



DOI:10.12404/j.issn.1671-1815.2400465

引用格式:谢晓兰,谭舒孺,王楠.对偶结构下的多标签半监督课程学习[J].科学技术与工程,2025,25(5):1980-1987.

Xie Xiaolan, Tan Shuru, Wang Nan. Semi-supervised curriculum learning of multi-label under dual structure[J]. Science Technology and Engineering, 2025, 25(5): 1980-1987.

自动化技术、计算机技术

对偶结构下的多标签半监督课程学习

谢晓兰^{1,2}, 谭舒孺^{1,2}, 王楠^{3*}

(1. 桂林理工大学信息科学与工程学院, 桂林 541004; 2. 广西嵌入式技术与智能系统重点实验室, 桂林 541004;

3. 吉林财经大学管理科学与信息工程学院, 长春 130117)

摘要 多标签学习是现实应用场景中的一个常见问题。大规模多标签数据集的构建往往意味着高昂的成本,因此出现了半监督学习技术。目前,大多数半监督学习主要用于单标签分类领域,尽管半监督学习在多标签分类领域取得了一些进展,但在训练时间消耗、训练效果和标签之间潜在关系的利用方面仍有很大的改进空间。针对上述问题,提出了一种二元结构下的多标签半监督课程学习模式(semi-supervised course learning under dual structure,SSCD)。首先,设计了一种基于对偶差分的课程学习方案,大大减少了训练时间,提高了模型的稳健性;其次,设计了一个单一注意力机制来探索标签之间的潜在相关性。在3个开放测试数据集上评估了SSCD在预测任务中的性能,并与4个基准模型进行了比较,结果表明SSCD的综合指标在各个方面都是最优的;最后,通过结构消融实验验证了所提出的单注意力机制的有效性。

关键词 多标签学习;半监督学习;自注意力机制;课程学习

中图分类号 TP391;

文献标志码 A

Semi-supervised Curriculum Learning of Multi-label under Dual Structure

XIE Xiao-lan^{1,2}, TAN Shu-ru^{1,2}, WANG Nan^{3*}

(1. College of Information Science and Engineering, Guilin University of Technology, Guilin 541004, China;

2. Guangxi Key Laboratory of Embedded Technology and Intelligent Systems, Guilin 541004, China;

3. School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun 130117, China)

[Abstract] Multi-label learning is a common problem in real application scenarios. The construction of large-scale multi-label datasets often means high cost, so semi-supervised learning technology appears. At present, most semi-supervised learning is mainly used in the field of single label classification. Although semi-supervised learning in the field of multiple labels classification has made some progress, there is still much room for improvement in training time consumption, training effects and the use of potential relationships between labels. A multi-label semi-supervised curriculum learning model was proposed under the dual structure semi supervised course learning under dual structure(SSCD) to solve the above problems. Firstly, a curriculum learning scheme based on dual difference was designed, which greatly reduces the training time and improves the robustness of the model. Secondly, a single attention mechanism was designed to explore the potential relevance between labels. The performance of SSCD in the prediction task was evaluated on three open test datasets, and the results compared with four benchmark models show that the comprehensive indicators of SSCD are optimal in all aspects. Finally, the structure ablation experiment was carried out to prove the effectiveness of the proposed single attention mechanism.

[Keywords] multi-label learning; semi-supervised learning; self-attention mechanism; curriculum learning

人们在生活中经常面临一个样本对应多个标签的情况^[1],例如,一段话可能有多种情感含义;一部电影也可能属于多种类别,可以包含推理、悬疑、恐怖等元素;一幅画可能描绘了雨和雷两种事物。因此,多标签分类任务已成为生活中不可或缺的一部分。目

前,多标签学习仍然面临一些挑战^[2-4]。首先,现有的多标签数据集不多,如CUB、Yeast,其主要原因是构建大规模的多标签数据集需要很高的成本,并且随着标签的增加,构建将变得越来越困难^[5]。同时,对于主观标签,如判别一段话中是否存在积极的情绪倾

收稿日期:2024-01-16 修订日期:2024-11-12

基金项目:国家自然科学基金(62262011);广西自然科学基金(2021JJA170130);国家社会科学基金(22BTQ048);吉林省教育厅科学研究项目(JJKH20250758KJ)

第一作者:谢晓兰(1974—),女,汉族,广西桂林人,博士,教授。研究方向:大数据及云计算。E-mail:xie_xiaolan@foxmail.com。

*通信作者:王楠(1980—),女,汉族,吉林长春人,博士,教授。研究方向:网络舆情、自然语言处理。E-mail:119014@jlufe.edu.cn。

向,其标准难以统一,可能会使得数据集中出现噪声。通过语义关系知识构建数据集是一种解决方案。然而,建立语义关系需要特殊的语义知识。即使构建成功,构建的关系图也只服务于特定任务^[6],难以应用于其他任务。在网络中标记数据的获取难度大,但未标记数据获取则相对简单。因此,学者们提出了使用少量标记数据进行训练的半监督学习^[7]解决数据问题。传统的半监督方法主要研究特征空间中的数据分布^[8-15]。

文献[10]在一些任务中使用管道来加强学习实现半监督学习。文献[11]使用基于图的方法来设计图卷积神经网络来进行半监督学习。文献[13]提出了一种新的基于自适应图的多视图降维方法解决标签问题。文献[15]提出了一种增强的分类对齐策略,探索了多种相互补充的技术。

目前,多标签半监督学习^[12]主要基于图结构传导策略和基于双分类结构的固定比例选择策略。但是基于图结构的传导策略对计算资源要求较高,计算方法也较为复杂。而基于双分类器结构的半监督学习方法需要花费较长的训练时间。

在单标签学习中^[16],一些学者提出了一种课程学习的训练方法,该方法在算资源与训练时间上的表现不亚于甚至超过当前主流的监督和半监督学习。然而,这种课程学习目前很少用于多标签学习领域。主要原因是多标签半监督学习不能将模型分类结果直接作为评价标准来评估伪标签样本。同时,如何控制学习节奏,为模型安排合适的课程,以及如何学习标签信息,也是多标签半监督学习面临的重要挑战。

因此,现借鉴单标签领域课程学习的思想^[17-18],并受到多标签半监督学习中的双重分类结构的启发,提出一种用于多标签分类的双重结构下的半监督课程学习(semi supervised course learning under dual structure, SSCD)模型,使用两个分类器的差异值作为设计课程的评分标准(即选择要添加到训练集中的伪标签样本),并根据数据的比例自适应地调整所选伪标签数据的差分阈值。具体地,基于双分类器结构,提出一种多标签分类的半监督课程学习框架,以解决基于双分类器的模型的训练速度和训练效果之间的平衡问题。提出一种单一注意力机制来学习标签之间的关系,解决现有模型中多个标签之间关系挖掘不足的问题。

1 多标签半监督学习相关研究工作

多标签学习即模型在学习时面对的样本具有一个及以上的标签。现实生活中有许多场景都适

用于多标签学习,例如文献[3]采用了双向 LSTM 算法实现了多标签的文本分类问题,文献[4]提出了一种基于 XLCLS 融合法律解释和相似性 XLNET-CNN 模型的法条多标签分类方法对公诉案件进行了分类研究,但是由于多标签数据集获取难度远高于无标记数据集,构建多标签数据集往往需要大量的人力成本,因此许多学者们开始注意到了已有的单标签学习领域的半监督学习^[19-21]。

在半监督学习中,数据集由标记数据和未标记数据组成。简单来说,半监督学习的过程为模型通过标签数据进行训练,随后预测无标签数据集后产生伪标记数据,根据某种准则筛选伪标记数据继续训练模型直到模型满足要求或所有未标记数据集全部加入训练集合。

但是,多标签半监督学习不同于单标签半监督的学习,多个标签导致在训练时很难基于单个标签的预测结果划分下一轮的伪标签训练数据。目前,大量的多标签半监督学习算法都是基于图模型^[23-23]和低秩假设^[23-24]的策略,其中,基于图的方法在多标签半监督学习中非常流行。CMLP^[23]使用协作技术设计了一种可扩展的多标签传播方法。文献[25]提出了一种低秩正则化方法,其中使用低秩特征恢复矩阵来动态更新图。SLRM^[26]强制分类器为低秩,基于图结构添加流形正则化以确保样本平滑。ML-TLLT^[27]在低秩图的基础上提出了一种迫使教师模型生成类似课程的训练方式用于模型训练。

除了基于图的半监督学习方法外,还有另一种基于双分类器结构的半监督学习策略。COIN^[28]结构将大规模协作训练策略应用于多标签半监督学习,通过成本函数对伪标记数据进行排序以选择参与训练的伪标记样本。基于 COIN 结构,文献[29]进一步提出了一种流式多标签数据学习方法。在此基础上,DRML^[30]设计了一个双分类器自适应网络来对齐标记和未标记数据的分布。DRML 计算了两个分类器的预测标签向量的外积,然后通过关系提取网络生成最终预测。随后在 DRML 的基础上,DRML 的简化模型 SDRL^[31]提出,简化了 DRML 的标签相关性复杂计算,在降低了模型的训练时间的同时提高了模型的识别准确率。

以上方法均有各自的局限性,基于图结构的传导预测策略对计算资源要求较高,计算方法也较为复杂。而基于双分类器结构(COIN)的半监督学习方法需要花费较长的训练时间。例如,文献[37]在每轮学习中只提取 1% ~ 3% 的伪标记样本加入训练集合。若无满足条件样本,则只向训练集合中添加一个样本。这种方案虽然减少了训练噪声,但使

得训练时间大幅增加。因此,通过适当的节奏使模型选择适当的伪标记数据一直是半监督学习的一项重要任务。因此,基于双分类器结构和课程学习的思想^[18],提出 SSCD 模型,该模型基于 COIN 结构,不需要图结构的大量计算资源,同时解决 COIN 结构训练时间较长的问题。

2 SSCD 模型

2.1 多标签半监督学习

SSCD 模型框架如图 1 所示。SSCD 包含一个特征提取器 $E(\cdot)$,两个多标签分类器 $C_1(\cdot)$ 和 $C_2(\cdot)$,以及一个最终分类器 $Cre(\cdot)$ 。 $E(\cdot)$ 用于提取特征, $C_1(\cdot)$ 和 $C_2(\cdot)$ 通过自注意力机制提取标签的相关性,然后将原始标签信息和提取的标签相关信息相加。最后,通过 $Cre(\cdot)$ 对样本进行分类。在课程学习过程中,通过 $C_1(\cdot)$ 和 $C_2(\cdot)$ 计算每个样本的差异选择添加到训练集中的伪标记样本。

设 $\{X_1, Y_1\}$ 为给定的标记数据; $X_1 \in \mathbf{R}^{n_1 \times d}$ 为标记样本的特征向量矩阵; $Y_1 \in \mathbf{R}^{n_1 \times d_1}$ 为相应的标签矩阵; n_1 为标签的样本数量; d 和 d_1 分别为特征和标签类别的数量; $x_i \in \mathbf{R}^d$ 为第 i 个样本,即 X_1 的一行; y_i 为 x_i 的对应标签; $X_u \in \mathbf{R}^{n_u \times d}$ 和 $Y_u \in \mathbf{R}^{n_u \times d_1}$ 分别为未标记样本的特征矩阵和标签矩阵。

使用特征提取器 $E(\cdot)$ 将所有样本编码到相同的特征空间中,如式(1)所示。

$$\begin{cases} S_1 = E(X_1) \\ S_u = E(X_u) \end{cases} \quad (1)$$

式(1)中: $S_u \in \mathbf{R}^{n_u \times d_s}$ 和 $S_1 \in \mathbf{R}^{n_1 \times d_s}$ 为编码特征矩

阵; d_s 为编码特征尺寸; $C_1(\cdot)$ 和 $C_2(\cdot)$ 用来获得初始预测结果,每个分类器关于标签的损失函数如式(2)所示。

$$L_1(X_1, Y_1) = \frac{1}{2} \left[\|C_1(S_1) - Y_1\|_F^2 + \|C_2(S_1) - Y_1\|_F^2 \right] \quad (2)$$

因此,第一个优化目标为

$$\min[L_1(X_1, Y_1)] \quad (3)$$

由于 X_1 和 X_u 可以通过多种方式获得,因此这两个数据集的分布存在一些差异。采用双重分类器结构来对齐实现标记和未标记数据的分布。第一部分训练分类器 $C_1(\cdot)$ 和 $C_2(\cdot)$ 及编码器 $E(\cdot)$ 以实现对齐分布(最大化 S_u 的方差),两个分类器预测结果的方差可以使用 L_1 范数来估计具体如式(4)所示。

$$d(p_{1i}, p_{2i}) = \frac{1}{d_1} \|p_{1i} - p_{2i}\|_1 \quad (4)$$

式(4)中: $\|\cdot\|_1$ 为 L_1 范数计算; $p_{1i} \in \mathbf{R}^{d_1}$ 和 $p_{2i} \in \mathbf{R}^{d_1}$ 为 $C_1(\cdot)$ 和 $C_2(\cdot)$ 的预测结果。根据式(4)可以得到测量两个分类器方差的第二损失函数 L_2 , 损失函数如式(5)所示。

$$L_2(X_u) = d[C_1(S_u), C_2(S_u)] \quad (5)$$

由于第一部分的训练目标是最大化分类差异,因此损失函数如式(6)所示。

$$\min[L_2(X_u) + \lambda L_1(X_1, Y_1)] \quad (6)$$

式(6)中: λ 为用于控制两个损失函数训练权重的超参数。同时 $E(\cdot)$ 为需要学习子空间中的编码,以减少分类结果的差异。因此,更新 $E(\cdot)$ 的损失函数如式(7)所示。

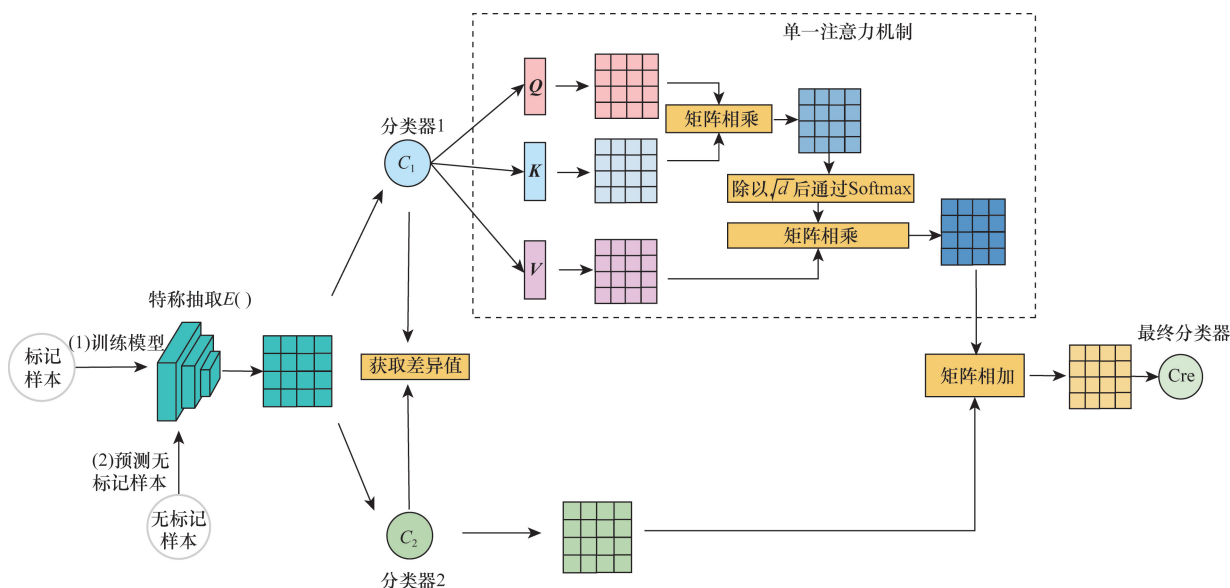


图1 SSCD 的模型框架

Fig. 1 Model Structure of SSCD

$$\min L_2 \quad (7)$$

标签之间的关系是提高多标签预测效果的重要因素。 $C_1(\cdot)$ 、 $C_2(\cdot)$ 只做了简单的预测结果,没有使用标签之间的关系。因此,提出了一种单一注意力机制来捕捉标签之间的关系。具体来说,使用预测结果中的一个矩阵 $\mathbf{RE}_1 \in \mathbf{R}^{n_1 \times d}$ 计算自注意力(基于 transformer 结构)来捕获标签的关系信息,然后将 \mathbf{RE}_2 和 \mathbf{RE}_1 相加。该步骤的目的是将标签的原始信息与标签的注意力信息相结合,并获得可训练的标签关系张量。这一结构的有效性将在 4.3 节中得到证明。为了获得标签关系向量,将对 $C_1(\cdot)$ 和 $C_2(\cdot)$ 的预测结果进行排列,以获得结果矩阵 \mathbf{RE}_1 和 \mathbf{RE}_2 。矩阵的具体形式如式(8)所示。

$$\begin{cases} \mathbf{RE}_1 = [r_{11}, r_{12}, \dots, r_{1i}, \dots, r_{1n_1}] \\ \mathbf{RE}_2 = [r_{21}, r_{22}, \dots, r_{2i}, \dots, r_{2n_1}] \end{cases} \quad (8)$$

式(8)中: $r_{1i} \in \mathbf{R}^d$ 为 $C_1(\cdot)$ 对第 i 个样本的预测结果。类似地; r_{2i} 为 $C_2(\cdot)$ 对第 i 个样本的预测结果。具体地说, $\mathbf{RE}_1(m, n)$ 表示由 $C_1(\cdot)$ 获得的第 m 个样本的第 n 个标签的预测值。类似地, $\mathbf{RE}_2(m, n)$ 表示由 $C_2(\cdot)$ 得到的第 m 个样本的第 n 个标签的预计值。通过 Transformer 中的注意力计算机制对 $\mathbf{RE}_1(m, n)$ 进行处理,得到用于计算注意力机制的 \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 矩阵,具体计算公式如式(9)所示。

$$\begin{cases} \mathbf{Q} = \mathbf{RE}_1 \mathbf{W}_Q \\ \mathbf{K} = \mathbf{RE}_1 \mathbf{W}_K \\ \mathbf{V} = \mathbf{RE}_1 \mathbf{W}_V \end{cases} \quad (9)$$

用 Softmax 函数规范标签的注意力,表达式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (10)$$

随后将 \mathbf{RE}_2 与 Attention 相加,得到包含标签关系信息和原始标签信息的向量 \mathbf{F} 。

$$\mathbf{F} = \text{Relu}[\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) + \mathbf{RE}_2] \quad (11)$$

最后,使用神经网络 Cre(\cdot)得到标签的最终预测。损失函数如式(12)所示。

$$L_{\text{Cre}} = \sum_{i=1}^{n_1} \|y_i - \text{Cre}(\mathbf{F})\|_2^2 \quad (12)$$

在训练的第二部分,所有网络都参与迭代更新,因此最终的损失函数如式(13)所示。

$$\min \left[\frac{\alpha}{2} L_1 + (1 - \alpha) L_{\text{Cre}} \right] \quad (13)$$

式(13)中: $\alpha \in [0, 1]$ 为用于控制两个损失函数的比例的超参数,两个损失函数同等重要,因此 α 取值为 0.5。

2.2 多标签半监督学习

在单标签半监督训练中,由于模型的预测结果可以直接用于选择伪标签样本,因此衍生出了许多伪标签训练样本选择策略。然而,在多标签学习中,一个样本包含多个标签,所以单个标签预测结果不能作为选择伪标签训练样本的依据。因此,借鉴双分类器结构的思想,提出了一种基于对偶结构差异值的课程半监督学习方法来解决上述问题。使用 $C_1(\cdot)$ 和 $C_2(\cdot)$ 的差异值作为伪标签样本选择的评估指标,计算公式为

$$\text{dif}(x_i) = \|C_1[E(x_i)] - C_2[E(x_i)]\|_2^2 \quad (14)$$

训练包括两个部分。在第一部分中,使用现有的标记数据迭代训练 E 、 C_1 和 C_2 大约 50 次。第二部分是基于对偶差分的半监督学习。由于基于对偶差分的概率分布遵循文献[33]中提到的极端事件的概率分布,因此,采用了与固定阈值不同的选择方法。基于文献[19]的结论,设计了一种自适应调整筛选样本差异阈值 dv 的半监督学习方法来选择伪标记训练样本,表达式为

$$dv = \text{Thr}(\mathbf{Y}_u, t) \quad (15)$$

阈值基于 t ($t \in [0.2, 0.4, 0.6, 0.8, 1]$) 的值自适应地调整。具体来说, $\text{Thr}(\mathbf{Y}_u, t)$ 的流程为当课程学习开始时, t 的初始值为 0.2, 然后对于课程学习的每次迭代, t 增加 0.2。首次由 C_1 和 C_2 计算出差值 $\text{dif}(x_i)$ 。之后,基于模型对每个样本预测的差值对伪标记样本进行排序。然后,提取排名在第 t 个百分位处的样本的差值作为阈值 dv 。最后,将差值小于 dv 的所有样本与模型预测出的标签信息添加到训练数据集。因为 t 在每次训练中增加 0.2, 所以每次选择大于等于 20% 的伪标记样本加入训练集合中。

具体训练过程步骤如下。

步骤 1 使用标记样本训练 SSCD。

步骤 2 使用在步骤 1 中训练的 SSCD 预测所有未标记的样本,以生成伪标记样本。

步骤 3 由 $C_1(\cdot)$ 和 $C_2(\cdot)$ 计算每个伪样本的 dif 。

步骤 4 根据 dif 对样本进行排序。

步骤 5 提取排名在 t 百分位的样本的差值,并将该差值用作阈值 dv 。

步骤 6 将差值小于该阈值 dv 的所有伪样本与 SSCD 在第一步的预测标签添加到训练集。

步骤 7 重新训练 SSCD。重复步骤 2 ~ 步骤 7, 直到所有未标记的样本都添加到训练集中。

课程学习的过程如图 2 所示。

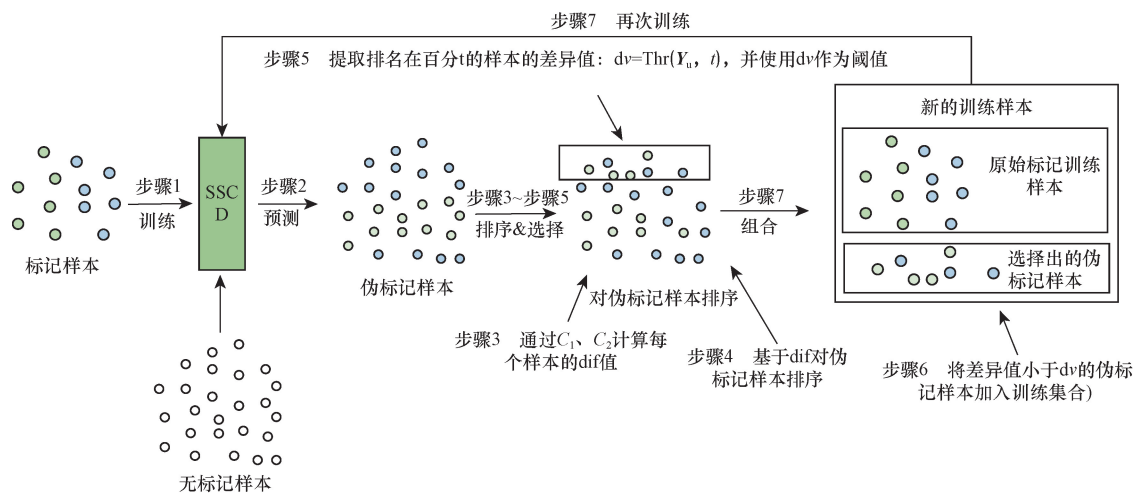


图2 SSCD 课程学习流程

Fig. 2 Curriculum learning of SSCD

3 实验分析

通过4种实验来验证SSCD的有效性:模型的分类能力、模型在面对不同比例有标签数据训练时的泛化性能、模型结构的有效性以及模型在训练时间上的优势。在实验过程中,使用如下3个公共的多标签数据集。

Corel5K 数据集^[21]:该数据集包含来自Corel数据库的照片。分别有4 500和499个样本用于训练和测试。总标签为260个,平均每个样本包含3.4个标签。

CUB 数据集^[1]:该数据集涉及200种鸟类的图像。使用了其中的10 000张进行训练和测试。共有312个标签,每个样本平均有31.4个标签。

Yeast 数据集^[5]:Yeast是一个包含酵母图像的数据集。总共有2 417个样本用于训练和测试。样本的标签总数为14,每个样本平均有4.23个标签。

通过部署VGG19网络^[33]作为 $E(\cdot)$ 来获得图像特征。提取每个样本的特征为长度25 088的一维向量。VGG在ImageNet^[31]上进行过预训练。

SSCD的参数设置如下: C_1 和 C_2 是4层神经网络,包括输入层、输出层和两个线性层;激活函数为ReLU; Q 、 K 、 V 的结构与 C_1 和 C_2 相同,各包含两个没有激活函数的线性层;Cre是一个3层神经网络,包括输入层、输出层、线性层。其中线性层激活函数为ReLU,每个线性层中的神经元数量与样本标签相同。

对于基本的预测能力,选择4种基准方法与本文提出的模型进行比较。

Fast Tag^[20]:该模型专门为处理噪声和不完整的训练样本而设计。模型一部分使用缺失标签训

练来学习标签信息,另一部分用于学习图像的特征信息。这两部分分别完成了缺失标签的线性投影和预测。

SAE (semantic auto encoder):使用线性自动编码器策略来解决标签预测问题。编码器和解码器共享相同的权重,将特征空间投影到标签空间中,然后返回到特征空间。

DRML (dual relation multi-label learning)^[31]:基于COIN结构提出并设计了一个基于矩阵的关系学习网络用于学习标签关系。

SDRL (semi-supervised dual relation learning):基于COIN结构提出了一种简单网络用于学习标签之间的潜在关系。

3.1 模型分类性能评估

为了评估模型之间的差异,使用了5个指标:精确率、召回率、 F_1 ^[22]、绝对匹配率和Hamming损失。精确率、召回率、绝对匹配率和 F_1 越高,表示模型的性能越好。相反,Hamming损失越小,模型表现越好。实验结果如表1所示。

从表1中可以看出,与基准模型相比,所提出的模型在3个数据集的综合能力上有一定的提高。在最简单的Yeast数据集上,本文模型在4个指标上全部优于其他模型。在Corel5k数据集上,尽管SSCD在绝对匹配率上低于SAE,但精确率、 F_1 和召回率指标上分别比单项最佳模型高出2%、4%和6%。在更复杂的CUB数据集上,SSCD在精确率、召回率和 F_1 分别比当前最佳模型高出2%、3%和2%。

3.2 不同比例标记数据的下的学习能力评估

为了进一步验证SSCD模型的有效性,还评估了在不同比例标记数据的情况下,模型的稳健性。结果如表2所示。以Corel5k数据集为例,在原有

表 1 每个模型对每个数据集的预测性能

Table 1 Prediction performance of each model on each data set

| 数据集 | 基准方法 | 绝对匹配率 | 汉明损失 | 精确率 | 召回率 | F_1 |
|---------|---------|-----------------|----------|----------|----------|----------|
| Corel5K | FastTag | 0.008 01 | 0.012 53 | 0.570 39 | 0.067 01 | 0.097 05 |
| | SAE | 0.188 37 | 0.016 89 | 0.218 47 | 0.132 01 | 0.118 06 |
| | DRML | 0.008 01 | 0.012 83 | 0.189 41 | 0.077 51 | 0.077 25 |
| | SDRL | 0.016 03 | 0.011 82 | 0.601 41 | 0.264 97 | 0.301 82 |
| | SSCD | 0.030 06 | 0.006 89 | 0.631 41 | 0.320 63 | 0.341 51 |
| CUB | FastTag | 0.002 51 | 0.082 45 | 0.174 51 | 0.036 46 | 0.027 78 |
| | SAE | 0.031 11 | 0.051 59 | 0.193 68 | 0.119 11 | 0.101 21 |
| | DRML | 0.005 11 | 0.098 33 | 0.378 34 | 0.022 38 | 0.032 62 |
| | SDRL | 0.008 51 | 0.051 88 | 0.335 71 | 0.162 25 | 0.171 17 |
| | SSCD | 0.010 51 | 0.048 78 | 0.388 63 | 0.195 43 | 0.194 75 |
| Yeast | FastTag | 0.015 17 | 0.228 07 | 0.733 69 | 0.999 81 | 0.800 16 |
| | SAE | 0.282 75 | 0.119 31 | 0.328 15 | 0.325 28 | 0.284 82 |
| | DRML | 0.015 17 | 0.228 07 | 0.743 34 | 0.999 81 | 0.806 44 |
| | SDRL | 0.137 93 | 0.139 81 | 0.704 04 | 0.550 84 | 0.509 84 |
| | SSCD | 0.018 17 | 0.108 71 | 0.754 18 | 0.999 81 | 0.806 44 |

表 2 不同比例标签训练数据下各模型的性能

Table 2 Performance of each model under different proportion of label training data

| 标记数据集 | 基准方法 | 绝对匹配率 | 汉明损失 | 精确率 | 召回率 | F_1 |
|-------|---------|----------|----------|----------|----------|----------|
| 30% | FastTag | 0.008 01 | 0.012 53 | 0.570 39 | 0.067 01 | 0.097 05 |
| | SAE | 0.188 37 | 0.010 81 | 0.218 47 | 0.132 01 | 0.118 06 |
| | DRML | 0.008 01 | 0.012 83 | 0.189 41 | 0.077 51 | 0.077 25 |
| | SDRL | 0.016 03 | 0.011 82 | 0.601 41 | 0.264 97 | 0.301 82 |
| | SSCD | 0.030 06 | 0.006 89 | 0.631 41 | 0.320 63 | 0.341 51 |
| 20% | FastTag | 0.016 03 | 0.101 25 | 0.909 09 | 0.032 15 | 0.056 14 |
| | SAE | 0.112 22 | 0.010 99 | 0.136 79 | 0.101 52 | 0.073 47 |
| | DRML | 0 | 0.013 53 | 0 | 0 | 0 |
| | SDRL | 0.066 13 | 0.010 19 | 0.308 75 | 0.205 01 | 0.170 62 |
| | SSCD | 0.028 05 | 0.009 51 | 0.477 53 | 0.370 34 | 0.318 43 |
| 10% | FastTag | 0.012 60 | 0.012 60 | 0.320 06 | 0.090 82 | 0.108 91 |
| | SAE | 0.202 40 | 0.010 78 | 0.148 88 | 0.169 1 | 0.090 17 |
| | DRML | 0 | 0.011 53 | 0 | 0 | 0 |
| | SDRL | 0.038 07 | 0.010 16 | 0.167 60 | 0.179 24 | 0.173 22 |
| | SSCD | 0.026 05 | 0.005 31 | 0.417 67 | 0.195 31 | 0.225 81 |

4 500 个训练集上进行了 3 个实验。每次标价样本的训练数据集数目占比分别为 30%、20% 和 10%。其余的训练数据集被用作未标记数据集,其中测试集为 499 张图片。

从表 2 可以看出,当标记数据占总训练数据的 30% 时,本文模型在综合性能方面与基准模型拉开了一定的差距。当标记数据占总训练数据的 20% 时,SSCD 的准确率和 F_1 比其他表现最好的模型高 17% 和 14%。当标记数据进一步减少到 10% 时,本文模型在召回率和 F_1 上仍然表现良好,与性能最好的 SDRL 相比,增幅分别高达 2% 和 5%。同时,当标记数据量逐渐减少时,模型各指标的下降幅度低于其他模型,这也证明了 SSCD 具有更可靠的半监督学习能力能够应对逐步增大的半监督数据集。

3.3 结构消融实验

为了证明单一注意力机制是该框架下相对最有效的机制,对 3 种结构进行了对比测试(以 Yesat 数据集为例)。

对比结构 1:在 C1 和 C2 之后,分别计算自注意力机制,将注意力结果相加,通过最分类器预测。

对比结构 2:在 C1 和 C2 之后,两个分类器进行相互计算,将各自的计算结果相加,通过最终分类器预测。

对比结构 3:在 C1 和 C2 之后,两个分类器将各自的计算结果相加,通过最终分类器预测。

对比结构如表 3、图 3 ~ 图 5 所示。从表 3 中可以看出,当绝对匹配率和汉明损失与其他结构基本相同时,本文结构在剩余项指标上均优于其他结构。

表3 SSCD在不同结构下的性能

Table 3 Performance of SCD under different structures

| 基准方法 | 绝对匹配率 | 汉明损失 | 精确率 | 召回率 | F_1 |
|---------|----------|----------|----------|----------|----------|
| 单一注意力机制 | 0.018 17 | 0.108 71 | 0.754 18 | 0.999 81 | 0.806 44 |
| 对比结构1 | 0.015 17 | 0.108 71 | 0.733 69 | 0.984 71 | 0.800 16 |
| 对比结构2 | 0.015 17 | 0.108 71 | 0.733 69 | 0.982 31 | 0.800 16 |
| 对比结构3 | 0.355 11 | 0.337 85 | 0.743 34 | 0.195 76 | 0.044 13 |

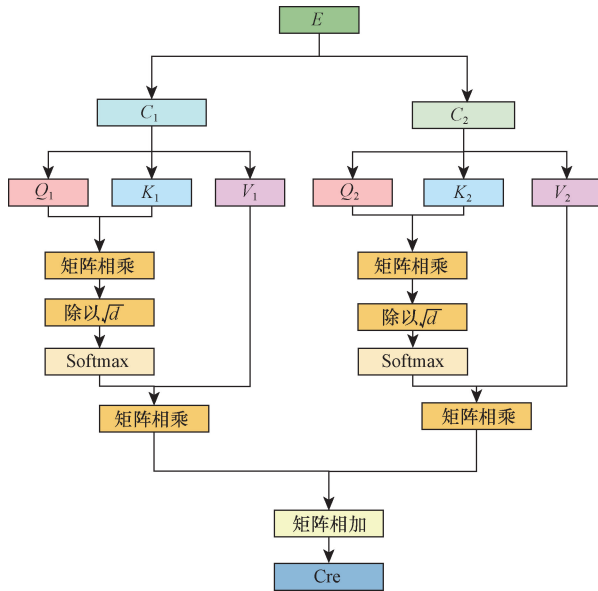


图3 对比结构1

Fig. 3 Comparison structure 1

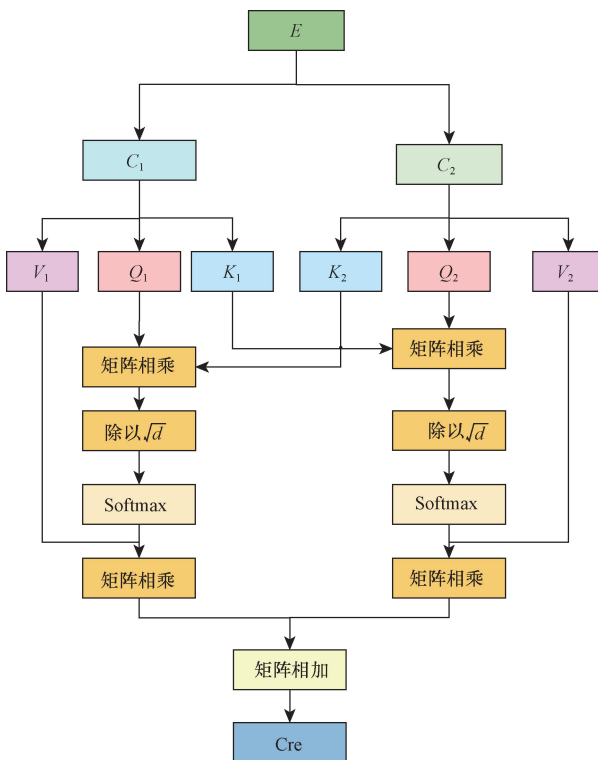


图4 对比结构2

Fig. 4 Comparison structure 2

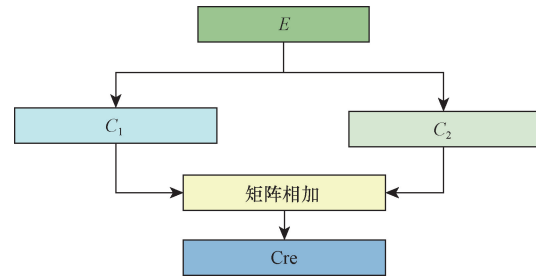


图5 对比结构3

Fig. 5 Comparison structure 3

因此可以佐证本文结构的有效性。根据结果分析可以得出,单一注意力机制优于其他结构的原因很大可能是标签之间的相关性信息含量有限,过于复杂的结构可能会导致模型的过拟合,从而导致模型在测试集上表现不佳。

3.4 模型训练花费时间对比

训练模型消耗的时间也是该模型能否在工业中应用的一个重要因素。以 CUB 数据集为例,比较了训练基准模型和 SSCD 所需的时间。从表 4 中可以看出,SSCD 在具有较好泛化性能的同时,在时间资源消耗方面具有较大优势。

表4 每个模型的训练时间

Table 4 Training time of each model

| 基准方法 | 训练花费时间/s |
|---------|----------|
| FastTag | 67 |
| SEA | 13 147 |
| DRML | 84 302 |
| SDRL | 21 284 |
| SSCD | 7 400 |

4 结论

针对多标签分类半监督学习存在的问题,提出了一种二元结构下的多标签半监督课程学习模型 SSCD。SSCD 旨在解决目前单标签领域的课程学习方案难以应用于多标签领域,在 COIN 下采用的半监督学习方案进行训练时间过长,以及现有模型中多个标签之间关系挖掘不足的问题。在 3 个数据集上与 4 个当前主流的基准方案进行了模型比较,实验结果表明,在训练时间和模型泛化性能的平衡上,SSCD 大大超过了基准方案。同时,通过消融实验证明了单注意力机制的有效性,并通过使用不同比例的标记数据训练进一步验证了模型的可靠性。

虽然模型目前取得了较好的效果,但在伪标记样本的选取准则上还有提高的空间,目前仅依据差异值选取加入训练集合的伪标记样本,未来考虑拓展评估标准,力求对加入训练集的伪标记样本做出更为全面的评估。

参 考 文 献

- [1] 武红鑫, 韩萌, 陈志强, 等. 监督和半监督学习下的多标签分类综述[J]. 计算机科学, 2022, 49(8): 12-25.
Wu Hongxin, Han Meng, Chen Zhiqiang, et al. Survey of multi-label classification based on supervised and semi-supervised learning [J]. Computer Science, 2022, 49(8): 12-25.
- [2] Wah C, Branson S, Welinder P, et al. The Caltech-UCSD Birds-200-2011 Dataset[J]. Computer Science and Biology, 2011(7): 16119123.
- [3] 宋宇婷, 余本功. 融合多粒度特征和标签语义共现的多标签分类[J]. 科学技术与工程, 2023, 23(16): 6959-6966.
Song Yuting, Yu Bengong. Multi-label classification with multi-granularity and label semantic co-occurrence features [J]. Science Technology and Engineering, 2023, 23(16): 6959-6966.
- [4] 张青, 王肖霞, 杨风暴. 基于 XLCLS 模型的法条多标签分类方法[J]. 科学技术与工程, 2022, 22(11): 4417-4423.
Zhang Qing, Wang Xiaoxia, Yang Fengbao. Research on multi-label classification method of law for public interest litigation cases [J]. Science Technology and Engineering, 2022, 22(11): 4417-4423.
- [5] Athitsos V, Sclaroff S. Boosting nearest neighbor classifiers for multi-class recognition[C]//IEEE Computer Society Conference on Computer Vision & Pattern Recognition. San Diego: IEEE, 2005: 45.
- [6] Wu B Y, Chen W D, Sun P, et al. Tagging like humans; diverse and distinct image annotation (supplementary material) [J]. Computer Science, 2018(8): 260497212.
- [7] Wu B Y, Jia F, Liu W. Multi-label learning with missing labels using mixed dependency graphs[J]. International Journal of Computer Vision, 2018, 126(8): 875-896.
- [8] Liu C, Zeng X, Wang K, et al. Multi-task learning for macromolecule classification, segmentation and coarse structural recovery in cryo-tomography[J]. ArXiv, 2018: 1805.06332.
- [9] Li R, Wang S, Zhu F, et al. Adaptive graph convolutional neural networks[J]. ArXiv, 2018: 1801.03226.
- [10] 郑伟, 王朝坤, 刘璋, 等. 一种基于随机游走模型的多标签分类算法[J]. 计算机学报, 2010, 33(8): 1418-1426.
Zheng Wei, Wang Chaokun, Liu Zhang, et al. A multi-label classification algorithm based on random walk model [J]. Chinese Journal of Computers, 2010, 33(8): 1418-1426.
- [11] 陆泉, 何超, 陈静, 等. 基于两阶段迁移学习的多标签分类模型研究[J]. 数据分析与知识发现, 2021, 5(7): 91-100.
Lu Quan, He Chao, Chen Jing, et al. A multi-label classification model with two-stage transfer learning [J]. Data Analysis and Knowledge Discovery, 2021, 5(7): 91-100.
- [12] Xu X, Deng C, Nie F. Adaptive graph weighting for multi-view dimensionality reduction[J]. Signal Processing, 2019, 165: 186-196.
- [13] Li K, Liu C, Zhao H, et al. ECACL: a holistic framework for semi-supervised domain adaptation[J]. ArXiv, 2021: 2104.09136.
- [14] Lee D H. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks[J]. Computer Science, 2013(7): 18507866.
- [15] Cascante-Bonilla P, Tan F, Qi Y, et al. Curriculum labeling: revisiting pseudo-labeling for semi-supervised learning. [C]//National Conference on Artificial Intelligence. New York: AAAI, 2021: 6912-6920.
- [16] Chang W C, Yu H F, Zhong K, et al. Taming pretrained transformers for extreme multi-label text classification [J]. ArXiv, 2019: 1905.02331.
- [17] Chen M, Zheng A, Weinberger K Q. Fast image tagging[C]//30th International Conference on Machine Learning. Atlanta Georgia: ICML, 2013: 2311-2319.
- [18] Ma J H, Chow T W S. Robust non-negative sparse graph for semi-supervised multi-label learning with missing labels [J]. Information Sciences, 2018(8): 336-351.
- [19] Xu Z, Kusner M J, Weinberger K Q, et al. Classifier Cascades and trees for minimizing feature evaluation cost [J]. Journal of Machine Learning Research, 2014, 15: 2113-2144.
- [20] Kong X, Ng M K, Zhou Z H. Transductive multilabel learning via label set propagation [J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 25(3): 704-719.
- [21] Li Y, Hu L, Gao W. Robust sparse and low-redundancy multi-label feature selection with dynamic local and global structure preservation [J]. Pattern Recognition, 2022, 134: 109120.
- [22] Sun L, Feng S, Lu G, et al. Robust semi-supervised multi-label learning by triple low-rank regularization [C]//Pacific-Asia Conference on Knowledge Discovery & Data Mining. Berlin: Springer, Cham, 2019: 269-280.
- [23] Jing L, Yang L, Yu J, et al. Semi-supervised low-rank mapping learning for multi-label classification [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 1483-1491.
- [24] Gong C, Tao D, Yang J, et al. Teaching-to-learn and learning-to-teach for multi-label propagation [C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix Arizona: AAAI, 2016: 1610-1616.
- [25] Liu D, Yan S, Hua X S, et al. Image retagging using collaborative tag propagation [J]. IEEE Transactions on Multimedia, 2011, 13(4): 702-712.
- [26] Zhan W, Zhang M L. Inductive semi-supervised multi-label learning with co-training [C]//Knowledge Discovery and Data Mining. Nanjing: ACM, 2017: 1220-1229.
- [27] Chu Z, Li P, Hu X. Co-training based on semi-supervised ensemble classification approach for multi-label data stream [C]//2019 IEEE International Conference on Big Knowledge (ICBK). Beijing: IEEE, 2019: 58-65.
- [28] Wang L, Liu Y, Qin C, et al. Dual relation semi-supervised multi-label learning [C]//National Conference on Artificial Intelligence. New York: AAAI, 2020: 6227-6235.
- [29] Broadwater J B, Chellappa R. Adaptive threshold estimation via extreme value theory [J]. IEEE Transactions on Signal Processing: A Publication of the IEEE Signal Processing Society, 2010, 58(2): 490-500.
- [30] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv, 2014: 1409.1556.
- [31] 高晗, 田育龙, 许封元, 等. 深度学习模型压缩与加速综述 [J]. 软件学报, 2021, 32(1): 68-92.
Gao Han, Tian Yulong, Xu Fengyuan, et al. Survey of deep learning model compression and acceleration [J]. Journal of Software, 2021, 32(1): 68-92.
- [32] Li S, Lin Y H, Zhou K. Semantic adversarial autoencoder for zero-shot learning [J]. ArXiv, 2024: 1704.08345.
- [33] Wang L, Liu Y, Di H, et al. Semi-supervised dual relation learning for multi-label classification [J]. IEEE Transactions on Image Processing, 2021, 30: 9125-9135.