

Minería de procesos de software: una revisión de experiencias de aplicación

Brian Keith Norambuena¹, Vianca Vega Zepeda¹

brian.keith@ucn.cl, vvega@ucn.cl

¹ Departamento de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Angamos 0610, 1270709, Antofagasta, Chile.

DOI: 10.17013/risti.21.51-66

Resumen: La minería de procesos es un campo que utiliza elementos de la minería de datos y el modelamiento de procesos para realizar tareas de descubrimiento, evaluación de conformidad y mejora de procesos. El objetivo de este artículo es mostrar la aplicación y utilidad de la minería de procesos en la ingeniería de software. Se presenta una serie de casos de estudio que ilustran aplicaciones sobre el proceso y el producto. Además, se muestran los desafíos a los que se enfrenta la minería de procesos en la ingeniería de software. El principal resultado de esta investigación corresponde a que el uso de minería de procesos facilita la evaluación y auditoría de procesos de software. Se propone como trabajo futuro el desarrollo de una metodología para la aplicación de la minería de procesos en la ingeniería de software.

Palabras-clave: minería de procesos; desarrollo de software; evaluación de procesos; mejora de procesos.

Process mining applications in software engineering

Abstract: Process mining is a field that uses elements from data mining and business process modeling to do tasks such as process discovery, conformance checking, and process improvement. The objective of this paper is to show the importance and practical usefulness of applying process mining approaches in software engineering. It shows a series of case studies that illustrate possible applications in the process and the product. Also, the main current challenges in applying process mining in software engineering are described. The main result of this study is the fact that using process mining facilitates software process evaluation and auditing. The development of a methodology for applying process mining in software engineering is proposed as future work.

Keywords: process mining; software development; process assessment; process improvement.

1. Introducción

La minería de procesos es una disciplina relativamente joven que se encuentra entre el aprendizaje automático y la minería de datos, por una parte, y entre el modelamiento y

análisis de procesos, por otra. Su objetivo principal es realizar tareas de descubrimiento, revisión de conformidad y mejora de procesos. Específicamente, la minería de procesos busca automatizar estas tres tareas mediante la aplicación de técnicas de minería de datos especialmente diseñadas para tratar con datos de procesos (Van der Aalst et al. 2011; Van der Aalst, 2016).

Esta capacidad de analizar automáticamente un proceso y determinar dónde se producen las desviaciones puede ser utilizada por las organizaciones para realizar acciones correctivas sobre sus procesos. Esto corresponde a una estrategia de mejora de procesos, donde en función de la información obtenida se proponen las mejoras correspondientes (Van der Aalst, 2016).

Uno de los primeros trabajos realizados en minería de procesos corresponde al artículo de Cook y Wolf en el año 1998, quienes intentaron construir modelos de procesos de software a partir de los datos contenidos en registros de eventos (Cook & Wolf, 1998; Kimble et al., 2016). Desde el planteamiento original realizado por Cook y Wolf, la minería de procesos ha evolucionado y se ha convertido en un campo en sí misma, teniendo ahora aplicaciones que van más allá de la ingeniería de software.

Considerando que uno de los primeros trabajos de minería de procesos corresponde a una aplicación concreta en la ingeniería de software, es natural preguntarse qué otras aplicaciones se le ha dado a la minería de procesos en este campo desde entonces. La minería de procesos aplicada a la ingeniería de software busca en una primera instancia resolver el problema del monitoreo, control y mejora de los procesos de desarrollo, uno de los problemas críticos dentro del desarrollo de software (Samalikova, Kusters, Trienekens, Weijters, & Siemons, 2011). En general, se espera que una mejora en la calidad del proceso de desarrollo produzca una mejor calidad en el producto. Esto refuerza la utilidad que tiene el aplicar técnicas de minería de procesos en este campo, pues eventualmente permitiría desarrollar software de mejor calidad.

Otro aspecto relevante que se debe considerar, es que los procesos de software no suelen encontrarse explícitamente modelados, y los manuales que soportan el trabajo de desarrollo contienen guías y procedimientos abstractos. En consecuencia, hay una gran brecha entre el proceso “real” y el proceso “oficial” (Humphrey, 1995; Samalikova et al., 2011). Esta brecha puede ser medida mediante técnicas de minería de procesos, permitiendo esto corregir y mejorar el proceso.

Analizando lo anterior y bajo la perspectiva de que los procesos de software se pueden considerar como software en sí mismos (Osterweil, 1987), la aplicación de técnicas de minería de procesos a éstos se podría ver como el equivalente a la aplicación de métodos de testing automatizados en el desarrollo de software.

En función de la revisión sistemática realizada sobre los trabajos de minería de procesos en ingeniería de software, se observa que existen dos dimensiones principales de aplicación: en el proceso y en el producto. En la Tabla 1 se observan los detalles de las posibles aplicaciones.

La falta de un proceso bien definido y formalizado es un problema presente en múltiples organizaciones y en particular en las pequeñas y medianas empresas de desarrollo de software. La construcción de herramientas que apoyen la gestión de conocimiento de

Dimensión	Aplicación	Descripción	Ejemplo
<i>Proceso</i>	Evaluación y mejora de procesos	Enfoque tradicional de minería de procesos en los procesos de desarrollo de software.	Encontrar discrepancias entre el proceso formal y real.
	Apoyo a procesos	Aplicación de técnicas de minería de procesos para apoyar el desarrollo de una actividad del proceso de desarrollo de software.	Usar descubrimiento de procesos para complementar la elicitación de requisitos.
<i>Producto</i>	Sobre el producto	Aplicación de técnicas de minería de procesos sobre el software mismo.	Usar minería de procesos para analizar el comportamiento de los usuarios.

Tabla 1 – Aplicaciones de la minería de procesos a la ingeniería de software.

los procesos es útil para estas organizaciones (Mejia & Uribe, 2015). El modelamiento del proceso permitiría a las organizaciones guiar las actividades para mejorarlo (Muñoz, Gasca & Valtierra, 2014). La aplicación de la minería de procesos, específicamente la tarea de descubrimiento de procesos, puede ser de utilidad para las organizaciones que se enfrenten a este desafío.

El propósito de este trabajo es estudiar las aplicaciones de la minería de procesos en la ingeniería de software. Existen estudios que analizan el uso de técnicas de minería de datos en ingeniería de software en general (Halkidi, Spinellis, Tsatsaronis, & Vazirgiannis, 2011). No obstante, ninguno se centra en la aplicación de técnicas de minería de procesos en ingeniería de software propiamente tal.

Habiendo hecho esta breve introducción de la problemática que se abordará, el resto de este documento se estructura de la siguiente forma: primero, se explican los conceptos fundamentales de la minería de procesos para luego presentar una revisión de las principales aplicaciones de la minería de procesos a la ingeniería de software. A continuación, se detallan algunas aplicaciones específicas como casos de estudio, analizando sus principales resultados. Luego, se explicarán algunos de los desafíos que enfrenta la minería de procesos en este campo. Finalmente, se resumen los principales resultados y se muestran las conclusiones obtenidas en el desarrollo de este trabajo de investigación.

2. Minería de procesos

La tarea fundamental de la minería de procesos es el descubrimiento de procesos. Esta tarea toma como entrada un conjunto de datos conformado por un registro de todos los eventos (event log) realizados en el desarrollo del proceso. En su forma más simple debe tener un identificador de la instancia del proceso, un identificador de la tarea actual y un tiempo. Con suficientes de estos datos es posible inferir el modelo de proceso, con niveles ajustables de detalle (Van der Aalst, 2016). La automatización del descubrimiento de procesos permite a las organizaciones eliminar el problema en el modelamiento de procesos realizados por expertos pues en muchos casos estas evaluaciones tienen influencia de los propios sesgos del modelador (Gottschalk, Van der Aalst, & Jansen-Vullers, 2008).

La calidad de los modelos de procesos resultantes es función de la calidad y cantidad de datos disponibles (Bose, Mans, & Van der Aalst, 2013). Estos datos pueden ser aumentados con información adicional tales como los recursos utilizados o los tiempos de ejecución de cada tarea. Por otra parte, dada la naturaleza de los datos disponibles, los algoritmos de minería de procesos sólo tienen acceso a ejemplos que han ocurrido, pero no tiene ejemplos de instancias de procesos que no pueden ocurrir. La falta de ejemplos negativos es una de las debilidades que presenta la minería de procesos (Van der Aalst, 2016).

La revisión de conformidad (conformance checking en inglés) toma como entrada un modelo de proceso formal definido, una instancia del proceso real y genera un análisis de la desviación del proceso real con respecto al ideal. Esta tarea es de gran utilidad a la hora de auditar procesos, y ha sido aplicada con éxito en diversas organizaciones (Van der Aalst, Van Hee, Van Werf, & Verdonk, 2010). La revisión de conformidad permite determinar qué tan fielmente se está siguiendo el modelo de proceso según se ha definido formalmente en la realidad. Para realizar esta tarea existe una serie de métodos, todos estos tienen en común que permiten definir el nivel de desviación de un proceso real con respecto al proceso formal y específicamente que pueden determinar en qué parte del proceso ocurre la desviación. Esto tiene su aplicación en el control y auditoría de los procesos dentro de las organizaciones (Van der Aalst, Van Hee, et al., 2010; Van der Aalst, 2016).

La mejora de procesos se basa en los resultados obtenidos en las tareas anteriores. Esta tarea toma como entrada el modelo de proceso actual y un indicador de rendimiento y genera como salida un nuevo modelo de proceso que debiese mejorar los resultados con respecto al modelo original (Van der Aalst, 2016). El ejemplo más simple corresponde a la eliminación de un cuello de botella mediante la detección de este en una tarea de descubrimiento de procesos. En este enfoque, es posible utilizar técnicas de simulación para analizar las mejoras propuestas y compararlas en función de una métrica bien definida (Van der Aalst, 2012).

Los modelos se pueden trabajar mediante diferentes notaciones formales, yendo desde los modelos de más bajo nivel correspondientes a diagramas de transición, a notaciones más avanzadas tales como las redes de Petri, BPMN y diagramas UML (Kalenkova, Van der Aalst, Lomazova, & Rubin, 2016).

La minería de procesos ha sido utilizada con éxito en diversos dominios de aplicación, tales como el área de la salud, turismo, la educación, el comercio en línea y la auditoría de sistemas (Zinski, 2001; Trcka & Pechenizkiy, 2009; Van der Aalst, Van Hee, et al., 2010). Existe una comunidad IEEE de minería de procesos que incluye a más de 60 organizaciones (Samalikova, Trienekens, Kusters, & Weijters, 2009). A pesar del éxito de la minería de procesos, una limitación existente para su aplicación es que las técnicas rara vez son utilizadas en ambientes operacionales debido a la complejidad asociada. No obstante, se ha demostrado que es posible utilizar la minería de procesos como un apoyo a las decisiones operacionales (Van der Aalst, Pesic, & Song, 2010).

3. Aplicaciones

La gran cantidad de datos existentes hoy en día como un subproducto del proceso de desarrollo de software permite aplicar diferentes técnicas de análisis de datos y utilizar dichos

resultados para guiar la optimización de los distintos procesos. Existen varios ejemplos de técnicas de minería de datos para el análisis y mejora de los procesos de software. Los lectores interesados pueden referirse al trabajo realizado por Halkidi et al. (2011). En particular, el presente trabajo se centra en las aplicaciones de la minería de procesos en este campo.

La evaluación de procesos clásica y la aplicación de acciones correctivas y de mejora a los procesos de software se basa en modelos obtenidos mediante entrevistas y documentación (e.g., manuales de calidad). No obstante, estos modelos no son completamente adecuados debido a que no consideran las instancias en que los procesos formales no son respetados dentro de una organización (e.g., no realizar una actividad por presiones de tiempo). Entender dónde se producen estas desviaciones y comprender sus causas e impactos, permitiría a la organización mejorar sus procesos teniendo en cuenta la realidad. También es importante destacar que, en muchas organizaciones, los procesos de desarrollo de software no se encuentran modelados explícitamente y en muchos casos son caóticos (Rubin et al., 2007). Es aquí donde entra el uso de la minería de procesos para detectar estos elementos.

Para poder aplicar las técnicas de minería de procesos, es necesario que los datos disponibles se ajusten al meta modelo descrito en la Figura 1. Específicamente, el registro de eventos debe contener datos de una cierta cantidad de procesos (generalmente uno). Cada proceso se compone de actividades y tiene asociada una serie de instancias de proceso. Cada instancia de proceso se compone de uno o más eventos. Estos eventos deben indicar la actividad a la que corresponden, deben tener una descripción, un tiempo de ejecución y un responsable. Es posible enriquecer el registro de eventos con más información según se tenga disponibilidad (Poncin, Serebrenik, & Van Den Brand, 2011).

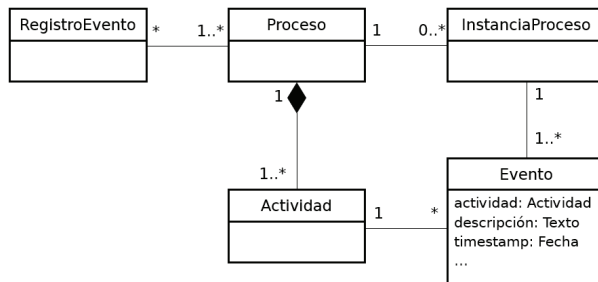


Figura 1 – Meta modelo de minería de procesos (traducido de Poncin et al., 2011).

En el caso de la ingeniería de software existen varias fuentes de donde se podría obtener datos (e.g., repositorios de control de versiones, sistemas de bug tracking y correos electrónicos). Debido a la multiplicidad de fuentes disponibles, se requiere una fuerte etapa de preparación de los datos antes de aplicar los procedimientos de minería de datos (Poncin et al., 2011).

La amplia aplicabilidad de la minería de procesos en la ingeniería de software se debe a lo genérico de sus técnicas. Éstas pueden ser aplicadas a cualquier tipo de proceso, siempre que los datos se ajusten al meta modelo de la Figura 1. La minería de procesos se

puede utilizar también para minar el uso mismo de los softwares desarrollados, un caso de estudio analiza dos aplicaciones industriales, específicamente un sistema de reservas y un portal de viajes (Rubin, Mitsyuk, Lomazova, & Van der Aalst, 2014).

Una posible aplicación de la minería de procesos sobre el proceso de desarrollo de software es la exploración de los procesos históricos con el propósito de determinar sus variaciones y poder realizar adaptaciones y mejoras del proceso teniendo en cuenta la realidad. A diferencia de una revisión de conformidad, el objetivo es directamente mejorar el proceso, en vez de medir la diferencia con el proceso idealizado. Esta información puede ser utilizada por los gestores de proyectos e ingenieros de proceso que deben determinar qué variaciones se incluirán en el proceso final (Santos & Oliveira, 2015).

Otra de las aplicaciones sobre el proceso es la auditoría de procesos de desarrollo. Aunque la minería de procesos puede agregar valor a la auditoría de procesos, y se espera que con el avance de las tecnologías y la maduración de éstas los auditores de procesos las asimilen adecuadamente. Este aspecto no ha sido estudiado exhaustivamente en todas sus posibles ramificaciones, debido a que la aplicación de estas técnicas requiere como punto de partida procedimientos de auditoría manuales (Van der Aalst, Van Hee., 2010; Jans, Alles, & Vasarhelyi, 2013).

La minería de procesos puede utilizarse como apoyo en el proceso de ingeniería de requisitos. Específicamente, se puede utilizar para regular que se esté siguiendo el proceso de ingeniería de requisitos adecuadamente (e.g., monitorear los cambios de los documentos de cada fase) y también puede utilizarse para verificar que los modelos propuestos por los ingenieros de desarrollo concuerden con instancias reales de los procesos que ejecutará el software (Van der Aalst, 2005).

Por otra parte, en el campo de la ingeniería de software dirigida por modelos, la minería de procesos también se ha aplicado con éxito para acortar la brecha existente entre la fase de diseño y la fase de ejecución. Si bien los resultados hasta ahora son prometedores, existen todavía desafíos pendientes en dicho campo (Mazak & Wilmer, 2016).

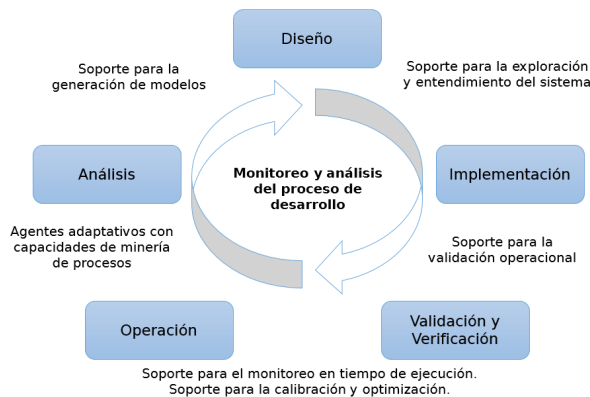


Figura 2 – Aplicaciones de minería de procesos en RUP (traducido de Cabac & Denz, 2015).

La minería de procesos puede aplicarse en el proceso de diversas formas para apoyar el proceso de desarrollo (Cabac & Denz, 2015). La Figura 2 ilustra algunas de las posibles aplicaciones de la minería de procesos en el contexto del desarrollo de software. El ciclo de vida presentado es genérico y sigue las etapas del Proceso Racional Unificado (RUP) (Kruchten, 2004).

Existen propuestas metodológicas para la aplicación de minería de procesos en la ingeniería de software, además de frameworks diseñados específicamente para facilitar estas tareas (Rubin et al., 2007; Poncin et al., 2011; Sureka, Kumar, & Gupta, 2015; Stolfa, Stolfa, Kosinar, & Snasel, 2016). No obstante, el desarrollo de una metodología que se considere estándar todavía es un problema abierto en esta área. Esto se debe a que la inteligencia de procesos de software es un campo todavía muy reciente y que se encuentra aún en evolución (Sureka et al., 2015).

4. Casos de estudio

En esta sección se muestran los resultados de algunos de los trabajos presentes en la literatura. Estos estudios han sido seleccionados debido a su valor ilustrativo. Los casos de estudio han sido clasificados de acuerdo al proceso involucrado. Solo los resultados principales se presentan aquí, el detalle de las implementaciones y el modelamiento se pueden encontrar en las referencias entregadas.

4.1. Proceso de control de configuración

El primer caso estudiado corresponde a una aplicación sobre el proceso de control de configuración. Con el propósito de llevar un control de todos los documentos y archivos necesarios, los ingenieros utilizan sistemas de gestión de configuración de software (SCM, por sus siglas en inglés). En su funcionamiento, estos sistemas recolectan y almacenan información sobre el proceso de software en sí. Luego, la información de los SCM puede ser utilizada para construir modelos de proceso explícitos, esto corresponde a un ejemplo de descubrimiento de procesos. (Rubin et al., 2007)

Utilizando la información de auditoría generada por el sistema de gestión de configuración, que corresponden a instancias del proceso (ejecuciones particulares del proceso de desarrollo), y luego aplicando los algoritmos de minería de procesos se deriva el modelo a partir de éstos. Este modelo puede ser verificado por el ingeniero de procesos, que debe luego decidir qué cambios deberían realizarse para optimizarlo y gestionarlo de mejor forma (Rubin et al., 2007).

Complementando la información del proceso con los responsables de cada actividad, es posible construir el modelo de red social del proceso y realizar un análisis sobre quiénes realizan cada actividad (Van der Aalst, 2016). La aplicación de esto reveló que en algunos casos los encargados del desarrollo del código eran los mismos encargados de las pruebas de éste. Esto por supuesto no es deseable pues en situaciones ideales serían personas diferentes (Rubin et al., 2007).

4.2. Proceso de pruebas

Los siguientes casos estudiados corresponden a aplicaciones sobre el proceso de pruebas y sobre el producto desarrollado. En particular, es posible aplicar las técnicas de minería de

procesos para obtener modelos a partir de los scripts de prueba utilizados durante la fase de testing en el desarrollo de software. Esto permite generar nuevos casos de prueba y facilita el entendimiento de los distintos casos de prueba utilizados (Shah, Khadke, & Rana, 2015).

En otro trabajo se propone utilizar técnicas de minería de procesos para analizar sistemas de software en circunstancias reales en línea. El objetivo es aprender lo máximo posible de la información obtenida de estos datos en tiempo de ejecución, y subsecuentemente utilizar este conocimiento para diagnosticar problemas y recomendar acciones (Van der Aalst, 2015).

A diferencia de los enfoques tradicionales de testing (tanto enfoques de caja negra como de caja blanca), el trabajo se focaliza en el producto trabajando en su ambiente natural y no en un experimento controlado. Además, se consideran muchas variantes del mismo software siendo ejecutadas en situaciones posiblemente muy diferentes. La mayoría de los enfoques de testing no son aplicados en vivo y limitan al sistema utilizando entradas usuales e inusuales de manera controlada. Esta variedad de condiciones permitiría realizar un proceso de pruebas más robusto (Van der Aalst, 2015).

El primer trabajo (Shah et al., 2015) corresponde a un ejemplo de aplicación de la minería de procesos como una herramienta de soporte dentro del proceso, pues el propósito es apoyar el desarrollo de la fase de testing y no analizar el proceso. Por otra parte, el segundo trabajo (Van der Aalst, 2015) corresponde también a un ejemplo de apoyo al proceso, pero además se aplica una aplicación sobre el producto, pues las herramientas de minería de procesos se aplican directamente sobre el software con el propósito de apoyar la fase de pruebas.

4.3. Proceso de control de cambios

El siguiente caso estudiado corresponde a una aplicación sobre el proceso de control de cambios. En particular, el estudio detalla la aplicación de técnicas de minería de procesos sobre los datos generados en el proceso de control de cambios a cargo de un Comité de Control de Cambios en el contexto de un proyecto de desarrollo de software. Uno de los principales resultados obtenidos reveló que en la mayoría de las instancias de proceso (alrededor del 70%), se omitía la fase de análisis y que las solicitudes de cambio eran directamente solucionadas sin realizar un análisis previo (Samalikova et al., 2009).

En general, la cantidad de casos que no cumplen con la especificación del modelo de proceso es alta. A lo largo del proyecto la cantidad de casos que no realizaban la tarea de análisis se mantiene relativamente constante. Esto indica que la omisión de la tarea de análisis es un problema estructural dentro de esta organización (Samalikova et al., 2009). Este tipo de discrepancias son justamente una de las que se busca detectar de manera automática con la aplicación de la minería de procesos al proceso de desarrollo de software, pues entregan conocimiento relevante sobre el funcionamiento de los procesos de desarrollo en las organizaciones.

4.4. Evaluación de conformidad de CMMI

El último caso de estudio que se abordará corresponde a una aplicación para la evaluación del proceso. En particular, correspondería a la aplicación de técnicas de revisión de conformidad para evaluar un proceso según CMMI (Samalikova et al.;

Astromskis et al., 2014). El modelo CMMI describe los elementos típicos que deben poseer los procesos para ser efectivos (Glazer, Dalton, Anderson, Konrad, & Shrum, 2008). Luego, la comparación entre el proceso de desarrollo de una organización dada y las recomendaciones del CMMI ayuda a evaluar la madurez del proceso analizado, encontrar las brechas entre las actividades realizadas y las sugeridas, e identificar las oportunidades de mejora.

De forma muy general, de acuerdo al método de evaluación estándar CMMI para la mejora de procesos (SCAMPI, por sus siglas en inglés), una evaluación de CMMI consiste de dos pasos: recolectar los datos y luego interpretarlos de acuerdo al modelo CMMI (Team. S. U., 2011). La evaluación de procesos puede ser tanto cualitativa como cuantitativa, y puede basarse en el juicio de expertos (Bourque & Fairley, 2014). La minería de procesos puede ayudar en la fase de evaluación de los datos. El uso de herramientas automatizadas de minería de procesos puede ayudar a diseñar evaluaciones objetivas mediante la reducción de la intervención humana (Astromskis, Janes, Sillitti & Succi, 2014; Van der Aalst, 2016).

La idea de aplicar técnicas de este campo para la evaluación de ciertos estándares de desarrollo no es nueva. De hecho, se han realizado trabajos anteriormente con el propósito de evaluar la conformidad con respecto a otros estándares. Por ejemplo, en el año 2006 se evaluó una compañía de desarrollo de software australiana tomando como referencia el estándar ISO/IEC 12207 (ISO, 2008). La aplicación de técnicas de minería de procesos como una herramienta de soporte para la evaluación de procesos de software bajo el modelo CMMI, tanto en principio como en la práctica, se ha demostrado factible (Samalikova, Kusters, Trienekens, & Weijters, 2014). En dicho trabajo se han identificado los elementos que pueden ser utilizados para realizar estas tareas de minería de procesos, las perspectivas de minería de procesos relevantes y las técnicas de minería de procesos asociadas. En función de éstas, la minería de procesos puede ofrecer información valiosa, esto es, evidencia objetiva, para la evaluación y mejora de procesos.

Otro de los aspectos en los que la minería de procesos puede aportar, corresponde a la reducción de costos del proceso de evaluación CMMI. Una de las limitantes que tienen los entornos pequeños en la aplicación de evaluaciones de procesos CMMI es el alto costo de éstas (Astromskis et al., 2014). De acuerdo a los Glazer et al. (2008), una evaluación CMMI para un negocio pequeño con menos de 100 desarrolladores que está completamente lista para ser evaluada puede costar (en dólares) entre \$36.000 y \$60.000, el precio exacto depende de una variedad de factores. Si el equipo necesita preparar todo para la evaluación, los clientes gastan típicamente alrededor de \$125.000 en un horizonte temporal de un año o menos. La aplicación de técnicas de minería de procesos podría reducir los costos tanto en entornos pequeños como en grandes organizaciones (Astromskis et al., 2014).

En el mismo trabajo se sugiere que los practicantes de las metodologías ágiles evalúen sus procesos con respecto a una perspectiva CMMI. Específicamente, proponen una herramienta basada en minería de procesos (Jidoka4CMMI), que una vez configurada, evalúa de manera constante la conformidad de los procesos de desarrollo con CMMI, sin generar costos adicionales, de tal forma que pueda ser utilizada fácilmente por entornos

más pequeños y de menores recursos. Esta herramienta provee un framework que define casos de pruebas y verifica si un proceso se adhiere a las recomendaciones de CMMI (Astromskis et al., 2014).

Además, en otro trabajo de minería de procesos para CMMI se aplicó ésta para evaluar el proceso real seguido por los desarrolladores de una organización. Luego, las discrepancias encontradas entre el proceso real y el modelo descrito fueron evaluadas. Los resultados de esta evaluación se tradujeron en un plan de trabajo y mejora del proceso (Huo, Zhang, & Jeffery, 2006).

Finalmente, se debe destacar que recientemente se ha desarrollado una extensión de SCAMPI. Esta extensión aborda alguna de las limitaciones del método tradicional, tales como: la dependencia del evaluador, el costo y complejidad de la evaluación, la subjetividad en la evaluación, y la confianza en la muestra y su representatividad (do Valle, Santos & de FR Loures, 2017).

5. Desafíos

A continuación, se presenta una lista no exhaustiva de los desafíos importantes a los que se enfrenta la minería de procesos en el campo de la ingeniería de software. Algunos de estos desafíos son generales a la minería de procesos y algunos son particulares al campo de desarrollo de software.

5.1. Conocimiento de los procesos

En muchos casos los sistemas de información de las empresas son agnósticos a los procesos de negocio (principalmente debido al hecho que en muchos casos son soluciones ad hoc que deben permitir un alto grado de flexibilidad). En muchos casos las acciones ejecutadas por el sistema se guardan en algún archivo de registro, pero aun así en los sistemas no identifican el proceso o la instancia del proceso a la cual corresponde (Burattin, 2015; Van der Aalst, 2016).

El problema de aplicar minería de procesos a registros generados por sistemas agnósticos al proceso es uno de los principales desafíos a los que se enfrenta la minería de procesos a la hora de aplicarse en el mundo empresarial. Algunos de los problemas que ocurren durante la adquisición de datos pueden ser solucionados mediante la definición y uso de estándares, tales como MXML (Van der Aalst et al., 2007) u OpenXES (Günther & Verbeek, 2009), ambos basados en XML (Burattin, 2015). Una compañía que tenga una infraestructura adecuada y conozca además sus procesos, puede beneficiarse de la minería de procesos de diversas maneras, aplicando cualquiera de las técnicas existentes (Burattin, 2015).

En el caso de los sistemas agnósticos al proceso, puede que sus registros no tengan suficiente información para aplicar algoritmos de minería de procesos. En algunos casos puede ocurrir lo contrario, en que la cantidad de detalles es demasiado alta y la complejidad del modelo resultante lo hace poco práctico, pues todo buen modelo debe mantener un balance entre precisión con respecto a la representación de la realidad y simplicidad computacional (Van der Aalst et al., 2011; Burattin, 2015).

5.2. Calidad de los datos

La minería de procesos depende fuertemente de la calidad de los datos recolectados y almacenados. Muchas veces se deben filtrar grandes volúmenes de datos debido a instancias incompletas de los procesos. No obstante, mediante preparación cuidadosa de los datos se puede llegar a conjuntos de datos útiles, a pesar de que originalmente no fueran datos almacenados con el propósito de aplicar minería de procesos. Aun así, la eficiencia y efectividad de la minería de procesos puede verse beneficiada sustancialmente de conjuntos de datos bien estructurados y definidos, además de guías de recolección que permitan obtener conjuntos de datos de alta calidad (Samalikova et al., 2011; Bose et al., 2013).

Otro de los problemas en los conjuntos de datos es la existencia de ruido. Este puede ocurrir algunas veces debido a la existencia de procesos que no son los esperados (actividades que no se supone ocurran, o que ocurren en posiciones incorrectas) (Burattin, 2015). La integridad y calidad de los datos es un factor importante en la calidad de los resultados de la evaluación de procesos (Bourque & Fairley, 2014), esto se vuelve aún más importante en conjunción con las necesidades naturales de las técnicas de minería de procesos.

5.3. Integración con las herramientas

Existen algunas consideraciones propias de este campo de aplicación de la minería de procesos debido a la naturaleza de los procesos de software. Particularmente, todas las metodologías de mejora de procesos de software comparten un prerrequisito común: el proceso y el producto deben ser continuamente medidos. Para realizar esto, los datos acerca del proceso de desarrollo deben ser recolectados y analizados de manera continua. No obstante, la recolección de datos manual requiere de un esfuerzo adicional por parte de los desarrolladores, y el análisis de todos estos datos toma aún más tiempo (Astromskis et al., 2014).

Un requerimiento clave para el seguimiento de los procesos de software es que estos no debieran interferir con el proceso que está siendo evaluado. Esto rara vez es el caso para los enfoques tradicionales (e.g., basados en cuestionarios), pues requieren de la participación activa de los distintos actores del proceso para que reporten sus actividades. Un enfoque alternativo es la recolección automática de datos dentro de los ambientes de trabajo (e.g., IDEs y herramientas de diseño). Estas herramientas integradas facilitarían la obtención de datos sin interferir con el trabajo de los desarrolladores de software (Astromskis et al., 2014).

5.4. Evaluación de resultados

Otro problema corresponde a la evaluación del proceso obtenido, pues es necesario definir un procedimiento estándar y riguroso para evaluar la calidad de la salida generada por el algoritmo de minería de procesos (Van der Aalst, 2016). Es posible medir la calidad del modelo mediante las discrepancias con los registros de eventos originales y el modelo obtenido. Otra problemática es escoger una representación adecuada para el modelo, pues finalmente el modelo no será útil si es que no es posible entenderlo y analizarlo adecuadamente (Burattin, 2015).

Otro de los desafíos de implementación de técnica de minería de procesos, es que los algoritmos deben ser parametrizados adecuadamente. La búsqueda de la parametrización óptima puede ser un trabajo complejo debido a la cantidad de parámetros disponibles (Burattin, 2015; Van der Aalst, 2016).

5.5. Usabilidad

Finalmente, el último de los desafíos abordados corresponde a la usabilidad. Según el Manifiesto de Minería de Procesos (Van der Aalst et al., 2011), uno de los desafíos que debe ser abordado para mejorar la usabilidad de la minería de procesos es su integración con otras metodologías y técnicas de análisis. Un ejemplo claro es provisto por las herramientas de simulación, que pueden complementar el diagnóstico de procesos y el análisis mediante la evaluación de implementaciones alternativas (Aguirre, Parra, & Alvarado, 2012).

En general, el proyecto de rediseño de un proceso requiere más de un único instrumento para alcanzar los resultados deseados. Si bien la minería de procesos provee herramientas para el diagnóstico y análisis de procesos. Éstas deben ser complementadas mediante otras metodologías, tales como la simulación u otros enfoques que permitan entender y planificar el proceso de rediseño adecuadamente (Aguirre et al., 2012; Van der Aalst et al., 2011; Van der Aalst, 2016).

6. Conclusiones

En el transcurso de esta investigación se han abordado los principales aspectos necesarios para comprender la importancia de la minería de procesos y su aplicación en el campo de la ingeniería de software, en particular al área de procesos de desarrollo de software. Se identificaron además los distintos desafíos y problemas a los que se enfrenta la minería de procesos a la hora de aplicarse en un ámbito real, tanto en general como en el caso de los procesos de software. Por otra parte, se ha visto que la minería de procesos puede ser aplicada sobre el proceso de desarrollo y sobre el producto desarrollado. Dada la generalidad de los algoritmos de minería de procesos, es posible aplicarlos sobre cualquier entidad que siga una secuencia de pasos lógicos y de la cual se tengan los datos en un formato adecuado.

En función de la revisión anterior se puede ver que la minería de procesos es un área con un alto potencial de aplicaciones en el campo de la ingeniería de software, específicamente al estudio de los procesos de desarrollo de software. Además, al ser la minería de procesos un campo relativamente joven en comparación a las disciplinas que lo componen, su potencial de crecimiento, tanto a nivel teórico como práctico es bastante alto.

La capacidad de descubrir los procesos reales, de evaluar su conformidad con respecto al modelo oficial y de encontrar oportunidades de mejora, hacen que la minería de procesos tenga un alto impacto a nivel organizacional. El desarrollo de sistemas de información que incluyan en su diseño la perspectiva de la minería de procesos entrega un valor agregado a las organizaciones de desarrollo de software. Es posible usar la minería de procesos para detectar las discrepancias en los procesos, obteniendo de esta tarea una

serie de información que puede ser utilizada por los gestores del proceso para planificar las eventuales correcciones y mejoras. El ejemplo que es conveniente destacar es el que muestra que alrededor de 70% de los cambios solicitados al comité de control de cambios no eran sometidos a una fase formal de análisis. La detección de problemas estructurales en los procesos de desarrollo permite a los encargados de proyectos y a la gerencia superior aplicar medidas correctivas según sea necesario.

Se han detallado también los desafíos que debe abordar el campo de la minería de procesos. Uno de los principales desafíos es la necesidad de disponer de datos adecuados, provistos por sistemas de información conscientes de los procesos de negocio. Este es un tema transversal que no sólo se da con los procesos de software.

Como se vio en uno de los casos de estudio, la minería de procesos puede ser utilizada en línea. Esta idea puede generalizarse para evaluar de manera continua el rendimiento de los procesos. Esto permitiría a los gestores realizar de mejor manera su labor, al ser notificados oportunamente de las posibles desviaciones del proceso y de los problemas que vayan surgiendo en el proceso en tiempo real.

Si bien existen metodologías para aplicar técnicas de minería de procesos en general, el desarrollo de una metodología estándar para la implementación de técnicas de minería de procesos en las organizaciones se considera un problema abierto todavía, aunque ya existen avances que abordan esta problemática. Específicamente, se considera la necesidad de una metodología que explicita el desarrollo de la actividad de incorporación de conocimiento del proceso a sistemas de información que son naturalmente agnósticos a éste. Una vez realizado esto, las tareas en realidad seguirían, en cierta forma, el enfoque tradicional de minería de datos, para el cual existen varias metodologías (e.g., CRISP-DM).

Finalmente, se afirma que en función de todo lo anterior, se ha mostrado la importancia de la minería de procesos y de los enfoques automáticos de evaluación de procesos, tanto a nivel general, como en los procesos de desarrollo de software.

Referencias

- Aguirre, S., Parra, C., & Alvarado, J. (2012, June). Combination of process mining and simulation techniques for business process redesign: a methodological approach. In *International Symposium on Data-Driven Process Discovery and Analysis* (pp. 24–43). Springer Berlin Heidelberg.
- Astromskis, S., Janes, A., Sillitti, A., & Succi, G. (2014). Continuous CMMI assessment using non-invasive measurement and process mining. *International Journal of Software Engineering and Knowledge Engineering*, 24(09), 1255–1272.
- Bose, R. J. C., Mans, R. S., & van der Aalst, W. M. (2013, April). Wanna improve process mining results?. In *Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on* (pp. 127–134). IEEE.
- Bourque, P., & Fairley, R. E. (2014). *Guide to the software engineering body of knowledge (SWEBOK (R)): Version 3.0*. IEEE Computer Society Press.

- Burattin, A. (2015). Obstacles to applying process mining in practice. In *Process Mining Techniques in Business Environments* (pp. 59-63). Springer International Publishing.
- Cabac, L., & Denz, N. (2008). Net components for the integration of process mining into agent-oriented software engineering. In *Transactions on Petri Nets and Other Models of Concurrency I* (pp. 86-103). Springer Berlin Heidelberg.
- Cook, J. E., & Wolf, A. L. (1998). Discovering models of software processes from event-based data. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 7(3), 215-249.
- do Valle, A. M., Santos, E. A., & de FR Loures, E. (2017). Applying Process Mining Techniques in Software Process Appraisals. *Information and Software Technology*.
- Glazer, H., Dalton, J., Anderson, D., Konrad, M. D., & Shrum, S. (2008). CMMI or agile: why not embrace both!.
- Gottschalk, F., van der Aalst, W., & Jansen-Vullers, M. (2008). Mining reference process models and their configurations. In *On the Move to Meaningful Internet Systems: OTM 2008 Workshops* (pp. 263-272). Springer Berlin/Heidelberg.
- Günther, C., & Verbeek, H. XES Standard Definition. www.xes-standard.org (2009). Cited on, 72.
- Halkidi, M., Spinellis, D., Tsatsaronis, G., & Vazirgiannis, M. (2011). Data mining in software engineering. *Intelligent Data Analysis*, 15(3), 413-441.
- Humphrey, W. S. (1995). *A discipline for software engineering*. Addison-Wesley Longman Publishing Co., Inc.
- Huo, M., Zhang, H., & Jeffery, R. (2006, May). An exploratory study of process enactment as input to software process improvement. In *Proceedings of the 2006 international Workshop on Software Quality* (pp. 39-44). ACM.
- ISO, I. (2008). *IEC 12207 Systems and software engineering-software life cycle processes*. International Organization for Standardization, Geneva, Switzerland.
- Jans, M., Alles, M., & Vasarhelyi, M. (2013). The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems*, 14(1), 1-20.
- Kalenkova, A. A., van der Aalst, W. M., Lomazova, I. A., & Rubin, V. A. (2016, October). Process mining using BPMN: relating event logs and process models. In *Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems* (pp. 123-123). ACM.
- Kimble, C., de Vasconcelos, J. B., & Rocha, Á. (2016). Competence management in knowledge intensive organizations using consensual knowledge and ontologies. *Information Systems Frontiers*, 18(6), 1119-1130.
- Kruchten, P. (2004). *The rational unified process: an introduction*. Addison-Wesley Professional.

- Mazak, A., & Wimmer, M. (2016). On Marrying Model-driven Engineering and Process Mining: A Case Study in Execution-based Model Profiling. In Proceedings of the 6th IFIP International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA 2016) (pp. 1–10).
- Mejia, J., & Uribe, G. (2015). Extracción del Conocimiento tácito como base para el establecimiento de mejora de procesos de software en las Organizaciones de desarrollo de Software. RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação, (SPE3), 17–28.
- Muñoz, M., Gasca, G., & Valtierra, C. (2014). Caracterizando las necesidades de las pymes para implementar mejoras de procesos software: Una comparativa entre la teoría y la realidad. RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação, (SPE1), 1–15.
- Poncin, W., Serebrenik, A., & Van Den Brand, M. (2011, March). Process mining software repositories. In Software Maintenance and Reengineering (CSMR), 2011 15th European Conference on (pp. 5–14). IEEE.
- Rubin, V. A., Günther, C. W., Van Der Aalst, W. M., Kindler, E., Van Dongen, B. F., & Schäfer, W. (2007, May). Process mining framework for software processes. In International Conference on Software Process (pp. 169-181). Springer Berlin Heidelberg.
- Rubin, V. A., Mitsyuk, A. A., Lomazova, I. A., & van der Aalst, W. M. (2014, September). Process mining can be applied to software too!. In Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (p. 57). ACM.
- Osterweil, L. (1987, March). Software processes are software too. In Proceedings of the 9th international conference on Software Engineering (pp. 2–13). IEEE Computer Society Press.
- Santos, R., & Oliveira, T. C. (2015, April). Mining software development process variations. In Proceedings of the 30th Annual ACM Symposium on Applied Computing (pp. 1657-1660). ACM.
- Samalikova, J., Kusters, R. J., Trienekens, J. J., & Weijters, A. J. M. M. (2014). Process mining support for Capability Maturity Model Integration-based software process assessment, in principle and in practice. *Journal of Software: Evolution and Process*, 26(7), 714–728.
- Samalikova, J., Trienekens, J. J., Kusters, R. J., & Weijters, A. T. (2009, September). Discovering changes of the change control board process during a software development project using process mining. In European Conference on Software Process Improvement (pp. 128–136). Springer Berlin Heidelberg.
- Samalikova, J., Kusters, R. J., Trienekens, J. J., Weijters, T., & Siemons, P. (2011). Toward objective software process information: experiences from a case study. *Software Quality Journal*, 19(1), 101–120.

- Shah, V., Khadke, C., & Rana, S. (2015, April). Mining process models and architectural components from test cases. In *Software Testing, Verification and Validation Workshops (ICSTW), 2015 IEEE Eighth International Conference on* (pp. 1–6). IEEE.
- Stolfa, J., Stolfa, S., Kosinar, M. A., & Snasel, V. (2016). Introduction to integration of the process mining to the knowledge framework for software processes. In *Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015* (pp. 21–31). Springer International Publishing.
- Sureka, A., Kumar, A., & Gupta, S. (2015, February). Ahaan: Software process intelligence: Mining software process data for extracting actionable information. In *Proceedings of the 8th India Software Engineering Conference* (pp. 198–199). ACM.
- Team, S. U. (2011). *Standard CMMI Appraisal Method for Process Improvement (SCAMPI) A, Version 1.3: Method Definition Document*.
- Trcka, N., & Pechenizkiy, M. (2009, November). From local patterns to global models: Towards domain driven educational process mining. In *Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on* (pp. 1114–1119). IEEE.
- Van Der Aalst, W. M., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Burattin, A. (2011, August). Process mining manifesto. In *International Conference on Business Process Management* (pp. 169–194). Springer Berlin Heidelberg.
- Van der Aalst, W. M., Pesic, M., & Song, M. (2010, June). Beyond process mining: from the past to present and future. In *International Conference on Advanced Information Systems Engineering* (pp. 38–52). Springer Berlin Heidelberg.
- Van Der Aalst, W. M. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2), 7.
- Van der Aalst, W. M. (2005). Business alignment: using process mining as a tool for delta analysis and conformance testing. *Requirements Engineering*, 10(3), 198–211.
- Van der Aalst, W. M. (2016). *Process mining: data science in action*. Springer.
- Van der Aalst, W. M. (2015, August). Big software on the run: in vivo software analytics based on process mining (keynote). In *Proceedings of the 2015 International Conference on Software and System Process* (pp. 1–5). ACM.
- Van der Aalst, W. M. (2014). Process mining in the large: a tutorial. In *Business Intelligence* (pp. 33–76). Springer International Publishing.
- Van der Aalst, W. M., van Dongen, B. F., Günther, C. W., Mans, R. S., De Medeiros, A. A., Rozinat, A., ... & Weijters, A. J. M. M. (2007, June). ProM 4.0: comprehensive support for real process analysis. In *International Conference on Application and Theory of Petri Nets* (pp. 484–494). Springer Berlin Heidelberg.
- Van der Aalst, W. M., van Hee, K. M., van Werf, J. M., & Verdonk, M. (2010). Auditing 2.0: Using process mining to support tomorrow's auditor. *Computer*, 43(3).
- Zinski, C. (2001). *Internet Banking: The Trade Name Trend*. Banking & Financial Services Policy.