受限玻尔兹曼机

玻尔兹曼机是一种存在隐节点的无向图模型。在图模型中最简单的是朴素贝叶斯模型（朴素贝叶斯假设），引入单个隐变量后，发展出了 GMM，如果单个隐变量变成序列的隐变量，就得到了状态空间模型（引入齐次马尔可夫假设和观测独立假设就有HMM，Kalman Filter，Particle Filter），为了引入观测变量之间的关联，引入了一种最大熵模型-MEMM，为了克服 MEMM 中的局域问题，又引入了CRF，CRF 是一个无向图，其中，破坏了齐次马尔可夫假设，如果隐变量是一个链式结构，那么又叫线性链 CRF。

在无向图的基础上，引入隐变量得到了玻尔兹曼机，这个图模型的概率密度函数是一个指数族分布。对隐变量和观测变量作出一定的限制，就得到了受限玻尔兹曼机（RBM）。

我们看到，不同的概率图模型对下面几个特点作出假设：

1. 方向-边的性质
2. 离散/连续/混合-点的性质
3. 条件独立性-边的性质
4. 隐变量-节点的性质
5. 指数族-结构特点

将观测变量和隐变量分别记为 $v,h,h=\{h_1,\cdots,h_m\},v=\{v_1,\cdots,v_n\}$。我们知道，无向图根据最大团的分解，可以写为玻尔兹曼分布的形式 $p(x)=\frac{1}{Z}\prod\limits_{i=1}^K\psi_i(x_{ci})=\frac{1}{Z}\exp(-\sum\limits_{i=1}^KE(x_{ci}))$，这也是一个指数族分布。

一个玻尔兹曼机存在一系列的问题，在其推断任务中，想要精确推断，是无法进行的，想要近似推断，计算量过大。为了解决这个问题，一种简化的玻尔兹曼机-受限玻尔兹曼机作出了假设，所有隐变量内部以及观测变量内部没有连接，只在隐变量和观测变量之间有连接，这样一来：

$$
p(x)=p(h,v)=\frac{1}{Z}\exp(-E(v,h)) \tag{1}
$$

其中能量函数 $E(v,h)$ 可以写出三个部分，包括与节点集合相关的两项以及与边 w 相关的一项，记为：

$$
E(v,h)=-(h^Twv+\alpha^Tv+\beta^Th) \tag{2}
$$

所以：

$$
p(x)=\frac{1}{Z}\exp(h^Twv)\exp(\alpha^Tv)\exp(\beta^Th)=\frac{1}{Z}\prod\limits_{i=1}^m\prod\limits_{j=1}^n\exp(h_iw_{ij}v_j)\prod\limits_{j=1}^n\exp(\alpha_jv_j)\prod\limits_{i=1}^m\exp(\beta_ih_i) \tag{3}
$$

上面这个式子也和 RBM 的因子图一一对应。

推断

推断任务包括求后验概率 $p(v|h),p(h|v)$ 以及求边缘概率 $p(v)$。

$p(h|v)$

对于一个无向图，满足局域的 Markov 性质，即
$p(h_1|h-\{h_1\},v)=p(h_1|Neighbour(h_1))=p(h_1|v)$。我们可以得到：

$$p(h|v)=\prod_{i=1}^m p(h_i|v) \tag{4}$$

考虑 Binary RBM，所有的隐变量只有两个取值 $0,1$：

$$p(h_l=1|v)=\frac{p(h_l=1,h_{-l},v)}{p(h_{-l},v)}=\frac{p(h_l=1,h_{-l},v)}{p(h_l=1,h_{-l},v)+p(h_l=0,h_{-l},v)} \tag{5}$$

将能量函数写成和 l 相关或不相关的两项：

$$E(v,h)=-(\sum_{i=1,i\neq l}^m\sum_{j=1}^n h_iw_{ij}v_j+h_l\sum_{j=1}^n w_{lj}v_j+\sum_{j=1}^n\alpha_jv_j+\sum_{i=1,i\neq l}^m\beta_ih_i+\beta_lh_l) \tag{6}$$

定义：$h_lH_l(v)=h_l\sum\limits_{j=1}^n w_{lj}v_j+\beta_lh_l,\overline{H}(h_{-l},v)=\sum\limits_{i=1,i\neq l}^m\sum\limits_{j=1}^n h_iw_{ij}v_j+\sum\limits_{j=1}^n\alpha_jv_j+\sum\limits_{i=1,i\neq l}^m\beta_ih_i$。

代入，有：

$$p(h_l=1|v)=\frac{\exp(H_l(v)+\overline{H}(h_{-l},v))}{\exp(H_l(v)+\overline{H}(h_{-l},v))+\exp(\overline{H}(h_{-l},v))}=\frac{1}{1+\exp(-H_l(v))}=\sigma(H_l(v)) \tag{7}$$

于是就得到了后验概率。对于 v 的后验是对称的，所以类似的可以求解。

$p(v)$

$$\begin{aligned}p(v)&=\sum_h p(h,v)=\sum_h\frac{1}{Z}\exp(h^Twv+\alpha^Tv+\beta^Th)\\&=\exp(\alpha^Tv)\frac{1}{Z}\sum_{h_1}\exp(h_1w_1v+\beta_1h_1)\cdots\sum_{h_m}\exp(h_mw_mv+\beta_mh_m)\\&=\exp(\alpha^Tv)\frac{1}{Z}(1+\exp(w_1v+\beta_1))\cdots(1+\exp(w_mv+\beta_m))\\&=\frac{1}{Z}\exp(\alpha^Tv+\sum_{i=1}^m\log(1+\exp(w_iv+\beta_i)))\end{aligned} \tag{8}$$

其中，$\log(1+\exp(x))$ 叫做 Softplus 函数。