一种基于多示例多标记学习的新标记学习方法

朱越^{1,2}, 姜远^{1,2*}, 周志华^{1,2}

1. 计算机软件新技术国家重点实验室 (南京大学), 南京 210023

2. 软件新技术与产业化协同创新中心, 南京 210023

* 通信作者. E-mail: jiangy@lamda.nju.edu.cn

收稿日期: 2018-05-31; 接受日期: 2018-08-14; 网络出版日期: 2018-12-13

国家自然科学基金 (批准号: 61673201, 61333014) 资助项目

摘要 多标记学习是一种应用非常广泛的学习范式, 其中, 一个对象可能同时与多个标记相关联. 传统的多标记学习研究多假设训练数据中观察到的标记分布与测试数据的真实标记分布一致. 但在实际应用中, 训练数据中可能存在一些从未被标注出的新标记. 在预测时, 不仅希望能够在目标标记集合 (已知标记) 上取得好的性能, 还要求能够检测出样本是否存在新标记. 针对这种多标记新标记学习问题, 本文提出了一种端到端的多视图多示例多标记学习方法 EM3NL. 该方法基于卷积神经网络产生多示例包, 并通过最小化包上观察标记的错分损失和对新标记预测值排序损失的惩罚以及对多视图预测不一致的惩罚同时学习图像, 文本两个视图的特征表示以及已知标记和新标记的预测函数. 在大规模图片-文本真实数据集上验证了 EM3NL 在已知标记学习和新标记检测任务上的有效性.

关键词 多标记新标记学习, 多示例多标记学习, 多视图学习, 深度学习

1 引言

多标记学习^[1~6]是一种非常重要的学习范式. 其中, 一个对象可能同时与多个标记相关联. 传统多标记学习的研究多假设训练数据中观察到的标记分布与测试数据的真实标记分布一致. 但在实际应用中, 训练数据中可能存在一些从未被标注出的新标记, 而对测试数据需要预测这些新标记. 例如图片标注任务要求标注图片是否有建筑和车辆, 那么标注者可能只关心建筑和车辆的语义, 而不会对图片出现的天空、树木等其他语义进行标注. 如何预测样本中是否有天空、树木这样在训练集中存在相关语义但从未被标注出的新标记呢?

直接在多标记学习的框架下检测新标记是非常困难的. 现有一类多标记弱标记学习工作^[7]考虑标记部分缺失的情况: 某些样本上缺失的标记在另一些样本上可以观测到. 在这种情况下, 可以利用低秩结构特性, 通过优化矩阵核范数^[8,9]或者最小化重构误差进行低秩矩阵分解^[10,11], 从而对标记矩

引用格式: 朱越, 姜远, 周志华. 一种基于多示例多标记学习的新标记学习方法. 中国科学: 信息科学, 2018, 48: 1670–1680, doi: 10.1360/N112018-00143
Zhu Y, Jiang Y, Zhou Z H. Multi-instance multi-label new label learning (in Chinese). Sci Sin Inform, 2018, 48: 1670–1680, doi: 10.1360/N112018-00143

阵进行补全. 但是在多标记新标记学习中, 由于在整个训练集中新标记对应的语义均未被标注, 即真实标记矩阵整列缺失, 无法利用低秩矩阵补全.

如果直接采用异常检测的方法判断是否有新标记出现也可能失败^[12, 13]: (1) 一个样本可能不仅仅有新标记, 还可能同时与多个已知标记相关联, 因而很难从特征上将同时有新标记和已知标记的样本和那些只有相同已知标记的样本区分开; (2) 即使一个样本被检测算法认为是一个异常样本, 可能由于该样本真的与新标记相关联, 也有可能是因为出现了训练集中没有出现过的已知标记的组合.

因此, 为了解决多标记新标记学习问题, 考虑应用表示能力更强的学习范式, 主要包括结构和特征两个层面. 从结构上, 学习框架采用多示例多标记学习框架^[14~19]. 其中, 每个对象由一个多示例包表示, 包中的每个示例对应某一个语义概念; 包的标记即为包中样本标记的并集 (在训练过程中, 只能观察到包的标记, 而具体的包中示例的标记不可见). 以图像标注任务为例, 一个多示例包对应的是一幅图像, 而包中的示例则对应图像分割成的各个小块. 直觉上看, 在多示例多标记学习框架下, 只要能够预测各个示例的标记, 如果其中某个示例不属于已知标记, 则该示例以及它所在的包存在新标记. 但是通过穷举方式搜索所有样本-标记组合的复杂度关于包中示例的个数和标记的个数均为指数级增长. 因此, 实际应用多示例多标记学习框架需要精心设计高效算法. 大部分现有的多示例多标记学习算法不能处理新标记学习问题, 且由于较高的计算复杂度, 很难处理大规模数据. 从利用表示能力更强的特征角度, 希望能够利用深度学习的分层卷积网络结构, 自动学习出好的特征表示. 之前的深度学习的研究主要集中在监督学习、半监督学习和无监督学习^[20~22], 而对于这种特殊的整列标记缺失的弱监督学习问题^[7], 很少见相关工作.

进一步地, 除了利用多示例多标记学习框架, 深度神经网络获得更强地表示能力之外, 开放的网络环境使得学习器能够获取很多额外的信息, 作为原始信息的补充. 例如对于图片标注任务, 尽管训练集标记有缺失, 但是可以获得对图片的文本描述. 相较于原始图片, 这些文本描述中有更加简明凝练的语义信息, 甚至可能包含了缺失标注所对应的语义. 把图片作为原始视图, 获取相应的文本描述作为增广视图, 通过多视图学习框架^[23~27], 利用辅助信息进一步提升模型性能.

为解决多标记新标记学习问题, 本文提出了一种端到端的多视图多示例标记新标记学习方法 EM3NL (end-to-end multi-view multi-instance multi-label learning with new labels), 将深度神经网络与多视图、多示例多标记学习结合, 输入原始图片和文本, 输出已知标记和新标记的预测. 本文主要贡献如下:

- (1) 将深度神经网络与多视图、多示例多标记学习结合, 设计了一种直接输入图像-文本视图, 输出已知标记和新标记预测的端到端统一解决方案;
- (2) 提出了一种新型的多示例包级的损失函数, 以及对新标记预测的正则化项;
- (3) 改变了 DeepMIML (deep multi-instance multi-label network) 工作^[28] 中深度模型的上层网络结构, 使其适用于新标记发现.

本文的组织结构如下: 第 2 节介绍了与本工作相关的已有研究工作; 第 3 节给出了问题的形式化; 第 4 节详细介绍了 EM3NL 方法; 第 5 节汇报了在 MS-COCO 数据上的实验结果; 最后, 对全文进行总结.

2 相关工作

多示例多标记学习最早由 Zhou 等^[14, 15] 提出, 其中, 每个对象是由多个示例组成的表示, 且可能同时与多个标记相关联. 文献 [16] 首先将 KNN 方法推广到多示例多标记学习设置; 文献 [17] 提出了

一种优化排序损失的 MIML (multi-instance multi-label) 学习方法; 文献 [18] 提出了一种快速 MIML 算法, 并提出了标记子语义概念. 这些工作大多假设所有的包属于固定的目标标记集合, 而不考虑新标记的存在, 且特征需要事先给定.

最近, Feng 和 Zhou^[28] 提出了 DeepMIML 方法, 设计了一种基于深度神经网络的多示例多标记学习方法. 该工作将两个表示能力强大的学习框架相结合, 能够自主提取特征, 自动发现关键区域, 且能处理较大规模的数据. 该工作主要应用于标记完整的训练数据, 并假设训练集与测试集标记分布一致, 而不考虑数据中可能存在的新语义. 本工作借鉴了 DeepMIML 将多示例多标记学习与深度神经网络结合的思想, 针对新标记检测任务, 将 DeepMIML 网络结构进行了扩展.

多标记新标记检测问题有两个相关工作, 分别是 MIMLNC (multi-instance multi-label learning in the presence of novel class instances)^[29] 和 DMNL (discovering multiple novel labels)^[12]. 这两个工作均基于多示例多标记学习框架, 发现训练集中隐藏的新语义, 并预测测试集样本是否包含这些新语义. MIMLNC 方法通过动态规划似然估计, 在每个多示例包上都考虑两种情况: 包含/没有新标记, 然后最大化似然估计, 这样迭代优化模型, 最终在新标记和已知标记上均取得了很好的效果. 但存在的问题是这种训练模型的方法, 计算开销关于标记数目指数级增长, 且每一轮模型更新都需要遍历所有样本. 该局限性使得 MIMLNC 难以处理大规模的数据.

DMNL 方法对 MIMLNC 做出了改进, 提出了一种包损失函数和样本聚类结构正则化项, 使得其计算复杂度关于标记是线性的. 但是 DMNL 每一次更新模型仍然需要遍历全部数据, 仍然是其处理大规模数据的瓶颈. 如果仅仅将 DMNL 的目标函数迁移到深度学习框架下, 损失函数与样本之间相似度有很大关联, 在深度学习框架下, 样本特征是通过学习得到的, 将会有很大的过拟合风险. 此外, DMNL 中需要计算示例的协方差矩阵, 通过小批量对协方差矩阵进行估计往往不准, 从而影响学习性能. 因此, 针对多标记新标记检测任务的端到端学习方法无法通过简单结合现有深度网络结构和损失函数得到.

多视图学习用于处理由多个视图描述的数据, 即多个特征集合. 这里, 每个视图是一个特征集合, 每个样本在每个视图上都有相应的特征向量表示. 现有的多视图学习研究工作^[23~27] 通过利用多个视图之间的关系提升学习性能或是降低样本复杂度. 前文所提到的多标记新标记检测问题的相关工作均是基于单一视图, 本文工作引入了多视图信息, 希望利用额外的信息提升学习性能. 同时借鉴了 OPMV (one-pass multi-view learning) 工作^[27] 的思想, 通过要求对同一对象在多个视图上的预测保持一致性, 从而约束模型. 将这种一致性约束写成正则化项的形式加入损失函数, 使得非常易于在线优化.

3 问题形式化

令 $\mathbf{v} = \{1, 2, \dots, l\}$ 表示已知的标记集合, $\mathcal{X}$ 和 $\mathcal{X}^*$ 分别表示原始视图和增广视图的样本空间. 给定一个输入序列 $\{\mathbf{X}_i, \mathbf{X}_i^*, \mathbf{y}_i\}_{i=1}^T$, 其中 $\mathbf{X}_i \in \mathcal{X}$ 表示原始视图的一个样本, $\mathbf{X}_i^* \in \mathcal{X}^*$ 表示相应的增广视图, $\mathbf{y}_i \in \mathcal{Y} = \{0, 1\}^l$ 表示观察到的标记向量, $y_{i,j} = 1$ 表示第 i 个样本有标记 v_j , 否则没有. 需要特别指出, 这里的原始视图和增广视图可以是但不仅限于原始的图片 and 文本输入. 本工作虽以图片与文本视图为例, 但可以通过改变特征学习的网络结构扩展到其他输入类型. 令 $\mathbf{v}' = \{1, 2, \dots, l, l+1\}$ 表示测试集标记集合, 其中, $v'_{l+1} \notin \mathbf{v}$ 表示存在新语义 (新标记). 令 $\mathcal{Y}' = \{0, 1\}^{l+1}$ 表示测试集的标记空间. 目标是学习一个映射 $f: \mathcal{X} \times \mathcal{X}^* \rightarrow \mathcal{Y}'$, 预测测试样本的标记 (包括新标记).

4 EM3NL 方法

本节主要介绍了 EM3NL 方法, 包括目标函数设计、多示例多标记学习网络结构设计, 以及特征提取网络结构设计, 分别对应 4.1~4.3 小节.

4.1 目标函数设计

在 EM3NL 中, 利用卷积及池化操作建立深度神经网络, 分别为图像视图和文本视图学得多示例特征, 记为 $\mathbf{B}_i = \phi(\mathbf{X}_i)$ 和 $\mathbf{B}_i^* = \psi(\mathbf{X}_i^*)$. 令 $\hat{\mathbf{y}}_i = \mathcal{H}(\mathbf{B}_i)$ 和 $\hat{\mathbf{y}}_i^* = \mathcal{H}^*(\mathbf{B}_i^*)$ 分别表示图像视图和文本视图多示例包上标记的预测 (包括新语义); 令 $\Pi_i^j(\cdot)$ 表示取矩阵的第 i 到 j 列, 则 EM3NL 优化了下述目标函数:

$$\min_{\phi, \psi, \mathcal{H}, \mathcal{H}^*} \sum_{i=1}^T \mathcal{L}(\Pi_1^l(\hat{\mathbf{y}}_i), \mathbf{y}_i) + \mathcal{L}(\Pi_1^l(\hat{\mathbf{y}}_i^*), \mathbf{y}_i) + \lambda_1(\mathcal{R}_1(\hat{\mathbf{y}}_i) + \mathcal{R}_1(\hat{\mathbf{y}}_i^*)) + \lambda_2 \mathcal{R}_2(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_i^*), \quad (1)$$

其中, $\mathcal{L}(\cdot, \cdot)$ 表示在已知标记上的分类误差损失函数; $\mathcal{R}_1(\cdot)$ 是对预测值的正则化项, 由于没有关于新标记的监督信息, 希望通过该正则化项利用标记之间的关系进一步约束标记的预测值; $\mathcal{R}_2(\cdot, \cdot)$ 是视图一致性正则化项, 约束同一对象的不同视图上的预测值尽可能接近.

由于 $\mathcal{Y}' \in \{0, 1\}^{l+1}$, 且一个对象可能同时与多个对象相关联, 因此对于 $\mathcal{L}(\cdot, \cdot)$, 采用深度学习中常用的二值交叉熵损失函数, 即

$$\mathcal{L}(\Pi_1^l(\hat{\mathbf{y}}_i), \mathbf{y}_i) = - \sum_{j=1}^l y_{i,j} \log \hat{y}_{i,j} + (1 - y_{i,j}) \log(1 - \hat{y}_{i,j}),$$

将 $\hat{\mathbf{y}}_i$ 替换为 $\hat{\mathbf{y}}_i^*$ 即可得 $\mathcal{L}(\Pi_1^l(\hat{\mathbf{y}}_i^*), \mathbf{y}_i)$.

进一步地, 通过 $\mathcal{R}_1(\cdot)$ 利用标记排序对新标记的预测值进行约束: 假设 $\forall v'_j \in \mathbf{v}_i'^+ \succ v'_{l+1} \succ \forall v'_j \in \mathbf{v}_i'^-$ 对 $i \in [T]$ 成立, 其中 $\mathbf{v}_i'^+ = \{v'_j | y_{i,j} = 1\}$, $\mathbf{v}_i'^- = \{v'_j | y_{i,j} = 0\}$ 分别表示第 i 个样本上观察到的相关标记集合和不相关标记集合, $a \succ b$ 表示 a 排在 b 之前. 这样令新标记排序在相关标记与不相关标记之间, 无论新标记预测为 1 或 0, 均不会使得标记排序损失增大. 为了对新标记排在相关标记之前或是不相关标记之后进行惩罚, 定义

$$\mathcal{R}_1(\hat{\mathbf{y}}_i) = - \sum_{j=1}^l (1 - y_{i,j}) \log(\sigma(\hat{y}_{i,l+1} - \hat{y}_{i,j})) + y_{i,j} \log(1 - \sigma(\hat{y}_{i,l+1} - \hat{y}_{i,j})),$$

其中 σ 是 sigmoid 函数.

最后, 通过最小化预测值的欧式距离约束同一对象两个视图上的预测尽可能相近, 定义

$$\mathcal{R}_2 = \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_i^*\|^2.$$

4.2 包级预测与示例级预测

为了方便描述使符号简洁, 本小节中省略了多示例包以及相应预测值的下标. 对于样本级的预测, 类似 DeepMIML^[28], 首先将示例映射到子标记语义: 假设每个标记 (包括新标记) 有 k 个子语义, 采用 3 层全连接层 (FC, fully connected layer) $\mathcal{F}$, 将多示例包中的每个示例 $\mathbf{b}_i \in \mathbf{B}$ 映射到 $\mathbf{g} = \mathcal{F}(\mathbf{b}_i) \in \mathbb{R}^{(l+1)k}$.

图 1 (网络版彩图) 示例级标记预测

Figure 1 (Color online) Instance level prediction

图 2 (网络版彩图) 包级标记预测

Figure 2 (Color online) Bag level prediction

前两层全连接层均采用 Relu 函数作为激活函数, 与 DeepMIML 不同的是, 最后一层采用 Softmax 函数作为激活函数而非 Sigmoid 函数. 应用 Softmax 函数相比于 Sigmoid 函数更加适合新标记检测任务, 应用 Softmax 函数由于要求所有子语义上的预测之和为 1, 通过优化往往会得到在某些语义上有比较大的预测值, 而在其他语义上的预测值均非常接近 0. 当已知标记的语义的预测值都非常小的时候, 则更有可能将示例预测为包含新标记的示例.

同样地, 利用 Reshape 操作将 g 转为一个 $(l+1) \times k$ 的矩阵作为 2D - 子语义层 (2D-subconcept layer). 由于利用 Softmax 函数得到的预测值天然有概率的含义, 表示属于某个子语义的概率, 则根据加法原理, 2D - 子语义层按行相加则得到了对该示例每个标记上的预测值 f_i . 单个示例的预测如图 1 所示.

在现有的多示例学习方法中, 根据示例级预测得到包级预测的方法主要包括: (1) 将包中所有示例的预测值平均作为最终对包的标记预测; (2) 将包中所有示例在每个标记上预测的最大值作为最终对包的标记预测. 对于 (1), 如果对于某个标记, 无关示例的数量远大于相关示例的数量, 那么通过平均得到的预测值将接近 0, 与真实情况相反. 对于 (2), 只有部分示例对包的标记预测做出了贡献, 而实际上, 包中与标记相关的示例可能不止一个, 它们对包的标记预测的贡献都应该考虑进去.

由于每个包与多个标记相关联, 且假设包中每个示例对应一个标记, 本工作对包标记预测的基本思想是着重考虑每个示例上最大预测值对应的标记预测对包的标记预测的贡献: $F = [f_1; \dots; f_m]$ 表示多示例包 B 中所有示例的示例级预测矩阵, 其中预测值经由全连接层和 2D - 子语义层输出, m 是

图 3 (网络版彩图) 辅助文本视图特征学习网络结构

Figure 3 (Color online) Text-view network architecture

包中示例的个数. 定义示例预测值权重矩阵为

$$\mathbf{W} = \left[\frac{1}{m} \right]^{m \times (l+1)} + \mathbf{S}, \quad (2)$$

其中 $\mathbf{S} = [\mathbf{s}_1; \dots; \mathbf{s}_m]$, $\mathbf{s}_i = \mathbf{e}_{\arg \max_j f_{i,j}}$, $\mathbf{e}_k$ 是第 k 位为 1, 其余位为 0 的行向量. 这里, 式 (2) 中的第 1 项是每个示例对每个标记的预测值对包上标记预测的基本贡献权重, 第 2 项代表每个样本在预测值最大的标记上对包上相应标记预测作出更多的贡献. 则对包中示例预测加权求和后可得包上的预测为

$$\hat{\mathbf{y}} = \frac{\mathbf{1}^T (\mathbf{W} \odot \mathbf{F})}{\mathbf{z}}, \quad (3)$$

其中 $\mathbf{z} = \mathbf{1}^T \mathbf{W}$ 为归一化系数向量, $\odot$ 表示矩阵对应元素相乘, $\div$ 表示对应元素相除.

综合样本级和包级预测, 设计了顶层网络结构 $\mathcal{H}$, 将多示例包映射到 $\mathcal{Y}'$, 如图 2 所示, 其中 $\mathbf{B}$ 到 $\mathbf{F}$ 映射网络结构的横截面即为图 1 展示的示例级标记预测网络结构, 且对于包中每个示例子语义 (subconcept) 的预测中, 全连接层对所有包中示例共享权重. 增广视图多示例包到标记预测网络 $\mathcal{H}^*$ 的设计与 $\mathcal{H}$ 类似, 只需根据输入示例的维度调整每一层的大小即可.

4.3 特征提取网络

对文本特征提取, 为了便于获取多示例包, 没有采用常用的 LSTM (long short-term memory) 结构, 而是基于一维卷积设计了卷积神经网络结构. 对于每段文本输入, 采用定长为 64 的截断策略, 即单词串长度超过 64 部分进行截断, 不足 64 的进行补 0 操作. 利用 Embedding 层将其转换为 64×1024 的实值矩阵. 考虑 4 种不同大小卷积核的一维卷积层, 分别对应每 1~4 个单词对应的向量上做卷积操作, 然后将输出拼接作为该层的表示. 这样做考虑了单词以及不同长度短语的结构 (不同卷积核卷积在边界处可能长度不足, 此时利用 Padding 操作进行补 0). 交替增加一维卷积层和最大池化层 (max pooling layer), 得到 16 个 2048 维示例组成的示例包, 如图 3 所示. 其中粉色/蓝色/橙色/绿色的矩形分别表示在 1/2/3/4 个单词对应的向量上做 1 维卷积, 红色矩形代表最大池化层, 输出每两个单词对应表示的最大值. 最后得到的每个示例为一个长度为 4 的单词子串的相关语义的表示.

对图像特征提取, 采用标准的 VGG (visual geometry group) 网络结构, 输入为一张 $224 \times 224 \times 3$ 的彩色图像, 输出为 49 个 512 维示例组成的多示例包. 其中每个示例对应的是原始图片 $1/49$ 的一个小块. 图像特征提取网络结构如图 4 所示.

图 4 (网络版彩图) 原始图像视图特征学习网络结构
Figure 4 (Color online) Image-view network architecture

5 实验

本节在 MS-COCO 数据集上验证了 EM3NL 方法在新标记检测以及图片多标记分类任务上的有效性.

5.1 数据集

实验数据集为 MS-COCO^[30] 数据集, 一共包括超过 12 万张图片, 且每张图片关联 5 句文本描述. MS-COCO 数据中一共有覆盖了 80 个不同标记, 平均每个样本有 2.91 个相关标记. 随机将 MS-COCO 划分训练集和测试集, 其中 2/3 样本作为训练数据, 剩余的 1/3 作为测试数据. 为了验证 EM3NL 发现新标记的性能, 分别将不同数量的标记从训练集样本标记中去除 (随机去除 1/5/10 个标记) 作为新标记; 而测试集中仍保留这些标记, 将去除标记后的数据集分别命名为 MS-COCO-1, MS-COCO-5, MS-COCO-10.

5.2 评价指标

从两个方面评价模型的性能, 包括: (1) 新标记检测的性能; (2) 在已知标记集合上的性能. 对于 (1), 如果样本包含训练集标记集合中所去除的任意标记, 则该样本被认为存在新标记, 采用 AUC 值作为评价指标. 对于 (2), 分别采用多标记学习中的 4 个常用指标^[5] Hamming loss, ranking loss, coverage 和 average precision (AP) 对模型进行评价, 其中 Hamming loss, ranking loss 和 coverage 越小越好, AP 越大越好.

5.3 对比方法

由于之前没有端到端的多视图多示例标记新标记学习研究工作 (由于 MIMLNC 和 DMNL 方法无法处理大规模的数据, 因而也无法在提取的特征上应用它们作为对比), 为了验证 EM3NL 的有效性, 对比以下方法:

(1) EM3NL-gt: 采用 EM3NL 方法进行训练, 但是在训练时能够看到新标记作为监督信息. 此方法用于显示 EM3NL 没有新标记监督信息进行训练与利用了真实新标记标记信息之间的性能差距;

(2) EM3NL-im: 仅使用 EM3NL 的图像视图训练模型, 优化 $\min_{\phi, \mathcal{H}} \sum_{i=1}^T \mathcal{L}(\Pi_1^l(\hat{\mathbf{y}}_i), \mathbf{y}_i) + \lambda_1 \mathcal{R}_1(\hat{\mathbf{y}}_i)$;

图 5 (网络版彩图) 新标记检测 AUC 对比结果

Figure 5 (Color online) New label detection performance in AUC

(3) EM3NL-txt: 仅使用 EM3NL 的文本视图训练模型, 优化 $\min_{\psi, \mathcal{H}^*} \sum_{i=1}^T \mathcal{L}(\Pi_1^l(\hat{\mathbf{y}}_i^*), \mathbf{y}_i) + \lambda_1 \mathcal{R}_1(\hat{\mathbf{y}}_i^*)$;

(4) DeepMIML-mv: 利用 $\mathcal{R}_2$ 将 DeepMIML 扩展至多视图模型; 由于 DeepMIML 方法不考虑新标记, 则将多示例包中所有预测值均低于 0.5 的示例当作新标记示例, 认为包含该示例的包有新标记.

实验中, λ_1 参数的取值范围为 $\{0.1 : 0.05 : 0.5\}$, λ_2 的取值范围为 $\{0.001, 0.005, 0.01, 0.05\}$, 参数选择通过 3 折交叉验证. 优化算法采用 Adadelta^[31]. 实验环境基于 Python 3.6 + keras 2.1.2 + tensorflow 1.4.1 + cuda 6.0; 使用 GPU 为单块 TITAN Xp.

5.4 实验结果

新标记预测的结果 (AUC) 如图 5 所示. EM3NL 在新标记预测任务上的结果仅次于 EM3NL-gt, 它超过两个单视图版本 EM3NL-im 和 EM3NL-txt, 远超 DeepMIML-mv. DeepMIML-mv 的设计不考虑新标记, 因此它倾向于将包中示例预测为已知标记, 因此对新标记的预测接近随机猜测. EM3NL-gt 从训练阶段就能利用真实新标记信息, 因此在新标记预测的结果最好. EM3NL 尽管在训练时没有新标记的监督信息, 但在测试集上对新标记进行预测时, 还是取得了接近 EM3NL-gt 的性能. 这验证了 EM3NL 在新标记预测任务上的有效性.

进一步观察 MS-COCO-1, MS-COCO-5 和 MS-COCO-10 在示例级对新标记的预测结果, 并以此绘制热度图. 图 6 展示了 3 例原始图片和相应新标记预测的热度图, 每个数据集对应一例, 热度图颜色由深蓝至浅黄表示预测值从 0 到 1. 尽管示例级的新标记预测与原图的相关语义的位置并不完全吻合, 但是二者之间有着非常强的相关性. 考虑到这些新标记在训练的时候完全未被标注出来, 说明 EM3NL 能够相对准确地地在图片中定位到新语义的位置.

在已知标记上评价指标的结果 Hamming loss, ranking loss, coverage 和 AP 分别如图 7 所示. 同样在已知标记上, EM3NL 在大部分指标上取得了除 EM3NL-gt 以外最好的性能. 这说明 EM3NL 预测新标记, 并不会损害对已知标记的预测效果. 这进一步说明了 EM3NL 的有效性.

6 总结

本文针对多标记新标记学习问题, 提出了一种基于多视图多示例多标记学习方法 EM3NL, 主要

图 6 (网络版彩图) 上部为原始图片, 下部为示例级的新标记预测热度图, 颜色由深蓝至黄色代表热度由 0 到 1
Figure 6 (Color online) The upper shows the original images, the lower is heat map of the new label prediction

图 7 (网络版彩图) 在已知标记上的性能比较

Figure 7 (Color online) The comparison results on known labels. (a) Hamming loss; (b) ranking loss; (c) coverage; (d) average precision

解决了多标记样本新标记的发现与预测问题. 主要方法是将深度神经网络与多视图、多示例多标记学习相结合, 设计了一种直接输入图像 - 文本视图, 对已知标记预测和新标记检测的端到端统一解决方案. 具体包括: 设计了新的目标函数考虑包级的预测损失、对新标记预测的正则化项, 以及多视图一致

性正则化项;改进设计了包级预测与示例级预测算法及相关网络结构.在MS-COCO数据集上验证了本文方法预测已知标记和新标记的有效性.

参考文献

- 1 Boutell M R, Luo J, Shen X, et al. Learning multi-label scene classification. *Pattern Recogn*, 2004, 37: 1757–1771
- 2 Zhang M L, Zhang K. Multi-label learning by exploiting label dependency. In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, 2010. 999–1008
- 3 Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification. *Mach Learn*, 2011, 85: 333–359
- 4 Huang S J, Zhou Z H. Multi-label learning by exploiting label correlations locally. In: *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, Toronto, 2012. 949–955
- 5 Zhang M L, Zhou Z H. A review on multi-label learning algorithms. *IEEE Trans Knowl Data Eng*, 2014, 26: 1819–1837
- 6 Zhou Z H, Zhang M L. *Multi-Label Learning*. Berlin: Springer, 2017
- 7 Zhou Z H. A brief introduction to weakly supervised learning. *Natl Sci Rev*, 2018, 5: 44–53
- 8 Xu M, Jin R, Zhou Z H. Speedup matrix completion with side information: application to multi-label learning. In: *Proceedings of Advances in Neural Information Processing Systems*, Lake Tahoe, 2013. 2301–2309
- 9 Xu L L, Wang Z, Shen Z F, et al. Learning low-rank label correlations for multi-label classification with missing labels. In: *Proceedings of the 14th IEEE International Conference on Data Mining*, Shenzhen, 2014. 1067–1072
- 10 Yu H F, Jain P, Kar P, et al. Large-scale multi-label learning with missing labels. In: *Proceedings of the 31th International Conference on Machine Learning*, Beijing, 2014. 593–601
- 11 Zhu Y, Kwok J T, Zhou Z H. Multi-label learning with global and local label correlation. *IEEE Trans Knowl Data Eng*, 2018, 30: 1081–1094
- 12 Zhu Y, Ting K M, Zhou Z H. Discover multiple novel labels in multi-instance multi-label learning. In: *Proceedings of the 31th AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 2977–2984
- 13 Zhu Y, Ting K M, Zhou Z H. Multi-label learning with emerging new labels. In: *Proceedings of the 16th International Conference on Data Mining*, Barcelona, 2016
- 14 Zhou Z H, Zhang M L. Multi-instance multi-label learning with application to scene classification. In: *Proceedings of Advances in Neural Information Processing Systems*, Vancouver, 2007. 1609–1616
- 15 Zhou Z H, Zhang M L, Huang S J, et al. Multi-instance multi-label learning. *Artif Intell*, 2012, 176: 2291–2320
- 16 Zhang M L. A k-nearest neighbor based multi-instance multi-label learning algorithm. In: *Proceedings of the 22nd IEEE International Conference on Tools with Artificial Intelligence*, Arras, 2010. 207–212
- 17 Briggs F, Fern X Z, Raich R. Rank-loss support instance machines for miml instance annotation. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, 2012. 534–542
- 18 Huang S J, Gao W, Zhou Z H. Fast multi-instance multi-label learning. In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, Quebec City, 2014. 1868–1874
- 19 Pham A T, Raich R, Fern X Z. Dynamic programming for instance annotation in multi-instance multi-label learning. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39: 2381–2394
- 20 Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35: 1798–1828
- 21 Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*, 2015, 61: 85–117
- 22 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521: 436–444
- 23 Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: *Proceedings of the 11th Conference on Computational Learning Theory*, Madison, 1998. 92–100
- 24 Wang W, Zhou Z H. A new analysis of co-training. In: *Proceedings of the 27th International Conference on Machine Learning*, Haifa, 2010. 1135–1142
- 25 Guo Y H, Xiao M. Cross language text classification via subspace co-regularized multi-view learning. In: *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, 2012. 1615–1622
- 26 Guo Y H. Convex subspace representation learning from multi-view data. In: *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, Bellevue, 2013. 387–393
- 27 Zhu Y, Gao W, Zhou Z H. One-pass multi-view learning. In: *Proceedings of the 7th Asian Conference on Machine*

- Learning, Hong Kong, 2015. 407–422
- 28 Feng J, Zhou Z H. Deep MIML network. In: Proceedings of the 31th AAAI Conference on Artificial Intelligence, San Francisco, 2017. 1884–1890
- 29 Pham A T, Raich R, Fern X Z, et al. Multi-instance multi-label learning in the presence of novel class instances. In: Proceedings of the 32nd International Conference on Machine Learning, Lille, 2015. 2427–2435
- 30 Lin T Y, Maire M, Belongie S J, et al. Microsoft COCO: common objects in context. In: Proceedings of the 13th European Conference on Computer Vision, 2014. 740–755
- 31 Zeiler M D. Adadelata: an adaptive learning rate method. 2012. ArXiv:1212.5701

Multi-instance multi-label new label learning

Yue ZHU^{1,2}, Yuan JIANG^{1,2*} & Zhi-Hua ZHOU^{1,2}

1. *National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China;*

2. *Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210023, China*

* Corresponding author. E-mail: jiangy@lamda.nju.edu.cn

Abstract In the studies of traditional multi-label learning, they assume the observed label distribution in the training data is the same as the true label distribution in the test data. In real applications, however, it is not the case; some existing semantics are not labeled throughout the dataset. In the testing phase, the learning system not only requires a good performance on the known target labels but also needs to detect the potential new labels. This paper proposes an end-to-end multi-view multi-instance multi-label learning approach, EM3NL, combining deep learning with multi-view multi-instance multi-label learning framework to address the new label learning problem. The performance on MS-COCO datasets validates the effectiveness of EM3NL on both the performance on the known labels and the new labels.

Keywords multi-label learning with new labels, multi-instance multi-label learning, multi-view learning, deep learning

Yue ZHU was born in 1989. He received the B.Sc. degree in computer science and technology from Nanjing Normal University, Nanjing, China, in 2011. In the same year, he was admitted to further study in Nanjing University, Nanjing, China. Now, he is currently a Ph.D. candidate and a member of LAMDA group. His research interests mainly include machine learning and data mining.

Yuan JIANG was born in 1976. She obtained her Ph.D. degree in computer science from Nanjing University, China, in 2004. She is now a professor at Nanjing University. Her research interests include machine learning, information retrieval, and data mining.

Zhi-Hua ZHOU was born in 1973. He obtained his Ph.D. degree in computer science from Nanjing University, China, in 2000. Currently, he is a professor at Nanjing University. His research interests mainly include artificial intelligence, machine learning, and data mining.