

一种基于直推判别字典学习的零样本分类方法*

冀中, 孙涛, 于云龙



(天津大学 电气自动化与信息工程学院, 天津 300072)

通讯作者: 于云龙, E-mail: yuyunlong@tju.edu.cn

摘要: 零样本分类的目标是对训练阶段未出现过的类别的样本进行识别和分类,其主要思路是,借助类别语义信息,将可见类别的知识转移到未见类别中.提出了一种直推式的字典学习方法,包含以下两个步骤:首先,提出一个判别字典学习模型,对带标签的可见类别样本的视觉特征和类别语义特征建立映射关系模型;然后,针对可见类别和未见类别不同引起的域偏移问题,提出了一个基于直推学习的修正模型.通过在3个基准数据集(AwA, CUB和SUN)上的实验结果,证明了该方法的有效性和先进性.

关键词: 零样本分类;图像分类;字典学习;直推学习
中图分类号: TP18

中文引用格式: 冀中,孙涛,于云龙.一种基于直推判别字典学习的零样本分类方法.软件学报,2017,28(11):2961-2970. <http://www.jos.org.cn/1000-9825/5338.htm>

英文引用格式: Ji Z, Sun T, Yu YL. Transductive discriminative dictionary learning approach for zero-shot classification. Ruan Jian Xue Bao/Journal of Software, 2017, 28(11): 2961-2970 (in Chinese). <http://www.jos.org.cn/1000-9825/5338.htm>

Transductive Discriminative Dictionary Learning Approach for Zero-Shot Classification

Ji Zhong, SUN Tao, YU Yun-Long

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: Zero-Shot classification aims at recognizing instances from unseen categories that have no training instances in the training stage. To address this task, most existing approaches resort to class semantic information to transfer knowledge from the seen classes to the unseen ones. In this paper, a transductive dictionary learning approach is proposed to facilitate the task in two steps. A discriminative dictionary learning model is first proposed for constructing the relations between the visual modality and the class semantic modality with the labeled seen instances. Then a transductive modified model is used to alleviate the domain shift issue caused by the disjointness between the seen classes and the unseen classes. Experimental results on three benchmark datasets (AwA, CUB and SUN) demonstrate the effectiveness and superiority of the proposed approach.

Key words: zero-shot classification; image classification; dictionary learning; transductive learning

传统的视觉分类方法无法对未见类别(训练阶段未出现过的类别)的样本进行识别和分类,而人类却能根据类别的描述信息辨认出未见过类别的样本.例如,假设一个小孩从未见过斑马(斑马对小孩来说是未见类),但他知道斑马是一种身上有黑白条纹的外形像马的动物(马对小孩来说是可见类).那么当他第一次看到斑马的时候,一般都能认出这是斑马.受此启发,零样本分类利用类别的语义描述信息将知识从可见类中转移到未见类中,以实现未知类别的预测和分类,在学术界引起了广泛的关注^[1-3].

现有的零样本分类方法大多借助类别的语义特征将可见类别的信息转移到未见类别中,以实现分类任务.

* 基金项目: 国家自然科学基金(61771329, 61472273)

Foundation item: National Natural Science Foundation of China (61771329, 61472273)

本文由复杂环境下的机器学习研究专刊特约编辑胡清华教授推荐.

收稿时间: 2017-03-16; 修改时间: 2017-06-16, 2017-08-23; 采用时间: 2017-09-06

目前,常用的类别语义特征为属性特征^[1-6]和词向量特征^[2,7-10].属性特征主要是通过人为定义或学习而得到的一些类别间共享的属性,例如“颜色”“纹理”“形状”等.类别的词向量特征利用自然语言处理技术从语料库中抽取,每个类别都可以用词向量空间中的一个向量来表示.与属性特征相比,词向量特征是以无监督学习的方式从语料库中提取语义信息,因此具有较好的扩展性.

零样本分类主要是通过构建视觉和语义间的映射关系实现^[2,3,6,8,11,12],这通常包括以下3个步骤:(1) 将可见类别标签嵌入到类别语义空间中;(2) 学习一个从视觉空间到类别语义空间的映射函数;(3) 利用学习到的映射函数将未见类别样本视觉特征映射到同一类别语义空间,在类别语义空间中,用最近邻方法完成对未见类别样本的分类.因此,如何构建一个有效的映射关系模型是实现零样本有效分类的重要因素.此外,在步骤(3)中,训练类别和测试类别的不同导致可见类别和未见类别的数据分布不同,因此,从视觉空间到语义空间的映射函数也不同.将从训练集学习到的模型直接应用于测试集,会导致测试集样本在语义空间的映射与真实类别语义存在较大的偏差,这种现象称为域偏移^[5].如何缓解域偏移问题,也是零样本分类中的关键研究方向之一.

针对上述两点,本文分别提出了相应的解决方案.为了提高映射关系的有效性,本文提出首先将类别语义特征映射到一个能够保持语义信息的隐空间中,从而减少语义特征存在的冗余信息;然后结合字典学习方法利用隐空间中映射特征对视觉特征进行重构.传统的基于视觉特征映射的方法是将视觉特征映射到类别语义空间,与这类方法不同,本文所提出的方法是将类别语义特征在隐空间中的向量表征嵌入到视觉空间中,是一种反向嵌入的过程.本文称其为判别字典学习方法(discriminative dictionary learning,简称 DDL).针对域偏移问题,在DDL的基础上,进一步提出了一种直推式DDL方法(transductive discriminative dictionary learning,简称 TDDL):首先,利用训练阶段学习到的模型预测未见类别样本的类别标签从而得到其预测类别语义特征;然后利用已预测的未见类别样本重新优化模型,从而提高模型对未见类别样本的分类能力.本文方法的流程图如图1所示.

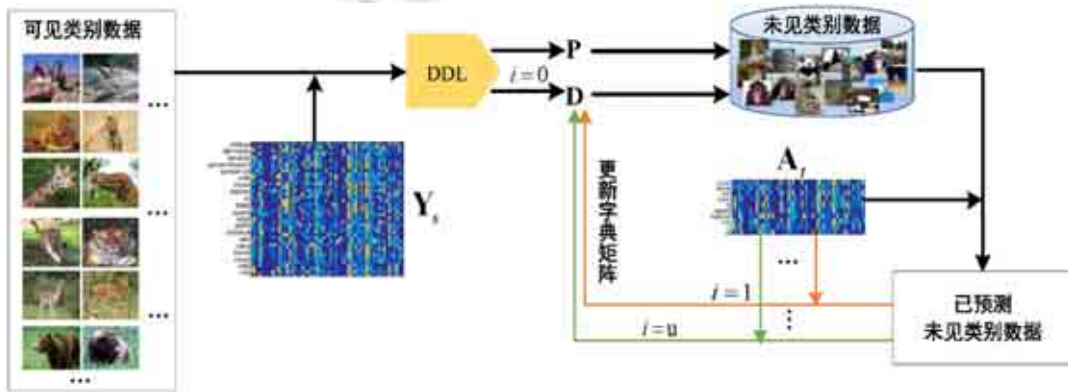


Fig.1 Flowchart of the TDDL approach for zero-shot classification

图1 基于TDDL方法的零样本分类方法流程图

图1中, Y_s 表示可见类别样本的类别语义特征, A_u 表示未见类别的类别语义特征, P 表示转移矩阵, D 表示字典矩阵, u 是迭代次数, $i=0$ 时利用可见类别样本学习 P 和 D ,当 $i=1$ 时,利用已预测的未见类别样本迭代更新模型,以学习一个更普适的字典矩阵 D ,用于未见类别样本的分类.

本文的主要贡献如下:

- (1) 提出了一种用于构建零样本分类中映射关系模型的判别字典学习方法 DDL.与传统字典学习模型相比,该方法将稀疏约束转换成一种线性描述,并利用类别语义特征在隐空间中的映射特征对原始视觉特征进行重构,显著提高了训练效率.
- (2) 进一步提出了一种用于修正零样本分类中的域偏移问题的直推式判别字典学习方法 TDDL.该方法通过利用预测的测试集分类结果对学习到的模型进行重新优化,从而学习得到一个更具普适性的映

射关系.实验结果表明,此直推学习框架可以显著提升零样本分类性能,并优于现有的相关方法.

本文第 1 节简要介绍相关工作.第 2 节具体描述本文提出的方法.第 3 节通过大量对比实验验证本文提出的方法的性能.第 4 节总结全文.

1 相关工作

字典学习模型通过求解字典矩阵以及样本在字典矩阵下的编码系数得到样本的稀疏表示,并广泛应用于图像分类、人脸识别等领域^[13,14].零样本分类中的视觉特征映射可以看作是字典学习问题,将类别语义特征看作样本的稀疏表示,从而学习映射矩阵将视觉特征(高维)映射到类别语义空间(低维),转化为求解字典矩阵以计算样本的稀疏表示.传统的基于字典学习的零样本分类方法^[15-17]利用可见类别训练样本的类别语义特征,学习一个有效的数据表征模型,然后对未见类别样本进行分类.例如,Kodirov 等人^[16]提出了一个字典学习的框架,将类别语义特征看作是重建样本视觉特征的编码参数.其未见类别的字典通过正则稀疏编码学习得到,从可见类别学习到的字典作为其正则项.Zhang 等人^[17]基于字典学习的框架提出了一个隐空间概率模型,将样本的视觉特征和类别语义特征同步映射到不同的隐空间中,然后学习一个跨空间的相似矩阵,对不同空间的特征进行匹配.

零样本分类方法可分为归纳零样本分类方法和直推零样本分类方法.在归纳零样本学习中,仅用可见类别的训练样本训练分类模型,然后再利用训练所得的分类模型对未见类别样本进行分类.这类方法实际上就是建立映射关系模型,因此接下来本文称这类方法为基于归纳的方法.例如,Lampert 等人^[1]提出了两种实现零样本学习的概率转移模型:直接属性预测(direct attribute prediction,简称 DAP)和间接属性预测(Indirect attribute prediction,简称 IAP).其中,

- DAP 是直接学习建立视觉样本与属性特征中每一种属性之间关系的模型,并利用所学习到的模型预测测试样本的属性特征.
- IAP 利用可见类别将测试样本和未见类别联系起来,其过程可以分成两步:首先,将测试样本分到可见类别中;然后,利用可见类别和未见类别之间的关系对测试样本进行预测.其中,可见类别和未见类别之间的语义关系是通过类别之间的语义描述建立起来的.

由于可见类别和未见类别不相交,归纳零样本学习方法存在域偏移问题.直推零样本学习方法^[18]可以有效地解决域偏移问题并且提升模型的分类表现.例如,Li 等人^[19]提出了一个半监督框架,利用先验知识学习一个新的类别语义空间.现有很多直推零样本学习方法^[16,19]利用未见类别样本重新定义映射函数.通过利用基于图的标签拓展方法来挖掘未见类别样本的流型结构,Rohrbach 等人^[20]提出了一种复杂的直推零样本学习方法.与上述方法不同,Xu 等人^[21]提出了一种数据增强方法,通过增加辅助数据到可见类别训练一个更普适的分类模型.自训练修正方法^[22]是一种后处理技术,其主要的思路是,根据所有未见类别的样本在语义空间中的分布修正未见类别的类别语义嵌入特征.

2 TDDL 方法

2.1 问题定义

$S=\{X_s, Y_s, z_s\}$ 和 $T=\{X_t, Y_t, z_t\}$ 分别表示来自 M 个可见类别的 m 个样本和来自 N 个未见类别的 n 个样本,其中, $X_s=[x_1^s, \dots, x_m^s] \in \mathbb{R}^{d \times m}$ 和 $X_t=[x_1^t, \dots, x_n^t] \in \mathbb{R}^{d \times n}$ 分别表示可见类别和未见类别样本的视觉特征, d 是视觉特征的维度; $Y_s=[y_1^s, \dots, y_m^s] \in \mathbb{R}^{p \times m}$ 和 $Y_t=[y_1^t, \dots, y_n^t] \in \mathbb{R}^{p \times n}$ 分别表示可见类别和未见类别样本的类别语义特征, p 是类别语义特征的维度; $z_s \in \mathbb{R}^m$ 和 $z_t \in \mathbb{R}^n$ 分别表示可见类别和未见类别样本的类别标签,并且有 $z_s \cap z_t = \emptyset$; $A_t=[a_1^t, \dots, a_n^t] \in \mathbb{R}^{p \times n}$ 表示未见类别的类别语义特征.

零样本分类的目标是,对未见类别的样本 X_t 进行预测和分类.

2.2 DDL方法

在零样本分类中,由于样本的类别语义特征之间存在冗余信息,直接将类别语义特征作为样本的编码系数的字典学习方法存在训练效率低的问题.为了提高类别语义信息的判别性,首先将类别语义特征映射到一个能够保持语义信息的隐空间中:

$$\min_{P, c_i} \sum_{i=1}^m \|Py_i^s - c_i\|_2^2 \quad (1)$$

其中, P 为转移矩阵, y_i^s 为可见类别样本的类别语义特征, c_i 为 y_i^s 在隐空间中的特征表征.

字典学习的目标是通过求解字典矩阵,学习样本在字典矩阵下的编码系数.为了提高模型的可预测性,传统字典学习方法对编码系数进行稀疏性约束.DDL 将样本类别语义特征在隐空间中的映射特征作为样本的编码系数,不需要传统字典学习模型中的稀疏约束,而是将稀疏约束转换为一种线性描述,利用样本的视觉特征和映射特征学习字典矩阵,其目标函数可以表示为

$$\min_{D_s, P, c_i} \sum_{i=1}^m \|x_i^s - D_s c_i\|_2^2 + \lambda \|PY_s - C_s\|_F^2 + \|d_i\|_2^2 \quad (2)$$

其中, D_s 为可见类别的字典矩阵, λ 为用于平衡目标函数前后两项表达式之间重要性的参数.目标函数(2)的矩阵形式为

$$\min_{D_s, P, C_s} \|X_s - D_s C_s\|_F^2 + \lambda \|PY_s - C_s\|_F^2 + \|d_i\|_2^2 \quad (3)$$

其中, $C_s = [c_1^s, \dots, c_m^s] \in \mathbb{R}^{q \times m}$ 表示 m 个可见类别样本的类别语义特征在隐空间中的映射特征, q 为隐空间的维度.

目标函数(3)中,第1项是用字典矩阵 D_s 和样本在隐空间的映射矩阵 C_s 的线性组合尽可能地还原样本的视觉特征 X_s ,第2项是将样本的类别语义特征 Y_s 映射到一个公共隐空间中,从而得到更具有判别性的映射特征 C_s .目标函数(3)的目标是利用训练样本学习一个能够将类别语义特征 Y_s 映射到更具判别性的隐空间中的转移矩阵 P 和一个能够将隐空间中的映射特征 C_s 重构为视觉特征 X_s 的字典矩阵 D_s .

同时,相对于 D_s, P 和 C_s 三者,目标函数(3)不是一个凸函数;但分别相对于 D_s, P 和 C_s ,目标函数(3)是一个凸函数.因此,可以利用迭代优化算法求解目标函数(3),其优化过程分为以下3步.

- 固定字典矩阵 D_s 和转移矩阵 P ,对隐空间中的映射特征 C_s 进行更新:

$$C_s^* = \arg \min_{C_s} \|X_s - D_s C_s\|_F^2 + \lambda \|PY_s - C_s\|_F^2 \quad (4)$$

目标函数(4)是一个最小二乘优化问题,方程的解为

$$C_s^* = (D_s^T D_s + \lambda I)^{-1} (D_s^T X_s + \lambda P Y_s) \quad (5)$$

- 固定映射矩阵 C_s ,对转移矩阵 P 进行更新:

$$P^* = \arg \min_P \lambda \|PY_s - C_s\|_F^2 \quad (6)$$

方程的解为

$$P^* = \lambda C_s Y_s^T (\lambda Y_s Y_s^T + I)^{-1} \quad (7)$$

- 固定映射矩阵 C_s 和转移矩阵 P ,对字典矩阵进行更新:

$$D_s^* = \arg \min_{D_s} \|X_s - D_s C_s\|_F^2 + \|d_i\|_2^2 \quad (8)$$

利用交替方向乘法求解目标函数(8).

求解出 D_s^* 和 P^* 后,利用转移矩阵 P^* 求出未见类别在隐空间中的映射矩阵,并利用字典矩阵 D_s^* 对未见类别的视觉特征进行重构.

$$B_i = D_s^* \times P^* \times A_i \quad (9)$$

其中, $B_i = [b_1^i, \dots, b_N^i] \in \mathbb{R}^{d \times N}$ 表示 N 个未见类别的重构视觉特征.

最后,通过分别计算未见类别样本 x_i' 与 N 个未见类别的重构视觉特征 b_j^i 的相似度实现对未见类别样本的

分类,即

$$j^* = \max_j ((x_i^t)^T \times b_j^t) \tag{10}$$

判别字典学习方法的算法流程图如图 2 所示.

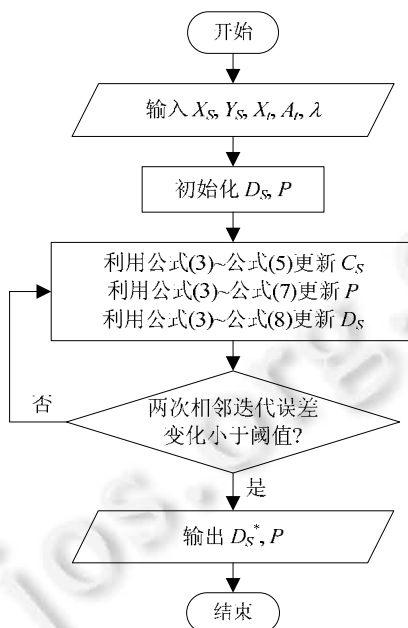


Fig.2 Algorithm flowchart of the DDL approach
图 2 DDL 方法的算法流程图

2.3 TDDL方法

为了修正零样本分类中存在的域偏移问题,本文在 DDL 方法的基础上进一步提出了 TDDL 方法,除了利用可见类别样本学习模型以外,还利用未见类别样本对模型进行修正.TDDL 方法首先利用 DDL 方法对未见类别样本进行预测,获得初始的分类结果,从而求出未见类别样本的预测类别语义特征 Y ,然后利用初始的预测结果对分类模型进行修正,即

$$\min_{D_i, C_i} \| X_i - D_i C_i \|_F^2 + \mu \| P Y - C_i \|_F^2, \| d_i \|_2^2 = 1 \tag{11}$$

其中, X_i 表示未见类别样本的视觉特征, P 表示利用可见类别的标注样本学习到的转移矩阵, D_i 表示未见类别的字典矩阵, C_i 表示未见类别样本在隐空间中的映射特征.

利用迭代优化算法求解目标函数(11),其优化过程分以下两步进行.

- 固定字典矩阵 D_i ,对映射特征 C_i 进行更新:

$$C_i^* = \arg \min_{C_i} \| X_i - D_i C_i \|_F^2 + \mu \| P Y - C_i \|_F^2, \| d_i \|_2^2 = 1 \tag{12}$$

方程的解为

$$C_i^* = (D_i^T D_i + \mu I)^{-1} (D_i^T X_i + \mu P Y) \tag{13}$$

- 固定映射特征 C_i ,对字典矩阵 D_i 进行更新:

$$D_i^* = \arg \min_{D_i} \| X_i - D_i C_i \|_F^2, \| d_i \|_2^2 = 1 \tag{14}$$

利用交替方向乘法求解目标函数(14).

求解出 D_i^* 后,按照公式(9)和公式(10)重新预测未见类别样本的类别标签,从而得到未见类别样本的预测类

别语义特征,利用未见类别样本重新修正分类模型.重复上述预测和修正过程(收敛条件为:两次相邻迭代目标函数的变化之差小于设定的阈值,如 0.01.).基于直推判别字典学习方法的具体步骤见算法 1.

算法 1. 基于直推判别字典学习方法(TDDL)的步骤.

输入:可见类别样本的视觉特征 X_v 和类别语义特征 Y_v ,未见类别样本的视觉特征 X_u ,未见类别的类别语义特征 A_u ,参数 λ 和 μ .

输出:未见类别样本的预测类别标签.

训练阶段:

1. 初始化 D_s 和 P ;
2. **repeat**
3. 利用公式(5)更新 C_s ;
4. 利用公式(7)更新 P ;
5. 利用公式(8)更新 D_s ;
6. **until** 两次相邻迭代目标函数(2)的变化之差小于设定的阈值.
7. 返回 D_s^* 和 P^* ;
8. 固定转移矩阵 P 不变,初始化未见类别字典矩阵 $D_u^* = D_s^*$;
9. **repeat**
10. 按照公式(9)和公式(10)预测未见类别样本的类别标签;
11. 根据未见类别样本的预测类别标签求出未见类别样本的预测类别语义特征 Y_u ;
12. 按照公式(13)更新 C_u ;
13. 按照公式(14)更新 D_u ;
14. **until** 两次相邻迭代目标函数(11)的变化之差小于设定的阈值.
15. 返回 D_u^* .

测试阶段:

16. 按照公式(9)和公式(10),预测未见类别样本的类别标签.

2.4 算法的收敛性和计算复杂性

分别分析我们所提出的 DDL 方法的收敛性和 TDDL 方法的计算复杂性.

在 AwA 数据集上利用属性特征作为类别语义特征,利用交叉验证法选取参数 λ 的值为 0.01,如图 3 所示,DDL 方法的目标函数(2)经过 10 次迭代优化已收敛,这说明 DDL 方法收敛迅速.

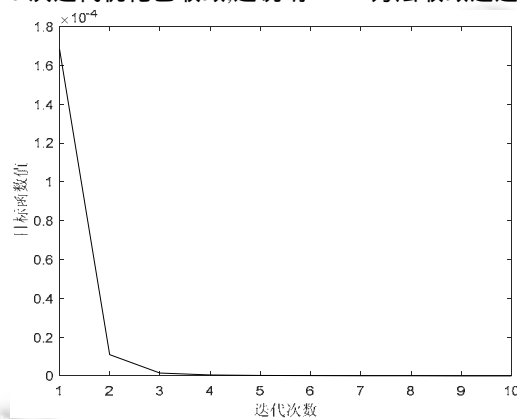


Fig.3 Convergence curve of DDL on the AwA dataset with attribute features

图 3 DDL 方法利用属性特征在 AwA 数据集上的收敛曲线

DDL方法在训练阶段迭代优化 D, P 和 C_s ,每次优化 D, P 和 C_s 的计算复杂度分别为 $O(qmp+p^3+p^2q), O(qdm+q^3+q^2m)$ 和 $O(K(dmq+q^3+q^2d+d^2q))$,其中, K 为交替方向乘子法迭代次数.在直推学习中,迭代优化 D_i 和 C_i ,每次优化 D_i 和 C_i 的计算复杂度分别为 $O(qdn+q^3+q^2n)$ 和 $O(K(dnq+q^3+q^2d+d^2q))$.假定 $M \ll q, p \ll q$ 和 m, d, n, q 为同一数量级,且算法收敛迅速,TDDL方法的计算复杂度为 $O(q^3)$.

3 实验

3.1 实验数据集及实验参数

实验数据集:DDL和TDDL方法分别在以下3个数据集上进行实验:Animals with Attributes(AwA)^[1], Caltech-UCSD Bird2011(CUB)^[23]和SUN Attribute Dataset^[24].其中,AwA包含来源于50个类别的30475张动物图片,每类动物用一个85维的属性向量表示;CUB是一个细粒度的数据集,包含来源于200个类别的11788张鸟类图片,每个类别用一个312维的属性向量表示;SUN包含717个用102维的属性向量表示的场景类别,其中每个类别有20张图片.在可见/未见类别的划分上,对于AwA数据集,利用标准的40/10划分方法^[1];对于CUB数据集,利用与文献[3]一样的150/50划分方法;对于SUN数据集,利用与文献[17]相同的划分方法,即将707个类别作为可见类别,10个类别作为未见类别.

特征:DDL和TDDL方法利用上述数据集提供的VGG-verydeep-19特征作为样本的视觉特征.对于类别语义特征,在AwA和CUB数据集上,分别利用属性特征和词向量特征作为类别语义特征;在SUN数据集上,由于对比算法很少提取其词向量特征,因此仅利用属性特征作为类别语义特征.

实验参数:DDL和TDDL方法中有两个参数 λ 和 μ ,实验中利用交叉验证法选择参数,从 $\{0.01, 0.1, 1, 10, 100\}$ 中选取参数 λ 和 μ 的最优值.

3.2 DDL实验结果与比较

在此实验中,分别利用属性特征和词向量特征作为类别的语义特征验证DDL方法的性能.首先利用属性特征作为类别语义特征,并选择4种基于归纳学习的零样本分类方法进行比较,分别为Direct Attribute Prediction(DAP)^[1],Structural Joint Embedding(SJE)^[2],Latent Embeddings(LatEm)^[7]和Synthesized Classifiers(SC)^[4].对比算法的实验结果来源于原始论文,实验结果见表1,其中,“F”,“V”和“G”分别表示视觉特征、VGG特征和GoogleNet特征,“-”表示在原始论文中对比算法没有在此数据集上进行实验.

Table 1 Experimental comparison of DDL and other approaches with attribute features

表1 利用属性特征,DDL与其他方法的实验比较

方法	F	AwA (%)	CUB (%)	SUN (%)
DAP ^[6]	V	57.5	-	72.0
SJE ^[2]	G	66.7	50.1	-
LatEm ^[7]	G	71.9	45.5	-
SC ^[4]	G	72.9	54.7	-
DDL	V	77.3	52.6	86.0

从表1可以看出,与对比算法相比,DDL方法在AwA和SUN数据集上都取得了比较好的实验结果.具体为:在AwA数据集上,DDL达到了77.3%的分类准确率,优于所有对比算法,比经典DAP提高了19.8%;在SUN数据集上,DDL达到了86.0%的分类准确率,比经典DAP提高了14.0%;在CUB数据集上,优于除SC外的其他对比算法.文献[1]利用流行结构实现视觉特征到语义空间的映射,与DDL相比,SC是一个更加复杂的模型,并且SC利用了更强大的图像特征,即GoogleNet特征.

然后,利用词向量特征作为类别语义特征,并选择3种基于词向量特征的方法进行比较,分别为Canonical Correlation Analysis(CCA)^[25],Structural Joint Embedding(SJE)^[2]和Latent Embeddings(LatEm)^[7],见表2.与对比算法相比,DDL在AwA数据集上的表现有了很大的提升.具体为:DDL相对于CCA,SJE和LatEm分别提高了10%,24.4%和14.5%.同时,在CUB数据集上,DDL表现一般,但只比对比算法中表现最好的低3.3%.

Table 2 Experimental comparison of DDL and other approaches with word vector features**表 2** 利用词向量特征,DDL 与其他方法的实验比较

方法	F	AwA (%)	CUB (%)
CCA ^[25]	V	65.6	30.4
SJE ^[2]	G	51.2	28.4
LatEm ^[7]	G	61.1	31.8
DDL	V	75.6	28.5

3.3 TDDL实验结果与比较

由于基于直推学习的相关方法的实验多集中在 AwA 和 CUB 数据集上,且多利用属性特征,故本实验使用相同的实验设置验证 TDDL 方法的性能,并选择 3 种主流的基于直推学习的方法进行比较,分别为 Transductive Multi-view Hypergraph Label Propagation(TMV-HLP)^[18],Shared Model Space Learning(SMS)^[26]和 Unsupervised Domain Adaptation(UDP)^[16].TMV-HLP 和 SMS 都利用可见类别样本和未见类别样本共同学习一个普适的模型.UDP 和 TDDL 都利用未见类别样本重新学习分类模型,不同的是,文献[16]是一个无监督模型,利用未见类别样本的结构信息,TDDL 利用未见类别样本在隐空间中的映射分布来更新模型.实验结果见表 3,从表中可以看出,TDDL 方法明显优于其他对比算法.具体为:TDDL 在 AwA 数据集上达到了 93.3%的分类准确率,相对于 TMV-HLP,SMS 和 UDP 分别提高了 12.8%,14.8%和 17.7%;在 CUB 数据集上,TDDL 达到了 55.3%的分类准确率,相对于 TMV-HLP 提高了 7.4%.

Table 3 Experimental comparison of TDDL and other transductive approaches with attribute features (%)**表 3** 利用属性特征,TDDL 与其他直推方法的实验比较

(%)

方法	AwA	CUB
TMV-HLP ^[18]	80.5	47.9
SMS ^[26]	78.5	-
UDP ^[16]	75.6	40.6
TDDL	93.3	55.3

最后,将 DDL 方法和 TDDL 方法进行比较,比较结果见表 4,其中,A 和 W 分别表示属性特征和词向量特征.从表中可以看出,在 AwA 数据集上分别利用属性特征和词向量特征作为类别语义特征,TDDL 相对于 DDL 分别提高了 16.0%和 16.3%.由此得出,TDDL 通过利用已预测的未见类别样本修正模型,显著提高了模型对未见类别样本的分类能力.在 CUB 数据集上,分别利用属性特征和词向量特征作为类别语义特征,TDDL 相对于 DDL 分别提高了 2.7%和 2.4%,提高程度不明显.原因是 DDL 的分类准确率比较低,导致很多被错误预测的未见类别样本去修正模型,进而导致 TDDL 对 DDL 的提升效果不明显.从表中还可以看出,利用属性特征作为类别语义特征,DDL 方法性能表现更好.原因可能是词向量特征利用无监督方式学习得到,没有人为标注信息.

Table 4 Experimental comparison of DDL and TDDL (%)**表 4** DDL 和 TDDL 的实验比较

(%)

方法	AwA		CUB	
	A	W	A	W
DDL	77.3	75.6	52.6	28.5
TDDL	93.3	91.9	55.3	30.9

3.4 方法训练时间

在一台 CPU 频率为 3.20GHz、安装内存为 8GB 的计算机上,DDL 方法和 TDDL 方法的训练时间见表 5,其中,A 和 W 分别表示属性特征和词向量特征.从表中可以看出,DDL 方法和 TDDL 方法运行非常高效.

Table 5 Training times of DDL and TDDL (s)

表 5 DDL 和 TDDL 的训练时间 (s)

方法	AwA		CUB	
	A	W	A	W
DDL	29.97	29.49	50.46	50.10
TDDL	47.83	56.34	59.00	60.82

3.5 讨论

TDDL 是在 DDL 方法获得比较好的分类效果的前提下,为进一步增强分类效果而提出的.实验结果表明:DDL 分类准确率较高时,TDDL 提高程度明显(DDL 在 AwA 数据集上达到了 75%的分类准确率,TDDL 相比于 DDL 提高了 16%);DDL 分类准确率较低时,TDDL 提高程度不明显(将属性特征作为类别语义特征,DDL 在 CUB 数据集上的分类准确率为 55.3%,TDDL 相对于 DDL 只提高了 2.7%).因此,TDDL 的效果取决于 DDL 的分类效果.如何提高 DDL 的分类精度,使得 TDDL 有效,是一个可研究的方向.在未来的工作中,可以对 DDL 方法的隐空间增加正交限制,进一步去除类别语义特征之间的相关性,提高 DDL 的分类精度,从而提高 TDDL 的有效性.

4 结论

本文首先提出了一种用于构建零样本分类中映射关系的判别字典学习模型(DDL).该模型首先将类别语义特征映射到一个更具有判别性的隐空间中,并结合字典学习方法利用隐空间中的映射特征对原始视觉特征进行重构.实验结果表明,与几种先进归纳学习方法相比,DDL 方法在 3 个标准数据集(AwA,CUB 和 SUN)上都取得了比较好的实验结果,证实了 DDL 方法在建立视觉描述和类别语义描述之间关联关系的有效性.然后,针对零样本分类中的域偏移问题,本文进一步提出了一种直推式学习方法(TDDL).其主要思想是:首先,利用可见类别样本学习分类模型;然后,利用已预测的未见类别样本不断修正模型.实验结果表明,TDDL 可以显著提升归纳分类方法的性能.与几种先进直推学习方法相比,TDDL 在 AwA 和 CUB 数据集上都取得了比较好的实验结果.

References:

- [1] Lampert CH, Nickisch H, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2009. 951–958. [doi: 10.1109/CVPR.2009.5206594]
- [2] Akata Z, Reed S, Walter D, Lee H, Schiele B. Evaluation of output embeddings for fine-grained image classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 2927–2936. [doi: 10.1109/CVPR.2015.7298911]
- [3] Akata Z, Perronnin F, Harchaoui Z, Schmid C. Label-Embedding for attribute-based classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2013. 819–826. [doi: 10.1109/CVPR.2013.111]
- [4] Changpinyo S, Chao WL, Gong B, Sha F. Synthesized classifiers for zero-shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 5327–5336. [doi: 10.1109/CVPR.2016.575]
- [5] Fu Y, Hospedales TM, Xiang T, Fu Z, Gong S. Transductive multi-view embedding for zero-shot recognition and annotation. In: Proc. of the European Conf. on Computer Vision. 2014. 584–599. [doi: 10.1007/978-3-319-10605-2_38]
- [6] Lampert CH, Nickisch H, Harmeling S. Attribute-Based classification for zero-shot visual object categorization. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2014,36(3):453–465. [doi: 10.1109/TPAMI.2013.140]
- [7] Xian Y, Akata Z, Sharma G, Nguyen Q, Hein M, Schiele B. Latent embeddings for zero-shot classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 69–77. [doi: 10.1109/CVPR.2016.15]
- [8] Frome A, Corrado GS, Shlens J, Bengio S, Dean J, Ranzato MA, Mikolov T. Devise: A deep visual-semantic embedding model. In: Proc. of the Advances in Neural Information Processing Systems. 2013. 2121–2129.
- [9] Elhoseiny M, Saleh B, Elgammal A. Write a classifier: Zero-Shot learning using purely textual descriptions. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2013. 2584–2591. [doi: 10.1109/ICCV.2013.321]

- [10] Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 1717–1724. [doi: 10.1109/CVPR.2014.222]
- [11] Jayaraman D, Grauman K. Zero-Shot recognition with unreliable attributes. In: Proc. of the Advances in Neural Information Processing Systems. 2014. 3464–3472.
- [12] Fu Z, Xiang T, Kodirov E, Gong S. Zero-Shot object recognition by semantic manifold distance. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 2635–2644. [doi: 10.1109/CVPR.2015.7298879]
- [13] Yang M, Zhang L, Feng X, Zhang D. Sparse representation based fisher discrimination dictionary learning for image classification. Int'l Journal of Computer Vision, 2014,109(3):209–232. [doi: 10.1007/s11263-014-0722-8]
- [14] Jiang Z, Lin Z, Davis LS. Learning a discriminative dictionary for sparse coding via label consistent K -SVD. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2011. 1697–1704. [doi: 10.1109/CVPR.2011.5995354]
- [15] Wang Z, Hu R, Liang C, Yu Y, Jiang J, Ye M, Chen J, Leng Q. Zero-Shot person re-identification via cross-view consistency. IEEE Trans. on Multimedia, 2016,18(2):260–272. [doi: 10.1109/TMM.2015.2505083]
- [16] Kodirov E, Xiang T, Fu Z, Gong S. Unsupervised domain adaptation for zero-shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 2452–2460. [doi: 10.1109/ICCV.2015.282]
- [17] Zhang Z, Saligrama V. Zero-Shot learning via joint latent similarity embedding. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 6034–6042. [doi: 10.1109/CVPR.2016.649]
- [18] Fu Y, Hospedales T, Tao X, Gong S. Transductive multi-view zero-shot learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(11):2332–2345. [doi: 10.1109/TPAMI.2015.2408354]
- [19] Li X, Guo Y, Schuurmans D. Semi-Supervised zero-shot classification with label representation learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 4211–4219. [doi: 10.1109/ICCV.2015.479]
- [20] Rohrbach M, Ebert S, Schiele B. Transfer learning in a transductive setting. In: Proc. of the Advances in Neural Information Processing Systems. 2013. 46–54.
- [21] Xu X, Hospedales T, Gong S. Semantic embedding space for zero-shot action recognition. In: Proc. of the IEEE Conf. on Int'l Conf. on Image Processing. 2015. 63–67. [doi: 10.1109/ICIP.2015.7350760]
- [22] Wang Q, Chen K. Zero-Shot visual recognition via bidirectional latent embedding. arXiv: 1607.02104, 2016.
- [23] Wah C, Branson S, Welinder P, Perona P, Belongie S. The caltech-UCSD Birds-200-2011 dataset. In: Proc. of the California Institute of Technology. 2011. 1–8.
- [24] Patterson G, Xu C, Su H, Hays J. The sun attribute database: Beyond categories for deeper scene understanding. Int'l Journal of Computer Vision, 2014,108(1-2):59–81. [doi: 10.1007/s11263-013-0695-z]
- [25] Lazaridou A, Bruni E, Baroni M. Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world. In: Proc. of the ACL, Vol.1. 2014. 1403–1414. [doi: 10.3115/v1/P14-1132]
- [26] Guo Y, Ding G, Jin X, Wang J. Transductive zero-shot recognition via shared model space learning. In: Proc. of the American Association for Artificial Intelligence. 2016. 1–8.



冀中(1979 -),男,河北保定人,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为多媒体理解,计算机视觉.



于云龙(1988 -),男,博士生,主要研究领域为计算机视觉,模式识别.



孙涛(1992 -),男,硕士生,主要研究领域为计算机视觉,模式识别.