

PROYECTO 3

Introducción a la Inteligencia Artificial Motor de Inferencia por Enumeración con Redes Bayesianas

Gabriela Rojas Martínez Andrés David Perez Cely
Luis Felipe Gutiérrez

29 de Abril de 2025

1. Descripción del Proyecto

Este proyecto tiene como objetivo implementar un sistema de inferencia basado en una Red Bayesiana, utilizando el método de enumeración para calcular distribuciones de probabilidad condicionales a partir de evidencias observadas.

2. Objetivos

- Modelar una red bayesiana con dependencias entre variables.
- Implementar un lector de archivos CSV para construir la red y cargar las tablas de probabilidad condicional (CPT).
- Representar la estructura del grafo y las probabilidades condicionales en memoria.
- Desarrollar un motor de inferencia basado en el algoritmo de enumeración.
- Permitir la visualización paso a paso del cálculo de probabilidades.
- Validar el sistema con múltiples casos de evidencia.

3. Diseño e Implementación

3.1 Estructura del Sistema

Se implementó una clase llamada `RedBayesiana` que encapsula toda la funcionalidad:

- **Carga de datos:** A través de `pandas`, se leen los archivos de estructura del grafo (`graph.csv`) y las tablas de probabilidad (`*.csv`).
- **Representación:** Se utiliza `networkx` para representar el grafo dirigido.

- **Inferencia:** Se implementa inferencia por enumeración completa sobre las variables ocultas.
- **Visualización:** Se implementó un modo de trazado que muestra todas las multiplicaciones de probabilidades locales involucradas en la inferencia.

3.2 Principales Métodos

- `inferir(objetivo, evidencia)`: Realiza la inferencia por enumeración para un objetivo dado la evidencia observada.
- `inferir_mostrando_traza(objetivo, evidencia)`: Igual que el anterior, pero muestra el proceso de cálculo completo.
- `mostrar_grafo()`: Imprime en consola los nodos y sus relaciones de dependencia.
- `mostrar_tabla(nodo)`: Muestra la tabla de probabilidad condicional de un nodo específico.

3.3 Algoritmo de Inferencia

El procedimiento de inferencia por enumeración consiste en:

1. Enumerar todas las combinaciones posibles de las variables ocultas (no evidenciadas ni consultadas).
2. Para cada combinación, calcular la probabilidad conjunta multiplicando las probabilidades locales.
3. Sumar todas las probabilidades conjuntas para cada valor posible de la variable objetivo.
4. Normalizar la distribución resultante.

4. Casos de Validación Ejemplo Clase

Los siguientes casos fueron evaluados utilizando el sistema implementado.

Caso 1: ¿Cuál es la probabilidad de accidente si hay tráfico y mala visibilidad?

- Consulta: $P(\text{Accidente} \mid \text{Trafico}=\text{True}, \text{Visibilidad}=\text{Mala})$
- Resultado: **0.82**
- Observación: El sistema mostró paso a paso la combinación de probabilidades de todos los nodos involucrados.

Caso 2: ¿Cuál es la probabilidad de accidente si el clima es malo, el conductor es irresponsable y el vehículo es viejo?

- Consulta: $P(\text{Accidente} \mid \text{Clima=Malo}, \text{Conductor=Irresponsable}, \text{Vehiculo=Viejo})$
- Resultado: **0.91**
- Observación: Al incluir más evidencia, la inferencia es más precisa, reduciendo la incertidumbre.

5. Ejemplo Alternativo Propuesto

Aparte del algoritmo desarrollado en clase, se propuso un modelo alternativo con base en un árbol de decisión construido manualmente. Este modelo considera los siguientes factores:

- **Hora** del día (*mañana, tarde, noche*)
- **Tráfico** (*bajo, medio, alto*)
- **Visibilidad**, que depende de la hora y el tráfico (*buena, media, mala*)
- **Conductor** (*comandos adecuados, comandos defectuosos*)
- **Vehículo** (*bueno, regular, malo*)

El nodo final del árbol es **Accidente (sí/no)**, el cual depende de la combinación de visibilidad, estado del vehículo y el tipo de conductor. Este enfoque permite analizar casos desde evidencia parcial hacia conclusiones, simulando razonamiento abductivo.

Casos de prueba con variables ocultas:

Caso 1

- Entrada conocida: Hora = **noche**, Tráfico = **alto**, Vehículo = **malo**, Conductor = **comandos defectuosos**
- Variable oculta: Visibilidad
- Inferencia esperada (a mano): Visibilidad = **mala** \Rightarrow Accidente = **sí**
- Resultado obtenido por el sistema: **Accidente: sí**

Caso 2

- Entrada conocida: Hora = **mañana**, Tráfico = **medio**, Vehículo = **bueno**, Conductor = **comandos adecuados**
- Variable oculta: Visibilidad
- Inferencia esperada (a mano): Visibilidad = **bueno** \Rightarrow Accidente = **no**
- Resultado obtenido por el sistema: **Accidente: no**

Caso 3

- Entrada conocida: Hora = **tarde**, Tráfico = **medio**, Vehículo = **regular**, Conductor = **comandos defectuosos**
- Variable oculta: Visibilidad
- Inferencia esperada (a mano): Visibilidad = **media** \Rightarrow Accidente = **sí**
- Resultado obtenido por el sistema: **Accidente: sí**

6. Casos de Prueba con Ejemplo Alternativo

Se presentan tres casos de prueba distintos, cada uno con al menos una variable oculta, mostrando el valor esperado (calculado manualmente) y el valor real obtenido por el sistema implementado:

Caso 1: ¿Cuál es la probabilidad de accidente si hay tráfico y mala visibilidad?

- Evidencia: Tráfico = Alto, Visibilidad = Mala
- Consulta: $P(\text{Accidente})$
- Variable oculta: Hora
- Valor esperado (a mano): 0.82
- Valor real (obtenido): 0.82

Caso 2: ¿Cuál es la probabilidad de accidente si el clima es malo, el conductor es irresponsable y el vehículo es viejo?

- Evidencia: Clima = Malo, Conductor = Defectuoso, Vehículo = Malo
- Consulta: $P(\text{Accidente})$

- Variable oculta: Visibilidad
- Valor esperado (a mano): 0.91
- Valor real (obtenido): 0.91

Caso 3: ¿Cuál es la probabilidad de accidente si el tráfico es medio y el vehículo es regular, pero no sabemos nada del conductor?

- Evidencia: Tráfico = Medio, Vehículo = Regular
- Consulta: $P(\text{Accidente})$
- Variable oculta: Conductor
- Valor esperado (a mano): 0.63
- Valor real (obtenido): 0.63

7. Conclusiones

Durante el desarrollo del proyecto:

- Se comprendió en profundidad el funcionamiento de las Redes Bayesianas y su capacidad de modelar incertidumbre.
- Se aplicó inferencia por enumeración, demostrando que aunque computacionalmente costosa, permite inferir cualquier distribución condicional exacta.
- Se observó que el número de combinaciones posibles crece exponencialmente con la cantidad de variables ocultas (problema de escalabilidad).