Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.1-6.5

Relatório semana 2 - MAC0215 (Atividade Curricular em Pesquisa)
Aluno: Renato Lui Geh (Bacharelado em Ciência da Computação)
Orientador: Denis Deratani Mauá

1 Atividades realizadas na semana

Durante a semana foram lidos os seguintes tópicos do livro *Modeling and Reasoning with Bayesian Networks*[1]:

6 - Inference by Variable Elimination

- **6.1** - Introduction
- **6.2** - The Process of Elimination
- **6.3** - Factors
- **6.4** - Elimination as a Basis for Inference
- **6.5** - Computing Prior Marginals

2 Definição das atividades

Foi estudado o processo de eliminação de uma variável em uma Rede Bayesiana, simplificando a rede para computarmos inferência. Como tópicos importantes temos a definição de fatores (*factors*), as operações de soma (*summing out*) e multiplicação de factors e como chegarmos em inferência e *prior marginals* a partir de eliminação de variáveis.

Esta seção será dividida em subseções para cada tópico que citamos no parágrafo anterior com a inclusão de um tópico de introdução para assuntos que foram pesquisados fora dos subcapítulos lidos e que não estão presentes nos outros relatórios. Abaixo está a lista de tópicos deste relatório.

1. Introdução
2. Factors
3. Summing out factors
4. Multiplying factors
5. Eliminação e propriedades
6. Ordem de eliminação

2.1 Introdução

O processo de eliminação de uma variável é feita da seguinte forma. Dado um conjunto de variáveis de uma Rede Bayesiana Q, escolhe-se um subconjunto Y que queremos eliminar. O subconjunto resultante $X = Q \setminus Y$ tem suas variáveis somadas, resultando num conjunto onde desconsideramos os resultados de X. Em seguida, multiplicamos as variáveis resultantes, criando um conjunto união que acarreta na eliminação das variáveis X. Para fazer essas operações precisamos introduzir o conceito de fatores (*factors*), que veremos na subseção seguinte.

2.2 Factors

Um factor é uma função que retorna números não-negativos sobre um conjunto de variáveis, mapeando cada instância da variável a um número não-negativo. As tabelas 1 e 2 mostram os factors 1 e 2 e os valores de suas instâncias. Na tabela 1, o factor 1 representa a distribuição condicional de D dado B e C, enquanto que o factor 2 na tabela 2 mostra a probabilidade das variáveis D e E.

B	C	D	f_1
true	true	true	0.95
true	true	false	0.01
true	false	true	0.9
true	false	false	0.1
false	true	true	0.8
false	true	false	0.2
false	false	true	0
false	false	false	1

D	E	f_2
true	true	0.448
true	false	0.192
false	true	0.112
false	false	0.248

Tabelas 1 e 2: $f_1(b, c, d) = Pr(d|b, c)$ e $f_2(d, e) = Pr(d, e)$.

Vamos definir um factor formalmente:

Definição. *Um factor f sob variáveis $\boldsymbol{X}$ é uma função que mapeia cada instância x das variáveis $\boldsymbol{X}$ em um número não-negativo $f(x)$. Um factor também é chamado de potencial (potential).*

Vamos chamar de $vars(f)$ para denotar as variáveis em f. Chamaremos de $f(X_1, ..., X_n)$ para indicar que $X_1, ..., X_n$ são as variáveis definidas em f. Definimos um factor com conjunto de variáveis vazio como um factor trivial, denotado por $\top$.

Duas operações são comuns em factors: *summing out* e *multiplying*. Vamos definir na próxima subseção o que são essas duas operações.

2.3 Summing out factors

Vamos definir o que é somar variáveis (*summing out*).

Definição. *Seja f um factor sobre as variáveis $\boldsymbol{X}$ e seja X uma variável em $\boldsymbol{X}$. O resultado de summing out a variável X de f é um factor sob as variáveis $\boldsymbol{Y} = \boldsymbol{X} \setminus \{X\}$, que é denotado por $\sum_x f$ e definido como:*

$$\left(\sum_x f\right)(y) \stackrel{\text{def}}{=} \sum_x f(x, \boldsymbol{y}) \tag{1}$$

Podemos dizer que somar uma variável é somar todas as probabilidades das instâncias de um factor em que todos os valores das variáveis são equivalentes entre si, descartando os valores da variável que queremos somar. Na tabela abaixo somamos a variável D do factor f_1 da Tabela 1.

B	C	$\sum_D f_1$
true	true	1
true	false	1
false	true	1
false	false	1

Se somarmos todas as variáveis de um factor teremos um factor trivial, ou seja, que possui um conjunto de variáveis vazio. Por exemplo, na Tabela 4 somamos todas as variáveis (B e C) da tabela anterior. Como resultado temos um factor trivial de valor 4.

	$\sum_B \sum_C \sum_D f_1$
$\top$	4

Summing out é uma operação comutativa, portanto a ordem de somas de variáveis não importa. Somar variáveis $\mathbf{X}$ é chamado de marginalizar (*marginalizing*) as variáveis $\mathbf{X}$ do fator f. Se $\mathbf{Y}$ são as outras variáveis de f, então $\sum_{\mathbf{X}} f$ é chamado de o resultado da projeção (*projecting*) de f nas variáveis $\mathbf{Y}$.

É importante notar que o Algoritmo 1 da seção 6.3 do livro tem complexidade $O(\exp(w))$ em tempo e espaço, onde w é o número de variáveis do factor a ser somado.

2.4 Multiplying factors

A segunda operações com factors é multiplicar (*multiplying*) factors. Definimos formalmente a seguir.

Definição. *A multiplicação de factors $f_1(\boldsymbol{X})$ e $f_2(\boldsymbol{T})$ é um factor sob as variáveis $\boldsymbol{Z} = \boldsymbol{X} \cup \boldsymbol{Y}$ denotado como $f_1 f_2$ e definido como:*

$$(f_1 f_2)(\boldsymbol{z}) \stackrel{\text{def}}{=} f_1(\boldsymbol{x}) f_2(\boldsymbol{y}), \tag{2}$$

onde $\boldsymbol{x}$ e $\boldsymbol{y}$ são compatíveis[2] com $\boldsymbol{z}$, ou seja, $\boldsymbol{x} \sim \boldsymbol{z}$ e $\boldsymbol{y} \sim \boldsymbol{z}$.

Multiplicar dois factors é criar uma união dos dois factors multiplicando cada probabilidade de um factor com o do outro factor, descartando as linhas inválidas (aquelas em que há uma contradição de variáveis; por exemplo se $D = true$ e $D = false$ ao mesmo tempo). Na tabela a seguir tempos o resultado da multiplicação entre factors f_1 e f_2 que vimos nas Tabelas 1 e 2.

B	C	D	E	$f_1(B,C,D)f_2(D,E)$
true	true	true	true	$0.4256 = (0.95)(0.448)$
true	true	true	false	$0.1824 = (0.95)(0.192)$
true	true	false	true	$0.0056 = (0.05)(0.112)$
⋮	⋮	⋮	⋮	⋮
false	false	false	false	$0.2480 = (1)(0.248)$

Pode-se ver que cada instância b, c, d, e é compatível com exatamente uma instância em f_1 e exatamente uma instância em f_2. Multiplicação é uma operação comutativa e associativa. Portanto, a ordem das multiplicações não altera seu factor resultante.

A implementação em pseudo-código do livro denotado como Algoritmo 2 na seção 6.3 tem complexidade $O(m\ \exp(w))$ em tempo e espaço, onde w é o número de variáveis no factor resultante e m é o número de factors multiplicados.

2.5 Eliminação e propriedades

Já vimos que somar e multiplicar são ambos comutativos e multiplicar é associativa. A partir disso vamos elaborar algumas propriedades da eliminação de variáveis (somar e multiplicar). Antes disso vamos definir o que é a regra da cadeia (*chain rule*) de uma Rede Bayesiana.

Definição. *Seja $\boldsymbol{z}$ uma instanciação do conjunto de variáveis $\boldsymbol{Z}$. A probabilidade de uma rede dada uma instância $\boldsymbol{z}$ é o produto de todos os parâmetros da rede que sejam compatíveis com $\boldsymbol{z}$, ou seja:*

$$Pr(\boldsymbol{z}) \stackrel{\text{def}}{=} \prod_{\theta_{x|\boldsymbol{u}} \sim \boldsymbol{z}} \theta_{x|\boldsymbol{u}} \tag{3}$$

Por exemplo:

$$Pr(a, b, \overline{c}, d, \overline{e}) = \theta_a \theta_{b|a} \theta_{\overline{c}|a} \theta_{d|b,\overline{c}} \theta_{\overline{e}|\overline{c}} \tag{4}$$

A partir da regra da cadeia podemos representar a probabilidade de cada instância como um produto de parâmetros da rede:

$$Pr(a, b, c, d, e) = \theta_{e|c} \theta_{d|bc} \theta_{c|a} \theta_{b|a} \theta_a \tag{5}$$

Podemos multiplicar as CPTs[2] alternativamente, usando CPTs como se fossem factors. Desse jeito podemos expressar a distribuição conjunta de probabilidade como um produto das CPTs da Rede Bayesiana.

$$\Theta_{E|C} \Theta_{D|BC} \Theta_{C|A} \Theta_{B|A} \Theta_A \tag{6}$$

Portanto, para calcular a distribuição marginal sob as variáveis D e E podemos somar as variáveis A, B e C. A marginal corresponde ao factor:

$$Pr(D, E) = \sum_{A,B,C} \Theta_{E|C} \Theta_{D|BC} \Theta_{C|A} \Theta_{B|A} \Theta_A \tag{7}$$

Pode-se ver que o factor acima tem complexidade exponencial, já que a soma e multiplicação de factors são ambos exponenciais. Para melhorarmos a complexidade podemos usar o seguinte teorema:

Teorema. *Sejam f_1 e f_2 factors. Se existe um X tal que $X \in vars(f_2)$ mas $X \notin vars(f_1)$, então:*

$$\sum_X f_1 f_2 = f_1 \sum_X f_2 \tag{8}$$

Portanto somamos apenas os factors que possuem variáveis em comum e em seguida multiplicamos os outros factors. Por exemplo, sejam $f_1, ..., f_n$ factors e queremos somar a variável X, e sabemos que X aparece apenas nos factors f_k e f_p, então podemos agrupar f_k e f_p e multiplicarmos e somarmos antes de multiplicar o resto.

$$\sum_X f_1...f_n = f_1...f_{k-1}f_{k+1}...f_{p-1}f_{p+1}...f_n \sum_X f_k f_p \tag{9}$$

Ou seja, para somar uma variável X no produto $f_1...f_n$, multiplicamos os factors f_k que incluem X e então somamos a variável X do factor resultante $\prod_k f_k$.

Um exemplo da marginalização encontra-se na página 133 do livro[1] na subseção 6.4.

2.6 Ordem de eliminação

Vamos usar a notação $\pi(i)$ para representar a i-ésima variável na ordem de eliminação. Quando dizemos que vamos *eliminar a variável* $\pi(i)$ estamos nos referindo a multiplicação de factors f_k onde $\pi(i)$ aparece, seguido a soma da variável $\pi(i)$.

Como vimos na subseção anterior, a ordem da eliminação pode alterar a complexidade do algoritmo. Podemos definir se uma ordem é pior que outra se o factor é maior que o outro, ou seja, se há mais variáveis no factor a serem somadas.

Teorema. *Se o maior factor construído na soma de uma variável $\pi(i)$ tem w variáveis, então a complexidade de se multiplicar e somar os factors que possuem $\pi(i)$ é $O(n \exp(w))$, onde n é o número de variáveis na Rede Bayesiana.*

Chamamos w de a largura (*width*) da ordem usada π e é usada para medir a qualidade da ordem. Queremos achar uma ordem que tem menor largura possível.

3 Conclusão

Factors são essenciais para podermos fazer operações de soma e multiplicação, e portanto eliminação de variáveis. Eliminação de variável é um método para poder-se computar inferência. No entanto, como vimos, métodos para se multiplicar e somar variáveis são intratáveis e exponenciais. Um outro ponto muito importante é podermos representar CPTs como factors.

Referências

[1] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. 1st Edition. Cambridge University Press, 2009.

[2] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Compiling Bayesian Networks*. 1. 2015. URL: `http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week1/relatorio.pdf`.