Redes Bayesianas Causales

Modelos Gráficos Causales

Sebastián Bejos

8 de julio de 2019

Instituto Nacional de Astrofísica Óptica y Electrónica

Introducción

Las ventajas de construir modelos (basados en un representación por medio de DAGs) sobre información causal por sobre información asociativa son varias:

1. Los juicios para la construcción del modelo son más significativos, más accesibles y, por lo tanto, más confiables.
2. Permiten representar y responder a cambios externos o espontáneos. Cualquier reconfiguración local de los mecanismos en el entorno se puede traducir, con solo una pequeña modificación, en una reconfiguración isomórfica de la topología de la red (DAG). (Agentes deliberativos vs reactivos)

Redes Causales como Oráculos para Intervenciones

- Esta flexibilidad se basa en el supuesto de que cada relación padre-hijo en la red representa un mecanismo físico estable y autónomo, en otras palabras, que es posible cambiar una de esas relaciones sin cambiar las demás.
- Organizar el conocimiento con estas configuraciones modulares, permite predecir el efecto de las intervenciones externas con un mínimo de información adicional.

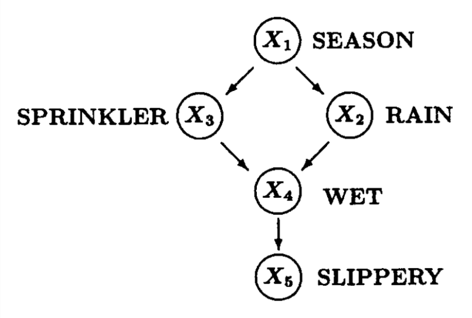
Redes Causales como Oráculos para Intervenciones

- Una distribución conjunta nos dice que tan probable son los eventos y como cambiarán las probabilidades con observaciones posteriores.
- Un modelo causal, además nos dice cómo cambiarían estas probabilidades como resultado de intervenciones externas.

Redes Causales como Oráculos para Intervenciones

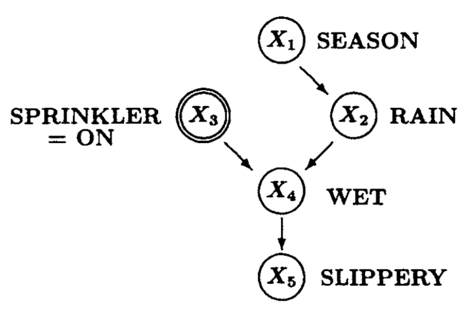
- En lugar de especificar una nueva función de probabilidad para cada una de las muchas intervenciones posibles.
- El efecto general de una intervención se puede predecir modificando la factorización de la distribución conjunta y utilizando el producto modificado se calcula una nueva función de probabilidad.

Redes Causales como Oráculos para Intervenciones



$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1)P(x_4 \mid x_2, x_3)P(x_5 \mid x_4)$$

Redes Causales como Oráculos para Intervenciones



$$P_{X_3=On}(x_1, x_2, x_4, x_5) = P(x_1)P(x_2 \mid x_1)P(x_4 \mid x_2, X_3 = On)P(x_5 \mid x_4)$$

Definición

Sea $P(\mathbf{v})$ *una distribución de probabilidad sobre un conjunto de variables* $\mathbf{V}$*, y sea* $P_{\mathbf{x}}(\mathbf{v})$ *la distribución resultante de la intervención* $do(\mathbf{X} = \mathbf{x})$ *que establece un subconjunto de variables* $\mathbf{X}$ *como contantes* $\mathbf{x}$*. Denotemos por* $\mathbf{P}*$ *al conjunto de todas las distribuciones intervencionales* $P_{\mathbf{x}}(\mathbf{v})$*,* $\mathbf{X} \subseteq \mathbf{V}$*, incluyendo a* $P(\mathbf{v})$*, que representa ninguna la intervención (i.e.,* $\mathbf{X} = \emptyset$*). Una DAG* G *es una* **Red Bayesina Causal** *compatible con* $\mathbf{P}*$ *si y solo si, se cumplen la siguientes tres condiciones para toda* $P_{\mathbf{X}} \in \mathbf{P}*$*:*

(i) $P_{\mathbf{X}}(\mathbf{v})$ *es Markov relativa a* G*.*

(ii) $P_{\mathbf{X}}(\mathbf{v}_i) = 1$ *para toda* $\mathbf{V}_i \in \mathbf{X}$ *cada vez que* $\mathbf{v}_i$ *es consistente con* $\mathbf{X} = \mathbf{x}$*.*

(iii) $P_{\mathbf{x}}(\mathbf{v}_i \mid \mathbf{pa}_i) = P(\mathbf{v}_i \mid \mathbf{pa}_i)$ *para toda* $\mathbf{V}_i \notin \mathbf{X}$ *cada vez que* $\mathbf{pa}_i$ *sea consistente con* $\mathbf{X} = \mathbf{x}$*, i.e., cada* $P(\mathbf{v}_i \mid \mathbf{pa}_i)$ *permanezca invariante a intervenciones que no involucren a* $\mathbf{V}_i$*.*

Redes Bayesianas Causales

- La definición anterior, impone restricciones en el espacio intervencionista $\mathbf{P}*$ que nos permite codificar económicamente este vasto espacio, en forma de una única red bayesiana G.
- Estas restricciones nos permiten calcular la distribución $P_X(v)$ resultante de cualquier intervención $do(X = x)$ como una factorización truncada.

$$P_X(v) = \prod_{\{i | V_i \notin X\}} P(v_i \mid pa_i), \text{para toda } v \text{ consistente con } x,$$

la cual se sigue (e implica) de las condiciones (i)-(iii).

Redes Bayesianas Causales

Si G es una red bayesiana causal con respecto a $P*$ se tienen la siguiente propiedades:

- **Propiedad 1** Para toda i, $P(v_i \mid pa_i) = P_{pa_i}(v_i)$.
  Esta propiedad hace que cada conjunto de padres PA_i sea exógeno en relación con su hijo V_i, asegurando que la probabilidad condicional $P(v_i \mid pa_i)$ coincida con el efecto (sobre V_i) de establecer PA_i en pa_i por control externo.
- **Propiedad 2** Para toda i y para todo sunconjunto S de variables, disjunto de $\{V_i, PA_i\}$, se tiene $P_{pa_i,S}(v_i) = P_{pa_i(v_i)}$.
  Esta propiedad expresa la noción de invariancia; una vez que controlamos sus causas directas PA_i, ninguna otra intervención afectará la probabilidad de V_i.