

Algoritmos de Minería de Proceso para el Descubrimiento Automático de Procesos

Marcos Hernán Rivas, Sussy Bayona-Oré

mrivasp@unmsm.edu.pe, sbayona@hotmail.com

Universidad Nacional Mayor de San Marcos, Av. Germán Amézaga s/n, Lima, Perú.

DOI: 10.17013/risti.31.33-49

Resumen: Un aspecto fundamental para gestionar y ejecutar procesos empresariales es el modelado de procesos. Para establecer las diferencias que existen entre los modelos preestablecidos y los modelos que han sido ejecutados se hace necesario revisar las huellas y los registros de eventos. Los registros de eventos son utilizados por la minería de procesos para descubrir el proceso real, mediante la extracción de conocimiento. Para conocer que algoritmos han sido desarrollados para el descubrimiento automático de procesos de negocios, se realizó una revisión de literatura de los artículos publicados en el periodo 2004-2017. Como resultado de la revisión 20 artículos primarios fueron identificados y analizados. Un total de 20 diferentes algoritmos fueron identificados. En el desarrollo de los algoritmos se han utilizado diferentes enfoques con predominio del enfoque general de algoritmo. Los algoritmos identificados en su mayoría utilizan redes Petri como técnica de modelamiento de procesos.

Palabras-clave: redes Petri; descubrimiento automático; minería de procesos; modelo de procesos; registro de eventos.

Process Mining Algorithms for Automated Process Discovery

Abstract: A fundamental aspect to manage and execute business processes is the process modeling. In order to establish the differences that exist between the pre-established models and the models that have been executed, the traces and records of events must be reviewed. Process Mining uses event logs to discover the real processes, through the extraction of knowledge. To know which algorithms have been developed for the automatic discovery of business processes, a literature review of articles published in the period 2004-2017 was carried out. As a result of the review, 20 primary articles were identified and analyzed. A total of 20 algorithms were identified using different approaches with predominance of the general algorithm approach. The algorithms identified mostly use Petri networks as a process modeling technique.

Keywords: Petri networks; automatic discovery; process mining; process model; event log.

1. Introducción

Una organización que desea supervivencia en el mundo competitivo empresarial del momento requiere efectividad, eficacia y eficiencia en sus procesos de negocios. Los procesos que no han sido bien diseñados tienen alta probabilidad de tener como resultado prolongados tiempos de respuesta, niveles de servicio por debajo de lo que ha sido establecido, utilización de recursos de forma desproporcionada, costos operativos en constante crecimiento y clientes descontentos con las entregas (Hermida, Rodríguez, Padrón, Domas, & López, 2016). Es por ello la relevancia de monitorear los procesos previos a su implementación, con la finalidad de descubrir las posibles fallas que podría haber en el diseño, fallas que se pueden presentar durante su ejecución e identificar posibles oportunidades de mejora de los procesos (van Der Aalst, 2015).

Para analizar y mejorar procesos se pueden utilizar como alternativas novedosas la Simulación y la Minería de Procesos, ambas soportadas por las Tecnologías de la Información y Comunicaciones (TICs) (Hermida et al, 2016). Agrawal, Gunopulos & Leymann (1998) fueron los primeros en presentar los conceptos de la minería de procesos. La minería de procesos puede ser visto desde dos puntos de vista: (1) el algorítmico, que ubica a la minería de procesos entre la minería de datos y la inteligencia computacional y (2) por la operatividad de los negocios que la ubica entre la modelación y el análisis de procesos (van der Aalst, 2016). El propósito es descubrir, monitorear y mejorar los procesos en ejecución (i.e., no los procesos supuestos), mediante la extracción de conocimiento de los registros de eventos dejados por los sistemas de información (van der Aalst et al., 2011).

Hoy para el análisis de procesos se usan técnicas que requieren mucho tiempo y los resultados que se esperan obtener dependen de la experiencia y las habilidades de los que realizan el estudio, con el riesgo de que se cometan errores. Sin embargo, los métodos empleados por la minería de procesos tienen menores costos en tiempo y recursos (Hermida et al, 2016). Para el descubrimiento del modelo de procesos se utilizan algoritmos que usan registros de eventos (representado como una tabla que contiene todos los eventos relacionadas con las actividades empresariales ejecutadas).

El modelo resultante puede ser una Red de Petri, un Modelo en Notación para la Gestión de Procesos de Negocio (BPMN, por sus siglas en inglés Business Process Model and Notation), una Cadena de Procesos Impulsadas por Eventos, (EPC, por sus siglas en inglés Event-driven Process Chain) o un diagrama en Lenguaje unificado de modelado (UML, por sus siglas en inglés, Unified Modeling Language) (Pourmasoumi & Bagheri, 2016).

Para determinar los algoritmos más utilizados en el descubrimiento automático de procesos se realizó una revisión sistemática de la literatura.

Este artículo está estructurado en 6 secciones incluido la Introducción. En la Sección 2 se describe los conceptos de minería de procesos, los tipos de minería, registros de eventos y el descubrimiento automático de procesos. En la Sección 3 se presenta la metodología utilizada para realizar la revisión de literatura. La Sección 4 muestra los resultados de la revisión de literatura. En la Sección 5 se discute los resultados. Finalmente, en la Sección 6 se presenta las conclusiones.

2. Marco Teórico

2.1. Minería de Procesos

Las organizaciones cada día requieren ser más competitivas y desean saber sobre la operatividad real de sus procesos, esta necesidad incentiva el desarrollo de nuevas técnicas de minería de procesos. 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 (Norambuena & Zepeda, 2017). Para satisfacer las necesidades de los especialistas en el campo de la minería de procesos, diversos algoritmos basados en algoritmos de otros campos de la informática han sido desarrollados (Tiwari, Turner & Majeed, 2008). La minería de procesos se inicia con Cook & Wolf (1998), quienes descubren modelos de procesos de software a partir de registros de eventos. Posteriormente Cook, Du, Liu & Wolf (2004) trasladaron sus resultados obtenidos en el campo de procesos de software a los sistemas de flujo de trabajo descubriendo modelos de comportamiento concurrentes. La minería de procesos puede ser utilizado para explorar los procesos históricos de la organización con el propósito de implementar mejoras (Méjia & Muñoz). Las técnicas de minería de procesos son utilizadas para descubrir, monitorear y mejorar los procesos reales en una organización. Estas técnicas facilitan la extracción de conocimiento de los registros de eventos dejados por los sistemas de información y como resultado brinda un modelo que describe los procesos. Para aplicar minería de procesos se requiere (a) contar con los registros de eventos de los procesos ejecutados, (b) identificar qué eventos pertenecen a la misma instancia del proceso y (c) que exista un orden de ejecución para los eventos en la misma instancia del proceso (van der Aalst, Weijters & Maruster, 2004).

2.2. Tipos de Minería de Procesos.

Dustdar, Hoffmann, & van der Aalst (2005) establecieron tres perspectivas en la minería de procesos de negocios: (a) perspectiva de tener un control de flujo, (b) perspectiva del desarrollo organizacional y (c) perspectiva de seguimiento del caso. El control de flujo es una perspectiva que se interesa por el orden de ejecución de las actividades, y el objetivo es poder mapear todas las rutas posibles que se pueden ejecutar en el proceso. La perspectiva organizativa tiene como objetivo construir la red social de la organización clasificando a las personas por su interacción teniendo en cuenta funciones y unidades organizativas a la que pertenecen. La perspectiva del caso se interesa por los elementos que pueden caracterizar el caso como la ruta de ejecución en el proceso, los recursos que se utilizan o por los valores que toman sus elementos de datos. La perspectiva de control de flujo puede extenderse con eventos que tienen marcas de tiempo. Van der Aalst & Vanthienen (2011) adiciona una perspectiva del tiempo que se ocupa de la sincronización y de la frecuencia de ejecución de las actividades dentro del proceso. Adicionalmente a las perspectivas descritas (Dustdar, Hoffmann, & van der Aalst, 2005), en (van der Aalst & Vanthienen, 2011), describen tres tipos básicos de minería de procesos: (a) *Descubrimiento* de procesos, cuando la entrada es un registro de eventos y la salida es un modelo de proceso, (b) *Conformidad*, cuando el modelo de proceso inicialmente diseñado y sus reglas se compara con los resultados obtenidos de un registro de eventos del mismo proceso, y (c) *Mejora*, cuando se pretende mejorar el modelo de proceso existente utilizando el conocimiento que se extrae del registro de eventos.

2.3. Registro de Eventos (RE)

La minería de Procesos utiliza un enfoque de registro de eventos para almacenar la información sobre la ejecución real del proceso de negocio, para su construcción extrae los datos almacenados por los sistemas de información que se crea durante el procesamiento de transacciones. Los registros de eventos representan la base de datos de entrada para ejecutar algoritmos de minería de procesos (van der Aalst et al., 2003). Un registro de eventos es representado como una tabla que contiene todos los eventos relacionadas con las actividades empresariales ejecutadas. Un caso tiene asignados eventos. Una sola ejecución de un proceso empresarial se denomina instancia (o caso) del proceso y se reflejan en el registro de eventos como un conjunto de eventos que se asignan al mismo caso. La secuencia de eventos registrados en un caso se llama traza (vanden Broucke & De Weerd, 2017). El modelo que describe la ejecución de una única instancia de proceso se denomina modelo de instancia de proceso. Casos y eventos se caracterizan por clasificadores y atributos. Los clasificadores aseguran la distinción de casos y eventos al asignar nombres únicos a cada caso y evento. Los atributos almacenan información adicional que se puede utilizar con fines de análisis.

Un modelo de la estructura de un registro de eventos se presenta en la Tabla 1. El registro de eventos de un proceso particular (Tabla 1), tiene 4 elementos de datos: Case ID, que implica una instancia de proceso, Activity, que es la tarea que se está realizando, Resource, se refiere al actor responsable de realizar la actividad (persona, grupo, departamento, etc.), timestamp indica el momento en que se registró un evento en el sistema. Un

Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	Register request	Pete	50	...
	35654424	31-12-2010:10.06	Examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	Check ticket	Mike	100	...
	35654426	06-01-2011:11.18	Decide	Sara	200	...
	35654427	07-01-2011:14.24	Reject request	Pete	200	...
2	35654483	30-12-2010:11.32	Register request	Mike	50	...
	35654485	30-12-2010:12.12	Check ticket	Mike	100	...
	35654487	30-12-2010:14.16	Examine casually	Pete	400	...
	35654488	05-01-2011:11.22	Decide	Sara	200	...
	35654489	08-01-2011:12.05	Pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	Register request	Pete	50	...
	35654522	30-12-2010:15.06	Examine casually	Mike	400	...
	35654524	30-12-2010:16.34	Check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	Decide	Sara	200	...
	35654526	06-01-2011:12.18	Reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	Examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	Check ticket	Pete	100	...
	35654531	09-01-2011:09.55	Decide	Sara	200	...
	35654533	15-01-2011:10.45	Pay compensation	Ellen	200	...
4	35654641	06-01-2011:15.02	Register request	Pete	50	...
	35654643	07-01-2011:12.06	Check ticket	Mike	100	...
	35654644	08-01-2011:14.43	Examine thoroughly	Sean	400	...
	35654645	09-01-2011:12.02	Decide	Sara	200	...
	35654647	12-01-2011:15.44	Reject request	Ellen	200	...
...

Tabla 1 – Estructura de un registro de eventos (van der Aalst, 2011)

proceso consiste en casos, un caso consiste en eventos tales que cada evento se relaciona precisamente con un caso, se ordenan los eventos dentro de un caso, los eventos pueden tener atributos (tiempo, actividad, costos y recursos) (van der Aalst, 2011).

2.4. Descubrimiento Automático de Procesos

El descubrimiento automático de procesos tiene como objetivo la construcción de un modelo de procesos ejecutando un algoritmo que tiene como entrada un registro de eventos y como resultado una representación del comportamiento del proceso. El modelo de proceso que se ha construido debería tener la capacidad de reproducir los casos del registro de eventos utilizado y prohibir el comportamiento de lo que no existe en los registros (van der Aalst et al., 2010). Teniendo en cuenta el requisito que cualquier evento debe estar vinculado a un caso (instancia de proceso) y a una actividad, y suponiendo que sólo esta información está disponible, un evento es descrito por un par (c, a) donde c se refiere al caso y a se refiere a la actividad. Por lo tanto, suponemos que cada caso se ejecuta independientemente de otros casos, es decir, el enrutamiento de un caso no depende del enrutamiento de otros casos (aunque pueden competir por los mismos recursos). Como resultado, podemos centrarnos en el ordenamiento de actividades dentro de casos individuales. Por lo tanto, un solo caso σ puede representarse como una secuencia de actividades, es decir, un trazo $\sigma \in A$, donde A es el conjunto de actividades. En consecuencia, un registro puede ser visto como una colección de trazas L , es decir, $(L \subseteq A)$. En la Figura 1, tomado de (van der Aalst et al., 2010), se muestra un ejemplo de registro y el correspondiente modelo de proceso descubierto utilizando el algoritmo α (van der Aalst, Weijters, & Maruster, 2004). Se visualiza que la red de Petri generada es capaz de reproducir el registro, es decir, hay un buen ajuste entre el registro y el modelo de proceso descubierto.

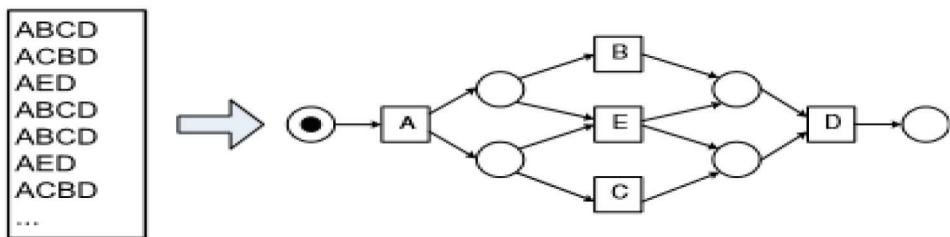


Figura 1 – Log representado por secuencias de actividades y el modelo de proceso que se descubre utilizando el algoritmo α (van der Aalst et al., 2010)

2.5. Enfoque en Desarrollo de Algoritmos de Descubrimiento.

Considerando los trabajos de (Tiwari, Turner & Majeed, 2008; Turner, Tiwari, Olaiya, & Xu, 2012; De Weerd, De Backer, Vanthienen, & Baesens, 2012; Pourmasoumi & Bagheri, 2016), existen diferentes enfoques en las técnicas de descubrimiento automático de procesos tales como: Enfoque algorítmico general como el algoritmo α , (b) Algoritmo Genético que utiliza un enfoque evolutivo, (c) Enfoque Markoviano que son algoritmos que examinan el comportamiento pasado y futuro para definir el estado actual (Cook

& Wolf, 1998), (d) Redes Neuronales caracterizados por su capacidad de “aprender” (Cook & Wolf, 1998), (e) Técnicas de agrupamiento que divide un grupo de soluciones en subgrupos homogéneos (Schimm, 2004), (f) Minería de Procesos basado en regiones que pueden ser de dos tipos, basados en el estado (van Dongen, Busi, Pinna & van der Aalst, 2007) y basados en el lenguaje (Bergenthum, Desel, Lorenz & Mauser, 2007), (g) Enfoque Heurístico mediante la representación de redes causales (Weijters & Ribeiro, 2011), (h) Enfoque de Minería declarativa que introduce eventos negativos generados artificialmente que muestran que no es posible una ruta particular a través de un proceso (Goedertier, Martens, Vanthienen & Baesens, 2009) o (i) Minería borrosa que es recomendado para procesos menos estructurados (Günther & van der Aalst, 2007).

3. Metodología

Para realizar la revisión de la literatura se ha utilizado la técnica de revisión sistemática sobre la base de las directrices originales propuestas por (Kitchenham, 2004). Una revisión sistemática de la literatura permite identificar, evaluar e interpretar la investigación disponible y relevante para dar respuesta a la pregunta de investigación en particular. Los estudios individuales que contribuyen a una revisión sistemática se denominan estudios primarios y la revisión sistemática es un estudio secundario (Kitchenham, 2004). Las etapas en el método de revisión sistemática de la literatura son: planificar la revisión sistemática, conducir la revisión sistemática y presentar los resultados. A continuación, se describen cada una de estas etapas.

3.1. Planificar la Revisión Sistemática

En esta etapa se elaboró el protocolo a ser usado para conducir la revisión sistemática. El protocolo incluye (i) la pregunta de investigación, (ii) la estrategia de búsqueda, (iii) los criterios de inclusión y exclusión, (iv) los criterios de calidad y (v) el método de extracción de datos. El objetivo de la investigación es determinar el estado actual de los métodos / técnicas / Algoritmos para realizar descubrimiento automático de proceso de negocios.

La pregunta principal de investigación definida en este estudio es:

P1: ¿Qué algoritmos de minería de procesos existen para realizar descubrimiento automático de procesos de negocios?”

Para crear la cadena de búsqueda se utilizaron las siguientes palabras clave: Process mining, data mining, machine learning, Workflow mining, event log, BPMN, Petri nets, Discovery, Techniques, Algorithms. En la construcción de la cadena de búsqueda se ha utilizado los operadores booleanos AND y OR.

(“Process mining” OR “data mining” OR “machine learning” OR “Workflow mining”) AND (“event log”) AND (“BPMN” OR “Petri nets”) AND (“Discovery”) AND (“Techniques” OR “Algorithms”)

Las bases de datos seleccionadas para la búsqueda de los artículos fueron ACM, IEEE Xplore, ScienceDirect, Scopus, Springer y Google Scholar.

3.2. Conducir la Revisión Sistemática

Según el protocolo establecido en la etapa de planificación, se procedió con la búsqueda de artículos en las bases de datos electrónicas seleccionadas, haciendo uso de la cadena de búsqueda (artículos publicados desde el año 2004 a mayo de 2017). Al aplicar la cadena de búsqueda en cada una de las bases de datos seleccionadas, se identificaron un total de 295 artículos. Después de aplicar los criterios de inclusión, exclusión (véase Tabla 2) y los criterios de calidad se seleccionaron 40 artículos potenciales.

Criterios de Inclusión	Criterios de Exclusión
(a) La investigación propone alguna técnica / algoritmo / Método de minería de procesos para el descubrimiento automático de procesos de negocios, esto es realiza la construcción de un modelo de procesos como una representación abstracta de los registros de eventos	(a) El estudio está centrado en otro objetivo de la minería de proceso diferente al de descubrimiento de procesos de negocios
(b) El modelo de procesos representado desde los registros de eventos utiliza BPMN o Redes de Petri	(b) El estudio sólo muestra una aplicación utilizando software sin evaluar el algoritmo
(c) El enfoque presentado es aplicación a casos de estudio o evaluación experimental.	

Tabla 2 – Criterios de Inclusión / Exclusión

Finalmente se hizo una revisión detallada de los algoritmos en los artículos quedando definido 20 artículos primarios a estudiar en el presente trabajo (véase Tabla 3). Los criterios de calidad para evaluar la calidad de los estudios fueron:

- a. Están los objetivos de la investigación claramente especificados?
- b. ¿Se diseñó el estudio para lograr estos objetivos?
- c. ¿Se describen claramente los métodos / técnicas/ algoritmos utilizados?
- d. ¿Se presenta de manera clara el registro de eventos utilizado en el estudio?
- e. ¿Los hallazgos del estudio están plenamente justificados?

En el proceso se detectó que varios artículos carecían de detalles suficientes sobre las limitaciones de algoritmo, muchos estudios se centran sólo en superar algunas deficiencias del algoritmo α y no evalúan limitaciones de otros problemas que fueron definidos en el descubrimiento automático de procesos. En la Figura 2 se presenta el diagrama del proceso para conducir la revisión.

En los 20 estudios primarios seleccionados se exponen algoritmos para el descubrimiento automático de procesos de negocios bajo diferentes enfoques. Se diseñó un formato de extracción de datos para registrar los siguientes datos: Algoritmo, Autor(s), Año, Característica funcional del Algoritmo, Enfoque utilizado, Algoritmo detallado (si/no), Limitaciones, Fortaleza, Técnica de modelamiento, Caso de estudio utilizado. Los resultados de la revisión sistemática se presentan con mayor detalle en la Sección 4.

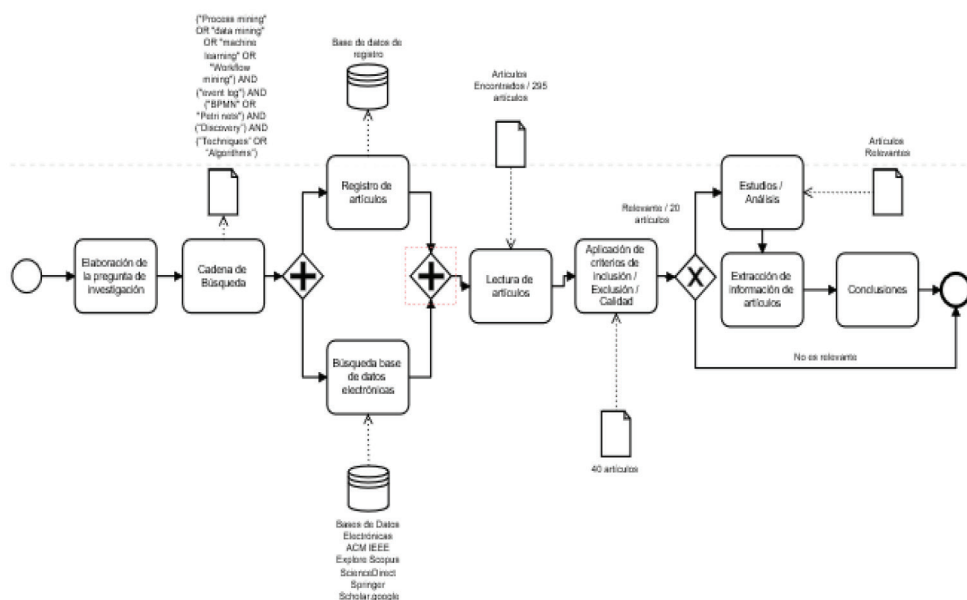


Figura 2 – Proceso para conducir la revisión

Bases de Datos	Estudios			
	Cadena Búsqueda	Inclusion / Exclusion	Calidad	Algoritmos
ACM	30	7	7	6
IEEE Xplore	13	4	4	1
Scopus	59	15	9	1
ScienceDirect	123	15	12	5
Springer	70	5	2	1
Scholar.google		7	6	6
TOTAL	295	53	40	20

Tabla 3 – Resumen de resultados de búsqueda

4. Resultados

Según los resultados de la revisión, el desarrollo de los algoritmos ha evolucionado a lo largo del tiempo (véase Figura 3).

En total se identificaron 20 algoritmos desarrollados en los últimos años (véase Tabla 4). En la Tabla 4, se muestra el enfoque utilizado para cada algoritmo identificado en

el estudio, que se representa con la siguiente nomenclatura: EAG: Enfoque algorítmico general, AE: Aproximación evolutiva y EH: Enfoque heurístico. La columna TMU representa la Técnica de modelado.

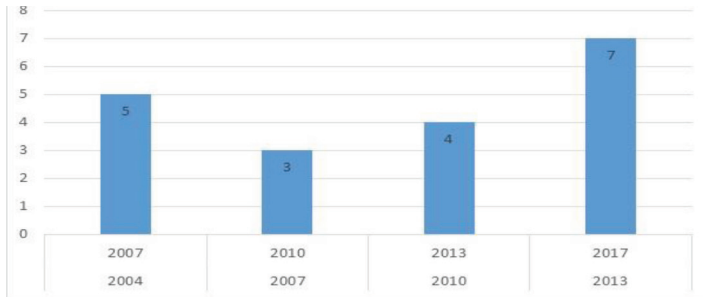


Figura 3 – Frecuencia de artículos por intervalos de años

De los resultados, se observa que la mayoría de los algoritmos fueron construidos tomando como base los resultados obtenidos en el algoritmo α . En varios estudios los autores se enfocan en solucionar los casos de estructuras de procesos donde el algoritmo α no da como resultado respuesta satisfactoria, y otros lo utilizan en la etapa de experimentación para evaluar las ventajas del nuevo algoritmo desarrollado. Se han utilizado diferentes enfoques de desarrollo y diferentes técnicas de modelado para el flujo de resultado. Todos los algoritmos funcionan para diferentes condiciones de entradas o suposiciones. En la Tabla 4 se presenta información relevante para obtener buenos resultados en el uso de los algoritmos.

ID	Algoritmos	TMU	Enfoque Utilizado	Información Importante
A1	α Algorithm	Petri net	EAG	Asegura buen resultado si no hay ruido y además se requiere que el registro de flujo de trabajo sea completo es decir todos los casos deben estar completos de inicio a fin (van der Aalst, Weijters & Maruster, 2004).
A2	$\alpha+$ Algorithm	Petri net	EAG	Presupone información perfecta: (i) el registro debe estar completo y (ii) no hay ruido en el registro (es decir, todo lo que está registrado en el registro es correcto) (Alves de Medeiros, Van Dongen, van der Aalst & Weijters, 2004).
A3	Genetic Algorithm	Petri net	AE	El principal valor agregado de usar un algoritmo genético es la capacidad de lidiar con el ruido y cuando el registro no está completo (van der Aalst, Alves de Medeiros, Weijters, 2005).
A4	The Heuristics Miner algorithm	Causal Net	EH	Puede tratar con ruido y pocos tipos de casos, es decir, puede trabajar con un registro de eventos con miles de trazas, pero con poca frecuencia de comportamiento y algo de ruido (Weijters, van der Aalst & Alves De Medeiros, 2006).

ID	Algoritmos	TMU	Enfoque Utilizado	Información Importante
A5	$\alpha++$ Algorithm	Petri net	EAG	Se asume información perfecta: (i) el registro debe estar completo y (ii) el registro está libre de ruidos (es decir, cada evento registrado en el registro es correcto y representativo del modelo que debe descubrirse) (Wen, van der Aalst, Wang & Sun, 2007).
A6	β -algorithm	Petri net	EAG	Suponiendo un registro completo, se puede demostrar que el algoritmo β puede descubrir correctamente cualquier red WorkFlow (Wen, Wang, van der Aalst, Huang & Sun, 2009).
A7	<i>Region-based workflow mining</i>	Petri net	Enfoque basado en regiones	El clásico algoritmo α es incapaz de redescubrir varias redes, mientras que el enfoque basado en regiones puede explotarlas correctamente, pero es demasiado complejo (Busi & Pinna, 2009).
A8	<i>Mining algorithm $\alpha\#$</i>	Petri net	EAG	La capacidad de extracción del algoritmo está limitada a las redes DIWF (es una subclase grande de redes WF) y se supone que el registro de eventos dado está completo (Wen, Wang, van der Aalst, Huang & Sun, 2010).
A9	<i>Mining Algorithm χ- (Algorithm Based on Event Type)</i>	Petri net	EAG	Propone una estructura novedosa de registros de eventos para descubrir las dependencias causales. Luego propone un nuevo algoritmo para redescubrir el modelo de proceso con este tipo de registros de eventos (Wang, Ge, Hu, & Luo, 2011).
A10	<i>Mining Algorithm λ</i>	Petri net	EAG	No se centra en los problemas de ruido (el registro de eventos es correcto) y se supone que el registro de eventos contiene información suficiente (el registro de eventos está completo) (Wang, Ge, Hu, Luo, & Huang, 2012).
A11	<i>Guided Local Search</i>	BPMN	Búsqueda local guiada	Con un mismo registro de eventos compara los resultados obtenidos en el descubrimiento del proceso con los resultados obtenidos con el algoritmo Heuristic Miner, obteniendo mejores resultados (Chinches & Salomie, 2013).
A12	<i>Ant Colony Optimization Algorithm</i>	BPMN	Optimización de la colonia de hormigas	Compara el proceso descubierto por ACO BP Miner con el proceso de descubrimiento realizado con Genetic Miner, ACO BP Miner encontró significativamente menos errores que Genetic Miner (Chinches & Salomie, 2013b).
A13	<i>Algorithm: Evolution Strategy</i>	BPMN	AE	La estrategia de evolución encuentra de manera eficiente los modelos de proceso que mejor representan un registro de eventos (Molka, Redlich, Drobek, Zeng, & Gilani, 2015).
A14	<i>The Fusion Miner algorithm</i>	Petri net	Fusión de algoritmos	Los resultados muestran que este enfoque produce modelos más precisos y completos, especialmente en entornos con múltiples capas de flexibilidad (De Smedt, De Weerd, & Vanthienen, 2015).

ID	Algoritmos	TMU	Enfoque Utilizado	Información Importante
A15	<i>Technique BPMN Miner</i>	BPMN	EH	Se requiere que los eventos en el registro incluyan atributos de datos. Además, el conjunto de atributos debe incluir una clave principal para cada subproceso a descubrir, así como una clave externa para cada Proceso - subproceso (Conforti, Dumas, García-Bañuelos & La Rosa, 2016).
A16	<i>Algorithm (τ) with token carried data</i>	Petri net	Utilización de Tokens	En el manejo de lugares paralelos con entradas múltiples, sus resultados son mejores que los algoritmos $\alpha +$, $\alpha ++$, β y λ (Li, C., Ge, J., Huang, L., Hu, H., Wu, B., Yang, H., et al., 2016).
A17	<i>Discovery algorithm of synchronization-core-based process models</i>	BPMN	Árbol de decisión	Se requiere que los registros del evento estén completos. Además, el enfoque propuesto no considera el ruido, los modelos de procesos múltiples y las tareas duplicadas (Lu, Zeng, & Duan, 2016).
A18	<i>Algorithm SLAD (Splitting Labels After Discovery)</i>	Petri net	EAG	Es un algoritmo para abordar tareas duplicadas en un modelo ya descubierto (Vázquez-Barreiros, Mucientes, & Lama, 2016).
A19	<i>Fodina</i>	Petri net	EH	Mejora y expande el algoritmo Heuristics Miner (vanden Broucke & De Weerd, 2017).
A20	<i>Mining $\lambda+$</i>	Petri net	EAG	Si el modelo contiene la subred simétrica que no está bien estructurada, podría no descubrir el modelo (Li, Ge, Huang, Hu, Wu, Hu et al., 2017).

Tabla 4 – Algoritmos utilizados en el descubrimiento automático de procesos

Los algoritmos han sido desarrollados utilizando diferentes enfoques tales como enfoque general, enfoque evolutivo, enfoque heurístico, enfoque basado en regiones entre otros (véase Tabla 5).

Enfoque Utilizado	Estudios	Descripción
<i>Enfoque Algorítmico General</i>	[A1] [A2] [A5] [A6] [A8] [A9] [A10] [A18] [A20]	Es un enfoque de descubrimiento que produce una red de Petri para el comportamiento del proceso registrado en los registros de eventos (van der Aalst, Weijters, & Maruster, 2004).
<i>Enfoque evolutivo</i>	[A3] [A13]	Utiliza un enfoque evolutivo (Alves de Medeiros, Weijters & van der Aalst, 2004a).
<i>Enfoque heurístico</i>	[A4] [A15] [A19]	Utiliza la representación de redes causales, en el que los nodos son actividades y los arcos son dependencias causales (Weijters & Ribeiro, 2011).
<i>Enfoque basado en regiones</i>	[A7]	Utiliza la Teoría de Regiones en la minería de procesos para modelar los conjuntos de estados que caracterizan en una red Petri.

Enfoque Utilizado	Estudios	Descripción
<i>Otros enfoques</i>	[A11] [A12] [A14] [A16] [A17]	Redes Neuronales como modelos de la mente humana en su capacidad de “aprender” e identificar patrones en los datos (Cook & Wolf, 1998). Técnicas de agrupamiento que divide un grupo de soluciones en subgrupos homogéneos. Enfoque de Minería declarativa que introduce eventos negativos generados artificialmente que muestran que no es posible una ruta particular a través de un proceso. Minería borrosa que es recomendado para procesos menos estructurados, permite simplificar la visualización a través de cluster.

Tabla 5 – Algoritmos y enfoques utilizados en su desarrollo

5. Discusión

5.1. Algoritmos desarrollados

De la revisión de los estudios primarios fueron identificados 20 algoritmos desarrollados utilizando diferentes enfoques tales como enfoque general, enfoque evolutivo, enfoque heurístico, enfoque basado en regiones entre otros. Un registro completo de eventos es un requisito para obtener buenos resultados, es decir, con casos que comienzan y terminan. Los algoritmos en su mayoría están enfocados en superar los problemas de manejo de bucle 1 y 2 que el algoritmo α no puede resolver. Una tendencia es utilizar BPMN que se ha convertido en el estándar para modelar los procesos, pero la mayoría de los algoritmos estudiados utiliza redes Petri.

5.2. Sobre el manejo de los algoritmos para registro de eventos con presencia de ruidos

El resultado de los algoritmos depende de un buen registro de eventos, la mayoría de ellos requieren registros de eventos sin ruido, por ejemplo, en los algoritmos que utilizan el enfoque evolutivo, a un mayor porcentaje de ruido en el registro de eventos hay menos probabilidad de que el algoritmo entregue como resultado una red de proceso correcta. El estudio muestra que no es posible descubrir arbitrariamente flujo de trabajo, los algoritmos funcionan para ciertos tipos de flujos de trabajo y en los nuevos desarrollos se observan que se enfocan en resolver los problemas encontrados básicamente por el algoritmo α . Todavía está pendiente resolver el manejo del registro de eventos con ruido. Los datos de los eventos suelen ser dispersos en diferentes fuentes de datos y, a menudo, se necesita un esfuerzo para recopilar los datos relevantes y filtrarlos para evitar el ruido. En (van der Aalst, Schonenberg & Song, 2011) se realiza una descripción general del flujo de trabajo para obtener un buen registro de eventos para la minería de procesos de fuentes de datos heterogéneas.

5.3. Sobre la técnica de modelamiento utilizada por el algoritmo

Los algoritmos para representar el proceso descubierto utilizan técnicas de modelamiento de procesos. Se ha observado que la técnica más utilizada es la de redes Petri ([A1] [A2] [A3] [A5] [A6] [A7] [A8] [A9] [A10] [A14] [A16] [A18] [A19] [A20]) y lo utilizan para la generación del modelo final como paso intermedio una relación de orden. Es decir, según la revisión, en el 70% de los estudios se usa la técnica de redes Petri. En segundo lugar, se utiliza la técnica de los árboles de procesos BPMN que hoy es el estándar en modelamiento de procesos ([A11] [A12] [A13] [A15] [A17]). La tendencia es que los algoritmos descubran el proceso en el lenguaje BPMN, estándar adoptado por la industria de proceso. En (Kalenkova, van der Aalst, Lomazova & Rubin, 2016) se proponen algunos algoritmos que permiten la conversión de formalismos conocidos como Redes de Petri, Redes Causales y Árboles de Proceso a BPMN que es el estándar que se está imponiendo para modelar procesos de negocios. De los artículos revisados el 25% usa el lenguaje BPMN. Finalmente, los algoritmos con enfoque heurístico se utilizan para representar el proceso descubierto Causal Net, el algoritmo SLAD (Splitting Labels After Discovery) usa Causal Net como paso intermedio para mejorar y obtener una Red Petri (Alves de Medeiros, Weijters & van der Aalst, 2004a). Solo en el 5% de los estudios seleccionados se utiliza la técnica Causal Net.

6. Conclusiones

La minería de procesos utiliza los registros de eventos para descubrir los procesos que se están enjutando en el mundo reales del negocio, mediante la extracción de conocimiento. Los resultados de la revisión de los estudios seleccionados (20) muestran que la mayoría de los algoritmos utilizan un enfoque general. Los resultados son aceptables cuando los registros de datos son completos y limpios. El algoritmo α es la base para el desarrollo de las nuevas propuestas que tratan de resolver problemas que en los inicios no pudieron ser resueltos. Para validar los algoritmos se utilizan casos cuyos resultados son comparados con los resultados obtenidos de otros algoritmos desarrollados previamente. Los algoritmos utilizan técnicas de modelamiento para descubrir los procesos y como resultado se cuenta con un flujo de trabajo mapeado en un lenguaje de modelado. Una tendencia es utilizar BPMN que se ha convertido en el estándar para modelar los procesos, pero la mayoría de los algoritmos estudiados utiliza redes Petri. A pesar de los avances mostrados, la presencia de ruido en los registros de eventos es un problema que no ha sido solucionado para descubrir los procesos de manera satisfactoria.

Referencias

- Agrawal, R., Gunopulos, D., & Leymann, F. (1998). Mining process models from workflow logs. Paper presented at the *International Conference on Extending Database Technology*, (pp. 467-483).
- Alves de Medeiros, A. A., Van Dongen, B., Van der Aalst, W., & Weijters, A. (2004). *Process Mining: Extending the α -Algorithm to Mine Short Loops*. Eindhoven: Technische Universiteit Eindhoven.

- Alves de Medeiros, A.K., Weijters, A.J.M.M., & van der Aalst, W.M.P. (2004). *Using genetic algorithms to mine process models: representation, operators and results*. In: Beta Working Paper Series, WP 124. Eindhoven: Eindhoven University of Technology.
- Bergenthum, R., Desel, J., Lorenz, R., & Mauser, S. (2007). Process Mining Based on Regions of Languages. In: G. Alonso, P. Dadam, & M. Rosemann, (Eds). *International Conference on Business Process Management (BPM 2007)*, volume 4714 of *Lecture Notes in Computer Science*, (pp 375–383). Berlin: Springer.
- Bratosin, C., Sidorova, N., & van der Aalst, W. (2010). Distributed genetic process mining. Paper presented at the *Evolutionary Computation (CEC), 2010 IEEE Congress on*, (pp. 1-8).
- Busi, N., & Pinna, G. M. (2009). Process discovery and petri nets. *Mathematical Structures in Computer Science*, 19(06), 1091–1124.
- Chinces, D., & Salomie, I. (2013). Business process mining using guided local search. In: *Proceedings of the Parallel and Distributed Computing (ISPDC), 2013 IEEE 12th International Symposium on*, (pp. 177-181).
- Chinces, D., & Salomie, I. (2013). Process discovery using ant colony optimization. In: *Proceedings of the Control Systems and Computer Science (CSCS), 2013 19th International Conference on*, (pp. 448-454).
- Conforti, R., Dumas, M., García-Bañuelos, L., & La Rosa, M. (2016). BPMN miner: Automated discovery of BPMN process models with hierarchical structure. *Information Systems*, 56, 284–303.
- Cook, J. E., & Wolf, A. L. (1998). Discovering models of software processes from event-based data. *ACM Transactions on Software Engineering and Methodology*, 7(3), 215–249.
- Cook, J. E., Du, Z., Liu, C., & Wolf, A. L. (2004). Discovering models of behavior for concurrent workflows. *Computers in Industry*, 53(3), 297–319.
- De Smedt, J., De Weerd, J., & Vanthienen, J. (2015). Fusion miner: Process discovery for mixed-paradigm models. *Decision Support Systems*, 77, 123–136.
- De Weerd, J., De Backer, M., Vanthienen, J., & Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, 37(7), 654–676.
- Dustdar, S., Hoffmann, T., & Van der Aalst, W. (2005). Mining of ad-hoc business processes with TeamLog. *Data & Knowledge Engineering*, 55(2), 129–158.
- Goedertier, S., Martens, D., Vanthienen, J., & Baesens, B. (2009). Robust process discovery with artificial negative events. *Journal of Machine Learning Research*, 10, 1305–1340.

- Günther, C., & Van der Aalst, W. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. En: Alonso, G., Dadam, P., & Rosemann, M. (Eds.), *Business Process Management* (Vol. 4714, 2007, p. 328-343). Berlín: Springer Berlin Heidelberg.
- Hermida, A. G., Rodríguez, M. S., Padrón, M. V., Domas, L. R., & López, R. R. (2016). Alternativa para realizar análisis de rendimiento de procesos automatizados empleando técnicas de minería de procesos y simulación sobre redes de petri. *Investigación Operacional*, 37(1), 93–103.
- Kalenkova, A. A., van der Aalst, W.M.P., Lomazova, I. A., & Rubin, V. A. (2016). 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).
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele University*, 33(2004), 1–26.
- Li, C., Ge, J., Huang, L., Hu, H., Wu, B., Hu, H., et al. (2017). Software cybernetics in BPM: Modeling software behavior as feedback for evolution by a novel discovery method based on augmented event logs. *Journal of Systems and Software*, 124, 260–273.
- Li, C., Ge, J., Huang, L., Hu, H., Wu, B., Yang, H., et al. (2016). Process mining with token carried data. *Information Sciences*, 328, 558–576.
- Lu, F., Zeng, Q., & Duan, H. (2016). Synchronization-core-based discovery of processes with decomposable cyclic dependencies. *ACM Transactions on Knowledge Discovery from Data*, 10(3), 31.
- Mejía, J., & Muñoz, M. (2017). Tendencias en Tecnologías de Información y Comunicación. RISTI - Revista Ibérica de Sistemas y Tecnología de información, 21, 51-66. Doi: 10.17013/risti.21.51-66.
- Molka, T., Redlich, D., Drobek, M., Zeng, X., & Gilani, W. (2015). Diversity guided evolutionary mining of hierarchical process models. In: *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, (pp. 1247-1254).
- Norambuena, B., & Zepeda, V. (2017). Minería de procesos de software: una revisión de experiencias de aplicación. RISTI - Revista Ibérica de Sistemas y Tecnología de información, 21, 51-66. Doi: 10.17013/risti.21.51-66.
- Pourmasoumi, A., & Bagheri, E. (2016). Business process mining. *Encyclopedia with Semantic Computing*, 1(1), 1–4.
- Schimm, G. (2004). Mining exact models of concurrent workflows. *Computers in Industry*, 53, 265–81.
- Tiwari, A., Turner, C. J., & Majeed, B. (2008). A review of business process mining: State-of-the-art and future trends. *Business Process Management Journal*, 14(1), 5–22.

- Turner, C. J., Tiwari, A., Olaiya, R., & Xu, Y. (2012). Process mining: From theory to practice. *Business Process Management Journal*, 18(3), 493–512.
- Van Der Aalst, W., Rubin, V., Verbeek, H. M., van Dongen, B. F., Kindler, E., & Günther, C. W. (2010). Process mining: A two-step approach to balance between underfitting and overfitting. *Software and Systems Modeling*, 9(1), 87–111.
- Van der Aalst, W., & Vanthienen, J. (2011). IEEE task force on process mining. *Lecture Notes in Business Information Processing*, 99, 169–194.
- Van Der Aalst, W., Adriansyah, A., & Van Dongen, B. (2011). Causal nets: A modeling language tailored towards process discovery. In: *Proceedings of International Conference on Concurrency Theory*, (pp. 28-42).
- Van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9), 1128–1142.
- Van der Aalst, W., & Weijters, A. (2004). Process mining: A research agenda. *Computers in Industry*, 53(3), 231–244.
- Van der Aalst, W., Alves De Medeiros, A. A., & Weijters, A. (2005). Genetic process mining. In: *Proceedings of International Conference on Application and Theory of Petri Nets*, (pp. 48-69).
- Van der Aalst, W., Schonenberg, M. H., & Song, M. (2011). Time prediction based on process mining. *Information Systems*, 36(2), 450–475.
- Van der Aalst, W., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A. J. (2003). Workflow mining: A survey of issues and approaches. *Data & Knowledge Engineering*, 47(2), 237–267.
- Van der Aalst, W. (2011). Getting the data. In: *Process mining* (pp. 95-123). Berlin: Springer.
- Van Der Aalst, W. (2015). Business process simulation survival guide. In: *Handbook on business process management 1* (pp. 337-370). Berlin: Springer.
- Van der Aalst, W. (2016). *Process mining: Data science in action*. Berlin: Springer.
- van Dongen, B.F., Busi, N., Pinna, G.M., & van der Aalst, W. (2007). An Iterative Algorithm for Applying the Theory of Regions in Process Mining. In: W. Reisig, K. van Hee, & K. Wolf, (eds). *Proceedings of the Workshop on Formal Approaches to Business Processes and Web Services (FABPWS'07)*, (pp.36– 55). Siedlce: Publishing House of University of Podlasie.
- Van den Broucke, S. K., & De Weerd, J. (2017). Fodina: A robust and flexible heuristic process discovery technique. *Decision Support Systems*.
- Vázquez-Barreiros, B., Mucientes, M., & Lama, M. (2016). Enhancing discovered processes with duplicate tasks. *Information Sciences*, 373, 369–387.

- Wang, D., Ge, J., Hu, H., & Luo, B. (2011). A new process mining algorithm based on event type. In: *Proceedings of Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*, (pp. 1144-1151).
- Wang, D., Ge, J., Hu, H., Luo, B., & Huang, L. (2012). Discovering process models from event multiset. *Expert Systems with Applications*, 39(15), 11970–11978.
- Weijters, A., & Ribeiro, J. (2011). Flexible heuristics miner (FHM). In: *Proceedings of Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on*, (pp. 310-317).
- Weijters, A., van Der Aalst, W., & Alves De Medeiros, A. A. (2006). Process mining with the heuristics miner-algorithm. *Tech.Rep.WP, 166*, 1–34.
- Wen, L., van der Aalst, W., Wang, J., & Sun, J. (2007). Mining process models with non-free-choice constructs. *Data Mining and Knowledge Discovery*, 15(2), 145-180.
- Wen, L., Wang, J., van der Aalst, W., Huang, B., & Sun, J. (2009). A novel approach for process mining based on event types. *Journal of Intelligent Information Systems*, 32(2), 163–190.
- Wen, L., Wang, J., van der Aalst, W., Huang, B., & Sun, J. (2010). Mining process models with prime invisible tasks. *Data & Knowledge Engineering*, 69(10), 999–1021.