

Modeling and Reasoning with Bayesian Networks: Compiling Bayesian Networks

Relatório semana 1 - 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]:

12 - Compiling Bayesian Networks

12.1 - Introduction

12.2 - Circuit semantics

12.3 - Circuit propagation

12.3.1 - Evaluation and differentiation passes

2 DEFINIÇÃO DAS ATIVIDADES

Foram estudados o processo de se compilar Redes Bayesianas em circuitos aritméticos, algumas notações usadas em Redes Bayesianas, a definição de uma *network polynomial*, algumas propriedades de Redes Bayesianas e diferenciação de uma rede a partir de uma evidência.

Esta seção será dividida em subseções para cada subtópico citado na seção anterior. Cada subseção é um resumo do que foi estudado em cada tópico, contendo os assuntos mais importantes para o projeto.

2.1 INTRODUCTION

Aqui apresentaremos algumas notações usadas em Redes Bayesianas - e que podem ser estendidas para outros Modelos Gráficos Probabilísticos (PGM).

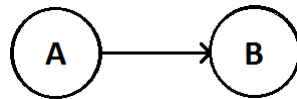


Figura 1: Uma Rede Bayesiana $A \rightarrow B$. Em Redes Bayesianas uma aresta representa uma dependência. No caso da imagem A depende de B . As CPTs associadas a esse grafo estão em Tabela 1 e Tabela 2.

Tabela 1 e Tabela 2

| | | A | B | $\Theta_{B A}$ |
|-------|--------------------------|-------|-------|----------------------------------|
| A | Θ_A | true | true | $\theta_{b a} = 0.1$ |
| true | $\theta_a = 0.3$ | true | false | $\theta_{\bar{b} a} = 0.9$ |
| false | $\theta_{\bar{a}} = 0.7$ | false | true | $\theta_{b \bar{a}} = 0.8$ |
| | | false | false | $\theta_{\bar{b} \bar{a}} = 0.2$ |

Antes de começarmos com circuitos aritméticos, vamos primeiro apresentar algumas definições importantes.

Chamamos de CPT (Conditional Probability Table) as tabelas que representam as probabilidades de uma rede (ex.: as CPTs da Rede Bayesiana na Figura 1 são as Tabelas 1 e 2). Pode-se claramente ver que para n nós de uma Rede Bayesiana, precisamos de uma quantidade exponencial 2^n de probabilidades para representar cada instância de variáveis.

Chamamos de MPE (Most Probable Explanation) a instância mais provável de variáveis que aceitam uma evidência. Mais formalmente dizemos que:

Definição. *Sejam X_1, \dots, X_n as variáveis da rede e e a evidência dada. Existe uma instância x_1, \dots, x_n onde $Pr(x_1, \dots, x_n|e)$ é maximal. Chamamos x_1, \dots, x_n a explicação mais provável (most probable explanation) dado e .*

Chamamos de MAP (Maximum A Posteriori hypothesis) quando a probabilidade de uma certa instância é maximal.

Definição. *Sejam X o conjunto de todas as variáveis da rede e M um subconjunto qualquer dessas variáveis. Dado uma evidência e , qualquer instância m de variáveis M onde $Pr(m|e)$ é maximal é uma hipótese máxima a posteriori (maximum a posteriori hypothesis).*

M também é chamado de variáveis MAP. MPE é uma MAP quando as variáveis MAP incluem todas as variáveis da rede.

Agora que temos uma base podemos voltar para circuitos aritméticos. Dado um circuito aritmético que representa uma Rede Bayesiana, teremos dois tipos de entradas: variáveis θ , que chamaremos de *parâmetros*, e as variáveis λ , que chamaremos de *indicadores*. Parâmetros são valorados de acordo com a CPT da rede, enquanto indicadores são valorados de acordo com a evidência. Nas próximas subseções veremos que podemos ter duas passagens pelo circuito. Uma em que vamos de baixo para cima (bottom-up) para calcular a probabilidade de uma dada evidência, e outra em que vamos de cima para baixo (top-down), chamada de *differentiation pass* para calcularmos as derivadas parciais de cada entrada do circuito.

2.2 CIRCUIT SEMANTICS

Ao compilarmos um circuito aritmético de uma Rede Bayesiana estamos representando, de forma compacta, a distribuição de probabilidade induzida da rede. No caso da Figura 1, a distribuição de probabilidade está representada na Tabela 3.

Tabela 3

| A | B | $Pr(A, B)$ |
|-----------|-----------|---|
| a | b | $\theta_a \theta_{b a}$ |
| a | \bar{b} | $\theta_a \theta_{\bar{b} a}$ |
| \bar{a} | b | $\theta_{\bar{a}} \theta_{b \bar{a}}$ |
| \bar{a} | \bar{b} | $\theta_{\bar{a}} \theta_{\bar{b} \bar{a}}$ |

Multiplicando cada $Pr(A, B)$ com as suas respectivas variáveis indicadoras temos que:

Tabela 4

| A | B | $Pr(A, B)$ |
|-----------|-----------|---|
| a | b | $\lambda_a \lambda_b \theta_a \theta_{b a}$ |
| a | \bar{b} | $\lambda_a \lambda_{\bar{b}} \theta_a \theta_{\bar{b} a}$ |
| \bar{a} | b | $\lambda_{\bar{a}} \lambda_b \theta_{\bar{a}} \theta_{b \bar{a}}$ |
| \bar{a} | \bar{b} | $\lambda_{\bar{a}} \lambda_{\bar{b}} \theta_{\bar{a}} \theta_{\bar{b} \bar{a}}$ |

Somando todas as probabilidades da distribuição temos:

$$f = \lambda_a \lambda_b \theta_a \theta_{b|a} + \lambda_a \lambda_{\bar{b}} \theta_a \theta_{\bar{b}|a} + \lambda_{\bar{a}} \lambda_b \theta_{\bar{a}} \theta_{b|\bar{a}} + \lambda_{\bar{a}} \lambda_{\bar{b}} \theta_{\bar{a}} \theta_{\bar{b}|\bar{a}} \quad (1)$$

Chamamos a função f de *network polynomial*, que representa a distribuição de probabilidade da rede na Figura 1. Para computar a probabilidade dada qualquer evidência e , definimos cada variável indicadora de forma consistente com a evidência. Vamos definir o que significa definir uma variável de forma consistente com uma evidência:

Definição. *Sejam um conjunto $X = \{x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n\}$ das variáveis da rede e uma evidência $e = \{e_p, \dots, e_q\}$. Dizemos que X está consistente com e quando para cada e_i , se $e_i = 1$ então $x_i = 1$ e $\bar{x}_i = 0$; e se $e_i = 0$ então $x_i = 0$ e $\bar{x}_i = 1$. Para todo j tal que $1 \leq j \leq n$ e que não exista um $e_j \in e$ definiremos $x_j = \bar{x}_j = 1$. Por questões de visibilidade, definimos x_i, \bar{x}_i ao invés de $\lambda_i, \lambda_{\bar{i}}$, porém ambas as notações são equivalentes.*

Portanto, se por exemplo tivermos evidência $e = \bar{a}$, teremos na *network polynomial* da Figura 1 os seguintes indicadores: $\lambda_a = 0, \lambda_{\bar{a}} = 1, \lambda_b = 1, \lambda_{\bar{b}} = 1$. Neste exemplo, o valor de f seria:

$$\begin{aligned} f(\mathbf{e} = \bar{a}) &= (0)(1)\theta_a \theta_{b|a} + (0)(1)\theta_a \theta_{\bar{b}|a} + (1)(1)\theta_{\bar{a}} \theta_{b|\bar{a}} + (1)(1)\theta_{\bar{a}} \theta_{\bar{b}|\bar{a}} \\ &= \theta_{\bar{a}} \theta_{b|\bar{a}} + \theta_{\bar{a}} \theta_{\bar{b}|\bar{a}} \\ &= Pr(\mathbf{e}) \end{aligned} \quad (2)$$

REFERÊNCIAS

- [1] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*.