Probabilistic Knowledge Representation

Raciocínio Probabilístico

- Exemplo simples: diagnósticos.
- Para odontologia:

$$\forall p \quad Sintoma(p, DorDeDente) \Rightarrow Enfermidade(p, Carie)$$

  $\rightarrow$ regra correta?
- Alternativa:
  $\forall p \quad Sintoma(p, DorDeDente) \Rightarrow Enfermidade(p, Carie) \vee Enfermidade(p, Gengivite) \vee Enfermidade(p, SisoIncluso) \ldots$
- Disjunção pode ser ilimitada.
- Outra alternativa, regra causal:

$$\forall p Enfermidade(p, Carie) \Rightarrow (p, DorDeDente)$$

  $\rightarrow$ regra correta?
- Conclusão: cáries e dores de dentes não estão necessariamente relacionados.

Lidando com Incerteza

- Como representar alguma relação? Graus de incerteza ou de credibilidade (*degrees of belief* ou *degrees of truth*).
- Lógica de primeira ordem não consegue representar incerteza.
- 3 razões:
  - **Laziness**: muito trabalhoso codificar listas completas de antecedentes ou consequentes necessários para representar todas as regras, e muito ineficiente utilizar o conjunto enorme de regras.
  - **Ignorância teórica**: por exemplo, a ciência médica não tem uma teoria completa sobre o domínio.
  - **Ignorância prática**: mesmo sabendo todas as regras, pode aparecer um paciente com dados novos q não estavam previstos.

Lidando com Incerteza

- Para lidar com incerteza: teoria das probabilidades: degree of belief, lógica fuzzy: degree of truth.
- Probabilidade fornece uma forma de resumir a incerteza proveniente de 'laziness' e ignorância.
- Por ex., *80% de pacientes já tratados que têm dor de dente, têm cárie.*
- 80% compreende todos os casos em que reunimos todas as condições necessárias para concluir que se um paciente tem dor de dente, também tem cárie, e todos os casos em que o paciente tem dor de dente e cárie, mas os dois não estão relacionados.

Lidando com Incerteza

- 20% compreende todas as possíveis causas de dor de dente não relacionadas por laziness ou ignorância.
- Prob = 0, sentença é categoricamente falsa.
- Prob = 1, sentença é categoricamente verdadeira.
- **evidência**: utilizada pelo agente para atribuir novas probabilidades às proposições a partir de percepções do mundo.
- **probabilidade incondicional**: inicial (prior).
- **probabilidade condicional**: posterior, após evidências serem obtidas.

Incerteza e Decisões Racionais

- Saída para o aeroporto.
- A_{90} ou A_{120} ou A_{1440}? O q é mais racional?
- Depende dos objetivos e do tempo q uma pessoa está preparada ou disposta a esperar pelo vôo no aeroporto.
- **Preferências** diferentes para as saídas do plano.
- Pe, Uma saída de um plano pode conter fatores tais como "agente vai chegar a tempo", e "tempo de espera no aeroporto".
- *Utility theory*: para atribuir graus de "utilidade" de um determinado estado.

Agente baseado em Teoria da Decisão

- **Teoria da Decisão** = teoria das probabilidades + Teoria da Utilidade.
- Idéia fundamental: agente é racional sss escolher uma ação que produza o caminho mais útil, em relação a todos os possíveis resultados da ação (princípio MEU, Maximum Expected Utility).
- Agente baseado em teoria da decisão: estrutura similar a de um agente lógico.
- Linguagem para representação: lógica proposicional com probabilidades. Diferença sintática entre probabilidade incondicional e condicional (como em probabilidade).

Revisão Probabilidade

- Probabilidade incondicional: P(A), probabilidade incondicional de que proposição A é verdadeira.
- Pe, P(Cárie) = 0.1, 10% de chance de que o paciente tem cárie, **na ausência de qq outra informação**.
- Variáveis aleatórias (ref. prob.):
  - P(Tempo=Ensolarado) = 0.7
  - P(Tempo=Nublado) = 0.08
- Domínio: possíveis valores $\langle x_1, x_2, \ldots, x_n \rangle$ de uma variável aleatória (conjunto discreto para os exemplos mostrados).
- Convenção: A, B... para variáveis booleanas. X, Y,... para variáveis com vários valores.
- **P**(Tempo) = (0.7,0.2,0.08,0.02): distribuição de probabilidade para a variável Tempo.

Revisão Probabilidade

- P(A|B): probabilidade condicional.
- Pe, P(Carie|DorDeDente) = 0.8: chance de 80% do paciente ter cárie, dado que **tudo que sabemos até o momento** é que ele tem dor de dente.
- $P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$, com $P(B) > 0$.
- Regra do produto: $P(A \wedge B) = P(A \mid B) P(B)$
- $\mathbf{P}(X,Y) = \mathbf{P}(X|Y)\mathbf{P}(Y)$
- Nem sempre evidência está disponível na base de dados: **inferência probabilística**.

Axiomas da Probabilidade

- $0 \leq P(A) \leq 1$
- $P(True) = 1,\ P(False) = 0$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$ (ref. diag Venn).
- Propriedades:

$$\begin{array}{ll} P(A \vee \neg A) = & P(A) + P(\neg A) - P(A \wedge \neg A) \\ P(True) = & P(A) + P(\neg A) - P(False) \\ 1 = & P(A) + P(\neg A) \\ p(\neg A) = & 1 - P(A) \end{array}$$

- de Finetti: Crenças inconsistentes levam agente sempre a perder ou produzir ações não desejadas. Axiomas são importantes, pq não deixam agentes criarem inconsistências.

Distribuição de Probabilidade Conjunta

- $\mathbf{P}(X_1, \ldots, X_n)$: atribui probabilidades a todos os possíveis *eventos atômicos*.
- *Evento Atômico*: atribuição de valores particulares para todas as variáveis, especificação completa do domínio.
- Pe:

	DorDeDente	$\neg$ DorDeDente
Carie	0.04	0.06
$\neg$ Carie	0.01	0.89

Utilização de Distribuição de Probabilidade Conjunta

- P(Carie): $0.06 + 0.04 = 0.10$
- P(Carie $\vee$ DorDeDente) $= 0.04 + 0.01 + 0.06 = 0.11$
- Mecanismo de inferência pode encontrar probabilidades de uma proposição A, dada a evidência B.
- P(Carie| DorDeDente) $= \frac{0.04}{0.04+0.01} = 0.80$
- Em problemas reais, não é aconselhável usar distribuição de probabilidade conjunta: não prático armazenar a tabela com 2^n entradas para n variáveis booleanas.
- sistemas de raciocínio modernos: lidam diretamente com probabilidades condicionais.

Regra de Bayes

- $P(B \mid A) = \frac{P(A|B)P(B)}{P(A)}$: teorema, lei ou regra de Bayes.
- Fundamenta a maior parte dos sistemas probabilísticos de inferência.
- Forma geral:

$$P(Y \mid X) = \frac{P(X \mid Y).P(Y)}{P(X)}$$

- Na prática regra de Bayes é útil, porque na maioria dos problemas temos estimativas para as duas probabilidades incondicionais e para a probabilidade condicional.
- Pe, um médico sabe que meningite causa rigidez nos músculos do pescoço do paciente, em 50% dos casos.
- Tb sabe dois fatos incondicionais: a prob de um paciente ter meningite é $\frac{1}{50.000}$, e a prob de um paciente ter rigidez dos músculos do pescoço é $\frac{1}{20}$.

Regra de Bayes

- S = paciente tem rigidez do pescoço, M = paciente tem meningite.

$$\begin{aligned} P(S \mid M) &= 0.5 \\ P(M) &= \frac{1}{50.000} \\ P(S) &= \frac{1}{20} \\ P(M \mid S) &= \frac{P(S|M)P(M)}{P(S)} = \frac{0.5 \times \frac{1}{50.000}}{\frac{1}{20}} = 0.0002 \end{aligned}$$

- ie, somente 1 em cada 5000 pacientes com rigidez do pescoço têm meningite.

Regra de Bayes

- conhecimento sobre diagnósticos é mais superficial do que conhecimento causal.
- se surgir uma epidemia de meningite, P(M) vai subir.
- um médico que não usa regra de Bayes não vai saber como atualizar seus dados puramente estatísticos. Um médico usando regra de Bayes sabe que $P(M|S)$ cresce proporcionalmente com P(M).

Regra de Bayes: Normalização

- Considere novamente a equação para calcular a probabilidade de meningite dado rigidez do pescoço: $P(M \mid S) = \frac{P(S|M)P(M)}{P(S)}$
- Considere a possibilidade do paciente ter rigidez do pescoço e isto causar uma distensão dos músculos: $P(W \mid S) = \frac{P(S|W)P(W)}{P(S)}$
- **probabilidade relativa**:
  $P(S \mid W) = 0.8$ e $P(W) = \frac{1}{1000}$
  $\frac{P(M|S)}{P(W|S)} = \frac{P(S|M)P(M)}{P(S|W)P(W)} = \frac{1}{80}$
- ie, distensão é 80 vezes mais frequente do que meningite, dado que o paciente tem rigidez dos músculos do pescoço.

Regra de Bayes – Normalização

- Em alguns casos, probabilidades relativas são suficientes para tomar decisões, mas em outros casos, há necessidade de calcular valores mais precisos, sem precisar utilizar P(S) (probabilidade incondicional): **normalização**.
  $P(M \mid S) = \frac{P(S|M)P(M)}{P(S)}$
  $P(\neg M \mid S) = \frac{P(S|\neg M)P(\neg M)}{P(S)}$
  Adicionando: (obs: $P(M \mid S) + P(\neg M \mid S) = 1$)
  $P(S) = P(S \mid M)P(M) + P(S \mid \neg M)P(\neg M)$
- Substituindo na regra de Bayes:
  $P(M \mid S) = \frac{P(S|M)P(M)}{P(S|M)P(M)+P(S|\neg M)P(\neg M)}$
- regra geral: $\mathbf{P}(Y \mid X) = \alpha \mathbf{P}(X \mid Y)\mathbf{P}(Y)$

Usando a Regra de Bayes: combinação de evidências

- Assuma:
  $P(Carie \mid DorDeDente) = 0.8$
  $P(Carie \mid MotorPrendeu) = 0.95$
- O que é que um(a) dentista pode concluir se o motor prendeu no dente em que o paciente sente dor?
- Usando tabela de distr. de prob. conjunta, bastaria consultar tabela para encontrar $P(Carie \mid DorDeDente \wedge MotorPrendeu)$.
- Usando Bayes:

$$P(Carie \mid DorDeDente \wedge MotorPrendeu) =$$

$$= \frac{P(DorDeDente \wedge MotorPrendeu \mid Carie)P(Carie)}{P(DorDeDente \wedge MotorPrendeu)}$$

Usando a Regra de Bayes: combinação de evidências

- esta forma pode levar a um número exponencial de valores de probabilidade se tivermos conjunções com mais variáveis. Por que não voltar a usar tabela de prob. conjunta neste caso?
- A utilização pura de teoria das probabilidades leva a que, para alguns casos, tenhamos que fazer um número exponencial de cálculos. Neste caso, métodos aproximados para tratar combinações de evidências são utilizados.

Usando a Regra de Bayes: combinação de evidências

- Em alguns domínios, Bayes pode ser simplificada e usar menos valores de probabilidades para produzir resultados: **atualização bayesiana** (*bayesian updating*).

$$P(Carie \mid DorDeDente) =$$

$$= P(Carie)\frac{P(DorDeDente \mid Carie)}{P(DorDeDente)}$$

Usando a Regra de Bayes: combinação de evidências

- qdo MotorPrendeu é observado, aplicamos novamente Bayes com DorDeDente como variável condicional de contexto:

$$P(Carie \mid DorDeDente \wedge MotorPrendeu) =$$

$$= P(Carie \mid DorDeDente)\frac{P(MotorPrendeu \mid DorDeDente \wedge Carie)}{P(MotorPrendeu \mid DorDeDente)}$$

- Em atualização bayesiana, cada vez que uma nova evidência é observada, a crença é multiplicada por um fator que depende da nova evidência.

Usando a Regra de Bayes: combinação de evidências

- Ainda está complicado!
- $P(MotorPrendeu \mid DorDeDente \wedge Carie)$ não é mais fácil de ser calculado do que $P(DorDeDente \wedge MotorPrendeu \mid Carie)$!
- observação chave: cárie é causa direta da dor de dente e do motor ter agarrado ao dente, neste exemplo.
- Simplificação: **independência condicional** de DorDeDente e de MotorPrendeu, dado Carie:

  $$P(MotorPrendeu \mid Carie \wedge DorDeDente) = P(MotorPrendeu \mid Carie)$$
  $$P(DorDeDente \mid Carie \wedge MotorPrendeu) = P(DorDeDente \mid Carie)$$

- Simplificando:

  $$P(Carie \mid DorDeDente \wedge MotorPrendeu) = P(Carie)\frac{P(DorDeDente|Carie)}{P(DorDeDente)}\frac{P(MotorPrendeu|DorDeDente \wedge Carie)}{P(MotorPrendeu|DorDeDente)}$$

Sistemas de Raciocínio Probabilístico

- Como construir sistemas de raciocínio utilizando modelos de redes e que usem incerteza de acordo com a teoria das probabilidades?
- **Rede de Crença** ou rede bayesiana: grafo com as seguintes características:
  - Nós representam variáveis aleatórias.
  - Arcos direcionados representam ligações diretas entre variáveis aleatórias.
  - Cada nó tem uma tabela de prob. cond.que quantifica o efeito dos pais deste deste nó.
  - O grafo não possui ciclos (DAG).
- relativamente fácil para um especialista definir as relações do que definir as probs.

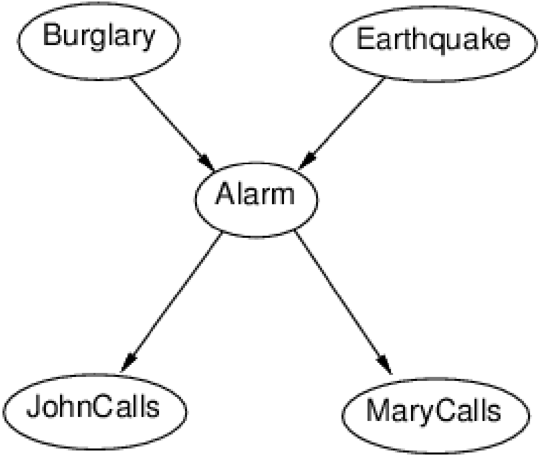
Sistemas de Raciocínio Probabilístico

- Redes de Crença:
  - Representação do Conhecimento
    - modelo gráfico de representação baseado no grafo acíclico
  - Dada a rede e as TPC:
    - podemos fazer consultas (inferência probabilística: exata ou aproximada)
  - Dada a rede e dados:
    - podemos "aprender" as TPC (parâmetros)
  - Se não tivermos a rede e nem as TPC:
    - podemos "aprender" as TPC (parâmetros) e a estrutura da rede

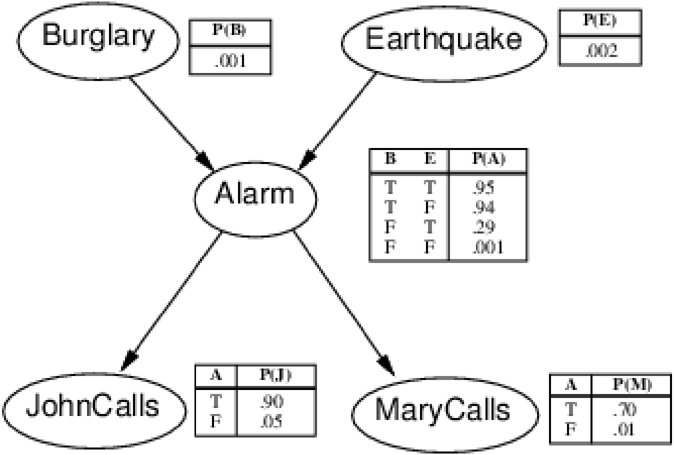
Sistemas de Raciocínio Probabilístico

- Exemplo: alarme contra roubo.
- Alarme toca em duas situações: tentativa de roubo e terremoto.
- John e Mary são os vizinhos que avisam ao dono da casa se o alarme estiver tocando.
- John liga toda vez q o alarme toca e tb qdo o tel toca.
- Mary somente liga qdo o alarme toca, mas não ouve algumas vezes.
- Dada a evidência de quem ligou para o dono da casa, queremos descobrir a probabilidade de ter havido roubo.

Sistemas de Raciocínio Probabilístico



Sistemas de Raciocínio Probabilístico



Sistemas de Raciocínio Probabilístico

- Rede somente representa ligações diretas, causais.
- Nada é informado sobre Mary ouvir música alta ou de John confundir o tel com o alarme.
- Tabela de probabilidades condicionais:

Roubo	Ter.	$\mathbf{P}(Alarme \mid Roubo, Terr)$
T	T	0.950 0.050
T	F	0.950 0.050
F	T	0.290 0.710
F	F	0.001 0.999

Semântica das Redes de Crenças

- Duas formas de entender:
  - representação da distribuição de probabilidade conjunta. Útil para **construir** redes.
  - conjunto de sentenças condicionalmente independentes. Útil para projetar procedimentos de inferência.
- Representando distribuição de probabilidade conjunta:
  - cada entrada na tabela pode ser calculada através da info na rede.
  - uma entrada genérica representa a probabilidade de uma conjunção de valores atribuídos a cada variável: $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$.
    $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid Pais(X_i))$

Inferência Probabilística

- Método de "Eliminação de variáveis" (**variable elimination**)
  - (aplicado ao exemplo Weather-Sprinkler-Lawn em aula teórico-prática)

Tipos de Inferência Probabilística

- Belief updating:

$$Bel(X_i) = P(X_i = x_i \mid evidence)$$

- Most Probable Explanation (MPE):

$$\overline{x}^* = \arg\max_{\overline{x}} P(\overline{x}, e)$$

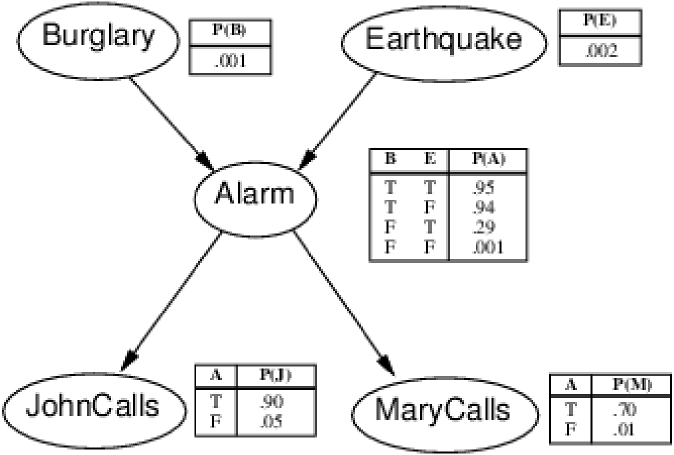
- Maximum A Posteriori hypothesis (MAP):

$$(a_1^*, \ldots, a_k^*) = \arg\max_{\overline{x}} \sum_{X/A} P(\overline{x}, e)$$

- Maximum Expected Utility (MEU) decision:

$$(d_1^*, \ldots, d_k^*) = \arg\max_{\overline{x}} \sum_{X/D} P(\overline{x}, e)\, U(\overline{x})$$

Redes de Bayes



Redes de Bayes

Calcular a probabilidade de ter havido um roubo dado que o alarme tocou.

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)} =$$

$$= \frac{P(A \mid B, E) \times P(B)}{P(A \mid B, E) \times P(B, E) + P(A \mid B, \neg E) \times P(B, \neg E) + P(A \mid \neg B, E) \times P(\neg B, E) + P(A \mid \neg B, \neg E) \times P(\neg B, \neg E)}$$

Construção das TPC

- dados completos: Maximum Likelihood (ML), MAP
- dados incompletos: Expectation Maximization
- quando faltam combinações de valores de variáveis: "imputação"
  - Maximum Likelihood
  - Laplace
  - m-estimate (precisa especificar um valor para m)

Construção de Redes de Crenças

- cada entrada na tabela é representada pelo produto dos elementos apropriados das tabelas de probabilidades condicionais (TPCs).
- TPCs fornecem uma representação decomposta da distribuição conjunta.
- Exemplo:
  $P(J \land M \land A \land \neg B \land \neg E) =$
  $= P(J \mid A)P(M \mid A)P(A \mid \neg B \land \neg E)P(\neg B)P(\neg E) =$
  $= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062$
- Métodos melhores do que obter a tabela inteira de distribuição de probabilidade conjunta.

Construção de Redes de Crenças

- Método para construir redes de crenças:
  - Rede é construída de forma que cada nó é condicionalmente independente dos seus predecessores, dada a probabilidade dos seus pais.
  - equação $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid Pais(x_i))$ usada para guiar o engenheiro de conhecimento a construir a topologia da rede.
  - Para construir uma rede de forma que esta tenha a estrutura correta para o domínio, escolhe-se nós pais adequados para garantir que cada nó é condicionalmente independente de seus antecessores.

Construção de Redes de Crenças

- Em geral:
  $\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = P(X_i \mid Pais(X_i))$, desde que $Pais(X_i)$ seja subconjunto ou igual a $\{x_{i-1}, \ldots, x_1\}$
- Esta condição é alcançada desde que se rotule os nós da rede em qq ordem q seja consistente com a ordem parcial implícita na estrutura do grafo.
- Ex: Mary telefona: não é **diretamente** influenciado por tentativa de assalto ou terremoto. É influenciado pelo efeito de tentativa de assalto ou terremoto, ou seja, soar o alarme.
- o Fato de John telefonar tb não tem influência direta sobre o fato de Mary telefonar. Neste caso temos independência condicional:
  $\mathbf{P}(MaryTel \mid JohnTel, Alarme, Terr, Assalto) = \mathbf{P}(MaryTel \mid Alarme)$

Construção de Redes de Crenças

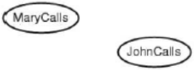
- Procedimento geral para construção de redes:
  1. Escolha o conj de variáveis X_i relevantes que descrevam o domínio.
  2. Escolha ordem para as variáveis.
  3. Eqto há vars:
     a) Pegue uma var X_i e adicione um nó na rede para X_i.
     b) Construir $Pais(X_i)$ com um conj mínimo de nós que já estejam na rede, tal que a prop de indep cond seja satisfeita.
     c) Defina a tabela de prob cond p/ X_i.

Construção de Redes de Crenças

- Procedimento garante que a rede é acíclica.
- Rede não contém valores de probabilidades redundantes.
- Garante que axiomas da prob não são violados.

Example of belief network construction

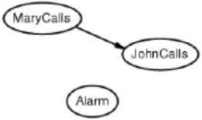
- Assuming the order: M, J, A, B, E



$P(J \mid M) = P(J)$?

Example of belief network construction

- Assuming the order: $M,\ J,\ A,\ B,\ E$



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)? P(A \mid J, M) = p(A)$?

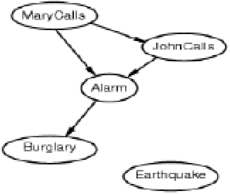
Example of belief network construction

- Assuming the order: M, J, A, B, E

$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)? P(A \mid J, M) = p(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$?
$P(B \mid A, J, M) = P(B)$?

Example of belief network construction

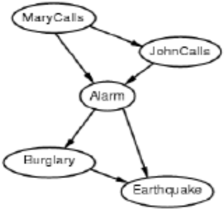
- Assuming the order: M, J, A, B, E



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)? P(A \mid J, M) = p(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$?
$P(E \mid B, A, J, M) = P(E \mid A, B)$?

Example of belief network construction

- Assuming the order: M, J, A, B, E



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)?P(A \mid J, M) = p(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$? No
$P(E \mid B, A, J, M) = P(E \mid A, B)$? Yes

Compactação e Ordenação de Nós

- Redes de crenças + **compactas** do que distribuição de prob conjunta.
- Sistemas **localmente estruturados** ou esparsos com info distribuída pelos nós.
- Crescimento polinomial.
- Em redes de crenças, podemos assumir que para a maioria dos domínios, cada variável aleatória é diretamente influenciada por no max k outras vars (nós pais).
- Qtde necessária de valores para a TPC de cada nó: 2^k.
- Para a rede completa (n nós): $n2^k$.

Compactação e Ordenação de Nós

- Ex concreto: rede com 20 nós e no max 5 pais p/ cada nó:
  - redes de crença: 640 valores.
  - tabela de distr de prob conj: ordem de 10^6 valores.
- Número de links extra na rede = maior precisão, mas pode não se justificar devido ao aumento do tamanho das tabelas.
- Regra geral: adicionar à rede primeiro os nós causadores de algum efeito e depois seus efeitos.

Representação de TPCs

- Problema: escolher as probs condicionais das TPCs.
- Relação entre pais e filhos pode se encaixar numa distribuição canônica. Neste caso, probs podem ser especificadas através de nomes e talvez parms adicionais.
- Ex mais simples: nós determinísticos, probs são iguais às probs dos pais.
- Nós não determinísticos: relação *ruidosa* OU.
- Representação das probs:
  - Se todos os pais F, nó de saída F, com 100% de certeza.
  - Se apenas 1 dos pais é V, nó de saída é F com prob = parâmetro ruidoso daquele nó pai que é V.

Representação de TPCs

- Ex: $P(Febre \mid Resf) = 0.4$, $P(Febre \mid Gripe) = 0.8$ e $P(Febre \mid Mal) = 0.9$

Resf	Gripe	Mal	P(Febre)	P(¬ Febre)
F	F	F	0.0	1.0
F	F	V	0.9	0.1
F	V	F	0.8	0.2
F	V	V	0.98	$0.02 = 0.2 \times 0.1$
V	F	F	0.4	0.6
V	F	V	0.94	$0.06 = 0.6 \times 0.1$
V	V	F	0.88	$0.12 = 0.6 \times 0.2$
V	V	V	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

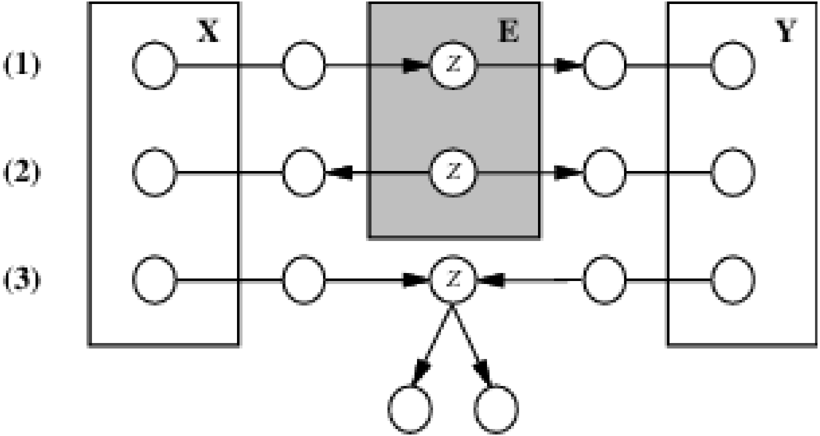
Relações de Independência Condicional

- Necessidade: saber se relações de independência condicional mais gerais (não somente de pais p/ filhos) existem para poder obter mecanismos de inferência capazes de responder consultas do tipo: "Existe algum conj de nós X independente de outro conj Y, dado conj de evidências E?"
- Método: **direction-dependent separation** ou **d-separation** (separação-d).
- separação-d: *se todo ramo não dirigido de um nó em X para um nó em Y é d-separado por E, então X e Y são condicionalmente independentes, dado E.*
- Um conj de nós E d-separa dois conj X e Y se todo ramo não dirigido de um nó em X para um nó em Y estiver **bloqueado**, dado E.

Relações de Independência Condicional

- Um ramo está bloqueado, dado um conj de nós E, se houver um nó Z no caminho tal que uma das três condições é satisfeita:
  1. Z está em E e Z tem um arco incidente e outro não incidente.
  2. Z está em E e Z tem ambos os arcos não incidentes.
  3. Nem Z nem nenhum descendente de Z estão em E, e ambos os arcos incidem em Z.

Inferência em Redes de Crenças



Inferência em Redes de Crenças

- Objetivo principal: computar distribuição de probabilidade posterior para um conj de variáveis de consulta, dados valores exatos para variáveis de evidência: $P(Cons \mid Evidencia)$.
- A princípio, qq nó pode servir como consulta ou evidência.
- Duas funções: BELIEF_NET_TELL p/ incluir novas evidências na rede e BELIEF_NET_ASK p/ computar novas probs p/ as variáveis de consulta.

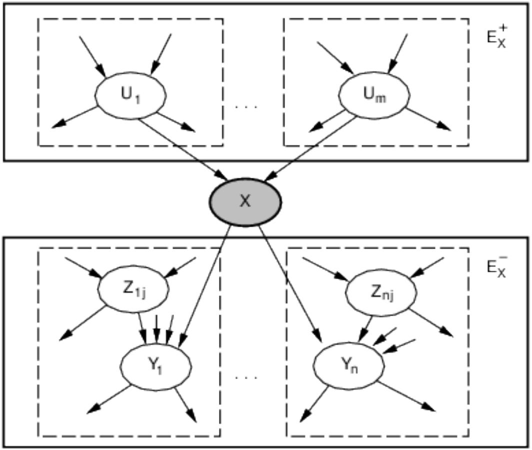
Inferência em Redes de Crenças

- **Diagnósticas**: $P(Assalto \mid JohnTel)$ (efeitos p/ as causas).
- **Causais**: $P(JohnTel \mid Assalto)$ (causas p/ os efeitos).
- **Inter-causais**: $P(Assalto \mid Alarme \land Terr)$.
- **Mistas**: combinação de 1 ou mais dos casos acima.
- Redes de crenças ainda podem ser usadas para:
  - tomar decisões baseadas em probs na rede e em funções de utilidade do agente.
  - decidir quais variáveis de evidência observar para obter info mais útil.
  - fazer "análise de sensibilidade" para entender aspectos do modelo que têm > impacto nas probs das vars de consulta.
  - Explicar os resultados da inferência probabilística para o utilizador.

Inferência em Redes de Crenças

- Algoritmo BELIEF_NET_ASK análogo ao backward-chain, faz inferências a partir das variáveis de consulta até encontrar alguma evidência.
- Algoritmo funciona somente para redes "singly connected", onde há no máximo um ramo não dirigido entre quaisquer dois nós da rede: **poli-árvores**.
- Algoritmos para redes mais gerais usam algoritmos poli-árvore como sub-rotinas.

Inferência em Redes de Crenças



Inferência em Redes de Crenças

- Nó X tem pais U e filhos Y.
- "singly connected" significa que todos os blocos são disjuntos e não têm links.
- X é a variável de consulta.
- Objetivo: computar $P(X \mid E)$.
- Conjunto de **suporte causal**: variáveis de evidência "acima" de X que estão conectadas a X através dos pais.
- Conjunto de **suporte evidencial**: variáveis de evidência "abaixo" de X que estão conectadas a X através dos filhos.
- $E_{U_i|X}$: evidências conectadas com todos os nós U_i, *exceto* via o ramo que passa por X.

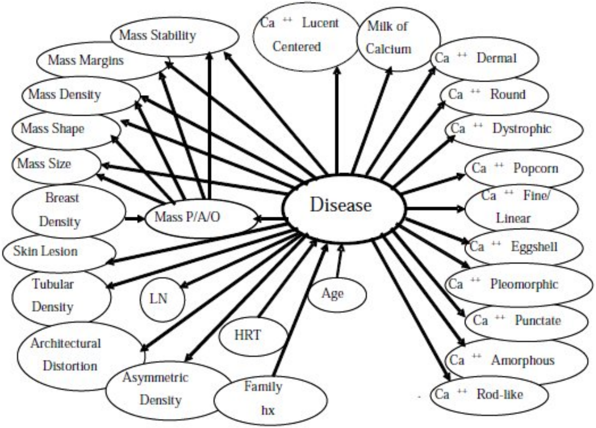
Inferência em Redes de Crenças

- Estratégia geral:
  - Representar $\mathbf{P}(X \mid E)$ em termos de contribuições de E_X^+ e E_X^-.
  - Computar a contribuição de E_X^+ através do seu efeito nos pais de X. Obs: computar os efeitos dos pais de X em X pode ser feito de forma recursiva.
  - Computar a contribuição de E_X^- através do seu efeito nos pais de X. Obs: computar os efeitos dos pais de X em X pode ser feito de forma recursiva.
- Método: aplicar Bayes, outros métodos padrão de prob, simplificações (independência condicional).

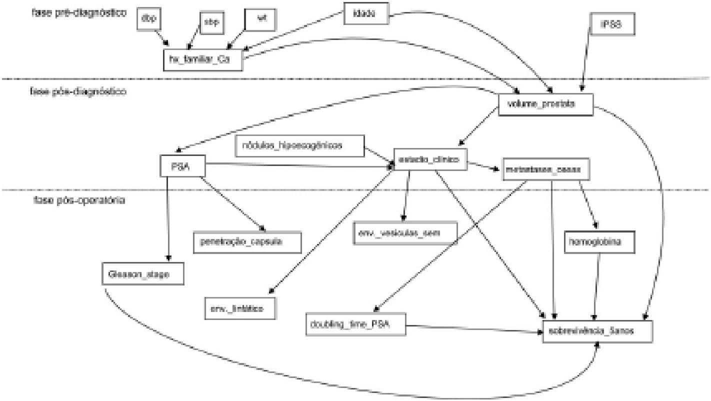
Inferência em Redes de Crenças

- $\mathbf{P}(X \mid E) = \mathbf{P}(X \mid E_X^-, E_X^+) = \frac{\mathbf{P}(E_X^- \mid X, E_X^+)\mathbf{P}(X \mid E_X^+)}{\mathbf{P}(E_X^- \mid E_X^+)}$
- Como X d-separa E_X^+ de E_X^- na rede, podemos usar indep cond p/ simplificar primeiro termo do numerador. Tb podemos usar $\frac{1}{\mathbf{P}(E_X^- \mid E_X^+)}$ como cte de normalização:
  $\mathbf{P}(X \mid E) = \alpha \mathbf{P}(E_X^- \mid X)\mathbf{P}(X \mid E_X^+)$
- $\mathbf{P}(X \mid E_X^+) = \sum_u \mathbf{P}(X \mid \mathbf{u}, E_X^+) P(\mathbf{u} \mid E_X^+)$
- $\mathbf{P}(X \mid E_X^+) = \sum_u \mathbf{P}(X \mid \mathbf{u}) \prod_i \mathbf{P}(u_i \mid E_X^+)$
- $\mathbf{P}(X \mid E_X^+) = \sum_u \mathbf{P}(X \mid \mathbf{u}) \prod_i \mathbf{P}(u_i \mid E_{U_i|X})$

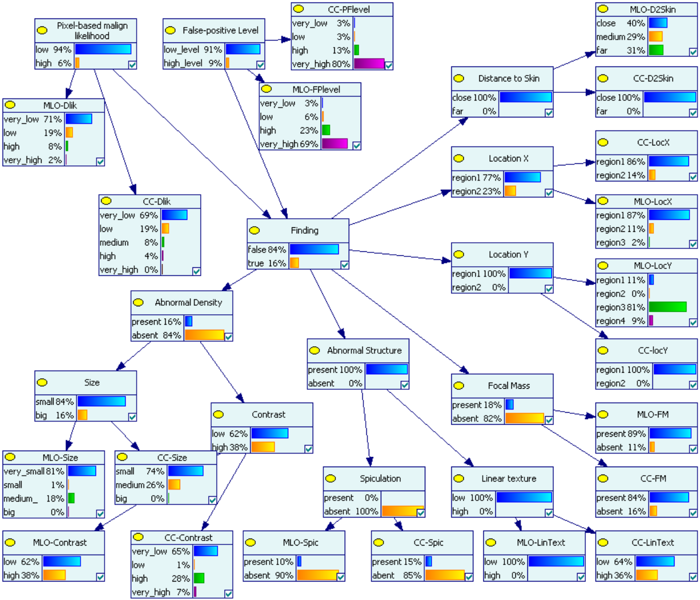
Redes de Bayes



Redes de Bayes



Redes de Bayes



Decision Theory steps

- **Create causal model**
  - Determine possible symptoms, disorders, treatments, and outcomes.
  - Draw arcs indicating what disorders cause what symptoms, and what treatments alleviate what disorders.
  - Often the model will match well informal graphical descriptions given in medical textbooks.

Decision Theory steps

- **Simplify to a qualitative decision model**
  - If we are using the model to make treatment decisions, we can often simplify the model by removing variables that are not involved in treatment decisions (discriminative model x generative model).

Decision Theory steps

- **Assign probabilities**
  - these can come from patient databases, literature studies or the expert's subjective assessments.
  - (arguably) easier for humans: assess the probability of a cause given an effect (for example, $P(S \mid M)$).
  - diagnostic reasoning ($P(S \mid M)$) performed by the Bayesian network inference algorithm.

Decision Theory steps

- **Assign utilities**
  - for a small number of possible of outcomes: create a scale from best to worst and give each a numeric value between 0 and 1 (0 can be healthy person and 1 a person with a disease).
  - this can be done by an expert, but it is recommended to involve the patient.
  - if there are exponentially many outcomes, combine them to use multiattribute utility functions (for example, we nay say that the costs of various comorbities are additive).

Decision Theory steps

- **Verify and refine the model**
  - need a set of correct (input, output) pairs: **gold standard** to compare against our outcomes.
  - medical conference meetings that decide on diagnosis and recommended treatment plan.
  - if the system does poorly, isolate parts that are going wrong and fix them.
  - verification can be done "backward": instead of presenting symptoms and asking for a diagnosis, present the system with the diagnosis and examine the predicted probability of symptoms and compare with the medical literature.

Decision Theory steps

- **Perform sensitivity analysis**
  - checks if the best decision is sensitive to small changes in the assigned probabilities and utilities by systematically varying those parameters and running the evaluation again.
  - if all variations lead to the same decision we can have some confidence that the system is taking the right decision.