Universidad de las Ciencias Informáticas

Facultad 5

Aplicación de las Redes Bayesianas para la toma de decisiones de los elementos en Entornos Virtuales

Trabajo de Diploma para optar por el título de

Ingeniero en Ciencias Informáticas

Autores: Reinier Bermudez González

Yurima Jorge Pupo

Tutora: Ing. Yenifer Del Valle Guevara

Ciudad de la Habana

Julio de 2008

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los ___ días del mes de __________ del año _____.

Reinier Bermudez Gonzalez **Yurima Jorge Pupo**

_________________ _________________

Firma del Autor Firma del Autor

Yenifer del Valle Guevara

____

Firma del Tutor

DATOS DE CONTACTO

Tutora:

Nombre y Apellidos: Yenifer del Valle Guevara.

Institución: Universidad de las Ciencias Informáticas (UCI).

Título: Ingeniera en Ciencias Informáticas.

Categoría Docente: Profesor Instructor.

E-mail: ydelvalle@uci.cu

Graduada de la UCI en el año 2007, líder del proyecto del polo de Realidad Virtual “Desarrollo de Elementos Virtuales Inteligentes” en la Universidad de las Ciencias Informáticas.

AGRADECIMIENTOS COMPARTIDOS

Al concluir la realización de este Trabajo de Diploma no podemos dejar de agradecer a las personas que nos brindaron su ayuda cuando fue necesario.

A nuestros padres por el apoyo en todo momento y la ayuda incondicional.

A la Revolución por darnos la oportunidad de estudiar en esta magnífica escuela.

A la UCI por ser nuestra casa durante estos 5 años.

A nuestra tutora Yenifer y a Alexey por su gran ayuda y por las largas horas de sueño entregadas para lograr este trabajo.

A nuestros compañeros de grupo, por permitirnos compartir tantos momentos buenos y por ser parte de esta gran familia.

Al profe Millet, por estar dispuesto cuando necesitamos su ayuda.

A Todos Muchas Gracias!!!

Yurima y Reinier

AGRADECIMIENTOS

No puedo dejar de agradecer a personas tan especiales sin las cuales no habría podido realizar este gran sueño:

A mis padres, por haberme guiado durante estos 23 años, aquí tienen el regalo que tanto esperaron…los quiero MUCHOOOO

A mi hermanito, por esa forma tan especial de quererme, por su sonrisa siempre oportuna y sus locas palabras de aliento.

A mis abuelos, los que están y los que ya se fueron, por ser tan especiales y hacerme sentir parte importante de su mundo.

A mis tíos, Enma, Yeyi, Magalis, Mireya, Delmis, Marlene, Javier, Eliecer y Tony por la ayuda infinita, por la preocupación, por su constante desvelo.

A mi Primo Danilito y a Yanet, por ser incondicionales, por los consejos y la ayuda que me dieron siempre.

A Rey, mi compañero de tesis por las noches de desvelo, por su confianza y optimismo, juntos alcanzamos este triunfo.

A Irina, por ser amiga y consejera, por estar siempre pendiente de mi, por esa frase que nunca olvido y que tanto me hace crecer…. "Voy a Ti".

A Naty, por ser la amiga de siempre, porque aunque estamos lejos sigue ahí.

A mis amigas Lisy, Yaima, e Isa, por los momentos buenos y las fiestas juntas, porque cuando las necesite siempre estuvieron dispuestas.

A mis amigos Luis Mario y Carlos Rafael por ser mis hermanos, por los buenos consejos.

A Yasel, por su ayuda y comprensión.

A todos sin que se me quede nadie... Gracias!!!.....Yuri

AGRADECIMIENTOS

Muchas son las personas a las que tengo que agradecer la realización de este proyecto

A mis padres, por su inmenso apoyo y sacrificio, por haber hecho hasta lo imposible para hacer este sueño realidad

A mis abuelos, por estar pendientes de mí a cada minuto a pesar de la distancia

A Yaima, por ser tan especial y darme tantos momentos de felicidad

A María del Carmen y Ariel, por ser como mis padres aquí en la Habana

A Isle, a Ivonne y a Flora, por ser mis "co-tutoras" en la realización de este trabajo

A mi primita Maye, a mi tía Maritza y a toda mi familia en general, por su apoyo incondicional en todos los momentos de mi vida

A mis amigos Yoelvis, Erick, Rafe… por estar siempre presentes en las buenas y en las no tan buenas

A mis compañeros de grupo, a los chacales del apartamento y a todos los que han tenido que ver de una forma u otra con la realización de este trabajo. Gracias!!!

Reinier

DEDICATORIA

A mis padres y a mi hermanito, porque este sueño es también de ellos,
Por desearlo tanto como yo.
A su comprensión y apoyo incondicional
En estos 5 años.
A su ejemplo de toda la vida.
Yuri

A mis padres, porque desde pequeño me decían
Que tenía que ser ingeniero y,
A mi hermanito, porque cuando sea grande
También tiene que serlo.
Reinier

RESUMEN

Para lograr un mayor realismo en los entornos de Realidad Virtual se utiliza en gran medida las técnicas de Inteligencia Artificial. Este trabajo se centra en una de ellas en específico: Las Redes Bayesianas.

En la presente investigación se realiza un estudio detallado sobre el funcionamiento de las Redes Bayesianas, partiendo desde enunciar un concepto de Red Bayesiana, hasta proponer algoritmos para la implementación de la misma teniendo en cuenta varias topologías de red.

Luego se llega a la conclusión de que topología sería mejor utilizar en las aplicaciones de Realidad Virtual teniendo en cuenta los requerimientos funcionales de este tipo de aplicaciones y se explica como se puede realizar la toma de decisiones en Entornos Virtuales a partir de una Red Bayesiana.

Por último se describe e implementa un caso de estudio sencillo de una Red Bayesiana aplicada a un Entorno Virtual a partir del cual se llegan a conclusiones.

PALABRAS CLAVES:

Inteligencia Artificial, Redes Bayesianas, Entorno Virtual, toma de decisiones, topología

INDICE

INDICE DE FIGURAS .... **VIII**

INTRODUCCIÓN .... **- 1 -**

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA .... **- 4 -**

INTRODUCCIÓN .... - 4 -

1.1 HISTORIA DE LAS REDES BAYESIANAS .... - 4 -

1.2 TEORÍA DE GRAFOS .... - 5 -

1.2.1 Conceptos básicos de Grafos no Dirigidos .... *- 6 -*

1.2.2 Conceptos básicos de Grafos Dirigidos .... *- 8 -*

1.2.3 Conceptos básicos de Grafos Mixtos .... *- 11 -*

1.3 TEORÍA DE LA PROBABILIDAD .... - 12 -

1.4 MODELOS GRÁFICOS PROBABILÍSTICOS .... - 14 -

1.4.1 Modelos gráficos probabilísticos no dirigidos. Redes de Markov .... *- 14 -*

1.4.2 Modelos gráficos probabilísticos dirigidos. Redes Bayesianas .... *- 15 -*

1.4.3 Modelos gráficos probabilísticos mixtos. Redes Cadenas .... *- 15 -*

1.5 FUNDAMENTOS TEÓRICOS DE LOS MODELOS GRÁFICOS PROBABILÍSTICOS DIRIGIDOS .... - 15 -

1.6 ESTUDIO SOBRE REDES BAYESIANAS EN CUBA .... - 18 -

1.7 APLICACIONES DE LAS REDES BAYESIANAS .... - 19 -

CAPÍTULO 2. REDES BAYESIANAS .... ¡ERROR! MARCADOR NO DEFINIDO.

INTRODUCCIÓN .... - 21 -

2.1 REDES BAYESIANAS .... - 21 -

2.2 TIPOS DE REDES BAYESIANAS .... - 23 -

2.3 TEOREMA DE BAYES .... - 23 -

2.4 PROPAGACIÓN DE LA EVIDENCIA .... - 29 -

2.4.1 Algoritmos de Propagación y Estructuras Gráficas .... *- 30 -*

2.5 EL APRENDIZAJE EN LAS REDES BAYESIANAS .... - 38 -

2.6 VENTAJAS Y DESVENTAJAS DE LAS REDES BAYESIANAS .... - 39 -

CAPÍTULO 3. PROPUESTA DE SOLUCIÓN .... **- 41 -**

INTRODUCCIÓN .... - 41 -

3.1 ALGORITMO DE PROPAGACIÓN .... - 41 -

3.2 TOMA DE DECISIONES .... - 45 -

3.3 CASO DE ESTUDIO - 45 -
3.4 RESULTADOS - 51 -
CONCLUSIONES - 52 -
RECOMENDACIONES - 53 -
REFERENCIAS BIBLIOGRÁFICAS - 54 -
ANEXO 1 FASE DE INICIALIZACIÓN - 57 -
ANEXO 2 FASE DE ACTUALIZACIÓN - 60 -
ANEXO 3 SEGUNDA FASE DE ACTUALIZACIÓN - 63 -
GLOSARIO DE TÉRMINOS ¡ERROR! MARCADOR NO DEFINIDO.

INDICE DE FIGURAS

Figura 1 Grafo - 6 -
Figura 2 Grafo dirigido ....Figura 3 Grafo no dirigido ....Figura 4 Grafo mixto - 6 -
Figura 5 Grafo completo - 7 -
Figura 6 Ciclado C={A,B,D,E} asociado a un grafo no dirigido - 8 -
Figura 7 Árbol - 8 -
Figura 8 Árbol simple................................... Figura 9 Poliárbol - 9 -
Figura 10 Numeración ancestral de un grafo dirigido - 10 -
Figura 11 Grafo Dirigido Acíclico (GDA) - 10 -
Figura 12 Grafo dirigido Acíclico .... Figura 13 Grafo no dirigido asociado al GDA de la Figura 12 - 11 -
Figura 14 Grafo cadena - 12 -
Figura 15 Grafo cadena...............................Figura 16 Grafo no dirigido asociado al - 12 -
Figura 17 Red Bayesiana - 17 -
Figura 18 Red Bayesiana que representa el Ejemplo 1 - 24 -
Figura 19 Red Bayesiana donde se muestran las probabilidades a Priori y probabilidades condicionales del Ejemplo 1 - 25 -
Figura 20 Red Bayesiana que representa el Ejemplo 2 - 27 -
Figura 21 Tipos de Estructuras Arboles...................... Poliarboles .... Redes Multiconectadas - 31 -
Figura 22 Red Bayesiana con topología de árbol - 32 -
Figura 23 Red Bayesiana con topología de poliárbol - 35 -
Figura 24 Estado Inicial de la red - 35 -
Figura 25 Instanciación del nodo E - 36 -
Figura 26 Intanciacion de los nodos E y B - 36 -
Figura 27 Instanciación de los nodos E, B y C - 37 -
Figura 28 Red Bayesiana aplicada al DEMO de tiro - 46 -
Figura 29 Red Bayesiana después de la Fase de Instanciación - 48 -
Figura 30 Red Bayesiana después de la primera Actualización - 49 -
Figura 31 Red Bayesiana después de la segunda actualización - 50 -

INTRODUCCIÓN

En la actualidad la Realidad Virtual se utiliza ampliamente en la Industria, permite minimizar los costes de la realización de prototipos, simulación, entrenamiento, marketing y otros usos. También la tecnología actual permite obtener un realismo sin precedentes.

En los últimos años en Cuba se ha comenzado a manejar este nuevo concepto y ya son muchas las personas y proyectos que se han propuesto desarrollar este campo, un ejemplo de ello es la Facultad 5 de la Universidad de las Ciencias Informáticas (UCI) que es la que se ha encargado de desarrollar Sistemas de Realidad Virtual dentro de la UCI. Actualmente en esta Facultad se llevan a cabo múltiples proyectos productivos en los cuales los elementos que interactúan dentro de un Entorno Virtual no son capaces de tomar las decisiones por si mismos en una situación determinada por lo que carecen de este nivel de inteligencia.

Para dar un mayor grado de realismo a los elementos que interactúan dentro del Entorno de Realidad Virtual en los proyectos desarrollados en esta facultad se pretende la utilización de las Redes Bayesianas para la toma de decisiones de estos elementos, para lo cual es necesario realizar una investigación sobre la utilización de este tipo de redes con este fin.

Dada la **Situación Problémica** planteada anteriormente, el **Problema Científico** de esta investigación es la "Falta de realismo en la toma de decisiones de los elementos en Entornos Virtuales en los proyectos de Realidad Virtual de la Facultad 5". Siendo el **Objeto de Estudio** "Las Redes Bayesianas" y el **Campo de Acción** específicamente la **"**Aplicación de las Redes Bayesianas en los proyectos de Realidad Virtual**".**

Una vez planteada la problemática principal se define como **Objetivo** del trabajo, **"**Proponer cómo aplicar las Redes Bayesianas para la toma de decisiones de los elementos en los proyectos de Realidad Virtual de la Facultad 5**"**. Para lograr el objetivo anterior será necesario cumplimentar las siguientes **Tareas de Investigación:**

- Investigar sobre la Teoría de Grafos y la Teoría de las Probabilidades.
- Investigar sobre las aplicaciones de las Redes Bayesianas.
- Describir el funcionamiento de las Redes Bayesianas.
- Describir como se realiza la toma de decisiones mediante Redes Bayesianas en Entornos Virtuales.
- Implementar un algoritmo de las Redes Bayesianas y aplicarlo a un DEMO de la Facultad.

Entre los **Métodos Teóricos** que serán utilizados para darle cumplimiento a estas tareas se encuentran: el Histórico-Lógico para estudiar la evolución y desarrollo del objeto de estudio de la investigación; y el Analítico-Sintético, haciendo énfasis en el análisis de las teorías, documentos, entre otros, que permite la extracción de los elementos más importantes para procesar la información y elaborar conclusiones.

Como **Resultados Esperados** en el presente trabajo de diploma se encuentran:

a) Una bibliografía completa sobre como aplicar las Redes Bayesianas para incorporarle un nivel de inteligencia en la toma de decisiones de los elementos en Entornos Virtuales a los proyectos de Realidad Virtual de la Facultad 5.

b) Un primer DEMO de Redes Bayesianas en la Facultad 5.

La presente investigación está estructurada en tres capítulos, en el primero se exponen los elementos que describen la fundamentación teórica a través de varios subepígrafes en los cuales se abordan temas como:

- Orígenes Históricos
- Teoría de Grafos
- Teoría de la Probabilidad
- Modelos Gráficos Probabilísticos
- Fundamentos Teóricos de los Modelos Gráficos Probabilísticos Dirigidos.
- Estudios sobre Redes Bayesianas en Cuba
- Aplicaciones de las Redes Bayesianas

En el capítulo 2 se abordan aspectos relacionados con las Redes Bayesianas haciéndose énfasis en las Redes Bayesianas Discretas y tratando temas como:

- Definición de Redes Bayesianas
- Tipos de Redes Bayesianas
- Teorema de Bayes
- Propagación de la Evidencia
- Aprendizaje
- Ventajas y desventajas de las Redes Bayesianas

En el capítulo 3 se expone como se realiza el proceso de la toma de decisiones utilizando un algoritmo de propagación de las Redes Bayesianas; y además se presenta un caso de estudio aplicado a un demo de tiro del proyecto Herramientas de Desarrollo para Sistemas de Realidad Virtual.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Introducción

El empleo de las Redes Bayesianas se presenta como una técnica para introducir conocimiento y realizar razonamiento en aplicaciones relacionadas con la toma de decisiones. Suponen una herramienta gráfica para el modelado del conocimiento como consecuencia de la unión de la Teoría de Probabilidad y la Teoría de Grafos, resultando un enfoque muy atractivo para tratar en general problemas de cierta complejidad en presencia de incertidumbre. El tratamiento de esta complejidad en un entorno gráfico permite que los problemas sean modularizables, utilizando la Teoría de la Probabilidad encargada de relacionar dichos módulos con los datos.

En este capítulo se presenta la historia de las Redes Bayesianas, donde se aborda el surgimiento y desarrollo del Teorema de Bayes. Posteriormente se introduce la Teoría de Grafos, muy necesaria para la comprensión de este tipo de redes ya que la forma de representarla e implementarla es mediante esta estructura. En otra de las secciones del capítulo se exponen conceptos y definiciones relacionados con la unión de la Teoría de Grafos y la Teoría de las Probabilidades llegando así a los Modelos Gráficos Probabilísticos Dirigidos.

En las secciones finales del capítulo se muestra una panorámica sobre estudios realizados en Cuba relacionados con el empleo de las Redes Bayesianas y además se expone la utilización de estas redes en los entornos virtuales.

1.1 Historia de las Redes Bayesianas

El método más antiguo para el tratamiento de la incertidumbre es la probabilidad. Una de las contribuciones más importantes de la Teoría de las Probabilidades es sin lugar a dudas el Teorema de Bayes. El surgimiento y desarrollo del Teorema de Bayes acentuó la importancia de un conjunto de conceptos utilizados en el proceso de búsqueda y dominio del conocimiento científico.

El Teorema de Bayes fue concebido por el famoso matemático inglés Thomas Bayes que estableció su teoría de la probabilidad en un ensayo que no se publicó sino hasta tres años después de su fallecimiento, en 1764. Laplace aceptó las conclusiones de Bayes en 1781, posteriormente las redescubrió Condorcet y permanecieron incontrovertibles hasta que Boole las cuestionó. Desde entonces las técnicas bayesianas han estado sujetas a controversia. (1)

Dentro del campo de la Inteligencia Artificial, surgieron críticas contra el uso de métodos probabilistas en sistemas expertos, especialmente porque las hipótesis necesarias para hacer tratable el método bayesiano clásico eran incorrectas en la mayor parte de los problemas del mundo real. Esto motivó el desarrollo de otros métodos, como los factores de certeza o la lógica difusa, en que se introducen implícitamente hipótesis y aproximaciones aún más exigentes. Afortunadamente, el desarrollo de las Redes Bayesianas en la década de los 80 como modelo probabilístico para el razonamiento con incertidumbre en Inteligencia Artificial, permitió refutar las objeciones anteriores contra el uso de la probabilidad, construyendo un modelo de razonamiento causal con un sólido fundamento teórico.

Por otro lado, los diagramas de influencia, que aparecen también en la década de los 80, pueden considerarse como una extensión de las Redes Bayesianas, que por tener nodos de decisión y nodos de utilidad, permiten resolver problemas de toma de decisiones. En la década de los 90 creció exponencialmente el número de investigadores, universidades y empresas dedicados a este tema. Actualmente existen sistemas expertos bayesianos en las especialidades más diversas. (2)

1.2 Teoría de Grafos

Muchos de los conceptos y definiciones relacionados con la Teoría de Grafos y que serán de gran ayuda para el estudio de las Redes Bayesianas, serán abordados en esta sección.

Se considera necesario en primer lugar conocer que la representación gráfica de una colección de objetos $V = \{V_1, V_2, \dots, V_n\}$ los cuales están relacionados entre si mediante aristas, siendo E_{ij} la arista que une los elementos V_i y V_j de V, define implícitamente un grafo, donde $V = \{V_1, V_2, \dots V_n\}$ son los nodos y E es el conjunto de aristas.

Definición de Grafo

Un grafo es la representación de un par $G = \{V, E\}$ donde $V = \{V_1, V_2, \dots, V_n\}$ es el conjunto finito de vértices o nodos y E es el conjunto de aristas, es decir, el conjunto de todos los pares ordenados de los distintos elementos de V que se relacionan. (3)

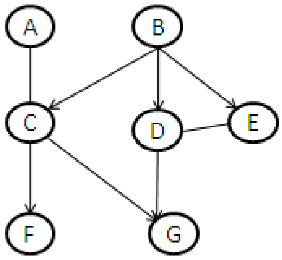


Figura 1 Grafo

En dependencia de la relación existente y el orden entre los nodos del grafo se pueden nombrar dos tipos de aristas: aristas dirigidas y aristas no dirigidas. Las aristas dirigidas se utilizan cuando $E_{ij} \in E$ pero $E_{ji} \notin E$, y se denota como $V_i \rightarrow V_j$, de forma que V_i se conecta con V_j y V_j no se conecta con Vi. Las aristas no dirigidas que se denota como $V_i - V_j$, se presenta cuando $E_{ij} \in E$ y $E_{ji} \in E$ quedando conectados los nodos V_i y V_j.

El tipo de arista puede definir el grafo, si un grafo tiene todas sus aristas dirigidas, el grafo se define como grafo dirigido; si el grafo tiene todas sus aristas no dirigidas, el grafo se denomina grafo no dirigido y cuando el grafo tiene aristas dirigidas y no dirigidas se dice que el grafo es un grafo mixto.

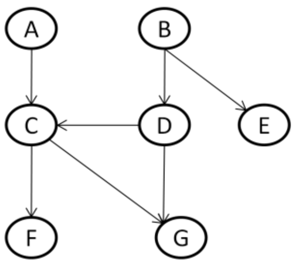


Figura 2 Grafo dirigido

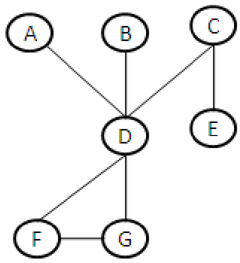


Figura 3 Grafo no dirigido

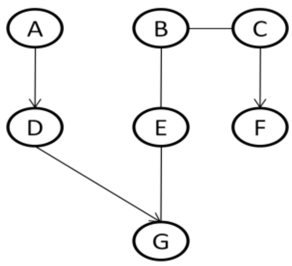


Figura 4 Grafo mixto

1.2.1 Conceptos básicos de Grafos no Dirigidos

En este subepígrafe se analizarán algunas definiciones y particularidades de los grafos no dirigidos $G = (V, E)$. Algunas de las cuales surgen de las relaciones existentes entre los nodos y de la forma en que se agrupan, surge así el concepto de vecinos de un nodo Vi como el conjunto de nodos directamente alcanzables desde V_i y se denota $vec(V_i)$. Un ejemplo de esto se puede observar en la Figura 3, donde los vecinos del nodo C son los nodos D y E, tal que $vec(C) = \{D, E\}$.

A continuación se mostrarán algunas definiciones fundamentales para tratar los grafos no dirigidos dependiendo de las relaciones que se establecen entre los nodos y las aristas:(3)

Definición de Bucle

En un grafo no dirigido se define un bucle como un camino cerrado, esto es una sucesión de nodos conectados tales que el nodo inicial coincide con el nodo final, por tanto, en la Figura 3 sólo existe un bucle que viene dado por los nodos $\{D,F,G\}$.

Definición de Subgrafo Asociado a un conjunto C

Sea $G = (V, E)$ un grafo no dirigido, y C un conjunto de nodos del mismo. Se define el subgrafo C asociado al conjunto C como el grafo no dirigido formado por los nodos de C y las aristas de E que unen a elementos de C.

Definición de Grafo Completo

Un grafo no dirigido $G = (V, E)$ es un grafo completo cuando existe una arista entre cada par de nodos. (Figura 5).

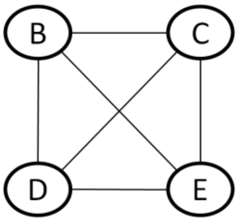


Figura 5 Grafo completo

Definición de Conjunto Completo

Sea C un subgrafo asociado a un subconjunto C de un grafo no dirigido, se dice que C es un subconjunto completo cuando existe una arista entre cada par de nodos de C.

Definición de Ciclado

Sea C un subconjunto completo de un grafo. Se dice que C es un ciclado cuando además C no es subconjunto propio de otro subconjunto completo, es decir, cuando C es maximal (Figura 6).

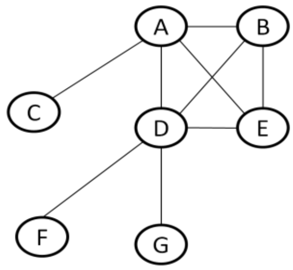


Figura 6 Ciclado C={A,B,D,E} asociado a un grafo no dirigido

Teniendo en cuenta los caminos que aparecen en un grafo, se pueden enunciar dos tipos de grafos: los grafos conexos no dirigidos que son aquellos en los que existe al menos un camino entre cada par de nodos, véase la Figura 6 como un ejemplo también de grafo conexo no dirigido, y los árboles que son grafos conexos no dirigidos en los que existe un único camino entre cada par de nodos. El grafo de la Figura 7 muestra un árbol obtenido a partir de la Figura 6 tras eliminar las aristas E_{AE}, E_{BD}, E_{DE}.

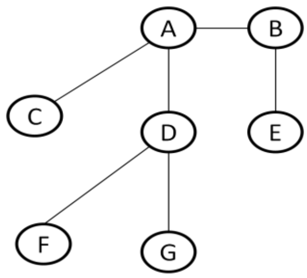


Figura 7 Árbol

1.2.2 Conceptos básicos de Grafos Dirigidos

De igual manera que en los grafos no dirigidos, las relaciones y agrupaciones que se establecen entre los nodos definen conjuntos de nodos específicos. En los grafos dirigidos las relaciones familiares marcan definiciones de igual carácter, entonces si, $Vi \rightarrow Vj$ se dice que Vi es *padre* de Vj, y se denota como $pa(V_j)$, y se dice además que V_j es *hijo* de Vi.

Así mismo al conjunto formado por un nodo Vi y sus padres $pa(Vi)$ se le llama familia del nodo Vi, *de* forma que $fa(V_i) = Vi \cup pa(V_i)$. En el grafo dirigido de la figura 2 se tiene que $pa(G) = \{C, D\}$ y los hijos de C son los nodos {F, G}, además $fa(G) = \{G, C, D\}$.

Dependiendo del número de padres de un nodo se presentan distintos tipos de grafos dirigidos, así, si cada nodo tiene como máximo un padre, el grafo dirigido se denomina grafo o árbol simple y en caso contrario poliárbol. (3)

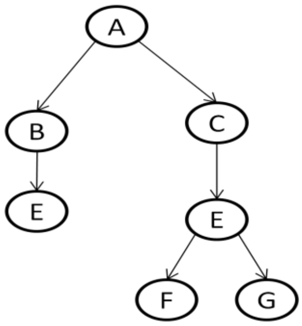


Figura 8 Árbol simple

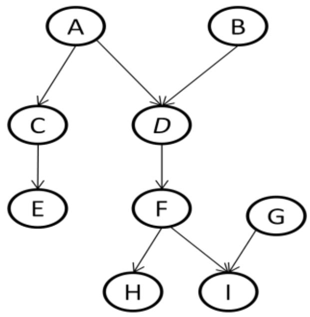


Figura 9 Poliárbol

Se define como los ascendientes de un nodo V_i, y se denota por $as(V_i)$, al conjunto de nodos que tienen un camino hasta Vi y se denominan descendientes del nodo Vi, denotado por $de(Vi)$ al conjunto de nodos a los que se puede ir desde V_i, así, en el poliárbol de la figura 9, el nodo D tiene de ascendientes a los nodos $\{\mathrm{A, B}\}$ y de descendientes a $\{\mathrm{F, H, I}\}$. Igualmente, se define el conjunto de no ascendientes de un nodo $Vi,\ na(V_i)$, como el conjunto de nodos V menos los ascendientes de Vi y el propio V_i, de forma que $na(V_i) = V \setminus (as(V_i) \cup V_i)$ y el conjunto de no descendientes de un nodo V_i, denotado como $nd(V_i)$, como el conjunto de nodos dado por $nd(V_i) = V \setminus (de(V_i) \cup V_i)$.

Además, se dice que un conjunto C es un conjunto ancestral, y se denota como $an(C)$, cuando C contiene todos los ascendientes de los nodos que conforman dicho conjunto. En la figura 9 el conjunto $\mathrm{C} = \{\mathrm{A, B, C, D}\}$ es un conjunto ancestral.

Como se ha comentado anteriormente, un grafo dirigido, cuya notación vendrá dada por $D = (V, E)$, refleja una ordenación entre sus nodos. Si se le asigna un número a cada uno de los nodos, se dice que se tiene una numeración ancestral cuando el número de cada nodo es menor que el correspondiente a sus hijos. En la figura 10 se presenta una numeración ancestral de un grafo dirigido.

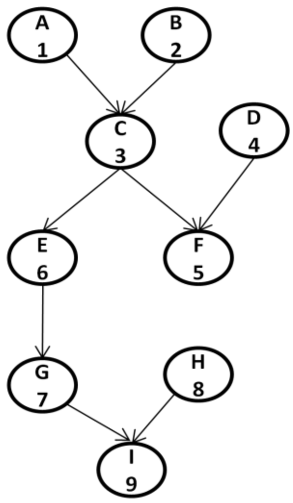


Figura 10 Numeración ancestral de un grafo dirigido

El camino cerrado en un grafo dirigido D se llama ciclo; (recuérdese que bucle es cuando el camino cerrado se encuentra en un grafo no dirigido). En función de la existencia o no de ciclos en el grafo se tiene la siguiente definición.

Definición de Grafo acíclico y cíclico

Un grafo dirigido $D = (V, E)$, es acíclico cuando no contiene ningún ciclo, en caso de contener al menos un ciclo el grafo es un grafo cíclico. (3)

En la Figura 11 se muestra un grafo dirigido acíclico (GDA). Esta estructura gráfica es básica para, posteriormente, poder especificar mediante un grafo un problema con incertidumbre en el que se relacionan un conjunto de variables.

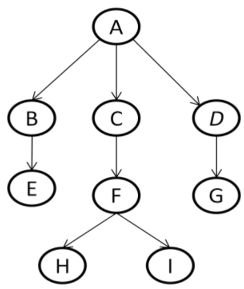


Figura 11 Grafo Dirigido Acíclico (GDA)

Para finalizar se tiene, que asociado a un grafo dirigido siempre existe un grafo no dirigido que se obtiene cambiando las aristas dirigidas del grafo por aristas no dirigidas, siendo el nuevo grafo el grafo no dirigido asociado. Ver figura 12 y figura 13

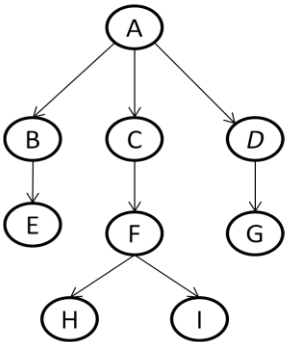


Figura 12 Grafo dirigido Acíclico

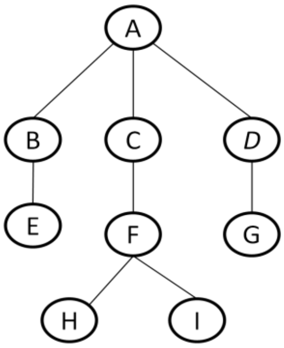


Figura 13 Grafo no dirigido asociado al GDA de la Figura 12

1.2.3 Conceptos básicos de Grafos Mixtos

Como se ha visto en la figura 4 los grafos mixtos tienen una parte de grafo dirigido y otra de grafo no dirigido. De esta forma, varias de las definiciones explicadas anteriormente para grafos dirigidos y no dirigidos también se pueden aplicar en este tipo de estructura. Ejemplo de ello son las definiciones presentadas para grafos dirigidos: padre e hijo de un nodo, padres de un conjunto de nodos, hijos de un conjunto de nodos; y las definiciones presentadas para grafos no dirigidos: vecino de un nodo y vecinos de un conjunto de nodos.

En la representación de un problema con incertidumbre, el grafo mixto que se utilice no puede tener ciclos dirigidos, es decir, que no puede presentar caminos cerrados formados por aristas dirigidas y esto se define como:

Definición de Grafo Cadena

Un grafo mixto $K = (V,E)$ es un grafo cadena cuando no contiene ningún ciclo dirigido. (3) Ver figura 14

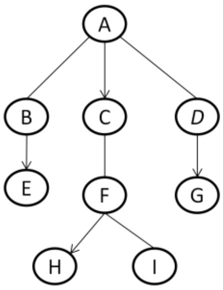


Figura 14 Grafo cadena

Al igual que sucede en los grafos dirigidos, es posible obtener el grafo no dirigido asociado a un grafo mixto. El grafo no dirigido asociado existe siempre y se obtiene tras cambiar las aristas dirigidas del grafo mixto por aristas no dirigidas.

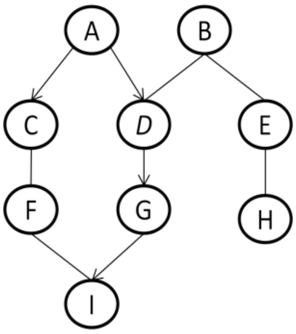


Figura 15 Grafo cadena

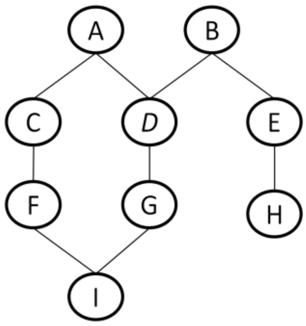


Figura 16 Grafo no dirigido asociado al grafo cadena de la Figura 15

1.3 Teoría de la Probabilidad

Luego de las definiciones básicas que hasta ahora se han visto, relacionadas con la Teoría de Grafos, en esta sección se expondrán una serie de conceptos y definiciones que permiten relacionar la Teoría de Grafos con la Teoría de la Probabilidad y que son necesarios para la comprensión de otras cuestiones que se verán con más profundidad en próximos capítulos. (4)

Cuando se aborda el tema de la probabilidad es necesario conocer qué son las variables y cuál es su clasificación al ser utilizadas. Por lo que primeramente se definirá qué es una variable:

Se le denomina **variable** al conjunto de los distintos valores numéricos que adopta un carácter cuantitativo y esta puede ser dependiente o independiente .

Una variable independiente es aquella que se utiliza para probar la dependencia de otra y se conoce que dos variables son independientes entre si por la definición que a continuación se enuncia*:*

Definición de Variables Independientes

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:

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

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

Las variables pueden ser además correlacionadas o dependientes, es decir, que a diferencia de las variables independientes en las variables correlacionadas la ocurrencia de una tiene relación directa con la ocurrencia de la otra. Se define como variable correlacionada:

Definición de Variables Correlacionadas o Dependientes

Dos variables X e Y están correlacionadas si no son independientes, es decir, si

$$P(x_i, y_j) \neq P(x_i)P(y_j)\ \exists i, j$$

Las variables pueden ser independientes dadas algunas condiciones específicas. La siguiente definición muestra las condiciones que deben darse para que dos variables sean condicionalmente independientes:

Definición de Variables Condicionalmente Independientes

Una variable X es condicionalmente independiente de otra variable Y dada una variable Z*,* si el conocer Z hace que X e Y sean independientes. Es decir, si conozco Z*,* Y no tiene influencia en X*. S*e puede decir que la independencia condicionada se cumple si y solo si se cumplen algunas de las siguientes condiciones:

$P(x|y, z) = P(x|z)$ con $P(z) > 0$

$P(x, y|z) = P(x|z)P(y|z)$ con $P(z) > 0$

$P(x, y, z) = P(x|z)P(y|z)p(z)$ con $P(z) > 0$

Esta definición se traduce a que cada variable es independiente de todas aquellas variables que no son sus “descendientes” una vez que se conocen sus propios nodos padres. 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.

1.4 Modelos Gráficos Probabilísticos

Cuando se construye un modelo matemático probabilístico es necesario tener la información cualitativa y cuantitativa relativa a la información de la cual se dispone y es por esta razón que al unir la Teoría de Grafos y la Teoría de la Probabilidad surgen los Modelos Gráficos Probabilísticos.

La información cualitativa del problema va a introducir la información asociada a las relaciones de dependencia que se establecen entre las variables del modelo, la cual se va a apoyar en la Teoría de Grafos. Esta información va a poder ser resumida en un grafo donde las variables del problema van a ser representadas por los nodos, y las aristas van a representar las dependencias y la causalidad entre las variables, así la falta de aristas induce a relaciones de independencia.

Al construir el modelo, se dispone además, de información relativa a la distribución de probabilidad de las variables del problema, denominada información cuantitativa del problema. Las distribuciones de probabilidad pueden ser estimadas a partir de la información que tienen los expertos relacionados con el problema que se está estudiando o mediante un conjunto de datos. Con la información cualitativa y cuantitativa se definirá el modelo gráfico probabilístico que estará asociado a dicho problema. (3)

En dependencia del tipo de grafo que se utilice a la hora de definir el problema se distinguen tres tipos de modelos gráficos probabilísticos:

1) Modelos gráficos probabilísticos no dirigidos. Redes de Markov.
2) Modelos gráficos probabilísticos dirigidos. Redes Bayesianas.
3) Modelos gráficos probabilísticos mixtos. Redes Cadena.

1.4.1 Modelos gráficos probabilísticos no dirigidos. Redes de Markov

Se emplean los modelos gráficos probabilísticos no dirigidos cuando las relaciones de dependencia entre las variables son relaciones de asociación o correlación, donde el grafo que va a representar la información cualitativa del problema es un grafo no dirigido.

Las Redes de Markov se utilizan básicamente en campos como la Física, la Robótica, para el análisis de imágenes y actualmente en el análisis de textos.

1.4.2 Modelos gráficos probabilísticos dirigidos. Redes Bayesianas

Cuando se conoce que el efecto de una variable X_j es producido por otra variable X_i , es decir, que las relaciones de dependencia son de tipo causal, se va a utilizar para representar la información cualitativa relacionada con el problema, un grafo dirigido en el cual las aristas dirigidas van a indicar la existencia de una relación causa-efecto entre las variables.

Considerando que la existencia de ciclos en el grafo dirigido hace mucho más engorrosa la obtención de la probabilidad conjunta, se trabaja con grafos acíclicos dirigidos para la representación de la información cualitativa del modelo gráfico probabilístico dirigido.

1.4.3 Modelos gráficos probabilísticos mixtos. Redes Cadenas

Cuando la información cualitativa del modelo indica la existencia de relaciones de dependencia causales y de asociación, el grafo adjunto al modelo gráfico probabilístico ha de ser un grafo mixto, con aristas dirigidas para representar las relaciones causales y aristas no dirigidas para mostrar las relaciones de asociación. En este caso, el grafo asociado al modelo probabilístico, es un grafo cadena.

1.5 Fundamentos teóricos de los modelos gráficos probabilísticos dirigidos

Como el modelo gráfico probabilístico a utilizar en el presente trabajo de diploma es el dirigido, a continuación se indican algunas definiciones y notaciones propias de la terminología de las Redes Bayesianas muy importantes para comprender el funcionamiento básico de los modelos gráficos probabilísticos dirigidos. (5)

A lo largo de este trabajo se denotarán las variables o nodos con letras mayúsculas como X y sus respectivos estados como x_i. Además, se debe tener presente que luego de imponérsele condiciones a una variable, es necesario en primer lugar saber que se está hablando de probabilidad condicional y luego conocer que se define como:

Definición de Probabilidad Condicional

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

La probabilidad condicional por definición es:

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

Análogamente, si se intercambia el orden de las variables:

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

A partir de las dos fórmulas anteriores se obtiene:

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

Luego de definir formalmente que es la probabilidad condicional se puede entonces hablar de probabilidad conjunta.

Definición de Probabilidad Conjunta

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

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

De igual forma la regla de la cadena sostiene que la probabilidad conjunta puede ser calculada como:

$$P(x_1, x_2, \dots, x_n) = \prod_{t-1}^{n} P(x_t|x_1, x_2, \dots, x_{t-1})$$

El siguiente ejemplo muestra el proceso para calcular la probabilidad conjunta de varias variables cuando se conoce la estructura gráfica de la red y sus respectivas probabilidades condicionales en función de sus padres:

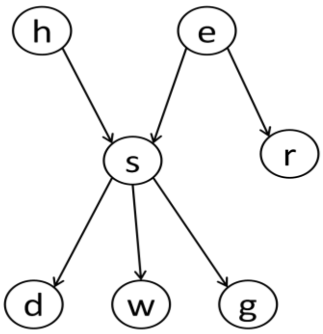


Figura 17 Red Bayesiana

La figura 17 representa una Red Bayesiana que contiene el conjunto de nodos {d, e, h, r, s, w, g}; si se elige una ordenación de los nodos de manera que se cumpla que todo nodo aparece antes que sus hijos se obtiene, por ejemplo, el conjunto {h, e, r, s, d, w, g}; aplicando la regla de la cadena y que cada nodo es independiente de sus predecesores conocidos sus padres se obtiene que:

$$P(h,e,r,s,d,w,g) = P(h)P(e|h)P(r|h,e)P(s|h,e,r)P(d|h,e,r,s)P(w|h,e,r,s,d) \ldots$$

$$P(h,e,r,s,d,w,g) = \cdots P(g|h,e,r,s,d,w)$$

$$P(h,e,r,s,d,w,g) = P(h)P(e)P(r|e)P(s|h,e)P(d|s)P(w|s)P(g|s)$$

Con la expresión anterior se calcula la probabilidad conjunta de todos los nodos que componen la red a partir de las probabilidades condicionales de cada nodo en función de sus nodos padres. Estas independencias condicionales son importantes porque simplifican la representación del conocimiento (menos parámetros) y el proceso de razonamiento o inferencia (propagación de probabilidades) que se verá más adelante.

Cuando es de interés conocer la distribución de alguna de las variables por separado, a partir de la información que se obtiene luego de calcular la probabilidad conjunta entonces se hace necesario enunciar otro concepto, el concepto de probabilidad marginal:

Definición de probabilidad marginal.

Dada una distribución de probabilidad conjunta $P(x_1, x_2, \ldots, x_n)$, la probabilidad marginal para un subconjunto de variables $X' = \{X'_1, X'_2, \ldots, X'_{n'}\} \subset X$ viene dada por:

$$P(x) = P(x'_1, x'_2, \ldots, x'_n) = \sum_{x_i | X_i \notin X'} P(x_1, x_2, \ldots, x_n)$$

Luego de enunciar los conceptos de probabilidad condicional, probabilidad conjunta y probabilidad marginal, es necesario conocer otras definiciones que se utilizarán posteriormente y que serán de mucha ayuda.

Para hacer referencia a definiciones como probabilidad a priori y probabilidad a posteriori es fundamental conocer que es la evidencia. Por lo tanto se dice que existe evidencia acerca de una variable cuando es conocido el estado en que se encuentra la misma.

Después de conocer que es la evidencia, se pueden enunciar entonces dos definiciones que se utilizarán con mucha frecuencia.

Definición de Probabilidad a Priori

Es la probabilidad de una variable en ausencia de evidencia

Definición de Probabilidad a Posteriori

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

1.6 Estudio sobre Redes Bayesianas en Cuba

Actualmente en Cuba en el ámbito de la Inteligencia Artificial se maneja con mayor frecuencia los conceptos de Realidad Virtual y Redes Bayesianas. En el Departamento de Ciencias de la Computación de la Universidad Central "Marta Abreu" de Las Villas se desarrolló un estudio basado en el uso de las Redes Bayesianas utilizando aspectos de Mapas Conceptuales como una perspectiva para los sistemas de enseñanza y aprendizaje inteligentes. (6)También en esta misma Universidad se llevó a cabo un estudio para la obtención de una Red Bayesiana a partir de factores de riesgo de la Hipertensión Arterial, en el cual se propone un método para obtener la topología de una Red Bayesiana. La Red Bayesiana obtenida es útil para realizar tanto el diagnóstico de la Hipertensión Arterial, como para analizar cual sería el comportamiento de los atributos predictores más significativos. (7)

Además, dentro del Plan Integral de Salud que ofrece Cuba, el Proyecto ELAM (Escuela Latinoamericana de Medicina) tiene la responsabilidad de formar recursos humanos para su sostenibilidad a largo plazo. Teniendo en cuenta la complejidad que para el Proyecto ELAM implica el seguimiento y atención de sus estudiantes y graduados, resulta necesario e imprescindible crear un grupo observatorio de recursos humanos que permita, además de generar una enorme cantidad de información, analizarla con elevado rigor científico, de manera que pueda ser ofrecida a los decidores una información oportuna, eficaz y pertinente con el fin de ser utilizada en la toma de decisiones.

Debido a que esto implica el almacenamiento de una gran cantidad de información y para evitar problemas de eficiencia y escalabilidad, se utiliza la minería de datos y dentro de esta una de sus técnicas más comunes; las Redes Bayesianas, las cuales han acaparado la atención de los investigadores por las enormes ventajas que presentan para tratar los problemas en presencia de incertidumbre.

1.7 Aplicaciones de las Redes Bayesianas

Muchas son las aplicaciones que poseen las Redes Bayesianas en la actualidad. Dichas redes son utilizadas en diversas áreas como la medicina, la ciencia y la economía. Pueden ser aplicadas en cualquiera de las siguientes actividades de negocio: prevención del fraude, prevención del abandono de clientes, blanqueo de dinero, marketing personalizado, mantenimiento preventivo y scoring de clientes. Además, son utilizadas en empresas como: Microsoft: Answer Wizard (Office): diagnóstico de problemas de impresora, Intel: Diagnóstico de fallos de procesadores, HP: Diagnóstico de problemas de impresora, Nasa: Ayuda a la decisión de misiones espaciales. (8)

En la medicina son utilizadas en los sistemas de apoyo a las decisiones clínicas (DSS) con los cuales se han logrado producir aplicaciones que compiten con la pericia diagnóstica clínica de un médico. Los DSS tienen un gran potencial como herramientas para capacitar a estudiantes de medicina y médicos jóvenes en razonamiento diagnóstico. Un ejemplo exitoso de ello es un sistema experto desarrollado por la Universidad de Harvard**.** Debido a que el razonamiento clínico es un proceso complejo que involucra numerosas variables y un considerable grado de incertidumbre, técnicas de inferencia probabilística como las Redes Bayesianas resultan de gran utilidad. Las Redes Bayesianas, a diferencia de las Redes Neuronales, tienen la ventaja de poder explicar su “lógica diagnóstica” al usuario. Este beneficio es considerable para lograr la aceptación de un sistema experto; los médicos difícilmente aceptarían un diagnóstico o una decisión que no pueda ser explicada racionalmente y de modo comprensible.

En el libro AI for Game Developers se propone el uso diversas técnicas para desarrollar juegos con Inteligencia Artificial, se dan ejemplos sobre la utilización de Redes Bayesianas para tomar decisiones cuando los estados del mundo de juego son inciertos. Se propone además ejemplos de como utilizar Redes Bayesianas en juegos que incluyen peleas de Kung Fu. (9)

En los juegos de aventuras existen múltiples posiciones y acciones que los jugadores pueden realizar, en (10) se propone un Modelo Bayesiano para lograr un primer acercamiento para un plan de reconocimiento de entrada para juegos de aventuras con el objetivo de identificar los planes de usuarios y las metas, en este libro se proponen estructuras de redes que representan la relación en el dominio para varias extensiones y comparan su poder para predecir la meta actual de un usuario, la siguiente acción y la siguiente posición.

CAPÍTULO 2. REDES BAYESIANAS

Introducción

En este capítulo se profundiza en las Redes Bayesianas de forma general y en los distintos tipos de redes que se obtienen en función de las variables del modelo; estudiándose con más profundidad las Redes Bayesianas Discretas, además se expone la definición de Red Bayesiana y se presenta la aplicación del Teorema de Bayes a través de varios ejemplos. En otra de las secciones se trata lo relacionado con la inferencia, y se ofrece una descripción de los algoritmos de propagación para cada una de las estructuras gráficas existentes. Para finalizar el presente capítulo se expone lo referente al aprendizaje en las Redes Bayesianas además de sus ventajas y desventajas.

2.1 Redes Bayesianas

Las Redes Bayesianas son una clase de modelos gráficos probabilísticos dirigidos, asociados a relaciones de dependencia de tipo causal. 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.

Definición de Red Bayesiana

Una Red Bayesiana es un grafo dirigido acíclico conexo en el que cada nodo representa una variable aleatoria y cada arco representa una dependencia probabilística; en la cual se especifica la probabilidad condicional de cada variable dados sus padres.

Las variables aleatorias que se utilizan en una Red Bayesiana pueden ser de dos tipos: variables discretas o variables continuas, las cuales se definen de la siguiente forma.

Definición de Variable Discreta

Una variable es discreta si el conjunto de valores posibles es finito, es decir, si los valores que asume se pueden contar.

Ejemplo: Presencia de una enfermedad, Número de asignaturas matriculadas, Sexo

Definición de Variable Continua

Una variable es continua si toma cualquier valor en un intervalo de los números reales, por lo cual tiene un número infinito de valores posibles.

Ejemplo: Distancia de un robot a la pared, Altura, Peso.

Una variable puede tener uno o varios estados, pero dichos estados deben cumplir con las siguientes dos propiedades.

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

Los arcos en una Red Bayesiana representan influencias causales, es decir, el que un nodo sea padre de otro implica que es causa directa del mismo. Por ejemplo, en el juego Grand Theft Auto, considerando las variables categóricas, robar auto (R) y persecución policial (P), existe una relación causal entre las mismas cuando se hace la suposición de que robar autos presenta un efecto directo sobre la existencia de una persecución policial (R→P).

Cada arco además representa una dependencia probabilística entre aquellos nodos que tienen una relación causal; por lo que cada nodo tendrá asociada una tabla de probabilidad condicional indicando la probabilidad de sus estados para cada combinación de los estados de sus padres. En el caso de que un nodo no tenga padres se indicarán sus probabilidades a priori.

De forma general una Red Probabilística se puede construir de la siguiente forma:

a) Se asigna un nodo a cada variable (X_i) y se indica de que otros nodos es una causa directa; a ese conjunto de nodos "causa del nodo X_i " se lo denota como el conjunto $pa(x_i)$ y se lo llamará "padres de X_i".
b) Se une cada padre con sus hijos, con arcos que parten de los padres y llegan a los hijos.
c) A cada variable X_i se le asigna una matriz $P(x_i|pa(x_i))$ que estima la probabilidad condicional de un evento X_i=x_i dada una combinación de valores de los $pa(x_i)$.

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 las variables de la red; para ello se debe calcular la probabilidad a posteriori de cada variable condicionada a la evidencia, la cual se podrá obtener haciendo uso del Teorema de Bayes, pero antes de comenzar a explicar el funcionamiento de dicho teorema se describirán los distintos tipos de Redes Bayesianas existentes, puesto que el cálculo de las probabilidades se realiza de forma diferente en cada caso.

2.2 Tipos de Redes Bayesianas

En dependencia de las variables aleatorias que se utilicen para modelar la red existen los diferentes tipos de Redes Bayesianas que a continuación se muestran.

Redes Bayesianas Discretas:

Las Redes Bayesianas Discretas se caracterizan porque todas las variables del modelo son discretas, de forma que cada variable solo puede tomar un conjunto finito de valores; y cuando además todas las variables del problema son binarias la red se denota como **Red Bayesiana Multinomial**.

Redes Bayesianas Gaussianas:

Una Red Bayesiana es gaussiana si todas las variables que intervienen en el problema son variables continuas, es decir, si puede asumir todos los valores posibles en un intervalo x-y.

Redes Bayesianas Mixtas:

Las Redes Bayesianas Mixtas, también denotadas como Redes Bayesianas Discretas-Gaussianas, se caracterizan por incluir variables discretas y continuas en el modelo gráfico probabilístico dirigido. Para poder especificar el modelo, las variables discretas deben preceder a las continuas en el grafo.

En el presente trabajo el tipo de red a utilizar son las Redes Bayesianas Discretas debido a que las variables que se van a emplear son discretas. Además, teniendo en cuenta que a medida que aumente la complejidad de los métodos bayesianos, también aumentan los requerimientos computacionales necesarios para el proceso de compilación; se llega a la conclusión de que para la toma de decisiones en entornos virtuales, no es recomendable la utilización de métodos complejos como los que implementan las Redes Bayesianas Gaussianas y las Redes Bayesianas Mixtas ya que en este tipo de entornos los algoritmos deben ser optimizados en cuanto al uso de los recursos de la computadora.

2.3 Teorema de Bayes

Las Redes Bayesianas toman dicho nombre debido a la utilización del Teorema de Bayes en el cálculo de las probabilidades a posteriori, dicho teorema se expresa como: .

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

Siendo la sumatoria que se encuentra en el denominador el **Teorema de la Probabilidad Total.**

A continuación se muestra un ejemplo de una Red Bayesiana donde se explica el significado de las variables que intervienen en el problema y como influyen las relaciones entre las variables en la aplicación del Teorema de Bayes; para de esta forma obtener las probabilidades a posteriori de una determinada variable.

Ejemplo 1

La Red Bayesiana más sencilla que se puede imaginar consta de dos variables, en este caso X e Y, y una arista dirigida desde X hasta Y.

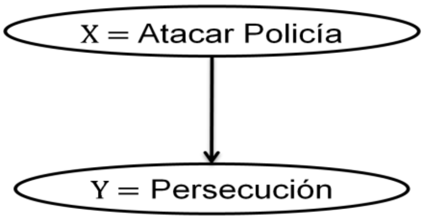


Figura 18 Red Bayesiana que representa el Ejemplo 1

El presente ejemplo está basado en el juego de consola Grand Theft Auto, para ello la variable X significa "Atacar Policía" y la variable Y representa "Persecución Policial". Para una mejor comprensión de este ejemplo, se puede decir que la agresión a un policía provoca que se lleve a cabo una persecución policial como se muestra en la figura 18.

En este caso las variables que se utilizan son binarias puesto que las variables van a tener solamente 2 estados que describen la ocurrencia o no de dicha variable. Por ejemplo, los estados de la variable X se denotan como x_1 y x_2 donde x_1 representará la probabilidad de que un policía sea atacado y x_2 lo contrario. De igual forma los estados de la variable Y se denotan como y_1 y y_2 donde y_1 significará la presencia de una persecución y y_2 la ausencia de la misma.

La información cuantitativa de una Red Bayesiana viene dada por:

a) La probabilidad a priori de los nodos que no tienen padres.
b) La probabilidad condicionada de los nodos con padres.

Ahora en este ejemplo, los datos que se conocen son $P(x)$ y $P(y|x)$.

La Red Bayesiana completa quedará como se muestra la figura 19:

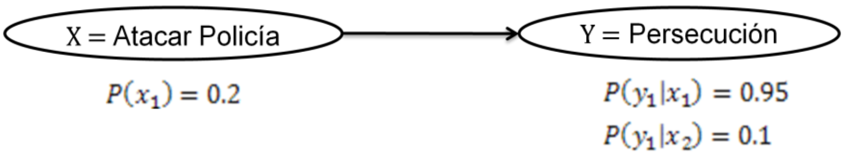


Figura 19 Red Bayesiana donde se muestran las probabilidades a Priori y probabilidades condicionales del Ejemplo 1

En este caso estos valores tienen un significado el cual se explicará a continuación:

a) $\mathrm{P}(\mathrm{x}_1) = 0.2$ indica que, a priori, existe un 20% de probabilidad de que un policía sea atacado.

b) $P(y_1|x_1) = 0.95$ indica que cuando un policía es atacado, se efectuará una persecución policial en el 95% de los casos.

c) $P(y_1|x_2) = 0.1$ indica que cuando un policía no es atacado, se llevará a cabo una persecución policial en el 10% de los casos.

Importante señalar que las probabilidades complementarias no se muestran en la figura 2.2 debido a que se pueden obtener fácilmente, por ejemplo:

$$P(x_2) = 1 - P(x_1) = 0.8$$

$$P(y_2|x_1) = 1 - P(y_1|x_1) = 0.05$$

$$P(y_2|x_2) = 1 - P(y_1|x_2) = 0.9$$

Luego de conocer los datos se puede entonces calcular:

a) La probabilidad a priori de Y

Dicha probabilidad se calcula mediante el Teorema de la Probabilidad Total de la siguiente forma:

$$\begin{aligned} P(y_1) &= P(y_1|x_1)P(x_1) + P(y_1|x_2)P(x_2) \\ &= 0.95 * 0.2 + 0.1 * 0.8 \\ &= 0.27 \\ P(y_2) &= P(y_2|x_1)P(x_1) + P(y_2|x_2)P(x_2) \\ &= 0.05 * 0.2 + 0.9 * 0.8 \\ &= 0.73 \end{aligned}$$

b) Las probabilidades a posteriori dada una evidencia observada *e,* $P^*(x)=P(x|e)$

Suponiendo que se está en presencia de una persecución policial ($e = y_1$). ¿Qué probabilidad habrá de que un policía sea atacado? Para calcularla se puede aplicar el Teorema de Bayes, por lo tanto se busca $P^*(x_1) = P(x_1|y_1)$.

$$P^*(x_1) = P(x_1|y_1) = \frac{P(y_1|x_1)P(x_1)}{P(y_1)} = \frac{0.95 * 0.2}{0.27} = 0.70$$

Lo que quiere decir que cuando se produce una persecución policial, hay un 70% de probabilidad de que un policía sea atacado.

De igual forma se puede calcular $P^*(x_2)$, que es la probabilidad complementaria.

$$P^*(x_2) = P(x_2|y_1) = \frac{P(y_1|x_2)P(x_2)}{P(y_1)} = \frac{0.1 * 0.8}{0.27} = 0.30$$

Como se puede observar la suma de las probabilidades a posteriori de la variable X es igual a uno, es decir, $P^*(X) = P^*(x_1) + P^*(x_2) = 1$

La expresión general del Teorema de Bayes que se ha utilizado es:

$$P^*(x) = P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

Luego se puede reescribir como sigue:

$$P^*(x|y) = \alpha P(y|x)P(x)$$

Donde $\alpha = [P(y)]^{-1}$

Con la fórmula expresada de esta forma, queda claro que la probabilidad a posteriori de la variable X depende fundamentalmente de la probabilidad a priori de X y de la probabilidad condicionada de Y dado X, puesto que α juega simplemente el papel de una constante de normalización, es decir, α es la encargada de que la suma de las probabilidades de la variable sea igual a uno.

Utilizando esta nueva expresión, se pueden repetir los cálculos:

$$P^*(x_1) = \alpha\, 0.95 * 0.2 = 0.19\, \alpha$$

$$P^*(x_2) = \alpha\, 0.1 * 0.8 = 0.08\, \alpha$$

Y al normalizar se obtiene el mismo resultado que antes .

Para el caso en que no exista una persecución policial ($e = y_2$), la probabilidad a posteriori de que el policía sea atacado se calcula de igual forma.

Ejemplo 2

En el ejemplo anterior se tiene que la agresión a un policía provoca que se desarrolle una persecución policial. Si se amplia este modelo y se le añade una nueva variable Z a la red como hija de X llamada "Policía Dispara", esto significa que el ataque a un policía va a causar además que dicho policía dispare o no. En este caso al ser Z una variable binaria se representarán los estados como z_1 cuando el policía dispara y z_2 cuando el policía no dispara. La red bayesiana modificada queda como se muestra en la figura 20.

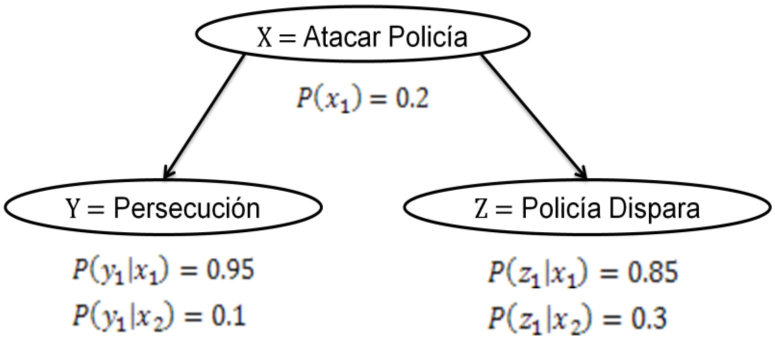


Figura 20 Red Bayesiana que representa el Ejemplo 2

Ahora en el ejemplo nótese que, dado que el policía es atacado, la probabilidad de que el policía dispare es menor que la probabilidad de que haya una persecución policial, pues en este modelo se tiene en cuenta que hay muchas otras causas por las que el policía puede disparar.

¿Qué tipo de conclusiones se pueden obtener a partir de esta información?

1. Si se supone que $e = \{z_1\}$. Entonces se puede calcular como antes la probabilidad a posteriori de que el policía sea atacado sabiendo que dicho policía ha disparado:

$$P^*(x_1) = P(x_1|z_1) = \alpha\, 0.85 * 0.2 = 0.17\, \alpha = 0.41$$

$$P^*(x_2) = P(x_2|z_1) = \alpha\, 0.3 * 0.8 = 0.24\, \alpha = 0.59$$

Como se puede observar solo hay un 41% de probabilidad de que el policía sea atacado, un resultado mucho más bajo que antes.

2. Suponiendo que $e = \{y_1, z_1\}$. ¿Cuál es ahora $P^*(x) = P(x|y_1, z_1)$?

Para realizar este cálculo se utiliza nuevamente le Teorema de Bayes:

$$P^*(x) = P(x|y_1, z_1) = \frac{P(y_1, z_1|x)P(x)}{P(y_1, z_1)}$$

Como se muestra en la expresión anterior ahora surgen nuevos datos del problema que no se conocen, como $\mathrm{P}(y_1, z_1)$ y $\mathrm{P}(y_1, z_1|\mathrm{x})$. Pero, aplicando las definiciones de independencia condicional explicadas en el capítulo anterior, se tiene que las variables Y y Z son independientes dados su padre común, X, es decir:

$$P(y_1, z_1|x) = P(y_1|x)P(z_1|x)$$

Entonces se puede continuar con los cálculos pues $\mathrm{P}(y_1, z_1)$ se obtendrá como constante de normalización.

$$\propto = [\, P(y_1|x_1)P(z_1|x_1)P(x_1) + P(y_1|x_2)P(z_1|x_2)P(x_2)]^{-1} = [0.186]^{-1} = 5.38$$

$$P^*(x_1) = P(x_1|y_1, z_1) = \propto P(y_1|x_1)P(z_1|x_1)P(x_1) = \propto 0.95 * 0.85 * 0.2 = 0.87$$

$$P^*(x_2) = P(x_2|y_1, z_1) = \propto P(y_1|x_2)P(z_1|x_2)P(x_2) = \propto 0.1 * 0.3 * 0.8 = 0.13$$

Como se puede apreciar, cuando se tienen dos evidencias a favor del ataque al policía, la probabilidad resultante es mayor que la correspondiente a cada uno de ellos por separado.

3. Ahora, cuando se tiene una evidencia a favor y otra en contra, se pueden calcular las probabilidades a posteriori de forma similar al caso anterior. Por ejemplo, suponiendo que se tiene la certeza de que no se está desarrollando una persecución policial y además se conoce que el policía ha disparado, es decir, $\mathrm{e} = \{\mathrm{y}_2, \mathrm{z}_1\}$, entonces ¿Cuál es la probabilidad de que el policía sea atacado?

Aplicando el Teorema de Bayes esto se calcula como:

$$P^*(x_1) = P(x_1|y_2, z_1) = \propto P(y_2|x_1)P(z_1|x_1)P(x_1) = \propto 0.0085 = 0.04$$

$$P^*(x_2) = P(x_2|y_2, z_1) = \propto P(y_2|x_2)P(z_1|x_2)P(x_2) = \propto 0.216 = 0.96$$

Como se puede apreciar hay mas evidencia a favor de x_2 que a favor de x_1, esto se debe fundamentalmente a la alta probabilidad de que se desarrolle una persecución policial cuando un policía es agredido por lo que cuando no se produce una persecución la probabilidad de que un policía sea agredido es muy baja

4. En este momento, todavía es posible extraer más información del ejemplo. Si ahora se supone que se tiene evidencia de que un policía ha disparado, es decir, $e = \{z_1\}$, ¿Cuál es la probabilidad de que se produzca una persecución policial?

Se busca ahora $P(y_1|z_1)$, esta probabilidad se calcula como:

$$P^*(y_1) = P(y_1|z_1) = \sum_X P(y_1|x, z_1)P(x|z_1) = \sum_X P(y_1|x, z_1)\frac{P(x, z_1)}{P(z_1)}$$

Teniendo en cuenta que

$$P(x, z_1) = P(z_1|x)P(x)$$
$$\alpha = [P(z_1)]^{-1}$$

La expresión anterior queda de la siguiente forma

$$P^*(y_1) = \alpha \sum_X P(y_1|x)\, P(z_1|x)P(x)$$

Entonces se está en condiciones de realizar los cálculos necesarios para obtener la probabilidad a posteriori de Y.

$$\alpha = [0.186\ + 0.2245]^{-1} = 2.436$$

$$P^*(y_1) = \alpha\, [P(y_1|x_1)(z_1|x_1)P(x_1) + P(y_1|x_2)P(z_1|x_2)P(x_2)] = \propto 0.186 = 0.45$$

$$P^*(y_2) = \alpha\, [P(y_2|x_1)(z_1|x_1)P(x_1) + P(y_2|x_2)P(z_1|x_2)P(x_2)] = \propto 0.2245 = 0.55$$

Resulta interesante que se comparen las expresiones utilizadas para calcular la probabilidad a priori $P(y_1)$ y la a posteriori $P^*(y_1)$. De esta forma queda claramente explicado el funcionamiento del Teorema de Bayes en el cálculo de las probabilidades.

2.4 Propagación de la evidencia

Una de las características fundamentales asociadas a las Redes Bayesianas es el proceso de propagación de la evidencia, este proceso también conocido como inferencia, consiste en actualizar la información probabilística de las variables de la red cuando se tiene evidencia acerca del estado de uno o varios nodos. Esta información actualizada de las variables es la probabilidad a posteriori de las mismas.

Un primer método de actualización de las probabilidades dada la evidencia disponible, se puede obtener a partir de la probabilidad marginal $\mathrm{P(X|Y)}$ que a su vez puede obtenerse de la probabilidad conjunta P($x_1, x_2, \dots, x_i$) sumando los valores para todas las variables que no pertenezcan al conjunto $X \cup Y$. Pero en la práctica, esto no es viable ya que el tiempo de procesamiento necesario para realizar los cálculos crece exponencialmente a medida que aumenta el número de variables de la red, por este motivo se han desarrollado varios algoritmos de propagación de la evidencia más eficientes. En el siguiente subepígrafe se desarrollará este tema con mayor profundidad. (5)

2.4.1 Algoritmos de Propagación y Estructuras Gráficas

La estructura gráfica de la red es una de las ventajas que brindan las Redes Probabilísticas debido a que ofrece información acerca de las relaciones entre las variables por lo que se puede reducir el número de operaciones necesarias para obtener las probabilidades a posteriori. Existen diversos 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 como calculen los valores de las probabilidades. Los métodos exactos calculan los valores por medio del Teorema de Bayes de forma precisa y sin error, mientras que los métodos aproximados utilizan técnicas de simulación para obtener valores aproximados de las probabilidades y se aplican cuando los exactos son muy costosos o inaplicables.

En las Redes Bayesianas existen tres tipos de topologías de red; dos de ellas son estructuras sencillamente conectadas como son: los árboles y los poliárboles. Los árboles son estructuras en las que habrá una trayectoria simple entre cada par de nodos, mientras que en el caso de los poliárboles son estructuras similares pero con la diferencia de que un nodo puede tener varios padres. La última de las estructuras es mucho más compleja que las dos anteriores puesto que no está conectada en forma sencilla, esta se denomina: redes multiconectadas. Para un mejor entendimiento de la estructura gráfica de la red, en la siguiente figura se muestran los tres tipos de topologías existentes.

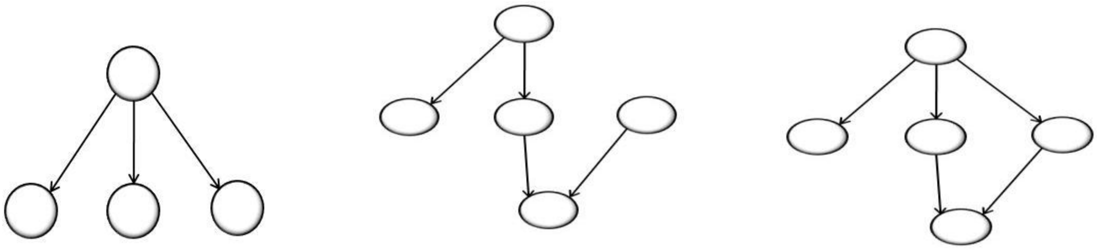


Figura 21 Tipos de Estructuras Arboles Poliárboles Redes Multiconectadas

Para cada tipo de estructura existen diferentes vías de hacer inferencia, dichos algoritmos de propagación de la evidencia se describen a continuación comenzando por la estructura más simple hasta llegar a la más compleja.

Propagación en árboles

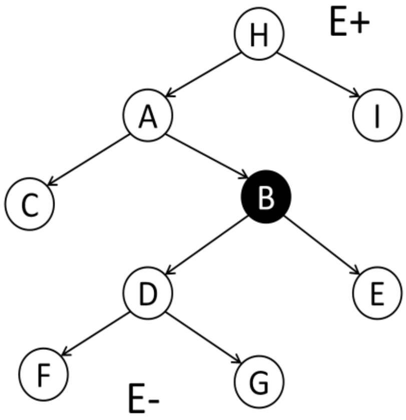
Para el caso en que la estructura gráfica de la red sea un árbol se utilizará el siguiente algoritmo de propagación: (11)

Como el tipo de red que se está analizando son las Redes Bayesianas Discretas, cada nodo corresponde a una variable discreta $A = \{A_1, A_2, \dots, A_n\}$ con su respectiva matriz de probabilidad condicional $P(B|A) = P(B_j|A_i)$. Dada cierta evidencia E (representada por la instanciación de ciertas variables) la probabilidad posterior de cualquier variable B es, por el Teorema de Bayes:

$$P(B_i|E) = \frac{P(E|B_i)P(B_i)}{P(E)}$$

Debido a que la estructura de la red es un árbol, el nodo B la separa en dos subárboles; de esta manera se puede dividir la evidencia en dos grupos:

- E^- : Datos en el árbol que cuya raíz es B.
- E^+ : Datos en el resto del árbol.



.

Figura 22 Red Bayesiana con topología de árbol

Entonces:

$$P(B_i|E) = \frac{P(E^-, E^+|B_i)P(B_i)}{P(E)}$$

Pero dado que ambos son independientes, se aplica nuevamente Bayes y se obtiene:

$$P(B_i|E) = \alpha P(B_i|E^+)P(E^-|B_i)$$

Donde α es la constante de normalización.

Esto separa la evidencia para actualizar la probabilidad de B; además se observa que no se requiere de la probabilidad a priori excepto en el caso de la raíz donde dado que dicho nodo no posee padres se tiene que:

$$P(A_i|E^+) = P(A_i)$$

Para simplificar el desarrollo se definen los siguientes términos:

$$\lambda(B_i) = P(E^-|B_i)$$

$$\pi(B_i) = (B_i|E^+)$$

Entonces:

$$P(B_i|E) = \alpha\pi(B_i)\lambda(B_i)$$

En base a la ecuación anterior se puede integrar un algoritmo distribuido para obtener la probabilidad de un nodo dada cierta evidencia. Para ello se descompone el cálculo de cada parte en:

- Evidencia de los hijos (λ)
- Evidencia de los demás nodos (π)

Debido a que por la propiedad de independencia condicional los hijos son condicionalmente independientes dado el padre:

$$\lambda(B_i) = \prod_k P(E_k^-|B_i) = \prod_k \lambda_k(B_i)$$

Donde E_K^- corresponde a la evidencia del subárbol del hijo k de B.

Condicionando cada término en la ecuación anterior respecto de todos los posibles valores de cada nodo hijo se obtiene:

$$\lambda(B_i) = \prod_k \left[\sum_j P\left(E_k^- \middle| B_i, S_j^k\right) P(S_j^k|B_i) \right]$$

Donde S_k es el hijo k de B, y la sumatoria es sobre los valores de dicho nodo (Teorema de la Probabilidad Total)

Dado que B es condicionalmente independiente de la evidencia dados sus hijos y sustituyendo la definición de λ *:*

$$\lambda(B_i) = \prod_k \left[\sum_j P\left(S_j^k \middle| B_i\right) \lambda(S_j^k) \right]$$

En forma análoga se obtiene una ecuación para π *;* primero se la condiciona sobre todos los posibles valores del padre:

$$\pi(B_i) = \sum_j P\left(B_i \middle| E^+, A_j\right) P(A_j|E^+)$$

Luego se puede eliminar E^+ del primer término debido a la independencia condicional. El segundo término representa la probabilidad posterior de A sin contar la evidencia del subárbol de B, por lo que se puede expresar usando la ecuación para $P\left(B_j \middle| E\right)$ y la descomposición de λ .

$$\pi(B_i) = \sum_j P(B_i|A_j)\left[\alpha\pi(A_j)\prod_k \lambda_k(A_j)\right]$$

Donde k incluye a todos los hijos de A excepto B.

Algoritmo

Mediante estas ecuaciones se integra un algoritmo de propagación de probabilidades en árboles donde cada nodo guarda los valores de los vectores π y λ así como las matrices de probabilidad P**;** la propagación se hace por un mecanismo de paso de mensajes en donde cada nodo envía los mensajes correspondientes a su padre e hijos:

Mensaje al padre (nodo B a su padre A*)*:

$$\lambda_B(A_i) = \sum_j P(B_j|A_i)\lambda(B_j)$$

Mensaje a los hijos (nodo B a su hijo S_k):

$$\pi_k(B_i) = \alpha\pi(B_j)\prod_{l\neq k} \lambda_l(B_j)$$

Al instanciarse ciertos nodos, éstos envían mensajes a sus padres e hijos y se propagan hasta llegar a la raíz u hojas o hasta encontrar un nodo instanciado. Esto se puede hacer en forma iterativa instanciando ciertas variables y propagando su efecto y luego instanciando otras variables y propagando la nueva información combinando ambas evidencias.

Propagación en poliárboles

Como se explicó al inicio del presente epígrafe, un poliárbol es una red conectada en forma sencilla, pero en la que un nodo puede tener varios padres. (11)

$$P(B_i|A_1, \dots \dots, A_n)$$

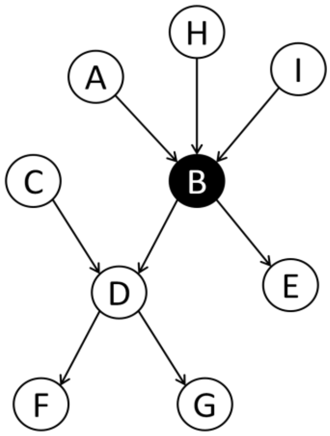


Figura 23 Red Bayesiana con topología de poliárbol

Algoritmo

El algoritmo de propagación en este caso es muy parecido al de árboles, con algunas consideraciones adicionales:

- Considerar la probabilidad condicional del nodo dados todos sus padres para el cálculo de π y λ.
- Enviar los λ-mensajes a cada uno de los padres de un nodo.

Ejemplo de propagación

1. Inicialmente se propagan todos los π-mensajes para inicializar la red como se muestra en la figura 24

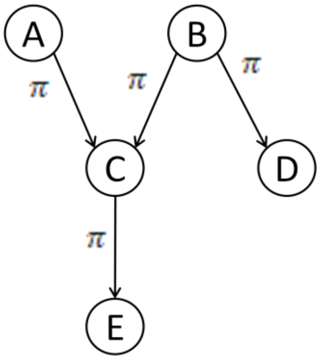


Figura 24 Estado Inicial de la red

- El nodo A envía un π-mensaje al nodo C y el nodo B envía π-mensajes a los nodos C y D
- Los nodos C y D recalculan sus probabilidades.
- El nodo C envía un π-mensaje al nodo E.
- El nodo E recalcula su probabilidad.

2. A continuación se procede a instanciar el nodo E

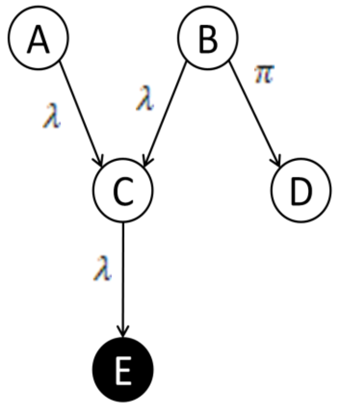


Figura 25 Instanciación del nodo E

- Un λ-mensaje es enviado del nodo E al nodo C.
- La probabilidad a posteriori del nodo C es recalculada a partir de la nueva evidencia.
- Se envían mensajes a todas las relaciones del nodo C excepto de nuevo al nodo E.
- Los nodos A y B recalculan sus probabilidades.
- El nodo B envía un π-mensaje al nodo D
- El nodo D recalcula su probabilidad.

3. Ahora se procede a instanciar el nodo B como se muestra en la figura 26.

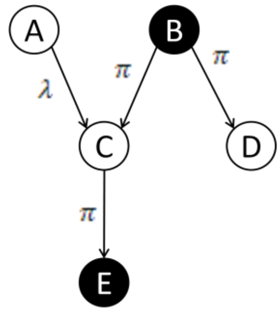


Figura 26 Instanciación de los nodos E y B

- El nodo B envía π-mensaje a los nodos C y D.
- Los nodos C y D recalculan sus probabilidades.
- El nodo C envía un λ-mensaje hacia el nodo A y un π-mensaje al nodo E.
- Como el nodo E se encuentra instanciado no modifica su estado.
- El nodo A recalcula su probabilidad basado en el mensaje recibido del nodo C.

4. Finalmente se instancia el nodo C, ver figura 27.

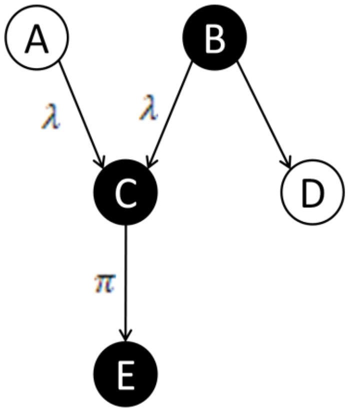


Figura 27 Instanciación de los nodos E, B y C

- Los λ-mensajes son enviados a los nodos A y B mientras que un π-mensaje se envía al nodo E.
- Los nodos B y E se mantienen sin modificaciones pero el nodo A recalcula su probabilidad.
- Dado que el nodo B se encuentra instanciado, éste no envía ningún π-mensaje al nodo D.

Propagación en redes multiconectadas

Las redes multiconectadas representan la estructura gráfica más compleja; puesto que esta topología de red es un grafo no conectado en forma sencilla, es decir, en el que hay múltiples trayectorias entre nodos; en este tipo de red probabilística los métodos anteriores ya no aplican pero existen otras técnicas alternativas que se detallan a continuación: (12)

- **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 nodos conectados en forma sencilla; 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. 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; por lo tanto no se recomienda la utilización de esta topología de red para su empleo en entornos virtuales por la complejidad de los métodos necesarios para realizar la inferencia.

2.5 El aprendizaje en las Redes Bayesianas

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. 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 es 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; a estas dos fases se las puede denominar respectivamente aprendizaje estructural y aprendizaje paramétrico. (5)

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. Consiste en encontrar los parámetros asociados a una estructura dada de una Red Bayesiana. Dichos parámetros consisten en las probabilidades a priori de los nodos raíz y las probabilidades condicionales de las demás variables dados sus padres; si se conocen todas las variables es fácil obtener las probabilidades requeridas ya que las probabilidades previas corresponden a las marginales de los nodos raíz y las condicionales se obtienen de las conjuntas de cada nodo con su(s) padre(s).

2.6 Ventajas y desventajas de las Redes Bayesianas

El hecho de que las Redes Bayesianas constituyan una mezcla de técnicas estadísticas y modelos gráficos les provee una serie de ventajas importantes. 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 representación 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 en un problema determinado, en el cual se obtienen las causas más probables dado un conjunto de variables; como para hacer razonamiento predictivo (obteniendo la probabilidad de ocurrencia de una de estas variables 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 (5). 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 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.

Aunque la utilización de las Redes Bayesianas se presenta actualmente como una de las técnicas de Inteligencia Artificial con gran aceptación, una de las grandes desventajas de su empleo es la gran cantidad de probabilidades numéricas que se necesita conocer para su utilización. Además, el empleo de estas redes implica un alto costo computacional y asimismo por la complejidad de la implementación de los algoritmos de inferencia se requiere de gran espacio en memoria. También para desarrollar una red de este tipo es necesaria la realización de estudios estadísticos.

CAPÍTULO 3. PROPUESTA DE SOLUCIÓN

Introducción

En este capítulo se expondrá la propuesta de solución del problema científico que consta de un algoritmo para propagar las probabilidades en la Red Bayesiana dado un evento. También se abordará específicamente la toma de decisiones de los elementos inteligentes en Entornos Virtuales mediante las Redes Bayesianas. Además, se describe un caso de uso donde se vincula una Red Bayesiana a un Entorno Virtual.

3.1 Algoritmo de propagación

La topología de Red Bayesiana que se utilizará para aplicar el algoritmo es una estructura de árbol, debido a que tiene la característica de ser una estructura conectada en forma sencilla, lo que la convierte en una estructura óptima para su aplicación en entornos virtuales; puesto que a medida que aumenta la complejidad de la red, también aumenta la complejidad de los métodos para realizar inferencia, por lo que los requerimientos computacionales son demasiado costosos. Sin embargo, manteniendo una estructura simple los métodos bayesianos se pueden utilizar mejor para tareas específicas y así dejar a los demás métodos hacer su trabajo. Esto hará más fácil el trabajo de compilación y prueba ya que se puede aislar el complicado código de Inteligencia Artificial del resto del código.

Este algoritmo de propagación consiste en calcular las probabilidades a posteriori de todas las variables contenidas en la red cuando se tiene cierta información o evidencia acerca del estado de alguna variable. Además, se puede agregar, que este método para realizar inferencias probabilísticas está basado en el paso de mensajes y es uno de los más rápidos entre los métodos exactos. Dicho algoritmo de propagación consta de dos fases fundamentales:

Fase de Inicialización: Se obtienen las probabilidades a priori de todos los nodos llegando a un estado inicial de la red.

Fase de Actualización: Cuando se tiene evidencia del estado de una determinada variable, se actualiza el estado de la red, obteniéndose las probabilidades a posteriori basadas en la evidencia considerada y llegando así a un nuevo estado de la red. Este proceso se repite cada vez que se instancia una variable obteniéndose los sucesivos estados de la red.

El proceso de actualización de la red funciona de la siguiente forma:

Cada vez que una variable se instancia, informa a sus nodos vecinos mediante el paso de mensajes, es decir, la variable envía a su padre un λ-mensaje y además envía a todos sus hijos un π-mensaje para informarles que su valor ha cambiado. Así la información se va propagando por la red tanto en sentido ascendente como descendente.

Estos mensajes asignan a cada variable unos valores que se llaman λ-valor y π-valor. Multiplicando estos valores y normalizando se obtendrán las probabilidades a posteriori de cada una de las variables de la red.

A continuación se presentan las 5 fórmulas necesarias para llevar a cabo el algoritmo:

1. Si B es un hijo de A, B tiene k valores posibles y A m valores posibles, entonces para $\mathrm{j} = 1, 2, \ldots, \mathrm{m}$, el λ-mensaje de B a A viene dado por:

$$\lambda_B(a_j) = \sum_{i=1}^{k} P(b_i/a_j)\,\lambda(b_i)$$

2. Si B es hijo de A y A tiene m valores posibles, entonces para $\mathrm{j} = 1, 2, \ldots, \mathrm{m}$, el π-mensaje de A a B viene dado por:

$$\pi_B(a_j) = \begin{cases} \pi(a_j) \prod\limits_{\substack{c \in s(A) \\ c \neq B}} \lambda_c(a_j) & si\ A\ no\ ha\ sido\ instanciada\ (*) \\ 1 & si\ A = a_j \\ 0 & si\ A \neq a_j \end{cases}$$

Donde $s(\mathrm{A})$ denota al conjunto de hijos de A.

(*)Esta fórmula es válida en todos los casos. Hay otra fórmula cuya aplicación resulta a veces más sencilla, pero que sólo es válida cuando todas las probabilidades P*(a_i) son no nulas, que es:

$P(a_j)\ /\ \lambda_B(a_j)$

Esta fórmula da un π-mensaje distinto (pero proporcional al de la otra fórmula) e iguales probabilidades a posteriori.

3. Si B tiene k valores posibles y $s(B)$ es el conjunto de los hijos de B, entonces para $i = 1, 2, \ldots, k$, el λ-valor de B viene dado por:

$$\lambda(b_i) = \begin{Bmatrix} \prod_{c \epsilon s(B)} \lambda_c(b_i) & si\ B\ no\ ha\ sido\ instanciada \\ 1 & si\ B = b_i \\ 0 & si\ B \neq b_i \end{Bmatrix}$$

4. Si A es padre de B, B tiene k valores posibles y A tiene m valores posibles, entonces para $i = 1, 2, \ldots, k$, el π-valor de B viene dado por:

$$\pi(b_i) = \sum_{j=1}^{m} P\left(b_i/a_j\right) \pi_B(a_j)$$

5. Si B es una variable con k posibles valores, entonces para $i = 1, 2, \ldots, k$ la probabilidad a posteriori basada en las variables instanciadas se calcula como:

$$P^*(b_i) = \alpha\lambda(b_i)\,\pi(b_i)$$

Conocidas estas fórmulas entonces se procederá a explicar por pasos y detalladamente el funcionamiento del algoritmo que dará solución a la problemática en cuestión: (13)

Fase de Inicialización.

A. Inicializar todos los λ-mensajes y λ-valores a 1.

B. Si la raíz A tiene m posibles valores, entonces para $j = 1,\ldots,m$, sea

$\pi(a_j) = P(a_j)$.

C. Para todos los hijos B de la raíz A, hacer

Enviar un nuevo π-mensaje a B usando la fórmula 2.

(En ese momento comenzará un flujo de propagación debido al procedimiento de actualización C).

Cuando una variable se instancia o una variable recibe un λ o π-mensaje, se usa uno de los siguientes procedimientos de actualización:

Fase de Actualización.

A. Si una variable B se instancia a un valor b_j, entonces:

BEGIN

A.1. Inicializar $\mathrm{P}^*(b_j) = 1$ y $\mathrm{P}^*(b_i) = 0$, para todo $i \neq j$.

A.2. Calcular $\lambda(\mathrm{B})$ usando la fórmula 3.

A.3. Enviar un nuevo λ-mensaje al padre de B usando la fórmula 1.

A.4. Enviar nuevos $\pi-$mensajes a los hijos de B usando la fórmula 2.

END

B. Si una variable B recibe un nuevo λ-mensaje de uno de sus hijos y la variable B no ha sido instanciada todavía, entonces:

BEGIN

B.1. Calcular el nuevo valor de $\lambda(\mathrm{B})$ usando la fórmula 3.

B.2. Calcular el nuevo valor de $\mathrm{P}^*(\mathrm{B})$ usando la fórmula 5.

B.3. Enviar un nuevo λ-mensaje al padre de B usando la fórmula 1.

B.4. Enviar nuevos $\pi-$mensajes a los otros hijos de B usando la fórmula 2.

END.

C. Si una variable B recibe un nuevo π-mensaje de su padre y la variable B no ha sido instanciada todavía, entonces:

BEGIN

C.1. Calcular el nuevo valor de $\pi(\mathrm{B})$ usando la fórmula 4.

C.2. Calcular el nuevo valor de $\mathrm{P}^*(\mathrm{B})$ usando la fórmula 5.

C.3. Enviar nuevos $\pi-$mensajes a los hijos de *B* usando la fórmula 2.

END.

3.2 Toma de decisiones

Para realizar la toma de decisiones haciendo uso de las Redes Bayesianas, primeramente se debe tener la estructura de la red con las respectivas tablas de probabilidades condicionales para cada nodo. Pero, con las probabilidades condicionales solamente no se pueden tomar decisiones, para ello se aplica el algoritmo de propagación descrito anteriormente.

Lo primero que se debe hacer es correr la fase de inicialización del algoritmo para así obtener un estado inicial de la red con las probabilidades de ocurrencia de cada variable a priori. Estas probabilidades a priori se utilizarán para tomar las decisiones en ausencia de evidencia, es decir, cuando no se tiene la certeza de que sucedió algún evento, las variables de la red se comportarán según sus probabilidades a priori.

Ahora, cuando se tiene evidencia del estado de alguna de las variables representadas en la red, se corre entonces la fase de actualización del algoritmo, dando como resultado un nuevo estado de la red con las probabilidades a posteriori de todas las variables condicionadas a la evidencia, quedando de esta forma actualizada toda la información de la red. Entonces, con la información actualizada de la red se está en condiciones de tomar las decisiones pertinentes, de esta forma las probabilidades a posteriori definirán el comportamiento de las variables.

A modo de resumen, para llevar a cabo la toma de decisiones, se tendrán en cuenta las probabilidades a priori o a posteriori según el estado en que se encuentre la red, es decir, el comportamiento de las variables en el entorno virtual dependerá de dichas probabilidades. Por ejemplo en el caso en que las variables posean más de un 50% de probabilidad de ocurrencia en uno de sus estados significa que la variable se va a comportar según dicho estado.

3.3 Caso de estudio

En el presente caso de estudio se tomó como base un demo de tiro del proyecto Herramientas de Desarrollo para Sistemas de Realidad Virtual. A dicho demo se le aplicó una Red Bayesiana que tiene una estructura gráfica en forma de árbol como se muestra en la siguiente figura.

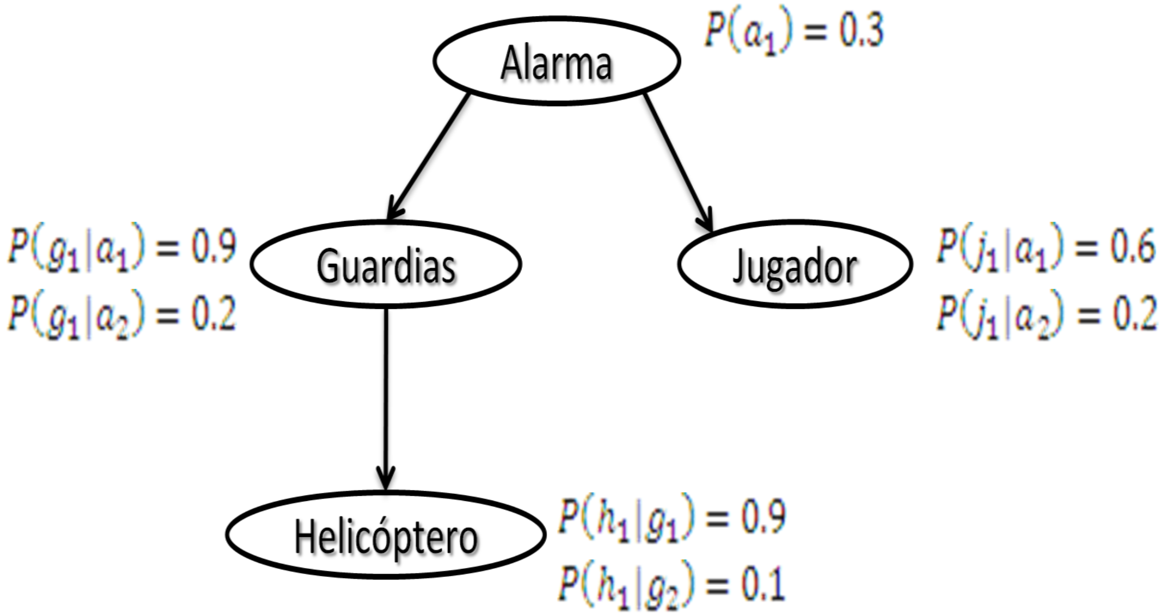


Figura 28 Red Bayesiana aplicada al DEMO de tiro

A lo largo de este epígrafe se utilizarán las siguientes variables:

A→Alarma

J→Jugador

G→Guardia

H→Helicóptero

Como se puede apreciar en la figura 28 la Alarma (A) influye en el comportamiento de los Guardias (G) y el Jugador (J); y la variable Guardias (G) influye en la disposición combativa del Helicóptero (H).

En este caso de estudio las variables utilizadas son binarias, es decir, que cada variable tendrá 2 estados. Los estados de la Alarma son a_1 que significa que la alarma suena y a_2 lo contrario, no suena. Los estados de la variable Jugador son j_1 y j_2, donde j_1 significa disparar y j_2 no disparar. De igual forma la variable Guardias tendrá 2 estados, g_1 cuando los guardias corren y g_2 cuando los guardias no corren. Por último, los estados de la variable Helicóptero son h_1 y h_2, el estado h_1 será cuando el helicóptero tiene una disposición combativa y el estado h_2 sería la acción inversa.

Como los datos que se tienen son las probabilidades a priori de la raíz y las probabilidades condicionales de las demás variables, cuando se carga el demo se hace necesario aplicar la Fase de Inicialización del algoritmo de propagación para así obtener el estado inicial de la red con sus respectivas probabilidades y poder realizar la toma de decisiones utilizando dichas probabilidades.

Para ello se enviarán π-mensajes desde los nodos padres hacia los nodos hijos comenzando por la raíz hasta llegar a las hojas. Por ejemplo, el nodo A envía π-mensajes a los nodos G y J, los cuales recalculan sus probabilidades y seguidamente el nodo G envía un π-mensajes al nodo H que también recalculará sus probabilidades. Para ver detalladamente como ocurre dicho proceso hasta obtener el estado inicial de la red, diríjase al Anexo 1.

Terminada la Fase de Inicialización, se está en condiciones de tomar la decisión de cómo se comportará cada variable utilizando las nuevas probabilidades como se muestra en la figura 29. Por ejemplo, la variable A como tiene una probabilidad de un 30% de que ocurra y un 70% de que no ocurra, entonces se obtiene como resultado que cuando se carga el demo la alarma no estará sonando. De igual forma sucede con el resto de las variables, por lo que cuando se carga el demo el jugador no estará disparando, los guardias no estarán corriendo y el helicóptero se encontrará en tierra y apagado, es decir, no estará en disposición combativa.

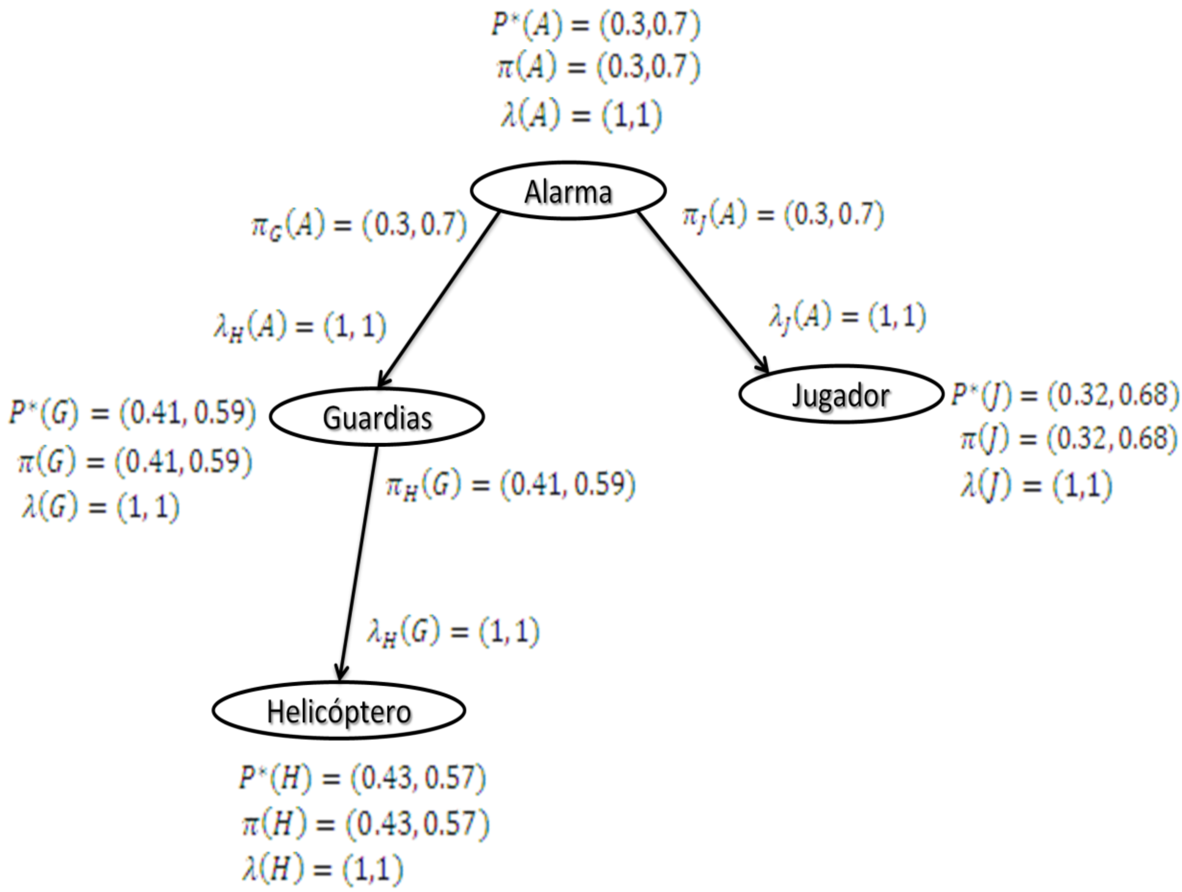


Figura 29 Red Bayesiana después de la Fase de Instanciación

Ahora, cuando el Jugador realiza un disparo comienza la Fase de Actualización de las probabilidades de la red, ya que existe una nueva evidencia, el jugador ha disparado, lo cual se representa como $J=j_1$

La actualización de las variables se lleva a cabo a través del algoritmo de paso de mensajes de la siguiente forma; el nodo J envía un λ-mensajes al nodo A, el cual recalcula sus probabilidades y envía un π-mensaje al nodo G, dicho nodo actualiza sus probabilidades y envía un π-mensaje al nodo H que también actualizará sus valores. Este proceso de actualización se describe en el Anexo 2.

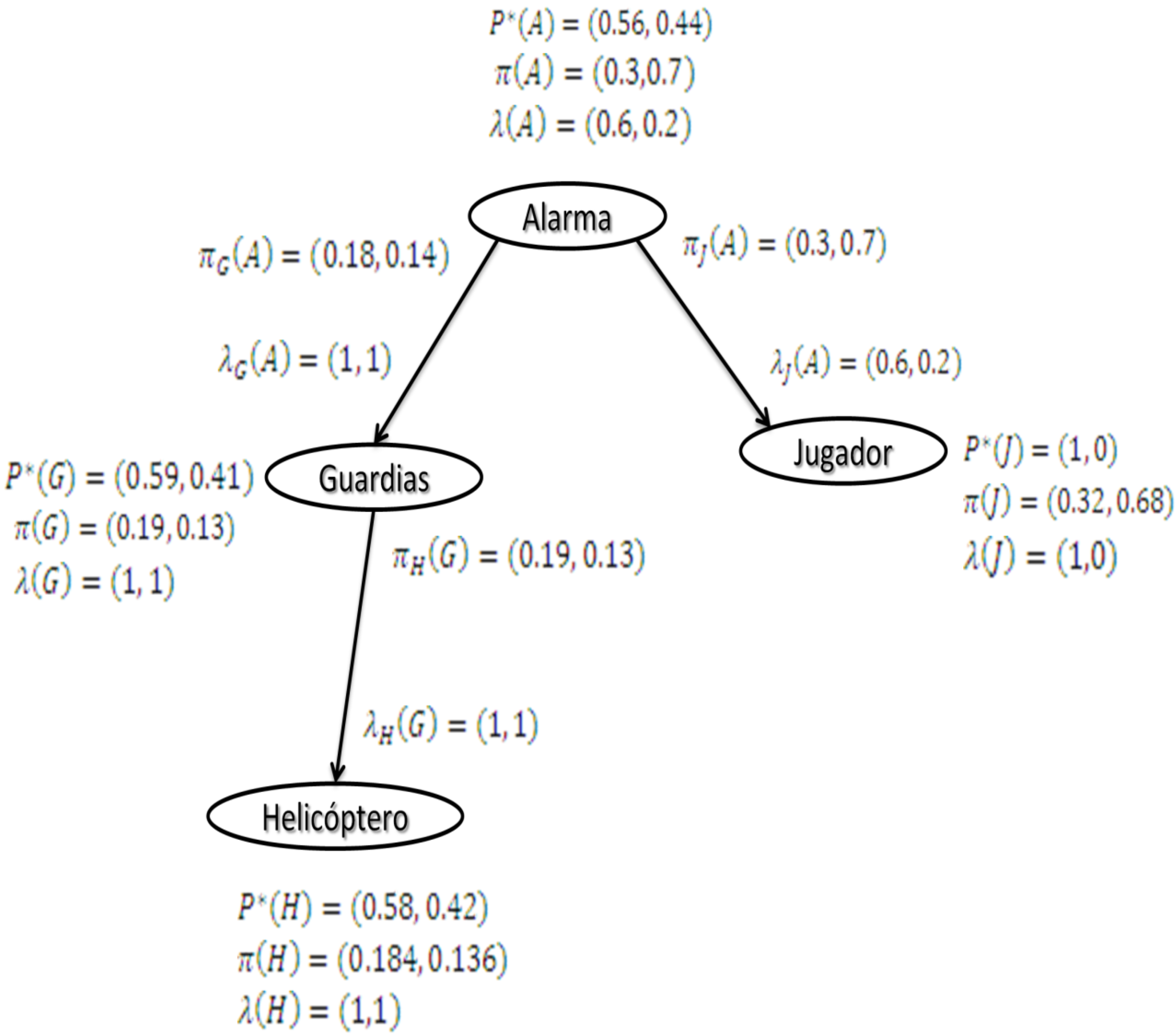


Figura 30 Red Bayesiana después de la primera Actualización

Como se muestra en la figura 30, una vez terminada la Fase de Actualización se obtendrán nuevas probabilidades para cada variable de la red. Por lo tanto, se está en condiciones de realizar la toma de decisiones. Las variables en el demo se comportarán del siguiente modo:

Alarma→comenzará a sonar

Guardias → correrán

Helicóptero→ disposición combativa (despega)

Dichas variables se mantendrán así mientras que el jugador esté disparando. Ahora, en caso de que se tenga la certeza de que el jugador dejó de disparar, comienza nuevamente el proceso de actualización de la red puesto que se tiene nueva evidencia ($J=j_2$). En este caso la red se actualiza de igual forma a través del envió de mensajes. Ver Anexo 3.

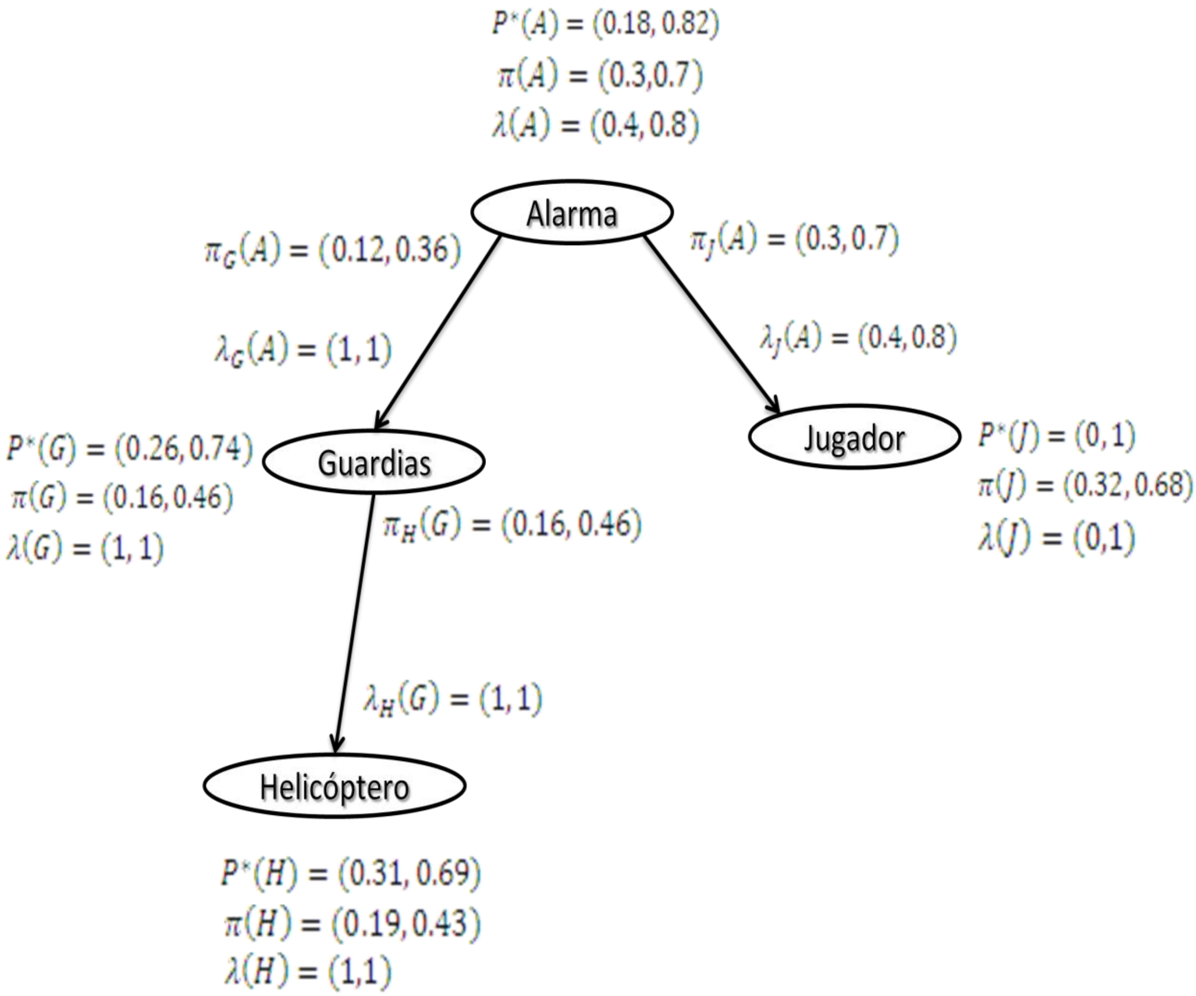


Figura 31 Red Bayesiana después de la segunda actualización

Una vez terminada la segunda actualización se obtienen nuevas probabilidades a posteriori de la red como se muestra en la figura 31, por lo que para realizar la toma de decisiones se tendrán en cuenta los nuevos valores de las variables, las cuales se comportarán de la siguiente manera.

Alarma→no sonará
Guardias → no correrán
Helicóptero→ no estará en disposición combativa (si está en el aire aterriza)

De igual forma se pueden continuar instanciando las variables según las evidencias que se obtengan aplicando el mismo algoritmo de propagación para actualizar los sucesivos estados de la red.

3.4 Resultados

Como se pudo apreciar a través del caso de estudio descrito anteriormente, queda demostrado que utilizando las Redes Bayesianas se puede llevar a cabo el proceso de toma de decisiones de los elementos en un Entorno Virtual para así dar solución a la problemática existente en los proyectos de Realidad Virtual de la Facultad 5.

CONCLUSIONES

A partir de esta investigación se obtuvo como resultado una abundante bibliografía sobre las Redes Bayesianas la cual será de mucha utilidad para los proyectos de Realidad Virtual de la Universidad de las Ciencias Informáticas.

Como resultado se expuso un algoritmo de una Red Bayesiana para una topología de árbol, la cual es muy efectiva para el trabajo en Entornos Virtuales por su rendimiento, y además se implementó dicha red demostrándose así la veracidad del mismo.

Asimismo se le dio cumplimiento al objetivo del trabajo al hacer una propuesta para realizar la toma de decisiones en Entornos Virtuales a través de una Red Bayesiana por lo que queda demostrado que se puede utilizar esta técnica de Inteligencia Artificial para incorporar inteligencia a los Sistemas de Realidad Virtual.

RECOMENDACIONES

Se recomiendan las siguientes acciones para investigaciones futuras de este tema:

Implementar otras topologías de red para ver en que medida estas se pueden utilizar en Entornos Virtuales.

Buscar otras vías para definir las probabilidades a priori y condicionales que no sea basado en el conocimiento de expertos (mediante técnicas de aprendizaje).

REFERENCIAS BIBLIOGRÁFICAS

1. **Robertson, J J O'Connor and E F.** Estadísticos Notables: Bayes, Thomas. [Online] Feb 16, 2007. http://www-history.mcs.st-andrews.ac.uk/Biographies/Bayes.html.

2. **García, Rosario Susi.** *Análisis de sensibilidad en Redes Bayesianas Gaussianas.* Madrid : s.n., 2007.

3. **Díez, F. J.** Razonamiento Bayesiano. [Online] Dic 20, 2005. http://www.ia.uned.es/~fjdiez/docencia/razbayes.

4. **Illera, Alvaro Marín.** Sistemas Expertos, Redes Bayesianas y sus aplicaciones. Deusto : s.n., 2005.

5. **Felgaer, Pablo Ezequiel.** Obtimización de Redes Bayesianas basado en técnicas de aprendizaje por inducción. Buenos Aires : s.n., 2005.

6. **Natalia Martínez Sánchez, Maikel León Espinosa, Denysde Medina Sotolongo, Zoila Zenaida Garcia Valdivia.** Red Bayesiana con aspecto de mapa conceptual: una perspectiva para los sistemas de enseñanza/aprendizaje inteligentes. Santa Clara : s.n.

7. **Maria del Carmen Chávez Cárdenas, Santiago Cuadrado Rodríguez, Gladys Casas Cardoso, Natalia Martínez Sánchez.** Red Bayesiana a partir de factores de riesgo de la Hipertensión Arterial. Santa Clara : s.n.

8. **Reina, Jose L. Ruiz.** Introducción a las Redes Bayesianas. Sevilla : s.n., 2005.

9. **Albrecht, D.W. Zukerman, I. Nicholson, E.** *Bayesian Models for Keyhole Plan Recognition in an Adventure Game.* 1998.

10. **Bourg, David M. Seeman, Glenn.** *AI for Game Developers.* s.l. : O'Reilly, 2004.

11. **Sucar, L. Enrique.** Modelos Gráficos Probabilísticos: Inferencia.

12. —. Modelos Gráficos Probabilísticos: Inferencia2.

13. **Millan, Eva.** Redes Bayesianas.

BIBLIOGRAFIA CONSULTADA

A.M. Gómez, M. Castillo. Modelo de decisión para el proceso de mercadeo de nuevos vehículos en GM Colmotores. Bogotá : s.n.

Albrecht, D.W. Zukerman, I. Nicholson, E. 1998. *Bayesian Models for Keyhole Plan Recognition in an Adventure Game.* 1998.

Bourg, David M. Seeman, Glenn. 2004. *AI for Game Developers.* s.l. : O'Reilly, 2004.

Dennis M. Buede, Joseph A. Tatman, Terry A. Bresnick. 1998. Introduction to Bayesian Networks. Monterrey : s.n., 1998.

Diaz, Jorge Luis Guevara. Inteligencia Artificial-Redes Bayesianas.

Díez, F. J. 2005. Razonamiento Bayesiano. [Online] Dic 20, 2005. http://www.ia.uned.es/~fjdiez/docencia/razbayes.

Felgaer, Pablo Ezequiel. 2005. Obtimización de Redes Bayesianas basado en técnicas de aprendizaje por inducción. Buenos Aires : s.n., 2005.

García, Rosario Susi. 2007. *Análisis de sensibilidad en Redes Bayesianas Gaussianas.* Madrid : s.n., 2007.

Gutiérrez, José Manuel. Introducción a la Inteligencia Artificial. Aplicaciones. Cantabria : s.n.

Illera, Alvaro Marín. 2005. Sistemas Expertos, Redes Bayesianas y sus aplicaciones. Deusto : s.n., 2005.

Jorge López Puga, Juan García García1. 2008. Sistemas de Tutorización Inteligente Basados en Redes Bayesianas. Almería : s.n., 2008.

Joseba Esteban López, José Javier Dolado. Estudio de los métodos de estimación: AHP y redes Bayesianas.

L, Carlos Hurtado. Redes Bayesianas. Chile : s.n.

Larrañaga, Pedro. 2002. Redes Bayesianas. 2002 .

Maria del Carmen Chávez Cárdenas, Santiago Cuadrado Rodríguez, Gladys Casas Cardoso, Natalia Martínez Sánchez. Red Bayesiana a partir de factores de riesgo de la Hipertensión Arterial. Santa Clara : s.n.

Millan, Eva. Redes Bayesianas.

Moore, Andrew W. 2001. Bayes Nets for representing and reasoning about uncertainty. 2001.

Moral, Serafin. Una introducción a las Redes Bayesianas. Granada : s.n.

Natalia Martínez Sánchez, Maikel León Espinosa, Denysde Medina Sotolongo, Zoila Zenaida Garcia Valdivia. Red Bayesiana con aspecto de mapa conceptual: una perspectiva para los sistemas de enseñanza/aprendizaje inteligentes. Santa Clara : s.n.

Nicandro Cruz Ramírez, Manuel Martínez Morales. Un Algoritmo para generar Redes Bayesianas a partir de datos.

Reina, Jose L. Ruiz. 2005. Introducción a las Redes Bayesianas. Sevilla : s.n., 2005.

Rendon, M. Valenzuela. 2006. bayes-foils. 2006.

Robertson, J J O'Connor and E F. 2007. Estadísticos Notables: Bayes, Thomas. [Online] Feb 16, 2007. http://www-history.mcs.st-andrews.ac.uk/Biographies/Bayes.html.

Scott Davies, Andrew Moore. Bayesian Networks: Independencies and Inference.

Sucar, L. Enrique. Modelos Gráficos Probabilísticos: Inferencia.

—. Modelos Gráficos Probabilísticos: Inferencia2.

—. Redes Bayesianas-Aprendizaje.

—. Redes Bayesianas-Representación.

—. Teoria de Grafos.

Vásquez, Carlos López de Castilla. 2005. Clasificadores por Redes Bayesianas. 2005.

Villanueva, David A. Velasco. 2007. Redes Bayesianas. 2007.

ANEXO 1 FASE DE INICIALIZACIÓN

1. Se inicializan todos los λ-valores y λ-mensajes a 1

$\lambda(A) = (1,1) \quad \lambda(G) = (1,1) \quad \lambda(J) = (1,1) \quad \lambda(H) = (1,1)$

$\lambda_H(G) = (1,1) \quad \lambda_G(A) = (1,1) \quad \lambda_J(A) = (1,1)$

2. Se hace $\pi\big(a_j\big) = P(a_j)$

$\pi(a_1) = P(a_1) = 0.3 \qquad \pi(a_2) = P(a_2) = 0.7$

3. Se envía π-mensajes a los hijos de A

Para $\pi_G(A)$

$\pi_G(a_1) = \pi(a_1)\lambda_J(a_1)$

$= 0.3 * 1$

$= 0.3$

$\pi_G(a_2) = \pi_G(a_2)\lambda_j(a_2)$

$= 0.7 * 1$

$= 0.7$

Para $\pi_J(A)$

$\pi_J(a_1) = \pi(a_1)\lambda_G(a_1)$

$= 0.3 * 1$

$= 0.3$

$\pi_J(a_2) = \pi(a_2)\lambda_G(a_2)$

$= 0.7 * 1$

$= 0.7$

4. Se calcula π(G)

$\pi(g_1) = P(g_1|a_1)\pi_G(a_1) + P(g_1|a_2)\pi_G(a_2)$

$= 0.9 * 0.3 + 0.2 * 0.7$

$= 0.41$

$$\pi(g_2) = P(g_2|a_1)\pi_G(a_1) + P(g_2|a_2)\pi_G(a_2)$$

$$= 0.1 * 0.3 + 0.8 * 0.7$$

$$= 0.59$$

5. Se calcula π(J)

$$\pi(j_1) = P(j_1|a_1)\pi_J(a_1) + P(j_1|a_2)\pi_J(a_2)$$

$$= 0.6 * 0.3 + 0.2 * 0.7$$

$$= 0.32$$

$$\pi(j_2) = P(j_2|a_1)\pi_J(a_1) + P(j_2|a_2)\pi_J(a_2)$$

$$= 0.4 * 0.3 + 0.8 * 0.7$$

$$= 0.68$$

6. Se calcula P*(J)

$$\alpha = [\lambda(j_1)\pi(j_1) + \lambda(j_2)\pi(j_2)]^{-1} = [1 * 0.32 + 1 * 0.68]^{-1} = 1$$

$$P^*(j_1) = \alpha\lambda(j_1)\pi(j_1)$$
$$= \alpha * 1 * 0.32$$
$$= 0.32$$

$$P^*(j_2) = \alpha\lambda(j_2)\pi(j_2)$$
$$= \alpha * 1 * 0.68$$
$$= 0.68$$

7. Se calcula P*(G)

$P^{*}(g_1) = \alpha\lambda(g_1)\pi(g_1)$

$= \alpha * 1 * 0.41$

$= 0.41$

$P^{*}(g_2) = \alpha\lambda(g_2)\pi(g_2)$

$= \alpha * 1 * 0.59$

$= 0.59$

8. Se envían π-mensajes desde G hasta H

$\pi_H(g_1) = \pi(g_1)$

$= 0.41$

$\pi_H(g_2) = \pi(g_2)$

$= 0.59$

9. Se calcula π(H)

$\pi(h_1) = P(h_1|g_1)\pi_H(g_1) + P(h_1|g_2)\pi_H(g_2)$

$= 0.9 * 0.41 + 0.1 * 0.59$

$= 0.428$

$\pi(h_2) = P(h_2|g_1)\pi_H(g_1) + P(h_2|g_2)\pi_H(g_2)$

$= 0.1 * 0.41 + 0.9 * 0.59$

$= 0.572$

10. Se calcula P*(H)

$P^{*}(h_1) = \alpha\lambda(h_1)\pi(h_1)$

$= \alpha * 1 * 0.428$

$= 0.43$

$P^{*}(h_2) = \alpha\lambda(h_2)\pi(h_2)$

$= \alpha * 1 * 0.572$

$= 0.57$

ANEXO 2 FASE DE ACTUALIZACIÓN

1. Se actualiza J.

$P^*(j_1) = 1$ $\qquad$ $P^*(j_2) = 0$

Se calcula $\lambda(J)$

$\lambda(j_1) = 1$ $\qquad$ $\lambda(j_2) = 0$

Se envía λ-mensaje

$$\lambda_J(a_1) = P(j_1|a_1)\lambda(j_1) + P(j_2|a_1)\lambda(j_2)$$
$$= 0.6 * 1 + 0.4 * 0$$
$$= 0.6$$

$$\lambda_J(a_2) = P(j_1|a_2)\lambda(j_1) + P(j_2|a_2)\lambda(j_2)$$
$$= 0.2 * 1 + 0.8 * 0$$
$$= 0.2$$

2. Se actualiza A

Se calcula $\lambda(A)$

$$\lambda(a_1) = \lambda_G(a_1)\lambda_J(a_1)$$
$$= 1 * 0.6$$
$$= 0.6$$

$$\lambda(a_2) = \lambda_G(a_2)\lambda_J(a_2)$$
$$= 1 * 0.2$$
$$= 0.2$$

Se calcula P*(A)

$$P^*(a_1) = \alpha\lambda(a_1)\pi(a_1)$$
$$= \alpha * 0.6 * 0.3 \qquad \alpha = [0.6 * 0.3 + 0.2 * 0.7]^{-1} = 3{,}125$$
$$= 0.56$$

$$P^*(a_2) = \alpha\lambda(a_2)\pi(a_2)$$
$$= \alpha * 0.2 * 0.7$$
$$= 0.44$$

Se envía π-mensaje a G

$$\pi_G(a_1) = \pi(a_1)\lambda_J(a_1)$$
$$= 0.3 * 0.6$$
$$= 0.18$$

$$\pi_G(a_2) = \pi(a_2)\lambda_J(a_2)$$
$$= 0.7 * 0.2$$
$$= 0.14$$

3. Se Actualiza G

Se calcula $\pi(G)$

$$\pi(g_1) = P(g_1|a_1)\pi_G(a_1) + P(g_1|a_2)\pi_G(a_2)$$
$$= 0.9 * 0.18 + 0.2 * 0.14$$
$$= 0.19$$

$$\pi(g_2) = P(g_2|a_1)\pi_G(a_1) + P(g_2|a_2)\pi_G(a_2)$$
$$= 0.1 * 0.18 + 0.8 * 0.14$$
$$= 0.13$$

Se calcula P*(G)

$$P^*(g_1) = \alpha\lambda(g_1)\pi(g_1)$$
$$= \alpha * 1 * 0.19$$
$$= 0.59$$

$$P^*(g_2) = \alpha\lambda(g_2)\pi(g_2)$$
$$= \alpha * 1 * 0.13$$
$$= 0.41$$

$$\alpha = 3.125$$

Se envían π-mensajes a H

$$\pi_H(g_1) = \pi(g_1) = 0.19 \qquad \pi_H(g_2) = \pi(g_2) = 0.13$$

4. Se actualiza H

Se calcula $\pi(H)$

$$\pi(h_1) = P(h_1|g_1)\pi_H(g_1) + P(h_1|g_2)\pi_H(g_2)$$
$$= 0.9 * 0.19 + 0.1 * 0.13$$
$$= 0.184$$

$$\pi(h_2) = P(h_2|g_1)\pi_H(g_1) + P(h_2|g_2)\pi_H(g_2)$$
$$= 0.1 * 0.19 + 0.9 * 0.13$$
$$= 0.136$$

Se calcula P*(H)

$$P^*(h_1) = \alpha\lambda(h_1)\pi(h_1) = \alpha * 1 * 0.184 = 0.58$$

$$P^*(h_2) = \alpha\lambda(h_2)\pi(h_2) = \alpha * 1 * 0.136 = 0.42$$

$$\alpha = 3.125$$

ANEXO 3 SEGUNDA FASE DE ACTUALIZACIÓN

1. Se actualiza J

$P^*(j_1) = 0$ $\qquad$ $P^*(j_2) = 1$

Se calcula λ-valores

$\lambda(j_1) = 0$ $\qquad$ $\lambda(j_2) = 1$

Se envía λ-mensaje

$$\lambda_J(a_1) = P(j_1|a_1)\lambda(j_1) + P(j_2|a_1)\lambda(j_2)$$
$$= 0.6 * 0 + 0.4 * 1$$
$$= 0.4$$

$$\lambda_J(a_2) = P(j_1|a_2)\lambda(j_1) + P(j_2|a_2)\lambda(j_2)$$
$$= 0.2 * 0 + 0.8 * 1$$
$$= 0.8$$

2. Se actualiza A

Se calcula λ-valores

$$\lambda(a_1) = \lambda_G(a_1)\lambda_J(a_1)$$
$$= 1 * 0.4$$
$$= 0.4$$

$$\lambda(a_2) = \lambda_G(a_2)\lambda_J(a_2)$$
$$= 1 * 0.8$$
$$= 0.8$$

Se calcula P*(A)

$$P^*(a_1) = \alpha\lambda(a_1)\pi(a_1)$$
$$= \alpha * 0.4 * 0.3$$
$$= 0.18$$

$$\alpha = 1.47$$

$$P^*(a_2) = \alpha\lambda(a_2)\pi(a_2)$$
$$= \alpha * 0.8 * 0.7$$
$$= 0.82$$

Se envía π-mensaje a G

$\pi_G(a_1) = \pi(a_1)\lambda_J(a_1)$
$= 0.3 * 0.4$
$= 0.12$

$\pi_G(a_2) = \pi(a_2)\lambda_J(a_2)$
$= 0.7 * 0.8$
$= 0.56$

3. Se actualiza G

Se calcula π-valor

$$\pi(g_1) = P(g_1|a_1)\pi_G(a_1) + P(g_1|a_2)\pi_G(a_2)$$
$$= 0.9 * 0.12 + 0.2 * 0.56$$
$$= 0.16$$

$$\pi(g_2) = P(g_2|a_1)\pi_G(a_1) + P(g_2|a_2)\pi_G(a_2)$$
$$= 0.1 * 0.12 + 0.8 * 0.56$$
$$= 0.46$$

Se calcula P*(G)

$P^*(g_1) = \alpha\lambda(g_1)\pi(g_1)$
$= \alpha * 1 * 0.16$
$= 0.26$

$P^*(g_2) = \alpha\lambda(g_2)\pi(g_2)$
$= \alpha * 1 * 0.46$
$= 0.74$

$\alpha = 1.6$

Se envían π-mensajes a H

$\pi_H(g_1) = \pi(g_1)$
$= 0.16$

$\pi_H(g_2) = \pi(g_2)$
$= 0.46$

4. Se actualiza H

Se calcula π-valores

$$\pi(h_1) = P(h_1|g_1)\pi_H(g_1) + P(h_1|g_2)\pi_H(g_2)$$
$$= 0.9 * 0.16 + 0.1 * 0.46$$
$$= 0.19$$

$$\pi(h_2) = P(h_2|g_1)\pi_H(g_1) + P(h_2|g_2)\pi_H(g_2)$$
$$= 0.1 * 0.16 + 0.9 * 0.46$$
$$= 0.43$$

Se calcula $P^*(H)$

$$P^*(h_1) = \alpha\lambda(h_1)\pi(h_1) \qquad P^*(h_2) = \alpha\lambda(h_2)\pi(h_2) \qquad \alpha = 1.61$$
$$= \alpha * 1 * 0.19 \qquad = \alpha * 1 * 0.43$$
$$= 0.31 \qquad = 0.69$$

GLOSARIO DE TÉRMINOS

Algoritmo: Un conjunto de reglas bien definidas para la solución de un problema en un número finito de pasos.

Elemento: En un Entorno Virtual un elemento es un objeto o una parte del entorno que interactúa o pertenece al mismo.

Entorno Virtual: Un Entorno Virtual puede definirse como una representación tridimensional por ordenador de un espacio en el que los usuarios pueden mover libremente su punto de vista en tiempo real. Es un sistema interactivo que permite sintetizar un mundo tridimensional ficticio, creando una ilusión de realidad.

Incertidumbre: Expresión del grado de desconocimiento de una condición futura. La incertidumbre puede derivarse de una falta de información o incluso por que exista desacuerdo sobre lo que se sabe o lo que podría saberse.

Inteligencia Artificial: Es una rama de la Informática que pretende desarrollar programas en los que las computadoras desarrollen conductas típicas de los seres inteligentes.

Sistema Experto: Es un sistema informático capaz de emular las prestaciones de un experto humano en un área concreta de conocimiento especializado.

1