GRUPO 5: REDES BAYESIANAS

Matheus Cafalchio de Oliveira

Paulo Muggler Moreira

Introdução

Os agentes lógicos tradicionais trabalham sob a premissa de que, dada uma proposição, o seu valor será sempre verdadeiro, falso ou desconhecido. Contudo, na realidade um agente inteligente dificilmente tem acesso a todas as informações sobre o ambiente em que atua com 100% de precisão. Nestes casos, o agente deve tomar decisões sem conhecer exatamente a) o estado do mundo em que se encontra e b) as conseqüências de suas ações sobre este mundo. Dizemos então que o agente deve atuar sob a incerteza.

Quando um agente lógico não consegue discernir uma sequência de ações que o levará até sua meta, ele fica incapacitado de agir, pois a forma como é construído lhe impede de tomar uma decisão nestes casos. Existem técnicas para se lidar com a incerteza dentro do planejamento condicional, porém estas técnicas em geral requerem a obtenção de informações desconhecidas pelo agente; se o mesmo é incapaz de obter estas informações, fica novamente incapacitado de atuar. Além disso, o grande número de contingências possíveis pode inviabilizar a construção e/ou operação de um agente, tornando-o excessivamente 'caro'. Uma solução é relaxar o modelo de mundo para um modelo mais restritivo, que permita que o agente derive um plano que funcione na maioria dos casos. Contudo, esta solução se mostra profundamente insatisfatória quando baseada puramente na lógica.

Os agentes de raciocínio probabilístico surgem para preencher esta lacuna deixada pelos agentes lógicos. Onde um agente lógico ficaria paralisado, incapacitado de tomar decisões, um agente probabilístico será sempre capaz de derivar um plano, pois se baseia em proposições que possuem um *grau de crença* associado a elas, ao contrário de um valor lógico totalmente determinado. Em um esquema de raciocínio probabilístico, uma decisão é tomada com base em compromissos diversos sobre os valores *prováveis* das variáveis do problema. Em outras palavras, a *decisão racional* do agente depende tanto da importância relativa de várias metas quanto da probabilidade e do grau em que elas serão alcançadas (Russel & Norvig, 2004).

Tratamento do Conhecimento Incerto

Ao aplicarmos o raciocínio puramente lógico a um problema, estamos supondo que todas as condições e causas relacionadas às proposições lógicas pertinentes ao problema são conhecidas e podem ser bem definidas utilizando as informações disponíveis. Além disso, estamos limitando o número das causas relevantes que tornam uma proposição lógica válida a uma quantidade razoavelmente fácil de lidar. Por exemplo, se tentarmos listar todas as causas que determinam se choverá em uma determinada data de forma a tornar esta regra logicamente exaustiva, chegaremos muito rapidamente a um número de causas tão grande a ponto de tornar o problema proibitivamente complexo pelo raciocínio estritamente lógico.

Existem três razões principais pelas quais não podemos utilizar a lógica de primeira ordem em um grande número de domínios de problemas práticos:

- **Preguiça**: O trabalho necessário para listar todo o conjunto de possíveis antecedentes e/ou conseqüentes de uma proposição lógica, de forma a assegurar uma regra sem exceções é demasiado grande, e gera regras proibitivamente difíceis de usar.
- **Ignorância Teórica**: Frequentemente, nós não possuímos um conhecimento teórico completo no domínio do problema com o qual estamos lidando.
- **Ignorância Prática**: Mesmo possuindo um modelo teórico completo no domínio do problema tratado e tendo construído regras logicamente exaustivas, certos dados ou testes podem ser muito custosos ou até mesmo impossíveis de serem coletados para seu uso na resolução das regras criadas.

Nos casos em que um ou mais dos fatores acima se aplicam, o conhecimento de um agente pode, na melhor das hipóteses, fornecer um **grau de crença** em cada uma das sentenças lógicas relevantes. A ferramenta matemática utilizada para lidar com este tipo de conhecimento será a teoria da probabilidade, atribuindo a cada sentença um grau de crença numérico situado no intervalo fechado entre 0 e 1.

A probabilidade fornece uma forma de resumir toda incerteza advinda de nossa *preguiça* e *ignorância.* Atribuir probabilidade 0 a uma sentença equivale à crença inequívoca de que a sentença é falsa, enquanto atribuir uma probabilidade 1 corresponde à crença inequívoca de que a sentença é verdadeira. Valores de probabilidade intermediários correspondem a graus de crença intermediários na veracidade da sentença. É importante observar que o grau de crença e o grau de verdade de uma sentença são conceitos distintos. Uma probabilidade de 0,8 sobre uma sentença não implica que a sentença é 80% verdadeira, mas sim que acreditamos com 80% de certeza que a sentença seja verdadeira. Ou seja, a sentença propriamente dita é *de* *fato* verdadeira ou falsa. Com a teoria da probabilidade, assumimos o mesmo compromisso que a lógica condicional, de que as sentenças são verdadeiras ou falsas, isto é, os fatos são ou não são válidos no mundo. A aplicação do conceito de *grau de verdade* é o assunto da *lógica difusa*.

Em lógica, uma sentença é verdadeira ou falsa dependendo de sua interpretação e do estado do mundo. Em teoria da probabilidade, uma sentença possui uma probabilidade associada a ela, portanto se refere às crenças do agente, e não diretamente ao estado do mundo. Estas crenças dependem de considerações iniciais, bem como de quaisquer percepções que o agente possa receber do mundo que venham a afetar o grau de crença em uma determinada sentença. Estas percepções constituem a **evidência** na qual se baseiam as asserções de probabilidade. Tomemos como exemplo um agente tentando adivinhar qual carta será retirada de um baralho de 52 cartas. Antes de olhar a carta retirada, o agente pode atribuir uma probabilidade de 1/52 de que a carta seja o ás de espadas. Após examinar a carta, a probabilidade atribuída à essa mesma proposição será 0 ou 1. Isto significa que as impressões recebidas pelo agente fazem com que os graus de crença nas sentenças que modelam o seu comportamento sejam atualizadas. A obtenção de conhecimento sobre o mundo tem um papel fundamental na inferência baseada em probabilidades; a atribuição de uma probabilidade a uma proposição implica em dizer o quanto a mesma é conseqüência (ou não) da base de conhecimento do agente. Assim como uma conseqüência lógica pode mudar com a introdução de novas sentenças na base de conhecimento, as probabilidades podem mudar conforme evidências são adquiridas.

Portanto, toda declaração de probabilidade deve indicar a evidência sobre a qual se baseia para a avaliação desta probabilidade. À medida que um agente atualiza suas percepções, suas avaliações de probabilidade também são atualizadas para refletir a expansão da base de conhecimento. A probabilidade atribuída a uma proposição antes da obtenção de quaisquer evidências é denominada probabilidade **a priori** ou **incondicional.** A probabilidade dada depois da obtenção de evidências é denominada probabilidade **posterior** ou **condicional.**

Decisões Racionais em Meio à Incerteza

Um agente puramente lógico toma decisões baseado no fato de que uma determinada ação pode ou não levá-lo a um estado-objetivo satisfatório para o problema em questão. Uma ação pode ser selecionada ou rejeitada se esta ação puder ou não levar o agente até a meta, independente de outras ações que também poderiam atingir a meta. Ao introduzirmos a incerteza e o conceito de grau de crença na modelagem do problema, a situação muda completamente de figura. Um problema pode ser resolvido de várias formas, e existirão vários estados finais que cumprem uma determinada função com graus de sucesso diferentes dependendo dos critérios envolvidos na seleção da solução.

Ou seja, para fazer escolhas, um agente lógico probabilístico deve ter **preferências** entre os possíveis **resultados** dos planos em questão. Um resultado específico é um estado completamente especificado, isto é, com a instanciação de todas as variáveis que compõem o problema. A formulação utilizada para lidar com preferências sobre os resultados será a **teoria da utilidade**, que atribui a cada resultado um grau de utilidade para um agente, que buscará sempre atingir o estado com a maior utilidade possível.

A utilidade de um estado para um agente será relativa às suas preferências, representadas pelas funções de utilidade. As preferências do agente são combinadas com as probabilidades, formando a **teoria da decisão**. A idéia fundamental da teoria da decisão é que *um agente é racional se e somente se ele escolhe a ação que resulta na mais alta utilidade esperada, calculada como a média sobre todos os resultados possíveis da ação*. Este princípio se chama princípio da **Utilidade Máxima Esperada (UME)**. (Russel & Norvig, 2004)

Noções Básicas de Probabilidade

Passaremos agora a descrever uma linguagem formal para lidar com o conhecimento incerto. Para ser útil, tal linguagem deverá ser capaz de:

a) Representar a natureza das sentenças às quais são associados graus de crença; e

b) Representar a dependência do grau de crença em uma determinada sentença com relação à experiência do agente.

Em nossa linguagem, a formulação tradicional da probabilidade será estendida com conceitos de lógica proposicional para formar suas sentenças. A dependência da experiência do agente é refletida pela distinção sintática entre declarações de probabilidade *a priori,* antes da obtenção de qualquer evidência e *condicionais,* incluindo explicitamente a evidência.

Os graus de crença serão sempre aplicados a **proposições** – afirmações de que tal situação está ou não ocorrendo. O elemento básico da linguagem é a **variável aleatória**, que representa uma noção a respeito do mundo cujo estado nos é inicialmente desconhecido. Toda variável aleatória possui um **domínio** de valores que ela pode assumir. Alguns domínios possíveis, por exemplo, são: *<verdadeiro, falso>*, *<quente, morno, frio>*, *[0, 100]*. As variáveis aleatórias dividem-se em três tipos de acordo com o tipo de seu domínio:

- **Booleanas:** Possuem domínio igual a *<verdadeiro, falso>*.

- **Discretas:** Admitem valores de um domínio enumerável. Os valores no domínio devem ser mutuamente exclusivos e exaustivos.

- **Contínuas:** Possuem o domínio igual à linha real inteira ou a um subconjunto da mesma, como por exemplo os intervalos [0,1], ou [0, +∞].

O estado do mundo sobre o qual um agente atua pode ser completamente especificado pela atribuição de valores a cada uma das variáveis que compõem o mundo. A isso chamamos um **evento atômico**. Os eventos atômicos têm algumas propriedades importantes:

- São *mutuamente exclusivos*: Apenas um evento atômico pode ocorrer em um determinado instante.

- O conjunto de todos os eventos atômicos possíveis é *exaustivo*. Pelo menos um evento atômico deve ocorrer em qualquer instante determinado. Isto equivale a dizer que a disjunção (OU lógico) entre todos os eventos atômicos é sempre igual a *verdadeiro*.

- Qualquer evento atômico específico impõe a verdade ou falsidade de toda proposição envolvendo as variáveis do mundo em questão. Por exemplo, o evento *atômico* $b \wedge \neg a$ impõe a verdade de b e a falsidade de $b \Rightarrow a$.

- Qualquer proposição é logicamente equivalente à disjunção de todos os eventos atômicos que impõem a verdade desta mesma proposição. Por exemplo, b é equivalente à disjunção entre $b \wedge a$ e $b \wedge \neg a$.

A **probabilidade incondicional** ou **a priori** associada a uma proposição é o grau de crença daquela proposição na ausência de quaisquer outras informações. Ela é representada por P(a). Assim que alguma informação sobre o mundo for obtida, passamos a raciocinar com a probabilidade *condicional* de a, dada esta nova informação.

Ao definirmos as probabilidades de todos os valores possíveis para uma determinada variável aleatória, estaremos definindo uma **distribuição de probabilidade** *a priori* para a variável aleatória. Por exemplo, sendo *a* uma variável aleatória com um domínio discreto de quatro valores, uma possível distribuição de probabilidade para *a* seria: P(a) = <0,7, 0,2, 0,08, 0,02>. Este vetor atribui uma probabilidade a cada valor do domínio de a, determinando a probabilidade de *a* assumir cada um destes valores.

Podemos definir também uma **distribuição de probabilidade conjunta** para denotar as probabilidades de todas as combinações de valores entre duas ou mais variáveis aleatórias. Por exemplo, P(a, b) pode ser representada em uma tabela de probabilidades de 4 x 2.

Estendendo o conceito de **distribuição de probabilidade conjunta** para todas as variáveis que descrevem o mundo, definimos a **distribuição de probabilidade conjunta total.** Uma distribuição conjunta total especifica a probabilidade de todo evento atômico, portanto especifica completamente a incerteza sobre o mundo em questão. Isto significa que qualquer consulta probabilística pode ser respondida a partir da distribuição conjunta total.

Assim que o agente obtém informações sobre o mundo em que atua, as probabilidades *a priori* não mais se aplicam. Neste caso, devemos utilizar as probabilidades **condicionais** ou **posteriores**, denotadas por P(a|b), lido como "a probabilidade de a, dado que *tudo o que sabemos* é b". Deve-se tomar muito cuidado com a interpretação dada a uma sentença de probabilidade condicional. Por exemplo, a sentença P(a|b) = 0,8 *não pode* ser interpretada com o significado "sempre que *b* for válida, conclua que *P(a) = 0,8.*" Tal interpretação seria errada porque P(a) sempre denota a probabilidade a priori de a. Além disso, a declaração P(a|b) = 0,8 é relevante apenas quando b é a única evidência disponível.

As probabilidades condicionais podem ser difnidas em termos de probabilidades incondicionais através da seguinte equação:

$$P(a|b) = P(a \wedge b) \;/\; P(b), \text{ ou}$$

$$P(a \wedge b) = P(a|b) * P(b)$$

À equação acima damos o nome de **Regra do Produto.**

Neste ponto, julgamos interessante recapitular alguns dos principais axiomas da probabilidade, que serão utilizados como base para o desenvolvimento dos conceitos apresentados adiante.

1. Uma probabilidade assume apenas valores entre 0 e 1. Para qualquer proposição *a,*
   $0 \leq P(a) \leq 1$;

2. Proposições necessariamente verdadeiras têm probabilidade igual a 1, enquanto proposições necessariamente falsas têm probabilidade igual a 0;

3. A somatória de uma distribuição de probabilidade sobre uma única variável é sempre igual a 1. Além disso, a somatória de uma distribuição de probabilidade conjunta sobre seu conjunto de variáveis também deve ser igual a 1;

4. A probabilidade de uma proposição é igual à soma das probabilidades dos eventos atômicos em que ela é válida;

5. A probabilidade de uma disjunção é dada por:
   P(a □ b) = P(a) + P(b) – P(a ^ b).

Inferência utilizando distribuições conjuntas totais

Inferência probabilística é a computação da evidência observada de probabilidades posteriores para proposições de consulta. A distribuição conjunta total pode ser utilizada como a "base de conhecimento" a partir da qual podemos derivar respostas a todas as perguntas. Vamos considerar como exemplo a distribuição conjunta total de um universo de três variáveis aleatórias booleanas, representada em uma tabela de 2 x 2 x 2:

	A		¬a	
	c	¬c	C	¬c
b	0,108	0,012	0,072	0,008
¬b	0,016	0,064	0,144	0,576

Utilizando esta tabela e com base nos axiomas da probabilidade expostos acima, podemos derivar uma séria de informações a respeito das variáveis aleatórias que compõe o universo em questão. Por exemplo, podemos calcular P(b v a) através da soma das probabilidades dos eventos atômicos em que esta proposição é válida:

P(b v a) = 0,108 + 0,012 + 0,072 + 0,008 + 0,016 + 0,064 = 0,28.

Podemos também extrair a probabilidade incondicional ou **probabilidade marginal** sobre uma variável (ou sobre um subconjunto de variáveis). A adição das entradas da primeira linha produz a probabilidade incondicional de *b:*

P(b) = 0,108 + 0,012 + 0,072 + 0,008 = 0,2.

A utilização de uma tabela para representar a distribuição conjunta total não é um método eficiente, pois não aumenta de escala muito bem. De fato, um universo descrito por *n* variáveis *booleanas* exigirá uma tabela com $O(2^n)$ entradas e o tempo de processamento da tabela será $O(2^n)$ para uma dada consulta. Visto um problema realista pode conter centenas ou até milhares de variáveis, torna-se impraticável o uso da forma tabular da distribuição conjunta total. Esta forma de representação servirá apenas como o embasamento teórico para a elaboração de abordagens mais efetivas.

Independência

Dois eventos são ditos **absolutamente independentes** se a ocorrência ou não-ocorrência de um não afeta a probabilidade do outro. A **independência absoluta** entre duas proposições *a* e *b* pode ser escrita como:

$P(a|b) = P(a)$ ou $P(b|a) = P(b)$ ou $P(a \wedge b) = P(a)*P(b)$.

As asserções de independência em geral se baseiam no conhecimento do domínio do problema. Estas asserções podem reduzir drasticamente a quantidade de informações necessárias para especificar a distribuição conjunta total. Se o conjunto completo de varáveis puder ser dividido em subconjuntos independentes, então a distribuição conjunta total poderá ser *fatorada* em distribuições conjuntas separadas sobre estes subconjuntos. Por exemplo, a distribuição conjunta sobre o resultado do lançamento de *n* moedas independentes $P(M_1, ..., M_n)$ pode ser representada como o produto de *n* distribuições de variáveis isoladas $P(M_i)$. Quando disponíveis, as asserções de independência podem ajudar na redução do tamanho da representação do domínio (por exemplo, dividindo as tabela de distribuição conjunta total em tabelas menores) e da complexidade do problema de inferência. Entretanto, esta abordagem é limitada no sentido de que é muitas vezes difícil a separação clara e completa de subconjuntos de variáveis independentes para um dado domínio, além do que mesmo estes subconjuntos podem ser excessivamente grandes para se lidar na prática com eles.

A regra de Bayes

Thomas Bayes (1702 - 17 de Abril de 1761) foi um matemático inglês e pastor presbiteriano, conhecido por ter formulado o caso especial do teorema nomeado em sua homenagem: o Teorema de Bayes, publicado postumamente. Este teorema pode ser derivado a partir da regra do produto:

$P(a \wedge b) = P(a|b)P(b)$
$P(a \wedge b) = P(b|a)P(a)$

Igualando os dois membros da direita e dividindo por P(a), obtemos:

$P(b|a) = P(a|b)P(b) / P(a)$.

Esta equação é conhecida como **regra de Bayes**, e é a base de todos os sistemas modernos de IA para inferência probabilística. A utilidade prática da regra de Bayes pode ser demonstrada pelo seguinte exemplo: Um paciente com meningite sofre de dores no pescoço em 50% dos casos; a probabilidade *a priori* do paciente ter meningite é de 1 em 50.000, e a probabilidade *a priori* do paciente ter uma rigidez no pescoço é de 1 em 20. Sabendo que um paciente sofre de rigidez no pescoço, qual é a probabilidade do mesmo ter meningite? Seja *s* a proposição de que o paciente tem uma rigidez no pescoço, e *m* a proposição de que o paciente tem mennigite:

$P(s|m) = 0{,}5$
$P(m) = 1/50.000$
$P(s) = 1/20$
$P(m|s) = P(s|m)P(m) \,/\, P(s) = 0{,}0002.$

Isto significa que ainda que o paciente sofra de dores no pescoço em 50% dos casos de meningite, esperamos que apenas um de cada 5.000 pacientes com dores no pescoço sofra de meningite. Isto ocorre porque a probabilidade *a priori* de uma rigidez no pescoço é muito maior que a probabilidade *a priori* de uma meningite. A regra de Bayes pode ser útil para responder a consultas probabilísticas condicionadas sobre uma única peça de evidência, com as informações probabilísticas frequentemente disponíveis na forma P(efeito|causa).

Para estender ainda mais a aplicabilidade da Regra de Bayes, vamos considerar dois ou mais itens de evidência possuindo alguma conexão entre si e que não podem ser considerados absolutamente independentes. Dada uma terceira evidência cujo valor é conhecido, a independência entre as evidências anteriores poderá em certos casos ser estabelecida. A isso chamamos **independência condicional**. Exemplo: Trovão é condicionalmente independente de Vento, dado Chuva. A definição geral de independência condicional entre duas variáveis aleatórias segue:

$P(a \wedge b \mid c) = P(a|c)*P(b|c).$

Assim como ocorre com as asserções de independência absoluta, as asserções de independência condicional permitem uma decomposição da distribuição conjunta total em fatores menores. De fato, para *n* efeitos condicionalmente independentes dada uma causa, o tamanho da representação cresce como $O(n)$, e não como $O(2^n)$. Desse modo, as asserções de independência condicional podem permitir o aumento de escala em sistemas probabilísticos; além disso, elas são muito mais comuns que as asserções de independência absoluta. Conceitualmente, no exemplo acima, *Chuva* separa *Vento* e *Trovão*, por ser uma causa direta de ambas. A decomposição de grandes domínios probabilísticos em subconjuntos conectados livremente por meio de independência condicional é um dos desenvolvimentos mais importantes na história recente da IA.

Quando uma única causa influencia de maneira direta vários efeitos, todos condicionalmente independentes dada a causa, podemos escrever a distribuição conjunta total como:

$P(\text{Causa}, \text{Efeito}_1, \ldots, \text{Efeito}_n) = P(\text{Causa})\prod_i P(\text{Efeito}_i|\text{Causa}).$

Esta distribuição de probabilidade é chamada modelo Bayesiano **ingênuo**, porque é usada com freqüência (como uma hipótese simplificadora) mesmo em casos nos quais as variáveis efeito não são condicionalmente independentes dada a variável causa. Na prática, sistemas de Bayes ingênuos podem funcionar muito bem mesmo quando a hipótese de independência condicional não é verdadeira.

Redes Bayesianas

Nem sempre é possível que um agente tenha todas as informações para tomar a decisão racional. A falta de informação pode ser tratada como incertezas. Nesse ambiente é necessário utilizar algo que possa representá-las. Um modo de representar um ambiente incerto é utilizando a teoria da probabilidade.

Uma rede Bayesiana é um grafo orientado onde cada nó é identificado por um valor de probabilidade. Uma definição para rede bayesiana:

1. Um conjunto de variáveis aleatórias constitui os nós da rede. As variáveis podem ser discretas ou contínuas.

2. Um conjunto de vínculos orientados ou setas conecta pares de nós. Se houver uma seta do nó X até o nó Y, X é denominado pai de Y.

3. Cada nó Xi tem uma distribuição de probabilidade condicional P((Xi)|pais(Xi)) que quantifica o efeito dos pais sobre o nó.

4. O grafo não tem nenhum ciclo orientado.

Uma rede Bayesiana é uma representação simples das relações de causalidade entre as variáveis. Para cada variável Xi que possui pais (Xi-1;Xi-2, ... , X1) é definida uma tabela de probabilidade condicional. Cada linha da tabela contém a probabilidade condicional de cada valor de nó para um caso de condicionamento. Um caso de condicionamento é uma combinação possível dos nós superiores. Cada linha deve somar 1 porque são eventos complementares.

Uma tabela de k pais boleanos tem 2^k probabilidades que podem ser de modo independente. Se um nó não tem pais, ele é especificado por apenas uma linha que representa sua probabilidade a priori.

A rede bayesiana faz uma representação completa do domínio. Uma entrada na rede pode ser representada por:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | pais(X_i)) \qquad (1)$$

Assim, cada entrada na distribuição conjunta é representada pelo produto de elementos apropriados das tabelas de probabilidade condicional.

Como construir uma rede Bayesiana ?

Escrevemos a distribuição conjunta em termos de uma probabilidade condicional usando a regra do produto.

$$P(A|B) = P(B|A)P(B)$$

$$P(X_1, ..., X_n) = P(X_n | X_n - 1, ..., X_1) P(X_n - 1, ..., X_1) \qquad (2)$$

Repetimos o processo, reduzindo cada probabilidade conjuntiva a uma probabilidade condicional e uma conjunção menor.

$$P(X_1, ..., X_n) = P(X_n | X_n - 1, ..., X_1) P(X_n - 1 | X_n - 2, ..., X_1) ... P(X_2 | X_1) P(X_1)$$

$$= \prod_{i=1}^{n} P(X_i | X_i - 1, ..., X_1) \qquad (3)$$

A rede deve ser ordenada de modo que os pais de um nó sejam inseridos antes na rede, assim a relações 1 e 3 são equivalentes. A relação 3 mostra que a rede bayesiana é uma representação correta do domínio somente se cada nó é condicionalmente independente de seus nós predecessores na ordenação dos nós.

A cada nó é atribuída uma tabela de probabilidade condicional, se o nó não tem predecessores, então a sua tabela tem apenas uma linha que representa a probabilidade do evento acontecer ou não. Cada linha de uma tabela para um nó X tem a combinação de um estado para todos os nós predecessores de X. Assim, com todas as combinações descritas na **tabela de probabilidade condicional** (TPC), a rede se torna uma representação fidedigna da probabilidade conjunta total.

Considere o exemplo onde um sistema de alarme tem uma boa eficiência na detecção de roubos e também dispara quando ocorre algum terremoto, quando o alarme dispara os vizinhos ligam para o dono da casa, ele tem dois vizinhos: João e Maria. João sempre liga quando escuta o alarme, mas algumas vezes confunde o alarme com o toque do telefone e liga avisando que o alarme soou. Maria escuta música em alto volume e às vezes não escuta o alarme tocar.

Seguindo as regras de inserção de nós, adicionaremos primeiro os nós "raízes" e depois os filhos. Neste caso os nós raízes são Terremoto e Roubo, agindo diretamente sobre Alarme, depois inserimos João Liga e Maria Liga.

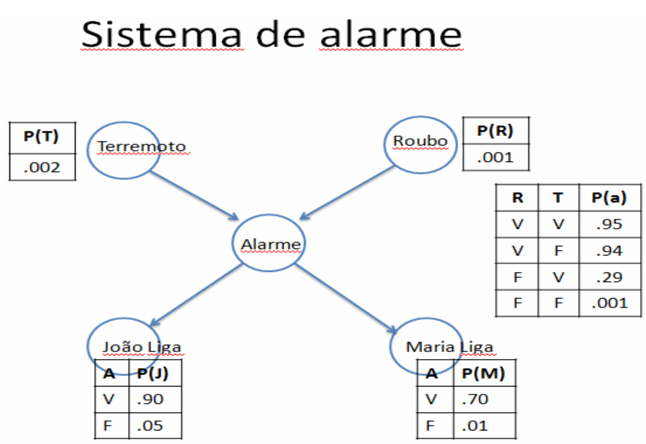

FIGURA 1

A partir da rede é possível calcular todas as probabilidades:

Exemplo:

Qual a probabilidade de João e Maria terem ligado dado que não ocorreu terremoto e não houve nenhum roubo ?

Para avaliar essa probabilidade utilizamos a regra:

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i \mid pais(X_i))$$

P(João liga, Maria liga, Alarme, ñ roubo, ñ terremoto)

= P(J | A)*P(M | A)*P(A | ñR, ñT)*P(ñT)*P(ñR)

Pelas tabelas:

= .90 * .70 * .001 * .998 * .999 = .00062

Densidade e ordenação dos nós

Além de ser uma descrição completa e não redundante do domínio, a rede pode ser mais compacta que a distribuição conjunta total, isso a torna aplicável para manipular domínios de muitas variáveis. A independência condicional faz com que a quantidade de variáveis utilizadas para caracterizar uma variável filho seja menor que a distribuição conjunta total. A densidade de redes bayesianas é um exemplo de **sistemas localmente estruturados** ou **sistemas esparsos.** Em um sistema localmente estruturado, cada subcomponente interage diretamente apenas com um número limitado de outros componentes, não importando o número total de componentes no sistema. Como a estrutura local está associada ao crescimento linear e não ao exponencial, a complexidade fica reduzida.

Representação eficiente de distribuições condicionais

Ainda que o número de pais k de uma variável seja reduzido, o preenchimento da tabela de probabilidade condicional exige até O(2^k) números e muito conhecimento em todos os casos de condicionamento. No caso onde o relacionamento entre pais e filhos é completamente arbitrário pode-se usar uma distribuição canônica que se ajusta de alguma forma-padrão. Em tais casos, a tabela completa pode ser especificada nomeandose o padrão e modificando alguns parâmetros se necessário. Um exemplo simples é a utilização de nós determinísticos. Um nó determinístico tem seu valor especificado exatamente pelos valores de seus pais, sem nenhuma incerteza.

Os relacionamentos incertos podem ser caracterizados pelos chamados relacionamentos lógicos ruidosos. A relação "OU-ruidoso" é uma generalização do OU lógico. O modelo utilizando OU-ruidoso faz duas suposições: pressupõe que todas as causas possíveis estão listadas e que a inibição de cada pai é independente da inibição de quaisquer outros pais. Com a utilização de "nós de vazamento" é possível agregar todas as causas ainda não consideradas. Em geral, relacionamentos lógicos ruidosos em que uma variável depende de k pais podem ser representadas com o uso de O(k) parâmetros, em vez de O(2^k) para a tabela de probabilidade condicional completa.

Redes Bayesianas com variáveis contínuas

Um modo de representar uma variável contínua em uma rede bayesiana é não representá-la, ou seja, discretizar a variável. Discretizar é repartir o valor contínuo em um número finito de valores, isso pode ser feito, por exemplo, dividindo uma variável em classes. Uma variável contínua que representa o ângulo de uma roleta poderia ser dividida de 10 em 10 graus e representada numa rede bayesiana. Outro modo é utilizar funções densidade de probabilidade que podem ser caracterizadas por parâmetros e no caso utilizar esses parâmetros na rede.

Uma rede que tem variáveis discretas e contínuas é denominada **rede híbrida.** Para especificar uma rede híbrida é necessário especificar distribuições condicionais discretas com pais contínuos e distribuições condicionais contínuas para pais discretos.

Conclusão

É utilizzada para representar uma probabilidade através do relacionamento entre proposições ou variáveis sempre que esta relação envolve incerteza ou imprecisão

A construção da rede depende do conhecimento do domínio a ser representado

As relações de independência diminuem o domínio do problema, possibilitando o cálculo das probabilidades e por isso as redes são empregadas em diversos problemas de IA.