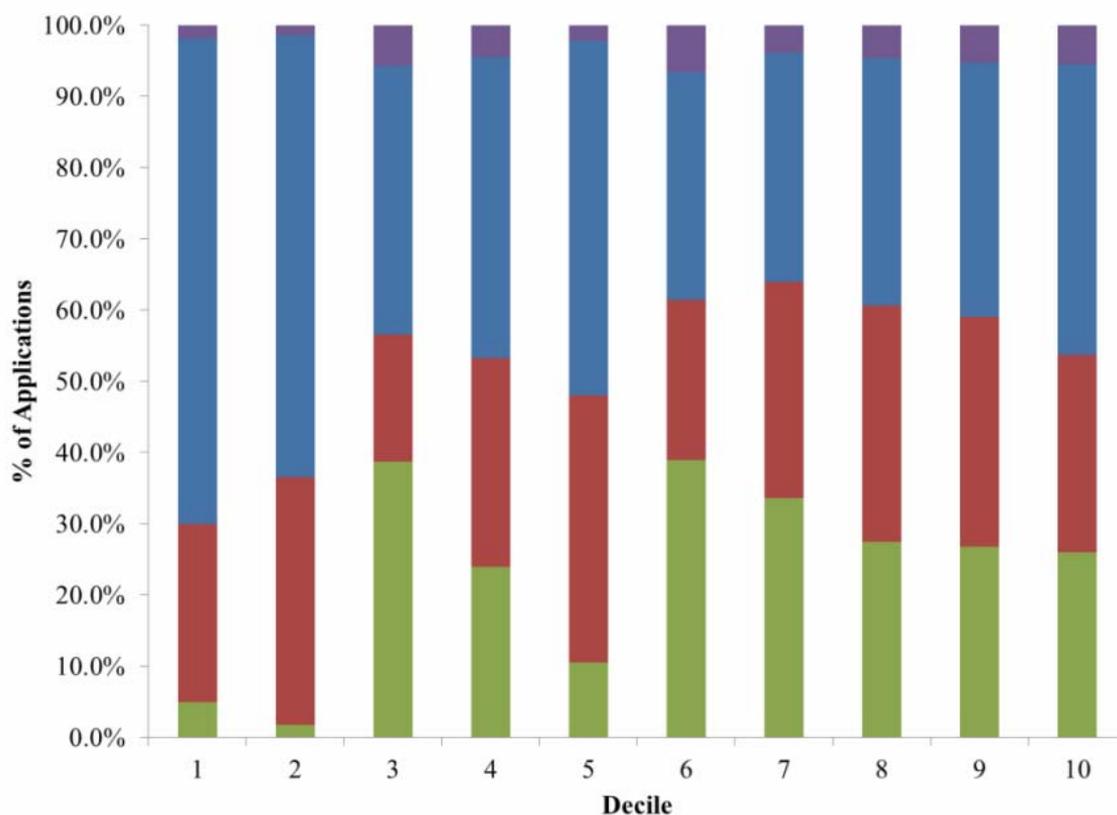


Figura 4. Puntos finales de los casos (eje izquierdo), segmentados por importe pedido por el solicitante. Verde: Aprobado, Rojo: Cancelado, Azul: Rechazado, Violeta: En Curso.

“ Estos resultados sugieren que una oficina de especialistas realizando actividades únicas puede ser más adecuada que un ejército de recursos asignado con una miríada de tareas ”

	<i>Afhandelen Leads</i>	<i>Beoordelen Fraude</i>	<i>Completeren aanvraag</i>	<i>Nabellen Offertes</i>	<i>Nabellen Incomplete Dossiers</i>	<i>Valideren Aanvraag</i>
Tiempo de Trabajo:						
Aprobado	13.659	23	45.909	68.473	89.204	121.099
Cancelado	14.601	2	119.497	94.601	25.633	7.775
Rechazado	67.560	2.471	63.052	30.870	26.993	29.946
Tiempo de Espera:						
Aprobado	198.916	8.456	1.873.537	34.972.224	5.980.887	10.537.938
Cancelado	300.062	28.763	16.582.465	42.630.195	2.006.774	678.105
Rechazado	986.421	236.115	3.294.367	13.542.054	1.001.354	3.227.252

Tabla 3. Trabajo total y tiempo de espera por tipo de evento (en minutos).

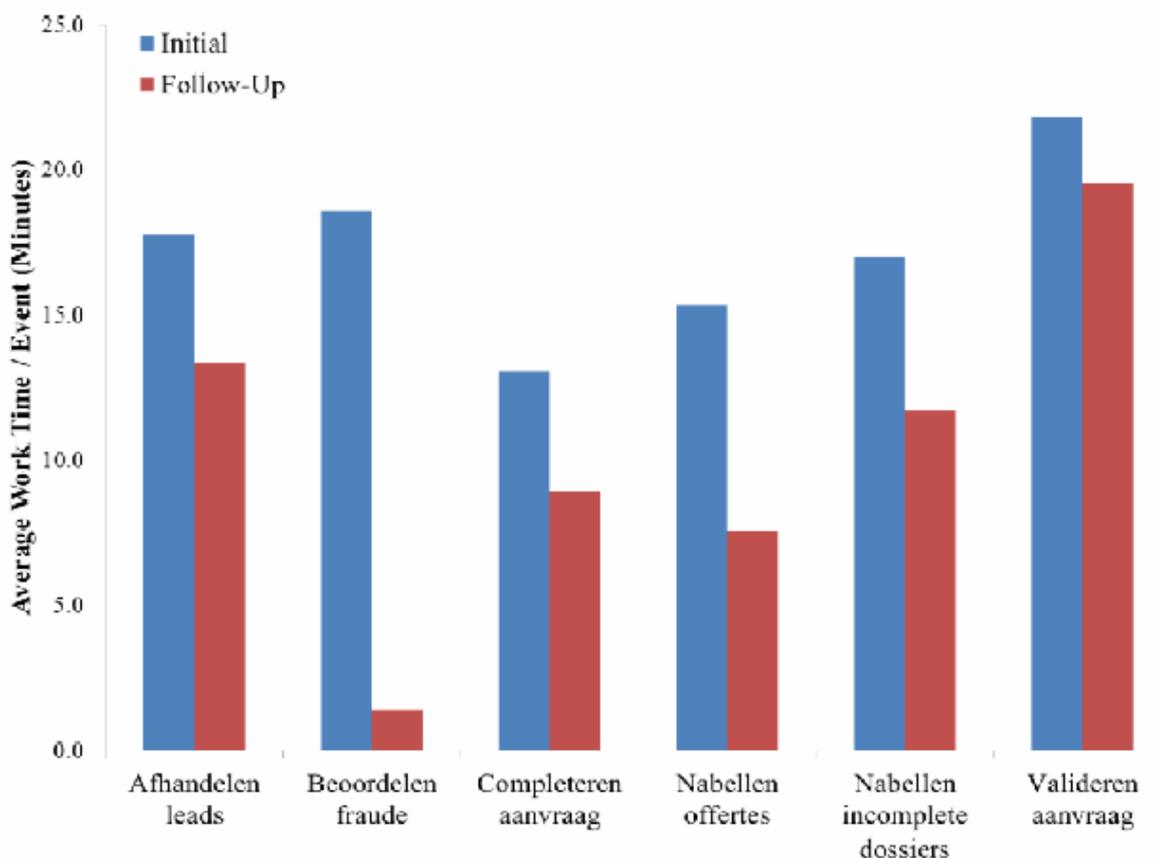


Figura 5. Comparación de tiempos de trabajo promedio, instancias de eventos iniciales vs. de seguimiento.

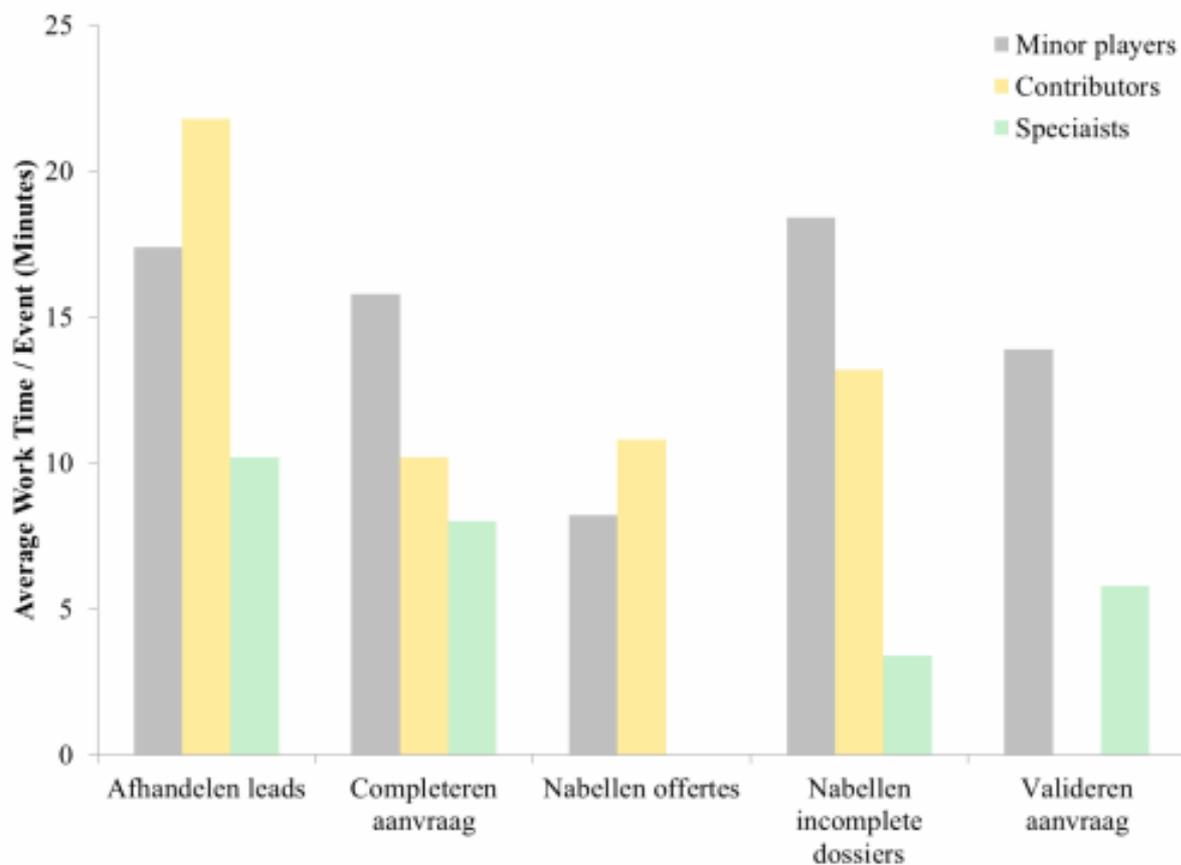


Figura 6. Tiempo de trabajo por evento, especialistas / contribuidores vs. colaboradores secundarios (*minor players*).

el caso de *Valideren aanvraag*, es probable que el banco sea lo más exhaustivo posible durante el proceso de validación, no importa cuántas veces antes haya visto una solicitud. Por otra parte, en la investigación de casos sospechosos de fraude, el banco puede haber llegado ya a una conclusión preliminar respecto de la solicitud y utilizar la instancia de seguimiento meramente para justificar su decisión.

Las instancias de seguimiento para aquellos eventos donde el banco debe contactar al solicitante a menudo tienen tiempos de trabajo promedio menores que sus contrapartes iniciales ya que estas actividades son las que tienen mayores posibilidades de quedar atrapadas en ciclos repetitivos, quizás debido a clientes que no responden. Uno puede aprovechar tales datos de eventos para entender el comportamiento del cliente y evaluar su utilidad potencial para la priorización del trabajo.

4.3. Análisis a nivel de recursos humanos

4.3.1. Actividades de trabajo dirigidas por especialistas vs. generalistas

Hemos construido el perfil de 48 personas (recursos humanos) que gestionaron al menos 100 eventos totales (excluyendo el recurso 112, ya que esta persona no gestiona otros eventos de trabajo que no sean la programa-

ción) y computamos el volumen de trabajo por el número de eventos gestionados por cada uno. Observamos nueve personas que invirtieron >50% de su esfuerzo en *Valideren aanvraag*, y un grupo destacado que realizó mayormente actividades de *Completeren aanvraag*, *Nabellen offertes* y *Nabellen incomplete dossiers*. Al parecer la validación la realiza un grupo dedicado de especialistas enfocado en este tipo de trabajo, mientras que las actividades de cara al cliente como *Completeren aanvraag*, *Nabellen offertes* y *Nabellen incomplete dossiers* podrían necesitar habilidades similares que son realizadas por otro grupo especializado.

A continuación examinamos el desempeño de recursos identificados como especialistas (>50% de los eventos de trabajo de un solo tipo) o contribuidores (25-50%) y los comparamos con aquellos que juegan solo un rol menor en actividades similares. A este fin tomamos el tiempo de trabajo total acumulado en una actividad por recursos pertenecientes a una categoría particular y calculamos promedios basados en el número total de eventos de trabajo realizados en esa categoría. Dos actividades, *Nabellen offertes* y *Valideren aanvraag*, no contienen especialistas o contribuidores, respectivamente y por lo tanto estas categorías fueron omitidas en las comparaciones para estas actividades.

Como se muestra en la figura 6, los especialistas dedicaron menos tiempo por instancia de evento que sus contrapartes, en algunos casos realizando las tareas hasta un 80% más eficientemente que los colaboradores secundarios. El desempeño de los contribuidores es mucho menos consistente, sin embargo, exhibiendo tiempos de trabajo promedio por caso que son más altos (*Afhandelen leads*, *Nabellen offertes*) o más bajos (*Completeren aanvraag*, *Nabellen incomplete dossiers*) que los de los colaboradores secundarios. Estos resultados sugieren que una oficina de especialistas realizando actividades únicas puede ser más adecuada para gestionar un número de casos más alto que un ejército de recursos asignado con una miríada de tareas.

4.4. Aprovechamiento de datos de comportamiento para priorizar el esfuerzo de trabajo

Uno de los objetivos de la minería de procesos es identificar oportunidades para dirigir la efectividad del proceso; esto es, lograr mejores resultados de negocio por el mismo o menor esfuerzo en un periodo igual o más corto. En particular, buscamos utilizar datos de eventos de proceso recogidos en una solicitud para priorizar mejor los esfuerzos de trabajo. Específicamente, nos propusimos entender si esto podría

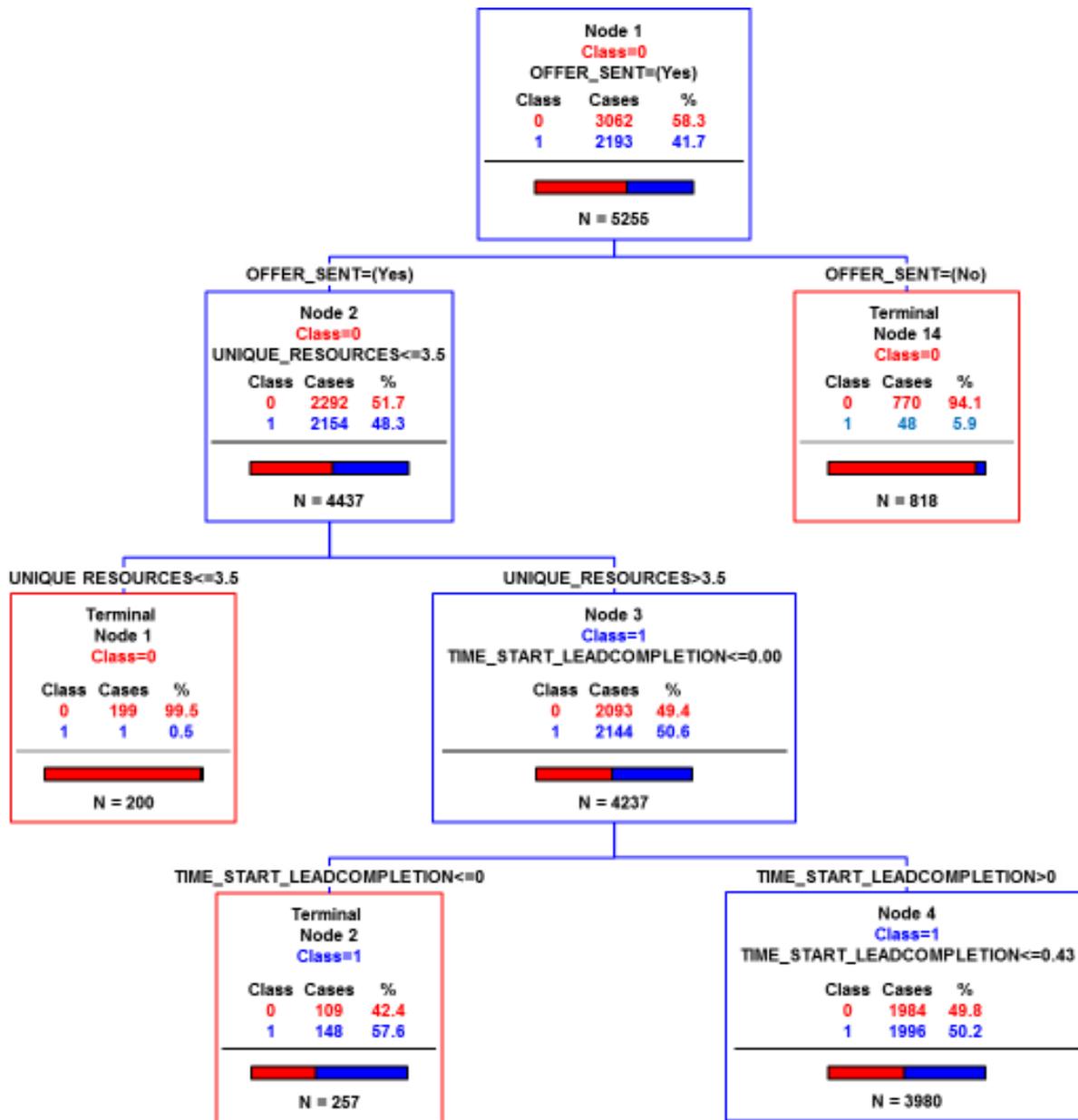


Figura 7. Vista parcial, árbol de segmentación CART.

hacerse en el quinto día a contar desde la presentación de la solicitud.

Con este fin, creamos un conjunto de datos a nivel de solicitud para 5.255 casos que duraron más de 4 días y donde el resultado final es conocido. Para estas solicitudes, capturamos todos los eventos desde la presentación hasta el final del día 4 y los usamos para calcular lo siguiente:

- Qué etapa había alcanzado la solicitud, y si había sido completada.
- Cuánto esfuerzo se había invertido en la solicitud.
- Cuántos eventos de cada clase ya habían sido registrados.
- Si la solicitud requería seguimiento de la oportunidad.

Intentamos encontrar segmentos clave en esta población con alta probabilidad de ser aprobados y aceptados, o bien con alta probabilidad de ser cancelados o rechazados. Lo hicimos segmentando los datos usando la técnica del Árbol de Clasificación y Regresión (*Classification and Regression Tree*, ver **figura 7**).

El árbol parcial anterior muestra dos segmentos con menos del 6% de tasas de aprobación. Los nodos terminales 1 y 14, con un total de 1.018 casos y con solamente 49 aprobaciones finales. El nodo 14, compuesto por 818 casos, muestra solicitudes incompletas donde el banco no pudo preparar ofertas para los clientes al final del Día 4. Tales solicitudes "lentas" tienen una probabilidad

menor del 6% de ser aprobadas, comparadas con un promedio del 42% para el grupo total de 5.255. El nodo 1 tiene solicitudes que son tocadas por 3 o menos recursos, uno de los cuales es el 112. Esto podría ser otro indicador de una solicitud lenta. Tales solicitudes carecen prácticamente de posibilidades de ser finalmente aprobadas.

Se podría repetir este análisis en diferentes etapas del ciclo de vida de la solicitud para colaborar con la priorización de esfuerzo. Este análisis preliminar indica un potencial significativo para reducir el esfuerzo en casos que podrían no llegar al estado final deseado. Análisis posteriores con demografía de clientes, detalles de la solicitud e información adicional acerca de los recur-

	<i>Afhandelen Leads</i>	<i>Completer en aanvraag</i>	<i>Nabellen Offertes</i>	<i>Nabellen Incomplete Dossiers</i>	<i>Valideren Aanvraag</i>
Tiempo de Trabajo Total (Minutos)	88.905	205.588	133.768	171.668	158.566
# Total de Tareas	5.041	20.830	10.426	19.748	7.819
Eficiencia Promedio del Especialista (Minutos / Evento)	10,2	8,0	10,8*	3,4	5,8
Tiempo Total de Trabajo Bajo la Eficiencia Promedio del Especialista (Minutos)	51.418	166.640	112.600	67.143	45.350
Ahorros de Tiempo Proyectados (Minutos)	37.487	38.949	21.167	104.525	113.216

Tabla 4. Ahorros potenciales de tiempo asociados con la conversión de los generalistas actuales en especialistas de una única actividad. *Ninguna de las personas que ejecutan *Nabellen offertes* fue identificada como especialista; por lo tanto se utilizó en su lugar a

los que trabajaron en tales casos podrían ayudar a refinar los hallazgos y sugerir pasos de acción específicos para mejorar la efectividad del proceso.

5. Discusión

5.1. Tratamiento de los desafíos que presentan los datos

5.1.1. Gestión de la complejidad de los eventos

La optimización del proceso de aprobación de préstamos es un ejercicio de racionalización de cada paso de la operación de principio a fin. Un punto notable que genera desafíos en la construcción de una visión del proceso racionalizada con herramientas de minería de procesos automatizadas es la cantidad y complejidad de los datos capturados. Si tales datos no se acompañan con un apropiado juicio de negocio, es posible perderse en la complejidad aparente (existen más de 4.000 variantes para un proceso que tiene 6-7 pasos clave). Hemos ilustrado este punto más arriba en nuestra discusión sobre eventos redundantes.

Recomendamos que esas complejidades sean abordadas en el momento del análisis, usando conocimiento del proceso y buen juicio empresarial, realizando pasos adicionales de pre-procesamiento de datos.

También es crítico escudriñar los datos de eventos por adelantado para entender todas las peculiaridades y construir formas de afrontarlas. Por ejemplo, una comparación del número de transiciones START y COMPLETE para eventos de tipo W-type en el conjunto de datos revela que hay 1.037 más transiciones COMPLETE que transiciones START. Como la marca temporal para estos eventos es única respecto de otras en el mismo *Case ID*, tienen el potencial de confundir mucho la acumulación de tiempo de trabajo y de espera para un caso particular y para recursos en la institución. Los designamos como errores de los sistemas y trabajamos con la primera transición COMPLETE que sigue a una START como la "correcta" para un tipo de evento de trabajo dado. En un proyecto real, validaríamos nuestra asunción revisando más en profundidad cómo surgen dichas instancias en el sistema y utilizando nuestra comprensión para tratar estas observaciones de forma correcta en nuestro análisis.

Como describimos en la **sección 3.1**, el registro de eventos también se beneficiaría de la consolidación de eventos que suceden concurrentemente, como los que ocurren cuando se aprueban solicitudes exitosas

(A_APPROVED, A_REGISTERED y A_ACTIVATED). Esto no solo reduciría el tamaño general del archivo (que pasa a ser importante cuando crece el volumen de los datos) sino que también reduce la complejidad del registro inicial.

5.2. Beneficios potenciales del re-despliegue de recursos

5.2.1. Redefinición de generalistas como especialistas

Como mencionamos antes, las tareas que están implicadas en el proceso de aprobación de préstamos son realizadas por una mezcla de especialistas y generalistas. A lo largo de nuestro análisis concluimos que el banco podría beneficiarse de la especialización del trabajo, de forma que los recursos actuales se reasignen a puestos únicos para maximizar la eficiencia. En la **tabla 4** mostramos las ganancias potenciales que pueden obtenerse mediante dicha reestructuración. Si el banco puede mejorar el desempeño de todos quienes ejecutan una tarea al mismo nivel que los especialistas, estimamos un ahorro sustantivo y general de tiempo.

También evaluamos los ahorros potenciales asociados con el ajuste a la baja del grupo general de personas asignadas a estas tareas. Utilizando la cantidad promedio de tiempo de trabajo para los recursos que manejan más de 100 eventos totales (aproximadamente 16.000 minutos; de nuevo excluyendo el recurso 112), estimamos la oportunidad de reducir el esfuerzo de trabajo en un 35%.

5.3. El poderío de la información adicional

5.3.1. Atributos adicionales a nivel de caso

En su formato bruto, el registro de eventos BPIC 2012 es una mina de oro de información que, una vez decodificada, ofrece una vista detallada de un proceso de aprobación de préstamos al consumo. Sin embargo, esta información se fortalecería enormemente si se añadieran unos pocos puntos de datos clave. Como cada caso trae consigo un único atributo (el importe pedido por el solicitante) no tenemos forma de saber por qué ciertos casos son aprobados mientras que otros con montos pedidos y caminos idénticos son rechazados. Por consiguiente, sería útil conocer la demografía de los clientes, las relaciones actuales o anteriores con los clientes y detalles adicionales sobre el personal que ejecuta estos procesos. Con esta información podríamos construir recomendaciones específicas para cambiar el proceso y estimar con mayor precisión los beneficios probables de tales cambios.

5.3.2. Rentabilidad del cliente y costes operativos

Un conjunto final de datos que está notoriamente ausente del registro BPIC 2012 provis-

to es el coste global asociado con el proceso de aprobación de préstamos y el valor de cada solicitud de préstamo para el banco. Merecería la pena entender cuánto cuesta operar cada recurso y si este coste varía en base a las actividades que el personal realiza o el número de eventos en los que participa. Esta información también nos permitiría calcular un coste de adquisición promedio para cada solicitante, y consiguientemente entender el umbral mínimo por debajo del cual no tiene sentido económico aprobar una petición de préstamo entrante.

6. Conclusiones

A través del análisis comprehensivo del registro de eventos de BPIC 2012, hemos convertido un conjunto de datos bastante complejo en un flujo de trabajo de extremo a extremos claramente interpretable para un proceso de aprobaciones de préstamos y descubiertos. Examinamos los datos en múltiples niveles de granularidad, descubriendo percepciones interesantes en todos los niveles. A través de nuestro trabajo sacamos a la luz mejoras potenciales en varias áreas, incluyendo la revisión de procesos automatizados, la reestructuración de recursos claves y la evaluación de los procedimientos actuales de tratamiento de casos. De hecho, el análisis futuro será ayudado en buena medida por la inclusión de datos adicionales tales como información del cliente, políticas de gobierno, costes operativos y valor relativo del cliente.

Como parte de nuestro análisis, realizamos un ejercicio predictivo rudimentario mediante el cual determinamos el estado actual de casos en diversos días a lo largo del proceso de aprobación y cuantificamos sus posibilidades de aprobación, cancelación o rechazo. Esto nos permitió estimar la suerte de un caso basado en su desempeño y adaptar el proceso general para minimizar el estancamiento en cuellos de botella tradicionales en los casos. Aunque sea preliminar por su naturaleza, esto abre la puerta para futuros ejercicios de modelado más elaborado, quizás dirigidos por algoritmos informáticos sofisticados.

Si bien cubrimos varias áreas en este ejercicio, hay otras donde no realizamos un análisis detallado. El banco se beneficiaría de manera significativa al explorar esas áreas adicionales, como por ejemplo el análisis de redes sociales.

En conclusión, los procedimientos resaltados por BPIC 2012 destacan el rol e importancia de la minería de procesos en el lugar de trabajo moderno. Pasos que con anterioridad solo podían ser elucidados después de años de práctica y observación pueden ahora ser examinados utilizando una muestra de datos existentes. A medida que la era del *Big Data* continúa su marcha hacia el mundo del nego-

cio, pronosticamos que la minería de procesos será un elemento central en el cambio que convertirá preguntas en soluciones y problemas en beneficio sostenible.

Reconocimientos

Agradecemos a la institución financiera que ha puesto a disposición estos datos para su estudio. Un agradecimiento especial a Fluxicon por facilitarnos una copia de evaluación de Disco acompañada de una copia del conjunto de datos BPIC 2012. También damos las gracias a Tom Metzger, Nicholas Hartman, Rolf Thrane y Pierre Buhler por sus útiles discusiones y percepciones. Nuestro agradecimiento se dirige también a Salford Systems, que puso a nuestra disposición su software en una versión de demostración.

Referencias

- [1] W. Van der Aalst, A. Adriansyah, A.K. Alves de Medeiros, F. Arcieri, T. Baier *et al.* Process Mining Manifesto. *Business Process Management Workshops 2011, Lecture Notes in Business Information Processing*, vol. 99. Springer-Verlag, 2011.
- [2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, 2011. <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>.
- [3] R. Adduci, D. Blue, G. Chiarello, J. Chickering, D. Mavroyiannis *et al.* *Big Data: Big Opportunities to Create Business Value*. Technical report, Information Intelligence Group, EMC Corporation, 2011.
- [4] R.P.J.C. Bose, W.M.P. van der Aalst. Analysis of Patient Treatment Procedures: The BPI Challenge Case Study. *First International Business Process Intelligence Challenge*, 2011. <<http://bpmcenter.org/wp-content/uploads/reports/2011/BPM-11-18.pdf>>.
- [5] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011. ISBN-10: 3642193447.

Santiago Aguirre Mayorga,
Carlos Alberto Parra Rodríguez
Departamento de Ingeniería Industrial,
Pontificia Universidad Javeriana, Bogotá
(Colombia)

<{saguirre,carlos.parra}@javeriana.edu.co>

Mejoramiento de procesos con técnicas de minería de procesos, simulación y optimización: Caso de estudio

1. Introducción

Para realizar un análisis integral de los procesos de negocio no basta con una sola metodología o herramienta. La minería de procesos provee herramientas para el diagnóstico y análisis de procesos, pero debe ser complementada con otras metodologías y técnicas como por ejemplo la simulación para poder construir modelos basados en la ejecución real del proceso y poder probar las diferentes alternativas de rediseño de los procesos antes de ponerlas en práctica.

La simulación y la optimización se han usado más que todo en procesos de producción y logística donde las rutas del proceso están predefinidas y se puede construir un modelo del proceso. En procesos de servicio, como por ejemplo procesar una queja, se pueden presentar muchas variaciones o rutas del proceso dependiendo del tipo de queja, por lo que es importante partir de analizar la bitácora de eventos del sistema de información para llegar a tener un modelo real del proceso de forma que la simulación se pueda realizar sobre este modelo y no sobre una versión idealizada del proceso.

La simulación puede beneficiarse de gran manera de la minería de procesos porque ésta le puede proporcionar los parámetros necesarios para construir el modelo del proceso a simular (proceso de ejecución real, tiempos de ejecución, tiempos de espera, distribución del arribo de nuevos casos, etc.) y esto se ve demostrado en los casos donde se combina minería de procesos con simulación [1][2][3].

2. Rediseño de procesos combinando minería de procesos, simulación y optimización

Para ilustrar la forma como las herramientas de modelamiento, simulación y optimización se pueden usar para la gestión y mejoramiento de procesos, se desarrolla un caso de estudio de rediseño de un proceso de abastecimiento, enmarcado en la metodología de rediseño de procesos del enfoque de BPM (*Business Process Management*) [4].

A continuación, presentamos las fases de la metodología enumerando actividades, herramientas y técnicas:

■ **Planeación del Proyecto.** En esta etapa se determina la oportunidad de mejora, se realiza un análisis de brechas entre el desempeño actual y el desempeño esperado y se determi-

Resumen: Este artículo presenta un caso de estudio donde se combinaron herramientas de minería de procesos, simulación y optimización para la mejora de un proceso. La minería de procesos proporcionó herramientas para el diagnóstico de los problemas del proceso y para encontrar los principales cuellos de botella y la simulación permitió determinar y evaluar las principales alternativas de mejora para su posterior implantación. Por otro lado, la optimización a través del Data Envelopment Analysis permitió encontrar la eficiencia relativa de los proveedores del proceso. Con esto se logró combinar diferentes herramientas para lograr los objetivos de mejora de procesos.

Palabras clave: Análisis de procesos, minería de datos, minería de procesos, rediseño de procesos, simulación.

Autores

Santiago Aguirre Mayorga es Ingeniero Industrial por la Pontificia Universidad Javeriana, Bogotá (Colombia) y *Magíster* en Ingeniería Industrial por la Universidad de los Andes (Uniandes – Colombia). Actualmente, es candidato a PhD. en Ingeniería por la Pontificia Universidad Javeriana, y trabaja como Profesor Asociado del Departamento de Ingeniería Industrial de dicha universidad. Es además consultor en el tema de *Business Process Management* (BPM) en diversas organizaciones.

Carlos Alberto Parra Rodríguez es Ingeniero Electrónico por la Pontificia Universidad Javeriana, Bogotá (Colombia) y *Magíster* en Ingeniería Eléctrica por la Universidad de los Andes (Uniandes – Colombia). Posee un DEA "*Diplome d'Etudes Approfondies*" en Informática Industrial y es Doctor por la *Université Paul Sabatier* de Toulouse (Francia). Trabaja como Profesor Titular y Director del Doctorado en Ingeniería de la Pontificia Universidad Javeriana. Entre sus temas de investigación se encuentra la inteligencia aplicada a procesos.

nan los objetivos, alcance, cronograma y equipo de proyecto.

■ **Análisis de Procesos (AS-IS).** En esta etapa se analizan detalladamente el proceso y sus oportunidades de mejora, se determinan los problemas y se analizan las causas.

■ **Rediseño de Procesos (TO-BE).** En la etapa de rediseño se evalúan las diferentes alternativas de mejoramiento de los procesos.

■ **Implantación.** Finalmente, se implantan los cambios y se evalúa el proyecto con respecto a sus objetivos.

Estas fases se aplicaron en el caso de estudio que se explica a continuación.

3. Caso de estudio: proceso de compra de bienes

El caso de estudio se desarrolló sobre un proceso de compra de bienes de una institución universitaria¹ donde se procesan aproximadamente 15.000 solicitudes de compra anualmente con un presupuesto estimado de 50 millones de dólares. El normal funcionamiento de la Universidad y sus proyectos dependen de la eficiencia con la que el Departamento de Suministros tramite los bienes y servicios requeridos por lo que en 2009 se

implantó un sistema integrado de información como soporte a este proceso.

3.1. Etapa de Planeación del Proyecto

3.1.1. Determinación de la oportunidad de mejora

A pesar de tener el soporte de un sistema integrado de información, este proceso de compras de bienes y servicios ha venido presentando problemas e inconvenientes que le restan eficiencia como son el exceso de control y aprobaciones del proceso así como la dificultad con el manejo de los documentos asociados, lo que hace que algunos pedidos se demoren más de lo esperado.

3.1.2. Análisis de brechas del proceso

En el diagrama de la **figura 1** se presentan las principales brechas de desempeño y de capacidades del proceso y las herramientas a usar en el análisis y mejoramiento.

3.1.3. Determinación de los objetivos del proyecto

El proyecto de rediseño de procesos tiene los siguientes objetivos:

■ Disminuir los tiempos de respuesta para lograr que el 70% de los pedidos se entregue en menos de 1 mes.

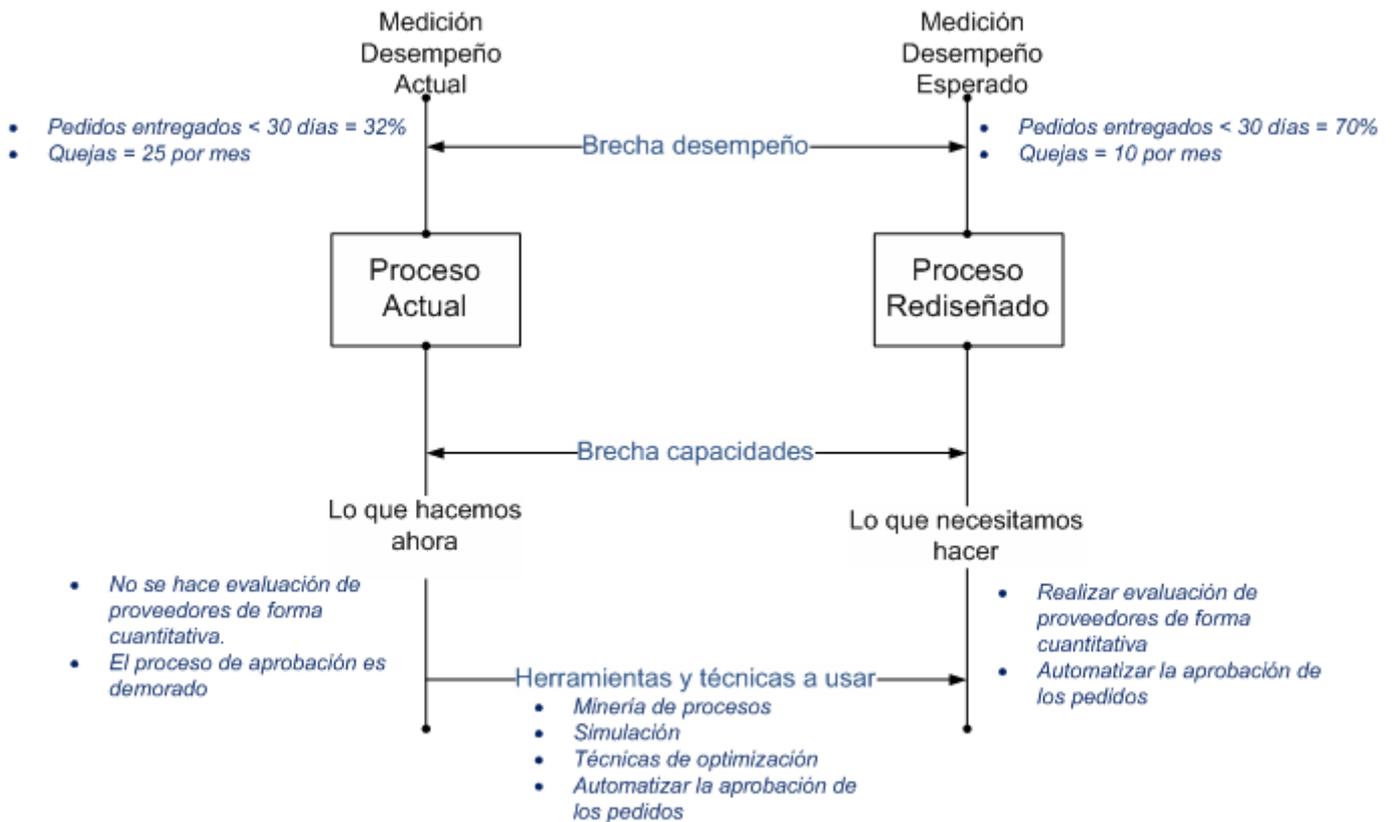


Figura 1. Análisis de brechas del proceso de compras.

- Disminuir las quejas de los usuarios a un máximo de 10 por mes.
- Generar mayores ahorros a la Universidad a través de una adecuada evaluación de los proveedores.

3.2. Etapa de Análisis de Procesos (AS-IS)

3.2.1. Extracción y análisis de la bitácora de eventos

Para la fase de análisis del proceso (AS-IS) se requiere extraer la bitácora de eventos del sistema de información donde se encuentran los datos con respecto a las actividades del proceso, los responsables, la fecha y hora de inicio y finalización de cada actividad, e información con respecto a cada solicitud de compra como el producto a comprar, el departamento solicitante y el proveedor, entre otros.

Este registro contiene el histórico de 1 año correspondiente a 15.091 registros. Una vez

obtenido el registro se procede a analizar la calidad de la información, en busca de datos perdidos y atípicos, para lo cual se puede usar la utilidad de análisis de datos perdidos de un paquete estadístico como SPSS².

Después de realizar la depuración de casos con datos perdidos y atípicos la base se redujo a 8.987 casos con 56 variables.

3.2.2. Descubrimiento del modelo real del proceso

A través de la aplicación de algoritmos de la minería de procesos como *alfa mining* [5], *heuristic mining* [6] o *genetic mining* [7], es posible descubrir de forma automática el modelo real del proceso utilizando la funcionalidad del software ProM o Disco.

El modelo de ejecución real del proceso se puede representar en una red de Petri, con la notación BPMN (*Business Process Modeling*

Notation) o UML (*Unified Modeling Language*), entre otras. La minería de procesos utiliza en sus aplicaciones las redes de Petri, debido a sus fundamentos matemáticos, lo que permite aplicar técnicas de análisis, y a que soporta características de los procesos como es la concurrencia [8]. Adicionalmente, se pueden usar las redes de Petri coloreadas (CPN) dada su capacidad de expresividad y de simulación en paquetes como CPN Tools [2].

En la figura 2 se presenta el proceso de modelado en una red de Petri. Por su parte, en la tabla 1 se presentan los principales hallazgos del descubrimiento del proceso.

3.2.3. Análisis de los indicadores de desempeño

De la bitácora de eventos del sistema se pueden extraer y analizar algunos indicadores de desempeño del proceso como son los

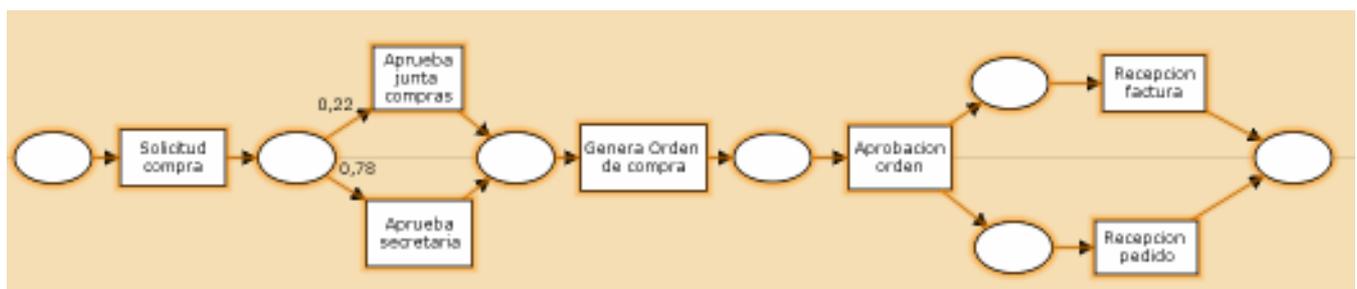


Figura 2. Proceso de abastecimiento representado en una red de Petri.

“ La simulación puede beneficiarse de gran manera de la minería de procesos porque ésta le puede proporcionar los parámetros necesarios para construir el modelo del proceso a simular ”

Hallazgos del descubrimiento del proceso

Existe un bucle (*loop*) entre la solicitud de compra y la aprobación por parte de la junta (en promedio se reprocesan el 24% de los casos que pasan a la junta de compras). Del total de los pedidos de bienes, el 22% pasa por la junta de compras.

El 100% de las órdenes de compra son aprobadas sin necesidad de modificaciones.

Tabla 1. Hallazgos de la fase descubrimiento del proceso.

tiempos por actividad y la productividad del personal.

En la **tabla 2** se presentan los tiempos de ciclo de cada actividad del proceso, los cuales se distribuyen de forma normal con su respectiva media y desviación estándar.

Para analizar los tiempos de ciclo de las compras asignadas a cada comprador se procedió a realizar un análisis donde se relacionó la variable tiempo de ciclo total con el comprador, tal como podemos observar en el diagrama de caja y bigotes mostrado en la **figura 3**.

Los principales hallazgos del análisis de indicadores de desempeño se encuentran documentados en la **tabla 3**.

3.2.4. Diagnóstico y análisis de los problemas

Para realizar un diagnóstico más detallado

Tiempo	Descripción	Tiempos W (en días) P (en minutos)
Tiempo elaboración solicitud	P: Tiempo para la elaboración de la solicitud en el sistema.	P: N(15,5)
Tiempo aprobación Junta	W: Tiempo entre la creación de la solicitud y el inicio de la aprobación de la junta. P: Tiempo de aprobación de los pedidos en la junta.	W: N(5,6) P: N(180,60)
Tiempo aprobación Secretaría	W: Tiempo entre la creación de la solicitud y el inicio de la aprobación por las secretarías de la facultad o de cargos directivos. P: Tiempo de aprobación de un pedido por las secretarías o cargos directivos.	W: N(14, 12) P: N(8, 5)
Tiempo elaboración orden de compra	W: Tiempo entre la finalización de la aprobación y el inicio de la elaboración de la orden de compra. P: Tiempo para la elaboración de la orden de compra en el sistema.	W: N(5, 3) P: N(10,8)
Tiempo aprobación orden de compra	W: Tiempo entre la finalización de la elaboración de la orden de compra y el inicio de su aprobación. P: Tiempo para aprobar la orden de compra en el sistema.	W: N(8, 4) P: N(12,10)
Tiempo de recepción	W: Tiempo desde el envío de la orden de compra al proveedor hasta la recepción del bien. P: Tiempo para realizar la entrada de la mercancía en el sistema.	W: N(25, 31) P: N(30,15)

Tabla 2. Descubrimiento de los tiempos reales de ejecución y espera de las actividades.

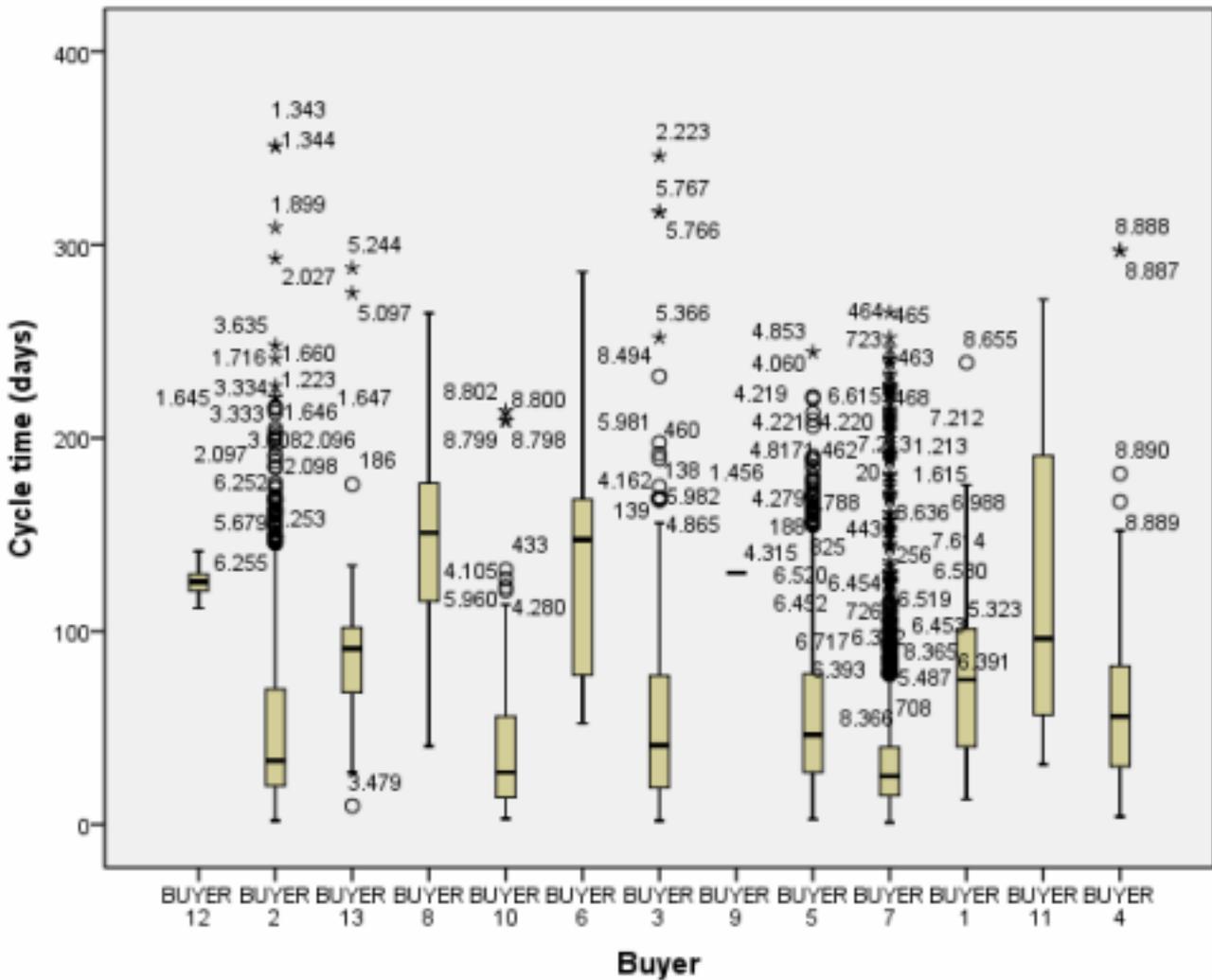


Figura 3. Tiempos de ciclo por comprador.

Hallazgos del análisis de indicadores

El proceso más demorado a nivel interno es la aprobación de la solicitud de compra por la secretaría de las facultades o por los cargos directivos de las unidades del gobierno central, lo que se constituye el cuello de botella del proceso.

Únicamente el 32% de los pedidos se entregan en menos de 30 días. El tiempo medio del ciclo del proceso es en promedio 50 días con una desviación estándar de 28 días.

Las compras de importación son en promedio 3 veces más demoradas que una compra nacional. El tiempo mínimo requerido para una importación son 40 días.

Gran parte de la variabilidad del tiempo de ciclo del proceso se ve explicada por el comprador (persona en el Departamento de Suministros) que gestiona cada solicitud.

Tabla 3. Hallazgos de la fase de análisis de indicadores.

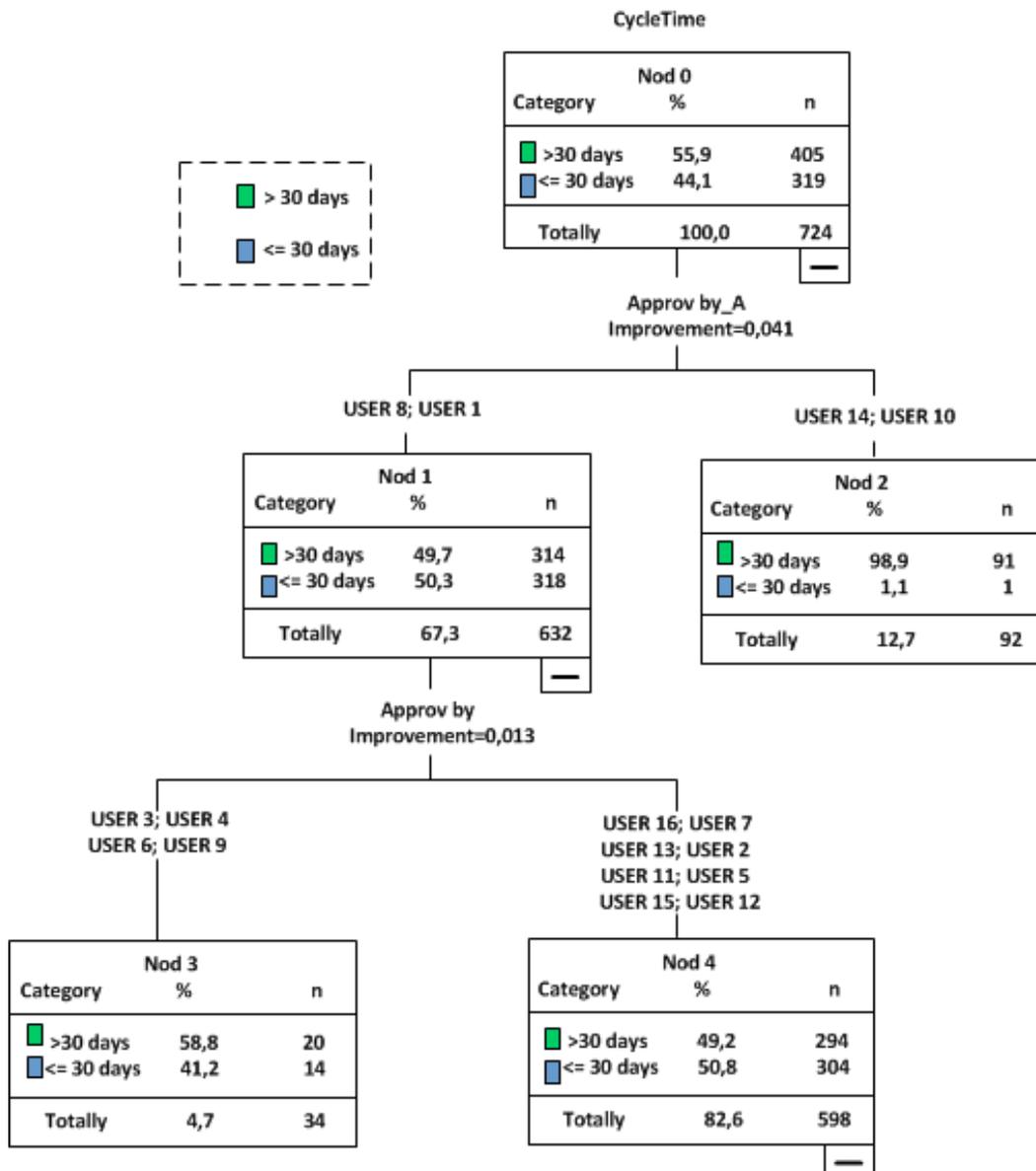


Figura 4. Árbol de decisión para el diagnóstico del proceso.

del subproceso de aprobación de la solicitud de compras por parte de la secretaria o de los cargos directivos, se realizó un análisis con árboles de decisión donde se encontró que cuando la solicitud de compra debe ser aprobada por los cargos del nodo 2, la probabilidad de que el pedido llegue antes de 30 días es del 1%. Cuando los aprobadores son los del nodo 1, la probabilidad de recibir el pedido en menos de 30 días aumenta a un 50% (ver figura 4).

El resumen de los principales hallazgos de esta etapa se muestra en la tabla 4.

3.2.5. Análisis de las causas de los problemas

Para el análisis de las causas del problema trabajamos sobre una espina de pescado, donde los involucrados en la ejecución del proceso determinaron las principales causas de la demora en la aprobación de las solicitudes. En la figura 5 se presenta este análisis.

En la tabla 5 se presentan los principales hallazgos de esta etapa.

3.3. Etapa de Rediseño de Procesos
3.3.1. Determinación de las alternativas de mejora

Una vez determinadas las causas de los problemas en el proceso se determinaron las alternativas de mejora. Para evaluar el impacto que tienen algunas de estas alternativas se usó la simulación, según se detalla en la tabla 6.

Hallazgos del diagnóstico y análisis

La persona que aprueba la solicitud de compra tiene una incidencia importante en la probabilidad de recibir el pedido antes de 30 días.

Tabla 4. Hallazgos de la fase de diagnóstico y análisis.

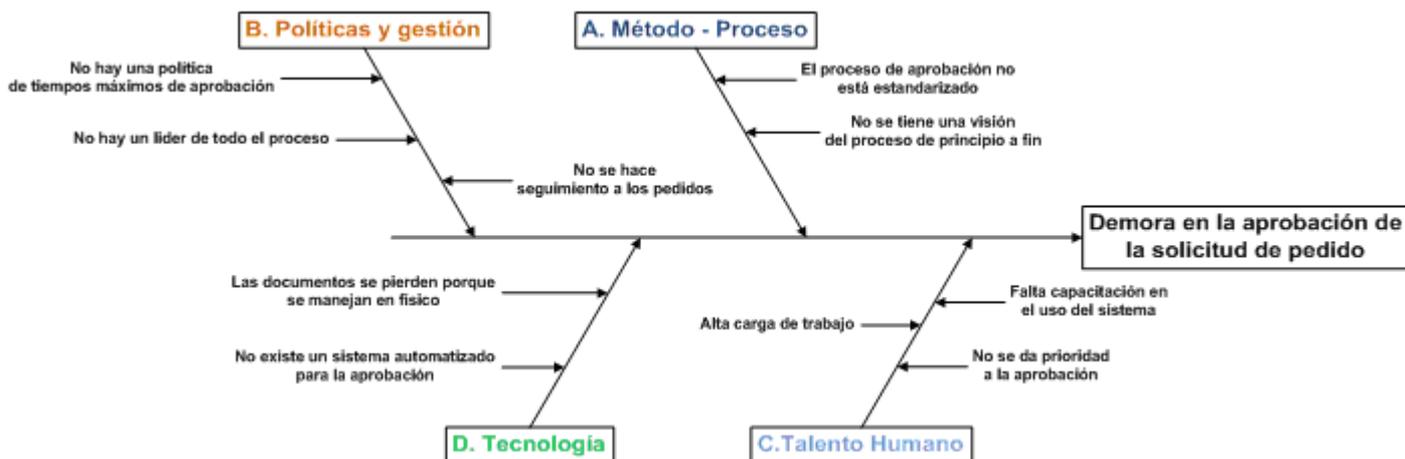


Figura 5. Diagrama de espina de pescado con la determinación de las principales causas de demora.

Hallazgos de las causas de los problemas
Una de las principales causas de la demora en la aprobación es que los documentos físicos que se manejan en el proceso muchas veces se pierden por la manipulación y transporte entre las diferentes dependencias.
La falta de un sistema automatizado para la aprobación de las compras hace que el proceso sea difícil de controlar.
No existe una política que determine un tiempo máximo para la aprobación de las solicitudes de compra.
Debido a su alta carga de trabajo, algunos cargos directivos no le dan prioridad a la aprobación de las compras.

Tabla 5. Hallazgos de la fase de análisis de causas.

Alternativa de mejora	Explicación
1. Eliminar el proceso de la aprobación de la orden de compra.	De los 8.987 casos analizados, ninguna orden de compra fue rechazada por lo que es un control que puede ser eliminado.
2. Establecer una política de tiempo máximo de aprobación de la solicitud	Establecer una política donde las secretarías, junta de compras y cargos directivos de la universidad tengan un máximo de 5 días para aprobar las solicitudes de compra.
3. Automatizar el proceso de aprobación de la solicitud de compra	Implantar un sistema BPMS para automatizar el proceso de aprobación de la solicitud, donde se manejen todos los documentos en forma digital. Este sistema genera alarmas cuando una solicitud de compra no se ha aprobado en cierto tiempo.

Tabla 6. Alternativas de mejora de la fase de rediseño de procesos.

“ A través de la aplicación de algoritmos de la minería de procesos como *alfa mining*, *heuristic mining* o *genetic mining*, es posible descubrir de forma automática el modelo real del proceso utilizando la funcionalidad del software ProM o Disco ”

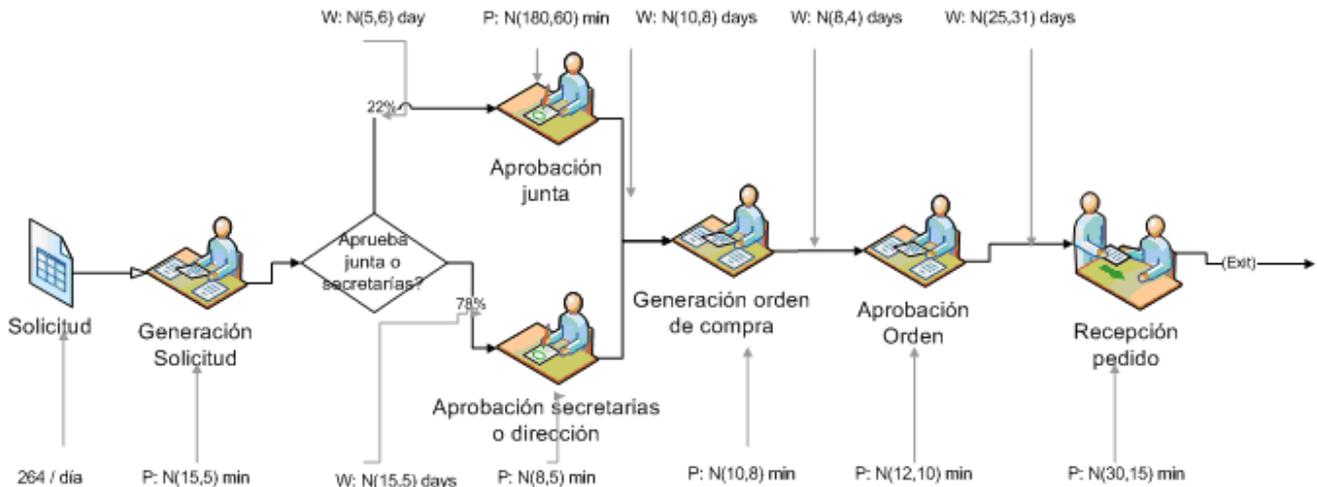


Figura 6. Modelo de simulación del proceso de compras de bienes.

3.3.2. Generación del modelo a simular

Se generó un modelo de simulación basado en el proceso descubierto referido en la sección 3.2.2 y con los datos de tiempos calculados con el análisis de la bitácora de eventos del sistema de información (ver tabla 2). En la figura 6 se presenta el modelo de la simulación, ejecutado en la herramienta *Process Modeler*.

Se realizó la simulación de 4.224 pedidos y el tiempo promedio fue de 50,72 días.

3.3.3. Simulación y evaluación de las alternativas de mejora

Mediante la simulación se evaluaron los siguientes escenarios que corresponden a las alternativas de rediseño del proceso:

- Escenario 1: Eliminar el proceso de aprobación de la orden de compra.

- Escenario 2: Establecer una política de tiempo máximo de aprobación de la solicitud de 5 días por parte de la junta o secretarías.

- Escenario 3: Escenario 1 + Escenario 2. En la tabla 7 se resumen los hallazgos de la simulación.

En esta fase se realiza un análisis coste-beneficio de cada una de las alternativas para poder soportar la toma de decisiones.

3.3.4. Optimización

En la fase de análisis de brechas del proceso (sección 3.1.2) se identificó que uno de los problemas del proceso de abastecimiento es que no se realiza una evaluación cuantitativa de los proveedores, por lo que no se dispone de la información para tomar decisiones de selección de un proveedor o de clasificar a los proveedores de acuerdo con su eficiencia.

Data Envelopment Analysis (DEA) es una técnica que utiliza la programación matemática para hallar la eficiencia relativa de los proveedores y esta ha sido aplicada para la evaluación y selección de proveedores [9][10]. Con esta técnica se pueden evaluar los proveedores con respecto a los tiempos de entrega, devoluciones por no conformidad y nivel de calidad de los productos teniendo en cuenta el precio del producto.

En la tabla 8 se presenta un ejemplo de la evaluación de 4 proveedores con respecto a tiempo, devoluciones y precio. El indicador de eficiencia evaluó como mejor al proveedor 3.

Al realizar un análisis de sensibilidad se pudo comprobar que si el proveedor 3 disminuye el porcentaje de entregas a tiempo del 98% al

Hallazgos de la simulación

Escenario 1: Si se elimina el paso de aprobación de la orden de compra se puede disminuir el tiempo a 43 días en promedio.

Escenario 2: Si se estable un tiempo máximo para la solicitud de compra de 5 días el tiempo se disminuye a 40 días en promedio.

Escenario 3: Si se elimina el paso de aprobación de la orden de compra y se establece un tiempo máximo de aprobación de la solicitud de 5 días se disminuye el tiempo a 35 días en promedio.

Tabla 7. Hallazgos de la fase de simulación.

	1 / Devoluciones	% Devoluciones	Entrega a tiempo	% Entrega a tiempo	Precio	Peso salidas	Peso entradas	Eficiencia
Proveedor 1	0,13	8%	0,90	90%	100	0,86	0,95	0,90
Proveedor 2	0,20	5%	0,95	95%	105	0,90	1,00	0,90
Proveedor 3	0,50	2%	0,98	98%	98	0,93	0,93	1,00
Proveedor 4	0,25	4%	0,92	92%	101	0,88	0,96	0,91
Peso	0,00011		0,95238		0,00952			

Tabla 8. Calificación de los proveedores con DEA.

	1 / Devoluciones	% Devoluciones	Entrega a tiempo	% Entrega a tiempo	Precio	Peso salidas	Peso entradas	Eficiencia
Proveedor 1	0,13	8%	0,90	90%	100	0,94	0,95	0,99
Proveedor 2	0,20	5%	0,95	95%	105	0,99	1,00	0,99
Proveedor 3	0,50	2%	0,85	85%	98	0,89	0,93	0,95
Proveedor 4	0,25	4%	0,92	92%	101	0,96	0,96	1,00
Peso	0,00000		1,04555		0,00952			

Tabla 9. Análisis de sensibilidad de la calificación de proveedores con DEA.

Hallazgos de la optimización

Mediante técnicas como *Data Envelopment Analysis (DEA)* se puede usar la programación matemática para hallar la eficiencia relativa de los proveedores para evaluarlos con respecto al precio, tiempos de entrega, devoluciones por no conformidad y nivel de calidad de los productos.

Tabla 10. Hallazgos de la fase de Optimización.

85% su eficiencia se ve afectada como se muestra en la **tabla 9**.

En la **tabla 10** se presentan los principales hallazgos de la optimización.

4. Conclusiones

Las técnicas de minería de procesos, simulación y optimización se pueden combinar con el objetivo de analizar y mejorar procesos de negocio. Éstas son herramientas complementarias que usadas en conjunto permiten el diagnóstico, análisis y rediseño de los procesos basados en datos reales de ejecución del proceso.

La minería de procesos provee técnicas para el descubrimiento del modelo real de ejecución del proceso y actualmente existen aplicaciones que soportan este análisis como son Disco y ProM, las que proveen herramientas gráficas de análisis, visualización y animación de los procesos. Las técnicas tradicionales de la minería de datos como árboles de decisión permiten predecir el desempeño en los procesos basados en encontrar las variables que inciden en los tiempos de ciclo.

La simulación se beneficia en gran manera de la aplicación previa de las técnicas de minería

de procesos, dado que ésta le puede proporcionar los parámetros y datos que son necesarios para la construcción del modelo de simulación basado en los datos de ejecución real del proceso. La simulación permite probar diferentes escenarios con respecto a las alternativas de mejora del proceso antes de su implementación, por lo que constituye una herramienta valiosa para la toma de decisiones.

Referencias

- [1] A. Rozinat. *Process Mining: conformance and extension*. PhD. Thesis. Eindhoven University of Technology, Eindhoven, 2010.
- [2] A. Rozinat, R. Mans, M. Song. Discovering simulation models. *Information Systems*, 34, pp. 305-327, 2009.
- [3] L. Maruster, L. Van Beest. Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, 21, pp. 267-297, 2009.
- [4] P. Harmon. *Business Process Change*. Morgan Kaufmann, Burlington, 2007.
- [5] W.M.P. Van der Aalst, A. Weijters, L. Maruster. Workflow mining: discovering process models from event logs. *IEEE Transactions on*

Knowledge and Data Engineering 16(9): pp. 1128-1142, 2004.

[6] A. Weijters J. Ribeiro. *Flexible heuristics miner*. BETA working Paper Series, Eindhoven University of Technology, Eindhoven, 2010.

[7] A. Medeiros, A. Weijters, W.M.P. Van der Aalst. Genetic process mining: An Experimental Evaluation. *Data mining and Knowledge Discovery* 14(2): pp. 245-304, 2007.

[8] C. Stahl. *Modeling Business Process: A Petri Net-Oriented Approach*. The MIT Press, Cambridge, 2011.

[9] Y. Chen. Structured Methodology for supplier selection and evaluation in a supply chain. *Information Sciences*, 181, pp. 1651-1670, 2011.

[10] M. Zeydan, C. Colpan, C. Cobanoglu. A combined methodology for supplier selection and performance evaluation. *Expert Systems with Applications* 38, pp. 2741-2751, 2011.

Notas

¹ Algunos datos del caso de estudio han sido modificados por motivos de privacidad y para ejemplificar la aplicación de las herramientas.

² <<http://es.wikipedia.org/wiki/SPSS>>.

Daniela Lorena Luengo Mundaca, Marcos Sepúlveda Fernández

Departamento de Ciencias de la Computación, Escuela de Ingeniería, Pontificia Universidad Católica de Chile

<dlluengo@uc.cl>, <marcos@ing.puc.cl>

1. Introducción

Hoy en día, en un mundo globalizado e hiperconectado, las organizaciones tienen la necesidad de mantenerse en permanente cambio para adaptarse a las necesidades del entorno, lo que implica que sus procesos de negocio también deban estar cambiando constantemente.

Para ilustrar esto, es posible considerar el caso de una tienda de juguetes, donde sus procesos de venta pueden variar de manera radical dependiendo de si son ejecutados en Navidad o en período de vacaciones, debido principalmente a los cambios en los volúmenes de la demanda. En Navidad se podría ejecutar un proceso que priorice la eficiencia y el volumen – *throughput* –, y en vacaciones se podría ejecutar un proceso con foco en la calidad de atención al cliente.

En este ejemplo es fácil identificar los períodos en los que la demanda cambia y, por lo tanto, es posible que el responsable de la gestión del proceso tenga claridad respecto a los cambios que sufre el proceso de venta a través del tiempo. Sin embargo, si una organización tiene un proceso que se realiza en distintas oficinas autónomas, por ejemplo, porque se encuentran en distintas ubicaciones geográficas, la evolución de los cambios en el proceso en cada oficina ya no es tan evidente para el responsable central del proceso; los cambios en cada oficina pueden ser distintos y estarse aplicando en distintos momentos en el tiempo.

Entender los cambios que están ocurriendo en las distintas oficinas, podría ayudar a entender mejor cómo mejorar el diseño global del proceso. Poder entender estos cambios y modelar las distintas versiones del proceso, permiten al responsable de su gestión contar con información más precisa y completa para tomar decisiones coherentes que redunden en una mejor atención o eficiencia.

Para lograr lo anterior, se han desarrollado diversos avances en la disciplina de gestión de procesos de negocio (BPM), disciplina que combina conocimiento sobre tecnologías de información y técnicas de gestión, las cuales son aplicadas a procesos de negocio operativos, con el objetivo de mejorar su eficiencia [1].

Dentro de BPM (*Business Process Management*), la minería de procesos se ha

Detección de cambios temporales en los procesos de negocio mediante el uso de técnicas de segmentación

Resumen: Hoy en día, las organizaciones tienen la necesidad de estar constantemente cambiando para ajustarse a las necesidades del entorno. Estos cambios se reflejan en sus procesos de negocio. Por ejemplo, un supermercado debido a cambios estacionales tendrá distinta demanda en distintos meses del año, por lo que sus procesos de abastecimiento o de reposición de productos podrían ser distintos en distintas épocas del año. Una forma de analizar con profundidad un proceso y entender cómo realmente se ejecuta en la práctica a través del tiempo, es en base al análisis de sus registros históricos almacenados en los sistemas de información, lo cual es conocido como minería de procesos. Sin embargo, en la actualidad la mayoría de las técnicas que existen para analizar y mejorar procesos consideran todos los registros de un proceso de manera estática, es decir, que el proceso no cambia a través del tiempo, lo cual en la práctica es poco realista dada la naturaleza dinámica de las organizaciones. Nuestro trabajo propone una técnica de segmentación que encuentra las distintas versiones de un proceso a través del tiempo. Esta técnica se basa en una ya existente de segmentación que solo considera características estructurales del proceso (flujo de actividades). Nuestra técnica incorpora de manera adicional la característica temporal de los procesos, de tal manera que los clusters que se generen al realizar la segmentación tengan una similitud estructural, pero también una cercanía temporal, de tal manera que representen distintas versiones del proceso. Este documento presenta el detalle de la técnica propuesta y un conjunto de experimentos que reflejan que nuestra propuesta entrega mejores resultados que las técnicas existentes de segmentación.

Palabras clave: Concept Drift, dimensión temporal, minería de procesos, segmentación.

Autores

Daniela Lorena Luengo Mundaca, nacida en Santiago de Chile, es Ingeniera Civil de Industrias mención en Tecnologías de la Información por la Pontificia Universidad Católica de Chile. Posee también el grado académico de Magíster en Ciencias de la Ingeniería de la misma casa de estudios. Adicionalmente tiene un Certificado académico en Actividad Física, Deporte, Salud y Educación. Del año 2009 al 2013 trabajó en el Centro de Estudios de Tecnologías de Información de la Pontificia Universidad Católica de Chile (CETIUC), como analista y consultora del área de excelencia de gestión de procesos, realizando levantamiento de procesos en terreno, propuestas de mejoras e investigaciones varias. Actualmente sus intereses están orientados en aplicar los conocimientos que posee, a través de distintos proyectos de emprendimiento.

Marcos Sepúlveda Fernández es Ingeniero Civil de Industrias mención en Computación de la Pontificia Universidad Católica de Chile. Posee también los grados académicos de Magíster y Doctor en Ciencias de la Ingeniería de dicha casa de estudios. Realizó un postdoctorado en la ETH Zürich, Suiza. Es profesor asociado en el Departamento de Ciencia de la Computación en la Escuela de Ingeniería de la Pontificia Universidad Católica de Chile. Desde el año 2001 es profesor en jornada completa de la Escuela de Ingeniería de la Universidad Católica, en el área de Tecnologías de Información. Sus intereses académicos están ligados a la minería de procesos, la modelación de procesos de negocio, el uso estratégico de las TI en las empresas, y la inteligencia de negocios. Es director del Centro de Estudios de Tecnologías de Información de la Pontificia Universidad Católica de Chile (CETIUC), el cual tiene como objetivo promover las mejores prácticas en el uso de las Tecnologías de Información y generar conciencia en los actores clave de la importancia de las TI para la generación de valor y para aumentar la eficiencia de los procesos de negocio.

posicionado como una disciplina emergente, proveyendo un conjunto de herramientas que ayudan a analizar y mejorar los procesos de negocio [1], en base al análisis de los registros de eventos que almacenan los sistemas de información durante la ejecución de un proceso. Sin embargo, a pesar de los avances desarrollados en este campo, aún existe un gran desafío, el cual consiste en incorporar el hecho de que los procesos cambian a lo largo

del tiempo, concepto que es conocido en la literatura como *Concept Drift* [2].

Dependiendo de la naturaleza del cambio, es posible encontrar diferentes tipos de *Concept Drift*, algunos de ellos son: *Sudden Drift* (cambio repentino y significativo a la definición del proceso), *Gradual Drift* (cambio gradual en la definición del proceso, permitiendo la existencia de dos definiciones de éste

“ Actualmente, uno de los problemas en la minería de procesos es que los algoritmos desarrollados suponen la existencia de información relativa a una única versión de un proceso en el *log* de eventos ”

de manera simultánea) o *Incremental Drift* (la evolución del proceso se realiza a través de pequeños cambios sucesivos a la definición del modelo).

Pese a que existen todas estas variantes de *Drift*, en la actualidad las técnicas de minería de procesos existentes están limitadas a encontrar los puntos en el tiempo en los que el proceso cambia, centrándose principalmente en cambios de tipo *Sudden Drift*. El problema de esta limitación es que en la práctica no es tan frecuente que los procesos de negocio muestren un cambio repentino de su definición.

Si aplicamos las técnicas de minería de procesos existentes en procesos que tengan cambios distintos a *Sudden Drift*, podríamos encontrarnos con resultados de poco sentido para el negocio.

En este documento proponemos un nuevo enfoque, permitiendo descubrir las versiones de un proceso cuando tiene distintos tipos de *Drift*, ayudando a entender cómo se comporta el proceso a través del tiempo. Para llevar a cabo esta tarea, se utiliza técnicas existentes de segmentación en minería de procesos, pero incorporando el tiempo como una variable adicional al control de flujo para generar los distintos segmentos o *clusters*. Se utilizan técnicas de *Trace Clustering*, las cuales, a diferencia de otras técnicas basadas en métricas para medir la distancia entre secuencias

completas, tienen una complejidad lineal, permitiendo la entrega de resultados en menores tiempos [3].

El enfoque de nuestro trabajo contribuye al análisis del proceso, permitiendo al responsable de la gestión del proceso tener una visión más realista de cómo se comporta el proceso en distintos intervalos de tiempo. Con este enfoque es posible determinar las distintas versiones del proceso, las características de cada una de ellas, e identificar en qué momento ocurren estos cambios.

Este artículo está organizado de la siguiente manera. La **sección 2** presenta el trabajo relacionado. La **sección 3** presenta el método de segmentación base y el modificado. La **sección 4** presenta experimentos y resultados obtenidos, para finalmente, presentar en la **sección 5** las conclusiones y el trabajo futuro.

2. Trabajo relacionado

La minería de procesos es una disciplina que ha concentrado gran interés en la actualidad. Esta disciplina asume que la información histórica almacenada en los sistemas de información sobre un proceso se encuentra en un registro, conocido como *log* de eventos [4]. Este registro contiene información histórica de las actividades que se llevan a cabo en cada ejecución del proceso, donde cada fila del registro está compuesta, al menos, por un identificador (id) asociado a cada ejecución

individual del proceso, el nombre de la actividad ejecutada, su marca de tiempo (día y hora en que ocurre la actividad) y, opcionalmente, información adicional, como el ejecutor de la actividad u otros. Adicionalmente, en la literatura [3] se define como traza de la ejecución de un proceso, a la lista ordenada de actividades invocadas por una ejecución en particular.

Actualmente, uno de los problemas en la minería de procesos es que los algoritmos desarrollados suponen la existencia de información relativa a una única versión de un proceso en el *log* de eventos. Sin embargo, esto muchas veces no se cumple, por lo que aplicar los algoritmos de minería de procesos a estos *logs* lleva a resultados poco representativos y/o de gran complejidad, que aportan poco a la tarea de análisis y mejora de procesos.

2.1. Segmentación del log de eventos

Para resolver el problema mencionado en la minería de procesos, se han propuesto técnicas de segmentación del *log* de eventos antes de aplicar técnicas de minería de procesos [5], las cuales consisten en dividir el *log* de eventos en *clusters* homogéneos, para luego aplicar de manera independiente las técnicas de minería de procesos sobre cada uno de ellos y así obtener información o modelos más representativos. La **figura 1** muestra la etapa de procesamiento del *log* y utiliza una técnica de descubrimiento como ejemplo de técnica de minería de procesos.

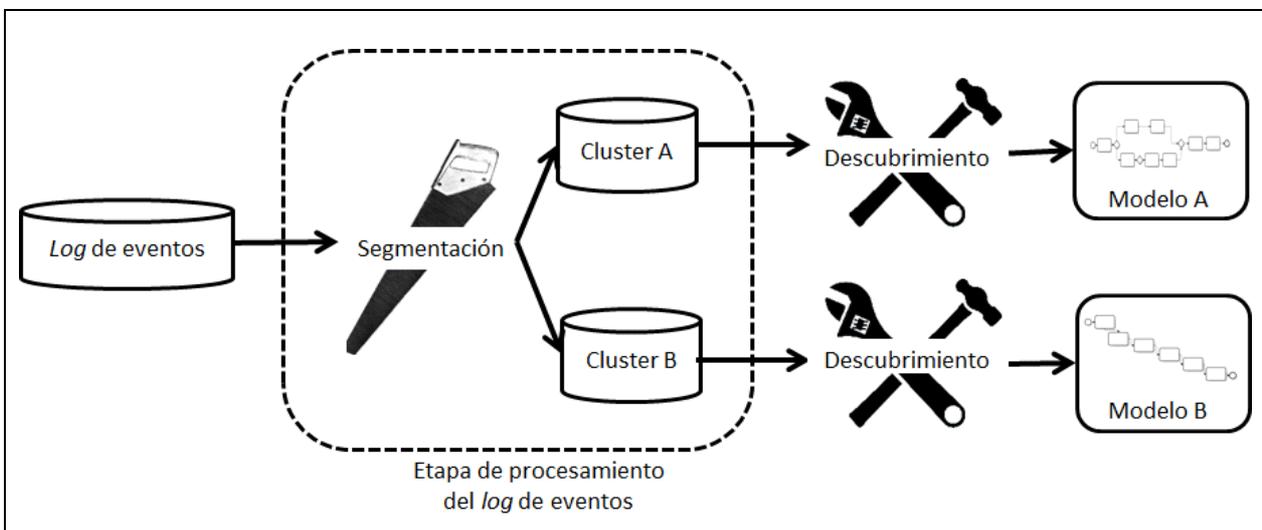


Figura 1. Etapa de procesamiento del *log* de eventos.

“*Concept Drift*, en BPM, se refiere a la situación en la cual un proceso ha sufrido cambios en su diseño dentro del periodo analizado y no se conoce el momento en que se produjeron los cambios”

Para realizar esta segmentación es necesario definir una manera de representar las trazas, de manera de poder agruparlas posteriormente de acuerdo a un criterio de similitud previamente definido. Actualmente existen varias técnicas de segmentación en la minería de procesos [3][6], donde la mayoría de ellas considera principalmente información del flujo de las actividades. Estas técnicas pueden ser clasificadas en dos categorías:

1) Técnicas que transforman las trazas en un espacio vectorial, en donde cada traza se convierte en un vector. La segmentación del *log* se puede hacer utilizando una variedad de técnicas de segmentación en el espacio vectorial, como por ejemplo: *Bag of activities*, *K-gram model* [6] y *Trace clustering* [5]. Sin embargo, estas técnicas tienen el problema que carecen de información de contexto, lo que se ha intentado resolver con la técnica *Trace clustering based on conserved patterns* [3].

2) Técnicas que operan con la traza completa. Estas técnicas utilizan métricas de distancia como *Levenshtein* y *Generic Edit Distance* [6], en conjunto con técnicas estándar de segmentación, asignando un costo a la diferencia entre trazas.

Sin embargo, las técnicas existentes en ambas categorías, a pesar de mejorar la segmentación a través de formar *clusters* de trazas estructuralmente similares, no consideran la dimensión temporal de la ejecución de los procesos, ni cómo el proceso va cambiando en el tiempo.

2.2. El desafío de *Concept Drift*

Concept Drift, en BPM, se refiere a la situación en la cual un proceso ha sufrido cambios en su diseño dentro del periodo analizado y no se conoce el momento en que se produjeron los cambios. Estos cambios se pueden deber a varios factores, pero principalmente se deben a la naturaleza dinámica de los procesos [7].

Los cambios en un proceso en los que se ha centrado el estudio de *Concept Drift*, tienen que ver con los cambios en la perspectiva de control de flujo, y pueden ser de dos tipos: cambios permanentes o cambios momentáneos, según la duración de los cambios.

Cuando ocurren cambios en periodos cortos de tiempo y pocas instancias se ven afectadas, entonces se habla de cambios momentáneos. Estos cambios también son reconocidos en el lenguaje de procesos, como ruido o anomalías.

Por otro lado, los cambios permanentes ocurren en periodos más prolongados de tiempo y/o hay una considerable cantidad de instancias afectadas por los cambios, lo cual hace referencia a un cambio en el diseño del proceso.

Nuestro interés se centra en los cambios permanentes en la perspectiva de control de flujo, los cuales pueden dividirse en las siguientes 4 categorías:

■ *Sudden Drift*: Se refiere a los cambios que ocurren de manera drástica, es decir, la forma de realizar el proceso cambia repentinamente de un momento a otro.

■ *Recurring Drift*: Cuando los cambios que sufre el proceso ocurren de manera periódica, es decir, una forma de hacer el proceso se repite en otro periodo de tiempo posterior.

■ *Gradual Drift*: Se refiere a cambios que no son drásticos, sino que en algún momento dos versiones del proceso se traslapan, ya que corresponde a una transición.

■ *Incremental Drift*: Es cuando un proceso tiene pequeños cambios incrementales. Este tipo de cambios es más frecuente en organizaciones que adoptan metodologías ágiles de BPM.

Para resolver el problema de *Concept Drift*, han surgido nuevos enfoques que analizan el dinamismo de los procesos.

Bose [2] propone métodos para manejar el *Concept Drift*, mostrando que los cambios en el proceso están indirectamente reflejados en el *log* de eventos, y la detección de estos cambios es factible examinando la relación entre las actividades. Para ello se definen distintas métricas. A partir de estas métricas se propone un método estadístico, cuya idea base es considerar una serie sucesiva de valores e investigar si hay una diferencia significativa entre dos series. Si es que existe, ésta correspondería a un cambio en el proceso.

Stocker [8] también propone un método para manejar *Concept Drift*, el cual considera las distancias entre pares de actividades de distintos trazos como una característica estructural para generar *clusters* cronológicamente subsecuentes.

Ambos enfoques se limitan a determinar el momento en el tiempo en el que el proceso cambia, por lo que se centran en procesos con cambios repentinos, dejando fuera otro tipo de cambios.

Para resolver esto, en un artículo anterior [9] propusimos un enfoque que utiliza técnicas de segmentación para descubrir los cambios

que puede sufrir un proceso a través del tiempo, pero sin limitarse a un tipo de cambio en particular. En este enfoque, la similitud entre dos trazas está definida por información del flujo de las actividades y por información del momento en el que se comienza a ejecutar cada traza.

En este trabajo, presentamos una extensión del artículo anterior [9], al incorporar una nueva forma de medir la distancia entre dos trazas.

3. Extendiendo la técnica de segmentación para incorporar la variable temporal

Como ya se mencionó, los enfoques existentes para tratar *Concept Drift* no son suficientemente efectivos para encontrar las versiones de un proceso cuando este tiene cambios de distintos tipos. Para resolver esta problemática recurrimos a la técnica *Trace Clustering* basado en conservación de patrones [3], que permite realizar segmentación del *log* de eventos considerando solo las secuencias de actividades de cada traza.

Nuestro trabajo se basa en esta técnica, y la extiende incorporando la variable temporal de manera adicional a las que ya utiliza para realizar la segmentación.

3.1. *Trace clustering* basado en conservación de patrones

La idea básica que planteamos en este artículo [3] es considerar subsecuencias de actividades que se repiten en múltiples trazos como conjuntos de características para realizar la segmentación. Cuando dos instancias tienen en común un significativo número de subsecuencias, entonces se asume que tienen una similitud estructural y estas instancias son asignadas al mismo *cluster*.

Se definen seis tipos de subsecuencias, pero haremos la definición formal solo de una (MR), ya que es la utilizada para desarrollar nuestro enfoque; el trabajo podría ampliarse y utilizar las otras subsecuencias.

■ *Maximal Repeat (MR)*: Un *Maximal Repeat* en una secuencia T, es definido como un par de subsecuencias idénticas, tal que los elementos inmediatamente a la derecha e inmediatamente a la izquierda sean distintos en T. Intuitivamente, una MR corresponde a una subsecuencia de actividades que se repite más de una vez en el *log*.

En la **tabla 1** se observa un ejemplo donde

“ La idea básica que planteamos en este artículo es considerar subsecuencias de actividades que se repiten en múltiples trazos como conjuntos de características para realizar la segmentación ”

Secuencia	Maximal Repeat	Conjunto de Características
bbbcd-bbbc-caa	{a, b, c, bb, bbbc}	{bb, bbbc}

Tabla 1. Ejemplo de Maximal Repeat y Conjunto de Características.

Traza\Conjunto de Características	bb	bbbc
bbbcd	2	1
bbbc	2	1
caa	0	0

Tabla 2. Matriz de características estructurales.

se determina las MR existentes en una secuencia. Lo que hace esta técnica es construir una única secuencia a partir del log de eventos, la cual es obtenida concatenando todas las trazas, pero incorporando un delimitador entre ellas. Luego, sobre esta única secuencia se aplica la definición de MR. El conjunto de todos las MR descubiertas en esta secuencia, con más de una actividad, es llamado Conjunto de Características.

A partir del Conjunto de Características, se crea una matriz que nos permite calcular la

distancia entre las distintas trazas. Cada fila de la matriz corresponde a una traza y cada columna a una característica del Conjunto de Características. Los valores de la matriz corresponden al número de veces que se encuentra cada característica en las distintas trazas (ver tabla 2). Esta matriz la llamaremos, matriz de características estructurales.

Este enfoque de segmentación basado en patrones, utiliza como técnica de segmentación el "Agglomerative Hierarchical Clustering" con criterio de mínima varianza, utilizando la

distancia euclidiana para medir la distancia entre trazas, la cual se define de la siguiente manera.

$$dist(A, B) = \sqrt{\sum_{i=1}^n (T_{Ai} - T_{Bi})^2}$$

Donde:

$dist(A, B)$ = distancia entre la traza A y la traza B
 n = número de características del Conjunto de Características
 T_{Ai} = número de veces que está la característica i en la traza A

3.2. Extendiendo Trace clustering para incorporar la variable temporal

Para identificar los distintos tipos de cambios que pueden ocurrir en los procesos de negocio, debemos buscar la manera de identificar las versiones de un proceso.

Si solo miramos las características estructurales dejamos fuera información respecto a su temporalidad. Ambas características, estructurales y temporales, son muy importantes, ya que la estructura nos indica cuán similar es una instancia a otra y la temporalidad nos indica qué tan cercanas en el tiempo están estas dos instancias. Nuestro enfoque busca identificar las distintas formas de ejecutar el proceso utilizando ambas características (estructurales y temporales) al mismo tiempo, tal como se ilustra en la figura 2.

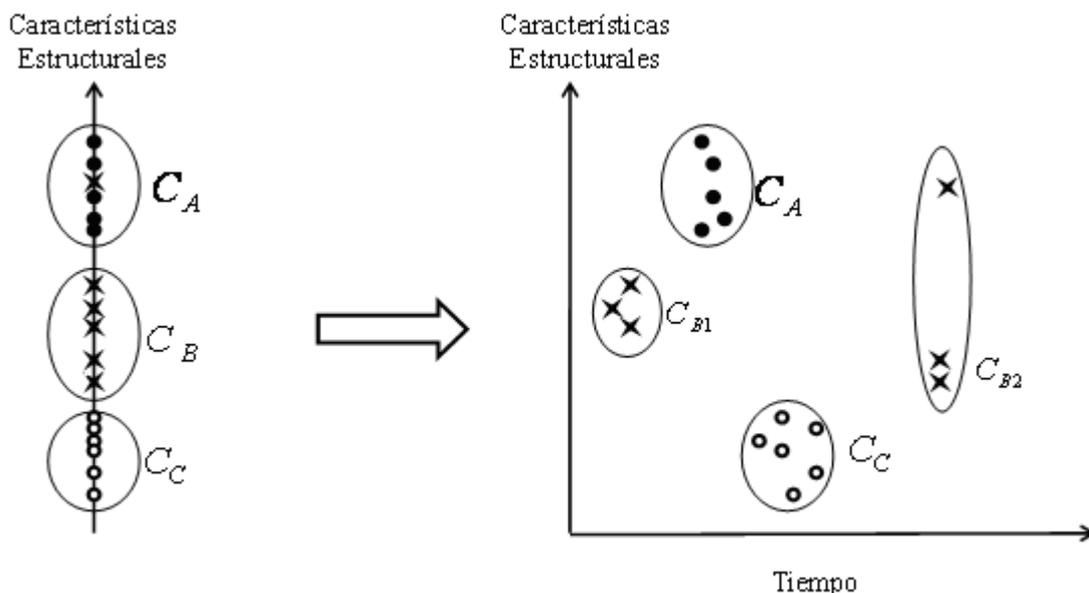


Figura 2. Ejemplo de la relevancia de considerar el tiempo en el análisis.

Traza\Conjunto de Características	bb	bbbc	Tiempo
bbbcd	2	1	tiempo 1
bbbc	2	1	tiempo2
caa	0	0	tiempo3

Tabla 3. Matriz de características estructurales más la dimensión temporal.

Consideramos que la variable temporal relevante de analizar es el tiempo en que se comienza a ejecutar cada instancia del proceso. Para cada traza, almacenamos en la dimensión tiempo, el número de días que han transcurrido desde una marca de tiempo de referencia, por ejemplo, días transcurridos desde el 1 de enero de 1970 hasta la marca de tiempo en la cual se comienza a ejecutar la primera actividad de la traza (ver tabla 3).

Este nuevo enfoque también utiliza "Agglomerative Hierarchical Clustering" con criterio de mínima varianza como técnica de segmentación.

Para calcular la distancia entre dos trazas utilizamos la distancia euclidiana, pero modificada, de tal manera que considere al mismo tiempo las características estructurales y la característica temporal.

Primero definimos T_{ji} , como la característica i de la traza J . Si la característica i no se encuentra en la traza J , su valor será 0, de lo contrario, su valor será el número de veces que la característica i se encuentra en la traza J .

$T_{j(n+1)}$ corresponde a la característica temporal de la traza J y su valor es el número de días (podrían ser horas, minutos o segundos, dependiendo del proceso) que han transcurrido desde una marca de tiempo de referencia. Se le da el índice $(n+1)$ para indicar que se agrega a las n características estructurales.

Definimos L como el conjunto de todas las trazas del log, luego la expresión $\max_{j \in L}(T_{ji})$ a la mayor cantidad de veces que está la característica i en alguna traza del log de eventos. De la misma forma, $\min_{j \in L}(T_{ji})$ corresponde a la menor cantidad de veces que está la característica i en alguna traza.

$\min_{j \in L}(T_{j(n+1)})$ y $\max_{j \in L}(T_{j(n+1)})$ corresponden al mayor y menor, respectivamente, instante de tiempo en que se comenzó a ejecutar una traza del log de eventos.

También, definimos $D_E(A, B)$ y $D_T(A, B)$, como la distancia estructural y la distancia temporal entre la traza A y la traza B respectivamente.

$$D_E(A, B) = \sqrt{\sum_{i=1}^n \left(\frac{T_{Ai} - T_{Bi}}{\max_{j \in L}(T_{ji}) - \min_{j \in L}(T_{ji})} \right)^2}$$

$$D_T(A, B) = \sqrt{\left(\frac{T_{A(n+1)} - T_{B(n+1)}}{\max_{j \in L}(T_{j(n+1)}) - \min_{j \in L}(T_{j(n+1)})} \right)^2}$$

Donde:

$n =$ número de características del Conjunto de Características Estructurales

Ambas distancias, D_E y D_T están normalizadas, sin embargo el dominio de D_E es mayor al de D_T . Es por ello que también definimos: Min_E, Max_E, Min_T and Max_T :

$$Min_E = \min_{A, B \in L} \sqrt{D_E(A, B)}, \quad A \neq B$$

$$Max_E = \max_{A, B \in L} \sqrt{D_E(A, B)}, \quad A \neq B$$

$$Min_T = \min_{A, B \in L} \sqrt{D_T(A, B)}, \quad A \neq B$$

$$Max_T = \max_{A, B \in L} \sqrt{D_T(A, B)}, \quad A \neq B$$

Min_E y Max_E corresponden a la distancia mínima y máxima (normalizada) entre todas las trazas, solo considerando las características estructurales.

Min_T y Max_T corresponden a la distancia mínima y máxima (normalizada) entre todas las trazas, solo considerando las características temporales.

La nueva forma para medir la distancia entre dos trazas, $dist(AB)$, incorpora el parámetro μ , al que llamaremos ponderador de la dimensión temporal, que sirve para ponderar las características estructurales y temporales. Adicionalmente, esta nueva forma para medir la distancia ajusta D_E y D_T , de tal manera que el peso de ambas distancias, D_E y D_T , sea equivalente.

$$dist(A, B) = (1 - \mu) \frac{D_E(A, B) - Min_E}{Max_E - Min_E} + \mu \frac{D_T(A, B) - Min_T}{Max_T - Min_T}$$

El ponderador de la dimensión temporal, μ , puede tener valores entre 0 y 1, según la relevancia que se le dé a la característica temporal.

4. Evaluación

Analizamos la técnica propuesta usando seis log de eventos obtenidos de distintos procesos sintéticos, los cuales se construyeron con CPN Tools [10][11].

Para medir el desempeño de la técnica utilizamos el *plug-in Guide Tree Miner* [3] disponible en ProM 6.1¹ y también una versión modificada de este *plug-in* que incorpora los cambios propuestos.

La evaluación se llevó a cabo usando distintas métricas para medir la efectividad de clasificación del nuevo enfoque versus el enfoque base.

4.1. Experimentos y resultados

En la figura 3 se muestra la secuencia de pasos realizada en los experimentos.

- 1) Se usó simulación a partir de modelos diseñados, en este caso M1 y M2, para crear el log sintético. Para esto se utilizó CPN Tools.
- 2) A partir del log sintético se utiliza alguna técnica de segmentación para generar clusters, en este caso dos (C_A y C_B). En los experimentos se aplican dos técnicas de segmentación:
 - Enfoque base: *Trace clustering* basada en conservación de patrones.
 - Nuestro enfoque: *Trace clustering* extendida, incorporando la variable temporal, utilizando el ponderador de la dimensión temporal, μ , con distintos valores entre 0 y 1.
- 3) Para cada uno de estos clusters se realiza descubrimiento de proceso, generando dos nuevos modelos (M_A y M_B).

El desempeño del enfoque puede ser medido en dos momentos:

- a) Conformidad 1: Las métricas se miden entre un modelo original y un cluster generado.
- b) Conformidad 2: Las métricas se miden entre un modelo original y un modelo generado.

Las métricas que se utilizan para medir Conformidad 1 son las siguientes:

- *Accuracy*: Número de instancias correctamente clasificadas en cada cluster, de acuerdo a lo que se conoce del log de eventos original. Sus valores están entre 0% y 100%, donde 100% corresponde a cuando la segmentación se hizo de manera exacta.
- *Fitness*: Indica cuánto del comportamiento observado en un log de eventos (por ejemplo, cluster C) es capturado por el modelo original del proceso (por ejemplo, modelo M_2) [12].

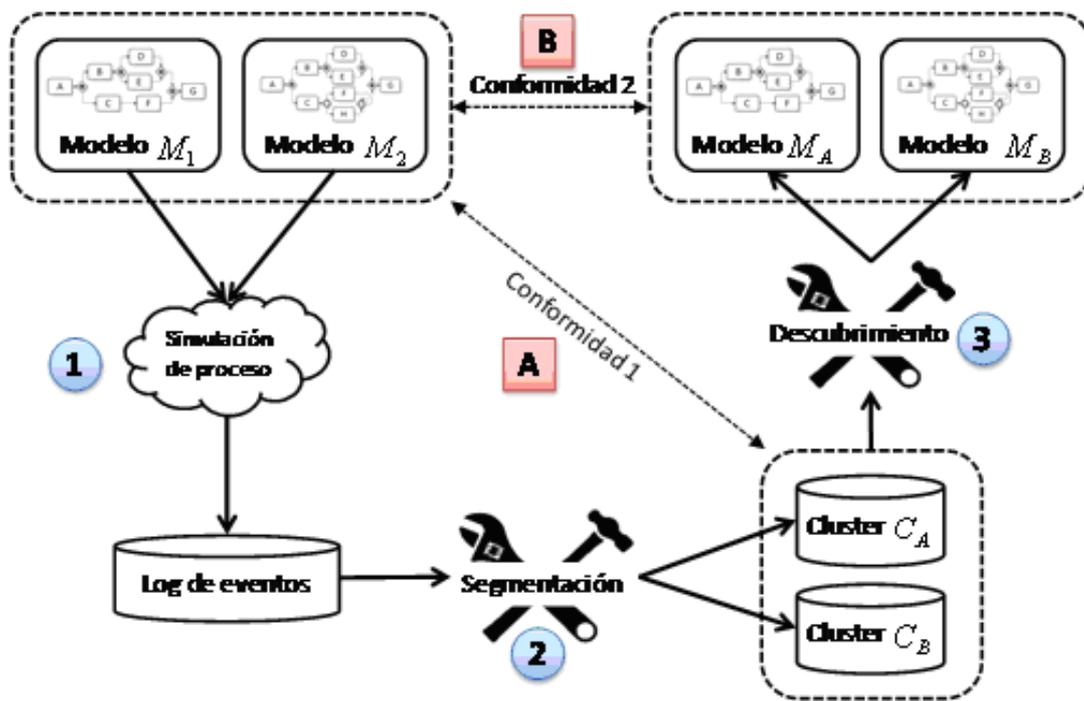


Figura 3. Pasos para realizar las pruebas experimentales.

Sus valores están entre 0 y 1, donde 1 corresponde a que el modelo es capaz de representar todas las trazas del log.

■ **Precisión:** Mide la generalidad del modelo, donde se prefiere modelos con un mínimo de comportamiento para representar lo mejor posible el registro del log de eventos. Sus valores están entre 0 y 1, donde 1 significa que el modelo no tiene comportamiento adicional a lo que indican los trazos [13].

Las métricas que se utilizan para medir Conformidad 2 son las siguientes [14]:

■ **Behavioral Precision (X) (B_p)**

■ **Structural Precision (X) (S_p)**

■ **Behavioral Recall (X) (B_r)^p**

■ **Structural Recall (X) (S_r)^R**

Estas métricas cuantifican precisión y generalidad de un modelo con respecto a otro. Sus valores están entre 0 y 1, donde 1 es el mejor valor esperado.

La **tabla 4** resume los resultados de aplicar el enfoque base y el nuevo enfoque (variando el valor del parámetro μ), a los distintos logs de eventos sintéticos creados. En esta tabla se muestra la métrica de *accuracy*, la cual nos

indica el porcentaje de trazas correctamente clasificadas.

Para cada log utilizado, el porcentaje de *accuracy* más alto se alcanza con nuestro enfoque, pero con distintos valores de μ (varía entre 0,2 y 0,9). La razón de esto se debe a que en cada log no existe la misma relevancia entre la distribución temporal de las trazas versus la estructura del proceso.

Utilizamos el Log (f) para hacer un análisis más profundo de los resultados, midiendo para este Log, todas las métricas definidas

Enfoque	μ	Log (a)	Log (b)	Log (c)	Log (d)	Log (e)	Log (f)
Base	-	57%	38%	55%	86%	99%	53%
Nuevo	0,0	59%	52%	49%	82%	59%	51%
	0,1	65%	52%	49%	82%	59%	68%
	0,2	87%	52%	63%	100%	59%	64%
	0,3	87%	52%	63%	100%	59%	62%
	0,4	95%	52%	63%	100%	59%	63%
	0,5	100%	52%	63%	100%	59%	95%
	0,6	100%	52%	73%	100%	100%	94%
	0,7	96%	89%	73%	100%	100%	67%
	0,8	78%	88%	73%	74%	84%	55%
	0,9	79%	77%	77%	68%	77%	78%
1,0	81%	78%	68%	73%	72%	55%	

Tabla 4. Métrica de *accuracy* calculada para los 6 logs sintéticos de prueba.

Enfoque	μ	Accuracy	Fitness	Precisión	Promedio B_P y S_P	Promedio B_R y S_R	Promedio
Base	-	53%	0,93	0,78	0,77	0,81	0,76
Nuevo	0,0	51%	0,93	0,73	0,73	0,70	0,72
	0,1	68%	0,93	0,81	0,86	0,86	0,83
	0,2	64%	0,92	0,80	0,83	0,82	0,80
	0,3	62%	0,92	0,80	0,80	0,78	0,78
	0,4	63%	0,92	0,79	0,83	0,86	0,81
	0,5	95%	0,95	0,85	0,97	0,96	0,94
	0,6	94%	0,95	0,86	0,95	0,94	0,93
	0,7	67%	0,92	0,84	0,84	0,89	0,83
	0,8	55%	0,94	0,87	0,88	0,94	0,84
	0,9	78%	0,94	0,81	0,89	0,95	0,87
1,0	55%	0,93	0,87	0,88	0,95	0,84	

Tabla 5. Detalle de métricas al analizar el Log (f).

tanto en Conformidad 1 como en Conformidad 2 (ver tabla 5).

Todas las métricas calculadas para el Log (f) tienen buen desempeño cuando el parámetro μ vale 0,5 o 0,6. Al ser promediadas las cinco métricas, el promedio más alto se alcanza con μ igual a 0,5.

5. Conclusiones y trabajo futuro

En este documento se presentan las limitaciones de las actuales técnicas de segmentación en minería de procesos, las cuales se centran en agrupar ejecuciones similares (estructuralmente) de un proceso, de tal manera que se formen grupos homogéneos de ejecuciones, para que el análisis sobre cada grupo sea de mayor simplicidad que si se analiza el conjunto de datos completo.

Al centrarse solo en la estructura de las ejecuciones, dejan de lado el comportamiento del proceso a través del tiempo. Para ello surgen nuevas técnicas, pero que también presentan limitaciones, ya que se centran en encontrar los puntos en que cambia el proceso, limitándose a un tipo de cambio, el *Sudden Drift*. Dada esta situación, presentamos un enfoque que utiliza la lógica que usan las técnicas de segmentación, de manera que se encuentren las distintas versiones de un proceso cuando éste presente distintos tipos de cambios, permitiendo entender las variaciones que ocurren en el proceso y cómo realmente se está ejecutando en la práctica a través del tiempo.

Nuestro trabajo se centra en la identificación de las versiones del proceso. La técnica que proponemos, es una herramienta que ayuda a las personas involucradas en el negocio a tomar decisiones. Por ejemplo, se puede determinar si los cambios que se producen en la ejecución del proceso, son realmente los esperados, y en base a esto, tomar medidas si se descubren comportamientos anormales.

También sirve para eventualmente identificar buenas y malas prácticas, las cuales son de alta utilidad al momento de querer mejorar o estandarizar los procesos.

En este documento presentamos un conjunto de métricas para medir el desempeño del enfoque. El desempeño se considera bueno, cuando el enfoque es capaz de segmentar los datos de la misma forma en la que fueron creados. Por lo tanto, estas métricas requieren información a priori del proceso, lo cual no es aplicable en casos reales.

Cada métrica aquí presentada mide distintos aspectos que son difíciles de analizar por sí solos. Sin embargo, al utilizarlos en conjunto, permiten tener una visión de distintas perspectivas del problema, haciendo más completo el análisis.

Un aspecto clave de nuestro enfoque de segmentación, es el valor que se le da al ponderador de la dimensión temporal, μ , el cual está estrechamente relacionado con la naturaleza del proceso. Valores altos de μ le dan mayor importancia al tiempo para realizar la segmentación, mientras que valores bajos de μ , le dan más importancia a las características estructurales del proceso.

Los resultados de los experimentos muestran que el enfoque propuesto en este documento tiene un mejor desempeño en la segmentación del log y que existe al menos un valor del parámetro μ que permite entregar mejores resultados en comparación a utilizar solo la técnica de segmentación estructural. Esto se logra, ya que nuestro enfoque es capaz de agrupar las trazas del log de tal manera que se identifiquen similitud estructural y cercanía temporal al mismo tiempo.

Una de las métricas utilizadas es el *accuracy*. En algunos experimentos, esta métrica, al-

canza el 100%, es decir, que todas las trazas son clasificadas correctamente.

Cuando no se alcanza el 100% de *accuracy*, se debe a que hay procesos que tienen versiones que pese a ser distintas, son similares entre sí, pudiendo incluso tener trazas ejecutables en las dos versiones del proceso, lo cual hace que la clasificación no sea exactamente igual a lo que se esperaba.

Nuestro trabajo futuro en esta línea de investigación es probar este nuevo enfoque con procesos reales. También queremos trabajar en desarrollar los algoritmos existentes para que sean capaces de determinar automáticamente el número óptimo de *clusters*; para ello es necesario definir nuevas métricas que nos permitan calcular el número óptimo de *clusters* sin saber a priori información de las versiones del proceso.

Referencias

- [1] **W. van der Aalst.** *Process Mining, Discovery, Conformance and Enhancement of Business Processes.* Springer, 2011. ISBN 978-3-642-19345-3.
- [2] **R.P. Bose, W. van der Aalst, I. Zliobaite, M. Pechenizkiy.** Handling Concept Drift in Process Mining. *23rd International Conference on Advanced Information Systems Engineering.* Londres, 2011.
- [3] **R.P. Bose, W. van der Aalst.** Trace Clustering Based on Conserved Pasterns : Towards Achieving Better Process Models. *Business Process Management Workshops*, pp. 170-181, 2010.
- [4] **W. van der Aalst, B. van Dongen, J. Herbst, L. Maruster, G. Schimm, A. Weijters.** *Data & Knowledge Engineering*, pp. 237-267, 2003.
- [5] **M. Song, C. Günther, W. van der Aalst.** Trace Clustering in Process Mining. *4th Workshop on Business Process Intelligence (BPI08)*, pp. 109-120. Milano, 2009.
- [6] **R. Bose, W. van der Aalst.** Context Aware Trace Clustering: Towards Improving Process Mining Results. *SIAM*, pp. 401-412, 2009.
- [7] **W. van der Aalst, A. Adriansyah, A.K. Alves de Medeiros, F. Arcieri, T. Baier, T. Blickle y otros.** *Process Mining Manifesto*, 2011.
- [8] **T. Stocker.** Time-based Trace Clustering for Evolution-aware Security Audits. *Proceedings of the BPM Workshop on Workflow Security Audit and Certification*, pp. 471-476. Clermont-Ferrand, 2011.
- [9] **D. Luengo, M. Sepúlveda.** Applying Clustering in Process Mining to find different versions of a business process that changes over time. *Lecture Notes in Business Information Processing*, pp. 153-158, 2011.
- [10] **A.V. Ratzer, L. Wells, H.M. Lassen, M. Laursen, J. Frank, M.S. Stissing y otros.** CPN Tools for editing, simulating, and analysing coloured Petri nets. *Proceedings of the 24th international conference on Applications and theory of Petri nets* pp. 450-462. Eindhoven: Springer-Verlag, 2003.
- [11] **A.K. Alves De Medeiros, C. Günther.** Process Mining: Using CPN Tools to Create Test Logs for Mining Algorithms. *Proceedings of the Sixth Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools*, pp. 177-190, 2005.
- [12] **A. Rozinat, W. van der Aalst.** Conformance testing: Measuring the fit and appropriateness of event logs and process models. *Business Process Management Workshops*, pp. 163-176, 2006.
- [13] **J. Muñoz-Gama, J. Carmona.** A fresh look at precision in process conformance. *Proceeding BPM'10 Proceedings of the 8th international conference on Business process management*, pp. 211-226, 2010.
- [14] **A. Rozinat, A.K. Alves De Medeiros, C. Günther, A. Weijters, W. van der Aalst.** *Towards an Evaluation Framework for Process Mining Algorithms.* Genetics, 2007.

Nota

¹ ProM es un *framework* extensible que soporta una variedad de técnicas de minería de procesos en forma de *plug-ins*. Se puede conseguir en <<http://www.processmining.org/>>.