Generación y simplificación automática de especificaciones de procesos de producción

Autor: Lucio Nardelli

Director: Hernán (Zeta) Ponce de León

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

Junio, 2016



Contenidos

- Introducción
- Nociones preliminares
- 3 Desarrollo
- 4 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?

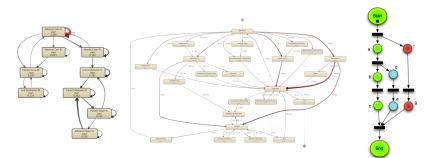
- Hoy en día existe un acceso masivo a sistemas informáticos.
- Esto genera inmensas cantidades de información.
- Dependencia total de estos sistemas, es esencial mantenerlos eficientes y seguros.
- ¿Cómo se puede asegurar esto?
- Utilizaremos métodos de representación formal.

Motivación Modelos formales

Utilizar un modelo gráfico, fácil de entender pero formal para que permitir su análisis.

Motivación Modelos formales

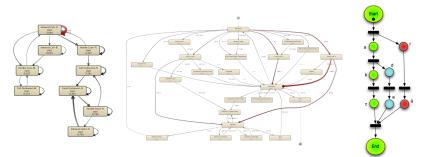
Utilizar un modelo gráfico, fácil de entender pero formal para que permitir su análisis.





Motivación Modelos formales

Utilizar un modelo gráfico, fácil de entender pero formal para que permitir su análisis.



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



Descubrimiento de procesos

 Para obtener el modelo recurrimos al descubrimiento de procesos.

Descubrimiento de procesos

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

Motivación Descubrimiento de procesos

- 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.

Descubrimiento de procesos

- 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.



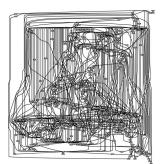
 Problema con las técnicas de descubrimiento: modelos spaghetti.

 Problema con las técnicas de descubrimiento: modelos spaghetti.

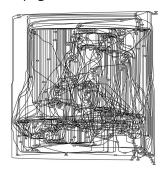


Los spaghetti pueden ser una muy buena idea para una cena romántica, pero no son una buena idea para un modelo formal...

 Problema con las técnicas de descubrimiento: modelos spaghetti.



 Problema con las técnicas de descubrimiento: modelos spaghetti.





¿Quién podría entender algo de un modelo como ese?



• 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 a partir del registro de actividades.

Motivación Objetivo general

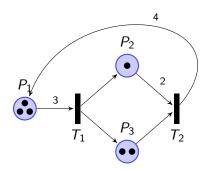
- Obtener un modelo formal a partir del registro de actividades.
- Simplificar el modelo.

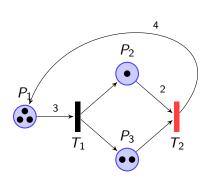
Motivación Objetivo general

- Obtener un modelo formal a partir del registro de actividades.
- Simplificar el modelo.
- Automatizar los puntos anteriores.

Contenidos

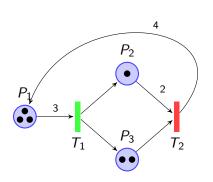
- Introducción
- 2 Nociones preliminares
- 3 Desarrollo
- 4 Conclusión y trabajos futuros





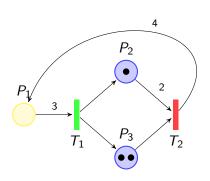
$$M_0(p_1) = 3$$

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



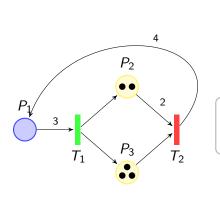
$$M_0(p_1) = 3$$

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



$$M_0(p_1) = 3$$

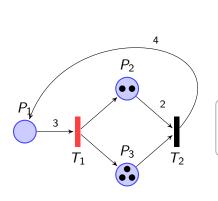
 $M_0(p_2) = 1$
 $M_0(p_3) = 2$





$$M_1(p_1) = M_0(p_1) - 3 = 0$$

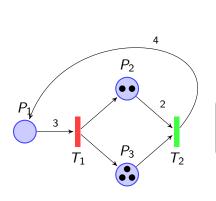
 $M_1(p_2) = M_0(p_2) + 1 = 2$
 $M_1(p_3) = M_0(p_3) + 1 = 3$





$$M_1(p_1) = M_0(p_1) - 3 = 0$$

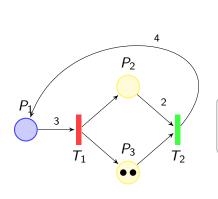
 $M_1(p_2) = M_0(p_2) + 1 = 2$
 $M_1(p_3) = M_0(p_3) + 1 = 3$





$$M_1(p_1) = M_0(p_1) - 3 = 0$$

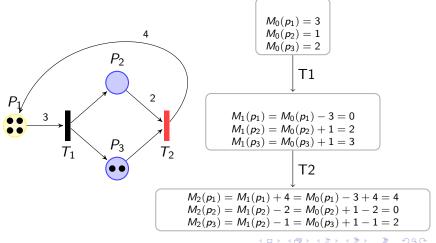
 $M_1(p_2) = M_0(p_2) + 1 = 2$
 $M_1(p_3) = M_0(p_3) + 1 = 3$





$$M_1(p_1) = M_0(p_1) - 3 = 0$$

 $M_1(p_2) = M_0(p_2) + 1 = 2$
 $M_1(p_3) = M_0(p_3) + 1 = 3$



Redes de Petri Evolución de una red de Petri

• Los markings pueden definirse de manera incremental como

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

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:

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



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.

De logs y vectores Parikh Conceptos

- 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 .

De logs y vectores Parikh

- 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 (4,2,1,0) y (0,0,0,2) respectivamente.

De logs y vectores Parikh

- 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 (4,2,1,0) y (0,0,0,2) respectivamente.

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

De logs y vectores Parikh Conceptos

- 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 (4,2,1,0) 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.

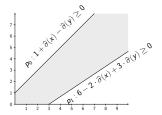
Enfoque del trabajo



- 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.

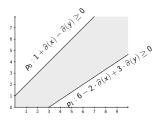
Dominios numéricos abstractos

Poliedros convexos



Dominios numéricos abstractos

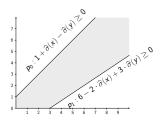
Poliedros convexos



• Los semi-espacios se representan mediante una inecuación lineal de la forma $a_1x_1 + a_2x_2 + \cdots + a_nx_n \ge b$.

Dominios numéricos abstractos

Poliedros convexos

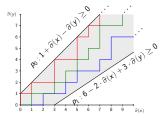


- Los semi-espacios se representan mediante una inecuación lineal de la forma $a_1x_1 + a_2x_2 + \cdots + a_nx_n \ge b$.
- Un poliedro convexo P puede representarse como una intersección de un conjunto de k hiper-espacios

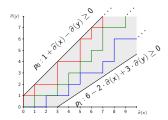
$$\mathcal{P} = \{ x \in \mathbb{R}^n \mid A \cdot x + b \ge 0 \}$$

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

Descubrimiento de procesos ¡Juntemos todo!

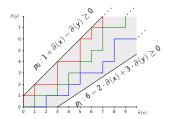


Descubrimiento de procesos ¡Juntemos todo!



$$\mathcal{P} = \left\{ \left[\begin{array}{c} \widehat{\sigma}(x) \\ \widehat{\sigma}(y) \end{array} \right] \ \mid \begin{array}{c} 1 + \widehat{\sigma}(x) - \widehat{\sigma}(y) \geq 0 \\ 6 - 2 \cdot \widehat{\sigma}(x) + 3 \cdot \widehat{\sigma}(y) \geq 0 \end{array} \right\}$$

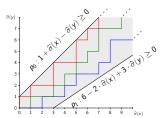
Descubrimiento de procesos ¡Juntemos todo!

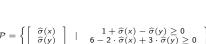


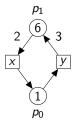
$$\mathcal{P} = \left\{ \left[\begin{array}{cc} \widehat{\sigma}(x) \\ \widehat{\sigma}(y) \end{array} \right] \ \mid \ \begin{array}{cc} 1 + \widehat{\sigma}(x) - \widehat{\sigma}(y) \geq 0 \\ 6 - 2 \cdot \widehat{\sigma}(x) + 3 \cdot \widehat{\sigma}(y) \geq 0 \end{array} \right\} \qquad M = \left[\begin{array}{cc} 1 \\ 6 \end{array} \right] + \left[\begin{array}{cc} 1 & -1 \\ -2 & 3 \end{array} \right] \cdot \left[\begin{array}{cc} \widehat{\sigma}(x) \\ \widehat{\sigma}(y) \end{array} \right] \geq \left[\begin{array}{cc} 0 \\ 0 \end{array} \right].$$

$$M = \begin{bmatrix} 1 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} \widehat{\sigma}(x) \\ \widehat{\sigma}(y) \end{bmatrix} \ge \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Descubrimiento de procesos ¡Juntemos todo!







$$\mathcal{P} = \left\{ \left[\begin{array}{cc} \widehat{\sigma}(x) \\ \widehat{\sigma}(y) \end{array} \right] \ \mid \ \begin{array}{cc} 1 + \widehat{\sigma}(x) - \widehat{\sigma}(y) \geq 0 \\ 6 - 2 \cdot \widehat{\sigma}(x) + 3 \cdot \widehat{\sigma}(y) \geq 0 \end{array} \right\} \qquad M = \left[\begin{array}{cc} 1 \\ 6 \end{array} \right] + \left[\begin{array}{cc} 1 & -1 \\ -2 & 3 \end{array} \right] \cdot \left[\begin{array}{cc} \widehat{\sigma}(x) \\ \widehat{\sigma}(y) \end{array} \right] \geq \left[\begin{array}{cc} 0 \\ 0 \end{array} \right].$$

Contenidos

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

- Modelos excesivamente complicados.
- Necesidad de simplificar.

- Modelos excesivamente complicados.
- Necesidad de simplificar.

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

- Modelos excesivamente complicados.
- Necesidad de simplificar.

- Buscamos nuevos coeficientes para el sistema de inecuaciones:
 - MINimización: Los nuevos coeficientes deben ser más "simples".
 - Preservación Trazas: 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 $\beta_{1,0},\beta_{1,1},\ldots,\beta_{m,n}$ tal que:

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

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

Dado un modelo de la forma:

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

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

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

- Para realizar el proceso de simplificación se utiliza SMT-Solver.
- Los SMT (Satisfability modulo theories) permiten definir un sistema de restricciones y encontrar una solución.

 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 \not\in L(N) 1: procedure \mathrm{DISCOVER}(\mathcal{L}^+, \mathcal{L}_-) 2: pp = \mathrm{PARIKH}(\mathcal{L}^+) 3: np = \mathrm{PARIKH}(\mathcal{L}_-) 4: H = \mathrm{ConvexHull}(pp) 5: H_{smt} = \mathrm{SHIFTROTATE}(H, np) 6: N = \mathrm{HULLNET}(H_{smt}) 7: return N 8: end procedure
```

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



PACH Resultados Experimentales

	Benchmark	Poliedro	SMT Pos.	SMT Neg.
A(42) 7813 4209 5009 CONFDIMB 33 29 29 CYCLES(5) 105 102 105 DOCUMENTFLOW 56 50 52 INCIDENT 406 216 292 RECEIPT 588 412 462	CONFDIMB CYCLES(5) DOCUMENTFLOW INCIDENT RECEIPT	33 105 56 406 588	4209 29 102 50 216 412	6489 5009 29 105 52 292 462 688

Resultados de complejidad de los modelos obtenidos mediante PACH.

Contenidos

- Introducción
- 2 Nociones preliminares
- 3 Desarrollo
- 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!