LBE-LF Vs. SAR-EM

LBE-LF

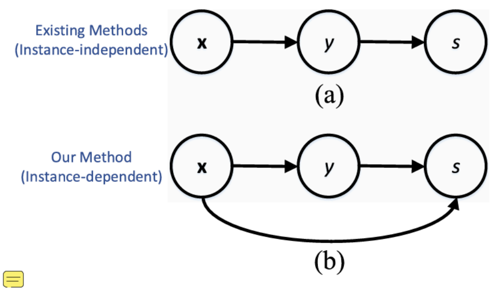
1. **常数标注概率 η 的情况**：
    - 如果 η 是常数并且与 x 无关，则 $\eta = P(s=1|y=1)$，可以直接通过数据估计出来。公式为：

$$\eta = \frac{P(s=1, y=1)}{P(y=1)} = \frac{P(s=1)}{P(y=1)}$$

    - 其中 $P(s=1)$ 和 $P(y=1)$ 可以直接从数据中估计。
2. **实例相关标注概率 $\eta(x)$ 的情况**：
    - 当 η 是与 x 相关的函数 $\eta(x)$ 时，情况变得复杂。在这种情况下，我们有：

$$\eta(x) = P(s=1|y=1,x) = \frac{P(s=1, y=1|x)}{P(y=1|x)} = \frac{P(s=1|x)}{P(y=1|x)}$$

    - 这表明 $\eta(x)$ 和类后验概率 $P(y=1|x)$ 是共存的。因此，为了估计 $\eta(x)$ 和 $P(y=1|x)$，需要找到一种新的方法来联合估计这两个概率。



由图(b)得到：

$P(y,s|x) = P(y|x)P(s|y,x)$

带入PU数据集后得到：

$P(y,s|x) = \prod_{i=1}^{n} P(y_i, s_i|x_i) = \prod_{i=1}^{n} P(y_i|x_i)P(s_i|y_i,x_i)$

构造分类器：

$h(x;\theta_1) = P(y=1|x) = (1+\exp(-\theta_1^\top x))^{-1}$

$\eta(x;\theta_2) = P(s=1|y=1,x) = (1+\exp(-\theta_2^\top x))^{-1}$

最大化似然函数：

$\arg\max_{\theta}\prod_{i=1}^{n}\sum_{y_i}P(s_i,y_i|x_i;\theta).$

对数似然：

$\arg\max_{\theta}\mathcal{L}(\theta)=\sum_{i=1}^{n}\log\sum_{y_i}P(s_i,y_i|x_i;\theta).$

因为存在未知变量y_i，通过EM算法求解。

LBE-LF E步

在E步中，希望计算每个数据点 i 的潜在变量 y_i 的后验概率 $\tilde{P}(y_i)=P(y_i|x_i,s_i)$。这是因为我们无法直接观测到 y_i，但可以通过观测到的变量 x_i 和 s_i 来推断 y_i 的概率。

联合概率的分解

根据条件独立性和贝叶斯定理，我们可以将联合概率 $P(y_i,s_i|x_i)$ 分解为：

$P(y_i,s_i|x_i)=P(s_i|x_i)P(y_i|x_i,s_i)$

这里我们主要关注 $\tilde{P}(y_i)$，即 $P(y_i|x_i,s_i)$。

通过贝叶斯定理，将 $P(y_i|x_i,s_i)$ 表示为：

$P(y_i|x_i,s_i)=\frac{P(s_i|y_i,x_i)P(y_i|x_i)}{P(s_i|x_i)}$

计算 $P(s_i|x_i)$ 是不必要的，因为它只是一个归一化常数。

使用比例表示法：

$P(y_i|x_i,s_i)\propto P(s_i|y_i,x_i)P(y_i|x_i)$

即：

$\tilde{P}(y_i)=P(y_i|x_i,s_i)\propto P(s_i|y_i,x_i)P(y_i|x_i)$

其中：

- $P(s_i|y_i,x_i)$：表示给定 y_i 和 x_i 后，观察到 s_i 的概率。**通过$\eta(x)$得到。**
- $P(y_i|x_i)$：表示给定 x_i 后，y_i 的先验概率。**通过$h(x)$得到。**

LBE-LF M步

期望对数似然函数的定义如下：

$\mathcal{J}(\theta) = \sum_i \mathbb{E}_{\hat{P}(y_i)}[\log P(y_i, s_i|x_i;\theta)]$

这个函数表示的是在E步计算出的隐变量 y_i 的期望值基础上，对参数 θ 的期望对数似然。

在M步中，我们的目标是最大化 $\mathcal{J}(\theta)$，即：

$\max_\theta \mathcal{J}(\theta)$

将期望对数似然函数 $\mathcal{J}(\theta)$ 展开，可以得到：

$\mathcal{J}(\theta) = \sum_i \mathbb{E}_{\hat{P}(y_i)}[\log P(y_i|x_i;\theta_1) + \log P(s_i|y_i, x_i;\theta_2)]$

对 θ_1 的梯度
梯度 $\nabla_{\theta_1}\mathcal{J}(\theta)$ ：

$\nabla_{\theta_1}\mathcal{J}(\theta) = \sum_i \nabla_{\theta_1}\mathbb{E}_{\hat{P}(y_i)}[\log P(y_i|x_i;\theta_1)]$

展开并化简得到：

$\nabla_{\theta_1}\mathcal{J}(\theta) = \sum_i \sum_{y_i} \hat{P}(y_i)\nabla_{\theta_1} \log P(y_i|x_i;\theta_1)$

进一步化简：

$\nabla_{\theta_1}\mathcal{J}(\theta) = \sum_i \hat{P}(y_i = 1)\nabla_{\theta_1} \log P(y_i = 1|x_i;\theta_1) + \hat{P}(y_i = 0)\nabla_{\theta_1} \log P(y_i =$

对 θ_2 的梯度
梯度 $\nabla_{\theta_2}\mathcal{J}(\theta)$ ：

$\nabla_{\theta_2}\mathcal{J}(\theta) = \sum_i \nabla_{\theta_2}\mathbb{E}_{\hat{P}(y_i)}[\log P(s_i|y_i, x_i;\theta_2)]$

展开并化简得到：

$\nabla_{\theta_2}\mathcal{J}(\theta) = \sum_i \sum_{y_i} \hat{P}(y_i)\nabla_{\theta_2} \log P(s_i|y_i, x_i;\theta_2)$

进一步化简：

$\nabla_{\theta_2}\mathcal{J}(\theta) = \sum_i \hat{P}(y_i = 1)\nabla_{\theta_2} \log P(s_i|y_i = 1, x_i;\theta_2) + \hat{P}(y_i = 0)\nabla_{\theta_2} \log P(s_i$

对于LBE-LF来说，

使用Adam优化器进行梯度更新，具体步骤如下：

1. 计算梯度 $\nabla_{\theta_1}\mathcal{J}(\theta)$ 和 $\nabla_{\theta_2}\mathcal{J}(\theta)$。
2. 使用Adam优化器根据计算出的梯度更新参数。

SAR-EM

。