DOI:10.11784/tdxbz201606003

基于典型相关分析和距离度量学习的零样本学习

冀 中，谢于中，庞彦伟
(天津大学电气自动化与信息工程学院，天津 300072)

摘 要：零样本学习是一类特殊的图像分类问题，是指测试数据的类别在训练数据中没有出现的情况. 为了更好地描述语义特征空间中图像特征和语义特征的距离关系，本文将距离度量学习引入零样本学习任务. 具体而言，首先利用典型相关分析将样本的图像特征和相应类别的语义特征映射至公共特征空间；然后，利用距离度量学习衡量图像特征和语义特征之间的距离；最后，使用最近邻分类器进行分类. 通过在流行的 AwA 和 CUB 数据集中的实验，证明了所提方法的有效性和鲁棒性.

关键词：零样本学习；典型相关分析；距离度量学习；图像分类
中图分类号：TP391.41 **文献标志码**：A **文章编号**：0493-2137(2017)08-0813-08

Zero-Shot Learning Based on Canonical Correlation Analysis and Distance Metric Learning

Ji Zhong，Xie Yuzhong，Pang Yanwei
(School of Electrical and Information Engineering，Tianjin University，Tianjin 300072，China)

Abstract：Zero-shot learning is a special case of image classification，whose test classes are absent in training samples. To better measure the distance between visual features and semantic features in the semantic embedding space，a distance metric learning based zero-shot learning method is proposed. Specifically，visual features and semantic features were first projected into a common semantic embedding space by use of canonical correlation analysis，then a distance metric learning method was employed to measure the distance between them. Finally，a nearest neighbor classifier was utilized to perform the classification. Experimental results on the popular AwA and CUB datasets demonstrate that the proposed approach is effective and robust.

Keywords：zero-shot learning；canonical correlation analysis；distance metric learning；image classification

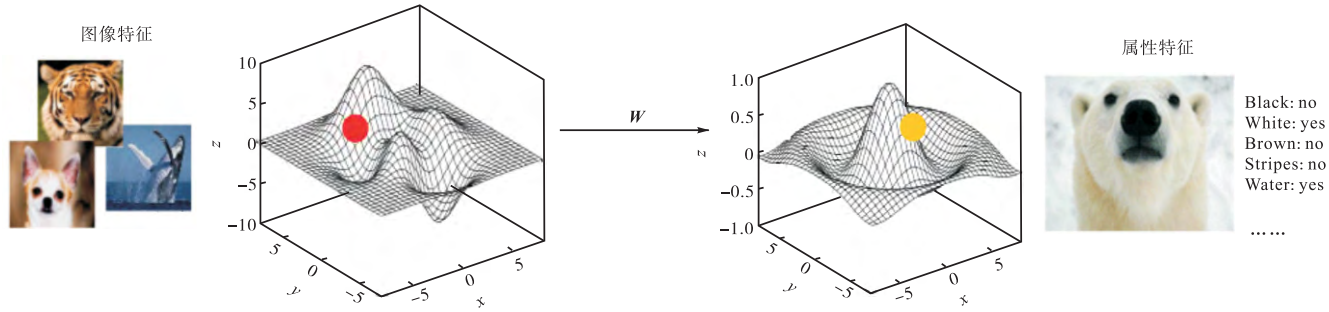
对于传统的分类系统，要想准确识别出某类数据，必须给出相应类别带标签的训练数据，但是训练数据的标签往往难以获得. 由于能够对在训练样本中没有出现过的类别进行分类，零样本学习(zero-shot learning)可以作为解决类别标签缺失问题的一种有效手段[1-2]. 不同于奇异值检测，零样本学习并非要单纯检测出新的类别，而是要在训练数据中未出现测试类别的情况下对测试类别进行分类[3]. 零样本学习的对象主要可以分为图像和视频 2 类，本文研究图像的零样本学习.

零样本学习可以看作是跨模态检索学习的一种特殊类型. 跨模态检索学习能被用于众多领域，如图像检索[4-5]、重排序[6]、动作识别[7]等. 关联学习是实现跨模态检索学习的一种重要方法，而典型相关分析(canonical correlation analysis，CCA)则是解决关联学习的一种经典方法[8]. Rasiwasia 等[9]提出了一种基于 CCA 的跨模态检索方法，用于获得不同模态特征间的共享描述. Zhang 等[10]使用 CCA 学习图像特征和视频特征之间的关联关系，然后使用基于相关性的距离度量方法，并使用该度量方法对图像和视频进行聚类. Hardoon 等[11]提出了核典型相关分析(kernel canonical correlation analysis，KCCA)方法以学习图片和文本描述之间的语义关系. 本文也将使用 CCA 进行跨模态学习.

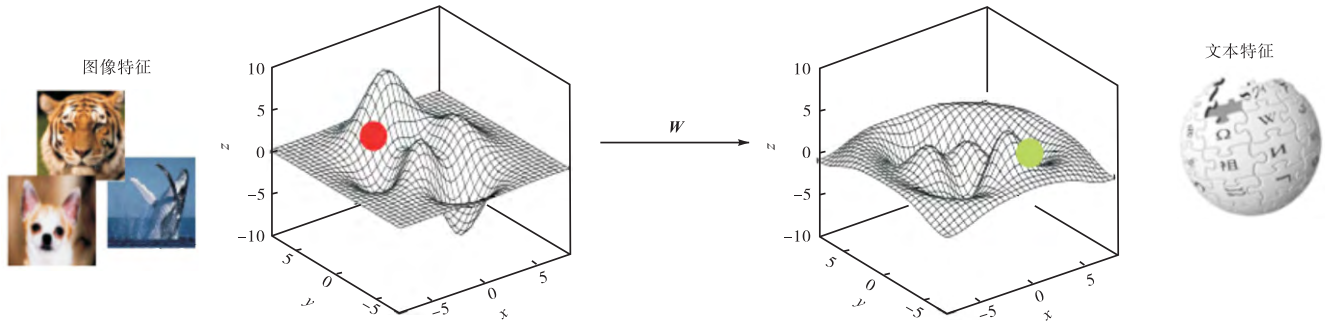
收稿日期：2016-06-03；**修回日期**：2016-11-24.
作者简介：冀 中(1979—)，男，博士，副教授.
通讯作者：冀 中，jizhong@tju.edu.cn.
基金项目：国家自然科学基金资助项目(61472273，61632018).
Supported by the National Natural Science Foundation of China(Nos. 61472273 and 61632018).

1　零样本学习

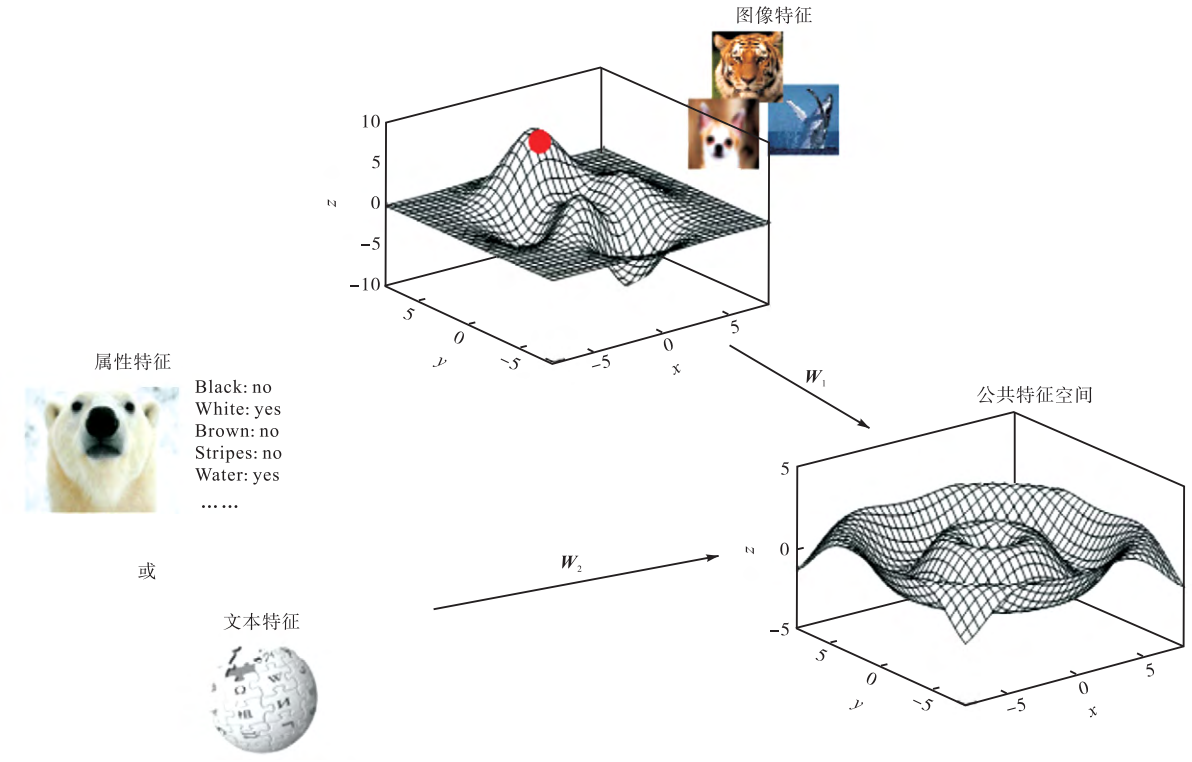
零样本学习的一般思路是通过构建语义特征空间，使得类别的语义特征和样本的特征之间的相似度可以直接进行度量，从而能够比较测试样本的特征与未出现过的类别的语义特征之间的相似度，进而找出最相近的类别[12]. 如图 1 所示，现有的语义特征空间可分为 3 类：①属性(attribute)特征空间[13-15]；②文本特征空间[1-2]；③公共特征空间[16-17].



(a) 属性特征空间



(b) 文本特征空间



(c) 公共特征空间

图 1　语义特征空间示例

Fig.1　Examples for different semantic embedding space

属性特征空间是直接将类别属性所在的空间作为特征空间. 属性是用于表示类别信息的一种特征，它的每一个维度表示该类别是否具有某种属性[13]. 属性通常由人工标注得到，因而能够有效地表示类别的信息并反映类别之间的关系. 但同时属性也具有耗费人力、可扩展性较差的缺点. 早期的方法包括 Lampert 等[13]提出的直接属性预测模型(direct attribute prediction，DAP). 该方法首先通过训练样本的图像特征和相应类别的属性特征，学习得到相应的属性分类器，从而得到测试样本属于各个未见过的类别的后验概率，最后使用最大后验估计对测试样本进行分类. 近期，Romera-Paredes 等[14]构造了一个 2 层的线性网络，第 1 层网络描述了图像特征和属性特征之间的关系，网络的权值在训练阶段学习得到；第 2 层网络由属性特征和类别标签之间的关系决定. 这种方法能够简单而高效地进行零样本学习. 也有一些文献对属性特征各元素之间的关系进行了探讨，Liu 等[15]提出了一个统一的架构，将属性间关系的学习和属性预测结合起来；巩萍等[16]通过属性间的正负相关性构建属性关系图，然后对图像特征进行图正则化特征选择，再将选择后的特征用于 DAP 模型的训练.

文本特征空间是直接将类别名称的文本特征所在的空间作为特征空间的方法. 文本特征通常通过自然语言处理技术从无标注的语料库中直接提取而来，该技术能将文本表示为特征向量，常见的模型包括WordVec[17]和Glove[18]. 使用词向量表示类别名称，词向量之间的相似度就能够较好地代表类别名称语义上的相似度. Socher 等[1]利用一个 2 层的神经网络训练一个映射函数，将图像特征映射至文本特征空间，使得图像特征和所属类别的文本特征距离最近. Frome 等[2]则直接将卷积神经网络的最顶层和 skip-gram 语言模型的输出层通过映射函数相连接，并使用合页损失函数(hinge loss function)学习得到映射函数.

此外，第 3 类方法是将样本的特征和类别的语义特征映射至一个公共空间. 例如，文献[19]提出了一种结构化联合嵌入模型(structured joint embedding，SJE)，通过映射矩阵将图像特征和语义特征嵌入公共特征空间，使得公共特征空间中的各模态特征内积和最大. Xian 等[20]则在 SJE 的基础上提出了一种隐藏嵌入模型(latent embeddings model，LatEm)，同样取得了良好的效果.

然而，现有的工作大部分着重研究使用何种语义特征或者如何更好地将数据特征映射至语义特征空间，关于语义特征空间中数据特征和语义特征距离度量方式的研究还鲜有报道. 事实上，在语义特征空间中，合理的距离度量方式能够准确地反映出各个模态特征之间的关系，从而有助于提高分类的性能. 现有的零样本学习方法通常使用传统的欧氏距离进行度量，假设样本特征的各个维度都同等重要，这往往不能有效地描述样本间的关系. 基于此，本文将距离度量学习(distance metric learning，DML)引入了零样本学习，可更好地描述图像特征和语义特征之间的距离. 具体地，本文首先利用典型相关分析将图像的视觉特征和类别的语义特征映射至公共特征空间，然后利用距离度量学习的方法衡量图像特征和语义特征之间的距离，最后使用最近邻分类器进行分类. 大量实验表明，将距离度量学习引入零样本学习能够显著提高性能.

2　典型相关分析

典型相关分析[8]最早由 Hotelling 提出，并用于研究 2 组随机向量之间的相关性问题. 给定 N 个样本对 $\{\boldsymbol{x}_i,\boldsymbol{y}_i\}(i=1,\cdots,N)$，其包含来自 2 个不同模态的特征向量 $\boldsymbol{x}_i$ 和 $\boldsymbol{y}_i$. 由 $\boldsymbol{x}_i$ 和 $\boldsymbol{y}_i$ 组成的特征矩阵可以表示为 $\boldsymbol{X}=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N]\in\mathbf{R}^{p\times N}$ 和 $\boldsymbol{Y}=[\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N]\in\mathbf{R}^{q\times N}$，其中 p 和 q 是特征的维数. CCA 的目标是寻找到一对映射矩阵 $\boldsymbol{W}_1\in\mathbf{R}^{p\times d}$ 和 $\boldsymbol{W}_2\in\mathbf{R}^{q\times d}$，其中 d 是公共特征空间的维数，使得映射后特征矩阵 $\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{X}$ 和 $\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{Y}$ 之间的相关系数最大，即

$$\max_{\boldsymbol{W}_1,\boldsymbol{W}_2}\frac{\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{C}_{xy}\boldsymbol{W}_2}{\sqrt{\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{C}_{xx}\boldsymbol{W}_1}\cdot\sqrt{\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{C}_{yy}\boldsymbol{W}_2}}\tag{1}$$

式中：$\boldsymbol{C}_{xx}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$、$\boldsymbol{C}_{yy}=\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}$ 分别是 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的协方差矩阵；$\boldsymbol{C}_{xy}=\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}$ 是 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的互协方差矩阵. 使用拉格朗日乘子法求解式(1)可以得到如下方程：

$$\begin{pmatrix}0 & \boldsymbol{C}_{xy}\\ \boldsymbol{C}_{xy}^{\mathrm{T}} & 0\end{pmatrix}\begin{pmatrix}\boldsymbol{W}_1\\ \boldsymbol{W}_2\end{pmatrix}=\lambda\begin{pmatrix}\boldsymbol{C}_{xx} & 0\\ 0 & \boldsymbol{C}_{yy}\end{pmatrix}\begin{pmatrix}\boldsymbol{W}_1\\ \boldsymbol{W}_2\end{pmatrix}\tag{2}$$

对式(2)进行广义特征值分解，由前 d 个最大的特征值对应的特征向量得到映射矩阵 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$.

3　距离度量学习

距离度量学习由 Xing 等[21]提出，其思路是对于 2 个特征向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，学习到的距离度量形式如下所示：

$$d(\boldsymbol{a},\boldsymbol{b})=d_{\boldsymbol{A}}(\boldsymbol{a},\boldsymbol{b})=\|\boldsymbol{a}-\boldsymbol{b}\|_{\boldsymbol{A}}=\sqrt{(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{a}-\boldsymbol{b})}\tag{3}$$

为了保证 $d_{\boldsymbol{A}}(\boldsymbol{a},\boldsymbol{b})$ 的非负性以及满足三角不等式，矩

阵 $\boldsymbol{A}$ 应该为半正定矩阵，即 $\boldsymbol{A}\geq 0$．当 $\boldsymbol{A}=\boldsymbol{I}$ 时，$d_{\boldsymbol{A}}(\boldsymbol{a},\boldsymbol{b})$ 为欧氏距离；当 $\boldsymbol{A}$ 被限制为对角阵时，$\boldsymbol{A}$ 的对角元素可以被认为是赋予各个维度的不同权重；当 $\boldsymbol{A}$ 为全矩阵时，学习得到的距离度量可以被认为是马氏距离.

进一步，目标函数可设计为[21]

$$\min_{\boldsymbol{A}} \sum_{(\boldsymbol{a}_i,\boldsymbol{b}_j)\in S} \left\|\boldsymbol{a}_i-\boldsymbol{b}_j\right\|_{\boldsymbol{A}}^2$$

$$\text{s.t.} \quad \sum_{(\boldsymbol{a}_i,\boldsymbol{b}_j)\in D} \left\|\boldsymbol{a}_i-\boldsymbol{b}_j\right\|_{\boldsymbol{A}}^2 \geq 1,\ \boldsymbol{A}\geq 0 \tag{4}$$

式中：$(\boldsymbol{a}_i,\boldsymbol{b}_j)\in S$ 表示 $\boldsymbol{a}_i$、$\boldsymbol{b}_j$ 是同一类别的样本；$(\boldsymbol{a}_i,\boldsymbol{b}_j)\in D$ 表示 $\boldsymbol{a}_i$、$\boldsymbol{b}_j$ 是不同类别的样本. 该目标函数的意义是在满足不相似样本对之间的距离之和大于一个固定值的前提下，寻找度量矩阵 $\boldsymbol{A}$，使得相似样本对之间的距离和达到最小. 对于度量矩阵 $\boldsymbol{A}$ 被限制为对角阵的情况，可以定义[21]

$$g(\boldsymbol{A})=g(\boldsymbol{A}_{11},\cdots,\boldsymbol{A}_{nn})=$$

$$\sum_{(\boldsymbol{a}_i,\boldsymbol{b}_j)\in S} \left\|\boldsymbol{a}_i-\boldsymbol{b}_j\right\|_{\boldsymbol{A}}^2 - \ln\left(\sum_{(\boldsymbol{a}_i,\boldsymbol{b}_j)\in D} \left\|\boldsymbol{a}_i-\boldsymbol{b}_j\right\|_{\boldsymbol{A}}^2\right) \tag{5}$$

则式(4)等效为式(5)中的 $g(\boldsymbol{A})$，从而可以采用 Newton-Raphson 法，更高效地学习得到度量矩阵 $\boldsymbol{A}$. 本文采用的即为这种方法.

4 基于典型相关分析和距离度量学习的零样本学习

在零样本学习任务中，图像特征和语义特征属于不同的模态，它们的维数、分布均不相同. 为了度量它们之间的相似度，本文使用典型相关分析构造映射矩阵，将它们映射至公共特征空间. 在公共特征空间中，现有零样本学习方法通常使用欧氏距离进行度量，但是欧式距离假设样本特征的每个维度都同等重要，这往往不能有效地描述样本间的距离关系. 基于此，本文将距离度量学习的方法和典型相关分析融合，引入到零样本学习中，提出了基于典型相关分析和距离度量学习的零样本学习算法(CCA-DML).

假设数据集中给定了 N_S 个已标注的训练样本 $S=\{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}$，$N_U$ 个未标注的测试样本 $U=\{\boldsymbol{X}',\boldsymbol{Y}',\boldsymbol{Z}'\}$. 其中 $\boldsymbol{X}\in\mathbf{R}^{p\times N_S}$ 和 $\boldsymbol{X}'\in\mathbf{R}^{p\times N_U}$ 分别是训练及测试样本的 p 维图像特征；$\boldsymbol{Z}$、$\boldsymbol{Z}'$ 分别是训练样本和测试样本的类别标签，因为测试样本中的类别在训练样本中没有出现，所以它们的交集为空，即 $\boldsymbol{Z}\cap\boldsymbol{Z}'=\varnothing$；$\boldsymbol{Y}\in\mathbf{R}^{q\times N_S}$ 和 $\boldsymbol{Y}'\in\mathbf{R}^{q\times N_U}$ 分别表示训练及测试样本其类别所对应的 q 维语义特征. 零样本学习的任务就是预测出测试样本的类别标签，其中测试样本的类别在训练样本中没有出现.

CCA-DML 的示意如图 2 所示，主要包括以下 5 个步骤.

步骤 1 提取样本的图像特征和类别的语义特征. 其中，类别的语义特征可以是人工标注的属性特征，也可以是由自然语言处理技术对类别名称提取的文本特征.

步骤 2 构建图像特征和语义特征的公共特征空间. 由训练样本的图像特征 $\boldsymbol{X}$ 和对应类别的语义特征 $\boldsymbol{Y}$，构造协方差矩阵 $\boldsymbol{C}_{xx}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$、$\boldsymbol{C}_{yy}=\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}$ 以及互协方差矩阵 $\boldsymbol{C}_{xy}=\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}$. 再将它们代入式(2)中，进行广义特征值分解，然后将前 d 个最大的特征值对应的特征向量组成映射矩阵 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$，其中 d 为指定的公共特征空间的维数.

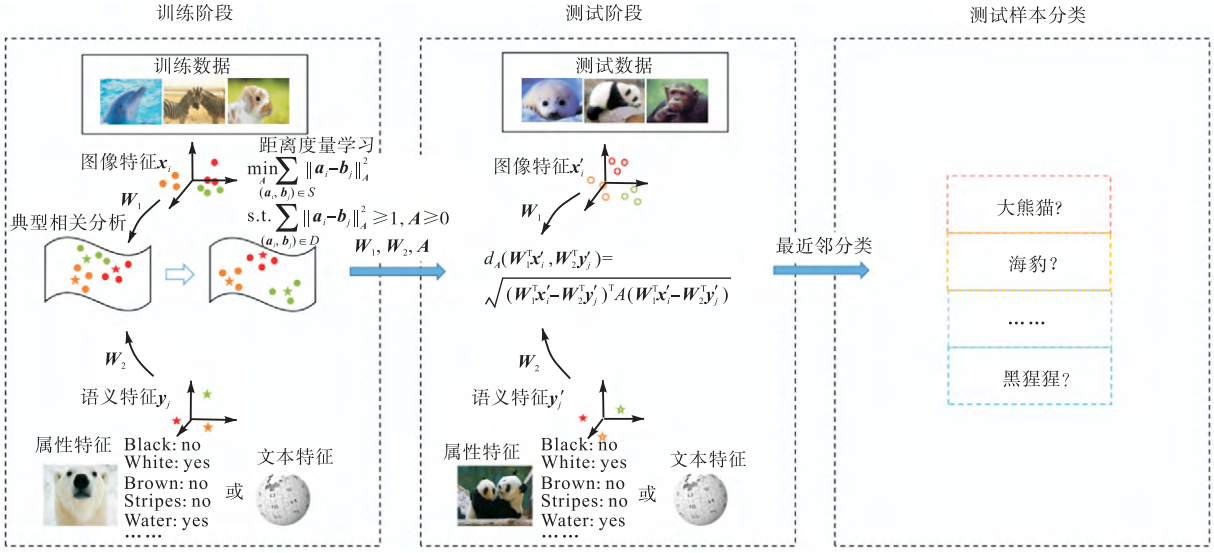


图 2 CCA-DML 零样本学习

Fig.2 Zero-shot learning with canonical correlation analysis and distance metric learning

步骤 3　在公共特征空间中进行距离度量学习. 将映射后的同类别图像特征和语义特征两两配对，组成相似样本集合；将映射后的不同类别图像特征和语义特征两两配对，组成不相似样本集合. 然后，代入式(5)中，求导后使用 Newton-Raphson 法，学习得到度量矩阵 $\boldsymbol{A}$.

步骤 4　相似距离的计算. 对于测试样本 $U=\{\boldsymbol{X}',\boldsymbol{Y}',\boldsymbol{Z}'\}$，虽然已知待测试类别的语义特征 $\boldsymbol{Y}'$，但它和图像特征 $\boldsymbol{X}'$ 之间的对应关系是未知的. 解决方法是通过公共特征空间，找出 $\boldsymbol{X}'$ 和 $\boldsymbol{Y}'$ 之间的对应关系. 首先，将 $\boldsymbol{X}'$ 和 $\boldsymbol{Y}'$ 映射至公共特征空间，得到 $\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{X}'$ 和 $\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{Y}'$. 然后，对于每个测试样本映射后的图像特征 $\boldsymbol{W}_1^{T}\boldsymbol{x}_i', i=1,\cdots,N_U$，计算它与测试类别数 M_U 个映射后语义特征 $\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{y}_j', j=1,\cdots,M_U$ 之间的距离 $d_{\boldsymbol{A}}(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{x}_i',\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{y}_j')$.

$$d_{\boldsymbol{A}}(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{x}_i',\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{y}_j')=\sqrt{(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{x}_i'-\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{y}_j')^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{x}_i'-\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{y}_j')} \tag{6}$$

步骤 5　测试样本分类. 使用最近邻分类器，令与测试样本在公共特征空间中距离最近的待测试类语义特征所对应的类别作为预测类别，即

$$\arg\min_j d_{\boldsymbol{A}}(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{x}_i',\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{y}_j') \quad j=1,\cdots,M_U \tag{7}$$

5　实验结果与讨论

5.1　数据集与实验设置

本文利用零样本学习领域常用的 AwA(animals with attributes)[13] 和 CUB(caltech-uCSD-birds-200-2011)[22]两个数据集中验证所提出的 CCA-DML 方法. 其中，AwA 数据集中包含了 50 个类别，共 30 475 张动物图片. 对于每一个类别，AwA 数据集提供了一个 85 维的属性特征，其图片及属性示例如图 3 所示. 在实验过程中，和其他对比算法相同[13]，本文选择了默认的训练/测试集划分方式，即用选定的 40 个类别作为训练样本，用剩下的训练样本中没有出现的 10 个类别作为测试样本.

CUB 数据集则包含了 200 个类别，共 11 788 张鸟的图片，并为每个类别提供了一个 312 维的属性特征. 和 AWA 数据集相比，CUB 数据集更具挑战性，因为它实际上是一个用于精细图像分类的数据集，各个类别之间的差异较小，其示例如图 4 所示. 此外，CUB 数据集包含了更多的类别，而每个类别中含有的样本数相对较少，这也增加了 CUB 数据集的难度. 与其他对比算法相同[19]，本文选择其中的 150 类作为训练数据，用剩下的 50 类作为测试数据.

图 3　AwA 数据集图像及相应属性标注示例

Fig.3　Examples of images with their corresponding attribute annotation on the AwA dataset

图 4　CUB 数据集示例

Fig.4　Examples of the CUB dataset

在 AwA 数据库中，图像特征采用和文献[23]相同的 CNN 特征(VGGNet-19). 在 CUB 数据集中，使用经过 Imagenet 预训练的 VGGNet-19 模型[24]，将隐藏层最顶层的 4 096 维输出作为图片的视觉特征. 除了数据集本身提供的属性特征外，本文还使用文本特征作为语义特征. 利用在维基百科语料库中训练好的 Word2Vec 模型[17]，为 AwA 和 CUB 数据集的类别名称分别提取 1 000 维和 400 维的文本特征.

在进行度量学习的过程中，为了减少计算时间，对训练集进行随机抽样. 对于 AwA 数据集，每类选择 30 个样本；对于 CUB 数据集，每类选择 2 个样本. 在实验过程中，使用上述的随机抽样方法，重复 10 次实验，使用 10 次实验结果的均值和方差作为算法的最终分类结果. 算法的性能通过平均分类准确率 $\bar{\eta}$ 来衡量，其由每个测试类别的分类准确率取平均

得到[12-13, 19], 即

$$\bar{\eta} = \frac{1}{M_U}\sum_{i=1}^{M_U}\frac{N_i'}{N_i} \tag{8}$$

式中: N_i' 是第 i 个测试类别中分类正确的样本数; N_i 是第 i 个测试类别的样本数.

5.2 距离度量学习对算法性能的影响分析

为了证明距离度量学习与典型相关分析相结合能够提高分类准确率, 将使用欧氏距离时的 CCA 分类结果(CCA-Euc)与 CCA-DML 的结果在各种情况下进行了对比, 如表 1 所示.

从表 1 的结果可以看出: ①在 2 个数据集中, 与使用 CCA-Euc 相比, CCA-DML 方法的性能在使用文本特征和属性特征 2 种方法上均有提升, 其中, 在 AwA 数据集中, 使用文本特征和使用属性特征时分别提高了 6.3% 和 1.8%; 在 CUB 数据集中, 使用文本特征和使用属性特征时分别提高了 6.5% 和 4.8%; ②文本特征较之属性特征性能提高较为明显, 这是因为文本特征是由无监督的方法直接从语料库中提取出来的, 包含的噪声较多, 而 DML 可以给这些噪声维度赋予较低的权重, 从而有效地克服噪声; 而属性特征是针对特定数据集人为手工设定的, 含有的噪声较少; 因而 CCA-DML 的性能提升的相对较小; ③在 CUB 数据集中, 使用属性特征时的性能提升也较为显著, 这是因为 CUB 数据集类别之间的差异较小, 分类难度较大, 只使用欧氏距离难以得到很好的性能, 而 DML 使类内元素靠近, 将类间元素拉远, 有助于区分这些类别.

表 1 CCA-Euc 与 CCA-DML 最佳分类性能比较

Tab.1 Performance comparison of CCA-Euc and CCA-DML

%

度量方式	AwA		CUB	
	文本特征	属性特征	文本特征	属性特征
CCA-Euc	66.3	75.5	27.2	49.3
CCA-DML	72.6	77.3	33.7	54.1

5.3 与当前先进算法的对比分析

表 2 给出了在 AwA 和 CUB 数据集中不同算法的平均分类准确率比较. 实验选取的对比算法主要有 SJE[19] 、LatEm[20] 、DAP[13] 、SSE-ReLU[23] 和 ESZSL[14]. 除 ESZSL 是自行实现之外, 其他对比算法的实验性能均为相应文章所提供的数值.

表 2 不同算法在 AwA 和 CUB 数据集中的性能比较

Tab.2 Performance comparison of different algorithms on AwA and CUB datasets %

图像特征	算法	AwA		CUB	
		文本特征	属性特征	文本特征	属性特征
GoogLeNet	SJE[19]	51.2	66.7	28.4	50.1
	LatEm[20]	61.1	71.9	31.8	45.5
VGGNet-19	DAP[13]		57.5		
	SSE-ReLU[23]		76.3 ± 0.8		30.4 ± 0.2
	ESZSL[14]	60.6 ± 1.6	74.6 ± 3.7	32.3 ± 0.8	50.8 ± 0.4
	CCA-DML	72.6 ± 1.4	77.3 ± 0.7	33.7 ± 0.2	54.1 ± 0.7

从表 2 可以看出, 无论使用文本特征还是属性特征, 本文所提出的 CCA-DML 方法均能取得最好的性能. 对于 AwA 数据集, 使用文本特征和属性特征时, CCA-DML 的性能分别比第 2 好的 LatEm 和 SSE-ReLU 提高了 11.5% 和 1.0%. 对于 CUB 数据集, 使用文本特征和使用属性特征时, CCA-DML 的性能比第 2 好的 ESZSL 分别提高了 1.4% 和 3.3%. 这些结果表明了所提 CCA-DML 方法的有效性.

5.4 公共特征空间维数对算法性能的影响

本节以在 CUB 数据集中使用属性特征作为语义特征时为例, 展示不同的公共特征空间维数对所提方法性能的影响. 从图 5 可以看出, 在维数较低时, CCA-Euc 的性能高于 CCA-DML. 这是因为这时映射后的特征包含的信息不足, DML 不能学习到有效的度量方式. 其次, 2 种方法的性能都是先逐渐增加, 随后 CCA-Euc 在某处达到峰值, 并随着特征维数的继续增加而减小, 而 CCA-DML 则继续保持上升趋势最后趋于平缓. 这是因为公共特征空间维数增加到一定程度之后, 对于 CCA 而言, 将会引入噪声, 导致性能有所下降; 而 CCA-DML 中的度量学习方法学习得到的度量矩阵 $\boldsymbol{A}$ 可以看作是不同维度的权重, 因此能够给噪声赋予较低的权重, 从而具有较好抗噪性能和鲁棒性.

实验中, AwA 数据集的公共特征空间维数均设为 50 维; CUB 数据集中使用文本特征和属性特征时, 公共特征空间维数分别设为 230 维和 70 维.

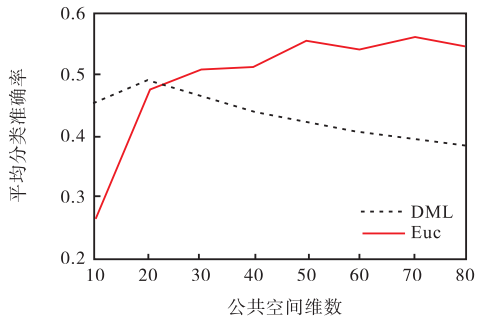


图 5　公共特征空间维数对算法性能的影响

Fig.5　Dimensionality impact of common feature space on algorithm performance

5.5　训练样本数对算法性能的影响

因为 CCA-DML 需要在整个数据集构造全部相似对和非相似对，当数据很多时，构造出的相似对集合和非相似对集合也会过大，将难以处理大规模的数据集问题，所以在实际工作中采用对训练集采样的方式减少训练样本. 图 6 是在 AwA 和 CUB 数据集中不同样本数下 CCA-DML 的性能变化.

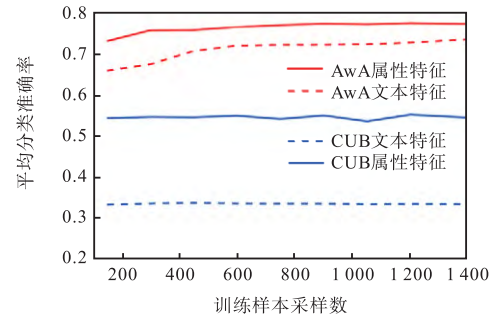


图 6　不同训练样本数下 CCA-DML 性能变化

Fig.6　Training samples impact on CCA-DML performance

从图 6 中可以看出，随着训练样本数的增加，平均分类准确率将会上升；当样本数高于一定值后，性能总体上趋于平稳. 在综合考虑分类性能和计算量之后，将 AwA 和 CUB 数据集中的随机抽样数分别定为 1 200 和 300.

6　结　语

本文提出了一种基于典型相关分析和距离度量学习的零样本学习方法. 通过与传统欧式距离的比较，可以看出距离度量学习能够有效地描述公共特征空间中图像特征和语义特征的距离关系，从而提高分类性能. 通过与当前主流方法的性能的比较，所提 CCA-DML 算法能够取得更好的性能.

本文下一步的主要研究方向是使用非线性典型相关分析(如 KCCA)对 CCA-DML 算法进行改进，以进一步提升算法性能.

参考文献：

[1] Socher R，Ganjoo M，Manning C D，et al. Zero-shot learning through cross-modal transfer[C]//*Advances in Neural Information Processing Systems*. Lake Tahoe，USA，2013：935-943.

[2] Frome A，Corrado G S，Shlens J，et al. Devise：A deep visual-semantic embedding model[C]//*Advances in Neural Information Processing Systems*. Lake Tahoe，USA，2013：2121-2129.

[3] Pimentel M A F，Clifton D A，Clifton L，et al. A review of novelty detection[J]. *Signal Processing*，2014，99(6)：215-249.

[4] Deng C，Tang X，Yan J，et al. Discriminative dictionary learning with common label alignment for cross-modal retrieval[J]. *IEEE Transactions on Multimedia*，2016，18(2)：208-218.

[5] Liu X，Deng C，Lang B，et al. Query-adaptive reciprocal hash tables for nearest neighbor search[J]. *IEEE Transactions on Image Processing*，2016，25(2)：907-919.

[6] Deng C，Ji R，Tao D，et al. Weakly supervised multi-graph learning for robust image reranking[J]. *IEEE Transactions on Multimedia*，2014，16(3)：785-795.

[7] Yang Y，Deng C，Tao D，et al. Latent max-margin multitask learning with skelets for 3-D action recognition [J]. *IEEE Transactions on Cybernetics*，2016：1-10.

[8] Hotelling H. Relations between two sets of variates[J]. *Biometrika*，1936，28(3/4)：321-377.

[9] Rasiwasia N，Costa Pereira J，Coviello E，et al. A new approach to cross-modal multimedia retrieval[C]// *ACM International Conference on Multimedia*. Firenze，Italy，2010：251-260.

[10] Zhang H，Zhuang Y，Wu F. Cross-modal correlation learning for clustering on image-audio dataset[C]// *ACM International Conference on Multimedia*. Augsburg，Germany，2007：273-276.

[11] Hardoon D R，Szedmak S，Shawe-Taylor J. Canonical correlation analysis：An overview with application to learning methods[J]. *Neural Computation*. 2004，16(12)：2639-2664.

[12] Fu Z，Xiang T，Kodirov E，et al. Zero-shot object recognition by semantic manifold distance[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Boston，USA，2015：2635-2644.

[13] Lampert C H, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(3): 453-465.

[14] Romera-Paredes B, Torr P H S. An embarrassingly simple approach to zero-shot learning[C]//*Proceedings of The* 32*nd International Conference on Machine Learning*. Lille, France, 2015: 2152-2161.

[15] Liu M, Zhang D, Chen S. Attribute relation learning for zero-shot classification[J]. *Neurocomputing*, 2014, 139: 34-46.

[16] 巩 萍, 程玉虎, 王雪松. 基于属性关系图正则化特征选择的零样本分类[J]. 中国矿业大学学报, 2015, 44(6): 1097-1104.
Gong Ping, Cheng Yuhu, Wang Xuesong. Zero-shot classification based on attribute correlation graph regularized feature selection[J]. *Journal of China University of Mining and Technology*, 2015, 44(6): 1097-1104(in Chinese).

[17] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//*Advances in Neural Information Processing Systems*. Lake Tahoe, USA, 2013: 3111-3119.

[18] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//*Conference on Empirical Methods on Natural Language Processing*. Doha, Qatar, 2014: 1532-1543.

[19] Akata Z, Reed S, Walter D, et al. Evaluation of output embeddings for fine-grained image classification[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 2927-2936.

[20] Xian Y, Akata Z, Sharma G, et al. Latent embeddings for zero-shot classification[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 69-77.

[21] Xing E P, Ng A Y, Jordan M I, et al. Distance metric learning with application to clustering with side-information[C]//*Advances in Neural Information Processing Systems*. Vancouver, Canada, 2003: 521-528.

[22] Wah C, Branson S, Welinder P, et al. The caltech-ucsd birds-200-2011 dataset[EB/OL]. http://www.vision.caltech.edu/visipedia/CUB-200-2011.html, 2011-01-15.

[23] Zhang Z, Saligrama V. Zero-shot learning via semantic similarity embedding[C]// *IEEE International Conference on Computer Vision*. Santiago, Chile, 2015: 4166-4174.

[24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C] // *International Conference on Learning Representations*. San Diego, USA, 2015: 1-13.

(责任编辑: 王晓燕)