

Analysis of Business Processes Variants

Damián Pérez-Alfonso, MSc.¹, Eudel Pupo Hernández, Ing.¹, and Juan Pedro Febles Rodríguez, PhD.¹

¹ Universidad de las Ciencias Informáticas, Cuba, dalfonso@uci.cu, epupoh@uci.cu, febles@uci.cu

Abstract— Process Mining research is a discipline that combines elements of machine learning, data mining and analysis and process modeling. The objective of this discipline is to extract knowledge from event logs generated during the execution of the processes information systems. The diagnosis is the phase that closes the lifecycle of the business process management with the analysis of the implementation thereof. At this stage the use of process mining techniques to identify patterns and contexts of execution, and deviations abnormalities is recommended. The models obtained by these techniques must balance the level of detail of the patterns that have the simplicity and accuracy of the model with respect to the event log. Therefore, in processes with variability in execution techniques available often offer dense or incomplete models, which is difficult to achieve the objectives of diagnosis. It proposes a new technique for diagnosis that allows for variations process models for the analysis of multiple alternatives run through a single representation. The application Technical Proposal using event logs of a hospital setting process evidence for its usefulness diagnosis.

Keywords— diagnosis process, mining processes, process management, model line

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2015.1.1.160>

ISBN: 13 978-0-9822896-8-6

ISSN: 2414-6668

13th LACCEI Annual International Conference: “Engineering Education Facing the Grand Challenges, What Are We Doing?”
July 29-31, 2015, Santo Domingo, Dominican Republic

ISBN: 13 978-0-9822896-8-6

ISSN: 2414-6668

DOI: <http://dx.doi.org/10.18687/LACCEI2015.1.1.160>

Análisis de Variantes de Procesos de Negocio

Damián Pérez-Alfonso, MSc., Eudel Pupo-Hernández, Ing., Juan Pedro Febles-Rodríguez, PhD

Universidad de las Ciencias Informáticas, La Habana, Cuba

dalfonso@uci.cu, epupoh@uci.cu, febles@uci.cu

Resumen—La minería de procesos es una disciplina de investigación que combina elementos del aprendizaje automático, la minería de datos y el análisis y modelado de procesos. El objetivo de esta disciplina es extraer conocimiento de los registros de eventos generados durante la ejecución de los procesos en los sistemas de información. El diagnóstico es la fase que cierra el ciclo de vida de la gestión de procesos de negocio con el análisis de la ejecución de los mismos. En esta fase se recomienda la utilización de técnicas de minería de proceso que permitan identificar patrones y contextos de ejecución, anomalías y desviaciones. Los modelos obtenidos por estas técnicas deben balancear el nivel de detalle de los patrones que presentan, la simplicidad y la exactitud del modelo respecto al registro de eventos. Por ende, en procesos con una variabilidad en su ejecución las técnicas disponibles suelen ofrecer modelos densos o incompletos, con los cuales resulta difícil lograr los objetivos del diagnóstico. Se propone una nueva técnica para el diagnóstico que permite obtener variantes de modelos del proceso, para el análisis de múltiples alternativas de ejecución a través de una representación única. La aplicación de la técnica propuesta utilizando registros de eventos de un proceso del entorno hospitalario evidencia su utilidad para el diagnóstico.

Keywords—diagnóstico de procesos, minería de procesos, gestión de procesos, variantes de modelos

I. INTRODUCCIÓN

La minería de procesos es una disciplina dedicada al análisis de los procesos, partiendo de sus trazas de ejecución en los sistemas de información [1]. Al conjunto de las trazas de un proceso se le denomina registro de eventos, ya que cada traza contiene una secuencia de eventos ordenados según su ocurrencia. Estos eventos reflejan la ejecución de las actividades del proceso al cual pertenece la traza [2].

El análisis automático de la información contenida en los registros de eventos, permite a las técnicas de minería de procesos ofrecer resultados en menor tiempo y con mayor fiabilidad que las técnicas tradicionales de análisis y mejora de procesos [3]. Debido a esto, dentro del área de gestión de procesos de negocio, ha aumentado sostenidamente el interés en la minería de procesos en la última década [4]. Por lo cual importantes proveedores de software, universidades y empresas consultoras se han aliado junto al IEEE para la investigación y el desarrollo de esta área [5].

La minería de procesos tiene como objetivo el diagnóstico de procesos, su descubrimiento y el chequeo de conformidad entre la ejecución real y sus modelos de referencia. Ha demostrado su utilidad en diferentes etapas de la gestión de procesos de negocio, incluyendo el diagnóstico, la ejecución, el ajuste y la (re)configuración. El diagnóstico del proceso ayuda a tener una visión general del proceso, de los aspectos más significativos del mismo y de las técnicas que pueden ser más útiles en su posterior análisis [6]. Esta etapa de la gestión

de procesos de negocio abarca el análisis de rendimiento, la detección de anomalías, la identificación de patrones comunes y su ubicación en el contexto de ejecución del proceso [7].

Se han desarrollado un conjunto de técnicas de minería de procesos útiles en la etapa de diagnóstico [8], [7], [6], [9]. Sin embargo, dichas técnicas suelen ser afectadas por el ruido, la ausencia de información, el número de tareas involucradas, las dimensiones y la complejidad del registro de eventos [10]. El ruido y la ausencia de información son características de los procesos cuya ejecución varía para distintas instancias del proceso. El 80 % de la variabilidad de un proceso suele encontrarse en menos del 20 % de las trazas del registro de eventos [1]. Aunque las técnicas de minería de proceso suelen descartar este comportamiento poco frecuente, en él se encuentran anomalías y desviaciones que son omitidas de los modelos resultantes. Por tanto dichos modelos son incompletos desde el punto de vista de los objetivos a lograr en el diagnóstico. En contraposición los modelos que reflejan todo el comportamiento suelen ser densos ya que no jerarquizan los comportamientos adecuadamente, dificultando la comprensión del proceso.

En la sección siguiente se muestra un análisis de las técnicas disponibles para el diagnóstico desde la perspectiva de su utilidad para la comprensión de los procesos con variabilidad en su ejecución. Partiendo de dicho análisis se presenta en la sección III Variants Miner, una nueva técnica que tiene como objetivo mejorar la comprensión de los procesos durante la etapa de diagnóstico, a partir del descubrimiento de múltiples variantes de modelos del proceso integradas una representación única. Los resultados de la aplicación de Variants Miner en un entorno real se muestran en la sección IV. Dichos resultados evidencian que la técnica propuesta permite identificar anomalías y desviaciones aún en procesos con variabilidad en su ejecución.

II. TÉCNICAS PARA EL DIAGNÓSTICO DE PROCESOS

La comprensión de un proceso pasa por la comprensión de sus representaciones, es decir, por la comprensión del o de los modelos que lo representan. Existen factores que afectan el entendimiento de los modelos de procesos tales como: notación de modelado [11], presentación visual [12] y complejidad del modelo de proceso [13].

La comprensión del modelo debe ser entendida como una escalera [14]: la sintaxis (cómo se combinan los elementos gramaticales en un modelo de proceso) debe estar clara antes de que la semántica pueda ser discutida y la semántica (qué significan los elementos gramaticales del modelo de proceso) debe estar clara antes de discutir los elementos pragmáticos [15]. Para facilitar la comprensión sintáctica debe considerarse la carga cognitiva extrínseca, la cual es determinada por la

forma en la que la información es representada [16]. Por tanto, lograr la comprensión del proceso en la fase de diagnóstico requiere de: la representación adecuada de los aspectos relevantes para el diagnóstico (sintaxis y carga cognitiva), la veracidad de dichos aspectos (semántica) y que los mismos conduzcan a interpretaciones y acciones prácticas (pragmática).

Con el objetivo de extraer y mostrar información útil para el diagnóstico de procesos se han desarrollado un conjunto de técnicas dentro del marco de trabajo para la minería de procesos ProM [2]. ProM cuenta con más de 600 complementos, cada uno de los cuales posibilita realizar diferentes análisis y visualizaciones [17]. Incluye además facilidades para la exportación, importación y visualización de modelos de procesos en múltiples formatos como Redes de Petri, BPMN, Redes Heurísticas, YAWL, Redes Causales, Mapas de Proceso, Árboles de Proceso, entre otros. A continuación se analizan las principales características de estas técnicas.

El *Análisis de Diagramas de Puntos* [18] brinda una visión general del registro de eventos, lo que facilita la identificación de los aspectos más significativos del proceso. Esta técnica presenta una “vista helicóptero” del registro de eventos desde la perspectiva temporal, ya que muestra los eventos del proceso como puntos en un plano, donde una dimensión representa las trazas y la otra se refiere al instante en el que se ejecutaron los eventos. Al no poseer mecanismos de abstracción y jerarquización los modelos que obtiene en procesos con variabilidad en su ejecución suelen ser densos, limitando la comprensión del mismo por su alta carga cognitiva.

Una técnica similar a la anterior es la *Alineación de Trazas* [7], basada en la Alineación de Múltiples Secuencias, una herramienta empleada en la bioinformática. Esta técnica presenta una alineación de las diferentes secuencias de actividades que se encuentran en las trazas, posibilitando la identificación de desviaciones en la ejecución del proceso. Sin embargo, esta técnica no permite detectar los subprocesos que conforman al proceso analizado para enmarcar en estos las anomalías y patrones identificados. Esto dificulta la ubicación en contexto del aspecto detectado y el entendimiento de las causas que lo originaron [6]. Por otra parte, la presencia de ruido en el registro de eventos puede conducir a la obtención de alineaciones con baja calidad [7] lo cual se traduce en una representación densa de las alineaciones.

La *Descomposición en Bloques de Construcción* se basa en la *Alineación de Trazas* para la construcción de una visión jerárquica del proceso, descomponiendo en subprocesos cada nivel según el patrón de control de flujo predominante [6]. Esta técnica considera en su descomposición ocho situaciones diferentes donde la ausencia de información genera un impacto estructural en el modelo [19]. Sin embargo, al no considerar el ruido en la identificación de los patrones de control de flujo puede obtener una descomposición que no representa adecuadamente el proceso.

La técnica que mejor reacciona antes registros de eventos con ruido es la *Minería Difusa* [9], ya que permite descartar el ruido del modelo obtenido, usando umbrales de frecuencia configurables por el usuario. Además utiliza similitudes con la forma de representar la información en la cartografía, aplicando conceptos como: abstracción, personalización, agregación

y énfasis [1]. Dicha técnica produce un modelo basado en grafos, donde existen dos tipos de nodos, los que representan una actividad y los que agrupan un conjunto de actividades denominados clústeres. Sin embargo este modelo no ofrece patrones de control de flujo y por ende no permite la identificación de subprocesos.

Otra enfoque es descubrir un modelo que represente el comportamiento común del proceso y luego identificar las desviaciones mediante el chequeo de conformidad entre el modelo descubierto y el registro de eventos. En este ámbito destaca el *Inductive Visual Miner (IvM)* [20], una técnica para la exploración de procesos. *IvM* combina el algoritmo de descubrimiento Inductive Miner [21] y la alineación entre el modelo y el registro de eventos [22] para obtener un modelo de proceso y sus desviaciones. Al igual que *Fuzzy Miner*, ofrece parámetros para configurar las actividades y relaciones a incluir en el modelo, a partir de su frecuencia de aparición en el registro de eventos.

Como el modelo obtenido por *IvM* debe balancear cuatro dimensiones de calidad: *Aptitud*, *Precisión*, *Generalización* y *Simplicidad* [1] no puede representar las anomalías del proceso. Para suplir este elemento se enriquece el modelo con las desviaciones identificadas a partir de los saltos realizados en la alineación, que pueden ser en el modelo (actividades omitidas) o en el registro de eventos (eventos omitidos). Las actividades omitidas son representadas mediante aristas discontinuas en el modelo que simbolizan caminos alternativos en el proceso. Mientras que los eventos omitidos se representan como lazos de líneas discontinuas a continuación de la última actividad del modelo que coincide con el registro de eventos. En este último caso no es posible conocer cuáles son los eventos omitidos ni su impacto en la estructura del proceso.

Las técnicas existentes suelen descartar el ruido presente en el registro de eventos o no identificarlo. Ambas alternativas tienen un impacto estructural sobre el modelo obtenido ya que pueden influir en el tipo de patrón de control de flujo identificado o sugerido por la técnica, en las secciones del registro de eventos donde está presente el ruido. Aún cuando el analista de procesos suele tener la posibilidad de establecer los umbrales de frecuencia, como en la *Minería Difusa* o en el *IvM*, no posee control sobre el impacto estructural de los comportamientos descartados en el modelo del proceso.

Para resolver la afectación que provocan el ruido y la ausencia de información en la comprensión del proceso es necesario controlar el impacto estructural de estas características. Sin embargo, considerar ciertos comportamientos infrecuentes como ruido y/o asumir que faltan determinadas evidencias de la ejecución del proceso, son estimaciones que sólo pueden ser confirmadas a partir del contexto de ejecución particular del proceso analizado. Por tanto se propone la construcción de varios modelos de procesos integrados en una única representación y con información sobre los comportamientos descartados y asumidos en la construcción de cada modelo. De esta manera el analista de procesos puede decidir en función del contexto cuáles son las anomalías y desviaciones, así como identificar su impacto en la estructura de control de flujo del proceso. El enfoque anteriormente planteado ha sido implementado en la técnica *Variants Miner*, la cual se describe en la sección siguiente.

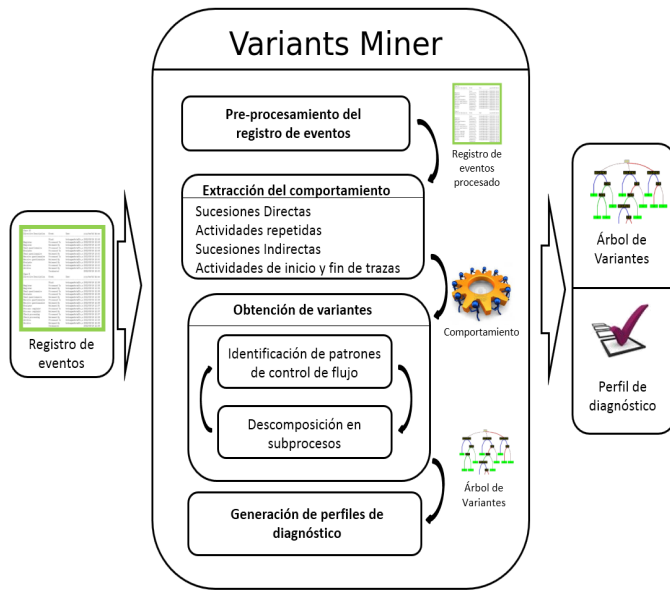


Figura 1. Esquema representativo de la técnica Variants Miner.

III. VARIANTS MINER

La Figura 1 muestra una representación de la técnica propuesta, nombrada *Variants Miner*. Está conformada por cuatro etapas que se ejecutan secuencialmente y al igual que el resto de las técnicas para el diagnóstico analizadas, utiliza como entrada un registro de eventos.

El objetivo perseguido por *Variants Miner* es construir variantes de proceso a partir de diferentes descomposiciones en subprocesos considerando patrones de control de flujo. Las variantes de descomposición obtenidas se muestran en un *Árbol de Variantes*, asociado al cual se obtiene un *Perfil de Diagnóstico*. *Variants Miner* ha sido implementada como complemento del marco de trabajo ProM. Utiliza como entrada un registro de eventos y genera como salida un *Árbol de variantes*, con su respectivo *Perfil de Diagnóstico*.

Para comprender el *Árbol de Variantes* que ofrece la técnica como resultado, es relevante conocer que la descomposición que se realiza en la técnica se basa en la Definición 3.1.

DEFINICIÓN 3.1: *Descomposición en subprocesos:* Se denota con S_P al conjunto de los subprocesos del proceso P , y por $W = \{w_s, w_x, w_o, w_p, w_l\}$ se denota al conjunto de los patrones de control de flujo. Un subproceso $s_i \in S_P$ se puede descomponer a través de diferentes patrones de control de flujo, $w_j \in W$, en varios subprocesos, $s_k \in S_P : k \neq i$, hasta el nivel de actividad.

La solución propuesta sigue un enfoque diferente al de otras técnicas de diagnóstico, al proponer varias descomposiciones alternativas para el mismo subproceso, utilizando diferentes patrones de control de flujo. Esto permite controlar el impacto estructural del ruido y la ausencia de información en la construcción de las alternativas. Las alternativas se construyen, descartando o no, determinados comportamientos presentes en el registro de eventos. También se pueden construir, considerando o no, determinados comportamientos

ausentes. Las diferentes alternativas de descomposición que pueden existir en cada subproceso conforman variantes del proceso [23], concepción expuesta en la Definición 3.2.

DEFINICIÓN 3.2: *Variantes de proceso:* Las variantes de un modelo de procesos o variantes de proceso, son modelos de un proceso que describen el mismo proceso de negocio, y poseen algunas diferencias estructurales. Las diferencias están dadas por los patrones de control de flujo que se utilizan en secciones equivalentes del proceso y la presencia de determinadas actividades.

Los elementos antes planteados permiten que la técnica ofrezca un modelo jerárquico en forma de árbol, que representa diferentes variantes del proceso cuya ejecución está contenida en el registro de eventos. Para esto se propone un modelo de proceso denominado *Árbol de Variantes*, formalizado en la Definición 3.3. Este árbol está compuesto por dos tipos de nodos, los nodos subproceso y los nodos patrón. Un nodo subproceso representa un subproceso y posee tantos nodos patrones de hijos, como posibles descomposiciones se hayan identificado para el subproceso. Un nodo patrón representa una descomposición de su padre, de acuerdo a un patrón de control de flujo, por lo que un nodo patrón posee dos o más nodos subproceso como hijos. El nodo raíz es un nodo subproceso y se refiere a todo el proceso. Los nodos hojas son siempre nodos de tipo subproceso.

DEFINICIÓN 3.3: *Árbol de variantes:* Un árbol de variantes (V_P) es una representación de diferentes descomposiciones en subprocesos aplicadas al proceso P a partir de un registro de eventos L_P .

Denotando a N_V como el conjunto de todos los nodos de V_P y por $\Gamma_V \subset N_V$ el conjunto de todos los nodos hoja. Siendo $n_{s_i} \in N_V : 0 < i \leq |S_P|$, un nodo subproceso que representa al subproceso $s_i \in S_P$ y l_i la sección del registro de eventos asociada a s_i . Siendo $n_{iw_j} \in N_V \wedge n_{iw_j} \notin \Gamma_V : 1 < j \leq |W|$, un nodo patrón que representa una descomposición del subproceso de acuerdo a cierto patrón de control de flujo $w_j \in W$.

El nodo raíz de V_P es n_{s_1} ya que $s_1 \in S_P \wedge s_1 = P$. Un nodo subproceso n_{s_i} puede tener ninguno o varios nodos hijos n_{iw_j} . Un nodo n_{iw_j} tiene dos o más nodos hijo n_{s_i} .

En la Figura 2 se muestran las variantes de descomposición para un mismo subproceso, que forman parte de un *Árbol de variantes*. La primera variante corresponde al patrón *Selección no Exclusiva*. La segunda variante representa una descomposición por el patrón de control de flujo *Secuencia*. Por último, se muestra la descomposición del subproceso utilizando el patrón *Lazo*. Los nodos hojas representan las actividades que conforman el proceso. Los nodos patrón, muestran el nombre del patrón de control de flujo utilizado para descomponer el subproceso en la variante correspondiente. Cada nivel en el árbol, de la raíz a las hojas, representa un nivel de abstracción inferior.

Asociado al *Árbol de Variantes* se propone un *Perfil de Diagnóstico* (Definición 3.4), el cual agrupa la información de diagnóstico asociada a cada nodo de tipo patrón del árbol.

DEFINICIÓN 3.4: *Perfil de Diagnóstico:* Se denota por

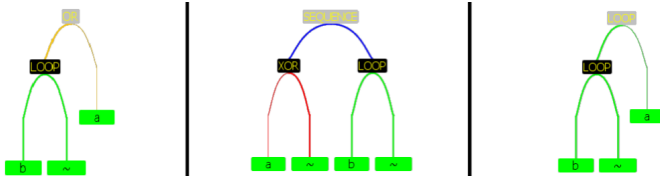


Figura 2. Variantes contenidas en un *Árbol de variantes*.

D_V al perfil de diagnóstico del árbol V_P tal que: $\forall n_{iw_j} \in N_V \exists d_{iw_j} \in D_V$.

Denotando por f a la estimación de aptitud, por ρ a la estimación de precisión, el comportamiento descartado como η_{iw_j} y el comportamiento asumido como θ_{iw_j} . La información de diagnóstico d_{iw_j} está constituida por la tupla $\{f, \rho, \eta_{iw_j}, \theta_{iw_j}\}$.

Los comportamientos η_{iw_j} y θ_{iw_j} se refieren al subproceso s_i del nodo n_{s_i} padre de n_{iw_j} .

Se denota por $\beta_{w_j l_i}$ al comportamiento del patrón w_j expresado en la sección del registro de eventos l_i . El valor de f expresa la relación entre η_{iw_j} y $\beta_{w_j l_i}$. De manera similar ρ expresa la relación entre θ_{iw_j} y $\beta_{w_j l_i}$.

En el *Perfil de Diagnóstico* se incluyen los comportamientos descartados, ya que estos pueden ser considerados como ruido, anomalías o desviaciones del proceso. A su vez, los comportamientos asumidos durante la descomposición, reflejan situaciones de ausencia de información, las cuales pueden estar asociadas también a anomalías durante la ejecución del proceso. La presentación de ambos tipos de comportamiento como resultado de la técnica, responde directamente a los intereses del diagnóstico. Se pretende, de esta manera, contribuir a la comprensión del proceso en la dimensión pragmática.

A partir de los comportamientos anteriormente descritos, se realiza una estimación de la calidad del modelo propuesto, en términos de su *Aptitud* y *Precisión*. Dicha estimación se incorpora en el *Perfil de Diagnóstico* con la intención de mejorar la comprensión del proceso, en el sentido de ofrecer información sobre la alineación que existe entre el modelo propuesto y la ejecución real del proceso.

En las secciones siguientes se presentan las etapas que componen la técnica *Variants Miner*. Se describen además, los algoritmos fundamentales que se proponen como parte de la solución.

III-A. Pre-procesamiento del registro de eventos

Esta etapa tiene como objetivo preparar el registro de eventos para el diagnóstico de variantes. El algoritmo que se ejecuta en esta etapa realiza la codificación de los eventos presentes en el registro de eventos. La codificación consiste en la asignación de una letra a cada evento contenido en el registro. Esto tiene como objetivo disminuir la carga cognitiva durante la comprensión del proceso, eliminando la información semántica de las etiquetas de los eventos. Mediante la codificación se pretende mejorar la comprensión sintáctica del modelo, antes de pasar a la comprensión semántica del mismo. Adicionalmente, la conversión de las trazas en secuencias codificadas facilita el procesamiento subsecuente.

13 th LACCEI Annual International Conference: “Engineering Education Facing the Grand Challenges, What Are We Doing?”
July 2931, 2015, Santo Domingo, Dominican Republic”

Posterior a la codificación se unifican las trazas cuyas secuencias coinciden, disminuyendo el número de trazas a procesar y agrupando los comportamientos coincidentes. Las secuencias resultantes de la agrupación son ordenadas descendientemente, por su frecuencia de aparición, para expresar su importancia relativa dentro del proceso.

Por último se extraen características generales presentes en el registro de eventos, como la cantidad de trazas diferentes y la cantidad de eventos diferentes.

III-B. Extracción de comportamiento

El propósito de esta etapa es la extracción de comportamientos representativos de los patrones de control de flujo, a partir del registro de eventos preprocesado. Para la técnica se considera el comportamiento del proceso según la Definición 3.5.

DEFINICIÓN 3.5: Comportamiento del proceso: Se denota por $l_i \subseteq L_P$ a la sección del registro de eventos relacionada con el subproceso $s_i \in S_P$. El comportamiento del patrón de control de flujo w_j para el subproceso s_i , denotado por $\beta_{w_j s_i} : w_j \in W$ está compuesto por todas las posibles:

- Sucesiones directas (*DS*), se denomina sucesión directa entre la actividad a y la actividad b , a la existencia en alguna traza de la secuencia ab .
- Sucesiones indirectas (*IS*), se denomina sucesión indirecta entre la actividad a y la actividad b , a la aparición en alguna traza de la actividad a seguida, inmediatamente o no, por la actividad b .
- Eventos que inician y/o finalizan trazas.
- Eventos repetidos en las trazas.

Siendo l_i la sección del registro de eventos relacionada con el subproceso s_i , el subconjunto de este comportamiento observable en l_i es el comportamiento del proceso $w_j \in W$: $\beta_{w_j l_i} \subseteq \beta_{w_j s_i} | w_j \in W$. Por tanto, el comportamiento presente en la sección del registro de eventos l_i es $\beta_{l_i} = \beta_{w_s l_i} \cup \beta_{w_x l_i} \cup \beta_{w_o l_i} \cup \beta_{w_p l_i} \cup \beta_{w_l l_i}$.

Por cada comportamiento extraído se obtiene la frecuencia de su aparición en las trazas codificadas. Esta frecuencia es utilizada para establecer cuán representativo es un comportamiento del proceso que lo genera.

Los comportamientos extraídos en esta fase se combinan durante la obtención de variantes de descomposición. La extracción de comportamiento permite identificar, con mayor precisión que a nivel de traza, los elementos de ejecución del proceso, cuya baja frecuencia de ejecución sugiere que deban ser considerados como ruido.

III-C. Obtención de variantes de descomposición

En la tercera etapa se obtienen las variantes de descomposición del proceso para su representación en un *Árbol de Variantes*. Siendo S_P el conjunto de todos los subprocesos del proceso P , para cada nuevo subproceso $s_i \in S_P$, se busca una variante de descomposición para cada uno de los patrones de control de flujo, dentro de un ámbito definido como

parámetros de entrada. El ámbito se especifica en términos de umbrales para ruido (T_n) y completitud (T_c), definidos como un número entre 0 y 1 para cada patrón de control de flujo. Por ejemplo, un umbral de 0.2 para ruido en el patrón secuencia significa que hasta el 20% del comportamiento expresado en el registro de eventos puede ser descartado para encontrar una descomposición mediante el patrón secuencia. De manera similar, un 0.8 de completitud para paralelismo indica que el subproceso puede ser descompuesto mediante ese patrón conteniendo solamente el 80% del comportamiento necesario para ser identificado.

Denotemos como $\beta'_{w_j s_i}$ al comportamiento del proceso para el patrón w_j en s_i que no está en β_{l_i} pero puede ser asumido dentro del umbral de completitud. Sea $\beta'_{l_i} \subset (\beta_{l_i} \setminus \beta_{w_j l_i})$ el comportamiento del proceso que está presente en β_{l_i} y puede ser descartado dentro del umbral de ruido. Para un patrón de control de flujo w_j es posible encontrar una variante si $(\beta_{w_j l_i} \setminus \beta'_{l_i}) \cup \beta'_{w_j s_i}$ es equivalente a $\beta_{w_j s_i}$.

Para encontrar la combinación de comportamientos que puede conducir a la descomposición de un subproceso por un patrón de control de flujo se utiliza la búsqueda de variantes [24]. Por cada patrón de control de flujo w_j y cada sección del registro de eventos l_i , cada elemento de comportamiento $b \in \beta_{w_j l_i}$ es procesado, en una búsqueda en orden descendente según la frecuencia de b (denotada por $b(\varphi(b))$).

En la búsqueda de variantes cada nodo en el espacio de búsqueda está definido por una posible descomposición $\Upsilon_{w_j s_i}$, el comportamiento no procesado ($\beta'_{w_j l_i} \subseteq \beta_{w_j l_i}$), el comportamiento descartado ($\eta_{i w_j}$) y el comportamiento asumido ($\theta_{i w_j}$). El costo del camino desde un nodo a alguno de sus vecinos se establece en función de la frecuencia relativa de los comportamientos que es necesario descartar o asumir para convertir la $\Upsilon_{w_j s_i}$ del nodo actual en la del nodo vecino. El objetivo se alcanza cuando $\beta_{w_j l_i}$ ha sido procesado completamente y la posible descomposición $\Upsilon_{w_j s_i}$ posee más de un conjunto disjunto de las actividades presentes en s_i .

En cada nodo de búsqueda se procesa paulatinamente el comportamiento del patrón presente en la sección del registro de eventos relativa al subproceso que se intenta descomponer. A continuación se exponen los principios básicos que guían este procesamiento para cada uno de los patrones de control de flujo que se intentan identificar.

1) *Secuencia*: El comportamiento procesado por este patrón son las sucesiones directas. Los subprocesos s_i y s_{i+i} están en secuencia si todas las *DS* dirigidas hacia actividades de s_{i+i} son solo de actividades de s_i o en s_{i+i} no existen relaciones directas hacia s_i . La búsqueda se realiza partiendo de una posible descomposición, donde se encuentra un subproceso el cual contiene todas las actividades presentes en l_i . El conjunto de actividades es procesado para crear el primer subproceso en secuencia con otro subproceso que contiene el resto de las actividades.

2) *Lazo*: Los comportamientos procesados para la identificación de este patrón son las *DS*, las actividades que inician y concluyen trazas y las actividades repetidas. Dos subprocesos nombrados **Do** y **Redo** están en lazo si se cumplen las siguientes condiciones:

- Todas las actividades que inician o terminan trazas pertenecen al **Do** y deben ser actividades repetidas.
- En todas las *DS* cuya actividad predecesora está en el **Do** y sucesora aparece en el **Redo**, la actividad predecesora debe iniciar al menos una traza.
- En todas las *DS* cuya actividad predecesora está en el **Redo** y sucesora aparece en el **Do**, la actividad sucesora debe terminar al menos una traza.

3) *Paralelismo*: El comportamiento procesado para este patrón son las sucesiones indirectas (*IS*). Los subprocesos *A* y *B* forman parte de un paralelismo si para cada actividad **a** que pertenece al subproceso *A* y para cada actividad **b** que pertenece a *B* existe una sucesión indirecta desde **a** hasta **b** y una desde **b** hasta **a**. Además, en cada una de las trazas debe aparecer una representación de los subprocesos *A* y *B*.

4) *Selección Exclusiva*: Los subprocesos *A* y *B* están en selección exclusiva si no existen *IS* desde **a** hasta **b** o desde **b** hasta **a**, para cada actividad **a** que pertenece al subproceso *A* y cada actividad **b** que pertenece a *B*. La primera descomposición posible contiene a cada una de las actividades como un subproceso. El procesamiento de las *IS*, en las cuales se violan las condiciones planteadas, provoca la unión de los subprocesos a los que pertenecen.

5) *Selección no Exclusiva*: Los subprocesos *A* y *B* están en selección no exclusiva si para cada actividad **a** perteneciente a *A* y para actividad **b** que pertenece a *B* existe una *IS* desde **a** hasta **b** y una *IS* desde **b** hasta **a**. Además, debe existir representación de solo uno de los subprocesos en al menos una traza.

III-D. Generación del perfil de diagnóstico

Esta es la última etapa que se ejecuta dentro de la técnica. El algoritmo implementado en esta etapa recorre el árbol de búsqueda generado y recopila los comportamientos considerados como ruido y ausencia de información durante la identificación de los patrones. A partir de estos comportamientos asumidos o descartados se establecen la estimación de aptitud (f) y precisión (ρ). El perfil se conforma con la información de diagnóstico de cada nodo de tipo patrón en el *Árbol de Variantes* obtenido.

Para la representación del *Árbol de Variantes* en la herramienta, se muestra en cada momento una única variante del proceso. En cada nodo visible se combinan, el nodo subproceso relativo al subproceso del nivel correspondiente del árbol, con el nodo patrón relativo a la descomposición seleccionada.

IV. APLICACIÓN DE *Variants Miner* A UN REGISTRO DE EVENTOS DEL SISTEMA XAVIA HIS

El Sistema de Información Hospitalaria XAVIA HIS es un sistema de gestión que permite la recolección, almacenamiento, procesamiento, recuperación y comunicación de información de atención al paciente para todas las actividades relacionadas con la institución de salud. Está concebido para registrar las actividades de salud orientadas a los pacientes, permitiendo además gestionar y controlar los recursos de cada una de las áreas de dichas instituciones.

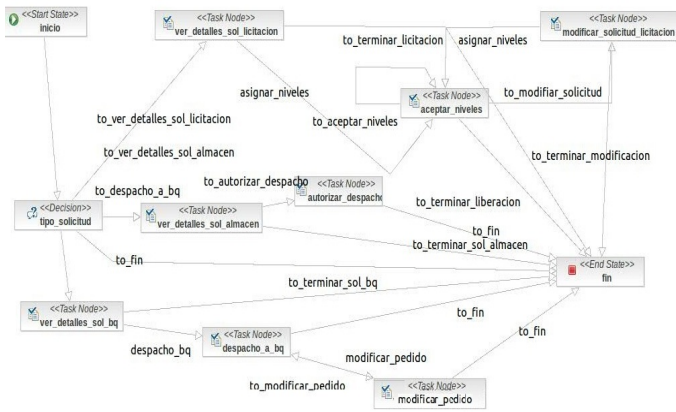


Figura 3. Modelo del proceso *Solicitar Producto*.

La técnica *Variants Miner* se ha aplicado a un registro de eventos que corresponde al módulo Almacén del sistema *XAVIA HIS*. Este registro de eventos recopila 1212 ejecuciones del proceso *Solicitar Producto* entre las fechas 09/01/2012 y 27/06/2013. El proceso *Solicitar Producto* cuenta con 8 actividades diferentes con un total de 2400 ocurrencias. El modelo de referencia de este proceso (Figura 3), posee dos tipos de nodos, los de tipo **Decision** y los de tipo **Task Node**. Los nodos del tipo **Decision** expresan el patrón de control de flujo *Selección exclusiva*.

Antes de cualquier **Task Node** aparece un nodo de tipo **Decision**, denominado *tipo_solicitud*, el cual expresa el patrón de control de flujo *Selección exclusiva* manifestando la existencia de diferentes alternativas. Dichas alternativas corresponden a los subprocesos *ver detalles_sol_licitacion*, *ver detalles_sol_almacen* y *ver detalles_sol_bq*.

La aplicación de *Variants Miner*, con los valores predeterminados en los parámetros de ruido y completitud, permite la obtención de dos variantes del proceso. Las variantes obtenidas corresponden a las descomposiciones mediante los patrones de control de flujo *Selección Exclusiva* y *Secuencia*, respectivamente.

En la primera variante se identifica la existencia de una selección exclusiva entre tres subprocesos. La presencia de este patrón de control de flujo, en este nivel de abstracción, indica que al inicio del proceso existe un punto de decisión a partir del cual solo se ejecuta uno de los subprocesos involucrados, lo cual es coherente con el modelo de referencia. En la Figura 4 se muestra el *Árbol de Variantes* identificado. A la derecha del modelo se puede observar la leyenda que contiene la codificación de las actividades.

En el *Perfil de Diagnóstico* asociado a esta variante se obtuvo una evaluación de 0.0 para el ruido y 1.0 para la completitud. Esto significa que para determinar esta variante no fue necesaria la consideración de comportamiento como ruido y que el comportamiento necesario para la identificación del patrón se encuentra presente en el registro de eventos. Las estimaciones de ruido y completitud obtenidas para esta descomposición corroboran que en el registro de eventos se expresa el comportamiento del patrón de control de flujo *Selección Exclusiva* en su totalidad.

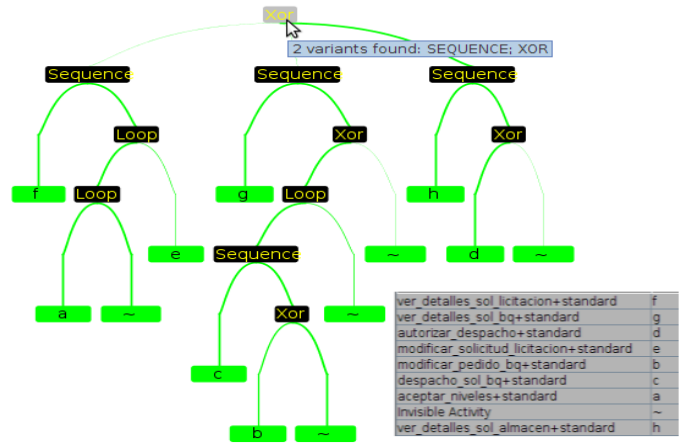


Figura 4. Variante de descomposición obtenida por Selección Exclusiva

El *Árbol de Variantes* proporciona una vista jerárquica del proceso, lo que posibilita el análisis de cada uno de los subprocesos que componen el proceso por separado y de acuerdo a diferentes niveles de abstracción. Se aprecia además, en el grosor de la última arista del primer nivel de abstracción, que la alternativa más ejecutada del proceso es la del subproceso “sol_almacen”, conformado por las actividades *ver detalles_sol_almacen* y *autorizar_despacho*, codificadas como *h* y *d* respectivamente. Seguidamente se realiza el análisis de los subprocesos identificados correspondientes al segundo nivel de abstracción.

IV-A. Subproceso “sol_almacen”

En la Figura 3 se puede apreciar una relación de las actividades *ver detalles_sol_almacen* y *autorizar_despacho* con el estado “fin” que indica el proceso puede concluir con cualquiera de estas actividades. Este comportamiento se encuentra reflejado en el registro de eventos y en la descomposición obtenida para esta sección del registro de eventos. La descomposición obtenida para este subproceso, las trazas y la codificación de sus actividades se muestran en la Figura 5

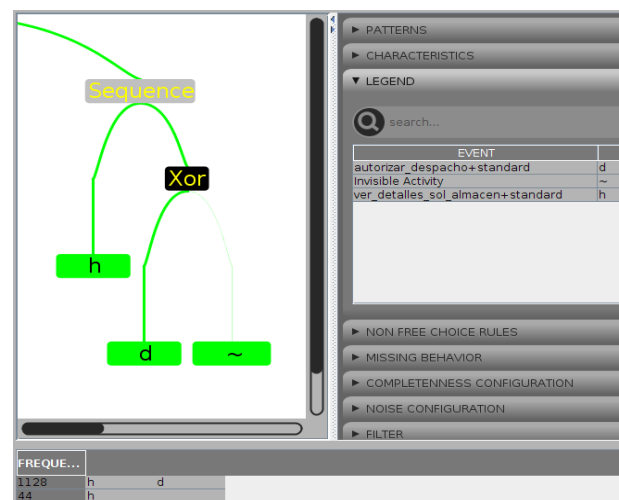


Figura 5. Subproceso *ver detalles_sol_almacen*

En esta sección del registro de eventos se identifica la existencia de la secuencia entre la actividad *ver_detalles_sol_almacen* y un subproceso descompuesto mediante el patrón de control de flujo *Selección Exclusiva*. La traza que contiene únicamente a la actividad *autorizar_despacho* representa una de las situaciones de ausencia de información reportadas en la literatura [25].

La ausencia de representación de uno de los subprocesos que se encuentran en secuencia en alguna traza es considerada ausencia de información para el patrón de control de flujo *Secuencia*. El análisis de la sección del registro de eventos correspondiente evidencia la ausencia de información en la segunda traza. La identificación de la ausencia de información en esta traza provoca la inserción de una actividad invisible inmediatamente después de la actividad *ver_detalles_sol_almacen*. Una tarea invisible se puede manifestar cuando se produce un salto de una o varias actividades en una situación de selección. La inserción de la actividad invisible en la traza posibilita la posterior identificación del patrón de *Selección Exclusiva*.

Se puede apreciar en el grosor de las aristas que la alternativa que contiene la actividad invisible insertada es menos frecuente. Esto indica que la ejecución más común del subproceso “sol_almacen” es la secuencia conformada por *ver_detalles_sol_almacen* y *autorizar_despacho*. Esto es coherente con la frecuencia de las trazas únicas de la sección del registro de eventos referente a este subproceso (Figura 5).

IV-B. Subproceso “sol_licitacion”

En la Figura 6 se muestra el resultado obtenido por *Variants Miner* en la sección del registro de eventos asociada a este subproceso. La técnica identificó una secuencia entre la actividad *ver_detalles_sol_licitacion* y un subproceso compuesto por las actividades *modificar_solicitud_licitacion* y *aceptar_niveles*. La secuencia identificada implica que ninguna instancia puede concluir su ejecución con *ver_detalles_sol_licitacion*. Esto contradice el modelo de referencia (Figura 3), pero es coherente con el comportamiento reflejado en las trazas.

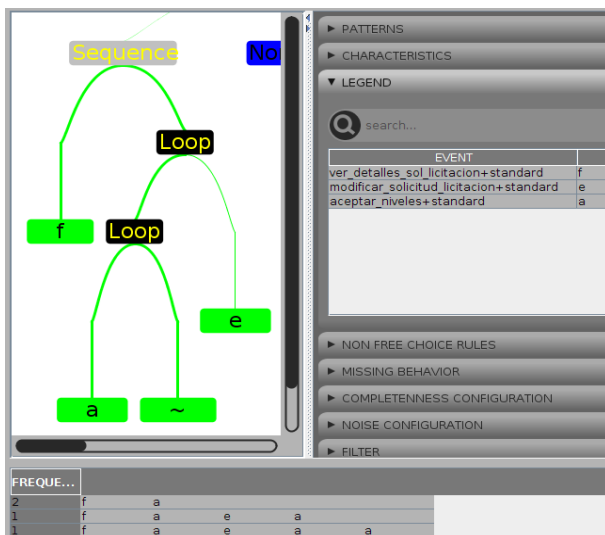


Figura 6. Subproceso *ver_detalles_sol_licitacion*

En el subproceso restante se identificó un lazo con la actividad *modificar_solicitud_licitacion* en el **Redo**, lo cual implica que nunca es la última actividad en ejecutarse. El subproceso **Do** se descompone en un lazo entre *aceptar_niveles* y una actividad invisible. Por ende, todas las instancias ejecutadas en el ámbito temporal que contiene el registro de eventos analizado han concluido con *aceptar_niveles*.

La descomposición obtenida no coincide completamente con el modelo de referencia, sin embargo es coherente con el registro de eventos. Se concluye entonces que en el registro de eventos no se encuentra registrado todo el comportamiento expresado en el modelo de referencia (Figura 3).

IV-C. Subproceso “sol_bq”

El *Perfil de Diagnóstico* para este subproceso (Figura 7) muestra una traza donde aparece solamente *ver_detalles_sol_bq* y no hay representación de las actividades que componen el segundo subproceso de la secuencia identificada. Esta situación genera la introducción de una actividad invisible para establecer que la ejecución puede concluir con *ver_detalles_sol_bq*. Se evidencia la baja frecuencia de esta alternativa en el grosor de la arista correspondiente en el *Árbol de Variantes*, lo cual sugiere una anomalía en el proceso.

En esta variante se identifica un lazo en el cual el subproceso **Redo** está compuesto por una actividad invisible. En el **Do** se identifica una secuencia entre *despacho_sol_bq* y un subproceso descompuesto por *Selección Exclusiva*, lo que indica que *despacho_sol_bq* siempre debe ejecutarse en este subproceso. El patrón *Selección Exclusiva* entre la actividad *modificar_pedido_bq* y una actividad invisible implica que la ejecución del lazo puede terminar tanto en la actividad *despacho_sol_bq* como en la actividad *modificar_pedido_bq*.

El análisis de la sección del registro de eventos correspondiente muestra que la descomposición obtenida representa el comportamiento presente en el registro de eventos. Además el *Árbol de Variantes* obtenido es coherente con el modelo de referencia (Figura 3).

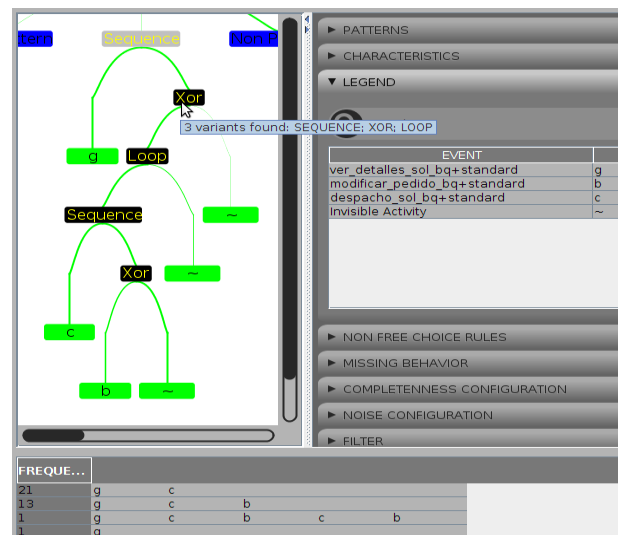


Figura 7. Subproceso *ver_detalles_sol_bq*

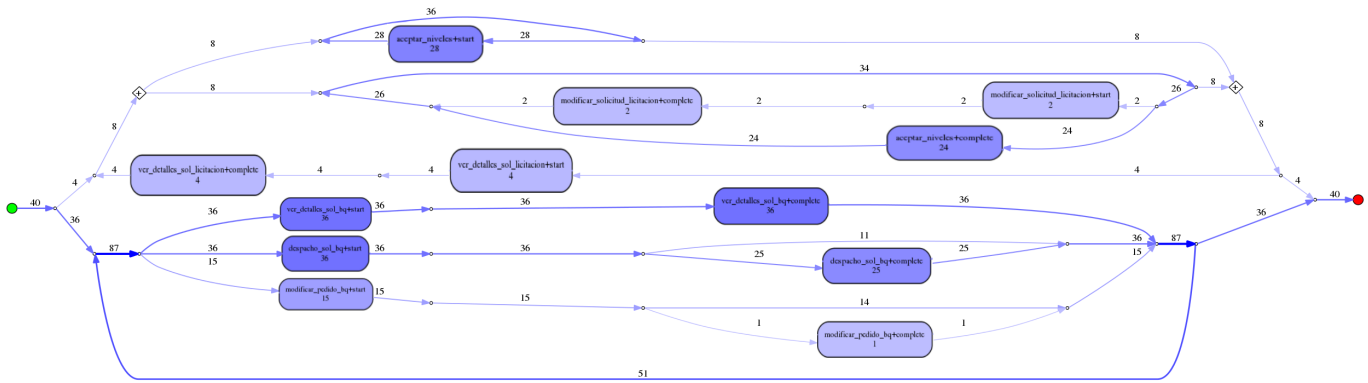


Figura 8. Modelo obtenido por *IvM* para las ejecuciones del proceso *Solicitar Producto* entre el 10/01/2012 y 01/02/2012

V. ANÁLISIS DE UN PROCESO CON VARIABILIDAD EN SU EJECUCIÓN

Para evidenciar la utilidad de realizar un diagnóstico basado en variantes de modelos se muestran los resultados de aplicar *Variants Miner* e *IvM* en un registro de eventos de un proceso con variabilidad en su ejecución. El registro de eventos recopila 40 ejecuciones del mismo proceso utilizado en la sección anterior (*Solicitar Producto*) pero en esta ocasión en otro hospital y entre las fechas 10/01/2012 y 01/03/2012. Este registro posee 16 trazas únicas, lo cual significa que 4 de cada 10 instancias de este proceso se ejecutaron con una secuencia de eventos diferente. Por cada actividad ejecutada se registró un evento al iniciar la misma y otro evento al completarse.

En la Figura 8 se muestra el modelo obtenido con *IvM* a partir de este registro de eventos. Los umbrales de *IvM* fueron configurados para garantizar una perfecta aptitud del modelo, lo que debe incluir en la visualización todos los caminos de ejecución. Se puede apreciar que se identifica una selección exclusiva entre los subprocesos “sol_licitacion” y “sol_bq”. La alternativa de ejecución que involucra al subproceso “sol_almacen” no aparece en este registro de eventos.

El *Árbol de Variantes* obtenido al ejecutar *Variants Miner* se muestra en la Figura 9. En este caso se visualiza la variante de modelo que posee mayor similitud con el modelo obtenido por *IvM*. A continuación se analizarán las variantes identificadas en el subproceso “sol_bq” y su relación con el modelo obtenido por *IvM* y con el registro de eventos.

V-A. Variantes en el subproceso “sol_bq”

La variante de descomposición de este subproceso utilizando un *Lazo* (Figura 9), es diferente de la encontrada por el *IvM* ya que ubica en el **Redo** a la actividad *modificar_pedido_bq+complete* (codificada como j), lo cual es coherente con el registro de eventos. En el **Do** de este lazo se identificó otro, con una actividad invisible en el **Redo** y una selección exclusiva entre *modificar_pedido_bq+start* y dos secuencias. Esta selección exclusiva coincide con el modelo de la Figura 8. Sin embargo esta variante de modelo, al igual que el que obtiene *IvM*, permite prácticamente cualquier comportamiento que involucre a estas actividades, incluyendo su repetición. Ambos modelos son poco precisos respecto al registro de eventos ya que en este no aparecen repetidas las actividades

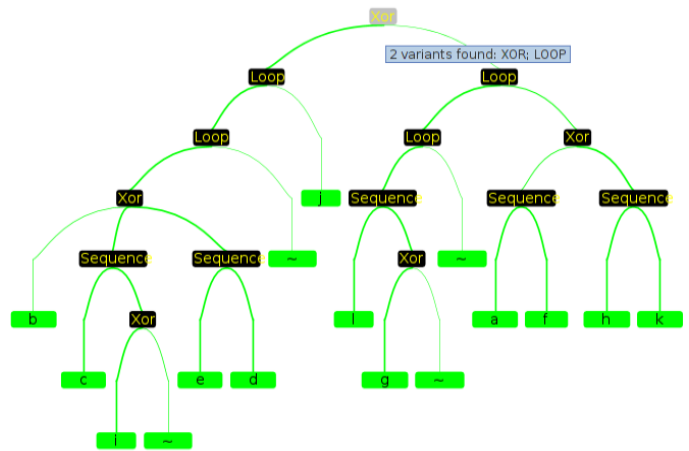
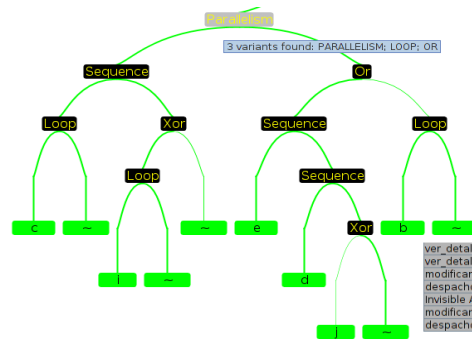


Figura 9. *Árbol de Variantes* obtenido con *Variants Miner*

modificar_pedido_bq+complete, *ver_detalle_sol_bq+start* ni *ver_detalle_sol_bq+complete*. En este caso el *IvM* al balancear los criterios de calidad sacrifica precisión para lograr una aptitud total del modelo.



3 variants found: PARALLELISM; LOOP; OR

| FREQUE... | b | e | d | c | i |
|-----------|---|---|---|---|---|
| 3 | c | i | b | e | d |
| 1 | b | e | d | c | i |
| 1 | b | i | b | e | d |
| 1 | c | i | e | d | b |
| 1 | e | d | i | b | c |

Figura 10. Variante del subproceso “sol_bq” con *Paralelismo*

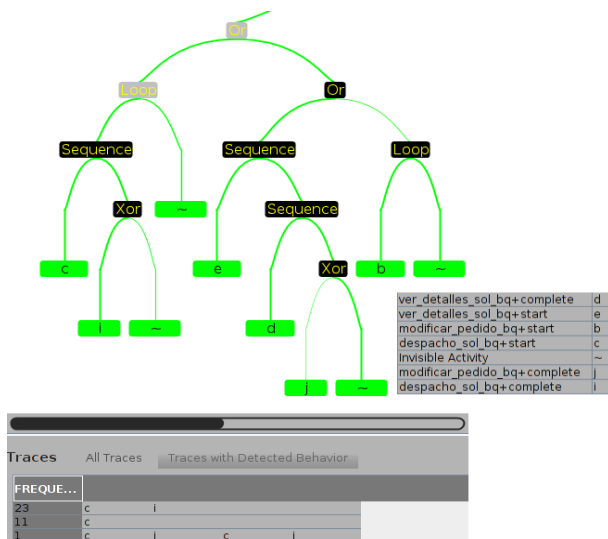


Figura 11. Variante del subproceso “sol_bq” con Selección no Exclusiva

Por otro lado, *Variants Miner* identificó para este subproceso dos variantes adicionales una para *Paralelismo* y otra para *Selección no Exclusiva*. En la Figura 10 se puede apreciar la variante para *Paralelismo*. Esta variante es más precisa que la del *Lazo* aunque descarta como ruido la traza marcada en azul en el *Perfil de Diagnóstico* que se puede apreciar en la figura. Este comportamiento es considerado por la variante de *Selección no Exclusiva*, que tiene una aptitud completa y es más precisa que la variante del *Lazo* (Figura 11).

En la variante de *Selección no Exclusiva* se identificó además un *Lazo* que incluye en el *Do* a la secuencia formada por *despacho_sol_bq+start* y *despacho_sol_bq+complete*. En la Figura 11 se muestra la sección del registro de eventos correspondiente a este subproceso. Se puede apreciar además por la *Selección Exclusiva* que se identificó al final de la secuencia, que la actividad *despacho_sol_bq+complete* ha sido saltada en ciertas ejecuciones, esto coincide con el modelo obtenido por *Ivm* y es coherente con el registro de eventos.

Otro elemento relevante en este subproceso es la *Selección no Exclusiva* que involucra a los eventos de las actividades *ver_detalle_sol_bq* y *modificar_pedido_bq* (Figura 11). En el *Árbol de Variantes* solamente se identifica un lazo sobre *modificar_pedido_bq+start*, lo cual no coincide con el modelo obtenido por *Ivm* pero es más preciso y coherente con el registro de eventos. Se puede apreciar además que es muy frecuente la secuencia de eventos *ver_detalle_sol_bq+start*, *ver_detalle_sol_bq+complete* y la baja ocurrencia de *modificar_pedido_bq+complete*. Esto último implica que en la mayoría de las instancias la actividad *modificar_pedido_bq* comienza pero no concluye.

V-B. Variantes en el subproceso “sol_licitacion”

La sección del modelo de la Figura 8 para este subproceso es poco precisa ya que muestra un ciclo que involucra a todas las actividades del subproceso, sin embargo en las trazas solamente aparecen repetidos los eventos asociados a la actividad *aceptar_niveles* (1 y g). En este modelo también

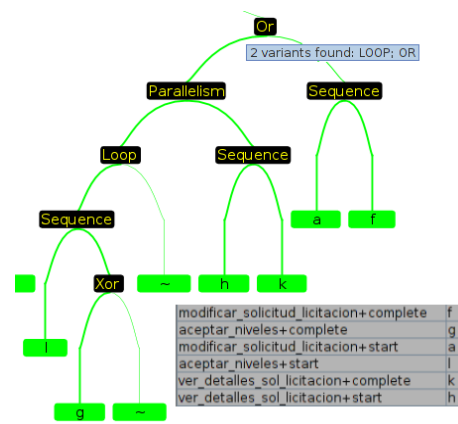


Figura 12. Variante de “sol_licitacion” con Selección no Exclusiva

se sacrifica precisión para lograr aptitud. En el *Árbol de Variantes* de la Figura 9, se muestra la variante de mayor similitud con el modelo de *Ivm*. La diferencia fundamental es que en este caso los eventos *aceptar_niveles+start* (1) y *aceptar_niveles+complete* (g) se ubican en un secuencia que incluye una actividad invisible. Esto permite detectar que la actividad *aceptar_niveles* no concluye su ejecución en todas las trazas.

La Figura 12, muestra una variante alternativa para este subproceso. En este caso sin perder aptitud se gana en precisión del modelo ya que se expresa la repetición únicamente de los eventos *aceptar_niveles+start* (1) y *aceptar_niveles+complete* (g). Además puede apreciarse la ejecución en paralelo de estos eventos con los asociados a la actividad *ver_detalle_sol_licitacion* (h y k), un comportamiento que se manifiesta en el registro de eventos sin embargo no es reflejado adecuadamente en el modelo obtenido por *Ivm*.

Se puede apreciar cómo los modelos alternativos contenidos en el *Árbol de Variantes* facilitan la exploración de un proceso con variabilidad en su ejecución. Esto brinda un mayor control al analista de procesos para decidir en función del contexto de ejecución cuáles son las anomalías y desviaciones y su impacto en la estructura de control de flujo del modelo que representa al proceso.

VI. CONCLUSIONES

La creación de un nuevo enfoque basado en la identificación de variantes de modelos de proceso permitió incluir las potencialidades y solventar las principales limitaciones de las técnicas disponibles, respecto a la comprensión del proceso en la etapa de diagnóstico. El desarrollo de la técnica *Variants Miner* a partir del enfoque propuesto permite contar con una nueva herramienta para el diagnóstico de procesos de particular utilidad en procesos que poseen variabilidad en su ejecución. La aplicación de *Variants Miner* en registros de eventos provenientes de entornos reales evidencia su capacidad para la identificación de patrones de control de flujo, control y visualización del ruido y la ausencia de información. Además se pudo comprobar en el caso de estudio analizado la utilidad del análisis de múltiples alternativas de ejecución a través del *Árbol de Variantes* obtenido por la técnica desarrollada.

AGRADECIMIENTOS

Agradecer al Ing. Arturo Orellana García por su colaboración en la extracción de los registros de eventos del sistema XAVIA HIS.

REFERENCIAS

- [1] van der Aalst, W.M.P., *Process Mining. Discovery, Conformance and Enhancement of Business Processes*. Springer, Heidelberg, Dordrecht, London et. al, 2011.
- [2] H. Verbeek, J. Buijs, B. Van Dongen, and W. Van Der Aalst, *XES, XESame, and ProM 6*, ser. CAISE Forum 2010 on Information Systems Evolution, Hammamet, 2011, vol. 72 LNBIP.
- [3] M. Jans, J. Van Der Werf, N. Lybaert, and K. Vanhoof, "A business process mining application for internal transaction fraud mitigation," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13 351–13 359, 2011.
- [4] W. van der Aalst, "A decade of business process management conferences: Personal reflections on a developing discipline," in *Business Process Management*, ser. Lecture Notes in Computer Science, A. Barros, A. Gal, and E. Kindler, Eds. Springer Berlin / Heidelberg, 2012, vol. 7481, pp. 1–16. [Online]. Available: <http://www.springerlink.com/content/0v72244588km1660/abstract/>
- [5] W. van der Aalst, A. Adriansyah, A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, A. Burattin, J. Carmona, M. Castellanos, J. Claes, J. Cook, N. Costantini, F. Curbra, E. Damiani, M. De Leoni, P. Delias, B. Van Dongen, M. Dumas, S. Dustdar, D. Fahland, D. Ferreira, W. Gaaloul, F. Van Geffen, S. Goel, C. Günther, A. Guzzo, P. Harmon, A. Ter Hofstede, J. Hoogland, J. Ingvaldsen, K. Kato, R. Kuhn, A. Kumar, M. La Rosa, F. Maggi, D. Malerba, R. Mans, A. Manuel, M. McCreesh, P. Mello, J. Mendling, M. Montali, H. Motahari-Nezhad, M. Zur Muehlen, J. Munoz-Gama, L. Pontieri, J. Ribeiro, A. Rozinat, H. Seguel Pérez, R. Seguel Pérez, M. Sepúlveda, J. Sinur, P. Soffer, M. Song, A. Sperduti, G. Stilo, C. Stoel, K. Swenson, M. Talamo, W. Tan, C. Turner, J. Vanthienen, G. Varvaressos, E. Verbeek, M. Verdonk, R. Vigo, J. Wang, B. Weber, M. Weidlich, T. Weijters, L. Wen, M. Westergaard, and M. Wynn, *Process mining manifesto*, ser. 9th International Conference on Business Process Management, BPM 2011P, Clermont-Ferrand, 2012, vol. 99 LNBIP.
- [6] R. Yzquierdo-Herrera, R. Silverio-Castro, and M. Lazo-Cortés, "Sub-process discovery: Opportunities for process diagnostics," in *Enterprise Information Systems of the Future*, ser. Lecture Notes in Business Information Processing, G. Poels, Ed. Springer Berlin Heidelberg, 2013, no. 139, pp. 48–57.
- [7] Bose, R.P.J.C. and van der Aalst, W.M.P., "Process diagnostics using trace alignment: Opportunities, issues, and challenges," *Information Systems*, vol. 37, no. 2, pp. 117–141, Apr. 2012.
- [8] T. Molka, W. Gilani, and X.-J. Zeng, "Dotted chart and control-flow analysis for a loan application process," in *Business Process Management Workshops*, ser. Lecture Notes in Business Information Processing, M. L. Rosa and P. Soffer, Eds. Springer Berlin Heidelberg, Jan. 2013, no. 132, pp. 223–224.
- [9] C. Günther and van der Aalst, W.M.P., "Fuzzy mining - adaptive process simplification based on multi-perspective metrics," in *Business Process Management*, ser. 5th International Conference on Business Process Management, BPM 2007, vol. 4714 LNCS, Brisbane, 2007, pp. 328–343. [Online]. Available: <http://www.springerlink.com/index/n7610472h7680875.pdf>
- [10] R. López-Jiménez, E. Pupo-Hernández, R. Yzquierdo-Herrera, and D. Pérez-Alfonso, "Técnicas de minería de proceso para el diagnóstico de procesos de negocio. process mining techniques for business processes diagnosis." in *I Conferencia Científica Internacional UCIENCIA*. La Habana, Cuba: Universidad de las Ciencias Informáticas, 2014.
- [11] K. Sarshar and P. Loos, "Comparing the control-flow of epc and petri net from the end-user perspective," in *Business Process Management*. Springer, 2005, pp. 434–439, 00061. [Online]. Available: http://link.springer.com/chapter/10.1007/11538394_36
- [12] H. A. Reijers, T. Freytag, J. Mendling, and A. Eckleder, "Syntax highlighting in business process models," *Decision Support Systems*, vol. 51, no. 3, pp. 339–349, 2011, 00020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167923611000042>
- [13] J. Mendling, *Metrics for process models: empirical foundations of verification, error prediction, and guidelines for correctness*. Springer, 2008, vol. 6, 00194.
- [14] R. K. Stamper, *Signs, Information, Norms and Systems*. Berlin: De Gruyter, 1996.
- [15] J. Mendling, M. Strembeck, and J. Recker, "Factors of process model comprehension. findings from a series of experiments," *Decision Support Systems*, vol. 53, no. 1, pp. 195–206, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167923612000085>
- [16] P. A. Kirschner, "Cognitive load theory: Implications of cognitive load theory on the design of learning," *Learning and instruction*, vol. 12, no. 1, pp. 1–10, 2002, 00408. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0959475201000147>
- [17] Jan Claes and Geert Poels, "Process mining and the ProM framework: An exploratory survey," in *LNBIP*. Springer, 2012. [Online]. Available: <http://processmining.ugent.be/pdf/ClaesPoels2012BPI@BPM.pdf>
- [18] M. Song and W. M. P. V. d. Aalst, "Supporting process mining by showing events at a glance," in *Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS)*, 2007, pp. 139–145. [Online]. Available: <http://www.wis.win.tue.nl/wvdaalst/publications/p433.pdf>
- [19] R. Yzquierdo-Herrera, R. Silverio-Castro, and M. Lazo-Cortés, "Tratamiento de la ausencia de información en la minería de procesos," *Revista Facultad de Ingeniería Universidad de Antioquia*, vol. 0, no. 69, pp. 67–78, 2013, 00000. [Online]. Available: <http://www.redalyc.org/articulo.oa?id=43029812005>
- [20] S. Leemans, D. Fahland, and W. van der Aalst, "Exploring processes and deviations," in *Business Process Management Workshops*, 2014. [Online]. Available: http://www.processmining.org/_media/blogs/pub2014/bpi2014leemans.pdf
- [21] S. J. Leemans, D. Fahland, and W. M. van der Aalst, "Discovering block-structured process models from event logs containing infrequent behaviour," in *BPM Conference*, 2013. [Online]. Available: <http://fluxicon.com/blog/wp-content/uploads/2013/09/Discovering-Block-Structured-Process-Models.pdf>
- [22] A. Adriansyah, J. Munoz-Gama, J. Carmona, B. F. v. Dongen, and W.M.P. van der Aalst, "Alignment Based Precision Checking," in *Business Process Management Workshops*, ser. Lecture Notes in Business Information Processing, M. L. Rosa and P. Soffer, Eds. Springer Berlin Heidelberg, 2013, no. 132, pp. 137–149, 00029. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-36285-9_15
- [23] C. Li, M. Reichert, and A. Wombacher, "Mining business process variants: Challenges, scenarios, algorithms," *Data and Knowledge Engineering*, vol. 70, no. 5, pp. 409–434, 2011.
- [24] D. Pérez-Alfonso, R. Yzquierdo-Herrera, R. López-Jiménez, and E. Pupo-Hernández, "Búsqueda de costo uniforme para el diagnóstico de variantes de procesos de negocio," in *I Conferencia Científica Internacional UCIENCIA*. La Habana, Cuba: Universidad de las Ciencias Informáticas, 2014. [Online]. Available: https://uciencia.uci.cu/sites/default/files/public/p3549-ponencia-1125_0.pdf
- [25] R. Yzquierdo-Herrera, "Modelo para la estimación de información ausente en las trazas usadas en la minería de proceso," PhD Thesis, Universidad de las Ciencias Informáticas, 2012.