A Linear Non-Gaussian Acyclic Model for Causal Discovery

Modelos Gráficos Causales

Sebastián Bejos

8 de julio de 2019

Instituto Nacional de Astrofísica Óptica y Electrónica

Introducción

En este trabajo se muestra cómo descubrir la **estructura causal completa** de datos de valores continuos, bajo los siguientes supuestos:

(a) El proceso de generación de datos es **lineal**,

(b) No hay **co-factores no observados** y

(c) Las variables de perturbación tienen distribuciones **no-Gaussianas** de varianza no cero.

Introducción

- Cuando se **omite** el supuesto de que los datos siguen una **distribución normal** se logra una **ventaja significativa**.
- Mientras el enfoque **lineal-Gaussiano** generalmente solo conduce a un **conjunto de modelos posibles**, equivalentes en su estructura de correlación condicional.
- Un enfoque **lineal-no-Gaussiano** permite estimar el **modelo causal completo, sin parámetros indeterminados**.

Redes Lineales Causales

Se asume que los datos observados son generados por un proceso con las siguiente propiedades:

- Las variables observadas $x_i, i \in \{1, \ldots, m\}$ se pueden ordenar de tal forma que las variables no causen ninguna variable anterior en el orden.
- Se le llama **orden causal** a tal orden y se denota por $k(i)$.

Redes Lineales Causales

- El valor asignado a cada variable x_i es una **función lineal** de los valores ya asignados a las variables anteriores, más un término 'perturbación' (ruido) e_i, y más un término constante opcional c_i, es decir

$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i + c_i$$

Redes Lineales Causales

- Las perturbaciones e_i son todas variables aleatorias continuas con **distribuciones no Gaussianas** de **varianzas diferente de cero**, y las e_i son **independientes entre sí**, es decir,

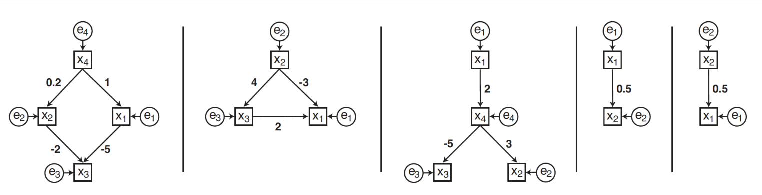
$$P(e_1, \ldots, e_m) = \prod_i P_i(e_i).$$

- Un modelo con estas tres propiedades se le llama un modelo **lineal, no gaussiano, acíclico**, abreviado **LiNGAM**.

Redes Lineales Causales

- Suponemos que podemos observar una gran cantidad de vectores de datos $\boldsymbol{x}$ (que contienen los componentes x_i)
- Cada uno de estos vectores $\boldsymbol{x}$ se genera de acuerdo con el proceso descrito anteriormente:
  1. Con el mismo orden causal $k(i)$,
  2. Los mismos coeficientes b_{ij},
  3. Las mismas constantes c_i y
  4. Las perturbaciones e_i se muestrean independientemente sobre la misma distribución.

Redes Lineales Causales



$$x_4 = e_4$$
$$x_2 = 0.2x_4 + e_2$$
$$x_1 = x_4 + e_1$$
$$x_3 = -2x_2 - 5x_1 + e_3$$

Identificación del Modelo utilizando ICA

- En notación matricial el modelo LiNGAM

$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i$$

puede ser escrito como:

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{e} \tag{1}$$

$$\begin{bmatrix} x_4 \\ x_2 \\ x_1 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -2 & -5 & 0 \end{bmatrix} \begin{bmatrix} x_4 \\ x_2 \\ x_1 \\ x_3 \end{bmatrix} + \begin{bmatrix} e_4 \\ e_2 \\ e_1 \\ e_3 \end{bmatrix}$$

Análisis de Componentes Independientes (ICA)

- El **modelo ICA** para las variables observables X_i con $i = 1, \ldots, d$ se define como sigue:

$$x_i = \sum_{j=1}^{d} a_{ij} s_j, \tag{2}$$

  donde las s_j son variables latentes continuas que son mutuamente independientes entre si.
- Estas variables latentes se les conoce como **componentes independientes** y siguen una distribución no-Gaussiana.

Análisis de Componentes Independientes (ICA)

- **ICA** representa un **proceso de generación de datos**, donde los componentes independientes s_j se suman con los coeficientes a_{ij} y se observan como x_i.
- En forma matricial, la ecuación (3) se puede representar como:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} \tag{3}$$

- Se puede mostrar que $\boldsymbol{A}$ es identificable salvo por el escalamiento y permutaciones sobre las columnas.

Análisis de Componentes Independientes (ICA)

- La matriz de mezcla identificada por los métodos ICA puede escribirse como:

$$\boldsymbol{A}_{ICA} = \boldsymbol{APD},$$

  donde $\boldsymbol{P}$ es una matriz de permutación no conocida y $\boldsymbol{D}$ es matriz diagonal no conocida.

Análisis de Componentes Independientes (ICA)

- La mayoría de los métodos de estimación de ICA estiman una matriz conocida como la matriz de separación $\boldsymbol{W} = \boldsymbol{A}^{-1}$.

$$\boldsymbol{W}_{ICA} = \boldsymbol{PDW} = \boldsymbol{PDA}^{-1}$$

Identificación LiNGAM utilizando ICA

Resolviendo la ecuación $\boldsymbol{x} = \boldsymbol{Bx} + \boldsymbol{e}$ de un LiNGAM para $\boldsymbol{x}$ se obtiene:

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{Bx} + \boldsymbol{e} \\ \boldsymbol{x} - \boldsymbol{Bx} &= \boldsymbol{e} \\ (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{x} &= \boldsymbol{e} \\ \boldsymbol{x} &= (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{e} \end{aligned}$$

Que se puede escribir como:

$$\boldsymbol{x} = \boldsymbol{Ae}, \text{ con } \boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{B})^{-1} \tag{4}$$

Identificación del Modelo utilizando ICA

- Dado que los componentes de $\boldsymbol{e}$ son independientes y no-Gaussianos, $\boldsymbol{x} = \boldsymbol{Ae}$ define un modelo **ICA**, que se sabe que es **identificable**.
- ICA es capaz de estimar $\boldsymbol{A}$ (y $\boldsymbol{W} = \boldsymbol{A}^{-1} = \boldsymbol{I} - \boldsymbol{B}$). Sin embargo, exhibe **indeterminaciones de permutación y escalado**.
- ICA regresa $\boldsymbol{W}_{ICA} = \boldsymbol{PDW}$, donde $\boldsymbol{P}$ es una **matriz de permutación desconocida** y $\boldsymbol{D}$ es una **matriz diagonal desconocida**.

Identificación del Modelo utilizando ICA

- La $\boldsymbol{P}$ correcta es la única que no contiene ceros en la diagonal de DW, dado que B debe ser una matriz que puede permutarse para conviértase una **matriz triangular inferior con ceros en la diagonal** y $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{B}$.
- Además, la **escala correcta** de los componentes independientes se pueden determinar utilizando los **unos en la diagonal de** $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{B}$.
- Para obtener $\boldsymbol{W}$ solo es necesario dividir los renglones de $\boldsymbol{DW}$ por sus elementos diagonales correspondientes.
- Finalmente, se puede calcular la **matriz de fuerza de conexión** $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{W}$.

Algoritmo LiNGAM ICA

Algoritmo LiNGAM ICA:

1. Dado un vector aleatorio d-dimensional $\boldsymbol{x}$ y su matriz $d \times n$ dimensional de datos observados, aplicar un algoritmo ICA para obtener una estimación de $\boldsymbol{A}$.
2. Encontrar la permutación única de las filas de $\boldsymbol{W} = \boldsymbol{A}^{-1}$ que produce una matriz $\tilde{\boldsymbol{W}}$ sin ceros en la diagonal principal.
3. Dividir cada fila de $\tilde{\boldsymbol{W}}$ por su elemento diagonal correspondiente para obtener una nueva matriz $\tilde{\boldsymbol{W}}'$ con una diagonal que consiste completamente de unos .
4. Calcular una estimación $\hat{\boldsymbol{B}}$ de $\boldsymbol{B}$ utilizando $\hat{\boldsymbol{B}} = \boldsymbol{I} - \tilde{\boldsymbol{W}}'$.
5. Finalmente, para estimar un orden causal $k(i)$, determinar la matriz de permutación $\tilde{\boldsymbol{P}}$ de $\hat{\boldsymbol{B}}$, obteniendo la matriz $\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{P}}\hat{\boldsymbol{B}}\tilde{\boldsymbol{P}}^T$ que esté lo más cerca posible a una estructura triangular inferior.