

Tipo de artículo: Artículo de revisión
Temática: Ingeniería y Gestión de software
Recibido: 29/11/2013 | Aceptado: 21/03/2014

Algoritmos y técnicas de descubrimiento de procesos poco estructurados: estado del arte

Algorithms and discovering techniques for unstructured processes: state of art

Dina Yaksilik Torres Sakipova

Facultad 3. Departamento de Programación. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370. Correo-e: dytorres@uci.cu

Resumen

La ejecución de los procesos de negocio genera un gran volumen de datos, de los que se puede extraer información que ayude a comprender fenómenos o a tomar decisiones aplicando técnicas y herramientas. La minería de procesos se encarga de descubrir, monitorear y mejorar los procesos reales que se llevan a cabo en una empresa, extrayendo conocimiento de los registros de eventos que manejan los sistemas informáticos actuales y generando modelos que describen los procesos. Cuando los procesos son poco estructurados se generan modelos poco comprensibles que dificultan su análisis. El objetivo de este trabajo es realizar un estudio de algunos algoritmos y técnicas de descubrimiento de mayor relevancia durante los últimos 10 años, enfocados en la etapa de descubrimiento de la minería de procesos, para identificar las principales dificultades existentes en esta área. La investigación mostró que existen varios aspectos relacionados con la interacción de los usuarios con los modelos durante su generación y análisis, y con la forma en que estos se presentan, que perjudican su comprensión. La identificación de estas dificultades permite direccionar las investigaciones hacia una mayor integración del usuario en la generación de los modelos, así como enfatizar la importancia de estudiar el área de las visualizaciones gráficas como variante para lograr mayor comprensibilidad en los mismos. Estos modelos son de gran ayuda para el análisis y la gestión de procesos, la toma de decisiones e incluso, las auditorías.

Palabras clave: minería de procesos, procesos poco estructurados, técnicas de descubrimiento, visualizaciones gráficas.

Abstract

The execution of the business processes generates a great volume of data, from where information can be extracted helping to understand phenomena or to make decisions applying techniques and tools. The process mining is in charge of discovering, monitoring and improving the real processes carried out in a company, extracting knowledge from the logs managed by the current computer systems and generating models that describe the processes. When the processes are unstructured, not very comprehensible models that hinder the analysis of the process, are generated. The goal of this research is to perform a study of some of the more relevant algorithms and techniques during the last 10 years, focused in the discovery stage of process mining to identify the main existing difficulties in this area. The research showed that many aspects related to the interaction of the users with the models during its generation and analysis, and the way they are exposed, disturb their understanding. Identifying these difficulties addresses the investigation toward the increase of user interaction during the models generation, and emphasize the importance of studying the graphic visualizations area as a variant to improve their understanding. These models provide a great help for the analysis and process management, decision-making and even audits.

Keywords: *discovering techniques, graphic visualizations, process mining, unstructured processes.*

Introducción

Los modelos de procesos de negocio constituyen “una representación abstracta (gráfica) de los procesos de una organización, que muestran principalmente cómo y por quién son llevadas a cabo las actividades que generan valor para la organización” (Ruiz, 2012).

Estos modelos se pueden obtener a través de la Minería de Procesos (MP), área joven que se encarga de descubrir, monitorear y mejorar los procesos reales que se llevan a cabo en una empresa, extrayendo conocimiento de los registros de eventos (en lo adelante RE) que manejan los sistemas informáticos actuales (van der Aalst, *et al.*, 2012). Este concepto se definió en el Manifiesto de Minería de Procesos, lanzado por el grupo de expertos de todo el mundo que investiga la MP, conocido como IEEE Task Force on Process Mining, en el 2012 y soportado por 53 organizaciones y bajo la contribución de 77 expertos en el tema (van der Aalst, 2013).

La MP, a diferencia de la modelación tradicional de procesos que emplea herramientas y donde interviene el factor humano y la concepción de lo que se debe hacer, muestra lo que realmente sucede. Esto permite realizar un estudio más analítico de los procesos y del funcionamiento de una empresa, basado en hechos que ya ocurrieron y que muestran la realidad de la ejecución de los procesos (van der Aalst, 2009) que se encuentran guardados como RE.

Cuando estos RE muestran que los procesos se caracterizan por una manera de ejecutarse bastante controlada y que poseen una clara estructura, dichos procesos son estructurados y son conocidos por la comunidad experta en el tema como “procesos lasaña” según (van der Aalst, 2011). Son frecuentemente menos interesantes debido a que es fácil aplicarle técnicas de MP y, debido a que están organizados, las mejoras son pequeñas.

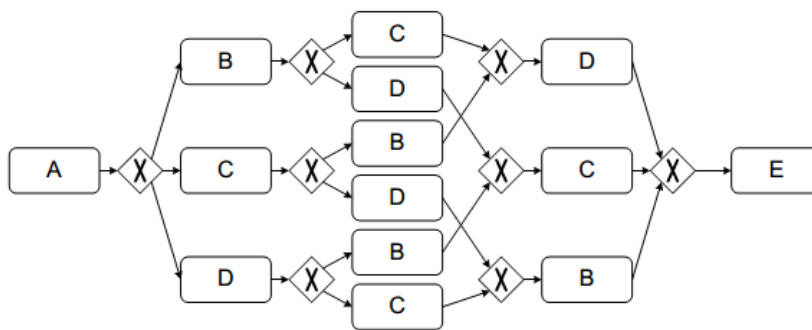


Figura 1. Modelo de un proceso estructurado (extraído de (van der Aalst, *et al.*, 2012)).

Por el contrario, cuando los RE contienen actividades ejecutadas con diversos órdenes, donde varía mucho el momento en el que se ejecutaron, las actividades precedentes y posteriores cambian, cambian los autores, o se evidencian muchas relaciones o variaciones entre las actividades, describen procesos poco estructurados conocidos como “espaguetti” según (van der Aalst, 2011) que son muy difíciles de comprender.

El objetivo de este trabajo es realizar un estudio de algunos algoritmos y técnicas de descubrimiento de mayor relevancia durante los últimos 10 años, enfocados en la etapa de descubrimiento de la MP, para identificar las principales dificultades existentes en esta área.

El uso de la MP permite no solo descubrir los procesos, sino también realizar un análisis de conformidad (investigando si la realidad se ajusta al modelo dado y viceversa) y una extensión (aumentando un modelo existente con información adicional extraída desde los RE).

Es importante en una organización identificar y extraer rápidamente la información importante para elevar los niveles de competencia del negocio. La MP ayuda a descubrir conocimiento, por lo que constituye un soporte para la toma de decisiones.

Además, los modelos de procesos basados en datos históricos pueden ayudar a realizar predicciones, por ejemplo, posibilitan estimar el tiempo restante de un proceso y la probabilidad de un resultado determinado. Igualmente, esta información puede emplearse para hacer recomendaciones como, por ejemplo, proponer una actividad que minimizara el tiempo y costo esperados (van der Aalst1, *et al.*, 2010).

De ahí la importancia de encontrar un tema de investigación que permita definir una técnica de descubrimiento que eleve el nivel de comprensibilidad de los modelos de procesos poco estructurados, de manera que se logre obtener modelos que apoyen a los analistas, auditores y directivos de las empresas en sus actividades de chequeo, análisis, mejora y toma de decisiones de los procesos de su entidad.

Materiales y métodos

Los RE permiten realizar tres tipos de MP. El primer tipo es el descubrimiento, que toma un RE y produce un modelo. Es frecuentemente usado como punto de partida para otros tipos de análisis (van der Aalst, *et al.*, 2012). El segundo tipo es la verificación de conformidad, que compara un modelo de proceso existente con un RE del mismo proceso, permitiendo chequear si la realidad, tal como está almacenada en el RE, es equivalente al modelo y viceversa. El tercer tipo es el mejoramiento, que consiste en extender o mejorar un modelo de proceso existente (Yzquierdo, 2012).

El objetivo fundamental del descubrimiento es revelar automáticamente modelos de procesos que describan procesos, con precisión, considerando solamente RE de sus procesos operacionales (Buijs, *et al.*, 2013). En este trabajo se estudia este proceso debido a que se realiza un estudio de técnicas que permiten obtener estos modelos. Por tanto no es relevante en esta investigación abordar el chequeo de conformidad y el mejoramiento.

Se realiza un análisis de la teoría sobre el descubrimiento en la MP de la última década y las principales técnicas de los autores fundamentales durante los últimos 5 años, con vistas a exponer los problemas más actuales de los modelos de procesos poco estructurados descubiertos con la MP.

Se exponen solamente algunos de los algoritmos y técnicas estudiados a pesar de haber realizado un estudio bibliográfico más extenso, debido a que no es objetivo repetir en qué consisten, sino realizar un análisis de las dificultades detectadas en todos ellos, que permita definir una línea de investigación a seguir.

Desarrollo

Durante la última década se han desarrollado varios algoritmos y técnicas de descubrimiento en la MP. A continuación se presentan algunos de los estudiados.

Algoritmos de descubrimiento

α

El algoritmo α , creado en 2004 por van der Aalst y otros, es uno de los primeros algoritmos de descubrimiento de procesos. Se basa en la captura, desde un RE, de las relaciones siguientes usando el sistema binario. Asume que el RE está completo, por lo que presenta problemas para lidiar con el ruido y el comportamiento infrecuente o incompleto (De Weerdt, *et al.*, 2012). No tiene en cuenta las frecuencias de las actividades. Además, manifiesta dificultades con el tratamiento de ciclos cortos, dependencias no locales, entre otros, generando modelos que no se ajustan adecuadamente a los eventos (van der Aalst, 2011).

A partir de este algoritmo surgieron el α^+ y el α^{++} , que incorporaban el tratamiento a ciclos cortos y constructores sin libre elección respectivamente. Posteriormente, en el 2009, surge el algoritmo β , que puede distinguir concurrencia y ciclos cortos.

Minería heurística

Este algoritmo, creado en 2006 por Weijters y otros, tiene en cuenta las frecuencias de los eventos. Este dato es usado para derivar las relaciones causales y de paralelismo. No obstante, no posee gran funcionalidad con respecto a las tareas duplicadas y se abstrae del comportamiento poco frecuente, es decir, no incorpora los eventos poco frecuentes al modelo, provocando que no se representen caminos que se ejecutan con poca frecuencia pero pudieran ser importantes (Günther, 2009) (Weijters, *et al.*, 2010) (van der Aalst, 2011).

Minería genética

Este algoritmo, presentado en (de Medeiros, 2006), simula la evolución en la naturaleza. Parte de una población inicial de individuos (en este caso procesos que son candidatos a la solución) a la que se le realizan varias iteraciones del algoritmo genético acercándose cada vez más a la solución apropiada.

Minería difusa

Este algoritmo (W. Günther, *et al.*, 2007), creado por Günther y van der Aalst, está basado en medidas de significancia y correlación para visualizar el comportamiento en los RE. Aunque este algoritmo es bastante aceptado como técnica de exploración de datos, posee una desventaja en el sentido de que un modelo difuso no puede transformarse a una Red de Petri, lo que limita una evaluación comparativa con otras técnicas de descubrimiento de procesos (De Weerd, *et al.*, 2012).

Todos estos algoritmos y muchos otros presentan dificultades en el tratamiento de diferentes aspectos que se muestran en la Tabla 1 de (De Weerd, *et al.*, 2012). Por ende fue necesario comenzar a crear técnicas que usaran estos algoritmos pero que intentaran atenuar esas dificultades.

Técnicas de descubrimiento

Los autores W. Günther y otros (2007) presentan una solución para la mejora de la visualización de los modelos de procesos poco estructurados, que permite que el analista de procesos pueda determinar con más claridad cuáles son sus procesos más importantes debido a que estos se muestran oscurecidos. Sin embargo, esto no resuelve del todo el problema, pues en un modelo como el que se muestra a continuación, tomado de (van der Aalst, 2011), aunque se destaquen relaciones oscuras, sigue siendo complicado descifrarlo debido a la gran cantidad de arcos que posee.

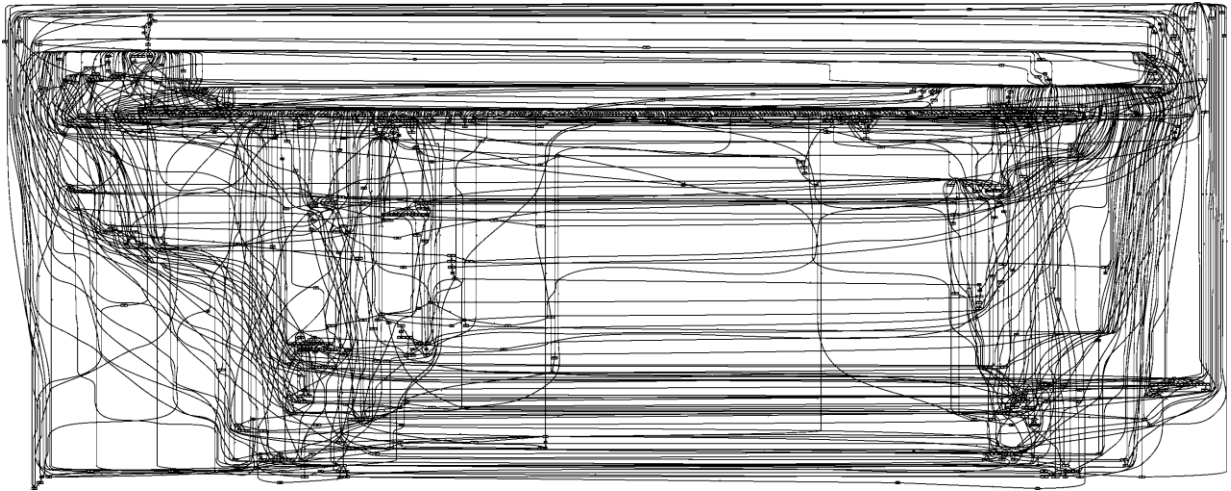


Figura 3. Técnica desarrollada por (W. Günther, *et al.*, 2007) que oscurece las relaciones más importantes (extraída de (van der Aalst, 2011)).

Además, en esa propuesta se expone cómo se usan varios conceptos cartográficos para mejorar la visualización. Ellos son Agregación (se fusionan algunas partes de la ciudad), Abstracción (los caminos insignificantes no se muestran), Énfasis (se destacan los caminos importantes con tamaño y color) y Personalización (enfoca detalles intencionados). Configurar este algoritmo denominado Fuzzy Mining y sus métricas ocupa tiempo y conocimiento profundo sobre el funcionamiento de la técnica (W. Günther, *et al.*, 2007).

Song y van der Aalst (2007) propusieron el análisis de mapa de puntos (análogo a los mapas de Gantt) para analizar el desarrollo del proceso dibujando eventos del proceso de manera gráfica. Esta propuesta presenta una vista “helicóptero” del RE e incluye algunas métricas como el mínimo, máximo e intervalo medio entre los eventos. Sin embargo, el analista del negocio necesita buscar manualmente en el mapa de puntos para identificar cualquier información potencial del proceso. Es por ello que el análisis se torna engorroso cuando los registros superan los cientos de actividades y muchas veces no es posible identificar patrones interesantes.

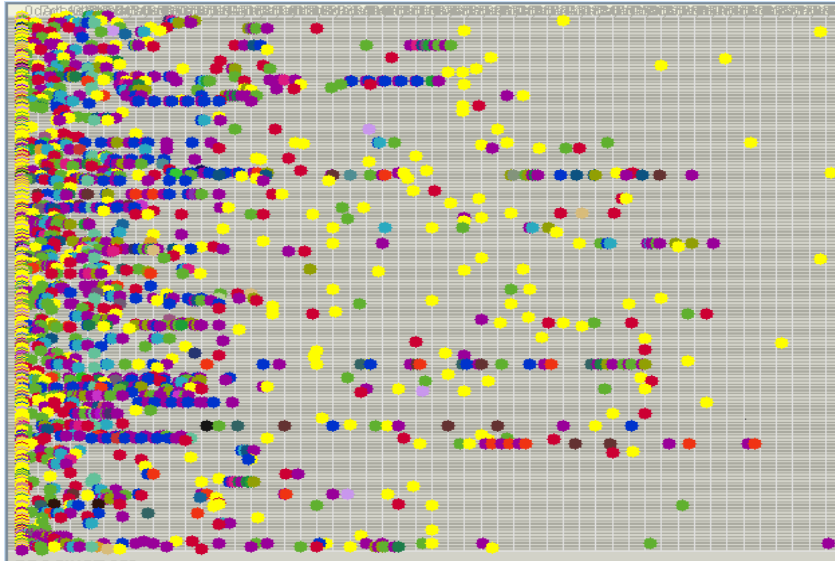


Figura 4. Modelo de procesos usando la técnica de Mapa de Puntos (extraído de (Song, *et al.*, 2007)).

El trabajo de W. Günther (2009) resalta el flujo principal del proceso. A pesar de ello, si el proceso es muy poco estructurado, el modelo resultante es tan inútil como el típico espagueti. Además, el resultado no permite derivar información útil del proceso debido a que solo muestra nodos y relaciones.

En la literatura se observa el término precisión, representado por dos métricas para evaluar la calidad de los modelos de procesos resultantes. La primera se denomina aptitud (*fitness* en inglés). Esta describe qué tan bien puede ser explicado el comportamiento observado por el modelo (Günther, 2009), (Rozinat, *et al.*, 2008). Cuando varios eventos extraídos de los RE no se permiten en el modelo, la aptitud decrece. La segunda métrica es denominada igualmente precisión (o comportamiento adecuado), que penaliza la generalización excesiva en un modelo, es decir, si un modelo permite más comportamiento del que debe, la precisión decrece (Günther, 2009), (Rozinat, *et al.*, 2008).

En (Günther, 2009) se propone que se debe sacrificar la precisión para lograr niveles de abstracción y obtener modelos más entendibles. Este autor afirma que reduciendo el alcance del modelo se simplifica la conducta del modelo, contribuyendo significativamente a la reducción de la complejidad del modelo. Esto es válido si solo se quiere ver una sección del modelo o el modelo con cierta abstracción, pero no permite obtener modelos con información intencionada.

Chandra y otros en (2009) proponen un método para pre procesar las trazas antes de obtener el modelo de procesos. Aplicando este método mejora la visualización del modelo, pero todavía se presenta bastante incomprensible, como se puede observar en la siguiente figura.

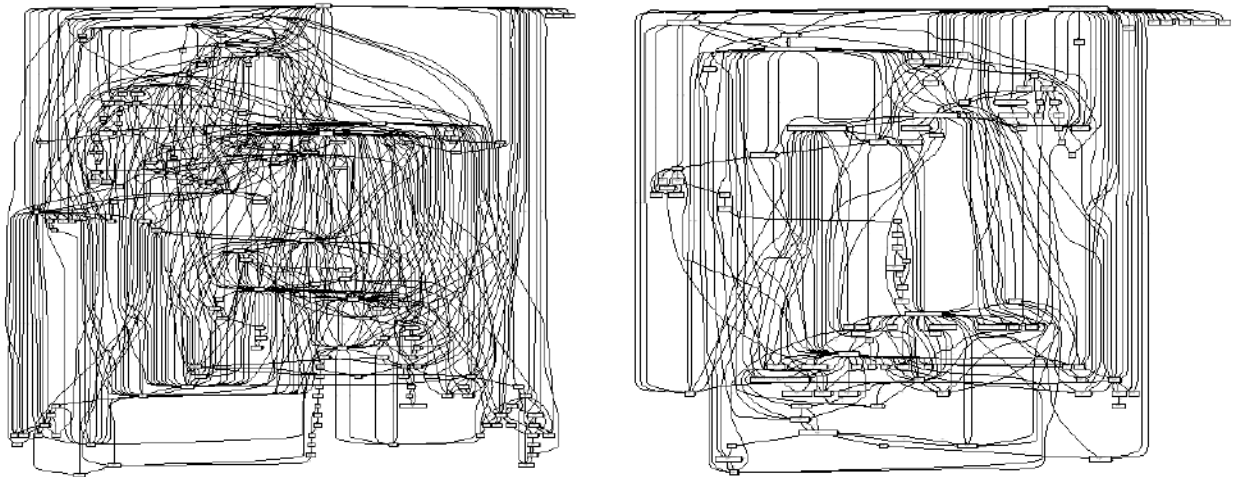


Figura 5. a) Muestra un modelo de procesos minado del registro de eventos original. b) Muestra la mejora del modelo de procesos minado del registro de eventos tras pre procesar las trazas (extraído de (Chandra, *et al.*, 2009)).

Raykenler Yzquierdo (2012) propone la descomposición de procesos en árboles de bloques de construcción. Esta técnica permite agrupar las tareas automáticamente y resalta los aspectos más significativos del proceso en cada momento. El descubrimiento de los subprocessos que componen el proceso analizado, sus dependencias y correlaciones, permiten una mayor precisión en el diagnóstico realizado. A pesar de ello, es difícil aplicarle métricas para comprobar su funcionamiento, debido a que la mayoría de estas funcionan sobre otras notaciones, como Redes de Petri.

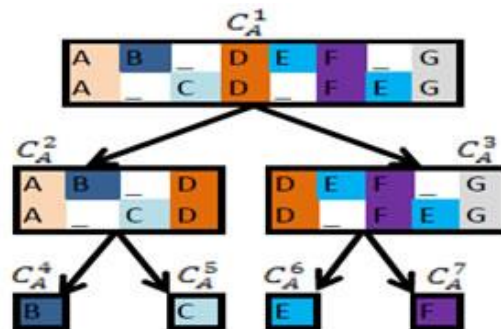


Figura 6. Árbol de descomposición en subprocessos (extraído de (Yzquierdo, *et al.*, 2012)).

Discusión

Dificultades en la minería de procesos

Las técnicas de MP son empleadas para descubrir el modelo de procesos del negocio a partir de los logs. Estas técnicas no manejan apropiadamente algunas situaciones como las que se muestran a continuación y que son tratadas en (Tiwari, *et al.*, 2008):

- Ruido: Los datos registrados pueden ser incorrectos o incompletos, creando problemas cuando los datos son minados. Es decir, existe un comportamiento infrecuente o excepcional (van der Aalst, 2011).
- Tareas escondidas: Tareas que existen pero que no se pueden encontrar en los datos.
- Tareas duplicadas: Dos nodos del proceso pueden referirse al mismo modelo de proceso.
- Constructores sin libre elección: Opciones controladas que dependen de selecciones hechas en otra parte del modelo de proceso.
- Ciclos minados: Un proceso puede ejecutarse varias veces, los ciclos pueden estar simplemente involucrando uno o varios eventos o pueden ser más complejos.
- Distintas perspectivas: Los eventos del proceso pueden añadirse con información adicional con propósitos específicos.
- Análisis Delta: Comparación entre el modelo de procesos y el modelo de referencia para chequear similitudes y disparidades.
- Visualización de resultados: Pueden presentarse de formas gráficas en paneles de gestión.
- Resultados heterogéneos: Acceso a sistemas de información basados en diferentes plataformas.
- Procesos concurrentes: Procesos que ocurren a la misma vez. Determinar tareas concurrentes durante la minería de RE incompletos puede ser un problema desafiante (Zareh, *et al.*, 2012).
- Búsquedas locales y/o globales: Las estrategias locales restringen el espacio de búsqueda y son menos complejas. Las estrategias globales son complicadas pero tienen mejor chance de encontrar una solución óptima.
- Redescubrimiento de proceso: La selección de un algoritmo de minería que pueda redescubrir una clase de modelo de proceso de un RE completo (Tiwari, *et al.*, 2008).
- Inexistencia de ejemplos negativos: Un RE muestra lo que ha pasado, pero no lo que debería haber pasado.

- Comportamiento incompleto: Debido a la concurrencia, los lazos y las selecciones, el espacio de búsqueda posee una estructura compleja y el RE frecuentemente solo contiene una fracción de todos los posibles comportamientos (van der Aalst, *et al.*, 2012a).

Estas situaciones provocan que el resultado sea un modelo inadecuado que no representa la realidad, que resulta difícil interpretarlo y por tanto, no suele ser útil.

Actualmente los modelos de procesos se representan estáticamente, es decir, no brindan prácticamente ninguna posibilidad de que el usuario pueda interactuar y modificar las vistas del modelo. Esto implica que toda la información extraída de los eventos se muestra en un modelo, lo que torna engorrosa la interpretación, principalmente en los modelos espaguetis, donde la interpretación se vuelve prácticamente nula. Además, no permiten percibir información que aparece incluida en el modelo pero no resaltada, de manera que el analista puede desaprovechar algo de información. Asimismo, sería conveniente poder filtrar la información que se muestra en el modelo convenientemente.

Las técnicas de MP que son convenientes para entornos de procesos poco estructurados padecen a menudo incomprendibilidad. Necesitan ser capaces de proporcionar un alto nivel de visibilidad en el proceso, abstrayéndose de detalles indeseados (W. Günther, *et al.*, 2007).

Otro elemento importante es que en los modelos espaguetis no se evidencia claramente el flujo principal de actividades del proceso. Muchas veces los nodos se encuentran conectados en ambas direcciones, lo que le dificulta al analista de procesos definir causa y efecto.

La mayoría de los algoritmos, una vez cargado y analizado el RE, no permiten al usuario modificar parámetros del resultado obtenido para mejorarlo, por lo que el usuario debe comenzar desde el principio definiendo otros parámetros al inicio y esperar el resultado. El resultado de cualquier algoritmo utilizado para minar los procesos generalmente presenta problemas, por lo que se le debería permitir al usuario usar su experiencia y conocimientos para mejorar el modelo resultante realizando ajustes manuales, de manera que el usuario no se quede detenido en un resultado indeseado. Esta propiedad se denomina inmutabilidad y la poseen la mayoría de los algoritmos.

Otra situación que torna indescifrables los modelos, es cuando los RE son muy grandes, millones de eventos, por ejemplo. En este caso, al igual que en las ideas analizadas anteriormente, se evidencia que se le debe brindar más protagonismo al usuario. Este debe poder obtener la perspectiva y el nivel de profundidad que desee.

Tabla 1. Principales dificultades de algunas técnicas de descubrimiento de la MP.

Algoritmos y Técnicas	Dificultades
Minería Heurística	<ul style="list-style-type: none"> • No posee gran funcionalidad con respecto a las tareas duplicadas. • No incorpora los eventos poco frecuentes al modelo, provocando que no se representen caminos que se ejecutan con poca frecuencia pero pudieran ser importantes.
Minería genética	<ul style="list-style-type: none"> • Debe mejorar la precisión para indicar una mejor relación entre el modelo y el registro de eventos.
Minería Difusa	<ul style="list-style-type: none"> • Su resultado no puede transformarse a una Red de Petri, lo que limita una evaluación comparativa con otras técnicas de descubrimiento de procesos.
(W. Günther, <i>et al.</i> , 2007) (Günther, 2009)	<ul style="list-style-type: none"> • Las relaciones oscuras no garantizan descifrar el modelo debido a la gran cantidad de arcos que posee.
(Song, <i>et al.</i> , 2007)	<ul style="list-style-type: none"> • Se necesita buscar manualmente en el mapa de puntos para identificar cualquier información potencial del proceso. • El análisis se torna engorroso cuando los registros superan los cientos de actividades. • Muchas veces es difícil o no es posible identificar patrones interesantes.
(Günther, 2009)	<ul style="list-style-type: none"> • Funciona si solo se quiere ver una sección del modelo o el modelo con cierta abstracción, pero no permite obtener modelos con información intencionada.
(Chandra Bose, <i>et al.</i> , 2009)	<ul style="list-style-type: none"> • El resultado continúa siendo en la mayoría de los casos un modelo incomprensible.
(Yzquierdo, <i>et al.</i> , 2012)	<ul style="list-style-type: none"> • Es difícil aplicarle métricas para comprobar su funcionamiento, debido a que la mayoría de estas funcionan sobre otras notaciones, como Redes de Petri.
Dificultades de la mayoría de las técnicas actuales	
No brindan casi ninguna posibilidad de que el usuario pueda interactuar y modificar las vistas del modelo.	
Toda la información extraída de los eventos se muestra en un modelo, lo que torna engorrosa su interpretación.	
Se dificulta percibir información que aparece incluida en el modelo pero no resaltada.	
No se evidencia claramente el flujo principal de actividades del proceso.	
Muchas veces los nodos se encuentran conectados en ambas direcciones.	

Presentan inmutabilidad, es decir, una vez obtenido un resultado, no permiten modificar parámetros para mejorarlo, por lo que el usuario debe comenzar desde el principio definiendo otros parámetros al inicio y esperar el resultado.

Con RE muy grandes se complejizan los modelos.
--

La Tabla 1 resume las dificultades de algunas de las técnicas abordadas y dificultades generales de la mayoría de las técnicas. De esta manera se observa con más claridad la idea de que resolver esas dificultades, debe mejorar la forma en que se visualizan estos modelos y su comprensión y, de esta forma, su utilidad, ya que las vistas obtenidas a través de la MP son usadas para optimizar los procesos de negocio y mejorar los servicios a clientes (Chandra, *et al.*, 2013).

Conclusiones

El resultado de minar los eventos de procesos poco estructurados es usualmente un modelo poco estructurado y difícil de entender, que resulta inútil para los analistas de procesos. Es necesario mejorar la forma de visualización de estos modelos, de manera que ayuden a identificar y extraer rápidamente información importante para apoyar la auditoría y la toma de decisiones, en aras de elevar los niveles de competencia de un negocio.

En este trabajo se hace un estudio de algunos algoritmos y técnicas que presenta hoy la MP, específicamente el área de descubrimiento, que tributan a un escaso o nulo entendimiento de los modelos de procesos obtenidos, fundamentalmente cuando los procesos son poco estructurados.

Se exponen algunos algoritmos y técnicas encaminados a mejorar el entendimiento de los modelos de procesos poco estructurados, evidenciando la necesidad de continuar las investigaciones en este sentido, debido a que resultan insuficientes.

Es importante explorar otras vías de solución a la dificultad visual que supone emplear estas técnicas y obtener un modelo de procesos espagueti. La representación del modelo en vistas constituye una gran ayuda para aliviar la incompreensión que brindan los modelos espaguetis. Una vista de un modelo de procesos de negocio muestra una representación parcial de acuerdo a alguna propiedad de forma similar a las vistas de tablas en bases de datos (Motahari, *et al.*, 2008) (Pérez-Castillo, *et al.*, 2012). Por ello se identifica el área de la visualización gráfica como perspectiva a investigar para descubrir cómo disminuir la dificultad para entender estos modelos espaguetis. El trabajo con vistas gráficas permite optimizar los procesos de negocio y mejorar los servicios (Chandra, *et al.*, 2013).

Referencias

- BUIJS, J. C. A. M., VAN DONGEN, B. F. Y VAN DER AALST, W.M.P. *Mining Configurable Process Models from Collections of Event Logs*. Eindhoven University of Technology : Springer-Verlag Berlin Heidelberg, 2013. págs. 33-48.
- CHANDRA B., J. Y VAN DER AALST, W. *Abstractions in Process Mining: A Taxonomy of Patterns*. En Business Process Management. Springer Berlin Heidelberg, 2009. p. 159-175. 2009.
- DE MEDEIROS, A.K.A. , WEIJTERS, A. J. M. M.. *Genetic Process Mining*. En Applications and Theory of Petri Nets 2005, volume 3536 of Lecture Notes in Computer Science. 2005.
- DE WEERDT, J, *et al.*. *A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs*. Bélgica : Elsevier. *Information Systems*. 2012. Vol. 37, No 7, págs. 654–676. www.elsevier.com/locate/infosys.
- GÜNTHER, CH. W. *Process Mining in Flexible Environments*. Eindhoven: Technische Universiteit : University Press Facilities, Eindhoven, 2009. ISBN 978-90-386-1964-4.
- JAGADEESH CH., B, R.P., MANS, R S. Y VAN DER AALST, WIL M.P. *Wanna Improve Process Mining Results? It's High Time We Consider Data Quality Issues Seriously*. Conference Paper : 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM'13, Singapore, April 16-19, 2013), (pp. 127-134). Piscataway: IEEE.
- MOTAHARI, H., *et al.*. Process spaceship: discovering and exploring process views from event logs in data spaces. *Proceedings of the VLDB Endowment*, 2008, vol. 1, no 2, p. 1412-1415.
- PÉREZ-CASTILLO, R., *et al.*. Modernización Software mediante Descubrimiento de Vistas de Procesos de Negocio. *XVI Jornadas de Ingeniería del Software y Bases de Datos (JISBD'11) - Track Desarrollo Software Dirigido por Modelos (DSDM)*. 2011. A Coruña, Spain: In Press.
- ROZINAT, A. y VAN DER AALST, W.M.P. Conformance Checking of Processes Based on Monitoring Real Behavior. *Information Systems*, 2008, vol. 33, no 1, p. 64-95.
- RUIZ, F. Proceso Software y Gestión del Conocimiento. *Grupo Alarcos. Dep. de Tecnologías y Sistemas de Información Esc. Sup. de Informática. Universidad de Castilla-La Mancha. Ciudad Real*. 2012. <http://alarcos.inf-cr.uclm.es>.

- SONG, M. Y VAN DER AALST, W.M.P. Supporting process mining by showing events at a glance. *En Proceedings of 17th Annual Workshop on Information Technologies and Systems (WITS 2007), Montreal, Canada.* 2007. p. 139-145.
- TIWARI, A., TURNER, C. J. Y MAJEED, B. . A review of business process mining: state-of-the-art and future trends. *Business Process Management Journal*, 2008, vol. 14, no 1, p. 5-22.
- VAN DER AALST, W.M.P. Y VERBEEK, H.M.W. *Process Discovery and Conformance Checking Using Passages*. BPM Center Report BPM-12-21, BPMcenter. org, 2012a.
- VAN DER AALST, W. *Business Process Management*, hoy. [entrev.] Ricardo Seguel P. 14 de Septiembre de 2009. BPM LATAM, <http://www.bpm-latam.org/entrevistas/entrevista14-04-09>.
- VAN DER AALST, W. M. P. *Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining*. En *Asia Pacific Business Process Management*. Springer International Publishing, 2013. p. 1-22.
- Van der Aalst, W. *Process Mining. Discovery, Conformance and Enhancement of Business Processes*. Springer, Heidelberg, 2011.
- VAN DER AALST, W. *et al.*. Manifiesto sobre Minería de Procesos. *IEEE Task Force on Process Mining*. BPM LATAM. 2012. <http://www.win.tue.nl/ieetfpm/lib/exe/fetch.php?media=shared:pmm-spanish-v1.pdf>.
- VAN DER AALST, W. M.P., Pesic, M. and Song, M.. *Beyond Process Mining: From the Past to Present and Future*. En *Advanced Information Systems Engineering*. Springer Berlin Heidelberg, 2010. p. 38-52.
- VAN DER AALST, W., ADRIANSYAH, A. Y VAN DONGEN, B. *Replaying History on Process Models for Conformance Checking and Performance Analysis*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2012, vol. 2, no 2, p. 182-192.
- W. GÜNTHER, CH. Y VAN DER AALST, W. Fuzzy Mining – Adaptive Process Simplification Based on Multi-Perspective Metrics. *En Business Process Management*. Springer Berlin Heidelberg, 2007. p. 328-343.
- WEIJTERS, A.J.M.M. Y RIBEIRO, J.T.S. *Flexible Heuristics Miner (FHM)*. En *Computational Intelligence and Data Mining (CIDM)*, 2011 IEEE Symposium on. IEEE, 2011. p. 310-317
- YZQUIERDO H., R. *Modelo para la estimación de información ausente en las trazas usadas en la minería de proceso*. La Habana, 2013. En *Compumat 2013*. https://compumat.uci.cu/sites/default/files/public/p1015-ponencia-3265_0.pdf.
- YZQUIERDO H., R. , *et al.*. *Diagnóstico de proceso basado en el descubrimiento de subprocessos*. *Ingeniería Industrial*, 2012, vol. 33, no 2, p. 133-141.

- ZAREH F., R. Y HASAN A., S. *A Two Phase Approach for Process Mining in Incomplete and noisy Logs*.
International Journal of Computer Science Issues (IJCSI), 2012, vol. 9, no 1.