Incerteza

- Bhatnagar (1986)
  - Incerteza

- Incerteza Tratada pela Matemática
  - Um evento particular dada determinadas observações

Regra de Bayes

- Probabilidades
  - Áreas de pesquisa

- Modelos Probabilísticos
- Teoria de Grafos
- Redes Bayesianas

IA – Origem e Contexto

- Uma máquina pensante e a curiosidade
  - Uniu duas áreas

- Divide-se em:
  - Sistemas Robótico
  - Sistemas Racionais

IA – Sistemas Racionais

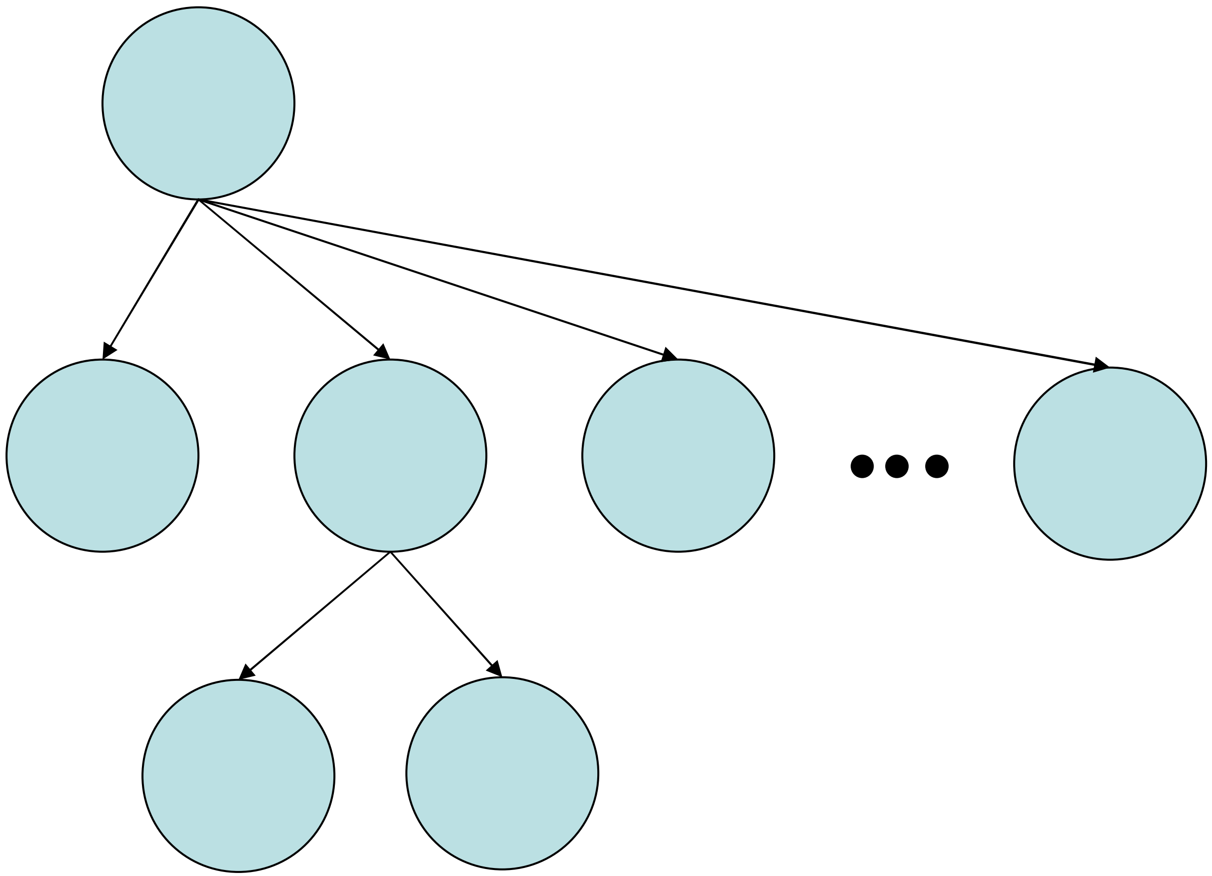
- Raciocínio Lógico
  - Problema conhecido

- Raciocínio Probabilístico
  - Problema pouco ou nada conhecido

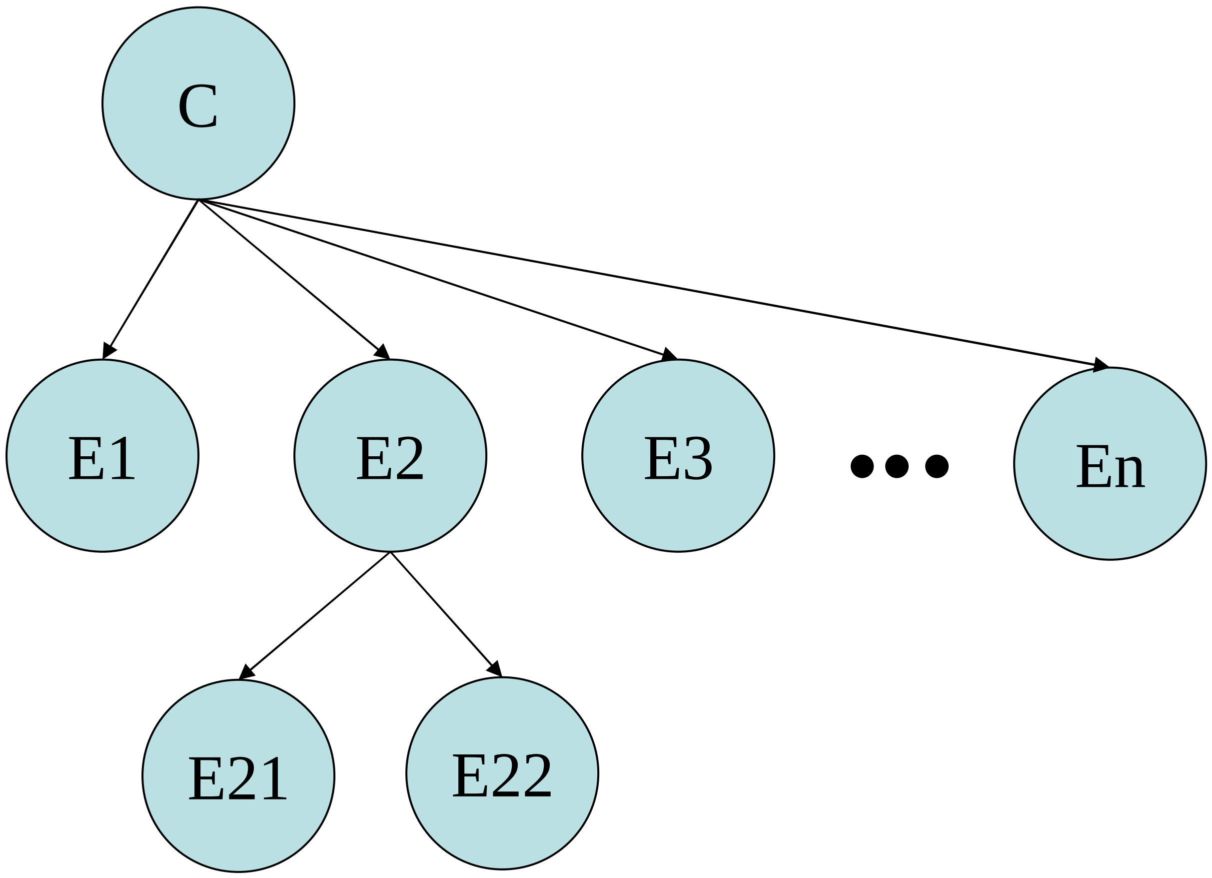
Redes Bayesianas

- Grafo orientado
- Nó = variável
- Arco = dependência probabilística

Redes Bayesianas



Redes Bayesianas



IA – Redes Bayesianas

- Informações com um CERTO nível de INCERTEZA

Raciocínio sobre Incerteza

- Impossibilidade de inserção de informações antecedentes
- Ignorância Teórica

Teorema de Bayes

•Definição:

- O modelo bayesiano interpreta a probabilidade de uma proposição como o grau de crença de um agente na veracidade dessa proposição.
- Exemplo: um dentista pode dizer: “na minha opinião, eu espero que a probabilidade de cárie seja aproximadamente 0.1”. $P(A|W)$ é o grau de crença do agente na veracidade da proposição A, dado sua experiência e conhecimento W.
- Nesse sentido, toda probabilidade é condicionada ao conhecimento do agente, diferentemente do enfoque em que se considera as probabilidades como sendo aspectos reais do universo não dependentes do conhecimento do agente

- **O teorema:**

- A inferência bayesiana repousa sobre o celebrado teorema de bayes, dado pela equação abaixo:

$$P(H \mid e) = \frac{P(e \mid H)P(H)}{P(e)}$$

Onde:

- *P(H)* é a probabilidade de ocorrer *H*;
- *P(H|e)* é a probabilidade condicional de *H*, isto é, a probabilidade de *H* após conhecer a evidência *e*;
- *P(e|H)* é a probabilidade condicional de *e,* isto é, probabilidade de *e* após conhecer a evidência *H*;

- Dado que $P(H|e)+P(\neg H|e) = 1$, tem-se que:

$$P(e|H)P(H)+P(e|\neg H)P(\neg H) = P(e).$$

- Deste modo, pode-se expressar a fórmula da inversão em termos proporcionais, sem o fator de normalização P(e), como na equação abaixo:

$$P(H|e)= \alpha\, P(e|H)P(H)$$

Aplicações do Teorema de Bayes

- O teorema de bayes é amplamente utilizado na prática, pois existem muitos casos em que fazemos boas estimativas probabilísticas. Para sua aplicação, requer-se:

- Conhecer as probabilidades *a priori* p(decisão);

- As probabilidades condicionais p(x|decisão);

- **Exemplo 1**

- Determine se um paciente tem cancro (h1)ou não (h2), sabendo que o paciente foi submetido a um exame de laboratório para detectar se tinha cancro, e o resultado do exame deu positivo. Este exame devolve um resultado positivo (+) correto em 98% dos casos em que a doença está efetivamente presente, e um resultado negativo (-) correto em 97% dos casos em que a doença não está presente. Além disso, 0.008 da população total sofre de cancro.

• Exemplo 1 – Resolução

Probabilidade de ocorrer o evento	**Evento**
$P(+ \mid h1) = 0.98$	Probabilidade do resultado do exame ser correto, dado que o paciente possui cancro.
$P(- \mid h1) = 0.02$	Probabilidade do resultado do exame ser incorreto, dado que o do paciente possui cancro.
$P(h1)=0.008$	Probabilidade de a população sofrer de cancro
$P(- \mid h2) = 0.97$	Probabilidade do resultado do exame ser correto, dado que o paciente não possui cancro.
$P(+ \mid h2) = 0.03$	Probabilidade do resultado do exame ser incorreto, dado que o do paciente não possui cancro.
$P(h2)=1-P(h1) = 0.992$	Probabilidade da população não possuir cancro.

• Exemplo 1 – Resolução

- Para um novo doente onde o teste é positivo. Qual deverá ser o diagnóstico?

  P(h1 | +) = ? P(h2 | +) = ?

Como:

$$P(h1 \mid +) = \frac{P(+ \mid h1)\ .\ P(h1)}{P(+)} = \frac{X}{K} \rightarrow P(h1|+) = P(+|h1)\ P(h1)$$

$$P(h2 \mid +) = \frac{P(+ \mid h2)\ .\ P(h2)}{P(+)} = \frac{Y}{K} \rightarrow P(h2|+) = P(+|h2)\ P(h2)$$

Então, a hipótese resume-se a:

P(h1 | +) = (+ | h1) . P(h1) → (0,98) X (0,008) = 0,00784

P(h2 | +) = P(+ | h2).P(h2) → (0,03) X (0,992) = 0,029

Logo, o doente não sofre de cancro, pois Hmáx = h2

- **Exemplo 2**

- Pacientes com problemas cardíacos são sujeitos a um eletrocardiograma (ECG). Os resultados são classificados:

  - positivos (+ECG) sugerindo doença cardíaca (+DC)

  - negativos (-ECG) no caso de não haver doença cardíaca (-DC)

Assumindo que um dado paciente realizou um eletrocardiograma positivo pretende-se saber qual a probabilidade deste paciente ter doença cardíaca ? → P(+DC | + ECD)

• Exemplo 2 – Continuação

- Sabendo que:
  - 10 pessoas em 100 têm um ataque cardíaco
  - 90 pessoas em 100 que tiveram doença cardíaca produziram um eletrocardiograma positivo
  - 95 pessoas em 100 que não tiveram doença cardíaca produziram um eletrocardiograma negativo

• Exemplo 2 - Resolução

- 10 pessoas em 100 têm um ataque cardíaco

$$\rightarrow P(+DC) = 0.1$$

$$\rightarrow P(-DC) = 1- P(+DC) = 1 - 0.1 = 0.9$$

- 90 pessoas em 100 que tiveram doença cardíaca produziram um eletrocardiograma positivo (+ECD)

$$\rightarrow P(+ECD \mid +DC) = 0.9$$

- 95 pessoas em 100 que não tiveram doença cardíaca produziram um eletrocardiograma negativo (-ECD)

$$\rightarrow P(-ECD \mid -DC) = 0.95$$

$$\rightarrow P(+ECD \mid -DC) = 1 - P(-ECD \mid -DC) = 1 - 0.95 = 0.05$$

•Exemplo 2 - Resolução

- $P(+ECD) = P(+ECD|+DC)*P(+DC)+P(+ECD|-DC)*P(-DC)$

  $= 0.9 * 0.1 + 0.05 * 0.9 = 0{,}135$

- $P(+DC \mid +ECD) = (P(+ECD \mid +DC) * P(+DC)) / P(+ECD)$

- $P(+DC \mid +ECD) = 0.1 * 0.9 / 0{,}135 = 0.67 \rightarrow 67\%$

Logo a probabilidade do paciente ter doença cardíaca, sabendo que seu exame foi positivo é de 67%.

- **Exemplo 3**

- Um médico sabe que a meningite faz o paciente ter uma rigidez no pescoço, durante 50% do tempo. O médico também conhece alguns fatos incondicionais: a probabilidade a *priori* de um paciente ter meningite é 1/50.000, e a probabilidade *a priori* de qualquer paciente ter uma rigidez no pescoço é 1/20. Sendo *s* a proposição de que o paciente tem uma rigidez no pescoço e *m* a proposição de que o paciente tem meningite, temos:

• Exemplo 3 - Resolução

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

0.02% = 1 a cada 5.000

• Exemplo 3 - Resolução

- Ou seja, esperamos que apenas 1 em 5.000 pacientes com rigidez no pescoço tenha meningite. Embora a rigidez no pescoço seja uma indicação bastante forte de meningite (com probabilidade 0,5), a probabilidade de o paciente estar acometido de meningite permanece pequena. Isso ocorre porque a probabilidade *a priori* sobre rigidez no pescoço é muito mais alta que a probabilidade *a priori* sobra a meningite.

Redes de Bayes

Redes de Bayes

Também conhecida como:

rede de crença bayesiana

ou diagrama de influência probabilística

Características

- Expressam variáveis do problema e suas dependências;

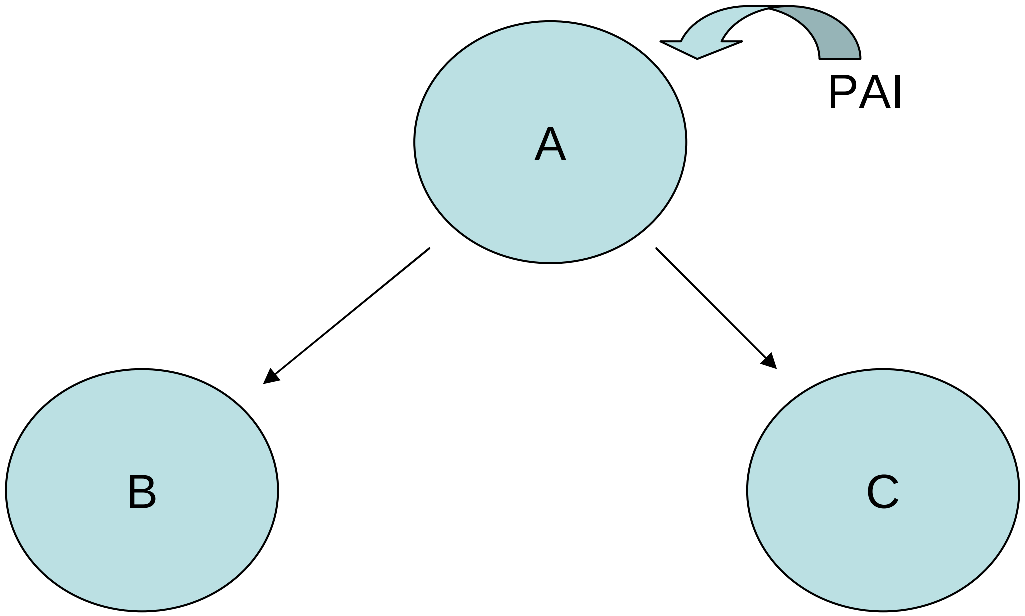
- As dependências entre os itens (variáveis) serão expressas por probabilidades condicionais (do tipo $P(A|B)$).

- Do ponto de vista de um especialista, redes bayesianas constituem um modelo gráfico que representa de forma simples as relações das variáveis de um sistema.

Consistem em:

- Um conjunto de variáveis e um conjunto de arcos ligando as variáveis.
- Cada variável possui um conjunto limitado de estados mutuamente exclusivos.
- As variáveis e arcos formam um grafo dirigido sem ciclos.
- Para cada variável A que possui como pais B1, ..., Bn, existe uma tabela P(A| B1, ..., Bn).

Ilustração



B e C são Variáveis

P(B|A)

- Probabilidade do evento B ocorrer, dado que se sabe que um evento A ocorreu.

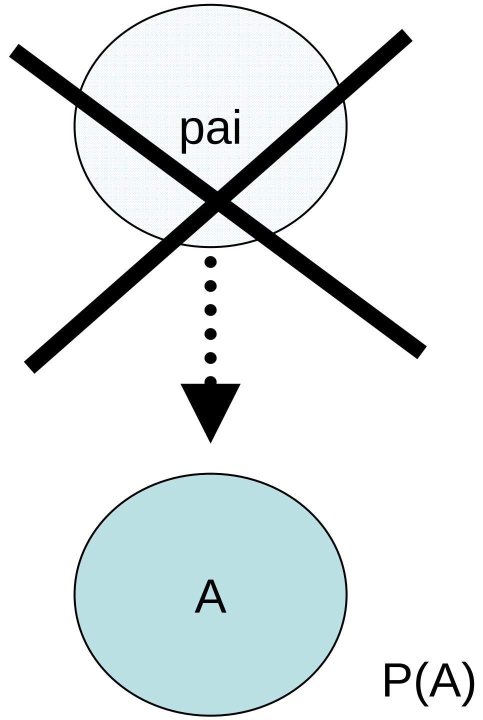
P(C|A)

- Probabilidade do evento C ocorrer, dado que se sabe que um evento A ocorreu.

Demonstra as probabilidades condicionais

Ilustração

- Caso A não possua um pai, a tabela de probabilidades é reduzida para uma probabilidade incondicional P(A).



Componentes de uma rede bayesiana

Possuem 2 componentes principais:

- uma estrutura, S, que define relacionamento entre os nós;
- parâmetros numéricos, Ɵ, que quantificam a relação probabilística causal entre os nós da estrutura.

Construíndo Redes Bayesianas

- Escolher o conjunto das variáveis relevantes que descrevem o domínio do problema;

- Escolher uma ordem para as variáveis (os pais devem surgir antes dos nós);

- Para cada variável (chamada de Xi):
  - Escolher o conjunto de nós pais para a variável Xi, de modo a que este conjunto seja mínimo e que :

    P(Xi | Pais(Xi))

  - Definir a tabela de probabilidades condicionada para Xi.

Vantagens de uma Rede Bayesiana

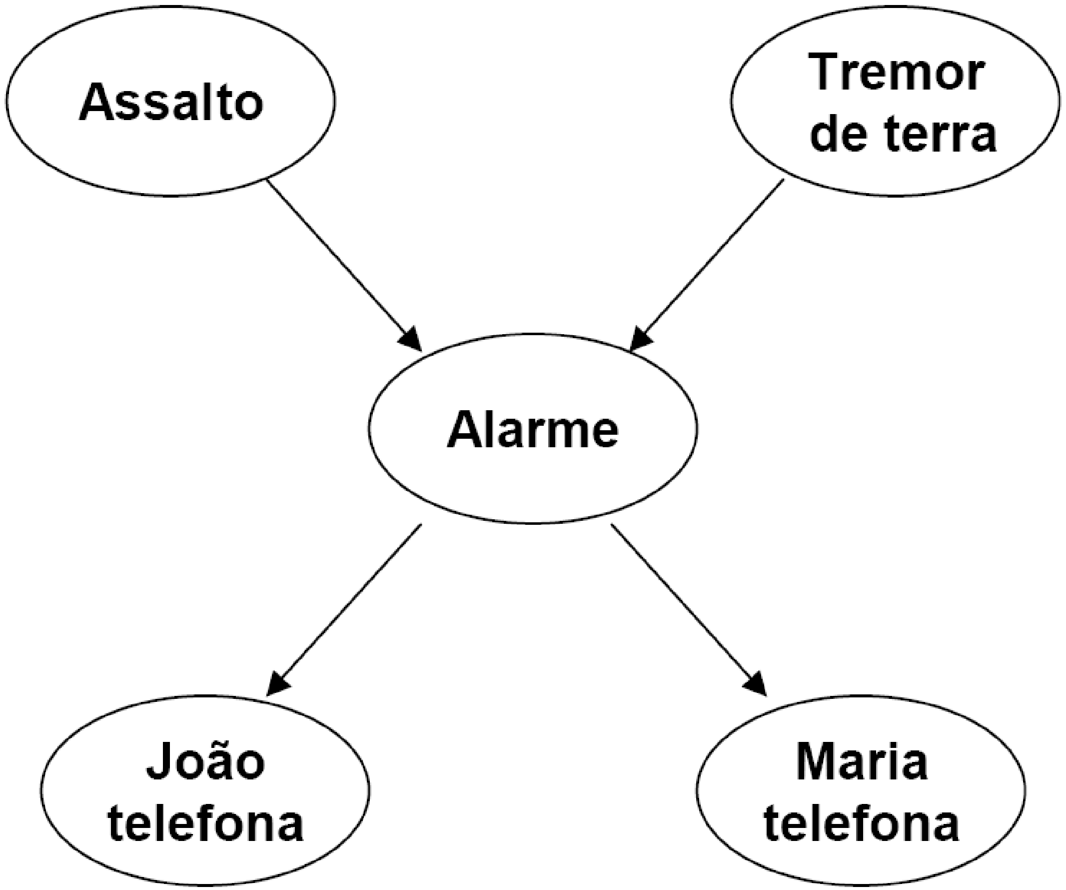
- Possuem grande campo de aplicações;
- As redes bayesianas oferecem uma estrutura unificada e intuitiva, onde é possível comparar diferentes hipóteses;
- Gráficos tornam fácil a compreensão do modelo, mesmo para usuários pouco familiarizados com a teoria Bayesiana;

Há várias outras vantagens, porém dependem do domínio do problema.

Exemplo 1 – Alarme

Você possui um novo alarme contra ladrões em casa. Este alarme é muito confiável na detecção de ladrões, entretanto, ele também pode disparar caso ocorra um terremoto. Você tem dois vizinhos, João e Maria, os quais prometeram telefonar-lhe no trabalho caso o alarme dispare. João sempre liga quando ouve o alarme, entretanto, algumas vezes confunde o alarme com o telefone e também liga nestes casos. Maria, por outro lado, gosta de ouvir música alta e às vezes não escuta o alarme. Este domínio pode ser representado de acordo com a figura seguinte.

Exemplo 1 – Alarme



Para que o alarme dispare, é necessário que haja um assalto ou tremor de terra...

Para que o João ou Maria telefone, é preciso que o alarma tenha disparado...

Exemplo 1 – Alarme

Lembre-se:

- Maria gosta de música alta; nem sempre pode ouvir o alarme!!
- João confunde o alarme com o telefone!!

Logo, a probabilidade que avisem ao escutar o alarme:

P(Alarme,João) e P(Alarme, Maria)

Não são 1 !!

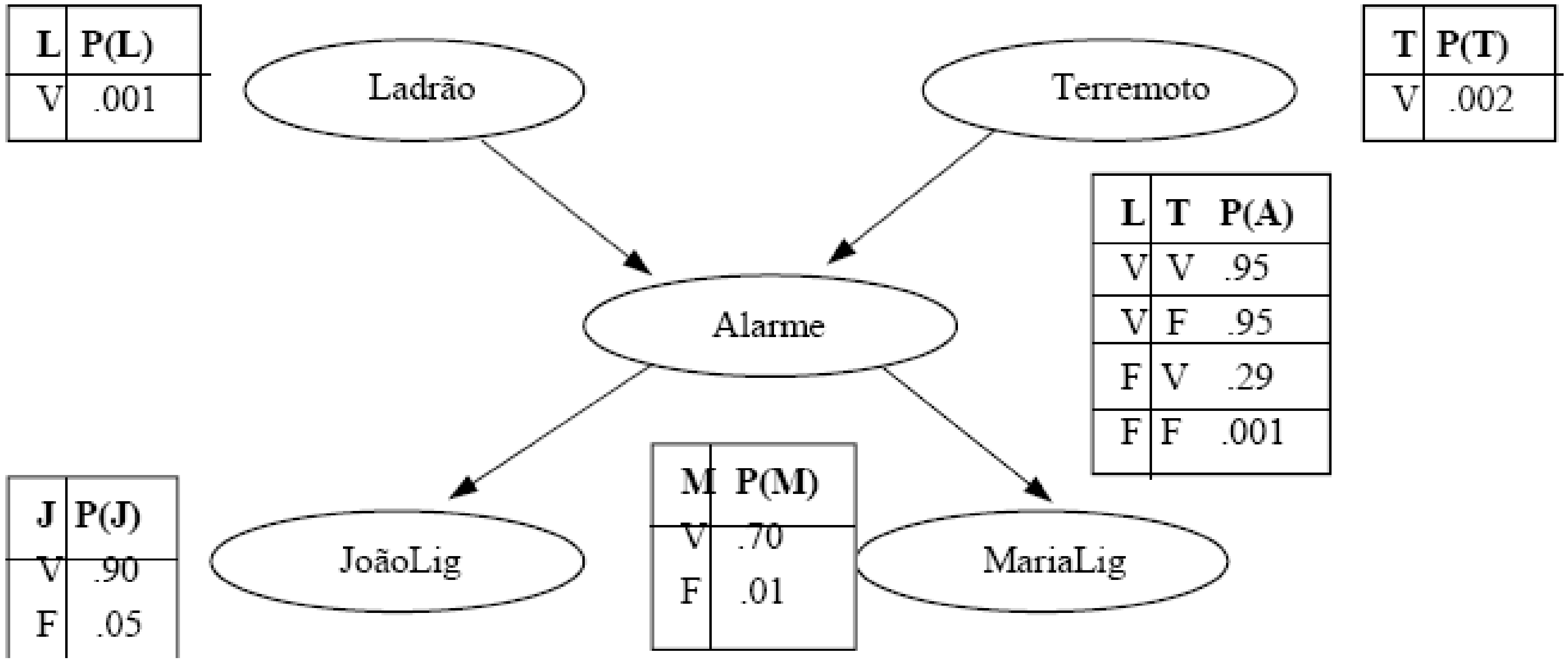
Exemplo 1 – Alarme

- Uma vez definido a topologia, é necessário se definir a tabela de probabilidades condicionais;
- Cada linha contém a probabilidade condicional para cada caso condicional dos nós pais;
- Um caso condicional é uma possível combinação dos valores para os nós pais.

Exemplo 1 – Alarme

Ladrão	Terremoto	$P(Alarme \mid Ladrão, Terremoto)$	
		Verdadeiro	Falso
Verdadeiro	Verdadeiro	0.95	0.050
Verdadeiro	Falso	0.95	0.050
Falso	Verdadeiro	0.29	0.71
Falso	Falso	0.001	0.999

Exemplo 1 – Alarme



Exemplo 1 – Alarme

- Diante disso tudo, pergunta-se: Qual a probabilidade do alarme ter tocado, mas, nem um ladrão nem um terremoto aconteceram, e ambos, João em Maria ligaram??

$$P(J \wedge M \wedge A \wedge \neg L \wedge \neg T).$$

Exemplo 1 – Alarme

$P(J \wedge M \wedge A \wedge \neg L \wedge \neg T) =$

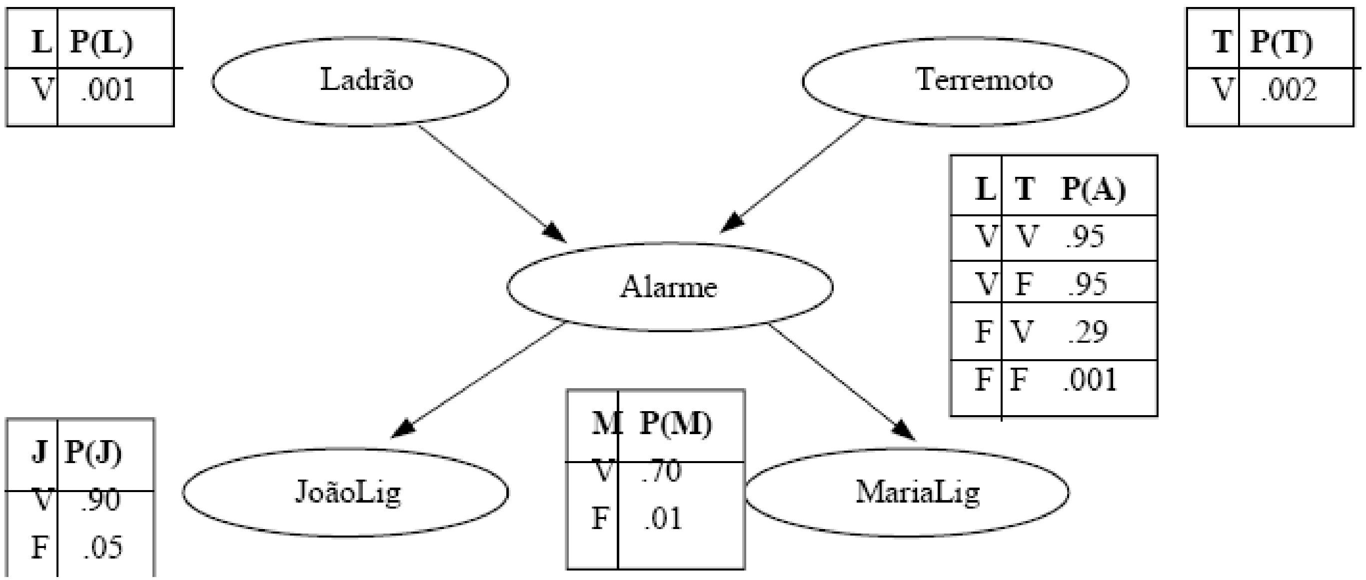
$P(J|A) \times P(M|A) \times P(A|\neg L \wedge \neg T) \times P(\neg L) \times P(\neg T) =$

$P(\neg L) = 1 - P(L)$

$P(\neg T) = 1 - P(T)$

$0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 =$

0.00062



Exemplo 2 – Sistema Especialista

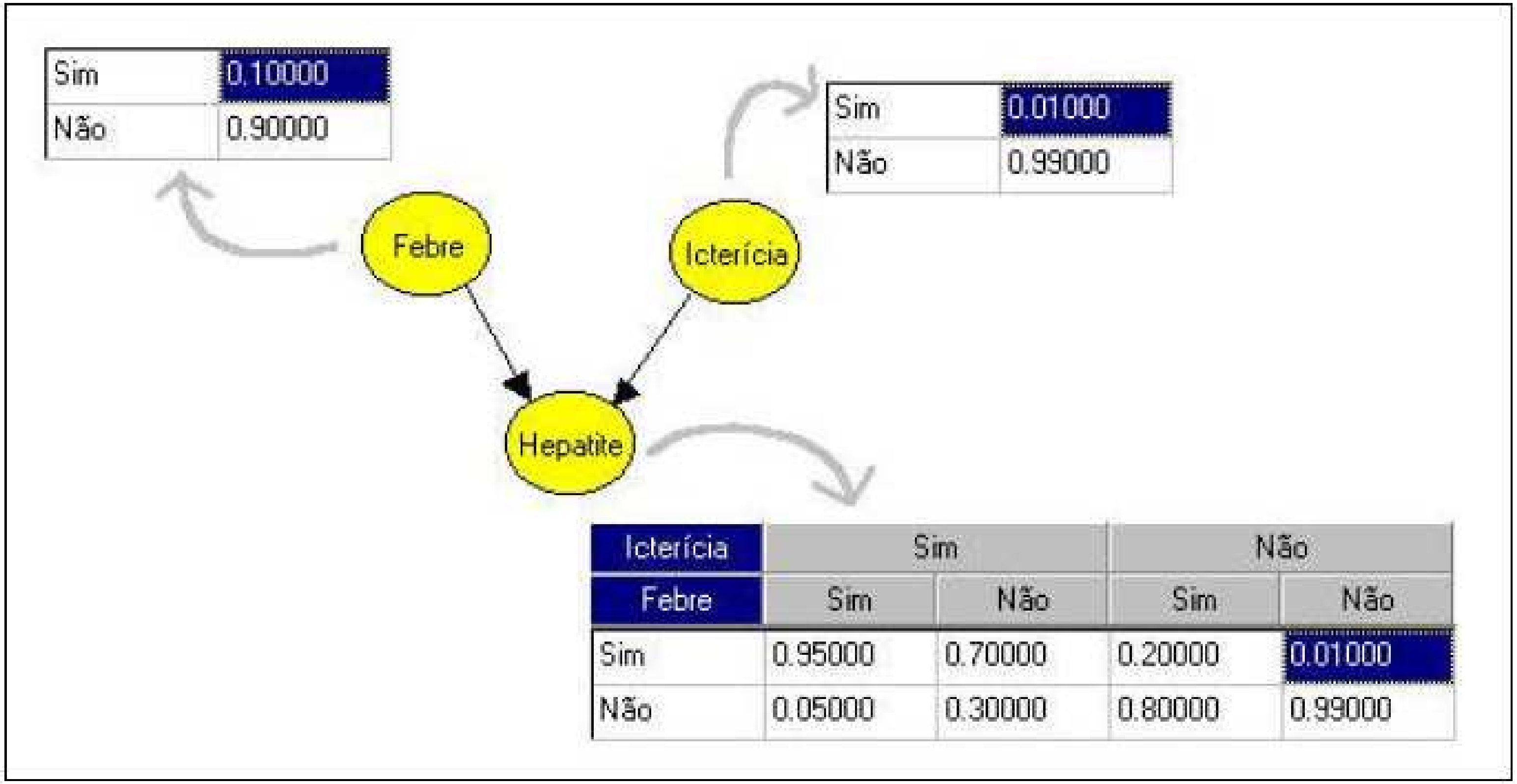
Nesse exemplo, podemos observar um domínio simples modelado a partir de uma Rede Bayesiana. Trata-se de um Diagnóstico da Hepatite. Possui três variáveis: *Febre*, *Icterícia* e *Hepatite*.Observe-se que as variáveis *Febre* e *Icterícia* estão associadas a tabelas de estados que indicam suas probabilidades a priori. As probabilidades a priori são o elemento responsável pelo tratamento da indeterminação. Quando o usuário não disser, ou não souber dizer o valor de uma evidência, a Rede Bayesiana usará, na inferência, essa probabilidade a priori, que foi definida pelo especialista como uma espécie de “valor padrão”.

Exemplo 2 – Sistema Especialista

- Variáveis: *Febre*, *Icterícia* e *Hepatite;*
- *febre e icterícia tem influência definida sobre* hepatite;

O valor das variáveis vão ser definidos conforme o valor das evidências.

Exemplo 2 – Sistema Especialista



Sim	0.10000
Não	0.90000

Sim	0.01000
Não	0.99000

Icterícia	Sim		Não	
Febre	Sim	Não	Sim	Não
Sim	0.95000	0.70000	0.20000	0.01000
Não	0.05000	0.30000	0.80000	0.99000

O especialista, durante a modelagem da rede, indicará, para cada combinação de respostas das evidências relacionadas a esse diagnóstico, um valor de resultado. A inferência que ocorre na rede durante uma consulta se dá, então, por algoritmos multiplicativos, que vão propagando as probabilidades e definindo os valores finais dos diagnósticos.

Exemplo 3 – Anti-SPAM

O que é SPAM?

- “Email comercial não solicitado”
- “**Não é** email comercial não solicitado”
- “Email não solicitado enviado automaticamente”

Exemplo 3 – Anti-SPAM

- **O que há de especial no problema?**

- SPAM incomoda as pessoas;

- pode ser facilmente identificado por seres humanos, mas computadores têm dificuldade para tratar o problema;

- em geral métodos tradicionais conseguem filtrar spams, mas apresentam um alto índice de erros;

- os SPAMERS se adaptam ao longo do tempo para burlar os filtros;

Exemplo 3 – Anti-SPAM

```
<br>
Ge<!--##email##-->ne<!--sMSe-->ric Vi<!
--sMSe-->ag<!--sMSe-->ra is now av<!--s
MSe-->ail<!--sMSe-->able to co<!--sMSe-
->nsu<!--sMSe-->me<!--sMSe-->rs<br> As
l<!--sMSe-->ow a<!--sMSe-->s $2.<!--sMS
e-->25 per do<!--sMSe-->se (5<!--sMSe--
><!--sMSe-->0 mg)<br> No Do<!--sMSe-->c
t<!--sMSe-->or’s Co<!--sMSe-->nsu<!--sM
Se-->lta<!--sMSe-->tion required<br>"S<
!--sMSe-->il<!--sMSe-->agra is as go<!-
-sMSe-->od as V<!--sMSe-->i<!--sMSe-->a
g<!--sMSe-->ra - ju<!--sMSe-->st ch<!--
sMSe-->ea<!--sMSe-->per!"<br>
<br>
```

Exemplo de SPAM html “adaptado”

Exemplo 3 – Anti-SPAM

- **tradução**

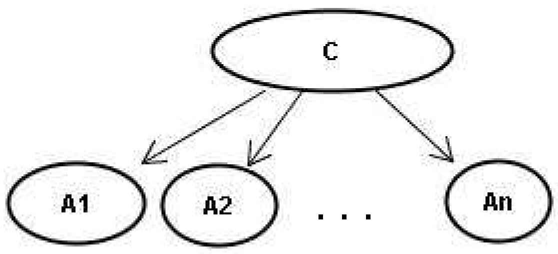
Generic Viagra is now
available to
consumers
As low as $2.5 per dose (50 mg)

No Doctor's
Consultation required

"Silagra is as good as
Viagra - just
cheaper!"

Exemplo 3 – Anti-SPAM

A detecção de SPAM, portanto passa a ser um caso típico de detecção de padrões, um problema que pode ser atacado por técnicas de Inteligência Artificial, utilizando-se de Redes Bayesianas.



Pode-se observar que o nodo “C" é o nodo que serve como variável de classificação. As folhas da árvores são independentes entre si, e representam as variáveis representativas e importantes para o mecanismo de classificação.

Exemplo 3 – Anti-SPAM

Nas folhas, foram inseridos nodos de palavras “boas”' e “ruins” com probabilidades associadas. Quando uma mensagem é submetida ao mecanismo, as palavras (ou *tokens*) são registrados como evidências, e para isto as variáveis (representadas nas folhas) são marcadas como “possui” e “npossui” conforme a mensagem. A entrada destas evidências fazem com que ocorra uma propagação das probabilidades, que influência a distribuição de probabilidades no nodo ``pai'' responsável final pela classificação.

Exemplo 3 – Anti-SPAM

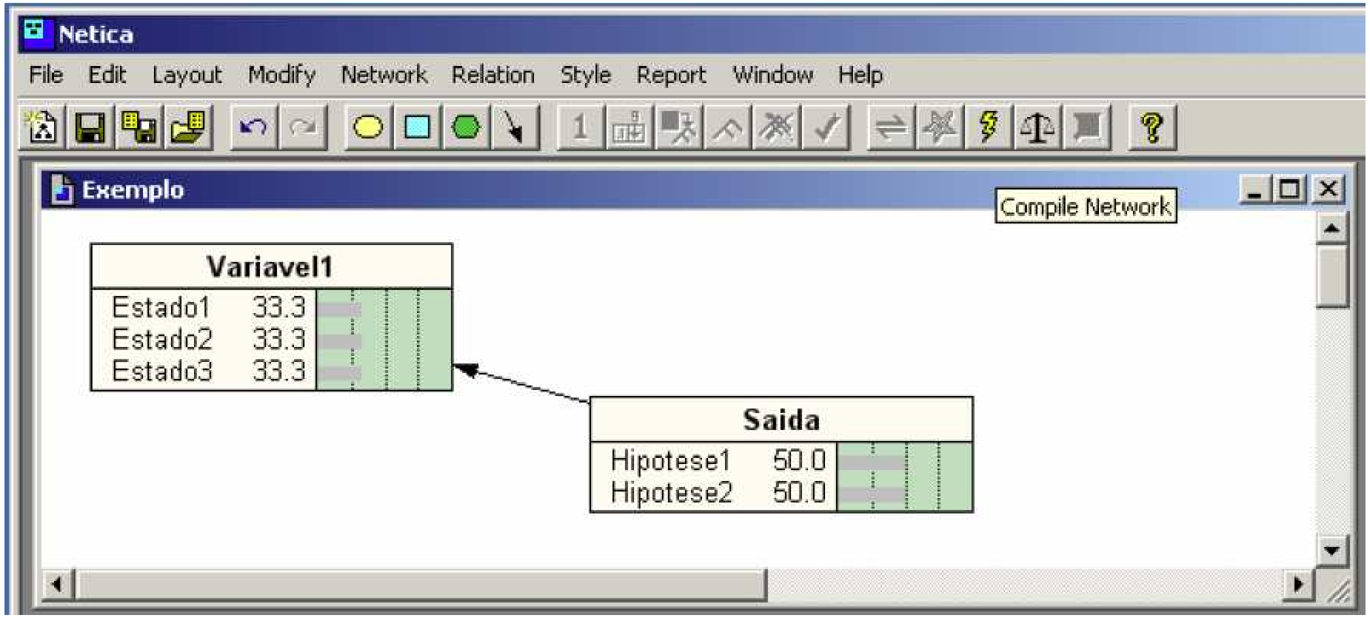
Diante desse modelo, foi implementado um sistema chamado BogoFilter, um software livre que nos mostrou características bastante importantes e desejáveis para identificar os padrões de SPAM, apresentando baixos índices de erros e altos índices de acerto em mensagens classificadas como SPAM.

Outras aplicações

- Extração de Conhecimento de Bases de Dados ;
- Aplicações Indústriais;
- Apoio ao Diagnóstico Médico;
- Previsão de vendas;
- Dentre outros....

Ferramentas

Software NETICA



Ferramentas

Hugin Expert