Introducción Nociones preliminares Desarrollo Resultados experimentales Conclusión y trabajos futuros

Generación y simplificación automática de especificaciones de procesos de producción

Autor: Lucio Nardelli

Director: Hernán Ponce de León

Departamento de Ciencias de la Computación Facultad de Ciencias Exactas, Ingeniería y Agrimensura Universidad Nacional de Rosario

Junio, 2016



Contenidos

- Introducción
- 2 Nociones preliminares
- 3 Desarrollo
- Resultados experimentales
- 5 Conclusión y trabajos futuros

Introducción Nociones preliminares Desarrollo Resultados experimentales Conclusión y trabajos futuros

• Hoy en día existe un acceso masivo a sistemas informáticos.

- Hoy en día existe un acceso masivo a sistemas informáticos.
- Esto genera inmensas cantidades de información.

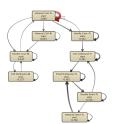
- Hoy en día existe un acceso masivo a sistemas informáticos.
- Esto genera inmensas cantidades de información.
- Dependencia total de estos sistemas, es esencial mantenerlos eficientes y seguros.

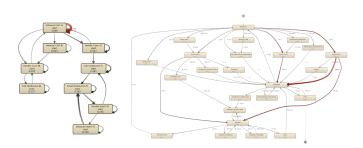
- Hoy en día existe un acceso masivo a sistemas informáticos.
- Esto genera inmensas cantidades de información.
- Dependencia total de estos sistemas, es esencial mantenerlos eficientes y seguros.
- ¿Cómo se puede asegurar esto?

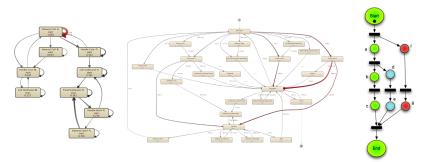
- Hoy en día existe un acceso masivo a sistemas informáticos.
- Esto genera inmensas cantidades de información.
- Dependencia total de estos sistemas, es esencial mantenerlos eficientes y seguros.
- ¿Cómo se puede asegurar esto?
- Utilizaremos métodos de representación formal.

Introducción Nociones preliminares Desarrollo Resultados experimentales Conclusión y trabajos futuros

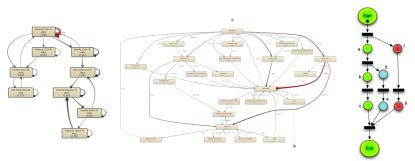
Motivación Modelos formales







Utilizar un modelo gráfico, que sea de fácil de entender pero formal para que se permita un análisis del sistema subyacente.



El problema es que estos modelos formales, ¡rara vez existen!

Introducción Nociones preliminares Desarrollo Resultados experimentales Conclusión y trabajos futuros

Introducción Nociones preliminares Desarrollo Resultados experimentales Conclusión y trabajos futuros

Motivación Descubrimiento de procesos

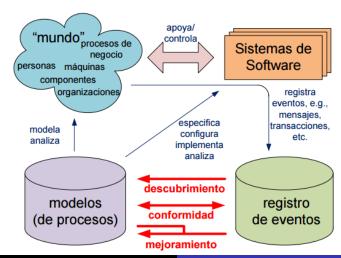
 Para obtener el modelo recurrimos al descubrimiento de procesos.

- Para obtener el modelo recurrimos al descubrimiento de procesos.
- Es una técnica de aprendizaje automatizado.

- Para obtener el modelo recurrimos al descubrimiento de procesos.
- Es una técnica de aprendizaje automatizado.
- Transforma un registro de acciones de un sistema en un modelo formal.

- Para obtener el modelo recurrimos al descubrimiento de procesos.
- Es una técnica de aprendizaje automatizado.
- Transforma un registro de acciones de un sistema en un modelo formal.

Motivación Minería de procesos gráficamente





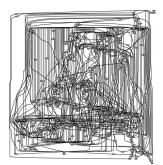
• Problema con las técnicas de descubrimiento: modelos *spaghetti*.

• Problema con las técnicas de descubrimiento: modelos *spaghetti*.



Los spaghetti pueden ser una muy buena idea para una cena romántica, pero no son una buena idea para un modelo formal...

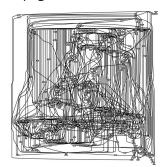
• Problema con las técnicas de descubrimiento: modelos *spaghetti*.



Introducción Nociones preliminares Desarrollo Resultados experimentales Conclusión y trabajos futuros

Motivación Mejora de modelos

 Problema con las técnicas de descubrimiento: modelos spaghetti.





¿Quién podría entender algo de un modelo como ese?



Introducción Nociones preliminares Desarrollo Resultados experimentales Conclusión y trabajos futuros

Motivación Mejora de modelos

• Para mejorarlos se aplican técnicas de mejora de modelos

- Para mejorarlos se aplican técnicas de mejora de modelos
- Puede lograrse considerando las trazas más frecuentes o bien información de comportamientos que *no* deben ocurrir.

Motivación Objetivo general

 Obtener un modelo formal de manera automática a partir de los logs de eventos de un sistema.

Motivación Objetivo general

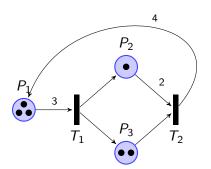
- Obtener un modelo formal de manera automática a partir de los logs de eventos de un sistema.
- Optimizar el modelo obtenido mediante técnicas de mejora de modelos para obtener un modelo que resulte simple.

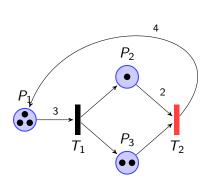
Motivación Objetivo general

- Obtener un modelo formal de manera automática a partir de los logs de eventos de un sistema.
- Optimizar el modelo obtenido mediante técnicas de mejora de modelos para obtener un modelo que resulte simple.
- Automatizar los puntos anteriores.

Contenidos

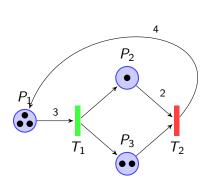
- Introducción
- 2 Nociones preliminares
- 3 Desarrollo
- Resultados experimentales
- 5 Conclusión y trabajos futuros





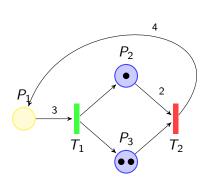
$$M_0(p_1) = 3$$

 $M_0(p_2) = 1$
 $M_0(p_3) = 2$



$$M_0(p_1) = 3$$

 $M_0(p_2) = 1$
 $M_0(p_3) = 2$

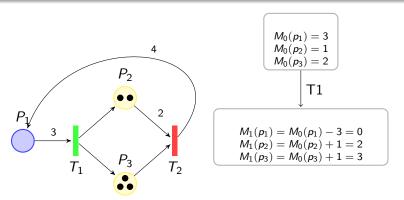


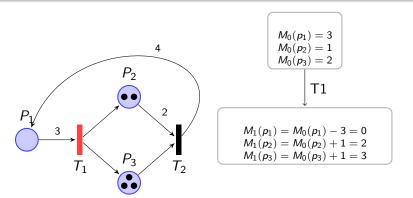
$$M_0(p_1) = 3$$

 $M_0(p_2) = 1$
 $M_0(p_3) = 2$

Redes de Petri

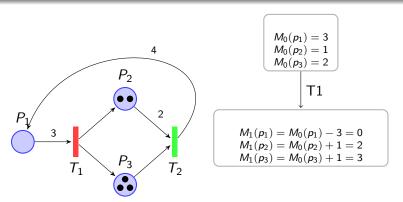
Ejemplo de la evolución de una red de Petri



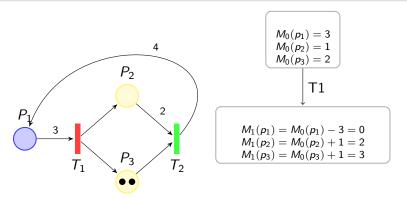


Redes de Petri

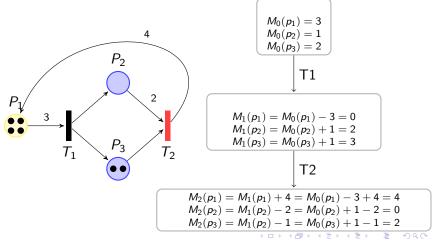
Ejemplo de la evolución de una red de Petri



Redes de Petri Ejemplo de la evolución de una red de Petri



Redes de Petri Ejemplo de la evolución de una red de Petri



Introducción **Nociones preliminares** Desarrollo Resultados experimentales Conclusión y trabajos futuros

Redes de Petri Evolución de una red de Petri

Redes de Petri Evolución de una red de Petri

• Los markings pueden definirse de manera incremental como

$$M'(p) = M(p) - F(p, t) + F(t, p).$$

Redes de Petri Evolución de una red de Petri

Los markings pueden definirse de manera incremental como

$$M'(p) = M(p) - F(p, t) + F(t, p).$$

• Más general, ante una sucesión de eventos:

$$M(p) = M_0(p) + \sum_{x_i} F(x_i, p) \cdot \widehat{\sigma}(x_i) - \sum_{y_i} F(p, y_i) \cdot \widehat{\sigma}(y_i).$$

Redes de Petri Evolución de una red de Petri

Los markings pueden definirse de manera incremental como

$$M'(p) = M(p) - F(p, t) + F(t, p).$$

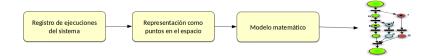
• Más general, ante una sucesión de eventos:

$$M(p) = M_0(p) + \sum_{x_i} F(x_i, p) \cdot \widehat{\sigma}(x_i) - \sum_{y_i} F(p, y_i) \cdot \widehat{\sigma}(y_i).$$

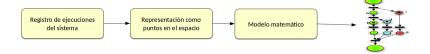
• Utilizando notación matricial, puede definirse para todos los places de una red:

$$M = M_0 + A \cdot \hat{\sigma}$$
.

Enfoque del trabajo



Enfoque del trabajo



- El registro de ejecuciones es generado por el sistema en los *logs de eventos*.
- La representación como puntos en el espacio corresponde al conjunto de vectores Parikh del log.

De logs y vectores Parikh

 Un log de eventos es un conjunto de trazas o sucesión ordenada de actividades relevantes de un sistema.

- Un log de eventos es un conjunto de trazas o sucesión ordenada de actividades relevantes de un sistema.
- Dada una traza $\sigma = \sigma_1 \cdot \sigma_2 \cdot \ldots \cdot \sigma_k$ sobre un alfabeto $T = \{t_1, t_2, \ldots, t_n\}$, el vector Parikh corresponde a la cantidad de ocurrencias de cada acción t_i en σ .

- Un log de eventos es un conjunto de trazas o sucesión ordenada de actividades relevantes de un sistema.
- Dada una traza $\sigma = \sigma_1 \cdot \sigma_2 \cdot \ldots \cdot \sigma_k$ sobre un alfabeto $T = \{t_1, t_2, \ldots, t_n\}$, el vector Parikh corresponde a la cantidad de ocurrencias de cada acción t_i en σ .

Ejemplo

Para las trazas $\sigma_1=t_1\cdot t_2\cdot t_1\cdot t_2\cdot t_1\cdot t_1\cdot t_3$ y $\sigma_2=t_4\cdot t_4$, sobre el alfabeto $T=\{t_1,t_2,t_3,t_4\}$ los vectores de Parikh de cada traza vienen dados por las tuplas (4,1,1,0) y (0,0,0,2) respectivamente.

- Un log de eventos es un conjunto de trazas o sucesión ordenada de actividades relevantes de un sistema.
- Dada una traza $\sigma = \sigma_1 \cdot \sigma_2 \cdot \ldots \cdot \sigma_k$ sobre un alfabeto $T = \{t_1, t_2, \ldots, t_n\}$, el vector Parikh corresponde a la cantidad de ocurrencias de cada acción t_i en σ .

Ejemplo

Para las trazas $\sigma_1=t_1\cdot t_2\cdot t_1\cdot t_2\cdot t_1\cdot t_1\cdot t_3$ y $\sigma_2=t_4\cdot t_4$, sobre el alfabeto $T=\{t_1,t_2,t_3,t_4\}$ los vectores de Parikh de cada traza vienen dados por las tuplas (4,1,1,0) y (0,0,0,2) respectivamente.

Entonces, teníamos un log de eventos y ahora podemos convertirlo en tuplas de enteros...

- Un log de eventos es un conjunto de trazas o sucesión ordenada de actividades relevantes de un sistema.
- Dada una traza $\sigma = \sigma_1 \cdot \sigma_2 \cdot \ldots \cdot \sigma_k$ sobre un alfabeto $T = \{t_1, t_2, \ldots, t_n\}$, el vector Parikh corresponde a la cantidad de ocurrencias de cada acción t_i en σ .

Ejemplo

Para las trazas $\sigma_1=t_1\cdot t_2\cdot t_1\cdot t_2\cdot t_1\cdot t_1\cdot t_3$ y $\sigma_2=t_4\cdot t_4$, sobre el alfabeto $T=\{t_1,t_2,t_3,t_4\}$ los vectores de Parikh de cada traza vienen dados por las tuplas (4,1,1,0) y (0,0,0,2) respectivamente.



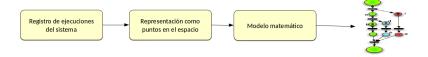
Entonces, teníamos un log de eventos y ahora podemos convertirlo en tuplas de enteros...

Enfoque del trabajo



- El registro de ejecuciones es generado por el sistema en los *logs de eventos*.
- La representación como puntos en el espacio corresponde al conjunto de vectores Parikh del log.

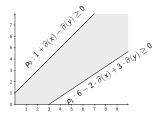
Enfoque del trabajo



- El registro de ejecuciones es generado por el sistema en los *logs de eventos*.
- La representación como puntos en el espacio corresponde al conjunto de vectores Parikh del log.
- El modelo matemático corresponde a un poliedro convexo que contenga el conjunto de vectores Parikh.

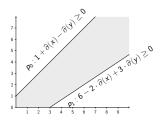
Dominios numéricos abstractos

Poliedros convexos



Dominios numéricos abstractos

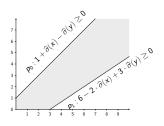
Poliedros convexos



 Los semi-espacios se representan mediante una inecuación lineal de la forma a₁x₁ + a₂x₂ + ··· + a_nx_n ≥ b.

Dominios numéricos abstractos

Poliedros convexos

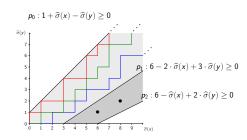


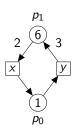
- Los semi-espacios se representan mediante una inecuación lineal de la forma a₁x₁ + a₂x₂ + · · · + a_nx_n ≥ b.
- Un poliedro convexo P puede representarse como una intersección de un conjunto de k hiper-espacios

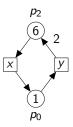
$$\mathcal{P} = \{x \in \mathbb{R}^n \mid A \cdot x + b \ge 0\}$$

donde $A \in \mathbb{R}^{k \times n}$ y $b \in \mathbb{R}^k$.

Descubrimiento de procesos ¡Juntemos todo!







Contenidos

- Introducción
- 2 Nociones preliminares
- 3 Desarrollo
- 4 Resultados experimentales
- 5 Conclusión y trabajos futuros

- Modelos excesivamente complicados y alejados de la realidad.
- Necesidad de simplificar.

- Modelos excesivamente complicados y alejados de la realidad.
- Necesidad de simplificar.

??? + ???
$$\cdot x_1$$
 + ??? $\cdot x_2$ + ??? $\cdot x_3 \ge 0$
??? + ??? $\cdot x_1$ + ??? $\cdot x_2$ + ??? $\cdot x_3 \ge 0$
??? + ??? $\cdot x_1$ + ??? $\cdot x_2$ + ??? $\cdot x_3$ ≥ 0

- Modelos excesivamente complicados y alejados de la realidad.
- Necesidad de simplificar.

- Buscamos nuevos coeficientes para el sistema de inecuaciones:
 - Los nuevos coeficientes deben ser más "simples".
 - Cada solución del sistema original tiene que continuar siendo una solución del nuevo sistema.

Dado un modelo de la forma:

Dado un modelo de la forma:

Se buscan nuevos coeficientes $\beta_{1,0}, \beta_{1,1}, \dots, \beta_{m,n}$ tal que:

$$|\beta_{i,j}| \le |\alpha_{i,j}|. \tag{MIN}$$

$$\bigwedge_{i=1}^{m} (\alpha_{i,0} + \sum_{j=1}^{n} \alpha_{i,j} \cdot x_{j}) \ge 0 \implies \bigwedge_{i=1}^{m} (\beta_{i,0} + \sum_{j=1}^{n} \beta_{i,j} \cdot x_{j}) \ge 0.$$
 (PC)

Dado un modelo de la forma:

Se buscan nuevos coeficientes $\beta_{1,0}, \beta_{1,1}, \dots, \beta_{m,n}$ tal que:

$$|\beta_{i,j}| \le |\alpha_{i,j}|. \tag{MIN}$$

$$\bigwedge_{i=1}^{m} (\alpha_{i,0} + \sum_{j=1}^{n} \alpha_{i,j} \cdot x_{j}) \ge 0 \implies \bigwedge_{i=1}^{m} (\beta_{i,0} + \sum_{j=1}^{n} \beta_{i,j} \cdot x_{j}) \ge 0.$$
 (PC)

- Para realizar el proceso de simplificación se utiliza SMT-Solver.
- SMT (Satisfability modulo theories) es una teoría que permite definir un sistema de restricciones v encontrar una solución. ◆ロト ◆卸ト ◆ 恵ト ◆ 恵ト ■ りゅつ 20/31



• Finalmente conseguimos el modelo simplificado.

 Finalmente conseguimos el modelo simplificado.



¡Qué bien!

- Finalmente conseguimos el modelo simplificado.
- Simplificar agregó nuevos puntos y por lo tanto comportamientos potencialmente indeseados.

- Finalmente conseguimos el modelo simplificado.
- Simplificar agregó nuevos puntos y por lo tanto comportamientos potencialmente indeseados.



¡Qué mal!

- Finalmente conseguimos el modelo simplificado.
- Simplificar agregó nuevos puntos y por lo tanto comportamientos potencialmente indeseados.
- Si contamos con información negativa podemos mejorarlo.

- Finalmente conseguimos el modelo simplificado.
- Simplificar agregó nuevos puntos y por lo tanto comportamientos potencialmente indeseados.
- Si contamos con información negativa podemos mejorarlo.



¡Qué bien!

- Finalmente conseguimos el modelo simplificado.
- Simplificar agregó nuevos puntos y por lo tanto comportamientos potencialmente indeseados.
- Si contamos con información negativa podemos mejorarlo.
- La información negativa representa conocimiento experto sobre comportamiento que el sistema no debe admitir.

- Finalmente conseguimos el modelo simplificado.
- Simplificar agregó nuevos puntos y por lo tanto comportamientos potencialmente indeseados.
- Si contamos con información negativa podemos mejorarlo.
- La información negativa representa conocimiento experto sobre comportamiento que el sistema no debe admitir.



. . .

Algoritmo de descubrimiento de procesos

Algoritmo completo de descubrimiento y simplificación supervisado

```
Entrada: trazas positivas \mathcal{L}^+ y trazas negativas \mathcal{L}_-
Salida: una red de Petri N donde \forall \sigma \in \mathcal{L}^+ : \sigma \in L(N) y \forall \sigma \in \mathcal{L}_- : \sigma \notin L(N)
 1: procedure DISCOVER(\mathcal{L}^+, \mathcal{L}_-)
 2:
           pp, np \leftarrow \emptyset
 3:
          for \sigma_p \in \mathcal{L}^+ do
 4:
                for \sigma prefix of \sigma_p do
 5:
                     add \widehat{\sigma} to pp
 6:
                end for
 7:
           end for
 8:
           for \sigma_n \in \mathcal{L}_- do
 9:
                add \widehat{\sigma}_n to np
10:
           end for
11:
           H = \text{ConvexHull}(pp)
12:
            H_{smt} = \text{ShiftRotate}(H, np)
13:
            N = \text{Hull2Net}(H')
14:
           return N
15: end procedure
```

- Desarrollo herramienta PACH en Python.
- Diferentes parámetros de configuración.
- Posibilita ser usada como herramienta de post procesamiento.
- Código disponible en GitHub.

Contenidos

- Introducción
- 2 Nociones preliminares
- 3 Desarrollo
- 4 Resultados experimentales
- 5 Conclusión y trabajos futuros

Resultados experimentales Sobre complejidad

A(32) 13129 10324 6489 A(42) 7813 4209 5009 ConfDimB 33 29 29 Cycles(5) 105 102 102 DocumentFlow 56 50 52 Incident 406 216 292 Receipt 588 412 462 Telecom 840 592 688	Benchmark	Poliedro	SMT Pos.	SMT Neg.
	A(42) CONFDIMB CYCLES(5) DOCUMENTFLOW INCIDENT RECEIPT	7813 33 105 56 406 588	4209 29 102 50 216 412	5009 29 102 52 292 462

Resultados de complejidad de los modelos obtenidos mediante PACH.



Resultados experimentales Sobre precisión

A(32) 0,16 0,15 0,15 A(42) 0,11 0,10 0,12 CONFDIMB 0,87 0,41 0,87 CYCLES(5) 0,23 0,23 0,23 DOCUMENTFLOW 0,04 0,04 0,04 INCIDENT 0,13 0,12 0,11 RECEIPT 0,15 0,13 0,15 TELECOM 0,08 0,08 0,08	Benchmark	Poliedro	SMT Pos.	SMT Neg.
	A(42) CONFDIMB CYCLES(5) DOCUMENTFLOW INCIDENT RECEIPT	0,11 0,87 0,23 0,04 0,13 0,15	0,10 0,41 0,23 0,04 0,12 0,13	0,12 0,87 0,23 0,04 0,11 0,15

Resultados de precisión de los modelos obtenidos mediante PACH.

Contenidos

- Introducción
- 2 Nociones preliminares
- 3 Desarrollo
- 4 Resultados experimentales
- 5 Conclusión y trabajos futuros

Conclusión

- Desarrollo de una herramienta que implementa minería de procesos de manera supervisada.
- Uso de SMT-Solver como herramienta de simplificación.
- La solución propuesta es independiente del algoritmo de descubrimiento.
- Buenos resultados experimentales.

Trabajo futuro

- Realizar pruebas con información negativa proporcionada por conocimiento experto.
- Descomposición en subproblemas ante información negativa dentro del poliedro positivo.
- Combinar con otras técnicas de simplificación.
- Obtención de poliedro de manera eficiente mediante SMT-Solver.

¿Preguntas?



¿Preguntas?



¡Gracias!