



Universidad
de Huelva



Actividad nº3. Ejercicio (inventado) de Red Bayesiana, aplicada a un diagnóstico médico.

Ejercicio de Sistema experto probabilístico tipo Red Bayesiana..

Inteligencia Artificial - Grado en Ingeniería Informática – 2019/2020

Estudiante: Borja López Pineda

Profesor: Luis Ignacio López Gómez

Índice

Resumen	2
1.-Introducción	2
2.- Descripción y utilidad de las redes bayesianas	3
3.- Red de diagnóstico médico	3
4.- Ejemplos	5
Anexo I. Automatización	8

Resumen

En este documento se realizará un análisis teórico de las redes bayesianas y se explotarán las capacidades de esta herramienta para construir un sistema de diagnóstico médico probabilístico.

Se aporta una red de pruebas y diversos ejemplos resueltos y explicados.

Introducción

Dentro del marco teórico de la probabilística, las redes bayesianas simplifican en gran medida los complejos modelos que, de otra forma, tendrían un coste computacional elevado.

Estas redes pueden ser construidas desde datos estadísticos y utilizarse para la toma de decisiones, creando así un sistema experto.

En nuestro caso buscamos una red que devuelva una probabilidad por cada enfermedad dados unos síntomas y unas causas o antecedentes. De la misma forma que un médico veterano en su materia identificaría la causa probable de los síntomas en base a su experiencia, nuestra red permitirá realizar inferencia sobre los datos conocidos y buscar la enfermedad más probable.

Descripción y utilidad de las redes bayesianas

Una red bayesiana es un grafo dónde los vértices (nodos) representan variables aleatorias y las aristas (uniones) representan dependencia condicional. Son grafos acíclicos y dirigidos. El sentido de la arista apunta al nodo condicionado.

A los nodos se les asocia una función de probabilidad con un valor para cada posible combinación de valores de los nodos de los que depende.

En su base encontramos el teorema de Bayes que sustenta el cálculo de la probabilidad condicionada.

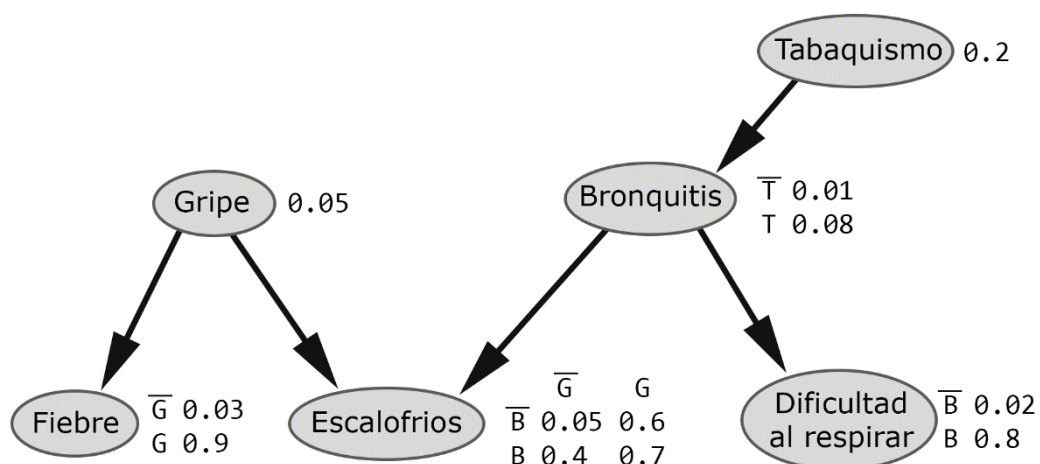
La mejora en eficiencia de estas redes reside en limitar la dependencia de las variables. En el mundo real todos los actos están condicionados, pero es viable realizar un modelo que solo contemple las relaciones más significativas. Por las propiedades de la D-Separación, no necesitamos tener en cuenta todas las variables.

Existen modelos con variables aleatorias continuas, aunque están limitados. Comúnmente se emplean variables discretas, en la red de prueba de este documento se utilizarán variables booleanas concretamente.

De entre los usos que reciben las redes bayesianas destacan: motores de búsqueda y el procesamiento de imágenes, así como algunos campos de la ingeniería, medicina y biología.

Red de diagnóstico médico

A partir de algunos datos sobre enfermedades comunes, he elaborado la siguiente red:



Las causas son: Tabaquismo

Las enfermedades son: Gripe y Bronquitis

Los síntomas son: Fiebre, Escalofríos y Dificultad al respirar

De las relaciones representadas en el árbol y las probabilidades asociadas a cada estado, se extraen las siguientes evidencias:

$$P(T) = 0.2$$

$$P(G) = 0.05$$

$$P(B|\bar{T}) = 0.01 ; P(B|T) = 0.08$$

$$P(F|\bar{G}) = 0.03 ; P(F|G) = 0.9$$

$$P(E|\bar{B}, \bar{G}) = 0.05 ; P(E|B, \bar{G}) = 0.4 ; P(E|\bar{B}, G) = 0.6 ; P(B, G) = 0.7$$

$$P(D|\bar{B}) = 0.02 ; P(D|B) = 0.8$$

Con estos datos se podemos calcular la probabilidad de cualquier estado condicionado a cualquier circunstancia.

Esta es una pequeña red de demostración muy simplificada y sin fundamento estadístico real, aún así intenta ser lo más fiel posible. Los síntomas pueden manifestarse sin necesidad de darse ninguna de las enfermedades, pero es poco probable y no siempre que se contrae una enfermedad, se manifiestan todos sus síntomas, pero suele ser lo normal.

Por ejemplo: es más probable que una persona con bronquitis tenga dificultad para respirar, pero puede haber personas con bronquitis sin problemas para respirar o personas con problemas para respirar que no se deban a una bronquitis.

Las causas o antecedentes son otros datos que un experto tendría en cuenta en su criterio. Por ejemplo: es más probable que un fumador desarrolle bronquitis, que un no fumador.

Todos los nodos que no son enfermedades corresponden a preguntas o pruebas que realizaría el experto (variables observadas) y las enfermedades son el objetivo que se obtendrá mediante inferencia. (variables ocultas)

Ejemplos

Se han realizado tres ejemplos, de cada uno se aporta un enunciado, un desarrollo matemático, una comprobación por ordenador y una breve descripción de lo que se pretende demostrar.

Ejemplo 1

La probabilidad de que una persona tenga gripe sabiendo que presenta los síntomas fiebre y escalofríos y no es fumador. No sabemos si tiene bronquitis.

$$P(G|F,E) = \frac{P(G,F,E)}{P(F,E)} = 0.9422$$

$$P(F,E) = P(G,F,E) + (\bar{G},F,E)$$

$$P(G,F,E) = P(G) \cdot P(F|G) \cdot P(E|G)$$

$$P(\bar{G},F,E) = P(\bar{G}) \cdot P(F|\bar{G}) \cdot P(E|\bar{G})$$

$$P(E|G) = P(E,G,B) + P(E,G,\bar{B}) = P(E|G,B) \cdot P(B) + P(E|G,\bar{B}) \cdot P(\bar{B})$$

$$P(E|\bar{G}) = P(E,\bar{G},B) + P(E,\bar{G},\bar{B}) = P(E|\bar{G},B) \cdot P(B) + P(E|\bar{G},\bar{B}) \cdot P(\bar{B})$$

$$P(B) = P(B,T) + P(B,\bar{T}) = P(B|T) \cdot P(T) + P(B|\bar{T}) \cdot P(\bar{T})$$

La probabilidad de que el paciente tenga gripe es del 94.22%.

$$p(G=1 \mid F=1, E=1) = 0.942153$$

Aquí se muestra un diagnóstico realizado a un paciente con algunos datos desconocidos. Mediante inferencia obtenemos una probabilidad bastante alta de que se trate de gripe. Si se tuviera que tomar una decisión, habiendo revisado antes las otras opciones, la gripe es un candidato muy probable.

Ejemplo 2

La probabilidad de que una persona tenga bronquitis frente a que tenga gripe si no es fumadora, pero presenta dificultad para respirar y escalofríos y no tiene fiebre. Considerando imposible tener las dos a la vez.

$$P(B|\bar{G}, D, E, \bar{F}, \bar{T}) = \frac{P(B, \bar{G}, D, E, \bar{F}, \bar{T})}{P(\bar{G}, D, E, \bar{F}, \bar{T})} = 0.7637$$

$$P(\bar{G}, D, E, \bar{F}, \bar{T}) = P(B, \bar{G}, D, E, \bar{F}, \bar{T}) + P(\bar{B}, \bar{G}, D, E, \bar{F}, \bar{T})$$

$$P(B, \bar{G}, D, E, \bar{F}, \bar{T}) = P(B|\bar{T}) \cdot P(\bar{G}) \cdot P(D|B) \cdot P(E|\bar{G}, B) \cdot P(\bar{F}|\bar{G}) \cdot P(\bar{T})$$

$$P(\bar{B}, \bar{G}, D, E, \bar{F}, \bar{T}) = P(\bar{B}|\bar{T}) \cdot P(\bar{G}) \cdot P(D|\bar{B}) \cdot P(E|\bar{G}, \bar{B}) \cdot P(\bar{F}|\bar{G}) \cdot P(\bar{T})$$

$$P(G|\bar{B}, D, E, \bar{F}, \bar{T}) = \frac{P(G, \bar{B}, D, E, \bar{F}, \bar{T})}{P(\bar{B}, D, E, \bar{F}, \bar{T})} = 0.06113$$

$$P(\bar{B}, D, E, \bar{F}, \bar{T}) = P(G, \bar{B}, D, E, \bar{F}, \bar{T}) + P(\bar{G}, \bar{B}, D, E, \bar{F}, \bar{T})$$

$$P(G, \bar{B}, D, E, \bar{F}, \bar{T}) = P(G) \cdot P(\bar{B}|\bar{T}) \cdot P(D|\bar{B}) \cdot P(E|G, \bar{B}) \cdot P(\bar{F}|G) \cdot P(\bar{T})$$

$$P(\bar{G}, \bar{B}, D, E, \bar{F}, \bar{T}) = P(\bar{G}) \cdot P(\bar{B}|\bar{T}) \cdot P(D|\bar{B}) \cdot P(E|\bar{G}, \bar{B}) \cdot P(\bar{F}|\bar{G}) \cdot P(\bar{T})$$

La probabilidad de que el paciente tenga bronquitis es del 76.37%, mientras que la probabilidad de que tenga gripe es del 6.11%.

$$p(B=1 \mid G=0 \ D=1 \ E=1 \ F=0 \ T=0) = 0.763723$$

$$p(G=1 \mid B=0 \ D=1 \ E=1 \ F=0 \ T=0) = 0.0611309$$

En este caso conocemos todas las variables excepto las enfermedades. Calculamos la probabilidad de que se den las dos de forma independiente y obtenemos que la bronquitis es mucha más probable que la gripe. No hemos tenido en cuenta la posibilidad de que el paciente no tenga ninguna enfermedad o que tenga las dos.

Ejemplo 3

La probabilidad de que una persona sea fumadora, sabiendo únicamente que tiene bronquitis.

$$P(T|B) = \frac{P(T, B)}{P(B)} = 0.666667$$
$$P(B) = P(T, B) + P(\bar{T}, B)$$
$$P(T, B) = P(T) \cdot P(B|T)$$
$$P(\bar{T}, B) = P(\bar{T}) \cdot P(B|\bar{T})$$

La probabilidad de que una persona sea fumadora sabiendo que tiene bronquitis es del 66.66%.

$$p(T=1 \mid B=1) = 0.666667$$

Ahora hemos utilizado la red sin la intención de realizar un diagnóstico, sino para obtener otro tipo de dato. Sabemos que el Tabaquismo influye en la bronquitis, hay relación entre las personas fumadoras y los enfermos de bronquitis. Aquí hemos obtenido la probabilidad de que un enfermo de bronquitis sea fumador.

En este ejemplo concreto, el teorema de Bayes habría sido suficiente, no se necesitaría una red. La utilidad reside en que ahora podemos escalar este ejemplo para hacer preguntas del tipo, ¿Cuál es la probabilidad de que una persona con dificultad para respirar sea fumadora?, ya que dificultad para respirar y tabaquismo se conectan mediante bronquitis.

Así se podría formar una red extensa con gran cantidad de enfermedades, síntomas y causas, rellena de datos estadísticos reales y, en teoría, debería ser capaz de diagnosticar pacientes.

Anexo I. Automatización

Esta vez no he realizado la implementación completa de un programa que resuelva el problema planteado debido a su excesiva complejidad.

En su lugar he utilizado la librería Dlib para C++ que incluye todas las herramientas necesarias para crear una red bayesiana y aplicar inferencia.

Se adjunta el código necesario para obtener los resultados de los ejemplos y otro correspondiente con el ejecutable entregado. En ese segundo programa se espera que el usuario introduzca sus síntomas y antecedentes para obtener un diagnóstico.