Uma Introdução a Sum-Product Networks

Relatório semana 9 - 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 as seguintes partes dos papers abaixo:

- *Learning the Structure of Sum-Product Networks*, [R. Gens, P. Domingos] [5]
  - Introduction
  - Sum-Product Networks
- *Sum-Product Networks: A New Deep Architecture*, [H. Poon, P. Domingos] [6]
  - Introduction
  - Sum-Product Networks
  - Sum-Product Networks and other models

2 Definição das atividades

Os tópicos mencionados na seção anterior referem-se à definição de uma Sum-Product Network e citam algumas semelhanças com outros modelos probabilísticos assim como suas diferenças.

Neste relatório vamos definir o que são Sum-Product Networks de uma forma mais didática e vamos supor que o leitor tenha conhecimento prévio de todo conteúdo coberto nos relatórios anteriores. Após termos definido Sum-Product Networks, vamos ver algumas propriedades e teoremas relacionados e em seguida vamos comparar, de forma sucinta, com outros modelos probabilísticos.

Vamos separar esta seção nos seguintes tópicos:

1. Introdução
   1.1. Distribuição normalizada de produtos de factors
   1.2. Função de partição
2. Definição
3. Propriedades
4. Comparação

2.1 Introdução

Um dos maiores problemas com modelos gráficos é a intractabilidade da inferência e aprendizado da estrutura. Inferência é sempre exponencial no pior caso, e como aprendizado usa inferência, a complexidade continua intratável. Além do mais, a amostragem necessária para aprendizado preciso é também exponencial no pior dos casos no tamanho do escopo. De fato existem modelos gráficos onde a inferência é tratável, no entanto elas são limitadas quanto às representações de distribuições de forma compacta.

Vamos mostrar que Sum-Product Networks (SPN), um novo tipo de arquitetura profunda, permite que computemos a função partição, a probabilidade de evidência e o estado MAP[2] com complexidade linear no número de arestas da SPN. Também vamos definir validade de uma SPN assim como completude e consistência. Depois vamos mostrar outras definições assim como alguns teoremas derivados dessas propriedades.

Antes de começarmos a definir Sum-Product Networks, precisamos antes explicar o que é uma distribuição normalizada de produtos de factors e definir uma função de partição.

2.1.1 Distribuição normalizada de produtos de factors

O objetivo de modelos gráficos probabilísticos é representar distribuições de forma compacta. Podemos representar tais distribuições como um produto normalizado dos factors[3] envolvidos. Tal representação é um jeito compacto de se representar as CPTs envolvidas.

Definição. *Sejam $x \in \mathcal{X}$ um vetor d-dimensional representando uma instância de d variáveis, ϕ_k uma função potential[3] do subconjunto $x_{\{k\}}$ de variáveis (ou seja, seu escopo[1]) e Z a função partição que veremos mais a frente. Representamos distribuições compactamente como o seguinte produto normalizado:*

$$P(X = x) = \frac{1}{Z} \prod_k \phi_k(x_{\{k\}}) \tag{1}$$

A representação acima é dita normalizada pois queremos representa-la como uma probabilidade, ou seja, um número real no intervalo $[0, 1]$. Como pode-se notar, dividimos o produtório por Z, a chamada função partição. De fato, como veremos a seguir, a função partição normaliza o produto dos factors.

2.1.2 Função de partição

Dizemos a função partição uma função que toma como argumentos todos os estados das variáveis e retorna a soma de todos os produtórios de todos os factors de cada estado.

Definição. *Seja ϕ_k uma função potential, dizemos que a função partição é*

$$Z = \sum_{x \in X} \prod_k \phi_k(x_{\{k\}}) \tag{2}$$

Portanto, é fácil notar que $\frac{1}{Z}\prod_k \phi_k(x_{\{k\}})$ é uma normalização por Z, já que Z é a soma de todos os possíveis resultados do produtório, e portanto será sempre maior ou igual ao valor do produtório normalizado, levando a $0 \leq P(X = x) \leq 1$, assumindo-se que $\phi_i \geq 0$.

No caso de $Z = 1$, então temos o caso trivial onde o produtório dos factors dada uma instância já está dentro do intervalo $[0, 1]$.

Uma das dificuldades de se computar inferência em modelos gráficos é a intractabilidade de Z, já que Z é a soma de um número exponencial de termos. Como todas as marginals[4] são somas de subconjuntos desses termos, computa-las é igualmente intratável. No entanto, se acharmos uma maneira eficiente de computar Z, então também podemos computar as marginals eficientemente. Mas Z é computado apenas com somas e produtos, e pode ser eficientemente computado se aplicarmos a distributiva em Z de tal forma que envolvamos um número polinomial de somas e produtos.

2.2 Definição

Assim como em [P. Domingos, H. Poon][6], vamos introduzir Sum-Product Networks com variáveis Booleanas. Mais para frente veremos que para variáveis discretas ou contínuas o processo é similar.

Antes de definirmos SPNs, vamos introduzir algumas notações:

- A negação de X_i é representada por $\overline{X_i}$.
- A função indicadora[2] [.] tem valor 1 se seu argumento é *true* e 0 caso contrário.
- Abreviaremos $[X_i]$ por x_i e $[\overline{X}_i]$ por $\overline{x_i}$.

Referências

[1] Renato Lui Geh. *Aprendizado Automático de Sum-Product Networks (SPN)*. 2015. URL: http://www.ime.usp.br/~renatolg/mac0215/doc/project/relatorio.pdf.

[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.

[3] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.1-6.5*. 2. 2015. URL: http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week2/relatorio.pdf.

[4] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.6-6.9*. 3. 2015. URL: http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week5/relatorio.pdf.

[5] Robert Gens e Pedro Domingos. "Learning the Structure of Sum-Product Networks". Em: *International Conference on Machine Learning* 30 (2013).

[6] Hoifung Poon e Pedro Domingos. "Sum-Product Networks: A New Deep Architecture". Em: *Uncertainty in Artificial Intelligence* 27 (2011).