



Optimización de Redes Bayesianas basado en técnicas de aprendizaje por inducción

Normandi Rocío Tirado Ríos¹, Freddy Enrique Triana Litardo², Jorge Wilson Saa Saltos³

1 Universidad Técnica Estatal de Quevedo, ntirado@uteq.edu.ec

2 Universidad Técnica Estatal de Quevedo, ftriana@uteq.edu.ec

3 Universidad Técnica Estatal de Quevedo, jsaa@uteq.edu.ec

RESUMEN

Una red bayesiana es un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependencia probabilística; son utilizadas para proveer: una forma compacta de representar el conocimiento y métodos flexibles de razonamiento. El obtener una red bayesiana a partir de datos es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico. En este trabajo se define un método de aprendizaje automático que optimiza las redes bayesianas aplicadas a clasificación mediante la utilización de un método de aprendizaje híbrido que combina las ventajas de las técnicas de inducción de los árboles de decisión (TDIDT - C4.5) con las de las redes bayesianas.

Palabras claves: Redes bayesianas. Aprendizaje por inducción. Clasificación. Sistemas inteligentes híbridos



Optimization of Bayesian Networks based on induction learning techniques

ABSTRACT

A bayesian network is a directed acyclic graph in which each node represents a variable and each arc a probabilistic dependency; they are used to provide: a compact form to represent the knowledge and flexible methods of reasoning. Obtaining a bayesian network from data is a learning process that is divided in two steps: structural learning and parametric learning. In this paper we define an automatic learning method that optimizes the bayesian networks applied to classification using a hybrid method of learning that combines the advantages of the induction techniques of the decision trees (TDIDT - C4.5) with those of the bayesian networks.

Keywords: Bayesian networks. Induction learning. Classification. Hybrid intelligent systems



1. INTRODUCCIÓN

Una red bayesiana es un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependencia probabilística; son utilizadas para proveer: una forma compacta de representar el conocimiento y métodos flexibles de razonamiento. El obtener una red bayesiana a partir de datos es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico. En este trabajo se define un método de aprendizaje automático que optimiza las redes bayesianas aplicadas a clasificación mediante la utilización de un método de aprendizaje híbrido que combina las ventajas de las técnicas de inducción de los árboles de decisión (TDIDT - C4.5) con las de las redes bayesianas.

2. METODOS

Esta investigación partió de una revisión bibliográfica previa sobre las redes bayesianas utilizando Scopus, ello permitió detectar los artículos que pueden considerarse seminales en esta temática. En la segunda etapa de la investigación se procedió a definir un método de aprendizaje automático a partir de un método de aprendizaje híbrido

3. RESULTADOS

Redes bayesianas

Las redes bayesianas o probabilísticas se fundamentan en la teoría de la probabilidad y combinan la potencia del teorema de Bayes con la expresividad semántica de los grafos dirigidos; las mismas permiten representar un modelo causal por medio de una representación gráfica de las independencias / dependencias entre las variables que forman parte del dominio de aplicación [Pearl, 1988].

Una red bayesiana es un grafo acíclico dirigido –las uniones entre los nodos tienen definidas una dirección– en el que los nodos representan variables aleatorias y las flechas



representan influencias causales; el que un nodo sea padre de otro implica que es causa directa del mismo. Se puede interpretar a una red bayesiana de dos formas:

1. Distribución de probabilidad: Representa la distribución de la probabilidad conjunta de las variables representadas en la red.
2. Base de reglas: Cada arco representa un conjunto de reglas que asocian a las variables involucradas. Dichas reglas están cuantificadas por las probabilidades respectivas.

A continuación se describirán los fundamentos teóricos de las redes bayesianas y distintos algoritmos de propagación.

Definición formal de las redes bayesianas

Una red bayesiana es un grafo acíclico dirigido en el que los nodos representan variables aleatorias que pueden ser continuas o discretas; en las siguientes definiciones se utilizarán letras mayúsculas para denotar los nodos (X) y las correspondientes letras minúsculas para designar sus posibles estados (i_x).

Los estados que puede tener una variable deben cumplir con dos propiedades:

1. Ser mutuamente excluyentes, es decir, un nodo sólo puede encontrarse en uno de sus estados en un momento dado.
2. Ser un conjunto exhaustivo, es decir, un nodo no puede tener ningún valor fuera de ese conjunto.

A continuación se indican algunas definiciones y notaciones propias de la terminología de las redes bayesianas:

Nodo

Un nodo X es una variable aleatoria que puede tener varios estados i_x .



La probabilidad de que el nodo X este en el estado x se denotará como $P(x) = P(X = x)$.

• **Arco**

Es la unión entre dos nodos y representa la dependencia entre dos variables del modelo.

Un arco queda definido por un par ordenado de nodos (X,Y) . .

Padre

El nodo X es un padre del nodo Y , si existe un arco (X ,Y) entre los dos nodos.

Probabilidad conjunta

Dado un conjunto de variables {X,Y,K,Z}, la probabilidad conjunta especifica la probabilidad de cada combinación posible de estados de cada variable $P(x_i, y_j, z_k, \dots)$, de manera que se cumple que:

$$\sum_{i,j,\dots,k} P(x_i, y_j, \dots, z_k) = 1$$

• **Probabilidad condicional**

Dadas dos variables X e Y, la probabilidad de que ocurra j y dado que ocurrió el evento i x es la probabilidad condicional de Y dado X y se denota como $P(y_j | x_i)$.

La probabilidad condicional por definición es:

$$P(y_j | x_i) = \frac{P(y_j, x_i)}{P(x_i)}, \text{ dado } P(x_i) > 0$$

Análogamente, si se intercambia el orden de las variables:

$$P(x_i | y_j) = \frac{P(y_j, x_i)}{P(y_j)}$$

A partir de las dos fórmulas anteriores se obtiene:



$$P(y_j | x_i) = \frac{P(y_j)P(x_i | y_j)}{P(x_i)}$$

Esta expresión se conoce como el Teorema de Bayes que en su forma más general es:

Al denominador se lo cor
$$P(y_j | x_i) = \frac{P(y_j)P(x_i | y_j)}{\sum_j P(x_i | y_j)P(y_j)}$$
 lad Total.

En las redes bayesianas el conjunto de valores que componen la probabilidad condicional de un hijo dados sus padres, se representa en las llamadas tablas de probabilidad condicional. ·

Independencia

Dos variables X e Y son independientes si la ocurrencia de una no tiene que ver con la ocurrencia de la otra. Por definición se cumple que Y es independiente de X si y sólo si:

$$P(y_j, x_i) = P(y_j)P(x_i) \forall i, j$$

Esto implica que:

$$P(y_j | x_i) = P(y_j) \forall i, j$$

$$P(x_i | y_j) = P(x_i) \forall i, j$$

• Observación

Es la determinación del estado de un nodo ($X = x$) a partir de un dato obtenido en el exterior del modelo.

• Evidencia



Es el conjunto de observaciones $e = \{X = x, Y = y, K, Z = z\}$ en un momento dado.

• **Probabilidad a priori**

Es la probabilidad de una variable en ausencia de evidencia.

• **Probabilidad a posteriori**

Es la probabilidad de una variable condicionada a la existencia de una determinada evidencia; la probabilidad a posteriori de X cuando se dispone de la evidencia e se calcula como $P(X | e)$.

Representación del conocimiento

Una red bayesiana representa relaciones causales en el dominio del conocimiento a través de una estructura gráfica y las tablas de probabilidad condicional entre los nodos, por lo tanto el conocimiento que representa la red está compuesto por los siguientes elementos:

1. Un conjunto de nodos $\{X_i\}$ que representan cada una de las variables del modelo. Cada una de ellas tiene un conjunto exhaustivo de estados $\{x_i\}$ mutuamente excluyentes.
2. Un conjunto de enlaces o arcos (X_i, X_j) entre aquellos nodos que tienen una relación causal. De esta manera todas las relaciones están explícitamente representadas en el grafo.
3. Una tabla de probabilidad condicional asociada a cada nodo X_i indicando la probabilidad de sus estados para cada combinación de los estados de sus padres. Si un nodo no tiene padres se indican sus probabilidades a priori.

La estructura de una red bayesiana se puede determinar de la siguiente manera:

1. Se asigna un vértice o nodo a cada variable (X_i) y se indica de qué otros vértices es una causa directa; a ese conjunto de vértices “causa del nodo X_i ” se lo denota como el conjunto X_{i-p} y se lo llamará “padres de X_i ”.



2. Se une cada padre con sus hijos con flechas que parten de los padres y llegan a los hijos.

3. A cada variable X_i se le asigna una matriz $(P_{ij})_{X_i}$ $P \times p$ que estima la probabilidad condicional de un evento i $X = x$ dada una combinación de valores de los X_i p .

Una vez que se ha diseñado la estructura de la red y se han especificado todas las tablas de probabilidad condicional se está en condiciones de conocer la probabilidad de una determinada variable dependiendo del estado de cualquier combinación del resto de variables de la red; para ello se debe calcular la probabilidad a posteriori de cada variable condicionada a la evidencia; estas probabilidades a posteriori se podrán obtener de forma inmediata a partir de la probabilidad conjunta de todas las variables

$$P(x_1, x_2, \dots, x_i) \cdot A$$

A continuación se indica cómo este proceso se ve simplificado al aplicar la propiedad de independencia condicional que permite obtener la probabilidad conjunta a partir de las probabilidades condicionales de cada nodo en función de sus padres.

Independencia condicional

Como se indicó anteriormente la topología o estructura de una red bayesiana no sólo representa explícitamente dependencias probabilísticas entre variables, sino que también describe implícitamente las independencias condicionales existentes entre ellas.

La siguiente definición muestra las condiciones que deben darse para que dos variables sean condicionalmente independientes: Una variable X es condicionalmente independiente de otra Y dada una tercer variable Z , si el conocer Z hace que X e Y sean independientes. Es decir, si conozco Z , Y no tiene influencia en X . Esto es: $P(X|Y, Z) = P(X|Z)$ Esta definición se traduce a que cada variable es independiente de todos aquellos nodos que no son sus “descendientes” una vez que se conocen sus propios nodos padres; a lo largo de este trabajo se utilizarán las palabras “nodos” y “variables” como sinónimos. Gráficamente se verifica en los casos en que los nodos X e Y están separados



por Z en el grafo. Esto implica que todos los caminos para ir de X a Y pasarán necesariamente por Z [Pearl, 1988].

Por ejemplo, en la red bayesiana de la figura 2.1, {E} es condicionalmente independiente de {A,C, D,F, G} dado {B}; con lo cual $P(E | A,B,C, D,F,G) = P(E | B)$; esto se conoce como Separación-D.

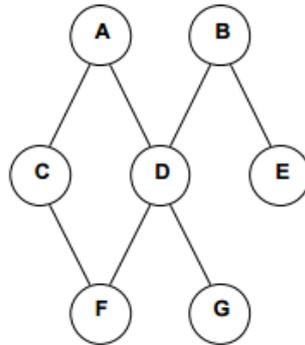


Figura 2.1: Ejemplo de red bayesiana.

Algoritmos de propagación

Una de las ventajas de disponer de una estructura gráfica de las relaciones entre las variables es que se puede utilizar esta información para reducir el número de operaciones necesarias para obtener las probabilidades a posteriori. Existen varios métodos computacionales que aprovechan la estructura gráfica para propagar los efectos que las observaciones del mundo real tienen sobre el resto de las variables de la red; las diferencias entre ellos se basan principalmente en la precisión de los resultados y en el consumo de recursos durante el tiempo de ejecución.

Los algoritmos de propagación se dividen inicialmente en “exactos” o “aproximados” según cómo calculen los valores de las probabilidades. Los métodos exactos calculan los valores por medio del teorema de Bayes mientras que los métodos aproximados utilizan técnicas iterativas de muestreo en las que los valores se aproximarán más o menos a los exactos dependiendo del punto en que se detenga el proceso. Los algoritmos de propagación dependen del tipo de estructura de la red bayesiana, existiendo las siguientes tres topologías de red:

- Árboles (sección 2.2.4.2.1),
- Poliárboles (sección 2.2.4.2.2), y



- Redes multiconectadas (sección 2.2.4.2.3).

Propagación en redes multiconectadas

Una red multiconectada es un grafo no conectado en forma sencilla, es decir, en el que hay múltiples trayectorias entre nodos (MCG); en este tipo de red probabilística los métodos anteriores ya no aplican pero existen otras técnicas alternativas que se detallan a continuación:

- **Condicionamiento:** Al instanciar una variable de la red, ésta bloquea las trayectorias de propagación lo cual implica que si se asumen valores para un grupo seleccionado de variables, es posible descomponer la gráfica en un conjunto de SCG; este método consiste en realizar el proceso de propagación de la evidencia para cada valor posible de dichas variables y luego promediar las probabilidades ponderadas.
- **Simulación estocástica:** Se asignan valores aleatorios a las variables no instanciadas, se calcula la distribución de probabilidad y se obtienen valores de cada variable dando una muestra; se repite el procedimiento para obtener un número apreciable de muestras y en base al número de ocurrencias de cada valor se determina la probabilidad de dicha variable.
- **Agrupamiento:** El método de agrupamiento consiste en transformar la estructura de la red para obtener un árbol mediante agrupación de nodos usando la teoría de grafos [Lauritzen, 1988].

Para ello se parte de la gráfica original y se siguen los siguientes pasos:

1. Se triangulariza el grafo agregando los arcos adicionales necesarios.
2. Se identifican todos los conjuntos de nodos totalmente conectados (cliques).
3. Se ordenan los cliques de forma que todos los nodos comunes estén en un solo clique anterior (su padre).



4. Se construye un nuevo grafo en que cada clique es un nodo formando un árbol de cliques.

Para la propagación de probabilidades se utiliza este árbol de macro nodos (cliques) obteniendo la probabilidad conjunta de cada clique, a partir de la cual se puede obtener la probabilidad individual de cada variable en el clique.

En general, la propagación en una red probabilística con una estructura compleja es un problema de complejidad NP-duro [Cooper, 1990];

sin embargo en muchas aplicaciones prácticas la estructura de la red no es tan compleja y los tiempos de propagación son razonables.

Propagación de mensajes

- Los mensajes π no pueden propagarse a través de nodos instanciados.

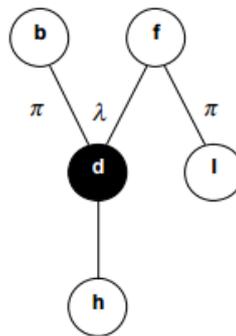


Figura 2.4: “The dog barking problem” – Instanciación del nodo **d**

El aprendizaje en las redes bayesianas

El aprendizaje es una de las características que definen a los sistemas basados en inteligencia artificial porque siendo estrictos se puede afirmar que sin aprendizaje no hay inteligencia; es difícil definir el término “aprendizaje”, pero la mayoría de las autoridades en el campo coinciden en que es una de las características de los sistemas adaptativos que



son capaces de mejorar su comportamiento en función de su experiencia pasada, por ejemplo al resolver problemas similares [Simon, 1983].

El aprendizaje suele ser imprescindible en aquellos sistemas que deben trabajar en entornos desconocidos o zonas de proceso poco frecuentes donde la adquisición de conocimiento de los expertos en una tarea difícil o incluso imposible; los sistemas de aprendizaje son capaces de generar nuevo conocimiento y de ajustar el conocimiento existente. El aprendizaje en la redes bayesianas consiste en definir la red probabilística a partir de datos almacenados en bases de datos en lugar de obtener el conocimiento del experto.

Este tipo de aprendizaje ofrece la posibilidad de inducir la estructura gráfica de la red a partir de los datos observados y de definir las relaciones entre los nodos basándose también en dichos casos; según Pearl [Pearl, 1988] a estas dos fases se las puede denominar respectivamente aprendizaje estructural y aprendizaje paramétrico.

A continuación se resume cada una de estas dos fases:

Aprendizaje estructural: obtiene la estructura de la red bayesiana a partir de bases de datos, es decir, las relaciones de dependencia e independencia entre las variables involucradas.

Las técnicas de aprendizaje estructural dependen del tipo de estructura o topología de la red (árboles, poliárboles o redes multiconectadas). Otra alternativa es combinar conocimiento subjetivo del experto con aprendizaje, para lo cual se parte de la estructura dada por el experto y se la valida y mejora utilizando datos estadísticos. · **Aprendizaje paramétrico:** dada una estructura y las bases de datos, obtiene las probabilidades a priori y condicionales requeridas.

Uno de los principales trabajos en el campo del aprendizaje de redes bayesianas es el de Herkovits y Copper [Herskovits & Copper, 1991]; el requisito principal para poder realizar la tarea de aprendizaje de redes bayesianas a partir de datos es disponer de bases



de datos muy extensas en las que esté especificado el valor de cada variable en cada uno de los casos.

Ventajas de las redes bayesianas

El hecho de que las redes bayesianas constituyan una mezcla de técnicas estadísticas y modelos gráficos les provee un serie de importantes ventajas. En primer lugar, el hecho de que las redes guarden información sobre las dependencias e independencias existentes entre las variables involucradas les permiten manejar situaciones donde exista incertidumbre; por otro lado la presentación gráfica de la red facilita la interpretación y obtención de conclusiones sobre el dominio en estudio por parte de la gente que lo analiza; también, debido a que estas redes combinan relaciones causales con lógica probabilística, permite combinar conocimiento experto con datos (dicho conocimiento experto generalmente viene dado en forma de relaciones de causalidad).

Las redes bayesianas permiten definir modelos y utilizarlos tanto para hacer razonamiento de diagnóstico (pues obtienen las causas más probables dado un conjunto de síntomas), como para hacer razonamiento predictivo (obteniendo la probabilidad de presentar un cierto síntoma suponiendo que existe una causa conocida).

Una de las características de las redes bayesianas es que un mismo nodo puede ser fuente de información u objeto de predicción dependiendo de cuál sea la evidencia disponible.

A continuación se muestran cuáles son las características de estos dos tipos de inferencia utilizando una red bayesiana :

Predicción

Si se supone que es cierto un hecho del mundo real que está representado en la red como un nodo padre, la red puede deducir cuáles serán sus efectos; para ello se debe introducir esta hipótesis en el nodo correspondiente y propagar esta información hacia el resto de los nodos. Este modo de razonamiento es de tipo predictivo y está regido por una



Optimización de Redes Bayesianas basado en técnicas de aprendizaje por inducción

Revista Publicando, 3(9). 2016, 41-60. ISSN 1390-9304

inferencia “deductiva” donde el conocimiento se puede expresar de la forma “si a entonces b” y se cumple que el hecho conocido es “a” y el hecho deducido es “b”.



Interpretación de datos

Las mismas relaciones representadas en la red en forma causal permiten hacer inferencias abductivas donde conocidos los síntomas se puede saber cuáles son sus posibles causas. El conocimiento es el mismo que en el caso anterior: “si a entonces b” pero ahora el hecho conocido es “b” y el hecho abducido es “es posible a”; este modo de razonamiento es el que permite la interpretación de las causas que generan determinados fenómenos.

Como se indicó anteriormente en este capítulo, las redes probabilísticas permiten ejecutar simulaciones efectuando hipótesis sobre cualquiera de los datos e incluso considerando ignorancia en alguna de las variables de entrada.

En el caso de los simuladores matemáticos sólo es posible establecer hipótesis sobre los datos de partida que son introducidos en el modelo como las condiciones iniciales y además es imprescindible que se introduzcan todas las variables de entrada del modelo para poder resolver las ecuaciones.

4. CONCLUSIONES

Como se puede observar todas las gráficas que representan el poder predictivo en función de la cantidad de casos de entrenamiento son crecientes. Este fenómeno se da independientemente del dominio de datos utilizado y del método evaluado (RB-Completa o RB-C4.5).

Del análisis de los resultados obtenidos en la experimentación podemos concluir que el método híbrido de aprendizaje propuesto en esta tesis (RB-C4.5) genera una mejora en el poder predictivo de la red respecto a la obtenida sin realizar el preprocesamiento de las variables (RB-Completa).

En otro aspecto, las RB-C4.5 poseen una cantidad de variables menor (o a lo sumo igual) que las RB-Completa; esta reducción de la cantidad de variables involucradas produce una simplificación en la conceptualización del dominio analizado, la cual trae aparejado dos importantes ventajas; por un lado, facilitan la representación e interpretación del



conocimiento eliminando parámetros que no repercuten de manera directa sobre el objetivo buscado (tarea de clasificación). Por el otro lado, simplifica y optimiza la tarea de razonamiento (propagación de las probabilidades) lo cual conlleva a la mejora de los tiempos de procesamiento.

En suma, basándonos en los resultados experimentales obtenidos concluimos que el método híbrido de aprendizaje propuesto en este trabajo optimiza las configuraciones de las redes bayesianas de tipo poliárbol aplicadas a tareas de clasificación

5. REFERENCIAS BIBLIOGRÁFICAS

Andersen, S.K., Olesen, K.G., Jensen, F. (1989). HUGIN – a Shell for Building Belief Universes for Expert Systems. In Proc. IJCAI, pages 1080-1085.

Beinlich, I.A., Suermondt, H.J., Chavez, R.M., Cooper, G.F. (1989). The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In proceedings of the 2nd European Conference on Artificial Intelligence in Medicine.

Bickmore, Timothy W. (1994). Real-Time Sensor Data Validation. NASA Contractor Report 195295, National Aeronautics and Space Administration.

Blurock, Eduard S. (1996). The ID3 Algorithm. Research Institute for Symbolic Computation, Austria.

Breese, John S., Blake, Russ (1995). Automating Computer Bottleneck Detection with Belief Nets. Proceedings of the Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, CA, pp 36-45.



Canavos, G.C. (1984). Probabilidad y Estadística, Aplicaciones y Métodos. Mc.Graw-Hill.

Carbajo, A., Curto, S., Schweigmann, N. (2003). Distribución espacio-temporal de *Aedes aegypti* (Diptera: Culicidae). Su relación con el ambiente urbano y el riesgo de transmisión del virus dengue en la Ciudad de Buenos Aires. Departamento de Ecología, Genética y Evolución. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Chen, M., Han, J., Yu, P. (1996). Data mining: An overview from database perspective. IEEE Transactions on Knowledge and Data Eng.

Cooper, G.F., Herskovits, E. (1992). A Bayesian Method for the Induction of Probabilistic Networks from Data. In Machine Learning 9, pages 54-62, Kluwer.

Cowell, R., Dawid, A., Lauritzen, S., Spiegelhalter, D. (1990). Probabilistic Networks and Expert Systems. Springer, New York, NY.

Díaz, F., Corchado, J.M. (1999). Rough sets bases learning for bayesian networks. International workshop on objective bayesian methodology, Valencia, Spain.

Díez Vegas, F.J. (1994). Sistema experto bayesiano para ecocardiografía . Tesis doctoral, Universidad Nacional de Educación a Distancia.

Evangelos, S., Han, J. (1996). Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Portland, EE.UU.

Ezawa, Kazuo J., Schuermann, Til (1995). Fraud/Uncollectible Debt Detection Using a



Bayesian Network Based Learning System: A Rare Binary Outcome with Mixed Data Structures. Proceedings of the Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, CA, pp 157-166.

Felgaer, P., Britos, P., Sicre, J., Servetto, A., García-Martínez, R., Perichinsky, G. (2003).

Optimización de redes bayesianas basado en técnicas de aprendizaje por inducción.

IX Congreso Argentino de Ciencias de la Computación. La Plata. Octubre 6 al 10.

Fritz, W., García-Martínez, R., Rama, A., Blanqué, J., Adobatti, R., Sarno, M. (1989).

The

Autonomous Intelligent System. Robotics and Autonomous Systems. Elsevier

Science Publishers. Holanda. Volumen 5. Número 2. Páginas 109-125.

Gallion, R., Clair, D., Sabharwal, C., Bond, W.E. (1993). Dynamic ID3: A Symbolic Learning Algorithm for Many-Valued Attribute Domains. Engineering Education Center, University of Missouri-Rolla, St. Luis, EE.UU.

García-Martínez, R. (1993). Aprendizaje Automático basado en Método Heurístico de Formación y Ponderación de Teorías. Revista Tecnología. Brasil. Volumen 15. Número 1-2. Páginas 159-182.

García-Martínez, R. (1995). Aprendizaje Automático. Enciclopedia Iberoamericana de Psiquiatría. Volumen II (Ed. G. Vidal, R. Alarcón & F. Lolas). Páginas 824-828. Editorial Médica Panamerica. ISBN 950-06-2311-0.

García-Martínez, R. (1997). Sistemas Autónomos. Aprendizaje Automático. 170 páginas. Editorial Nueva Librería. ISBN 950-9088-84-6.

García-Martínez, R., Borrajo, D. (2000). An Integrated Approach of Learning, Planning and Executing. Journal of Intelligent and Robotic Systems. Volumen 29, Número 1, Páginas 47-78. Kluwer Academic Press.

García-Martínez, R., Servente, M., Pasquini, D. (2003). Sistemas Inteligentes. 347 páginas. Editorial Nueva Librería. ISBN 987-1104-05-7.



Gowans, M. (2001). Bayesian Network Toolkit. Department of Computing, Imperial College.

Grosser, H., Britos, P., García-Martínez, R. (2005). Detecting Fraud in Mobile Telephony Using Neural Networks. Lecture Notes in Artificial Intelligence. Volumen 3533, Páginas 613-615.

Han, J. (1999). Data Mining. Urban and Dasgupta (eds.), Encyclopedia of Distributed Computing, Kluwer Academic Publishers.

Pablo Ezequiel Felgaer., Optimización de Redes Bayesianas Basado en Técnicas de Aprendizaje por Inducción Febrero 2005.