“Un Algoritmo para generar Redes Bayesianas a partir de datos estadísticos”.

Nicandro Cruz Ramírez[1]
(ncruz@mixteco.utm.mx)
Universidad Tecnológica de la Mixteca
Km. 2.5 Carretera Acatlima-Huajuapan de León, Oaxaca
Tel. y Fax (91-953) 2-03-99

Manuel Martínez Morales
(mmartine@mia.uv.mx)
Sebastián Camacho # 5, Zona Centro
C.P. 91000
Xalapa, Veracruz
Tel. (91-28) 17-29-57
Fax. (91-28) 17-28-55

Resumen

Una red probabilista es una representación gráfica para manejar incertidumbre en sistemas expertos. Dentro de este campo se tienen dos divisiones en cuanto a la forma de construir el sistema: el enfoque tradicional y el enfoque de aprendizaje. En el enfoque tradicional la determinación de la topología ó estructura de la red y de los parámetros asociados con dicha topología es propuesta por el experto humano. En el enfoque de aprendizaje tanto la topología como los parámetros son determinados a partir de una muestra de datos estadísticos sin la intervención directa del experto humano (en la mayoría de los casos). A la fecha, la búsqueda de esquemas de aprendizaje en redes probabilistas es un campo de investigación abierto y sumamente activo

En este trabajo proponemos un algoritmo original que “aprende” de una muestra de datos estadísticos la topología de la red probabilista a partir de una serie de pruebas estadísticas de independencia condicional. Este algoritmo se aplica a una clase de problemas en los cuales las variables se encuentran divididas en variables dependientes y variables independientes.

Palabras Clave.

Redes Bayesianas, aprendizaje, datos estadísticos, entropía, medidas de información.

[1] Este artículo es producto de la tesis de maestría realizada en la Maestría en Inteligencia Artificial que imparten conjuntamente la Universidad Veracruzana y el Laboratorio Nacional de Informática Avanzada (LANIA) y con el apoyo económico del Consejo Nacional de Ciencia y Tecnología (CONACyT, número de becario 70356).

1.- Antecedentes.

1.1- Sistemas expertos que manejan incertidumbre.

La tarea de un experto humano es la de resolver problemas dentro de un dominio específico de conocimiento. Un sistema experto es un programa basado en el conocimiento que no intenta imitar la estructura de la mente humana ni los mecanismos para inteligencia general sino más bien, llegar a soluciones que sean muy parecidas a las del experto humano.

“Un sistema experto es un programa de computadora que utiliza conocimiento y métodos de inferencia para resolver problemas en un dominio específico del conocimiento y que requieren de una pericia humana significativa”. [Weschler, 1990].

Los primeros sistemas expertos que se construyeron (y los más conocidos) son los sistemas basados en reglas. El conocimiento en estos sistemas está estructurado en forma de reglas del tipo si-entonces (if-then). La lógica clásica sirvió como herramienta para crear estos sistemas teniendo al modus ponens como regla de inferencia principal. Con este enfoque puede modelarse eficientemente el conocimiento categórico.

Sin embargo, el conocimiento humano está expresado generalmente de manera incierta. Un experto resuelve problemas donde el conocimiento que se tiene no es del todo seguro ni tampoco se tienen todos los datos necesarios para la solución del problema, por lo que el experto humano debe hacer algunas suposiciones y entonces así ser capaz de ofrecer soluciones a los problemas

presentados. Debido a que los primeros sistemas expertos fueron los sistemas basados en reglas, éstos fueron los primeros que se adaptaron para la representación de conocimiento incierto.

Para que un sistema experto pueda razonar bajo incertidumbre, se le debe dotar de ciertas capacidades. Muchos han sido los enfoques para incorporar incertidumbre en sistemas expertos, teniendo todos sus detractores y defensores. Algunos de estos enfoques son:

- *factores de certeza*
- *teoría matemática de la evidencia (Dempster-Shafer)*
- *lógica no-monótona*
- *lógica difusa (fuzzy logic)*
- *teoría de la probabilidad*

En años recientes ha habido un notable avance en el uso de la teoría de la probabilidad para el manejo de incertidumbre. Se han creado métodos que tienen su base en la teoría de la probabilidad para este objetivo.

Uno de estos métodos basado en la teoría de la probabilidad representa el conocimiento incierto por medio de una gráfica que a su vez describe relaciones de dependencia/independencia condicional ; a esto se le conoce con varios nombres: **redes causales**, **redes probabilistas**, **redes bayesianas**, **diagramas de influencia modelos gráficos y redes de creencia** [Pearl, 1988; Whittaker, 1990; Edwards, 1995]. El término que usaremos es el de ***redes bayesianas*.**

2.- Concepto de Independencia Condicional.

Antes de entrar de lleno con el tema central de este trabajo (las redes bayesianas), revisaremos primeramente el concepto de independencia condicional que ya habíamos mencionado con anterioridad y que nos servirá para explicar los fundamentos de dichas redes.

A continuación se definen los conceptos de independencia (marginal) e independencia condicional.

Sean *X, Y, Z* variables aleatorias discretas (con un número finito de valores posibles) con funciones de probabilidad marginal *P(x), P(y), P(z)* y funciones de probabilidad conjunta *P(x,y), P(x,z), P(y,z)* respectivamente. Entonces tenemos la siguiente definición:

Definición 1:

X es (marginalmente) independiente de *Y* si y solo si:

$P(x,y) = P(x)\ P(y)$ para todos los pares de valores (x, y).

Equivalentemente; si:

$P(x/y) = P(x)$ para todos los pares de valores (x, y) siempre que $P(y) > 0$.

Donde $P(x/y)$ denota la probabilidad condicional de $X=x$ dado $Y=y$.

Si *X* y *Y* son marginalmente independientes lo denotamos por :

$$X \perp\!\!\!\perp Y$$

Esta segunda versión de la definición nos ayuda a interpretar el significado del concepto de independencia en el contexto de representación del conocimiento en sistemas expertos, ya que nos indica que conocer el valor de la variable Y *no modifica* la función de probabilidad de X (i.e $X \perp\!\!\!\perp Y$). Si el conocimiento de Y modifica la probabilidad de ocurrencia de algún valor cualquiera de X entonces $P(x/y) \neq P(x)$ y diremos que Y y X *no son* independientes (i.e $Y \not\perp\!\!\!\perp X$)

Definición 2:

X es *condicionalmente independiente* de *Y* dado *Z* si y solo si:

$P(x,y/z) = P(x/z)\ P(y/z)$ para todos los valores *x, y, z*.

Equivalentemente; si:

$P(x/y,z) = P(x/z)$ para todos los valores *x, y, z*.

Si *X* y *Y* son condicionalmente independientes dado *Z* lo denotamos por :

$$X \perp\!\!\!\perp Y/Z$$

En una red probabilista los nodos representan variables aleatorias y la ausencia de un arco entre dos nodos (variables), digamos X, Y, representa el hecho de que X y Y sean condicionalmente independientes *dadas todas las otras variables* (nodos) en la red.

Un concepto importante asociado con una función de probabilidad es el de *entropía* ó *información*. Precisamente el algoritmo que se propone en este trabajo se basa en medidas informacionales para generar una red bayesiana. Las medidas de información utilizadas en este trabajo están detalladas en [Martínez-Morales, 1995].

Para determinar independencia (marginal) ó independencia condicional entre variables dada una muestra, se hacen pruebas estadísticas de independencia condicional. La que utilizamos en el algoritmo propuesto es la prueba propuesta originalmente por Kullback [Kullback, 1959] y que luego se ha empleado extensamente la asociación con modelos log-lineales [Edwards, 1995; Agresti, 1984]. Esta prueba está basada en medidas informacionales.

En nuestro caso, todas las probabilidades que necesitamos son calculadas de una muestra de datos.

Hacemos una prueba estadística de independencia condicional para verificar si las diferencias entre cifras son significativas ó no para aceptar ó rechazar la hipótesis de independencia condicional entre variables según sea el caso.

Para realizar una prueba de este tipo, necesitamos calcular las probabilidades condicionales y marginales de las variables en una determinada muestra. Después calculamos un estimador (de la ganancia en información) y así a partir de esta determinar la significancia entre cifras para la determinación de independencia ó no independencia condicional entre las distintas variables del problema que está siendo tratado. En nuestro caso utilizamos la prueba chi-cuadrada que consiste en comparar los datos que se tienen contra lo que esperamos observar si la hipotesis (independencia) fuera verdad.

3.- Definición de una Red Bayesiana.

"Una **red bayesiana** es una gráfica acíclica dirigida(GAD) en la que tenemos:

i) Un conjunto de nodos que representan variables aleatorias y un conjunto de arcos dirigidos entre variables.

ii) Cada variable tiene un conjunto finito de estados.

iii) Para cada nodo existe una función de probabilidad condicional la cual depende de los estados de los nodos padres del nodo considerado.

[Braganca, 1990] [Hugin, 1993].

La ausencia de un arco conectando a dos nodos indica que las variables asociadas con estos dos nodos son ***condicionalmente independientes*** dadas todas las demás variables.

Una red bayesiana tiene al menos un nodo raíz (sin padre alguno) y un nodo terminal (sin hijo alguno).

Ejemplo : La red bayesiana de la figura 3.1 representa las siguientes relaciones de independencia :

$L \perp\!\!\!\perp D/S,E,G$
$L \perp\!\!\!\perp G/S,D,E$
$E \perp\!\!\!\perp G/L,S,D$

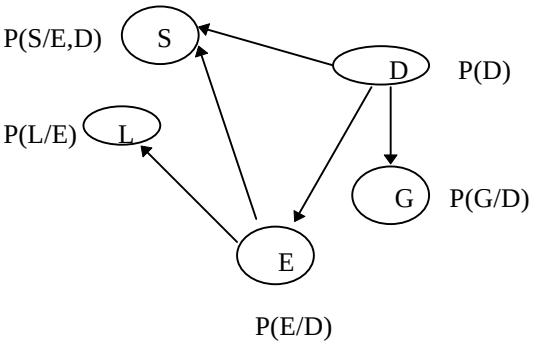

figura 3.1

Las redes bayesianas están siendo utilizadas cada día más para la representación de conocimiento incierto en sistemas expertos. Las áreas de aplicación (por mencionar algunas) para resolver problemas de la vida diaria son: diagnóstico y pronóstico, visión automática, control de producción, recuperación de información, lenguaje natural, planeación, control, reconocimiento de voz y en casi cualquier área donde tenemos información incompleta e incierta. [Pearl, 1988], [Heckerman et al, 1995].

3.1- Construcción de Redes Bayesianas.

Los problemas clásicos a los que nos enfrentamos en la construcción de una red probabilista son tres a saber:

a) Determinación de la estructura. La determinación de la estructura de una red consiste en encontrar su topología, es decir, las relaciones de dependencia entre las variables relevantes involucradas en un problema dado. La determinación de la topología de una red probabilista es, en muchos de los casos, proporcionada por el experto (enfoque clásico que veremos más adelante). Sin embargo, existen otros métodos (como el algoritmo que proponemos más adelante) que no extraen directamente del humano la topología sino de datos estadísticos. Más recientemente se han propuesto métodos que combinan estos dos enfoques.

b) Determinación de los parámetros. En una red probabilista, cada nodo tiene asociada una función de probabilidad (ya sea marginal ó condicional). Para cada diferente estructura de una red probabilista, debemos determinar u obtener las distribuciones de probabilidad en cada nodo para esa estructura.

c) Propagación de probabilidades. Si tenemos la topología de una red y las distribuciones de probabilidad para cada nodo, entonces queremos determinar el cambio en estas probabilidades cuando los valores de algunas variables llegan a ser conocidos. El proceso de instanciar estas variables de entrada y propagar sus efectos a través de la red es lo que se llama propagación de probabilidades [Sucar, 1994].

Básicamente existen dos enfoques para construir una red probabilista:

- *Enfoque tradicional*
- *Enfoque de aprendizaje*

Enfoque tradicional.

Con respecto al enfoque tradicional, en la década pasada, la construcción de sistemas expertos que manejaran incertidumbre utilizando el enfoque de redes bayesianas se empezó a hacer popular. Los primeros sistemas de este tipo se construían bajo una interpretación subjetivista, esto es, el experto humano definía las relaciones entre las variables

para un caso dado así como los parámetros numéricos que corresponden a los nodos en la red probabilista (como lo habíamos visto ya en la definición de red probabilista). Sin embargo, en muchos de los casos, el mismo experto no tiene bien definidas las relaciones de dependencia relevantes entre las variables del problema (a pesar de que el conjunto de ellas no sea muy grande) y en la asignación de probabilidades en los nodos de dicha red, el experto puede caer en inconsistencias. De hecho, la extracción del conocimiento de un experto para la determinación de los parámetros estructural y numéricamente (topología y distribuciones de probabilidad respectivamente), es una tarea compleja que consume mucho tiempo ya que como en la construcción de cualquier tipo de sistema experto, esta tarea es considerada como uno de los principales problemas para la construcción de éstos.

3.2- Enfoque de aprendizaje.

Por otra parte, y más recientemente, el enfoque de aprendizaje ha surgido como otro medio para la construcción de redes bayesianas.

El aprendizaje consiste en determinar tanto la topología como los parámetros así como determinar la existencia de variables escondidas y/o calcular los valores faltantes en algunos de los casos que están siendo tomados en cuenta a partir de un muestreo de datos estadísticos.

Dentro de este enfoque de aprendizaje podemos notar claramente 2 casos. En el primer caso, algunos algoritmos combinan el conocimiento del experto humano con los datos disponibles del problema; esto es, el experto construye la red probabilista y después se utiliza la base de datos para actualizarla añadiendo ó eliminando arcos entre variables ó incluso creando una ó más redes bayesianas diferentes.

En el segundo caso, algunos otros algoritmos utilizan para construir una red probabilista solamente los datos disponibles del problema a resolver. Por ejemplo, una base de datos de X número de pacientes con ciertos síntomas ó una base de datos de Y número de autos con ciertas fallas, etc. La inducción de la red se hace sin la interacción directa con el experto humano.

Dentro del enfoque de aprendizaje (para cualquiera de los casos), también se imcluye el escoger una estructura en un espacio de muchas posibles. Esto es un problema exponencial en función del número de variables (nodos) del problema. La fórmula de Robinson que propuso en 1977 [Cooper and Herskovits, 1992], determina el número de posibles estructuras con n número de nodos. Para n = 2, el número de posibles estructuras es 3, para n = 3, el número es 25, para n = 5, es 29,000 y para n= 10, es $4.2X10^{18}$. Por esta razón como se mencionó anteriormente, algunos algoritmos [Cooper and Herskovits, 1992] hacen uso de métodos heurísticos que reducen la complejidad exponencial a polinomial.

Debido a que la red probabilista se construye ó modifica (según sea el caso) a partir de datos estadísticos, un problema es el tamaño de la muestra de los datos para hacer esto. La determinación del número de casos requeridos para el entrenamiento de la red es otro problema. Para el caso de muestras pequeñas, los datos tienen que ser complementados por el experto para que así los resultados obtenidos sean útiles y confiables. Como ya hemos dicho varias veces, a veces los expertos pueden ser no muy buenos en plasmar su conocimiento en una red probabilista ni en estimar las probabilidades de dicha red. El hacer esto correctamente resulta primordial para el éxito del algoritmo. Mientras mayor sea el número de casos en la muestra es mucho más confiable (desde el punto de vista estadístico) la construcción de la red.

4.- El Algoritmo.

4.1- Bases Conceptuales.

En el algoritmo aquí propuesto usamos como criterio las medidas informacionales discutidas anteriormente que, como se ha señalado, están asociadas con hipótesis estadísticas de independencia e independencia condicional. Ideas similares, aunque en contextos distintos, han sido empleadas para generar redes por [Sucar, 1994] (árboles y poli-árboles) y en el paquete **MIM** elaborado por [Edwards, 1995] aunque en este último caso se usa para generar gráficas no dirigidas que permiten representar solamente independencia condicional, más no permiten inferencias ulteriores (pronósticos ó diagnósticos).

La idea básica es que las entropías medias H(X/Y), H(X/Y;Z) forman un sucesión decreciente, de tal manera que es de esperarse que si se considera solamente al conjunto de variables que den la mayor ganancia en información, las restantes aportan poco, una vez que las primeras han sido consideradas.

Además, como se ha mencionado, el uso de las ganancias en información permite inducir un orden completo en las variables independientes que es la base para garantizar que la red generada sea una gráfica acíclica dirigida.

El algoritmo es eficiente en el sentido que no prueba todas las posibles asociaciones (como hace MIM y otros algoritmos) sino solamente aquellas que pueden ser significativas a luz de las medidas informacionales.

El algoritmo propuesto hace las pruebas de hipótesis condicional agregando arcos entre las variables (si esta prueba es rechazada por supuesto). El algoritmo ordena las variables de mayor a menor relevancia (según la

información que proporcionan sobre la variable dependiente). A este tipo de algoritmos generalmente se les denomina **stepwise forward**. Otro algoritmo de este tipo es el de [Cooper and Herskovits, 1992].

4.2- Descripción del algoritmo.

En términos generales el algoritmo procede de la siguiente manera :

Sea Y la variable aleatoria dependiente y X_1, X_2, X_n las variables independientes.

1.- Se calcula la información de Y dadas todas las variables independientes: I(Y/Xi), donde i = 1,2,.., n y se ordenan las variables independientes de mayor a menor según estos valores: $X_{(1)}$, $X_{(2)}$, ..., $X_{(n)}$. Donde $X_{(1)}$ es la variable que proporciona mas información sobre la variable Y. Al ordenar las variables con respecto a la información proporcionada por cada una de ellas, inducimos un orden completo y nos aseguramos que obtendremos una GAD debido a los siguientes teoremas [Neapolitan, 1990] :

Teorema : Sea G = (V,E) una gráfica dirigida. Entonces un orden ancestral de nodos en V existe si y sólo si G es una gráfica acíclica dirigida. Un orden ancestral se refiere a que para cada nodo en V, todos los ancestros de ese nodo se encuentran ordenados antes que éste. Y el término ancestro se refiere a que un nodo cualquiera *X* es ancestro de un nodo cualquiera *Y* si existe una trayectoria (un camino) de *X* hacia *Y*.

Teorema : Sea G = (V,E) una GAD y $v \in V$. Entonces siempre es posible obtener un orden ancestral de los vértices en V tal que sólo los descendientes de v están etiquetados después de v.

2.- Sea Z el conjunto de las variables mas relevantes, Z = $\{X_{(1)}, X_{(2)}, ..., X_{(k)}; 1<=k<n\}$. Estas variables mas relevantes son aquellas para las cuales se ha rechazado la hipótesis de independencia condicional por lo que se traza un arco directo de ellas hacia Y. En el caso más simple solo una es la variable mas relevante y se traza un arco de ésta hacia la variable dependiente. (En la versión actual del algoritmo, el usuario puede decidir cuántas variables incluir en Z).

3.- Sea W el conjunto de las variables no relevantes, es decir, todas menos la variable que resultó ser la mas relevante; W = $\{X_{(k+1)},..., X_{(n)}\}$. Pruébese la hipótesis de independencia condicional condicionando con la variable mas relevante $X_{(1)}$, es decir, $Y \perp\!\!\!\perp X/Z$, utilizando I(Y/X;Z) para cada $X \in W$.

4.- Sean U y V dos conjuntos ajenos tales que $U \cup V = W$ y para cada variable que se encuentre en V se tiene que $Y \perp\!\!\!\perp V/Z$ y para cada variable que se encuentre en U se tiene que $Y \perp\!\!\!\perp U/Z$. Por lo que se traza un arco de cada U a Y. En otras palabras, para las variables que se encuentran en V se acepta la prueba de independencia condicional mientras que para las variables que se encuentran en U se rechaza esta misma hipótesis.

5.- Hágase Y: = $X_{(1)}$ y repítase el procedimiento hasta acabar con todas las variables.

4.3- Limitaciones del algoritmo.

El algoritmo tiene varias limitantes que deben ser consideradas :

i) La prueba estadística se basa en un nivel de significancia (α=0.01 ó α=0.05) en cada prueba incondicional . Como se realizan muchas pruebas sucesivamente, el nivel de significancia global es mucho mayor que el nominal. Entonces en cada caso particular y con mayor razón cuando las muestras son pequeñas, debe revisarse el nivel de significancia real que debe emplearse.

ii) La dirección de los arcos está determinada exclusivamente por el orden ancestral inducido por la ganancia en información principalmente por cada variable y no indica *causalidad* u orden temporal.

iii) Si la muestra es pequeña los resultados estadísticos son endebles, en el sentido que la distribución de probabilidad de la estadística posiblemente está lejos de parecerse a la de la variable χ^2. Por otra parte habrá muchos "huecos" en los datos, es decir, habrá muchas combinaciones de valores de las variables (*x, y, z,* ...) para los cuales no se contará con casos en la muestra. Esta situación también produce inestabilidad en los estimadores de las probabilidades y de las estadísticas.

La recomendación es que la red generada por el algoritmo sea sometida a la consideración del experto para su evaluación y modificación.

5.- Pruebas y resultados.

El algoritmo expuesto en este trabajo y explicado anteriormente, fue propuesto en su primera versión por el Dr. Manuel Martínez Morales [Martínez-Morales, 1995], investigador de la Maestría en Inteligencia Artificial de la Universidad Veracruzana. La implantación computacional de este algoritmo (en su versión mejorada) cuenta con las siguientes opciones:

1.- Editor de archivos de variables y validación de las mismas.
2.- Generación de la red bayesiana.
3.- Visualización de la red bayesiana resultante.
4.- Visualización paso a paso de la red bayesiana resultante.
5.- Cálculo de los parámetros asociados a cada nodo, cálculo de las informaciones proporcionadas por cada variable así como la justificación de rechazo o aceptación de las pruebas de hipótesis condicional.

6.- Opción para agregar o eliminar arcos entre variables.
7.- Módulo de propagación de probabilidades para casos sencillos (consulta de un caso específico con base en la(s) variable(s) que proporciona(n) mayor información a la variable dependiente).
8.- Posibilidad para combinar dos o más variables en una sola.

Esta versión mejorada difiere de la versión original en alguno aspectos que mencionamos en seguida.

a) Capacidad de interacción directa con el experto o el usuario en cuanto a la posibilidad de modificar la estructura de la red resultante eliminado o agregando arcos entre nodos. Al hacer alguna modificación, el cálculo de los parámetros asociados a cada nodo se vuelve a realizar (ya que la estructura de la red cambió). Además hay una fase de validación en este módulo que permite guardar la propiedad de que la gráfica sea una gráfica acíclica dirigida (GAD). Este módulo es de suma importancia ya que durante el proceso de validación de la red bayesiana resultante inducida a partir de datos estadísticos, el experto tendrá que hacer en muchas de las veces, cambios en la estructura de la misma de acuerdo con sus criterios para que la red refleje lo más fielmente posible las relaciones entre las variables del problema que está siendo considerado.

b) Mayor eficiencia en el cálculo de probabilidades asociadas a cada nodo. Este módulo fue completamente reprogramado para lograr reducir significativamente el tiempo de procesamiento.

c) Permite combinar dos o más variables en una sola y así poder condicionar a la variable dependiente con el número de variables independientes que uno desee para observar el comportamiento del algoritmo con respecto a las variables combinadas.

A continuación presentamos un ejemplo tomado de [Whittaker, 1990] de una encuesta sobre la cantidad de evasión de impuestos en la industria de la construcción, dibujando las gráficas resultantes (figuras 5.1 y 5.2) aplicando el algoritmo propuesto por [Whittaker, 1990] y el algoritmo propuesto en este trabajo [Martínez-Morales, 1995].

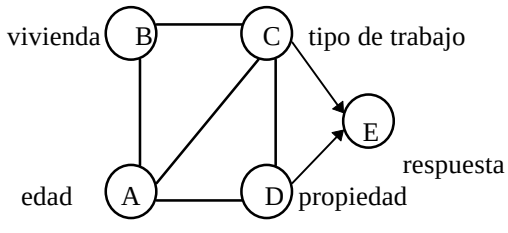

figura 5.1 (J. Whittaker)

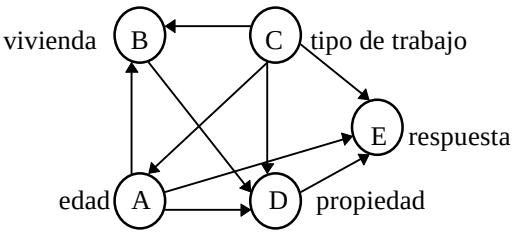

figura 5. (Algoritmo propuesto)

Como podemos observar las relaciones más importantes entre las variables son “descubiertas” por ambos algoritmos con la diferencia que la gráfica inducida por el algoritmo aquí propuesto es una gráfica acíclica dirigida permitiendo realizar inferencias tales como pronóticos y diagnósticos.

Ahora veamos otro ejemplo (figura 5.3) que se refiere a la determinación de la mortandad en pacientes infartados a partir de variables tales como diabetes, sexo, edad y localización del infarto en una muestra de 97 pacientes.

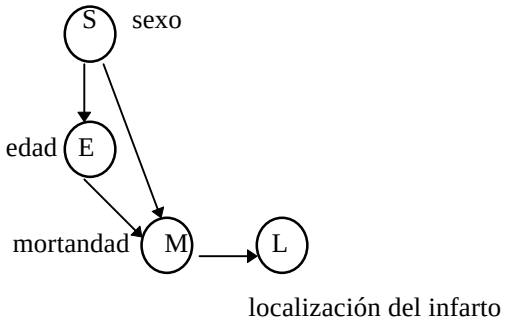

figura 5.3

La gráfica resultante nos muestra que tanto la edad como el sexo son factores que determinan la mortandad en esta muestra de pacientes. Este resultado fue validado por un especialista del área cuyas conclusiones coinciden con aquéllas a las que llegó el algoritmo.

En la gráfica 5.4 se muestra el resultado de la combinación entre las variables edad y sexo (llamada variable Z) del ejemplo anterior.

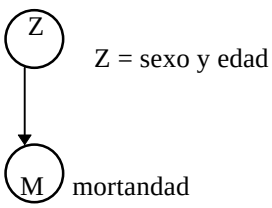

figura 5.4

Debemos mencionar que el algoritmo propuesto en este trabajo se ha aplicado a datos reales con buenos

resultados ; es decir, los resultados obtenidos han sido similares con aquéllos obtenidos por otros métodos, como el algoritmo propuesto por [Whittaker, 1990]; además los expertos han validado el desempeño del algoritmo obteniendo en muchos de los casos conclusiones semejantes. Estos datos son los siguientes:

- encuesta sobre preferencias electorales [Martínez-Morales, 1995].
- preferencias en la elección de carreras universitarias [Portilla Carolina, 1996].
- datos sobre ozono y contaminación en la cd. de México [Sucar, 1996].
- biodiversidad medida como abundancia de especies de escarabajos [Arellano, 1996].
- conductas sexuales en monos aulladores [Cortés Liliana, 1996].
- mortandad en pacientes infartados [Martínez-Morales, 1995].
- encuesta sobre la cantidad de evasión de impuestos en la industria de la construcción [Whittaker, 1990].

6.-Conclusiones, recomendaciones y trabajo futuro.

- Fue posible construir un algoritmo eficiente (se han manejado conjuntos de datos conteniendo hasta 45 variables y 10,000 casos) para generar la red y sus probabilidades asociadas sin haber ocupado excesivo tiempo (16 minutos en una sparc station 5 Sun).

- La red generada no necesariamente representa nexos causales por lo que su mayor utilidad se obtendría en el contexto de la adquisición del conocimiento colaborativo entre máquina y experto. La red debe ser presentada al especialista para su posible modificación y, de ser posible, debe validarse con nuevos datos y/o el análisis de los mismos datos con otros procedimientos.

- Por último, estamos trabajando en el análisis de la bondad del algoritmo utilizando conjuntos de datos simulados. Esto se refiere al hecho de que a partir de la determinación de la topología de una red bayesiana cualquiera se generen datos aleatoriamente, e introduciendo éstos al algoritmo, se compruebe si la gráfica inducida es semejante a la red con la cual se generaron dichos datos.

7.- Bibliografía.

[1] [Agresti, 1984] A. Agresti. ***Categorical Data Analysis***. John Wiley, New York, 1984.

[2] [Arellano et al., 1996] L. Arellano García, L.E. Rivera Cervantes, M. Martínez Morales. ***Comparación de la Composición y Abundancia Estacional de los Escarabajos Carroñeros (Coleoptera Silphidae) en Bosques Templados de Jalisco y Veracruz, México***. Reporte técnico. Instituto de Ecología. Xalapa, Veracruz, 1996.

[3] [Barlow and Braganca, 1990] Richard E. Barlow and Carlos Alberto de Braganca Pereira. Conditional Independence and Probabilistic Influence Diagrams, in Block, H.W. *et al.* ed. ***Topics in Statistical Dependence***. Lecture Notes-Monograph Series Vol. 16, 19-33; Institute of Mathematical Statistics, Hayward, Ca., 1990.

[4] [Cooper and Herskovits, 1992] Gregory F. Cooper and Edward Herskovits. A Bayesian Method for the Induction of Probabilistic Networks from Data. ***Machine Learning*** 9, 309-347, 1992.

[5] [Cortés, 1996] Liliana Cortés. Reporte no publicado.

[6] [Edwards, 1995] David Edwards. ***Introduction to Graphical Modelling***. Springer-Verlag, New York, 1995.

[7] [Heckerman *et al.*, 1995] David Heckerman, Abe Mandani and Michael P. Wellman. Real World Applications of Bayesian Networks. Introduction. ***Communications of the ACM*** (38), 3, 24-26. March 1995.

[8] [Hugin, 1993] Hugin Expert A/S Company and Jensen V. Finn. ***Manual on Bayesian Networks***. August, 1993.

[9] [Kullback, 1959] Solomon Kullback. ***Information Theory and Statistics***. Dover, New York, 1959.

[10] [Martínez-Morales, 1995] Manuel Martínez Morales. An Algorithm For The Induction of Probabilistic Networks From Data. ***Memorias de la XII Reunión Nacional de Inteligencia Artificial***, Soc. Mex. de Intel. Artif., Instituto Tecnológico de Estudios Superiores de Monterrey, Campus Morelos, Instituto de Investigaciones Eléctricas, Cuernavaca, Morelos. Limusa, 36-40, 1995.

[11] [Neapolitan, 1990] Richard E. Neapolitan**.** ***Probabilistic Reasoning in Expert Systems: Theory and Algorithms***. John Wiley, New York, 1990.

[12] [Pearl, 1988] Judea Pearl. ***Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference***. Morgan Kaufmann, San Mateo, Ca., 1988.

[13] [Portilla, 1996] Carolina Portilla, 1996. Reporte no publicado.

[14] [Spiegelhalter et al., 1993] David J. Spiegelhalter, A. Philip Dawid, Steffen L. Lauritzen and Robert G. Cowell. Bayesian Analysis in Expert Systems. ***Statistical Science*** ; Vol. 8, 3, 219-246, 1993.

[15] [Sucar, 1994] L. Enrique Sucar. Structure and Parameter Learning in Probabilistic Networks. ***Memorias de la XI Reunión Nacional de Inteligencia Artificial***, Soc. Mex. de Intel. Artif., Universidad de Guadalajara. Limusa, 23-41, 1994.

[16] [Wechsler, 1990] Harry Wechsler. ***Computational Vision***. Academic Press, Inc. 1990.

[17] [Whittaker, 1990] J. Whittaker. ***Graphical Models in Applied Multivariate Analysis***. John Wiley, Chichester, 1990.