CONHECIMENTO PROBABILÍSTICO: REPRESENTAÇÕES

MODELOS DE DEPENDÊNCIA

Sumário

- ◇ Inadequação de Representações Numéricas
- ◇Representação de Dependências
- ◇ Base Axiomática
- ◇ Dependências e Grafos Não-Direcionados
- ◇ Redes de Markov
- ◇ Redes Bayesianas
- ◇ Expressividade de Redes Bayesianas

Inadequação de Representações Numéricas

Distr. conjunta $P(X_1, \ldots, X_n)$: entidade básica em probabilidade.
Computacionalmente inadequado:

- 2^n endereços de memória para armazenamento (X_i booleanos).
- Cálculo de probabilidades marginais é tarefa muito árdua:

$$P(X_i = x) = \sum_{j \neq i} P(X_1, \ldots, X_i = x, \ldots, X_n)$$

- Cálculo de probabilidades condicionais é tarefa muito árdua:

$$P(X_i \mid X_j = x_j) = \alpha P\left(X_i, X_j = x_j\right) = \alpha \sum_y P\left(X_i, X_j = x_j, y\right)$$

Além disso, valores <u>exatos</u> de probabilidade não tem grande significado. Precisamos representar probabilidades (como medida de incerteza) de uma maneira mais eficiente e significativa.

CT215 3

Noção Qualitativa de Independência

O que nos interessa é a noção <u>qualitativa</u> de independência: raciocínio humano não mede independência via $P(x,y)=P(x)P(y)$.

Por outro lado, avaliamos facilmente a relação básica de independência condicional: Dado Z, Y é influenciado por X?
Exemplo:
X = horário em que o último ônibus deixou o ponto.
Y = tempo estimado para passar o próximo ônibus.
Z = local em que o próximo ônibus está agora.

$$P(Y|X,Z) = P(Y|Z)$$

Queremos definir uma representação que permita <u>rapidamente</u> visualizar e avaliar este tipo de independência.

CT215 4

Relação entre Dependência e Inferência

Seja *K* o conhecimento numa BC. Quero avaliar a certeza a respeito de *A*. Convém consultar uma nova proposição *B* não armazenada na BC?

Em outras palavras: antes de examinar *B*, quero saber se existe algo não codificado na BC que pode gerar conhecimento relevante a respeito de *A*.

Relevância da informação:permite separar o que interessa do que não interessa para fins de inferência e geração de novo conhecimento.

Relevância em Teoria da Probabilidade é formalizada pelo conceito de independência condicional:

A é independente de *B* dado *K* se **P(A|B,K) = P(A|K)** ou
B é irrelevante para inferir A dado K se **P(A|B,K) = P(A|K)**

É este tipo de relação que devemos representar de forma eficiente!

Dependências: Base Axiomática

Def. Seja $U=\{\alpha, \beta, \ldots\}$ um conjunto de variáveis discretas, e seja P uma função de probabilidade conjunta sobre as variáveis em U. Sejam X, Y e Z três subconjuntos de U. Diz-se que X e Y são condicionalmente independentes dado Z se

$$P(X=x|Y=y,Z=z)=P(X=x|Z=z)$$

Notação: $I(X,Z,Y) \Leftrightarrow P(X=x \mid Y=y, Z=z) = P(X=x|Z=z)$.

Algumas propriedades:

$$I(X,Z,Y) \Leftrightarrow P(x,y|z) = P(x|z)P(y|z)$$
$$I(X,Z,Y) \Leftrightarrow \exists f, g: P(x,y,z) = f(x,z)g(y,z)$$
$$I(X,Z,Y) \Leftrightarrow P(x,y,z) = P(x|z)P(y,z)$$

Teoremas

• **Simetria**: $I(X,Z,Y) \Leftrightarrow I(Y,Z,X)$. Dado *Z*, se *Y* é irrelevante para *X* então *X* é irrelevante para *Y*.
• **Decomposição**: $I(X,Z,Y \cup W) \Rightarrow I(X,Z,Y) \wedge I(X,Z,W)$.Se dois itens de informação são irrelevantes para *X*, então cada um isoladamente é irrelevante para *X*.
• **União Fraca**: $I(X,Z,Y \cup W) \Rightarrow I(X,Z \cup W,Y)$. Apreender a irrelevância de uma informação *W* não torna outras informações relevantes.
• **Contração**: $I(X,Z,Y) \wedge I(X,Z \cup Y,W) \Rightarrow I(X,Z,Y \cup W)$. Se descobrimos que *W* é irrelevante para *X* após aprendermos que *Y é* irrelevante para *X*, então *W* já era irrelevante antes..
• **Intersecção**: $I(X,Z \cup W,Y) \wedge I(X,Z \cup Y,W) \Rightarrow I(X,Z,Y \cup W)$.. Se *Y* não afeta *X* quando *W* não muda e *W* não afeta *X* quando *Y* não muda, então *W* e *Y* nunca afetam *X*

(Pearl: fig. 3.1, pág. 86)

Representação Gráfica de Dependências

Estrutura escolhida: grafos de dependência (nós e ligações).

◇ Grafos têm representação pictórica: metáfora conveniente para dependências conceituais (nós- conceitos, ligações - dependências).

◇ Grafos permitem operações de visualização imediata (estabelecimento de dependências via análise de caminhos, etc.)

◇Parecem ter correspondência com raciocínio humano (e.g., figuras de linguagem, redes semânticas).

Esquemas gráficos mais úteis: Redes de Markov e Redes Bayesianas.

Seja portanto um grafo *G* representando interação entre elementos.
Idealmente:

- Independência = falta de conexão entre nós correspondentes.
- Dependência = existência de conexão entre elementos correspondentes.

Dependências e Grafos Não-Direcionados

Problema: topologia baseada apenas em nós, ligações e conexões é pouco expressiva. Preciso considerar ligações indiretas como um meio para representar interação condicional.

Exemplo: Dois machos $\{M_1,M_2\}$ e duas fêmeas $\{F_1, F_2\}$ em uma sociedade heterossexual. Estrutura em diamante indica:

- Falta de contato direto entre dois machos ou duas fêmeas
- Dependência condicional. *Exemplo.*

Observe que:

- Falta de direcionalidade indica origem indiferente de doença.
- Assimetria pode ser codificada (via pesos).
- Semântica da topologia é definida pelas ligações ausentes.c

Modelos de Dependência

Def. Seja U=$\{\alpha,\beta\}$ um conjunto de variáveis discretas, e sejam *X*, *Y* e *Z* três subconjuntos de *U*. Seja *M* um modelo de dependência, que atribui *V* ou *F* para o predicado $I(X,Z,Y)_M$.

O modelo *M* basicamente determina o subconjunto *I* de triplas para as quais a afirmação "*X* é independente de *Y* dado *Z*" é verdadeira.

Observe que qualquer distribuição de probabilidade *P é* um modelo de dependência (Posso verificar sempre se $P(x|y,z) = P(x|z)$.

Uma representação gráfica de dependência deveria estabelecer uma correspondência direta entre os elementos *U* de *M* e o conjunto de nós desta representação:

$$\langle X|Z|Y\rangle_G \Rightarrow I(X,Z,Y)_M$$

e

$$I(X,Z,Y)_M \Rightarrow \langle X \mid Z \mid Y\rangle_G$$

Modelos de Dependências

Def. Um grafo não direcionado *G* é um mapa de dependência (D-mapa) de *M* se existir uma correspondência unívoca entre elementos de *U* e nós de *G*, de modo que

$$I(X,Z,Y)_M \Rightarrow \langle X \mid Z \mid Y\rangle_G$$

No grafo: Se Z não corta X, Y então x e Y são dependentes. Caso contrário, não posso dizer nada.

Similarmente, *G* é uma mapa de independência (I-mapa) se:

$$I(X,Z,Y)_M \Leftarrow \langle X \mid Z \mid Y\rangle_G$$

No grafo: Se Z corta X, Y então X e Y são independentes. Caso contrário, não posso dizer nada.

***G* é um mapa perfeito de M se for simultaneamente um I-mapa e D-mapa de M.**

Insuficiência de Grafos Não-Direcionados

Vários modelos de dependência não correspondem a mapas perfeitos.
Exemplo: Dependência Induzida - proposições não-relacionadas tornam-se relevantes entre si quando novos fatos são apreendidos:

$$I(X,Z_1,Y)_M \wedge \neg I(X,Z_1 \cup Z_2,Y)_M$$

Mas separações em grafos sempre satisfazem

$$\langle X|Z_1|Y\rangle_G \Rightarrow \langle X,Z_1 \cup Z_2,Y)_M$$

Exemplo: moedas e sino.

Grafos não-direcionados são portanto representacionalmente pobres. Vamos nos conformar com grafos não-direcionados que exibem apenas as dependências (I-mapas).

Redes de Markov

Perguntas:

1. Dada uma distribuição *P*, é possível construir um I-mapa *G* de *P* com número mínimo de vértices? SIM! (redes de Markov)
2. Dado um par *(P,G)*, é possível testar se *G* é um I-mapa de *P*? SIM!
3. Dado um grafo *G*, é possível construir uma distribuição *P* tal que *G* é mapa perfeito de *P*? SIM!

Número mínimo de vértices garante que pelo menos minimizaremos o número de independências não-representáveis, mas redes de Markov:

- Não podem representar dependências induzidas ou não-transitivas.
- Relacionam variáveis independentes caso estas influenciem uma terceira.

Muitas independências relevantes deixam portanto de ser representadas em redes de Markov, apesar dos nossos esforços em definí-las como I-mapas mínimos.

Precisamos enriquecer a semântica de grafos adicionando o conceito de direção à relações. Setas permitem discriminar entre dependências reais e dependências induzidas por hipóteses.

Exemplo: moedas e sino.

Redes Bayesianas

Perguntas:

1. Dada uma distribuição *P*, é possível se construir um **Grafo Acíclico Direcionado** (GAD) *D* que é um I-mapa mínimo de *P*? SIM!
2. Dado um par *(P,D)*, é possível testar se *D* é um I-mapa de *P*? SIM!
3. Dado um GAD *D*, é possível se construir uma distribuição *P* tal que *D* é um mapa perfeito de *P*? SIM!

◇Nós = variáveis

◇Arcos = influência direta entre variáveis

Em redes de Markov, tínhamos: $\langle X|Z|Y \rangle_G \Rightarrow I(X,Z,Y)$

EM REDES BAYESIANAS, O CRITÉRIO DE DEPENDÊNCIA É BASEADO EM SEPARAÇÃO-d.

Semântica de Redes Bayesianas

Def. Sejam *X*, *Y* e *Z* subconjuntos disjuntos de um GAD D. Diz-se que *Z* d-separa *X* de *Y* ($\langle X|Z|Y \rangle_D$) se, para qualquer caminho entre *X* e *Y*, existir um nó *w* satisfazendo uma das seguintes condições:

- *w* tem apenas setas convergentes e nem *w* nem nenhum de seus descendentes está em *Z*;
- *w* não tem apenas setas convergentes e está em *Z*.

Um caminho que satisfaz uma destas condições é dito bloqueado.
Um caminho que não satisfaz nenhuma destas condições é dito ativado (por *Z*).

semantica redes bayesianas. pdf

Redes Bayesianas como I-Mapas

Um GAD é um I-mapa de um modelo de dependência *M* se

$$\langle X|Z|Y \rangle_D \Rightarrow I(X,Z,Y)_M$$

para quaisquer subconjuntos disjuntos *X*, *Y* e *Z*.

Um GAD é um I-mapa mínimo de *M* se nenhuma de suas setas puder ser removida sem destruir a condição de I-mapa..

Def. Dada uma distribuição *P* e um conjunto de variáveis *U*, um GAD *D* é dita uma Rede Bayesiana de *P* se *D* é um I-mapa mínimo de *P*.

Construindo Redes Bayesianas

Dada uma distribuição *P* e um dado ordenamento de variáveis de *U*, existe um procedimento recursivo simples para construir uma rede Bayesiana. E dada uma rede Bayesiana, existe um procedimento simples para reconstruir $P(X_1, X_2, . . . ,X_n)$

Na prática, o problema é definir as variáveis e suas interdependências (quase nunca disponho de *P*). Entretanto, só preciso de relações diretas.A própria rede deve definir as influências indiretas.

EM REDES BAYESIANAS, JULGAMENTO LOCAL MANTÉM CONSISTÊNCIA GLOBAL.

construção redes bayesianas. pdf

Expressividade de Redes Bayesianas

◇ Redes Bayesianas são mais expressivas: indução e dependências não-transitivas.

◇ Redes Bayesianas capturam um conjunto maior de independências probabilísticas do que redes de Markov (especialmente aquelas que parecem ser mais úteis em problemas reais).

Entretanto:

◇ Redes Bayesianas não representam tudo que redes de Markov representam. Exemplo: estrutura em diamante.

◇ Nenhuma representação gráfica consegue distinguir **conectividade entre conjuntos** de **conectividade entre elementos**. Em probabilidade, independência entre elementos não implica independência entre conjuntos.