

# Definición de Redes Bayesianas y sus aplicaciones

**Autor: José Carlos Santiesteban Rojas - 21-06-2012**

<https://vinculando.org/articulos/redes-bayesianas.html>

## Resumen

Cada vez son más numerosas las investigaciones que trabajan con un amplio número de variables donde existen relaciones complejas entre ellas. Las redes bayesianas son herramientas estadísticas surgidas en el campo de la Inteligencia Artificial que permiten afrontar situaciones de investigación con estas características. Una red bayesiana es un modelo gráfico que permite representar las relaciones de dependencia entre un conjunto de variables, mediante el cual se puede dar información interesante en cuanto a cómo se relacionan las variables del dominio, las cuales pueden ser interpretadas en ocasiones como relaciones de causa-efecto.

En el presente trabajo se abordan algunas cuestiones básicas acerca de las redes bayesianas, así como se profundiza en las aplicaciones que poseen las mismas.

## Introducción

Las redes bayesianas proporcionan una representación gráfica para un conjunto de variables aleatorias y para las relaciones existentes entre ellas. La estructura de la red permite especificar la función de probabilidad conjunta de estas variables como el producto de funciones de probabilidad condicionadas, por lo general, más sencillas. Este enfoque representa una buena estrategia para hacer frente a problemas relacionados con la incertidumbre, donde las conclusiones no pueden ser construidas sólo a partir de un conocimiento previo sobre el problema.

Inicialmente, estos modelos eran construidos "a mano" basados en un conocimiento experto, pero en los últimos años se han desarrollado diversas técnicas para aprender a partir de datos, tanto la estructura como los parámetros asociados al modelo.

Las redes bayesianas se encuentran entre los modelos gráficos más populares. La principal diferencia, con respecto a otros modelos, está en que sus arcos son dirigidos y representan dependencia condicional entre las variables. El nombre proviene del hecho que gran parte de la teoría relevante con este tipo de redes se basa en la estadística bayesiana.

## Estudio teórico

### 1. Definición y conceptos básicos

Las redes bayesianas (también conocidas como redes causales probabilísticas, redes causales, sistemas expertos bayesianos, redes de creencia, sistemas expertos probabilísticas o diagramas de influencia) son herramientas estadísticas que representan un conjunto de incertidumbres asociadas sobre la base de las relaciones de independencia condicional que se establecen entre ellas.

Una red bayesiana es un grafo acíclico dirigido en el que cada nodo representa una variable aleatoria que tiene asociada una función de probabilidad condicional. La estructura de la red bayesiana provee información sobre las relaciones de dependencia e independencia condicional existentes entre las variables. Estas relaciones simplifican la representación de la función de probabilidad conjunta como el producto de las funciones de probabilidad condicional de cada variable.

Sea  $U = \{X_1, X_2, \dots, X_n\}$  un conjunto de variables aleatorias. Formalmente, una red bayesiana para  $U$  es un par  $B = \langle G, T \rangle$  en el que:

$G$  es un grafo acíclico dirigido en el que cada nodo representa una de las variables  $X_1, X_2, \dots, X_n$ , y cada arco representa relaciones de dependencia directas entre las variables. La dirección de los arcos indica que la variable "apuntada" por el arco depende de la variable situada en su origen.

$T$  es un conjunto de parámetros que cuantifica la red. Contiene las probabilidades  $P_B(x_i | p_{xi})$  para cada posible valor  $x_i$  de cada variable  $X_i$  y cada posible valor  $p_{xi}$  de  $P_{X_i}$ , donde éste último denota al conjunto de padres de  $X_i$  en  $G$ .

Así, una red bayesiana  $B$  define una distribución de probabilidad conjunta única sobre  $U$  dada por:

Es importante observar que la topología o estructura de la red no sólo proporciona información sobre las dependencias probabilísticas entre las variables, sino también sobre las independencias condicionales de una variable o conjunto de ellas dada otra u otras variables. Cada variable es independiente de las variables que no son descendientes suyas en el grafo, dado el estado de sus variables padre.

La inclusión de las relaciones de independencia en la propia estructura del grafo hace de las redes bayesianas una buena herramienta para representar conocimiento de forma compacta (se reduce el número de parámetros necesarios). Además, proporcionan métodos flexibles de razonamiento basados en la propagación de las probabilidades a lo largo de la red de acuerdo con las leyes de la teoría de la probabilidad.

La figura 2 presenta un ejemplo concreto de una red bayesiana que representa un cierto conocimiento sobre medicina. En este caso los nodos representan enfermedades, síntomas y factores que causan algunas enfermedades. Como ya se ha mencionado, la variable a la que apunta un arco es dependiente de la que está en el origen de éste. Por ejemplo, *fiebre* depende de *tifoidea* y *gripe*.

Figura 2 Ejemplo red bayesiana

Pueden observarse las suposiciones de independencia acertadas por la red, por ejemplo, *reacciones* es condicionalmente independiente de *comida*, *gripe*, *fiebre* y *dolor* (nodos no descendientes de *reacciones*) dado *tifoidea* (su único nodo padre).

Es decir, se observa en la red que:  $P(R | C, T, G, F, D) = P(R | T)$ , donde  $R$  es *reacciones*,  $C$  es *comida*,  $T$  es *tifoidea*,  $G$  es *gripe*,  $F$  es *fiebre*, y  $D$  es *dolor*.

## 1. Inferencia

A partir de una red ya construida, y dados los valores concretos de algunas variables de una instancia, podrían tratar de estimarse los valores de otras variables de la misma instancia aplicando razonamiento probabilístico.

El razonamiento probabilístico sobre las redes bayesianas consiste en propagar los efectos de las evidencias (variables conocidas) a través de la red para conocer las probabilidades a posteriori de las variables desconocidas. De esta forma se puede determinar un valor estimado para dichas variables en función de los valores de probabilidad obtenidos.

Cuando se conocen los valores observados para todas las variables de la red excepto para una, obtener una estimación para ésta es inmediato a partir de la fórmula de la distribución de probabilidad conjunta de la red.

En un caso más general interesaría obtener una estimación del valor de alguna variable dados valores observados para un subconjunto de todas las restantes. En general, una red puede usarse para calcular la distribución de probabilidad para cualquier subconjunto de variables dados los valores de cualquier subconjunto de las restantes.

El mecanismo de inferencia sobre redes bayesianas permite utilizarlas para construir clasificadores. Para ello, se ha de construir una red bayesiana en la que clase y atributos sean las variables interrelacionadas en el grafo. La clase corresponderá a la variable desconocida, objetivo de la inferencia. Dada una instancia cualquiera para la que se conozcan todos sus atributos, la clasificación se efectuará infiriendo sobre el grafo la probabilidad posterior de cada uno de los valores de la clase, y seleccionando aquel valor que maximice dicha probabilidad.

### 1.1 Clasificadores bayesianos

Un clasificador es una función que asigna un valor de un atributo, llamado *clase*, a instancias o ejemplos descritos mediante un conjunto de atributos. Cuando una red bayesiana se construye con la finalidad de predecir el valor de una variable *clase* dada cualquier configuración en el resto de las variables, se tiene un "clasificador bayesiano".

Los clasificadores bayesianos son ampliamente utilizados debido a que presentan ciertas ventajas:

1. Generalmente, son fáciles de construir y de entender.
2. Las inducciones de estos clasificadores son extremadamente rápidas, requiriendo sólo un paso para hacerlo.
3. Es muy robusto considerando atributos irrelevantes.
4. Toma evidencia de muchos atributos para realizar la predicción final.

Un sistema experto puede ser utilizado como clasificador. Así por ejemplo, un sistema experto en un hospital para determinar que para un conjunto de síntomas presentados por un individuo, no es probable que exista cáncer. O bien, el sistema experto para ayuda a la decisión en una entidad financiera aconsejar no otorgar un préstamo a un cliente porque la probabilidad de impago sea muy elevada, a partir de una serie de atributos, fundamentalmente financieros, del mismo.

Un clasificador bayesiano que suele ser bastante preciso a pesar de su simplicidad es el "clasificador bayesiano simple" (del inglés Naive Bayes classifier) (ver figura 3).

La estructura de esta red bayesiana se basa en una fuerte restricción: todos los atributos que describen los casos son independientes entre sí dado el valor de la clase (Figura 3).

O sea, el modelo de este clasificador asume la independencia de los atributos  $x_1, x_2, \dots, x_n$  dada la clase  $y$ .

Figura 3 Estructura de un clasificador bayesiano simple con 4 variables de entrada.

Han sido propuestos otros modelos más sofisticados. Entre ellos, las *redes bayesianas simples aumentadas* (del inglés *Augmented Naïve Bayesian networks (AN)*), que permiten arcos entre los atributos de entrada, de manera que se reduce la fuerte restricción propia de las redes simples.

Para decidir la estructura concreta de una red AN han sido propuestos diversos algoritmos, entre ellos cabe destacar el algoritmo de aprendizaje de *red bayesiana Simple Aumentada en árbol* (del inglés *Tree Augmented Naïve Bayesian network (TAN)*) y el de *red bayesiana Simple Aumentada Estructurada* (del inglés *Structured Augmented Naïve Bayesian network (SAN)*).

El algoritmo TAN construye una red bayesiana con una estructura TAN (Figura 4), es decir, una estructura en la que la variable clase no tiene padres y los atributos de entrada tienen como padres la clase y como máximo otro

atributo más de entrada.

Figura 4 Ejemplo de estructura TAN

Por otra parte, el algoritmo SAN es aun más flexible que TAN, en el sentido de que permite la construcción de estructuras AN menos restrictivas. Estas estructuras, llamadas también *SAN* (Figura 5) se caracterizan porque la clase no tiene padres y los atributos de entrada pueden tener como padres además de la clase, cualquier número de atributos de entrada, siempre que no haya ciclos dirigidos, pues la estructura de una red bayesiana es siempre un Grafo Dirigido Acíclico (GDA).

Figura 5 Ejemplo de estructura SAN

Sin embargo, cuanto más compleja es una estructura, es decir, cuantos más arcos existan en una estructura SAN, mayor es el riesgo de *sobreajuste*, es decir, mayor es el riesgo de que la estructura aprendida clasifique bien los casos usados para el aprendizaje pero tenga una baja eficiencia para casos nuevos. Así, su capacidad de generalización será baja y por tanto el aprendizaje no se puede considerar aceptable.

Para evitar el problema del sobreajuste en estructuras complejas el algoritmo de aprendizaje SAN utiliza un principio inductivo que favorece la creación de estructuras simples. Dicho principio es el llamado Minimización del Riesgo Estructural (del inglés *Structural Risk Minimization*). Así, este principio define un equilibrio entre la calidad de un modelo dado un conjunto de datos y la complejidad del mismo.

## 1.2 Aprendizaje de redes bayesianas

Durante mucho tiempo las redes bayesianas se construyeron a mano a partir del conocimiento de expertos. La pregunta a plantearse a continuación es la siguiente: ¿pueden inducirse a partir de conjuntos de datos clasificadores basados en redes bayesianas? Si no se restringe la topología de las redes inducidas, ¿pueden obtenerse clasificadores mejores que Naive Bayes?

El problema del aprendizaje bayesiano puede describirse informalmente como: dado un conjunto de entrenamiento  $D = \{u_1, u_2, \dots, u_n\}$  de instancias de  $U$ , encuéntrese la red  $B$  que se ajuste mejor a  $D$ .

Típicamente, este problema se divide en dos partes:

- *Aprendizaje estructural*: obtener la estructura de la red.
- *Aprendizaje paramétrico*: conocida la estructura del grafo, obtener las probabilidades correspondientes a cada nodo.

### 2.1.2. Aprendizaje paramétrico

Datos completos

El aprendizaje de los parámetros es simple cuando todas las variables son completamente observables en el conjunto de entrenamiento. El método más común es el llamado estimador de máxima verosimilitud, que consiste sencillamente en estimar las probabilidades deseadas a partir de la frecuencia de los valores de los datos de entrenamiento, de forma análoga a como se hace en Naive Bayes.

La calidad de estas estimaciones dependerá de que exista un número suficiente de datos en la muestra. Cuando esto no es posible se puede cuantificar la incertidumbre existente representándola mediante una distribución de probabilidad, para así considerarla explícitamente en la definición de las probabilidades. Habitualmente se emplean

distribuciones Beta en el caso de variables binarias, y distribuciones Dirichlet para variables multivaluadas. Esta aproximación es útil cuando se cuenta con el apoyo de expertos en el dominio de la aplicación para concretar los valores de los parámetros de las distribuciones.

Si existen variables de tipo continuo la estrategia más habitual es aplicar discretizarlas antes de construir el modelo estructural. Existen algunos modelos de redes bayesianas con variables continuas, pero están limitados a variables gaussianas relacionadas linealmente. Es posible también efectuar la discretización mientras se construye el grafo de la red, si éste se aprende utilizando el principio de longitud de descripción mínima (MDL) como medida de ajuste. Para más información sobre el aprendizaje con MDL véase la sección 2.1.3. Aprendizaje estructural, sobre el principio MDL.

### Datos incompletos

Aparecen mayores dificultades cuando los datos de entrenamiento no están completos. Pueden plantearse dos tipos de información incompleta:

- *Valores faltantes*: faltan algunos valores de uno o varias variables en algunos ejemplos.
- *Nodo oculto*: faltan todos los valores de una variable.

El primer caso es más sencillo, y existen varias alternativas, entre ellas:

- Eliminar los ejemplos con valores ausentes.
- Considerar un nuevo valor adicional para la variable: "desconocido".
- Considerar el valor más probable a partir de los datos de la misma en las demás instancias.
- Considerar el valor más probable en base a las demás variables.

Las dos primeras opciones son habituales en problemas de aprendizaje, y válidas siempre y cuando se cuente con un número elevado de datos completos. La tercera opción viene a ignorar las posibles dependencias de la variable con las demás, cuando ya se cuenta con la estructura que las describe en el grafo; no suele proporcionar los mejores resultados.

La cuarta técnica se sirve de la red ya conocida para inferir los valores desconocidos. Primero se rellenan las tablas de parámetros usando todos los ejemplos completos. Después, para cada instancia incompleta, se asignan los valores conocidos a las variables correspondientes en la red y se propaga su efecto para obtener las probabilidades a posteriori de las no observadas. Entonces se toma como valor observado el más probable y se actualizan todas las probabilidades del modelo antes de procesar la siguiente instancia incompleta.

La aparición de nodos ocultos requiere un tratamiento más complejo. Existen diferentes técnicas para estimar las probabilidades faltantes en este caso. Una habitual es la aplicación del algoritmo EM (Expectation Maximization), cuya aplicación ya se ha estudiado en la asignatura en el contexto de las técnicas de agrupamiento. Su aplicación al aprendizaje de parámetros se traduce en lo siguiente:

Asignar valores aleatorios (o basados en conocimiento experto, si se dispone de él) a las probabilidades desconocidas de la red.

- Utilizar los datos conocidos para estimar desconocidos infiriéndolos sobre el modelo con las probabilidades actuales.
- Completar el conjunto de datos con los valores estimados y volver a calcular las probabilidades de la red a partir de ellos.
- Repetir los dos pasos anteriores hasta que no haya cambios significativos en las probabilidades.

En [RUSSELL] se destaca que existe cierta similitud entre el aprendizaje de la red bayesiana cuando existen nodos ocultos y el aprendizaje de pesos en las capas ocultas de un perceptrón multicapa –en el que se conocen los valores de entrada y salida para cada ejemplo, pero ningún valor para los elementos de proceso intermedios. Basándose en esta idea se propone una técnica de gradiente similar a la usada en el algoritmo de retropropagación. La técnica trata de maximizar la probabilidad de los datos de entrenamiento conocida la hipótesis  $P(D | h)$ , considerando como espacio de hipótesis el conjunto de todas las posibles combinaciones de valores para las probabilidades que parametrizan la red. Para ello, sigue el gradiente de  $\ln P(D | h)$  con respecto a las probabilidades de la red, actualizando cada parámetro  $w_{ijk}$  desconocido de forma iterativa con el incremento:

Figura 6 Parámetro  $w_{ijk}$

donde  $w_{ijk}$  es el parámetro desconocido correspondiente a la probabilidad condicional de que la variable  $X_i$  tome el valor  $x_{ij}$  cuando sus padres  $P_i$  toman los valores  $p_{ik}$ , y  $k$  es una tasa de aprendizaje. En cada iteración las probabilidades  $w_{ijk}$  se renormalizan tras el incremento.

Tanto el algoritmo EM como el de gradiente ascendente encuentran soluciones que son sólo óptimas localmente, por lo que en ambos casos la calidad del resultado dependerá de la asignación inicial de las probabilidades desconocidas.

### 2.1.3. Aprendizaje estructural

El aprendizaje estructural conlleva explorar un espacio de grafos. Esta tarea es muy compleja. A poco que se incrementa el número de variables (nodos), el número de posibles grafos a construir con ellas se dispara. Por eso en muchas ocasiones se restringe el espacio de búsqueda a grafos con características concretas. Existen muchos algoritmos específicos para el aprendizaje de redes donde  $G$  se limita a un árbol, o a un poliárbol, o a otras estructuras menos generales.

No obstante, existen técnicas para aprender redes con estructuras generales. Trabajar sin restricciones debería permitir, idealmente, construir redes que ajusten mejor al conjunto de entrenamiento, por complejas que sean las dependencias entre los atributos.

Hay dos aproximaciones básicas al aprendizaje de redes sin restricciones. La primera de ellas reúne métodos que exploran las relaciones de dependencia existentes entre pares, tripletas u otros subconjuntos de variables para elegir la forma en que deben conectarlas. El estudio de esas relaciones requiere establecer un criterio cuantitativo para medir la dependencia entre variables, y es dicho criterio el que guía la construcción de la red. Un ejemplo de algoritmo que se engloba en esta familia de técnicas es el de construcción de TAN de Friedman y Goldzsmi th.

Dicho algoritmo crea una red con una topología restringida, pero el principio que guía la construcción es el mismo, una medida de información mutua que cuantifica la relación entre las variables.

La otra aproximación habitual al aprendizaje de redes consiste en realizar una búsqueda guiada por una medida global de calidad. Nótese que en la aproximación anterior el criterio guía es local, se aplica a subconjuntos reducidos de variables, no a toda la red. En esta otra aproximación, la operación general consiste en generar distintos grafos mediante un algoritmo de búsqueda y aplicar a cada uno de ellos una función de medida de calidad para decidir qué grafo conservar en cada paso.

Existen muchos algoritmos que siguen esta técnica, definidos a partir de la combinación de dos elementos:

- Algoritmo de búsqueda
- Medida global de ajuste

Es habitual emplear algoritmos de búsqueda heurística. Intentar una búsqueda exhaustiva por todo el espacio de grafos es sencillamente intratable. Algunas posibilidades son las técnicas de ascenso de colinas (*hill climbing*), algoritmos genéticos, búsquedas bidireccionales, etc...

Otra opción es aplicar una búsqueda voraz. Se comienza con una red vacía y se aplican sucesivas operaciones locales mejorando de forma maximal la medida de ajuste hasta que se encuentra un óptimo local. Las operaciones aplicadas incluyen la adición, borrado e inversión de arcos.

También hay muchas medidas de ajuste. Dos habituales son la medida bayesiana y el principio de mínima longitud de descripción.

La medida bayesiana trata de maximizar la probabilidad de la estructura dados los datos de entrenamiento  $P(B_i | D)$ . Como el objetivo de la medida es comparar el valor obtenido para distintas estructuras  $i$  y  $j$ , es habitual recurrir al cociente:

$$P(B_i | D) / P(B_j | D) = P(B_i, D) / P(B_j, D)$$

Considerando variables discretas y datos independientes, las probabilidades conjuntas del segundo cociente se pueden estimar utilizando las predicciones hechas por cada estructura ante los datos de entrenamiento.

Por su parte, el principio MDL caracteriza el aprendizaje en términos de compresión de los datos. El objetivo del aprendizaje es encontrar un modelo que facilite la obtención de la descripción más corta posible de los datos originales. La longitud de esta descripción toma en cuenta:

- La descripción del propio modelo, penalizando la complejidad del mismo.
- La descripción de los datos que usan el modelo, alentando su verosimilitud.

En el contexto de las redes bayesianas, el modelo es la red. Dicha red  $B$  describe la probabilidad condicional  $P_B$  sobre las instancias que aparecen en los datos. Usando esta distribución, puede construirse y codificarse un esquema que asigne palabras de código más cortas a las instancias más probables. De acuerdo con el principio MDL, debería escogerse una red  $B$  tal que la longitud combinada de la descripción de la red y los datos codificados (con respecto a  $P_B$ ) sea mínima.

A partir de este punto, distintos autores definen diferentes formas de medir cada elemento de la descripción a partir del esquema general:

$MDL(B | D) = \text{complejidad}(B) - \text{verosimilitud}(D)$ , en el que un valor menor  $MDL(B | D)$  es mejor.

Por ejemplo, en [FRIEDMAN] se define:

El primer término (complejidad  $(B)$ ) representa la longitud de una descripción de la red en la que se emplean  $\frac{1}{2} \log N$  bits para cada parámetro ( $N$  es el tamaño de  $D$ ), siendo  $|B|$  el número total de éstos. El segundo término mide cuántos bits se necesitan para describir todos los elementos  $d_i$  en  $D$  asignando las longitudes de código en función de la distribución de probabilidad  $P_B$ .

Tanto la medida bayesiana como MDL son bien conocidas y están bien estudiadas. Ambas funciones son asintóticamente equivalentes cuando aumenta el tamaño de la muestra, y además asintóticamente correctas: con probabilidad igual a uno la distribución aprendida converge a la distribución subyacente a medida que el número de muestras aumenta.

## 2. Ventajas

Entre las ventajas de las redes bayesianas se encuentra:

- Permiten representar al unísono la dimensión cualitativa y cuantitativa de un problema en un entorno gráfico inteligible.
- Pueden trabajar con datos perdidos de una manera eficiente, algo que en la práctica es deseable.
- Permiten reducir el sobre ajuste de los datos.
- Posibilitan el descubrimiento de la estructura causal subyacente en un conjunto de datos.
- Representan toda la información en un único formato (probabilístico y gráfico) lo que hace sencillas las interpretaciones, permite retractarse de conclusiones obtenidas con anterioridad ya que no son razonables a la luz de nuevas evidencias, proporcionan una visión general del problema, generan un conjunto de alternativas ordenadas y facilita la explicación de las conclusiones.
- Permiten realizar inferencias bidireccionales; esto es, desde los efectos a las causas y desde las causas a los efectos, etc.

## 3. Aplicaciones

En la actualidad, las redes bayesianas poseen numerables aplicaciones, sobre todo estas se ponen de manifiesto cuando la cantidad de datos manipulados aumenta a velocidades vertiginosas, haciéndose necesario procesarlos e interpretarlos de forma que sea posible extraer el conocimiento preciso para una adecuada toma de decisiones. A continuación se exponen algunos ejemplos de aplicaciones de las redes bayesianas:

- Tutores bayesianos inteligentes

**HYDRIVE:** El sistema HYDRIVE fue desarrollado por los Laboratorios Amstrong de las Fuerzas Aéreas de los Estados Unidos para simular el funcionamiento del avión de combate F-15. Los problemas se presentan en formato de video donde el piloto describe algunas deficiencias en el funcionamiento de un aparato que está aterrizando o que ya ha aterrizado (por ejemplo, el chequeo rutinario del timón de aterrizaje no responde correctamente). La interfaz gráfica permite al estudiante llevar a cabo una tarea de resolución de problemas revisando videos del aparato y actuando sobre ellos.

La red bayesiana generada a partir de HYDRIVE consta de 22 nodos organizados jerárquicamente en cuatro capas. En la cúspide de la jerarquía aparece la ejecución global del usuario. En un nivel inferior se reflejan los tres tipos de conocimientos que el sistema evalúa: conocimiento del sistema, conocimiento estratégico y conocimiento procedimental. En la tercera capa están los subcomponentes de cada tipo de conocimiento. La cuarta capa sirve para recoger los datos de la ejecución del alumno. Todos los arcos están orientados en sentido decreciente desde las capas superiores a las inferiores.

**Andes:** El rasgo que caracteriza a Andes es que no reduce la iniciativa del usuario estableciendo vías estrictas en el aprendizaje. Más bien, es un sistema abierto que permite aprender una habilidad o concepto por medio de diferentes procedimientos. Otra característica importante de Andes es que puede suministrar ayuda dinámica cuando el/la estudiante se encuentra en un callejón sin salida y no sabe como continuar en la resolución de un problema. Se desarrolló a partir de OLAE (acrónimo de la expresión anglosajona Off-Line Assessment of Expertise) y su objetivo es la enseñanza de física newtoniana a nivel universitario.

- Biomedicina

1. Predicción de la supervivencia en cáncer de mama.
2. Multiclasificador de dos niveles para supervivencia en Unidad de Cuidados Intensivos (UCI).



3. Clustering geográfico de la incidencia del cáncer.
4. Predicción de la estructura secundaria de las proteínas.
5. BayesChess: Programa de ajedrez capaz de adaptar su estrategia al usuario al que se enfrenta y de refinar la función de evaluación que guía el proceso de búsqueda en base a su propia experiencia de juego.
6. Investigaciones policiales

Se construyó un sistema informático que predice la probabilidad de robos en casas de una región metropolitana de Birmingham. El mismo está integrado por varios módulos que funcionan controlados por diversas herramientas estadísticas. Sin embargo, el motor de inferencia principal es una red bayesiana que estima la probabilidad de asaltos a casas en función de un conjunto de variables.

- Aplicaciones en empresas
- Microsoft: Answer Wizard (Office), diagnóstico de problemas de impresora, etc.
- Intel: Diagnóstico de fallos de procesadores.
- HP: Diagnóstico de problemas de impresora.
- Nasa: Ayuda a la decisión de misiones espaciales.
- Psicología

El Instituto de Investigación en Salud Mental de Victoria (Melbourne, Australia), trabaja con esta tecnología con el objetivo de comprender, tratar y prevenir trastornos psicológicos como la esquizofrenia, los trastornos del estado de ánimo, etc. De igual manera se piensa puede extenderse hacia la psicología educativa, forense y psicodiagnóstico.

## Conclusiones

La realización de este trabajo permitió abordar temas referentes a las redes bayesianas y sus aplicaciones en la vida real.

Muchos de los problemas cotidianos pueden ser resueltos con esta técnica de la Inteligencia Artificial por su capacidad de inferir conocimientos sobre otros previamente conocidos, su capacidad de aprendizaje la hacen una técnica muy utilizada a nivel mundial.

## Referencias bibliográficas

### [APRENDIZAJE]

- *Aprendizaje automático: conceptos básicos y avanzados: aspectos prácticos utilizando el software Weka. Madrid: Pearson Prentice Hall, D.L. 2006. Coordinador: Sierra Araujo, Basilio.*

### [FERNÁNDEZ]

- *Fernández, Enrique. Análisis de Clasificadores Bayesianos [on-line]. Facultad de Ingeniería de la Universidad de Buenos Aires. Disponible en: <http://www.fi.uba.ar/materias/7550/clasificadores-bayesianos.pdf>*

### [FRIEDMAN]

- *Friedman, Nir; et al. Bayesian Network Classifiers [en línea]. Cop. Kluwer Academic Publishers, Boston.*

[IRIS]

- UCI Machine Learning Repository. Acceso a la base de datos de flor de iris [en- línea].
- <http://www.ics.uci.edu/~mlearn/databases/iris/>

[KEOGH]

- Keogh E. & Pazzani M. (1999). *Learning Augmented Bayesian Classifiers: A Comparison of Distribution- based and Classification- based Approaches. Uncertainty 99, 7th. Int'l Workshop on AI and Statistics, Ft. Lauderdale, Florida, 225- 230. Disponible en: <http://www.ics.uci.edu/~pazzani/Publications/EamonnAIStats.pdf>*

[MITCHELL]

- Mitchell, Tom M. *Machine Learning* . New York [etc.]: MacGraw- Hill, 1997. ISBN- 0070428077.

[SOYBEAN]

- UCI Machine Learning Repository. Acceso a labase de datos de soja [on- line].
- <http://www.ics.uci.edu/~mlearn/databases/soybean/>

## **Bibliografía**

- *Aprendizaje automático: conceptos básicos y avanzados: aspectos prácticos utilizando el software Weka. Madrid: Pearson Prentice Hall, D.L. 2006. Coordinador: Sierra Araujo, Basilio.*
- Fernández, Enrique. *Análisis de Clasificadores Bayesianos [on-line]. Facultad de Ingeniería de la Universidad de Buenos Aires. Disponible en: <http://www.fi.uba.ar/materias/7550/clasificadores-bayesianos.pdf>*
- Friedman, Nir; et al. *Bayesian Network Classifiers [en línea]. Cop. Kluwer Academic Publishers, Boston. Disponible en: [r /Papers/FrGG1.pdf](http://papers.fr.gg1.pdf)*
- UCI Machine Learning Repository. Acceso a la base de datos de flor de iris [en- línea].
- [/databas es /iris/](http://www.ics.uci.edu/~mlearn/databases/iris/)
- Keogh E. & Pazzani M. (1999). *Learning Augmented Bayesian Classifiers: A Comparison of Distribution- based and Classification- based Approaches. Uncertainty 99, 7th. Int'l Workshop on AI and Statistics, Ft. Lauderdale, Florida, 225- 230. Disponible en <http://www.ics.uci.edu/~pazzani/Publications/EamonnAIStats.pdf>*
- Mitchell, Tom M. *Machine Learning. New York [etc.]: MacGraw-Hill, 1997. ISBN- 0070428077.*
- UCI Machine Learning Repository. Acceso a la base de datos de soja [en-línea].
- <http://www.ics.uci.edu/~mlearn/databases/soybean/>
- Jorge López Puga, Juan García García, Leticia de la Fuente Sánchez, Emilia Inmaculada de la Fuente Solana. *Las redes bayesianas como herramientas de modelado en Psicología. [en-línea: 2007].*
- Edwards, W. (1998). *Hailfinder. Tools for and experiences with bayesian normative modeling. American Psychologist.*

**Autores:** Ing. Dianet Utria Pérez.

Ing. José Carlos Santiesteban Rojas.

Ing. Carlos Enrique Hernández Reyes.