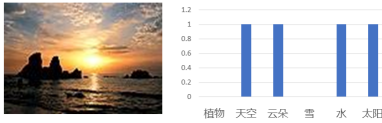


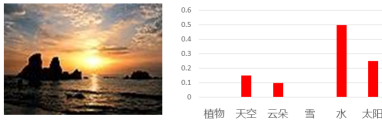
第一部分：研究动机

逻辑标记和标记分布是刻画标记多义性 (示例能被多个标记同时描述) 的两种常用标记形式

逻辑标记: 通过逻辑值来指示某个标记能否描述该示例



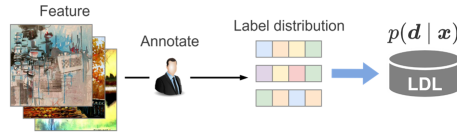
标记分布: 通过一个实值向量来刻画相关标记之间的相对重要程度



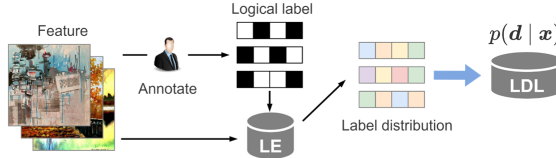
标记分布能够提供更丰富的标记多义性信息

为什么要预测标记分布

一种方法是LDL (标记分布学习): 要求专家以标记分布标注示例, 并从该训练样例中学习一个标记分布预测器, 从而可用以输出示例的标记分布



另一种方法是LE (标记增强): 要求专家以逻辑标记标注示例, 并设计算法来恢复训练数据的标记分布, 再从该训练样例中学习一个标记分布预测器



优缺点分析:

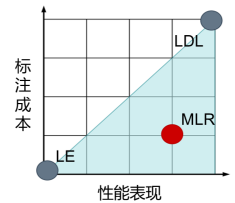
- LDL: 长于预测性能较好, 短于标注成本高且标注难度大
- LE: 长于标注成本低, 短于性能没有良好保证

现有有哪些方法能预测标记分布, 以及它们存在的问题

MLR (多标记排序) 是缓解上述问题的不错选择:

- 标注成本和难度均可接受
- 性能表现有良好的保障

MLR、LDL、LE的关系:



因此我们研究从MLR预测标记分布

- 我们从理论上研究了MLR和标记分布之间的关系
- 我们从算法上研究了如何从MLR标注的样例中预测标记分布

我们的研究

第二部分：理论研究

Definition 1 Suppose that an instance is annotated with an annotation τ ; δ and δ^* are implicit and explicit margins, respectively; S_τ^i and S_τ^e are the implicit and explicit ranges, respectively. Then the expected approximation error of τ to the internal label description vector is

$$e_{\tau}^{\delta, \delta^*} = \int_{z \in S_\tau^i} \int_{z \in S_\tau^e} \frac{1}{V_\tau^i V_\tau^e} \|z - \tau\|_2^2 dz, \quad V_\tau^i = \int_{z \in S_\tau^i} dz, \quad V_\tau^e = \int_{z \in S_\tau^e} dz. \quad (1)$$

Theorem 1 If an instance is annotated by a multi-label ranking σ , m is the number of relevant labels, δ and δ^* are the implicit and explicit margins, respectively, then the EAE of σ is

$$e_{\sigma}^{\delta, \delta^*} = \frac{m}{6(m+1)} \left((m+1)^2 (\delta^2 + \delta^2) - 2m(\delta + \delta^*) - (4m+2)\delta\delta^* + 2 \right). \quad (2)$$

Corollary 1 If an instance is annotated by a multi-label ranking σ , m is the number of relevant labels, the explicit margin δ^* minimizing the EAE of σ is $\delta^* = (2m+1)\delta + m(m+1)^{-2}$.

Corollary 2 If an instance is annotated by a multi-label ranking σ , m is the number of relevant labels, $0 \leq \delta \leq m^{-1}$, $m(m+1)^{-2} \leq \delta^* \leq m^{-1}$, then the EAE of σ is bounded by:

$$0 \leq e_{\sigma}^{\delta, \delta^*} \leq \frac{m(m^2 + 4m + 2)}{6(m+1)^3} < \frac{1}{5}. \quad (3)$$

Corollary 3 If an instance is annotated by a multi-label ranking σ , m is the number of relevant labels, δ and δ^* are uniform over $[0, m^{-1}]$ and $[m(m+1)^{-2}, m^{-1}]$, respectively, then we have:

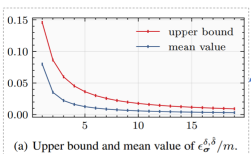
$$\mathbb{E} [e_{\sigma}^{\delta, \delta^*}] = \frac{2m^4 + 8m^3 + 8m^2 + 4m + 1}{36m(m+1)^3}. \quad (4)$$

Theorem 2 If an instance is annotated by a logical label vector l , m is the number of relevant labels, δ and δ^* are the implicit and explicit margins, respectively, then the EAE of l is

$$e_l^{\delta, \delta^*} = \frac{m}{6} (2\delta^2 + 2\delta^2 - \delta - \delta - 3\delta\delta^* + 1). \quad (5)$$

Corollary 4 Suppose that e_l^{δ, δ^*} and $e_{\sigma}^{\delta, \delta^*}$ are the EAE of the logical label vector l and the EAE of the multi-label ranking σ , respectively, we have the following inequality holds for $m \geq 3$:

$$e_l^{\delta, \delta^*} - e_{\sigma}^{\delta, \delta^*} \geq \frac{7m}{48} (\delta^2 - 2\delta) + \frac{m(m-1)(7m^2 + 20m + 9)}{48(m+1)^3} > \frac{7m^5 - m^4 - 46m^3 - 30m^2 + 7m + 7}{48m(m+1)^3} > 0. \quad (6)$$



MLR的期望近似误差与相关标记个数的关系

MLR相比逻辑标记的优势与相关标记个数的关系

定义了期望近似误差以衡量某种标注方式对真实标记分布的近似能力

MLR的期望近似误差

在什么范围内估计标记分布能够最小化MLR的期望近似误差

MLR的期望近似误差值的上下界是多少

MLR的期望近似误差值的平均值是多少

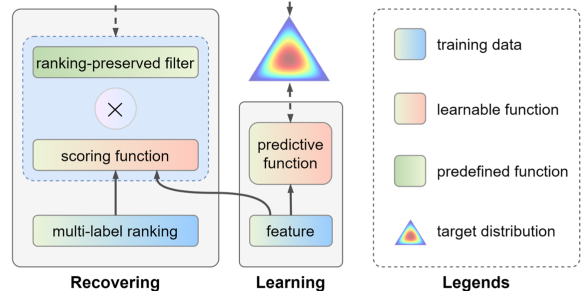
逻辑标记的期望近似误差是什么

就期望近似误差而MLR比逻辑标记好多少

第三部分：算法研究

- 我们提出了一个从MLR恢复、学习并预测标记分布的通用学习框架, 名为DRAM
- 首先通过一个形式预定义且参数自适应学习的评分函数为标记分布构造目标分布
- 进而通过条件狄利克雷混合物来表示预测分布
- 最后通过最小化预测分布相对于目标分布的交叉熵来学习模型参数

简介



DRAM学习框架的示意图

DRAM学习框架的细节

目标分布 $p^*(d) = \frac{1}{Z_p} \phi(d; \theta) \int_0^\infty \mathbb{I}(td \in S_\tau^e) dt$

预测分布 $p(d|x) = \sum_{k=1}^K f_k(x; \alpha) \text{Dir}(d | f(x; \beta^k))$

学习目标 $\mathbb{E}_{p(d)} [\ln p(d|x)] \approx \sum_{i=1}^L \frac{\phi(d^{(i)}; \theta) \ln p(d^{(i)}|x)}{\sum_{j=1}^L \phi(d^{(j)}; \theta)}$, $d^{(i)} = \frac{1}{Z_p} z^{(i)}$, $z^{(i)} \sim \text{Uni}(z | S_\tau^e)$

学习算法 EM算法

E-step $\gamma_{nk}^{(i)} \triangleq q_k(d^{(i)} = k) = \frac{f_k(x; \alpha) \text{Dir}(d^{(i)} | f(x; \beta^k))}{\sum_{j=1}^K f_j(x; \alpha) \text{Dir}(d^{(i)} | f(x; \beta^j))}$, $k \in [K]$

M-step $\arg \max_{\alpha, \beta^1, \dots, \beta^K} \sum_{i=1}^L \sum_{k=1}^K \frac{\phi(d^{(i)}; \theta)}{\sum_{j=1}^L \phi(d^{(j)}; \theta)} \left(\sum_{k=1}^K \gamma_{nk}^{(i)} \ln \frac{f_k(x; \alpha) \text{Dir}(d^{(i)} | f(x; \beta^k))}{\gamma_{nk}^{(i)}} \right)$

预测结果 $d^* = \sum_{k=1}^K \frac{1}{Z_k} f_k(x; \alpha) f(x; \beta^k)$, $Z_k = \sum_{i=1}^M f_i(x; \beta^k)$

- 具有成本效益: 它在多标记排名标注的示例上进行训练, 并取得了与LDL算法 (其需要昂贵的标记分布标注) 相当的性能
- 灵活的: 它允许用户轻松地通过评分函数来编码其先验知识
- 端到端: 它将恢复和学习标签分布的过程整合到一个学习准则中, 而非分别执行

DRAM学习框架的优点

第四部分：实验结果

对比算法的平均排名

Method	Cheb	Canber	Cosine	Rho
DRAM-LN	2.00	1.14	2.00	1.43
DT+GL+DM	5.00	4.14	4.29	2.29
DT+GL+SA	4.29	3.71	3.86	4.71
DT+VI+DM	6.43	5.43	6.29	4.29
DT+VI+SA	4.86	5.43	5.43	6.00
GT+DM	2.43	3.43	2.43	4.43
GT+SA	3.00	4.71	3.71	4.86

- 我们的方法: DRAM框架+线性学习器+无信息评分函数
- 在MLR标注的训练集上进行训练的算法
- 直接用真实的标记分布数据来训练的LDL算法