Aprendizado Automático de Sum-Product Networks (SPN)

Projeto de MAC0215 (Atividade Curricular em Pesquisa)
Aluno: Renato Lui Geh (Bacharelado em Ciência da Computação)
Orientador: Denis Deratani Mauá

1 Introdução

Para compreendermos melhor as aplicações de uma Sum-Product Network (SPN) precisamos primeiro entender distribuições de probabilidade multivariadas, o que é o escopo de uma distribuição, as dificuldades de representar todas as probabilidades no espaço e o que é inferência de uma distribuição.

1.1 Distribuições de probabilidade multivariadas

Uma distribuição de probabilidade multivariada é uma distribuição de probabilidade que resulta na probabilidade de que cada variável aleatória $X_1, ..., X_n$, para $n \geq 2$, esteja em um dado intervalo ou conjunto para cada valor x_i, $1 \leq i \leq n$ correspondente, sendo usadas em classificação de imagens em Sum-Product Networks. Se a distribuição possui apenas uma variável aleatória, dizemos que ela é monovariável. Segue abaixo um exemplo:

Consideremos um dado não viciado. Seja $A = 1$ se o número tirado é par e $A = 0$ caso contrário. Seja $B = 1$ se o mesmo número tirado em A é primo e $B = 0$ caso contrário. Podemos construir a tabela abaixo:

Tabela 1

	1	2	3	4	5	6
A	0	1	0	1	0	1
B	0	1	1	0	1	0

A distribuição de probabilidade multivariada resultante é:

$$P(A = 0, B = 0) = P\{1\} = \frac{1}{6}$$
$$P(A = 1, B = 0) = P\{4, 6\} = \frac{2}{6}$$
$$P(A = 0, B = 1) = P\{3, 5\} = \frac{2}{6}$$
$$P(A = 1, B = 1) = P\{2\} = \frac{1}{6}$$

1.2 Escopo de uma distribuição

Denotamos $P(X_1 = x_1, ..., X_n = x_n)$, dados valores $x_1, ..., x_n$, uma distribuição de probabilidade multivariada com variáveis aleatórias $X_1, ..., X_n$. Dada uma distribuição de probabilidade, chamamos de escopo da distribuição o conjunto de variáveis que definem o domínio da função, ou seja, $X_1, ..., X_n$. No caso da Tabela 1, o escopo da probabilidade é o conjunto $\{A, B\}$.

Definição. *Sejam $P_1(X_1, ..., X_k)$ e $P_2(X_p, ..., X_n)$ distribuições de probabilidade multivariadas cujos escopos têm mesmo tamanho.*

1. *Dizemos que P_1 e P_2 tem mesmo escopo se para toda variável aleatória X_i em P_1 existe uma variável aleatória X_j em P_2 tal que $i = j$.*
2. *Dizemos que P_1 e P_2 tem escopos disjuntos se para qualquer variável aleatória X_i em P_1, toda variável aleatória X_j em P_2 obedece a regra $i \neq j$.*

1.3 Espaço de uma distribuição multivariada

Vamos reutilizar a Tabela 1, renomeando A para X_1 e B para X_2. Portanto teremos $P(X_1 = x_1, X_2 = x_2)$. Montando a tabela temos que:

Tabela 2

X_1	X_2	$P(X_1 = x_1, X_2 = x_2)$
$X_1 = 0$	$X_2 = 0$	$\frac{1}{6}$
$X_1 = 0$	$X_2 = 1$	$\frac{2}{6}$
$X_1 = 1$	$X_2 = 0$	$\frac{2}{6}$
$X_1 = 1$	$X_2 = 1$	$\frac{1}{6}$

Pode-se ver que uma tabela com todas as possíveis probabilidades de todas as variáveis terá 2^n elementos, onde n é o número de variáveis aleatórias. Portanto o crescimento em espaço necessário é exponencial, o que é impraticável em conjuntos de dados maiores. A solução vem com o uso de Redes Profundas (ex.: Sum-Product Networks), que tornam $O(2^n)$ em $O(n)$ no tamanho do grafo.

1.4 Inferência

Diz-se inferência a probabilidade de uma distribuição multivariada dados valores fixos de certas variáveis aleatórias. Formalmente, dado um conjunto e de evidência que representa os valores $X_{i_1} = x_{i_1}, ..., X_{i_k} = x_{i_k}$, a inferência de $P(X_1, ..., X_n)$ dado e é a probabilidade normalizada $P(x) = \frac{P'(x)}{Z}$, onde Z é a função partição que normaliza $P'(x)$ em uma distribuição de probabilidade e $P'(x)$ é a função não normalizada de $P(X_1, ..., X_n)$ dado evidência e que representa a inferência. Em Sum-Product Networks, $Z = S(*) = S(1, ..., 1)$.

Como um exemplo, consideremos o caso em que as variáveis aleatórias representam variáveis Booleanas. Seja o conjunto evidência $e = \{X_1 = 1, X_2 = 0\}$ numa distribuição $P(X_1, X_2, X_3, X_4) = P(x_1, x_2, x_3, x_4, \overline{x_1}, \overline{x_2}, \overline{x_3}, \overline{x_4})$, associamos todas as ocorrências $x_i \in e$ às suas equivalentes em P e o resto a 0. Para as variáveis que não fixamos, associamos a 1. Então, a inferência é $P(x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1, \overline{x_1} = 0, \overline{x_2} = 1, \overline{x_3} = 1, \overline{x_4} = 1)$, que representa todas as combinações de probabilidade de $X_1, ..., X_4$ dados eventos conhecidos $X_1 = 1$ e $X_2 = 0$.

2 Definição de uma Sum-Product Network

Sum-Product Networks (SPN) são uma nova classe de modelos probabilísticos cuja inferência é sempre polinomial no tamanho da rede.

Definição. *Pela definição recursiva de Gens e Domingos[4]:*

1. *Uma distribuição monovariável polinomial é uma SPN.*
2. *Um produto de SPNs cujos escopos são disjuntos é uma SPN.*
3. *Uma soma de SPNs com peso não negativo cujos elementos tem mesmo escopo é uma SPN.*
4. *Nada mais é uma SPN.*

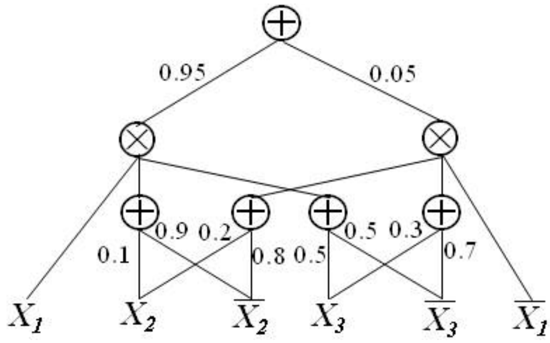

Figura 1: Um exemplo de uma SPN com variáveis Booleanas, onde $x_1, ..., x_d$ e $\overline{x_1}, ..., \overline{x_d}$ são folhas e o resto dos nós são somas ou produtos.[6]

Podemos definir graficamente uma SPN como um grafo direcionado acíclico (DAG) onde as folhas são sempre variáveis (ou distribuições monovariáveis), seus nós internos são somas ou produtos e para todo vértice conectando um nó soma com um nó filho há um peso não negativo. Também assume-se que toda soma e produto estão em alturas alternantes, ou seja, todo nó pai de um nó interno que é soma é um produto e vice-versa.

Tomando a Figura 1 como exemplo, considere variáveis aleatórias binárias $X_1, ..., X_n \in \{0, 1\}$. Para este exemplo $n = 3$. Para todo X_i, indicamos por x_i o indicador em que $X_i = 1$ e $\overline{x}_i$ para quando $X_i = 0$. Portanto, quando $X_i = 1$, $x_i = 1$ e $\overline{x}_i = 0$; e quando $X_i = 0$, $x_i = 0$ e $\overline{x}_i = 1$. A função multilinear do diagrama que representa a SPN é a seguinte:

$$\begin{aligned} S(X_1, X_2, X_3) = & \\ & = S(x_1, x_2, x_3, \overline{x}_1, \overline{x}_2, \overline{x}_3) = \\ & = 0.95x_1(0.1x_2 + 0.9\overline{x}_2)(0.5x_3 + 0.5\overline{x}_3) + \\ & \quad + 0.05\overline{x}_1(0.2x_2 + 0.8\overline{x}_2)(0.3x_3 + 0.7\overline{x}_3) \end{aligned} \tag{1}$$

Quando todos indicadores pertencentes ao escopo da SPN S são iguais a 1, diz-se que S é $S(*)$ e que ela é a função partição. $S(x)/S(*)$ é a distribuição de probabilidade normalizada da SPN S.

3 Características

Há várias vantagens de SPNs sobre outras redes de aprendizado:

1. SPNs têm estrutura parecida a outros Modelos Gráficos Probabilisticos (PGM), mas a representação de certos tipos de independência é mais fácil.
2. SPNs têm inferência polinomial no tamanho do grafo, enquanto que inferência em Redes Bayesianas é NP-difícil.
3. Experimentos mostram que aprendizado de arquiteturas SPN tiveram melhores resultados quando comparadas a arquiteturas estáticas.[2]

4 Aplicações

SPNs obtiveram resultados impressionantes em muitos conjuntos de dados[3], tais como:

- Reconstrução de imagens.
- Classificação.
- Reconhecimento de atividade.
- Logs click-through.
- Sequências de ácido nucleico.
- Filtragem colaborativa.

5 Proposta

O objetivo deste projeto é realizar uma comparação das principais técnicas de aprendizado automático de Sum-Product Network a partir de um conjunto de dados, avaliando o desempenho e os resultados das técnicas utilizadas.

Será estudada a estrutura e propriedades de uma Sum-Product Network e em seguida serão aplicadas as seguintes técnicas de aprendizado:

- Divisão em subconjuntos mutualmente independentes e EM clustering.[4]
- Busca gulosa.[1]
- Aprendizado com clustering de variáveis.[2]
- Aprendizado por meio de SPNs Bayesianas Não-Paramétricas.[5]

Depois o aluno irá comparar as técnicas implementadas e apresentar os resultados.

6 Acompanhamento

Relatórios semanais serão publicados no seguinte endereço:

`http://www.ime.usp.br/~renatolg/mac0215/doc/reports/`

É também possível acompanhar tanto os relatórios quanto as implementações pelo repositório do projeto:

`https://github.com/RenatoGeh/mac0215/`

7 Referências

Artigos

[1] Aaron Dennis e Dan Ventura. "Greedy Structure Search for Sum-Product Networks". Em: *International Joint Conference on Artificial Intelligence* 24 (2015).

[2] Aaron Dennis e Dan Ventura. "Learning the Architecture of Sum-Product Networks Using Clustering on Variables". Em: *Advances in Neural Information Processing Systems* 25 (2012).

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

[5] Sang-Woo Lee, Christopher Watkins e Byoung-Tak Zhang. "Non-Parametric Bayesian Sum-Product Networks". Em: *Workshop on Learning Tractable Probabilistic Models* (2014).

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

Websites

[3] University of Washington Department of Computer Science. *Sum-Product Networks*. URL: `http://spn.cs.washington.edu/index.shtml`.

Os artigos listados na referência acima podem ser encontrados em: `http://www.ime.usp.br/~renatolg/mac0215/articles/`