The Identification of Dynamics Plans

Modelos Gráficos Causales

Sebastián Bejos

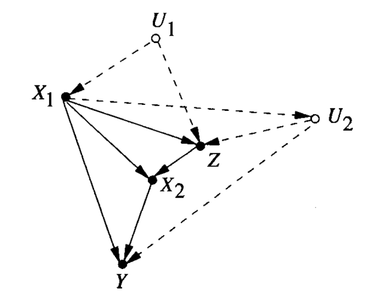
8 de julio de 2019

Instituto Nacional de Astrofísica Óptica y Electrónica

Resumen

- **Evaluación probabilística de planes en presencia de variables ocultas.**
    - Cada plan consta de varias acciones concurrentes o secuenciales.
    - Cada acción puede ser influenciada por sus acciones predecesores en el plan.
- Se establece un **criterio gráfico** para reconocer cuándo se pueden **predecir los efectos de un plan** dado a partir de observaciones pasivas sobre las variables medidas.
- Cuando se cumple el criterio, se proporciona una expresión de forma cerrada para la **probabilidad de que el plan logre un objetivo específico**.

Motivación



- X_1 y X_2 son **tratamientos prescritos** por un médico para un paciente en **dos momentos distintos**.
- Z representa **observaciones que el segundo médico** consulta para determinar X_2. Y representa la **supervivencia del paciente**.
- Las variables ocultas U_1 y U_2 representan, respectivamente, **parte de la historia del paciente** y la **disposición del paciente para recuperarse**.

Identificación de Planes: Notación y Suposiciones

Un **problema de control** consiste de un grafo acíclico dirigido (DAG) G con conjunto de vértices V, dividido en cuatro conjuntos separados $V = \{\mathbf{X}, \mathbf{Z}, \mathbf{U}, \mathbf{Y}\}$, donde:

$\mathbf{X}$ = el conjunto de variables de control (intervenciones, tratamientos, etc.);

$\mathbf{Z}$ = el conjunto de variables observadas, a menudo llamadas covariables;

$\mathbf{U}$ = el conjunto de variables no observadas (latentes);

$\mathbf{Y}$ = una variable de resultado.

Identificación de Planes: Notación y Suposiciones

- Se ordenan las **variables de control** $\mathbf{X} = X_1, \ldots, X_n$, de modo que cada X_k no es un descendiente de $X_{k+j}(j > 0)$ en G, y dejamos que el resultado Y sea un descendiente de X_n.
- Sea N_k el conjunto de nodos observados que no son descendiente de ningún elemento en el conjunto $\{X_k, X_{k+1}, \ldots X_n\}$

Definición

Un **plan** *es una secuencia ordenada* $[do(x_1), do(x_2), \ldots do(x_n)]$ *de asignaciones de valores a las variables de control.*

Notación y Suposiciones

Definición

Un **plan condicional** *es una secuencia ordenada* $[do(g_1(z_1)), do(g_2(z_2)), \ldots do(g_n(z_n))]$*, donde cada* g_k *es una función de un conjunto* Z_k *a* X_k*. El soporte* Z_k *de cada función* $g_k(z_k)$ *no debe contener ninguna variable que sea descendientes de* X_k *en* G*.*

Notación y Suposiciones

- El problema es **evaluar un plan** (incondicional) mediante el cálculo de
$$P(y \mid do(x_1), do(x_2), \ldots do(x_n)),$$
que represente el impacto del plan $[do(x_1), do(x_2), \ldots do(x_n)]$ en la variable resultado Y.
- Se dice que la expresión $P(y \mid do(x_1), do(x_2), \ldots do(x_n))$ es **identificable** en G si, para cada asignación $[do(x_1), do(x_2), \ldots do(x_n)]$, la expresión se puede determinar únicamente a partir de la distribución conjunta de los observables $\{\mathbf{X}, Y, \mathbf{Z}\}$.
- Un **problema de control es identificable** siempre que $P(y \mid do(x_1), do(x_2), \ldots do(x_n))$ sea identificable.

El Criterio de Puerta Trasera Secuencial

Teorema

La probabilidad $P(y \mid do(x_1), do(x_2), \ldots do(x_n))$ *es identificable si, para toda* $1 \leq k \leq n$*, existe un conjunto* $\mathbf{Z}_k$ *de covariables que satisfacen las siguientes condiciones:*

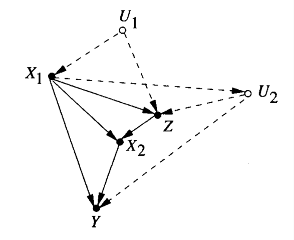
i $\mathbf{Z}_k \subseteq \mathbf{N}_k$ *(i.e.,* $\mathbf{Z}_k$ *consiste de nodescendientes de* $\{X_k, X_{k+1} \ldots X_n\}$*)*

ii $(Y \perp\!\!\!\perp X_k \mid X_1, \ldots, X_{k-1}, Z_1, \ldots, Z_k)_{G_{\underline{X}_k, \overline{X}_{k+1}, \ldots \overline{X}_n}}$.

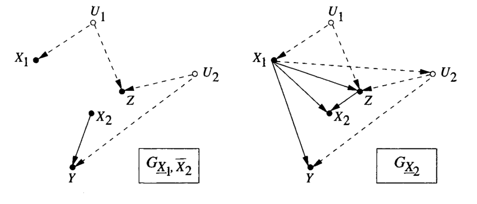
Cuando estas condiciones se satisfacen, el efecto del **plan** *está dado por:*

$$
\begin{aligned}
P(y \mid do(x_1), do(x_2), \ldots do(x_n)) = \sum_{z_1, \ldots, z_n} & P(y \mid z_1, \ldots, z_n, x_1, \ldots, x_n) \\
& \times \prod_{k=1}^{n} P(z_k \mid z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1})
\end{aligned}
$$

Ejemplo



U_1
X_1
U_2
Z
X_2
Y



U_1
X_1
U_2
Z
X_2
Y
$G_{\underline{X}_1, \overline{X}_2}$
U_1
X_1
U_2
Z
X_2
Y
$G_{\underline{X}_2}$

(a) (b)

Secuencia admisible y G-identificabilidad

Definición (Secuencia admisible y G-identificabilidad)
Cualquier secuencia $Z_1, \ldots, Z_n$ de covariables que cumplan las condiciones en (i) y (ii) del teorema 3 se llamará admisible, y cualquier expresión que sea identificable según el criterio del teorema 3 se llamará G-identificable.

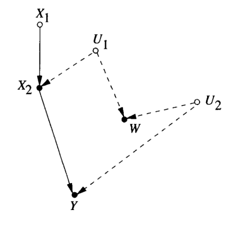


Figura 1: Una opción admisible $Z_1 = W$ que descarta cualquier opción admisible para Z_2. La elección $Z_1 = \emptyset$ permitiría la construcción de una secuencia admisible $(Z_1 = \emptyset, Z_2 = \emptyset)$.

Corolario: Puerta Trasera Secuencial

Corolario

Un problema de control es G-**identificable** *si y solo si este tiene una secuencia admisible.*

Sub-secuencia Admisible Minimal

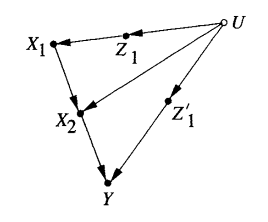


Figura 2: Ejemplo de la no unicidad de los conjuntos mínimos admisibles: Z_1 y Z'_1 son minimales y admisibles, dado que se cumple que $(Y \perp\!\!\!\perp X_1 \mid Z_1)$ y $(Y \perp\!\!\!\perp X_1 \mid Z'_1)$ en $G_{\underline{X_1}\overline{X_2}}$

Sub-secuencia Admisible Minimal

Teorema

Si existe una secuencia admisible $Z_1^*, \ldots Z_n^*$ *entonces, para toda secuencia admisible minimal* $Z_1, \ldots Z_{k-1}$ *de covariables, existe un conjunto admisible* Z_k*.*

Corolario

Un problema de control es G-identificable si y solo si el siguiente algoritmo termina con éxito.

1. *Hacer* $k = 1$
2. *Escoger cualquier minimal* $Z_k \subseteq N_k$ *que satisfaga (ii) del teorema 3.*
3. *Si no existe tal* Z_k *entonces terminar con falla; en otro caso, hacer* $k = k + 1$
4. *Si* $k{=}n{+}1$ *entonces terminar con exito; en otro caso, regresar al paso 2.*

G-identificable

Teorema

La probabilidad $P(y \mid do(x_1), do(x_2), \ldots do(x_n))$ es G-identificable si se satisfacen la siguiente condición para toda $1 \leq k \leq n$:

$$(Y \perp\!\!\!\perp X_k \mid X_1, \ldots, X_{k-1}, W_1, \ldots, W_k)_{G_{\underline{X}_k, \overline{X}_{k+1}, \ldots \overline{X}_n}},$$

donde W_k es el conjunto de covariables en G que son, no decendientes de $\{X_k, X_{k+1}, \ldots, X_n\}$ y tienen a Y o a X_k como descendientes en $G_{\underline{X}_k, \overline{X}_{k+1}, \ldots \overline{X}_n}$. Además, si esta condición se satisface, entonces los planes se evalúan como:

$$P(y \mid do(x_1), do(x_2), \ldots do(x_n)) = \sum_{w_1, \ldots, w_n} P(y \mid w_1, \ldots, w_n, x_1, \ldots, x_n) \times \prod_{k=1}^{n} P(w_k \mid w_1, \ldots, w_{k-1}, x_1, \ldots, x_{k-1})$$