RankNet To LambdaRank

RankNet是一种基于PairWise方式的排序算法。

在RankNet算法中将模型打分转换为概率，以两个文档u_i和u_j为例，默认对两个Doc的打分分别为s_i和s_j，则u_i应该排在u_j前面的概率可以表示为:

$$P_{ij} = P(u_i \rhd u_j) = \frac{1}{1+e^{-\sigma(s_i - s_j)}}$$

由概率公式可以看出，模型对文档u_i打分越高，则$s_i - s_j$的值就越大，则分母的值就越小，概率也就越接近于1。反之，分母的值就越大，概率就越小。

两个文档u_i和u_j真是的排序则由其标注值决定，即标注值大的Doc排在前面，对真实排序的概率值我们可以做如下定义:

$$\bar{P_{ij}} = 1 \text{，如果} u_i \text{的} Label \text{值大于} u_j \text{的} Label \text{值}$$

$$\bar{P_{ij}} = 0 \text{，如果} u_i \text{的} Label \text{值小于} u_j \text{的} Label \text{值}$$

$$\bar{P_{ij}} = 0.5 \text{，如果} u_i \text{的} Label \text{值与} u_j \text{的} Label \text{值相同}$$

则我们可以得到形式化的公式表达为:

$$\bar{P_{ij}} = \frac{1}{2} * (1 + S_{ij})$$

则S_{ij}的取值为当u_i的Label值大于u_j的值时为1，当u_i小于u_j的值时为-1，当两DocLabel值相同时为0。

根据模型得分得到的概率与真实的概率值，我们可以构建交叉熵，以此作为模型的损失函数:

$$C = -1 * \bar{P_{ij}} log P_{ij} - (1 - \bar{P_{ij}}) log(1 - P_{ij})$$

将$\bar{P_{ij}}$和P_{ij}分别带入上面的公式，可以得到:

$$C = -[\frac{1}{2} * (1 + S_{ij})] * log(\frac{1}{1+e^{-\sigma(s_i - s_j)}}) - [1 - \frac{1}{2} * (1 + S_{ij})] * [1 - \frac{1}{1+e^{-\sigma(s_i - s_j)}}]$$

下面我们逐步将公式进行化简:

$$C = -\frac{1}{2}(1 + S_{ij}) * [log1 - log(1 + e^{-\sigma(s_i - s_j)})] - [1 - \frac{1}{2} * (1 + S_{ij})] * log(\frac{e^{-\sigma(s_i - s_j)}}{1+e^{-\sigma(s_i - s_j)}})$$

$$C = \frac{1}{2}(1+S_{ij}) * log(1+e^{-\sigma(s_i-s_j)}) + \frac{1}{2}(S_{ij}-1)[log(e^{-\sigma(s_i-s_j)}) - log(1+e^{-\sigma(s_i-s_j)})]$$

$$C = log(1+e^{-\sigma(s_i-s_j)}) + \frac{1}{2}(1-S_{ij})\sigma(s_i-s_j)$$

当$S_{ij}=1$时,

$$C = log(1+e^{-\sigma(s_i-s_j)})$$

当$S_{ij}=-1$时,

$$C = log(1+e^{-\sigma(s_i-s_j)}) + \sigma(s_i-s_j)$$

$$C = log((1+e^{-\sigma(s_i-s_j)}) * e^{\sigma(s_i-s_j)})$$

$$C = log(e^{\sigma(s_i-s_j)}+1)$$

$$C = log(e^{-\sigma(s_j-s_i)}+1)$$

有损失函数的公式可以看出，在两个文档标签不同时，即使模型对两个文档打分相同，损失函数值仍为$log2$，这样就会使得模型趋于将标签不同的文档打分区分开。

利用上面得到的损失函数分别计算s_i和s_j的梯度，可以得到:

$$\frac{\partial C}{\partial s_i} = \sigma[\frac{1}{2}(1-S_{ij}) - \frac{1}{1+e^{-\sigma(s_i-s_j)}}]$$

$$\frac{\partial C}{\partial s_j} = \sigma[1+e^{-\sigma(s_i-s_j)} - \frac{1}{2}(1-S_ij)]$$

可以看出

$$\frac{\partial C}{\partial s_i} = -\frac{\partial C}{\partial s_j}$$