概率图推理 - 信念传播算法

作者：Calvin
QQ: 179209347
Mail: 179209347@qq.com

介绍

笔记简介：

- 面向对象：深度学习初学者
- 依赖课程：**线性代数，统计概率**，优化理论，图论，离散数学，微积分，信息论

知乎专栏：

https://zhuanlan.zhihu.com/p/693738275

Github & Gitee 地址：

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

*** 版权声明：**

- 仅限用于个人学习
- 禁止用于任何商业用途

概率图模型 – 推理（ Inference ）

推理（ Inference ）：指在已知部分变量的情况下，计算其他变量的概率分布或期望值。

常见的推理方法包括:

1. 精确推理（Exact Inference)

变量消除（Variable Elimination）：

- 通过逐步消除变量来计算目标变量的边缘概率分布。
- 对于贝叶斯网络，可以通过消除非目标变量来获得边缘概率。

信念传播（Belief Propagation, BP）：

- 在树形图中，通过消息传递算法来计算边缘概率分布。
- 针对一般图，可以使用近似信念传播（Loopy Belief Propagation）。

2. 近似推理（Approximate Inference)

蒙特卡罗方法（Monte Carlo Methods）：

- 通过随机采样来近似计算概率分布。
- 常用方法包括马尔可夫链蒙特卡罗（MCMC）、重要性采样（Importance Sampling）。

变分推理（Variational Inference）：

- 将复杂的概率分布近似为更简单的分布，通过优化方法进行推理。
- 常用技术包括变分贝叶斯（Variational Bayesian）、期望最大化（Expectation-Maximization）。

信念传播算法

信念传播（Belief Propagation, BP）算法是一种用于概率推理的消息传递算法，主要应用于贝叶斯网络和马尔可夫随机场等图模型中，也称为和积（Sum-Product）算法或消息传递（Message Passing）算法，将变量消除法中的和积（Sum-Product）操作看作是消息，并保存起来，这样可以节省大量的计算资源。该算法由Judea Pearl在20世纪80年代提出，主要用于计算隐变量的边缘概率分布。

算法流程

信念传播算法分为两种形式：标准信念传播（也称为精确信念传播）和近似信念传播。标准信念传播适用于树形结构的图，而近似信念传播则适用于一般图。

标准信念传播（树形图）

- **初始化**：为每个节点初始化信念（belief），通常是均匀分布或者根据先验概率设定。
- **消息传递**：
  1. 从叶节点向根节点传递消息（消息是关于相邻节点的边缘概率）。
  2. 根节点接收到所有子节点的消息后，向下传递消息，直到所有节点都接收到来自其父节点的消息。
- **更新信念**：每个节点根据接收到的消息和自身的观测值，更新其信念。
- **迭代**：重复消息传递和更新信念，直到收敛或者达到预设的迭代次数。

近似信念传播（一般图）

对于一般图，由于存在环，标准信念传播可能无法收敛。因此需要使用近似方法，例如：

- **Loopy Belief Propagation（循环信念传播）**：在存在环的图中使用信念传播算法，允许在固定次数的迭代后停止，即使消息传递可能没有完全收敛。
- **变分推理**：将概率推理问题转化为优化问题，通过最小化某种距离（如KL散度）来逼近真实的后验分布。

信念传播算法

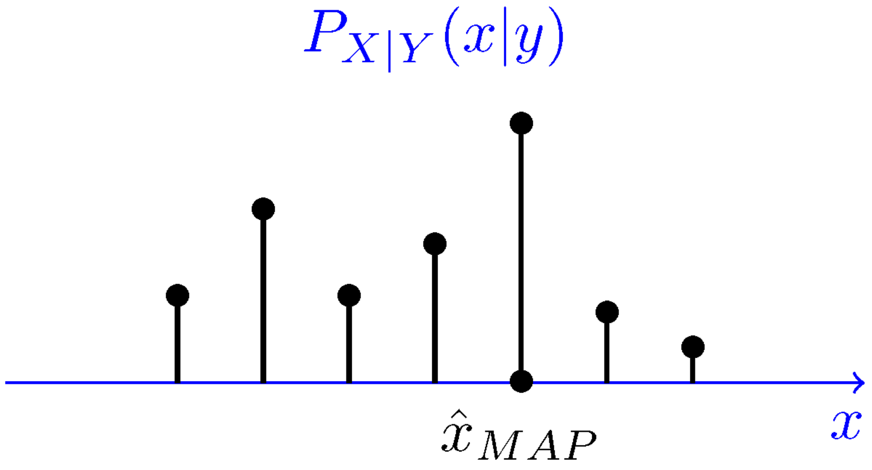
优点和局限性
优点:
- 适用于大规模稀疏图结构的概率推理。
- 对于树形结构，能精确计算边缘概率。

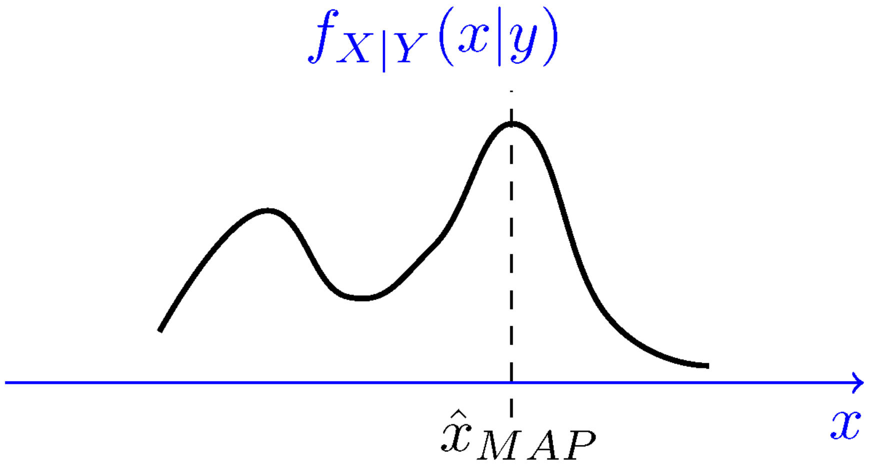
局限性:
- 对于含有大量环的图，标准信念传播可能无法收敛。
- 近似信念传播可能会得到次优解。

最大后验概率（MAP）

最大后验概率 (Maximum a posterior，MAP):
最大后验概率（MAP）是指在贝叶斯统计学中，给定观察到的数据，确定使后验概率最大化的状态。换句话说，它是在考虑了观察数据的情况下，所能获得的最有可能的状态。在贝叶斯推断中，后验概率是基于先验概率和观察数据计算得到的，表示在观察到数据后对参数或未知变量的信念程度。

$$\mathbf{x}^* = \arg \max_{\mathbf{X} \in \mathcal{X}} p(\mathbf{X})$$





最大后验概率（MAP）

根据贝叶斯公式，我们可以得到后验概率作为似然概率和先验概率的乘积：

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

$$\propto P(X|\theta)P(\theta) \qquad \text{我们忽略了归一化常数 } P(X)$$

我们将MLE公式中的可能性替换为后验，我们可以得到：

$$\begin{aligned}
\theta_{MAP} &= \arg\max_{\theta} P(X \mid \theta)P(\theta) \\
&= \arg\max_{\theta} \log P(X \mid \theta) + \log P(\theta) \\
&= \arg\max_{\theta} \log \prod_{i} P(x_i \mid \theta) + \log P(\theta) \\
&= \arg\max_{\theta} \sum_{i} \log P(x_i \mid \theta) + \log P(\theta)
\end{aligned}$$

比较MLE方程和MAP方程，唯一不同的是MAP中包含先验 $P(\theta)$ ，否则它们是相同的。这意味着，可能性现在被来自先验的一些权重加权。

最大后验概率（MAP）

如果在MAP估计中使用最简单的先验，比如均匀先验。这意味着可以在所有可能的θ值上分配相等的权重。这意味着可能性由一些常数等效加权。由于是常数，它可以从MAP方程中忽略，因为它不会对最大值有贡献。
假设可以为 θ 分配六个可能的值。现在，我验 $P(\theta)$ 在分布中处处都是 1/6。因此，可以在MAP估计中忽略该常数。

$$
\begin{aligned}
\theta_{MAP} &= \arg\max_{\theta} \sum_{i} log\, P\,(x_i|\theta) + log\, P\,(\theta) \\
&= \arg\max_{\theta} \sum_{i} log\, P\,(x_i|\theta) + const \\
&= \arg\max_{\theta} \sum_{i} log\, P\,(x_i|\theta) \\
&= \theta_{MLE}
\end{aligned}
$$

如果选择一个先验不是均匀的，比如高斯的，那么先验不再是一个常数了。概率可能很高或很低，但不会相同，因为这取决于分布的区域。
因此，可以清楚地得出结论：当先验是一致的时，MLE是MAP的特殊情况。

变量消元法示例1 - MRF （链状图模型） - 消除 A

MRF （链状图模型）： 求节点 E 的边缘概率

$$A \to B \to C \to D \to E$$

$$P(e) = \sum_d \sum_c \sum_b \sum_a P(a,b,c,d,e)$$

$$= \sum_d \sum_c \sum_b \sum_a P(a)P(b \mid a)P(c \mid b)P(d \mid c)P(e \mid d)$$

$$= \sum_d \sum_c \sum_b P(c \mid b)P(d \mid c)P(e \mid d) \sum_a P(a)P(b \mid a)$$

$$= \sum_d \sum_c \sum_b P(c \mid b)P(d \mid c)P(e \mid d)p(b)$$

$$P(e) \propto \sum_d \sum_c \sum_b \sum_a \tilde{P}(a,b,c,d,e)$$

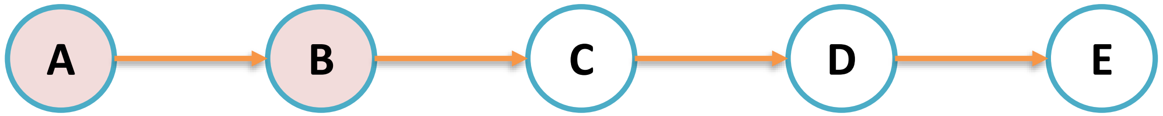
$$= \sum_d \sum_c \sum_b \sum_a \phi_1(a,b)\phi_2(b,c)\phi_3(c,d)\phi_4(d,e)$$

$$= \sum_d \sum_c \sum_b \phi_2(b,c)\phi_3(c,d)\phi_4(d,e) \sum_a \phi_1(a,b)$$

$$= \sum_d \sum_c \sum_b \phi_2(b,c)\phi_3(c,d)\phi_4(d,e)\tau_1(b)$$

变量消元法示例1 - MRF （链状图模型） - 消除 B

MRF （链状图模型）： 求节点 E 的边缘概率



$$P(e) = \sum_d \sum_c \sum_b P(c \mid b)P(d \mid c)P(e \mid d)p(b)$$

$$= \sum_d \sum_c P(d \mid c)P(e \mid d) \sum_b P(c \mid b)p(b)$$

$$= \sum_d \sum_c P(d \mid c)P(e \mid d)p(c)$$

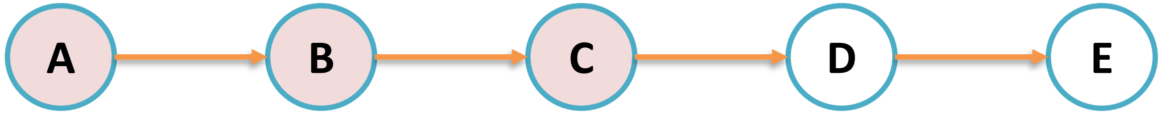
$$P(e) \propto \sum_d \sum_c \sum_b \phi_2(b,c)\phi_3(c,d)\phi_4(d,e)\tau_1(b)$$

$$= \sum_d \sum_c \phi_3(c,d)\phi_4(d,e) \sum_b \phi_2(b,c)\tau_1(b)$$

$$= \sum_d \sum_c \phi_3(c,d)\phi_4(d,e)\tau_2(c)$$

变量消元法示例1 - MRF （链状图模型） - 消除 C

MRF （链状图模型）： 求节点 E 的边缘概率



$$P(e) = \sum_{d} \sum_{c} P(d \mid c)P(e \mid d)p(c)$$

$$= \sum_{d} P(e \mid d) \sum_{c} P(d \mid c)p(c)$$

$$= \sum_{d} P(e \mid d)p(d)$$

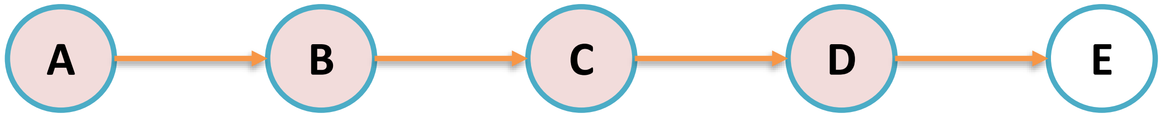
$$P(e) \propto \sum_{d} \sum_{c} \phi_3(c,d)\phi_4(d,e)\tau_2(c)$$

$$= \sum_{d} \phi_4(d,e) \sum_{c} \phi_3(c,d)\tau_2(c)$$

$$= \sum_{d} \phi_4(d,e)\tau_3(d)$$

变量消元法示例1 - MRF （链状图模型）- 消除 D

MRF （链状图模型）： 求节点 E 的边缘概率



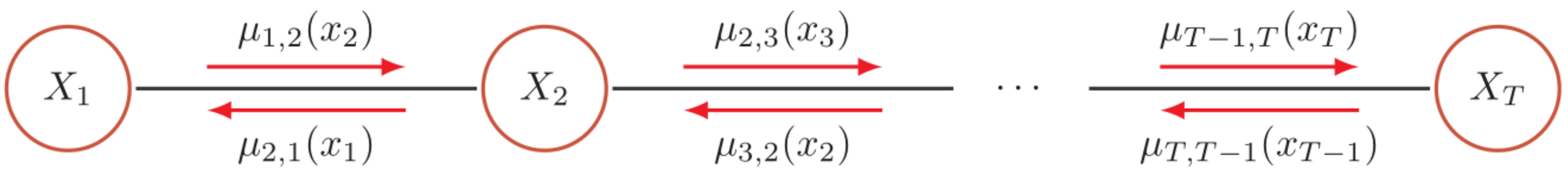
$$P(e) = \sum_{d} P(e \mid d)p(d)$$

$$= p(e)$$

$$P(e) \propto \sum_{d} \phi_4(d, e)\tau_3(d)$$

$$= \tau_4(e)$$

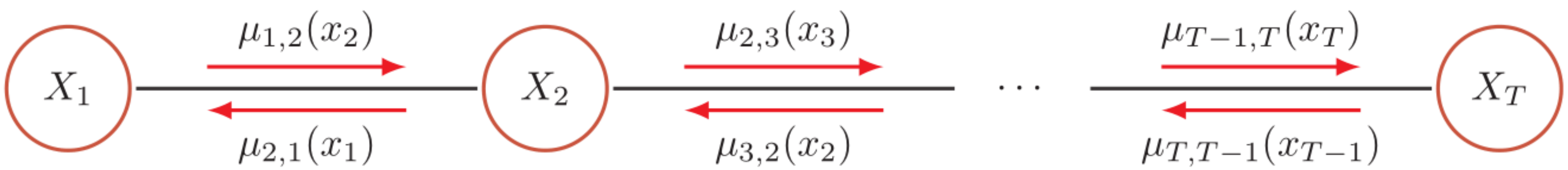
链上的消息传递



第 t 个变量的边际概率 $p(x_t)$ 为

$$\begin{aligned}
p(x_t) &= \sum_{x_1} \cdots \sum_{x_{t-1}} \sum_{x_{t+1}} \cdots \sum_{x_T} p(\mathbf{x}) \\
&= \frac{1}{Z} \sum_{x_1} \cdots \sum_{x_{t-1}} \sum_{x_{t+1}} \cdots \sum_{x_T} \prod_{t=1}^{T-1} \phi(x_t, x_{t+1}).
\end{aligned}$$

链上的消息传递



根据乘法的分配律，边际概率 $p(x_t)$ 可以通过下面方式进行计算：

$$
\begin{aligned}
p(x_t) &= \frac{1}{Z}\left(\sum_{x_1}\cdots\sum_{x_{t-1}}\prod_{j=1}^{t-1}\phi(x_j,x_{j+1})\right)\cdot\left(\sum_{x_{t+1}}\cdots\sum_{x_T}\prod_{j=t}^{T-1}\phi(x_j,x_{j+1})\right)\\
&= \frac{1}{Z}\left(\sum_{x_{t-1}}\phi(x_{t-1},x_t)\cdots\left(\sum_{x_2}\phi(x_2,x_3)\left(\sum_{x_1}\phi(x_1,x_2)\right)\right)\right)\cdot\\
&\quad\left(\sum_{x_{t+1}}\phi(x_t,x_{t+1})\cdots\left(\sum_{x_{T-1}}\phi(x_{T-2},x_{T-1})\left(\sum_{x_T}\phi(x_{T-1},x_T)\right)\right)\right)\\
&= \frac{1}{Z}\mu_{t-1,t}(x_t)\mu_{t+1,t}(x_t),
\end{aligned}
\tag{11.38}
$$

链式结构图模型的信念传播过程

1. 依次计算前向传递的消息 $\mu_{t-1,t}(x_t)$，$t=1,\cdots,T-1$；
2. 依次计算反向传递的消息 $\mu_{t+1,t}(x_t)$，$t=T-1,\cdots,1$；
3. 在任意节点 t 上计算配分函数 Z，

$$Z=\sum_{x_t}\mu_{t-1,t}(x_t)\mu_{t+1,t}(x_t).$$


Thank
You