

数学 机器学习 自然语言处理 机器视觉 概率图模型

如何理解马尔可夫随机场里因子的表达？

如何理解势函数和能量函数的物理意义？为什么定义成exp形态呢？为什么这样定义特征函数？为什么不像贝叶斯网络那样把因子直接定义成表格形态？为什么要...显示全部

关注问题

写回答

添加评论

分享

邀请回答

举报



斤木

自然语言理解/计算语言学/社会计算/书法

17 人赞同了该回答

谢邀。

现有答案都没答到点子上。简单梳理下涉及到的概念。

对于 n 个变量 $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ 构成的团集合为 C ，有：

$$P(\mathbf{X}) = \frac{1}{Z} \prod_{Q \in C} \psi_Q(\mathbf{x}_Q)$$

其中， $\psi_Q(\mathbf{x}_Q)$ 即为团 Q 的势函数（因子）， $Z = \sum_{\mathbf{x}} \prod_{Q \in C} \psi_Q(\mathbf{x}_Q)$

根据Hammersley-Clifford theorem，MRF与Gibbs distribution等价。可将该joint distribution改写成极大团的形式，有：

$$P(\mathbf{X}) = \frac{1}{Z^*} \prod_{Q \in C^*} \psi_Q(\mathbf{x}_Q)$$

其中 $Z^* = \sum_{\mathbf{x}} \prod_{Q \in C^*} \psi_Q(\mathbf{x}_Q)$

下面考察势函数。首先需要明确势函数规定是严格正。为何？

回到刚刚的joint distribution。为保证由势函数累乘后除以规范化因子得到的概率为正，我们规定势函数严格正。

为满足势函数严格正，常常选用指数势函数，即

$$\psi_Q(\mathbf{x}_Q) = e^{-H_Q(\mathbf{x}_Q)}$$

这样只需保证定义的能量函数 $H_Q(\mathbf{x}_Q)$ 为实值函数即可：

$$H_Q(\mathbf{x}_Q) = \sum_{u,v \in Q, u \neq v} \alpha_{u,v} t_{u,v}(x_u, x_v) + \sum_{v \in Q} \beta_v s_v(x_v)$$

这里第一项考虑结点间关系，第二项考虑结点本身。

为了统一形式，可将上式简写：

$$H_Q(\mathbf{x}_Q) = \sum_k^{K_Q} w_k^T \phi_k(\mathbf{x}_k)$$

这里， $\phi_k(\mathbf{x}_k)$ 为第 k 个特征函数， K_Q 为团 Q 中特征函数总数。

回代可得

Misplaced &

K 为整个马尔可夫随机场中特征函数总数。至此完成一个漂亮log-linear model的表示。

下面开始回答问题。

1) 为什么把马尔可夫随机场中的特征函数定义成exp形态?

答: 特征函数并没有定义成exp形态, 是势函数定义成了exp形态。原因已阐明。

2) 为什么马尔可夫随机场不像贝叶斯网络那样把因子定义成表格形态?

答: 三个字, 不够用。

BN有向, 因子形式为CPD (Conditional Probabilistic Distribution), 很自然被定义成表格形态;

MRF无向, 其因子表达只关注对称的affinities, 不关注causal relationship, 被定义为joint distribution。下面利用joint distribution定义证明MRF无法将因子定义成边上的表格形态。

证明:

一个有 n 个二值结点的全联接MRF, 其joint distribution需要 $f(n) = 2^n - 1$ 个parameters.

现在考虑另一种parameterization形式, 即将MRF的因子定义成边上的表格形态。如此, 每条边上有4个parameters, 该MRF共

计 $f'(n) = 4 \binom{n}{2}$ 个parameters。

易知当 $n \geq 1$ 时, $f'(n) > f(n)$ 。

故使用定义在边上的表格去表达MRF因子不可行。

证毕。

更具体来说, 边只表达了pairwise interactions, 无法表达更多相关变量的关联。基于最大团的势函数paramization可以解决这个问题。

3) 如何理解势函数和能量函数的物理意义?

答: Markov Network Model最早的一种叫伊辛模型, 用来在统计物理学中考察原子间相互作用。每个原子为一个变量

$X_i \in \{-1, +1\}$, 用来表征原子旋转方向。能量函数的定义很简单:

$$-\sum_{i < j} w_{ij} x_i x_j - \sum_{i \in Q} v_i x_i$$

当 $w_{ij} > 0$ 时为铁磁自旋, 反之, 反铁磁自旋。

玻尔兹曼机则令 $X_i \in \{0, 1\}$, 能量函数定义与伊辛模型一致。易知, 对于 $\{X_i, X_j\}$ 当且仅当 $X_i = X_j = 1$ 时有能量贡献。玻尔兹曼机原理与神经元的activation model类似, 使用基于邻接节点加权的sigmoid函数做为概率分布。在此不再详述。

END.

发布于 2016-06-09

▲ 17 ▼ 6 条评论 分享 收藏 感谢 ...

收起 ^



bhuztez

正在找工作 ...

3 人赞同了该回答

这个一开始我也很困惑, 在网上找到的很多教程在这里都是糊弄过去的。直到我发现了有一个Hammersley-Clifford theorem, 把这个定理推一遍, 这个问题就明白了。

比如说, 假设有事件 $X_1 X_2$, 你想计算出 $p(X_1) p(X_2) p(X_1|X_2) p(X_2|X_1) \dots$ (请自行脑补not X_1 , not X_2)

假设有事件 $X_1 X_2 X_3$, 你想计算出 $p(X_1) p(X_2) p(X_3) p(X_1|X_2X_3) p(X_2|X_1X_3) p(X_3|X_1X_2) \dots$

你并不知道, $X_1 X_2 \dots X_n$ 之间的关系, 你希望能有一个神奇的模型, 只需要比较少的参数, 就能计算所有关于 $X_1 X_2 \dots X_n$ 之间能定义出来的概率。

Hammersley-Clifford theorem说的是, 假如这个 $X_1 X_2 \dots X_n$ 符合Markov property, 那么就可以用Gibbs XXX来表示, 或者假如你发现可以用Gibbs XXX来表示, 那么他们就符合Markov property, 两者是等价的。