









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.

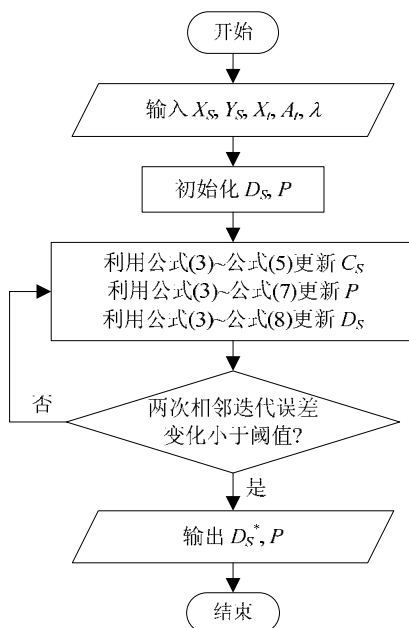


Fig.2 Algorithm flowchart of the DDL approach  
图 2 DDL 方法的算法流程图

### 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

$$\min_{D_t, C_t} \| X_t - D_t C_t \|_F^2 + \mu \| P Y - C_t \|_F^2, \| d_t \|_2^2 = 1 \tag{11}$$

其中,  $X_t$  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,  $D_t$  表示未见类别的字典矩阵,  $C_t$  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  $D_t$ ,对映射特征  $C_t$  进行更新:

$$C_t^* = \arg \min_{C_t} \| X_t - D_t C_t \|_F^2 + \mu \| P Y - C_t \|_F^2, \| d_t \|_2^2 = 1 \tag{12}$$

方程的解为

$$C_t^* = (D_t^T D_t + \mu I)^{-1} (D_t^T X_t + \mu P Y) \tag{13}$$

- 固定映射特征  $C_t$ ,对字典矩阵  $D_t$  进行更新:

$$D_t^* = \arg \min_{D_t} \| X_t - D_t C_t \|_F^2, \| d_t \|_2^2 = 1 \tag{14}$$

利用交替方向乘法求解目标函数(14).

求解出  $D_t^*$  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  $\lambda$  和  $\mu$ .

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  $D_i^* = D_s^*$ ;
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  $Y_i$ ;
12. 按照公式(13)更新  $C_i$ ;
13. 按照公式(14)更新  $D_i$ ;
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  $D_i^*$ .

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  $\lambda$  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.

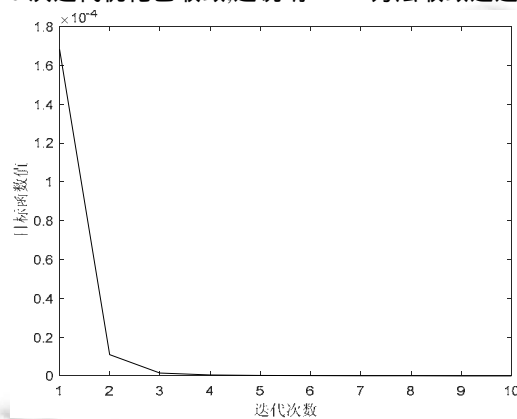


Fig.3 Convergence curve of DDL on the AwA dataset with attribute features

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)<sup>[1]</sup>, Caltech-UCSD Bird2011(CUB)<sup>[23]</sup>和SUN Attribute Dataset<sup>[24]</sup>.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<sup>[1]</sup>;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 $\lambda$ 和 $\mu$ ,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 $\lambda$ 和 $\mu$ 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)<sup>[1]</sup>,Structural Joint Embedding(SJE)<sup>[2]</sup>,Latent Embeddings(LatEm)<sup>[7]</sup>和Synthesized Classifiers(SC)<sup>[4]</sup>.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 <sup>[6]</sup>	V	57.5	-	72.0
SJE <sup>[2]</sup>	G	66.7	50.1	-
LatEm <sup>[7]</sup>	G	71.9	45.5	-
SC <sup>[4]</sup>	G	72.9	<b>54.7</b>	-
DDL	V	<b>77.3</b>	52.6	<b>86.0</b>

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)<sup>[25]</sup>,Structural Joint Embedding(SJE)<sup>[2]</sup>和Latent Embeddings(LatEm)<sup>[7]</sup>,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 <sup>[25]</sup>	V	65.6	30.4
SJE <sup>[2]</sup>	G	51.2	28.4
LatEm <sup>[7]</sup>	G	61.1	<b>31.8</b>
DDL	V	<b>75.6</b>	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)<sup>[18]</sup>,Shared Model Space Learning(SMS)<sup>[26]</sup>和 Unsupervised Domain Adaptation(UDP)<sup>[16]</sup>.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 <sup>[18]</sup>	80.5	47.9
SMS <sup>[26]</sup>	78.5	-
UDP <sup>[16]</sup>	75.6	40.6
TDDL	<b>93.3</b>	<b>55.3</b>

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.