

Minería de Procesos aplicada a la distribución de productos postales

Victor Martinez¹, Laura Lanzarini^{1,2}, and Franco Ronchetti^{1,2}

¹ *Facultad de Informática, Universidad Nacional de La Plata*

² *Instituto de Investigación en Informática LIDI (UNLP-CIC)*

`martinezvictor@hotmail.com, {laural, fronchetti}@lidiinfo.unlp.edu.ar`

Abstract. The abstract should summarize the contents of the paper and should contain at least 70 and at most 150 words. It should be written using the *abstract* environment.

Keywords: Process Mining, Data Mining, Distribución Postal, Procesos Postales, Business Process Management

1 Introduction

Centrar la explicación en la mejor de los procesos que utilizan eventos. Finalizar la sección diciendo que la Minería de Procesos permite modelarlos e identificar subprocesos o partes relevantes.

Las técnicas de Minería de Procesos permiten extraer conocimiento a partir de registros de eventos de los sistemas de información actuales. Con estas técnicas se puede descubrir, monitorear y mejorar procesos.

En el manifiesto de la Minería de Procesos generado por la IEEE Task Force on ProcessMining [5] se expresa que la minería de procesos es una disciplina de investigación relativamente joven que se ubica entre la Inteligencia Computacional y la Minería de Datos, por un lado, y el modelado y análisis de procesos por otro. La idea de la minería de procesos es descubrir, monitorear y mejorar los procesos reales a través de la extracción de conocimiento de los registros de eventos ampliamente disponibles en los actuales sistemas de información.

La minería de procesos incluye el descubrimiento automático de procesos, la verificación de conformidad (cuanto se cumple un proceso dado), la construcción automática de modelos de simulación, y la predicción de casos, entre otras cosas.

El punto de partida de la Minería de Procesos es el registro de eventos, se asume que se registran una serie de eventos secuenciales pertenecientes a una actividad y que se relacionan con un caso particular. Se pueden almacenar datos adicionales de cada evento y se podrán utilizar para diferentes tipos de análisis (por ejemplo la persona que realiza la actividad, si la actividad está completa o no, fecha de inicio y fecha de fin), pero principalmente debe tener la fecha de evento, el identificador del caso y el tipo de evento.

En el proceso de distribución postal se lleva un registro de todas las actividades, desde el ingreso del producto hasta su entrega al cliente. Estas actividades

pueden ser ingreso a oficina comercial o almacén e intentos de entrega, entre otras. Se deben cumplir en un orden y un intervalo de tiempo específico para brindar un servicio de calidad a los clientes.

A través de la minería de procesos se puede construir un modelo que permita identificar patrones, desvíos y cuellos de botella en el proceso de distribución postal. Estos pueden ocurrir debido a las características propias de la distribución, o debido a errores humanos estos producen retrasos o desvíos operativos que impactan en la calidad del servicio. Por tal motivo, se buscará detectar desvíos o retrasos en dicho proceso de distribución con el objetivo de sugerir posibles mejoras.

2 Trabajos Relacionados

- Research of Postal Data mining system based on big data[1] paper publicado en 3rd International Conference on Mechatronics, Robotics and Automation sobre la utilización de minería de datos y big data en el correo chino.
- Context Aware Process Mining in Logistics[2] artículo publicado en la The 50th CIRP Conference on Manufacturing Systems acerca de la implementación de process mining en logística.
- PM2: a Process Mining Project Methodology[4] paper acerca de la metodología de ejecución de process mining en un proyecto.
- Reinventing the Postal Sector in an Electronic Age[3] Libro donde se plantea la necesidad de reinención del negocio de correo con el advenimiento de las nuevas tecnologías.

3 Técnicas de Process Mining

El objetivo de la minería de procesos es descubrir, monitorear y mejorar los procesos reales a través de la extracción de conocimiento de los registros de eventos dejados por los sistemas de información actuales. La información debe estar almacenada con ciertas características que permitan su extracción y análisis, por ejemplo identificador de la traza, identificador del evento, timestamp del momento en que ocurre adicionalmente puede contener información de la completitud de la traza, la persona que realizó el evento o el timestamp de la fecha de finalización para más información.

Con la minería de procesos se puede realizar un análisis para responder las siguientes preguntas de un proceso real: ¿Qué fue lo que pasó realmente?, ¿Por qué pasó?, ¿Qué es lo que podría pasar en el futuro?, ¿Cómo se puede mejorar el control del proceso?, ¿Cómo se puede rediseñar el proceso para mejorar la performance?[6]

Existen tres tipos de minería de procesos:

- *Descubrimiento*: Se busca obtener automáticamente el modelo de negocio a través de los datos almacenados en el log de eventos sin ninguna información previa. Existen diferentes algoritmos que permiten realizar esta

tarea, los más comunes son el algoritmo Alpha [8], Heuristics Miner [9] y el Inductive Miner [10].

El proceso descubierto puede exportarse a una Red de Petri o a un modelo BPMN (Business Process Model and Notation). La calidad del proceso descubierto depende de la muestra utilizada, no se puede descubrir partes del proceso que no se encuentren en la muestra.

- *Verificación de Conformidad*: en contrapartida al descubrimiento, se compara un modelo existente (puede ser el descubierto anteriormente u otro) con el log de procesos para encontrar desvíos y verificar el funcionamiento. Permite obtener representaciones gráficas y animaciones que posibilitan observar el comportamiento real y ver cuanto se ajusta la realidad al proceso definido originalmente.
Su principal ventaja es que muestra la realidad, no es una simulación, lo que permite realizar un análisis mucho más preciso.
- *Mejoramiento*: Se busca extender o mejorar el proceso existente a través de la información almacenada en el log de eventos. A diferencia de la Verificación de Conformidad donde se comparan los datos con el modelo, aquí se busca modificar el proceso [5]

Se debe variar el grado de abstracción que se utilizará durante el análisis, con el objetivo de establecer un balance entre cuatro fuerzas para encontrar los mejores resultados.

Estas fuerzas son *Ajuste* (habilidad para explicar el comportamiento observado), *precisión* (evitando sobreajustar el modelo con un excesivo nivel de detalle), *generalización* (evitando un nivel de abstracción muy alto) y *simplicidad* (logrando una mejor comprensión del modelo)[6].

Finalmente es importante destacar que el resultado obtenido del análisis con minería de procesos está altamente vinculado a la calidad de los datos de entrada. De hecho se sabe que siempre hay una cantidad de ruido en los datos, el ruido pueden deberse a trazas incompletas, intervalos que no se hayan registrado correctamente o duplicidad de datos. Esta es información que puede distorsionar o falsificar el resultado del análisis.[7]

Algunos algoritmos son muy sensibles al ruido en los datos. Por eso es necesario verificar los datos de entrada y realizar un procesamiento previo para eliminar la mayor cantidad de ruido posible

4 Proceso de distribución postal

El proceso de distribución postal abarca diferentes tipos de productos, estos pueden ser paquetería (encomiendas o ecommerce), productos postales tradicionales (cartas, telegramas, cartas documento), productos de alta seguridad (tarjetas de crédito, medicamentos), entre otros.

En todos los casos el proceso trata de una recepción del producto del remitente (ya sea por un medio físico o un canal digital) y una entrega al destinatario, pudiendo esta ser exitosa o no. La no entrega del producto no implica un error en el proceso, puede haber varios motivos para ello (es probable

que el remitente no se encuentre, haya muerto o rechace el envío, entre otros)

En general el proceso se compone de los siguientes pasos:

1. Recepción: Se recibe el producto a través de uno de los canales habilitados.
2. Identificación: asignación de un identificador único de envío.
3. Ingreso a centro de procesamiento: se registra el ingreso del producto y se enruta hacia el destino
4. Envío a centro de distribución: Se envía a centro de distribución de destino para su entrega
5. Intento de entrega: se realizan uno o más intentos de entrega dependiendo del producto.
6. Espera en centro de distribución: según el tipo de producto se puede quedar un tiempo esperando a que el destinatario venga a retirarlo.
7. Confirmación de entrega - no entrega y devolución: confirmación de entrega - no entrega al remitente y devolución si corresponde.

En todos los casos se deja registro de cada uno de los pasos realizados asociados a un identificador único de envío, lo que permite conocer el estado actual del mismo en informarlo al cliente. Los problemas típicos que se pueden dar son envío a centro de distribución erróneo, demoras en un paso determinado, no cumplimiento de los intentos de entrega o tiempo de espera definidos para el producto, entre otras cosas.

5 Aplicación de la Minería de Procesos a la distribución Postal

Se aplicarán a continuación las técnicas de minería de procesos a la distribución postal para descubrir el proceso real que se utiliza y evaluar su funcionamiento.

Para este caso práctico se utilizó un producto que tiene dos intentos de entrega en el domicilio, si no se puede entregar se guarda un tiempo esperando que el destinatario venga a retirar, transcurrido ese tiempo se realiza la devolución al remitente. Todos los envíos que no se pueden entregar se devuelven al remitente. Se define una traza cada traza como un envío. Cada movimiento que se registre de ese envío es un evento. Se considera una traza completa cuando el envío tiene registrado su ingreso y su fin, con entrega exitosa o no.

5.1 Extracción de los datos

Se realiza la extracción primaria de los datos de un producto que contempla dos intentos de entrega y guarda para su retiro antes de hacer una devolución. Se obtienen los campos mínimos necesarios para el análisis, se obtiene identificador de traza, identificador de evento, fecha de evento, descripción del evento.

Se busca obtener una muestra heterogénea que sea lo más representativa posible y contenga todos los pasos del proceso.

	Identificador de evento		Identificador de traza	Fecha evento	
	123 trazalD	123 EveID	ABC eveDescrip	eveFecha	
31	481,053	0	INGRESADO	2017-08-16 10:45:22	
32	481,053	2	1 INTENTO DE ENTREGA	2017-08-18 11:15:00	
33	481,053	9	DEVOLUCION	2017-08-18 13:00:00	
34	481,054	0	INGRESADO	2017-08-16 10:45:28	
35	481,054	2	1 INTENTO DE ENTREGA	2017-08-22 12:05:00	
36	481,054	9	DEVOLUCION	2017-08-22 17:25:00	
37	481,055	0	INGRESADO	2017-08-16 10:45:27	
38	481,055	1	ENTREGADO	2017-08-22 15:13:00	

Fig. 1. Ejemplo de la muestra extraída para el análisis

5.2 Transformación a process log

Se almacenan los datos de la muestra en el formato estándar XES ¹ para agilizar su procesamiento. Este formato es el utilizado por la mayoría de los algoritmos de descubrimiento utilizados en la minería de procesos, teniendo la muestra a priori se reduce considerablemente el tiempo de procesamiento en las iteraciones del descubrimiento. **colocar tamaño muestra antes y después de filtrar**

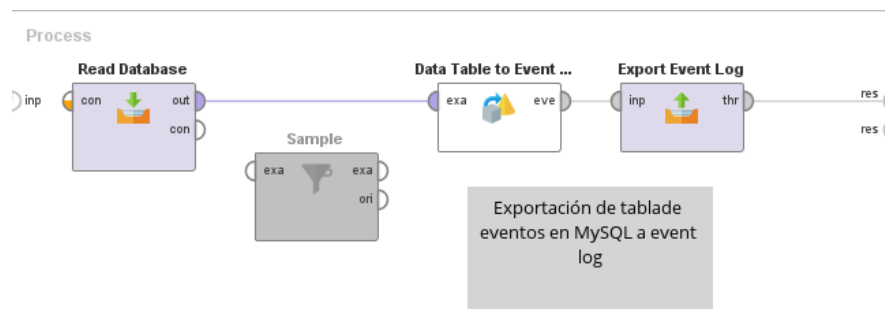


Fig. 2. Convertir datos de BD a formato XES

5.3 Filtrado del log

Se quitan de la muestra todas las trazas que estén incompletas y todas las trazas que no respeten el negocio para descubrir el proceso correcto. Se define como traza incompleta aquella que no tenga un evento de inicio (Recepción), por algún error de carga y las que no tengan un evento final (entrega, devolución, destinatario inexistente, fallecido, rechaza el envío o se mudó), ya sea por error

¹ <http://www.xes-standard.org/>

de carga o porque todavía no transcurrió el tiempo estipulado para finalizar el proceso.

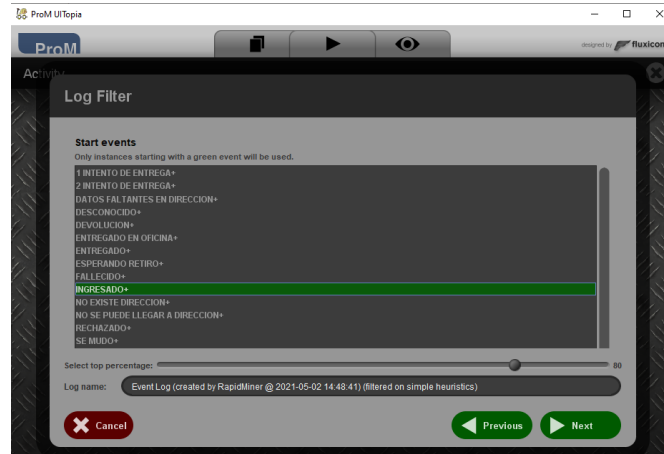


Fig. 3. Filtrar eventos incompletos - Inicio

Se conservan solo las que tengan un estado inicial válido y los estados finales esperados (solo aquellas donde el proceso se cumplió correctamente) Se utiliza el filtro de reglas heurísticas simples sobre el log para establecer cuales son las trazas completas, se dejan solo aquellas que tengan un estado inicial "Ingresado" y las que posean los estados finales válidos.

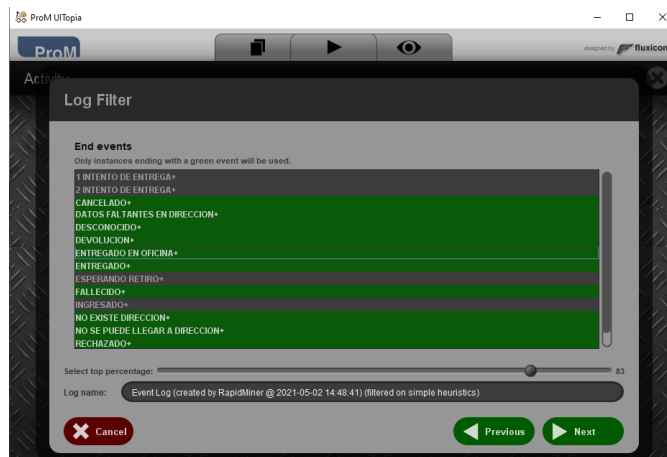


Fig. 4. Filtrar eventos incompletos - Fin

5.4 Descubrimiento del proceso real

Se realizan diferentes iteraciones con la muestra de datos para encontrar el proceso que mejor describe el negocio. De las pruebas con el algoritmo Alpha se obtiene el siguiente proceso:

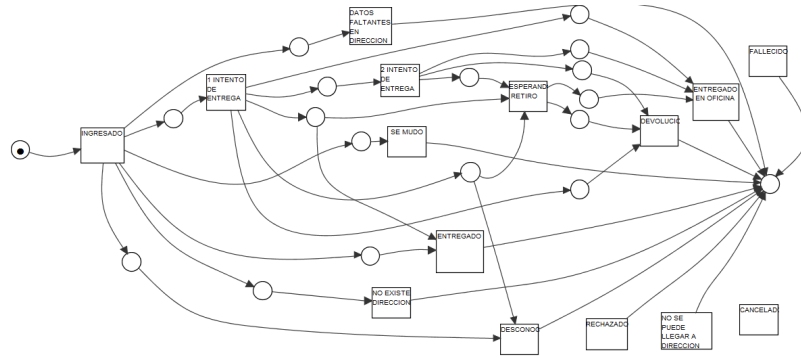


Fig. 5. Proceso descubierto con el algoritmo Alpha

De las pruebas utilizando el algoritmo Heuristics Miner se obtiene el siguiente proceso:

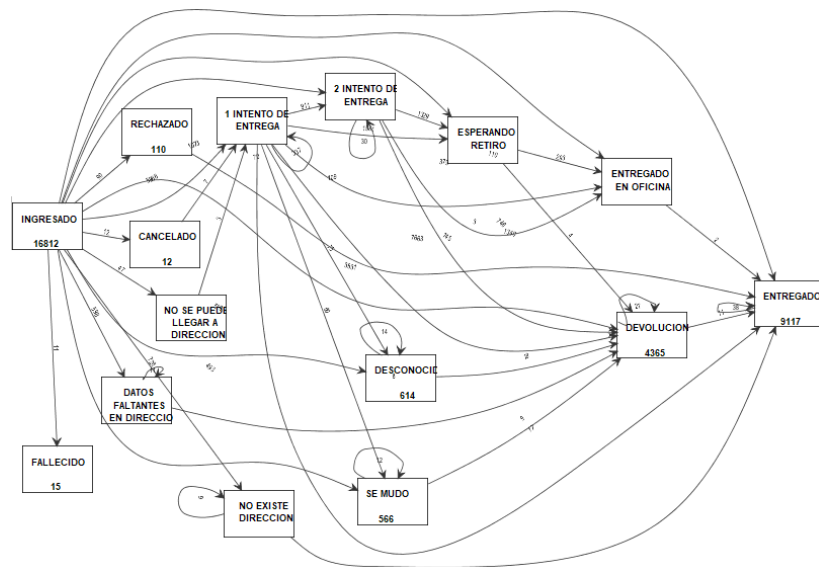


Fig. 6. Proceso descubierto con el algoritmo Heuristics Miner

Finalmente se descubrió el proceso con el algoritmo inductive miner

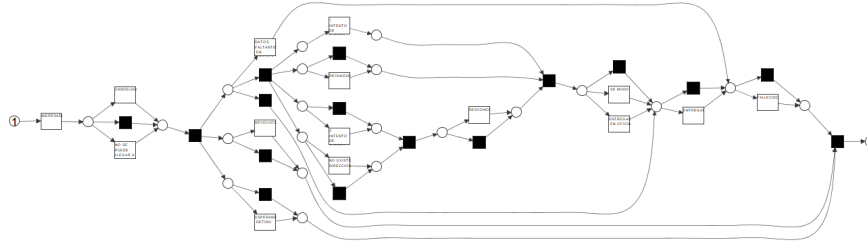


Fig. 7. Proceso descubierto con el algoritmo Inductive Miner

Luego de evaluar los diferentes modelos descubiertos con los diferentes algoritmos evaluados se decide utilizar el algoritmo **XXXXXX** por ser el que más se ajusta al modelo diseñado originalmente.

5.5 Verificación de conformidad

Se realiza la verificación de conoformidad, para ello se compara un set que contiene más de **30.000 Verificar!** trazas contra el proceso descubierto en el paso anterior. En primera instancia se realiza una representación visual de los casos con utilizado el algoritmo *Inductive Visual Miner* con este se observa una animación que muestra como se suceden los eventos de cada traza en una línea de tiempo, cada traza está simbolizada con un punto amarillo que va recorriendo las diferentes etapas del proceso.

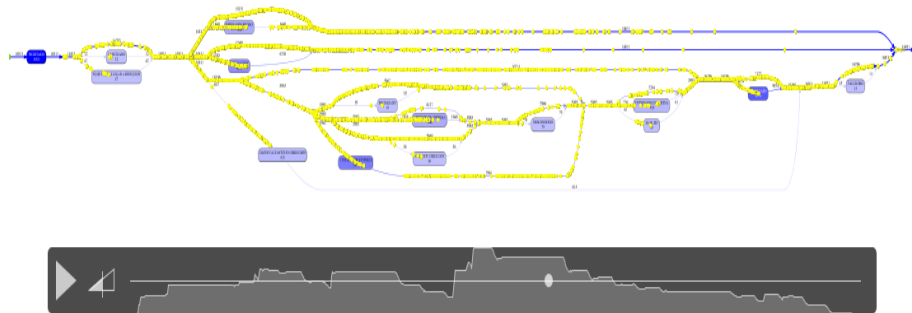


Fig. 8. Comparación de los casos contra el proceso

Se observa en el detalle la representación de cada caso.

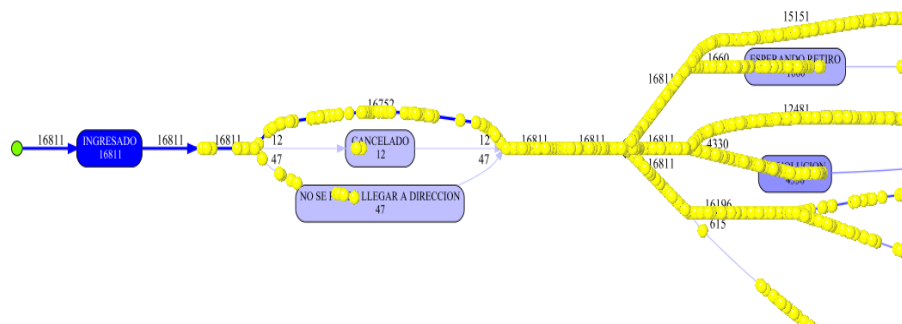


Fig. 9. Detalle de los casos contra el proceso

Luego se realizan verificaciones de los datos con el modelos con **xx,yy**

Finalmente se extraen del log los casos que se desvían en un **XX por ciento** del modelo descubierto para su posterior análisis en busca de mejorar conflictos de la distribución

6 Conclusión y futuras líneas de investigación

Se verifica que se pueden utilizar las técnicas de minería de procesos para el análisis y solución de problemas en el proceso de distribución postal.

- Continuar con las técnicas de minería de procesos para que permitan sugerir mejoras al proceso real.

- Establecer alertas automatizadas que se disparen ante casos que se desvíen en forma excesiva del modelo.
- Generar herramienta que permita análisis en tiempo real y no sobre casos finalizados.

References

1. Research of Postal Data mining system based on big data, Xia Hul; Yanfeng Jinl; Fan Wang, 3rd International Conference on Mechatronics, Robotics and Automation, 2015, 10.2991/icmra-15.2015.124, https://www.researchgate.net/publication/300483008_Research_of_Postal_Data_mining_system_based_on_big_data
2. Context Aware Process Mining in Logistics, Mitchell M. Tseng; Hung-Yin Tsai; Yue Wang, 2017, The 50th CIRP Conference on Manufacturing Systems, <https://www.sciencedirect.com/science/article/pii/S2212827117303311>
3. Reinventing the Postal Sector in an Electronic Age, Michael A. Crew; Paul R. Kleindorfer, 978-1849803601, 2011, Edward Elgar Publishing
4. PM2: a Process Mining Project Methodology, Maikel L. van Eck; Xixi Lu; Sander J.J. Leemans; Wil M.P. van der Aalst, Eindhoven University of Technology, The Netherlands, http://www.processmining.org/_media/blogs/pub2015/pm2_processminingprojectmethodology.pdf
5. Wil van der Aalst, The Process Mining Manifesto by the IEEE Task Force, 2012, <https://www.tf-pm.org/resources/manifesto>
6. Process Mining: Data Science in Action, Wil van der Aalst, 978-3-662-49850-7, 2016, Springer
7. Process mining in flexible environments, Christian Walter Gunther, 978-90-386-1964-4, 2009, Technische Universiteit Eindhoven, <https://research.tue.nl/en/publications/process-mining-in-flexible-environments>
8. Workflow mining: discovering process models from event logs, W. van der Aalst; T. Weijters; L. Maruster, 1041-4347, 2004, IEEE, <https://ieeexplore.ieee.org/document/1316839>
9. Flexible Heuristics Miner, A.J.M.M. Weijters; J.T.S. Ribeiro, 978-1-4244-9926-7, 2011, IEEE, <https://ieeexplore.ieee.org/document/5949453>
10. Discovering Block-Structured Process Models from Event Logs - A Constructive Approach, Sander J. J. Leemans Dirk Fahland Wil M. P. van der Aalst, 978-3-642-38696-1, 2013, Springer, https://link.springer.com/chapter/10.1007/978-3-642-38697-8_17