

一种基于视觉特征组合构造的零样本学习方法*

杨刚¹, 刘金露¹, 李锡荣^{1,2}, 许洁萍¹

¹(中国人民大学 信息学院,北京 100872)

²(数据工程与知识工程重点实验室(中国人民大学),北京 100872)

通讯作者: 李锡荣, E-mail: xirong@ruc.edu.cn



摘要: 零样本学习是机器学习和图像识别领域重要的研究热点.零样本学习方法通常利用未见类与可见类之间的类别语义信息,将从可见类样本学习到的知识转移到未见类,实现对未见类样本的分类识别.提出了一种基于视觉特征组合构造的零样本学习方法,采用特征组合的方式构造产生大量未见类样例特征,将零样本学习问题转化为标准的监督学习分类问题.该方法模拟了人类的联想认知过程,其主要包括4步:特征-属性关系提取、样例构造、样例过滤、特征域适应.在可见类样本上抽取类别属性与特征维度的对应关系;利用特征-属性关系,通过视觉特征的组合构造的方式,产生未见类样例;引入非相似表示,过滤掉不合理的未见类样例;提出半监督特征域适应和无监督特征域适应,实现未见类样例的线性转换,产生更有效的未见类样例.在3个基准数据集(AwA, AwA2 和 SUN)上的实验结果显示,该方法效能优越,在数据集 AwA 上获得了当前最优的 Top-1 分类正确率 82.6%.实验结果证明了该方法的有效性和先进性.

关键词: 零样本学习;图像分类;非相似表示;数据预处理

中文引用格式: 杨刚,刘金露,李锡荣,许洁萍.一种基于视觉特征组合构造的零样本学习方法.软件学报,2018,29(Suppl.(2)):16-29. <http://www.jos.org.cn/1000-9825/18014.htm>

英文引用格式: Yang G, Liu JL, Li XR, Xu JP. Visual feature combination approach for zero-shot learning. Ruan Jian Xue Bao/Journal of Software, 2018,29(Suppl.(2)):16-29 (in Chinese). <http://www.jos.org.cn/1000-9825/18014.htm>

Visual Feature Combination Approach for Zero-Shot Learning

YANG Gang¹, LIU Jin-Lu¹, LI Xi-Rong^{1,2}, XU Jie-Ping¹

¹(School of Information, Renmin University of China, Beijing 100872, China)

²(Key Laboratory of Data Engineering and Knowledge Engineering (Renmin University of China), Beijing 100872, China)

Abstract: Zero-Shot learning is an important research in the field of machine learning and image recognition. Zero-Shot learning methods normally use the semantic information among unseen classes and seen classes to transfer the knowledge which is learned from examples of seen classes to unseen classes, so as to recognize and classify the examples of unseen classes. In this study, a zero-shot learning approach based on construction of visual feature combination is proposed. The approach generates many examples of unseen class on visual feature level by the way of feature combination, which is first proposed, and thus transforms zero-shot learning problem to be a traditional classification problem solved by supervised learning. The approach mimics human cognition process of associative memory, and includes four steps: feature-attribute relation extraction, example construction, example screening, and domain adaption. On training examples of seen classes, the relationship between class attributes and dimensions of feature is extracted; on visual feature level, examples of unseen classes are generated by visual feature combination; dissimilarity representation is introduced to filter the generated examples of unseen classes; semi-supervised and unsupervised feature domain adaption are proposed to linearly transform the generated examples of unseen classes to be more effective. The proposed approach shows superior performance on three benchmark datasets (AwA, AwA2, and SUN), especially on dataset AwA, it obtains 82.6% top-1 accuracy which is the best result as far as we know. Experiment

* 基金项目: 国家自然科学基金(61773385, 61672523)

Foundation item: National Natural Science Foundation of China (61773385, 61672523)

收稿时间: 2018-04-14; 修改时间: 2018-06-13, 2018-09-30; 采用时间: 2018-10-31

results demonstrate the effectiveness and superiority of the proposed approach.

Key words: zero-shot learning; image classification; dissimilarity representation; data preprocessing

零样本学习(zero-shot learning)是图像语义理解领域的一类尤为重要的机器学习方法.随着图像标注成本的不断增加,待识别类别愈加繁杂,利用大量训练数据通过监督学习构建图像分类模型的方法,已逐渐突显不足.零样本学习能够利用类别关联等先验知识,实现对未见类别图像的识别,而无需依赖大量的目标类(未见类)训练数据.零样本学习作为一种解决未见类图像分类问题的有效方法,近年来倍受关注^[1].

零样本学习方法通常利用可见类(存在训练样本的类)与未见类(不存在训练样本的类)之间的类别语义信息,将从可见类样本学习到的知识转移到未见类,实现对未见类样本的分类识别^[2-10].零样本学习是一种特殊的迁移学习.迁移学习是指对先前领域中学到的知识进行识别,并将其应用到一个全新的领域,其中先前领域称为源领域(source domain),新领域称为目标领域(target domain)^[11],而零样本学习是利用从可见类中学到的知识去识别未见类,其中可见类分布于源领域,未见类分布于目标领域且未见类样例无标签.此外,在一定条件下,零样本学习可以转化成监督学习,例如在获取了有标签的未见类样例的前提下,可以通过学习这些样例去识别真实的未见类样例.当前,零样本学习方法主要分为两大类:概率推理方法和特征映射方法.DAP(direct attribute prediction)^[9]方法利用了跨类别的属性预测和最大似然估计推导未见类的类别概率,是概率推理方法的基础^[1,12].随后,语义属性(semantic attribute)和确定性属性(discriminative attribute)优化了概率推理方法^[13,14].然而,概率推理方法过于依赖属性预测的正确性,并且,类别属性是类别层次上的描述,实例层次的语义描述上存在很大的偏差^[1].这些缺点限制了概率推理方法的发展和應用.特征映射方法形成了一条在不同嵌入之间进行样例特征比较的有效途径.利用特征映射方法,图像的低层次特征能够映射到高层次语义特征空间,直接与未见类的语义特征比较,形成分类.而且,确定性嵌入(discriminative embedding)可以通过 Word2Vec^[15]将类别文本转化为高维词向量,形成词向量空间中有效语义表示,便于特征空间映射^[16].另外,一些中间嵌入(intermediate embeddings)被引入到特征映射方法中^[17,18],用于在视觉流空间(visual manifold space)表示具体实例^[4],由此,视觉特征和属性特征将被映射到该空间中用于比较并分类.大量的确定性嵌入使特征映射方法具有了更加丰富的特征表示,实现了在不同特征空间下的未见类和可见类的样例分类.然而,特征映射方法仍然需要大量的标注数据和耗时的运算来训练模型,并且此类方法在原理上与人类零样本学习的认知过程不一致.

零样本学习是人类拥有的基本认知能力^[19].人类能够利用已有知识,参考和组合已见类物体的视觉特征,想象构造未见类物体的视觉特征,从而实现未见类物体的识别.例如,当一个人学习了“老虎是生活在森林中的具有斑纹毛皮的最大的猫科动物”的文字表述后,他会以自己见到的家猫为基本参考,通过想象替换掉家猫的毛皮和周边环境,继而产生老虎图像;此人会乐观地认可并学习自己想象的老虎图像,形成对“老虎”的视觉认知.由此,他具有了对未见类“老虎”的认知能力.此未见类图像的想象构造过程使得人类具有了零样本学习能力.模拟此未见类图像的构造过程,将形成一种有效的零样本学习方法.

本文提出了一种基于视觉特征组合构造的零样本学习方法,简称为 ViFC 方法.ViFC 方法模拟了人类根据语义描述联想到未见类物体视觉特征的过程.在参考最相似类样例特征的基础上,ViFC 方法通过不同特征维度上的特征组合,构造未见类的样例特征;并在测试样例批量输入的假设条件下,通过类别间的空间结构匹配优化,转换已构造的未见类样例特征,形成有效样例.ViFC 方法将零样本学习问题转化为标准的监督学习分类问题.传统的监督学习分类算法可以直接学习 ViFC 方法产生的未见类样例,构建分类模型,实现对真实的未见类样例的分类识别.

本文的贡献主要有以下3点:(1) 本文提出了一种模拟人类认知想象过程的零样本学习方法.此方法提取了属性与特征维度之间的关联关系,通过不同维度的特征组合,在特征层次上产生未见类的样例,将零样本学习问题转化为传统的监督学习分类问题,是一种从新角度解决零样本学习问题的方法.(2) 在测试样例批量输入的前提假设下,本文提出了类别间的空间结构匹配策略,有效地解决了特征域迁移问题.该策略借鉴非相似表示(dissimilarity representation)的原理,利用已产生的未见类样例与测试样例的空间结构匹配关系,获得未见类的

真实特征中心,实现已产生的未见类样例的特征域适应,从而解决了特征域偏移问题。(3) 本文提出了利用样例特征组合构造的方式产生未见类样例,此构造方式与特征映射和概率推理方法存在着显著的不同。在 AwA2 等数据集上的实验结果表明,ViFC 方法能够获得当前最优的分类正确率,优于已知的先进的零样本学习方法。

本文第 1 节简要介绍相关工作。第 2 节详细描述本文提出的 ViFC 方法。第 3 节详述 ViFC 方法在经典数据集上的大量实验对比和效能分析。第 4 节给出结论。

1 相关工作

人类学习新事物和新概念时,常常参考相似事物和相似概念,从而形成认知^[8]。在视觉场景下,人们会通过解析和组合物体外观来想象新事物的外观形象^[5]。2015 年,利用概率规划归纳的人类层次概念学习(human-level concept learning)在图灵测试中已获得巨大的进步^[8];在人类层次概念学习上,解析和组装对象是一种强有力的样例产生方法^[6,8,19,20]。然而,此类样例产生方法大多依赖实例级别的标注,仅有少量工作是基于类别级别的标注^[4]。

零样本学习利用类间共享关系和辅助信息(side information),比如类属性信息,针对未见类样例构建分类模型。目前,通过挖掘类间辅助信息的隐藏关系,很多零样本学习方法已经实现了在特定类别范围内对未见类样例的分类识别^[5,9,10,21],类间辅助信息构成了描述类别和样例的语义空间,形成了对类和样例的特征表示。当前,常见的两类语义空间是语义属性空间(semantic attribute space)和词向量空间(word vector space)^[2,5,22-24]。处于语义属性空间的类别属性,由领域专家人工标注;处于词向量空间的词向量表示,构建于类别的外部文本描述,如维基百科。与词向量表示相比,类别属性对类别样例的描述更加准确,因此,类别属性已被广泛应用于零样本学习方法中,为零样本学习方法提供了可靠的外部辅助信息。传统的基于属性的零样本学习方法,利用可见类训练样本的语义特征,构建一个有效的数据表征模型,利用此模型和类别关联关系,实现对未见类别样本的分类。例如,Lampert 等人^[10]提出了直接属性预测(direct attribute prediction,简称 DAP)和间接属性预测(indirect attribute prediction,简称 IAP),两种方法直接和间接地利用类别属性间的关联关系,构建属性预测模型和类别预测模型。Zhang 等人^[23]提出了一个隐空间概率模型,利用类别属性的关联关系,将样本的视觉特征和类别语义特征映射到隐空间,通过学习一个隐空间转换矩阵,对不同空间的特征匹配分类。Xian 等人^[25]通过大量实验比较了多个传统的零样本学习方法,揭示出当前零样本学习问题上最优的算法,其中 SYNC^[4]、ALE^[26]、SJE^[27]等基于属性的零样本学习方法表现突出。本文提出的 ViFC 方法是利用类别属性实现的零样本学习方法。

利用类别属性作为外部辅助信息的零样本学习方法,可以划分为两类:基于嵌入(embedding)的方法和基于相似类推理的方法。基于嵌入的零样本学习方法,多数利用映射函数将视觉特征投影为语义嵌入表示,并在此新的语义空间中搜索匹配未见类^[16-18,28]。在此类方法中,类别属性能够监督指导特征映射过程。融合了多种语义嵌入的零样本学习方法将视觉嵌入(visual embedding)和文本嵌入(textual embedding)映射到共同的中间嵌入空间,通过在共同空间下的特征匹配实现对未见类样例的分类识别^[18,29]。语义嵌入表示形式的优劣决定了此类方法的效能。基于相似类推理的方法是通过在类别层次上的相似度判断,关联未见类与可见类,直接针对未见类构建分类器^[3]。此类方法利用特征空间的转换、分类器组合和知识转化,实现对分类器参数和分类模型关系的推理预测,并依据推理预测结果构建未见类的分类器。此类方法过于依赖数学推理过程,难以有效应用在实际系统中。

可见类别与未见类别的不相交造成了零样本学习存在明显的域偏移问题。利用可见类别样本的空间分布,通过类别属性关系的关联,可以近似地获得未见类样例的空间分布,缓解域偏移问题。当前已有此类方法被提出。例如,Guo 等人^[30]基于类样本高斯分布的假设,根据可见类别和类属性的先验知识,预测未见类别概率分布,实现了有效空间分布的样例特征构造。直推零样本学习方法(transductive zero-shot learning)利用未见类别样本获得明确的特征域空间和映射函数,Fu 等人^[31]通过构建直推式嵌入空间将语义属性空间和词向量空间联系在一起,解决未见类别样本在不同域之间的偏移问题。Kodirov 等人^[7]提出了基于无监督域适应的零样本学习方法,使用目标域类别标签投影规范化目标域投影函数,克服域偏移问题。Li 等人^[32]提出了一个半监督框架,自动学习一个类别语义嵌入,将未见类别和可见类别融合成一个多类别分类问题。本文引入了半监督的未见类域适

应策略,指导构造的未见类样例适应性地调整到特定类的域位置,有效地缓解和解决域偏移问题。

从构建未见类样例的角度出发,可以将零样本学习问题转化为监督学习分类问题,此类方法近年来广受关注,构造未见类样例通常可分为两种方式,一种是产生符合特定分布的数据作为未见类样例^[33,34],另一种是利用对抗神经网络结合语义信息和视觉特征产生未见类样例^[35,36]。Wang 等人提出了一种通过学习类间语义关联知识在视觉特征空间中构造未见类样本的方法^[33]。该方法将语义空间中可见类和未见类的分布特征和高斯分布进行结合,生成尽可能符合未见类分布特征的样例。Changpinyo 等人提出了针对零样本学习任务的未见类视觉原型方法,并在多个数据集上获得了目前较先进的分类结果^[4]。然而此方法受限于每个类仅有一个原型,此特点降低了该方法的泛化能力。Long 等人提出了依赖实例级别属性标注的样例产生方法,并在多个数据集上取得了优异的分类效果,然而此方法不适用于类级别属性标注的零样本分类问题^[17]。Zhu 等人利用对抗网络识别视觉特征和基于文本信息生成的特征^[36],该方法对文本信息精确度要求较高,在粗粒度分类问题上适应性不强。与之前的方法不同,本文提出了利用视觉特征组合构造的方式产生未见类样例,并依据广受认可的非相似表示(dissimilarity representation)^[37]的思想,筛选未见类样例,保留最有代表性的未见类样例。本方法将零样本学习问题有效地转化为传统的基于监督学习的分类问题,效果突出。

2 基于视觉特征组合构造的零样本学习方法

$S = \{X_s, A_s, C_s\} / U = \{X_u, A_u, C_u\}$ 表示可见类/未见类的样本,类属性和类标签的集合;其中, $X_s = [x_1^s, \dots, x_m^s] \in \mathcal{R}^{d \times m}$ 和 $X_u = [x_1^u, \dots, x_n^u] \in \mathcal{R}^{d \times n}$ 分别表示 k_s 个可见类的 m 个样本和 k_u 个未见类的 n 个样本的视觉特征集合,视觉特征由高维特征向量表示; $A_s = [a_1^s, \dots, a_{k_s}^s] \in \mathcal{R}^{p \times k_s}$ 和 $A_u = [a_1^u, \dots, a_{k_u}^u] \in \mathcal{R}^{p \times k_u}$ 分别表示可见类别和未见类别的属性语义特征, p 表示属性个数; $C_s \in \mathcal{R}^{1 \times m}$ 和 $C_u \in \mathcal{R}^{1 \times n}$ 分别表示可见类别样本和未见类别样本的类别标签,且 $C_s \cap C_u = \emptyset$; $A = A_s \cup A_u$, $C = C_s \cup C_u$, $k = k_s + k_u$ 。本文的研究目标是在外部信息类别属性 A 的作用下,通过学习可见类别样本 S ,实现对未见类样例 X_u 的分类识别。

本文提出的 ViFC 方法试图在视觉特征层次上模拟人类联想记忆的认知过程,实现对未见类图像的分类识别。ViFC 方法能够在特征层次上产生大量未见类样例,这些新产生的样例与真实的未见类样例在视觉特征空间中具有相近的分布,此外,一方面受到非相似性表示(dissimilarity representation)^[37]思想的启发,ViFC 方法利用空间结构匹配策略,筛选产生的未见类样例,削弱特征域迁移(feature domain shift)的影响;另一方面,在测试样例批量输入的前提假设下,利用空间结构匹配策略,搜索产生的未见类样例与真实测试样例的类别对应关系,采取特征域适应,可以有效地解决特征域迁移问题。ViFC 方法的流程如图 1 所示,主要包含 4 个基本步骤,依次为:特征-属性关系提取、样例构造、样例过滤和特征域适应。经过 ViFC 方法的 4 步处理,大量有效的未见类样例被构建出来。由此,传统的监督学习分类算法,如支持向量机(SVM)和 K 近邻算法(KNN)等,可以直接学习产生的未见类样例,直接构建出针对未见类的分类模型。

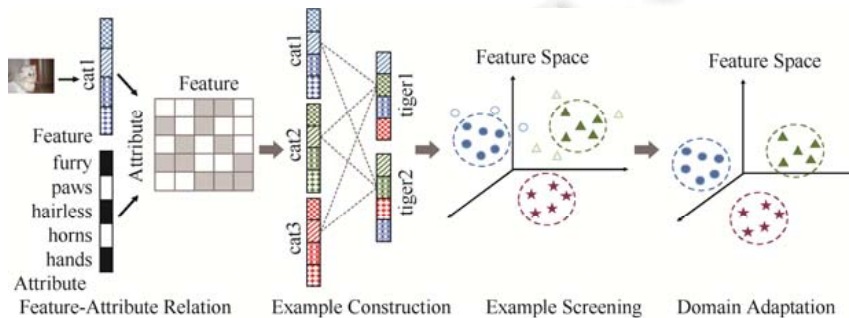


Fig.1 Flowchart of the visual feature combination approach for zero-shot learning

图 1 基于视觉特征组合构造的零样本学习方法流程图

2.1 特征-属性关系提取

物体的属性与其视觉特征的某些维度存在强关联关系^[18].在标注类别属性时,人们往往根据某类物体的部分视觉特征主观性地标注,并且标注的属性值存在模糊性.例如,在描述动物类别时,类别属性“四条腿”能够近似地对应到图像的某个视觉区域.Xian 等人已经通过实验验证了每个类别属性可以捕捉可见类图像的具体视觉区域^[21].因此,ViFC 方法利用此强关联性,获得特征-属性关系,并通过属性指导,构造未见类样例特征.

由于类别属性是模糊的,各属性对类别的描述准确性存在差异,ViFC 方法首先需要获得有辨别力的类别属性.设 $\beta = [\beta_1, \beta_2, \dots, \beta_p]$ 为属性系数向量, $\beta \in [0,1]$; \tilde{A} 表示权重转换后的类-属性矩阵,即 $\tilde{A} = [\beta_1 a_{1,1}, \beta_2 a_{2,1}, \dots, \beta_p a_{p,1}]$.为了获得有辨别力的类别属性,ViFC 方法希望找到一个属性系数向量 β ,使其非 0 的元素个数尽可能地少,并且使 \tilde{A} 能够尽可能地区分属性.属性系数向量 β 的求解可以转化为如下优化目标.

$$\operatorname{argmin} \|\beta\|_0 - \frac{1}{2} \sum_{\substack{i=1, j=1, \\ i \neq j}}^k \|\beta a_{i,1} - \beta a_{j,1}\| \quad (1)$$

优化目标(1)中, $\|\beta\|_0$ 表示 β 的稀疏度, $-\frac{1}{2} \sum_{\substack{i=1, j=1, \\ i \neq j}}^k \|\beta a_{i,1} - \beta a_{j,1}\|$ 是对类别间属性区分度的约束.ViFC 方法利

用演化算法求解此优化问题,得到有辨别力的类-属性关系矩阵 \tilde{A} .在属性系数向量 β 作用下,各个类的描述属性值发生变化,非强相关的属性被抑制或剔除,形成了对未见类和可见类具有更强辨识力的属性关系矩阵 \tilde{A} .例如,假设类-属性矩阵 A (行表示类,列表示属性),演化算法求解公式(1)获得属性系数向量 β ,则调整后具有更大辨识力的类-属性关系矩阵 \tilde{A} 如下所示.

$$A = \begin{bmatrix} 0.9 & 0.4 & 0.3 & 0.2 & 0.1 & 0.8 \\ 0.3 & 0.7 & 0.6 & 0.7 & 0.6 & 0.1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.4 & 0.2 & 0.8 & 0.2 & 0.6 & 0.7 \end{bmatrix}, \quad \beta = (0.7, 0.0, 0.9, 1.0, 0.0, 1.0), \quad \tilde{A} = \begin{bmatrix} 0.63 & 0.0 & 0.27 & 0.2 & 0.0 & 0.8 \\ 0.21 & 0.0 & 0.54 & 0.7 & 0.0 & 0.1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.28 & 0.0 & 0.72 & 0.2 & 0.0 & 0.7 \end{bmatrix}.$$

由此可知,第 2 列和第 5 列属性被剔除,不再参与后续的样例特征构造过程.

根据类-属性关系矩阵 \tilde{A} ,可以提取特征-属性关系.矩阵 \tilde{A} 的属性值处于 $[0,1]$ 范围内,代表了类和属性之间的关联程度.以 $[0,0.1]$ 的滑动窗口,将矩阵 \tilde{A} 的每个属性划分为 10 个不同属性区间,可见类样本在其对应类别上具有了离散的属性值,例如,属性值 0.25 对应了第 2 属性区间.针对每个属性,根据可见类样本的不同属性区间,利用 L_2 的线性支持向量分类(SVC)执行特征选择,产生特征-属性关系矩阵 D . D 是二值矩阵, $D_{ij}=1$ 表示属性 i 与特征维度 j 强相关, $D_{ij}=0$ 表示二者非强相关.有多种算法可以实现特征选择.通常情况下,当某类别属性描述的是该类别的局部视觉特性时,ViFC 方法采用 SVC 实现特征选择;当某类别属性表示的是该类别的整体特性时,例如属性“聪明”,ViFC 方法可采用 K 近邻算法(KNN).不同的特征选择算法会影响 ViFC 方法的最终分类结果.

例如,在利用特征提取获得特征-属性关系时, \tilde{A} 被转化为不同区间的离散矩阵,利用 SVC 执行特征选择,形成特征-属性对应关系矩阵. \tilde{A} 转化后的离散矩阵,针对每个属性产生的特征-属性矩阵 D (行表示属性,列表示特征维度)如下所示.

$$\text{离散化 } \tilde{A} = \begin{bmatrix} 6 & 0 & 2 & 2 & 0 & 8 \\ 2 & 0 & 5 & 7 & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 3 & 0 & 7 & 2 & 0 & 7 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

其中,离散化 \tilde{A} 采用整数区分不同的类别,SVC 针对每个属性在训练集上实现特征选择,得到各个属性对应的强相关维度; D 显示在 9 维表示的特征样例下,第 1 个属性与特征维度 0,1,4,6,7 强相关.

2.2 样例构造

在 ViFC 方法中,构建未见类的样例需要依据特征-属性关系 D 和相似类样例特征.特征-属性关系 D 建立了

未见类与可见类之间的信息共享桥梁,为构造未见类样例提供指导.

构建 C_u 中未见类 c_u 的样例时,其构造过程如下.

1. 确定 c_u 的相似类序列.根据类-属性关系 \tilde{A} 确定类别 c_u 在可见类集合 C_s 中的相似类,并依据 Euclidean 距离相似度计算,获得相似距离由小到大依此排列的相似类序列 $(s_0, s_1, \dots, s_{k_s})$.

2. 获得样例构造的基准参考样例 Y .根据 $(s_0, s_1, \dots, s_{k_s})$, 随机选择类别 s_0 中的样本,作为产生未见类 c_u 样例的基准参考样例集合 Y .

3. 确定类别 s_0 与 c_u 的差异属性.根据类-属性关系 \tilde{A} , 确定两个类别间的属性差异,即确定两类别在哪些属性上存在属性值的明显差异.当类-属性关系 \tilde{A} 属性值差值大于 0.2 时,称为存在属性值的明显差异.

4. 确定辅助类.根据步骤 3 确定的差异属性,从 $(s_0, s_1, \dots, s_{k_s})$ 中依次取可见类,得到最小辅助类集合 $\{s_1, \dots, s_i\}$.在 10 个不同属性区间上,此辅助类集合 $\{s_1, \dots, s_i\}$ 属性区间的并集覆盖了 c_u 的差异属性.例如, c_u 的某属性值为 0.25,类别 s_0 的此属性值为 0.92,存在属性值明显差异,在 $(s_0, s_1, \dots, s_{k_s})$ 中从小到大查找首个此属性值处于第 2 区间的类,置入 $\{s_1, \dots, s_i\}$.

5. 组合构造样例.根据不同选择策略,获得可见类 $\{s_1, \dots, s_i\}$ 的样本,根据 $\{s_1, \dots, s_i\}$ 与 c_u 的差异属性和特征-属性关系 D 确定特征替换维度;将可见类 $\{s_1, \dots, s_i\}$ 样本在替换维度上的特征值,替换掉基准参考样例 Y 的对应维度特征值.依次随机替换,直到覆盖全部差异属性.

6. 输出已完成全部差异属性特征替换的样例 Y .

在步骤 5 中,可以采用 3 种不同的选择策略,获得可见类样本.

1. 随机选择:随机选择可见类的样本.

2. 上下文相似选择:在特征-属性关系 D 中,确定了特征替换维度后,剩下的其他维度构成了一个样例的上下文环境维度.在上下文环境维度上,选择出与基准参考样例 Y 最相似的可见类样本.

3. 属性预测选择:根据特征-属性关系 D ,在特征替换维度上,选择预测的属性值与 c_u 的属性值最相似的可见类样本.此选择策略依赖 p 个属性分类器的属性值预测结果.

此构造过程的 5 步实现了大范围的合理性样例的特征构建.第 1 步实现了利用类别属性获得相似类别关系,确定基准参考类和辅助类.第 2 步获得与待构造的未见类最相似的可见类样本,作为构造未见类样例的基准参考样例,此基准参考样例往往与待构建的未见类样例非常相似,例如“杀人鲸”的样例与“蓝鲸”的样例.第 3 步用于确定需要改变基准参考样例的哪些属性和哪些对应的特征,才能构造出最吻合未见类属性描述的特征样例,以便于能够保留基准参考样例中与待构建类相似的特征,替换掉不吻合的特征.第 4 步用于确定采用哪些辅助类的样例特征,替换掉基准参考样例对应的特征,确保所有差异属性都能被有效地替换;选择辅助类时,仍需考虑类别相似性特点,例如构建“猪”的样例,替换“斑纹”属性时,优先采用“奶牛”而非“猫”的辅助类.第 5 步实施特征替换,随机次序地替换掉前面步骤确定的对应特征,构造出大量未见类的样例.在此步构造过程中,随机特性的加入增加了样本的多样性,构造的样本能够极大地吻合真实未见类的特征.

举例说明此样例构造过程:假设在动物的数据集上,构造未见类“老虎”的样例.根据类-属性关系矩阵 \tilde{A} 和欧式距离相似计算,获得“老虎”的可见类的相似类列表(猫、美洲豹、大猩猩、斑马),取可见类“猫”的一个样例作为基准样例;根据 \tilde{A} 确定“老虎”和“猫”两类之间的差异属性,如属性“在森林生活”上“老虎”的值为 0.92,而“猫”的值为 0.25,在“老虎”的相似类列表中,依次取相似类,获得在属性“在森林生活”上与“老虎”的属性值在同一个离散化类别的类,如“大猩猩”;基于 3 种不同选择策略从可见类“大猩猩”中选取样本,并根据特征-属性关系 D ,确定属性“在森林生活”与特征维度:0,1,4,6,7 强相关,由此用“大猩猩”样本在此维度上的特征值,替换“猫”的样例的同维度的特征值,至此完成该样例在属性“在森林生活”上的特征替换.依次进行差异属性的特征替换,构造出大量“老虎”样本.差异属性的替换次序对此构造过程存在一定的影响.

2.3 样例过滤

利用视觉特征组合构造的方式产生的样例中,存在大量的噪音样例,样例过滤能够剔除噪音样例.类似于非

相似表示(dissimilarity representation)^[37],ViFC 方法利用类-属性关系 \tilde{A} ,在类属性层次上,构建未见类在高维特征空间下的空间距离表示;同时在视觉特征层次上,形成已产生的未见类样例相对可见类样本中心的空间距离表示;在规范化空间距离表示后,计算样例的视觉特征层次上空间距离表示与属性层次上空间距离表示的相似度,通过阈值 d_{th} 过滤掉不相似的样例,保留相似的样例.

空间距离表示类似于非相似表示(dissimilarity representation),是采用当前样例与已知样例之间的相似距离形成的特征表示,图 2 示意性地说明了空间距离表示的原理.在属性空间中,“老虎”这一类可以由“老虎”与其他可见类的相似距离矢量来表示,即空间距离向量(0.20,0.80,0.70);在视觉特征空间中,第 2.2 节产生的未见类“老虎”的样例可以用该样例与可见类样本中心的相似距离矢量来表示,即空间距离向量(0.20,0.85,0.72).利用欧式距离计算,获得两个表示的相似度.视觉特征空间中相似度小于阈值上限 d_{th} 的“老虎”样例将被保留,作为未见类“老虎”的有效样本.

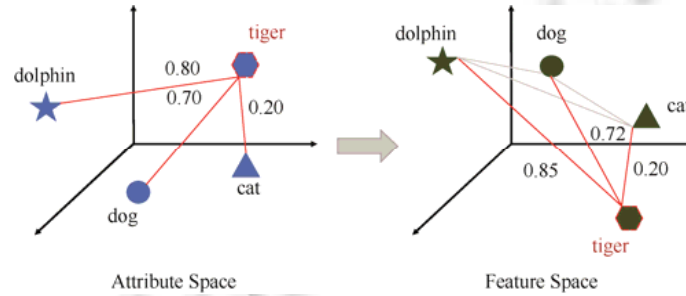


Fig.2 Space structure representation

图 2 空间结构表示图

2.4 特征域适应

因为特征域迁移问题普遍存在于零样本学习中^[7],所以解决特征域迁移问题是提升零样本学习方法有效性的必要条件.ViFC 方法采用特征域适应解决特征域迁移问题,通过样例特征的整体线性位移,产生了更为有效的未见类样例,从而缓解或解决特征域迁移问题.

ViFC 方法的特征域适应包含两种情况:无监督特征域适应和半监督特征域适应.

无监督特征域适应:在 ViFC 方法中,不增加测试样本的特征域适应.在 ViFC 方法中,针对初步构造出来的某个未见类的样例 Y ,其特征中心为 Y_c ,筛选后的样例特征中心为 Y_s ,无监督特征域适应通过 $Y = Y - (Y_c - Y_s)$ 获得整体样例的迁移.无监督特征域适应可以配合样例过滤,多次运行和调整,保留最接近未见类特征分布的样例.

半监督特征域适应:在 ViFC 方法中,假设未见类的测试样本是分批输入,并且分批输入的测试样本的类别数是已知的,在此条件下进行的特征域适应属于半监督域适应.前期的某些零样本学习方法,默认设定为测试样本的整体输入,其隐含了该假设,因此该假设在零样本学习中是合理的^[29].

在半监督特征域适应情况下,ViFC 方法利用聚类算法和空间距离表示实现对未见类的类别匹配,并根据匹配结果实现整体样例的迁移.具体过程如下.

1. 利用聚类算法,对输入的测试样本聚类,产生 k_u 个簇.
2. 分别计算 k_u 个簇的簇中心.
3. 在未见类范围内,计算 k_u 个簇中心的空间距离表示和样例 Y 中各类中心的空间距离表示,利用空间结构匹配,计算簇中心与类中心的匹配优化,确定测试样本簇中心和产生样例类中心的对应关系.

4. 筛选后的样例 Y ,线性迁移到匹配后的特征域.

半监督特征域适应的特征迁移受聚类算法和匹配优化的直接影响.在 ViFC 方法中,聚类采用 K-Means 算法,匹配优化采用演化算法.

3 实验

3.1 实验数据集和实验设置

实验数据集. ViFC 方法分别在 Aimals with Attributes(AwA)^[9]、Aimals with Attributes2(AwA2)^[38]和 SUN Attribute Dataset(SUN)^[39]上进行实验. 其中, AwA 是一个中等规模的粗粒度的动物图像数据集, 包含了 50 个类别的 30 475 张动物图片, 每类动物用一个 85 维的向量表示其属性; AwA2 是类似 AwA 的动物图像数据集, 其具有对外公开的原始图片, 同样采用 85 维属性向量表示 50 类动物, 共包含 37 322 张动物图片; SUN 包含了 717 个场景类别, 每类用 102 维属性向量表示, 每类含有 20 张图片. 针对零样本学习的验证需求, AwA 和 AwA2 采用 40/10 的可见/未见类的原始类别划分^[38]; SUN 数据集采用 707 个场景类别作为可见类, 10 个类别作为未见类^[39]. 以上 AwA、AwA2 和 SUN 数据集的划分是零样本学习研究中通用的划分方式.

视觉特征. 深度特征被证实具有有效的视觉特征表示能力, 已被广泛应用于图像识别和图像检索中. ViFC 方法使用 VGG19 网络 FC2 层提取的 4 096 维特征^[40]表示图片, 在视觉特征层面上构造未见类的样例. 同时, ViFC 方法中的属性值被转换到 [0, 1] 范围内, 用于表示某类中属性出现的概率, 并构成类-属性关系向量.

实施设置. 实验中, ViFC 方法采用 Euclidean 距离计算样例相似性; 类别属性被等分为 10 个区间, 利用多类支持向量机(multi-class SVM)实现对图片属性值的预测. 利用 DEAP^[41]的演化算法(evolutionary algorithm)在训练集上优化目标函数, 实现属性系数向量 β 的求解, 构建特征-属性关系矩阵. 在样例构造阶段, ViFC 方法融合了 3 种选择策略构造样例特征. 在特征域适应阶段, ViFC 方法分别实验了两种不同的特征域适应; 在半监督域适应中, 空间结构的匹配优化同样采用 DEAP 的演化算法^[41].

3.2 实验结果比较与分析

本实验验证了 ViFC 方法解决零样本学习问题的能力和效果. 为了直观地表示 ViFC 方法的分类效能, 本文引入了多个当前先进的零样本学习算法在实验数据集上的分类结果, 对比 ViFC 方法的分类结果, 形成分类效果的分析 and 评价.

表 1 比较了 ViFC 方法和当前多个先进的零样本学习算法在 3 个数据集上的 Top-1 分类正确率. 表 1 中的数据除特别标明以外, 都取自原始文章, “-”表示原始文章中并没有相应数据集的分类正确率, 其中, 为了统一对比标准, 增强实验结果的可比性, DEM^[28]等 4 种先进算法基于 VGG19 网络提取的特征在 3 个数据集上的复现结果被列入表 1, 复现过程中的参数与原始参数保持一致. 如表 1 所示, 与当前已知的先进的零样本学习算法相比, ViFC 方法在 AwA, AwA2 和 SUN 数据集上都取得了优异的分类结果.

无监督特征域适应的 ViFC 方法(ViFC_{unsuper-adapt}) 在 AwA 和 AwA2 数据集上分别实现了 82.6% 和 81.4% 的分类正确率, 分类结果均优于当前所有的先进算法. ViFC 方法能够较好的适应 AwA 和 AwA2 数据集, 其原因是这两个数据集的属性是对图像的部分视觉区域的描述, 这一特点与 ViFC 方法的特征组合构造的特点相吻合. 而 SUN 数据集的很多属性是对图片整体场景的描述, 往往表示的是抽象的视觉特性, 例如属性“Cold”和“Digging”, 这些属性与整张图片的视觉特征都存在关系. 此特点会恶化 ViFC 方法的域偏移问题. 因此, ViFC 方法在 SUN 数据集上的分类效果不如在 AwA 数据集上明显. 但是 ViFC_{unsuper-adapt} 方法和 ViFC_{non-adapt} 方法在 SUN 数据集上仍取得相对较高的分类正确率: 75.5% 和 82.0%, 尤其是与 GAZSL^[36]、CLSWGAN^[35]等其他通过生成样例解决零样本问题的算法相比, ViFC 方法在 SUN 数据集上表现更为优异, 而 GAZSL^[36]、CLSWGAN^[35]等算法在 ResNet101 网络提取的较低维度(2 048 维)的特征上表现更优. ViFC 方法采用的是特征组合的方式构造样例, 更高维度的视觉特征与属性的关系匹配更为精准, 构造的样例也更为有效, 因此区别于其他生成样例的算法^[35, 36], ViFC 采用的是更高维度(4 096 维)的视觉特征. 综合表 1 所示数据, ViFC 方法通过视觉特征组合构造的方式产生的未见类样例特征能够在一定程度上覆盖未见类的样例特征域, 构造过程合理而有效.

Table 1 Top-1 accuracy (%) comparison of ViFC with the state-of-the-art zero-shot learning approaches**表 1** ViFC 方法和当前先进的零样本学习方法的比较(Top-1 分类正确率(%))

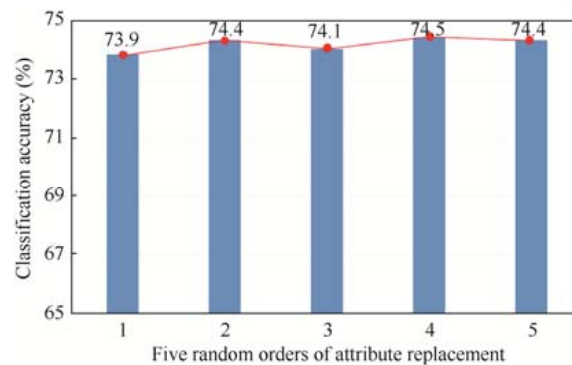
方法	特征	数据集		
		AwA	AwA2	SUN
JLSE ^{[24]-vi}	VGG19	80.5	—	82.8
JLSE ^{[24]-viii}	VGG19	79.1	—	83.8
SSE ^{[23]-INT}	VGG19	71.5	—	82.2
SSE ^{[23]-RELU}	VGG19	76.3	67.5 [§]	82.5
MLZSC ^[2]	VGG19	77.3	—	84.4
DAP ^[10]	VGG19	57.2	58.7 [§]	72.0
SynC ^{o-vs-o[3]}	VGG19	64.0	—	—
SynC ^{cs[3]}	VGG19	64.8	—	—
SynC ^{struct[3]}	VGG19	62.8	71.2 [§]	—
ESZSL ^[42]	VGG19	75.3	75.6 [§]	82.1
DEM ^[28]	VGG19 [#] /ResNet101	72.1/68.4	78.4/67.1	27.0/—
GAZSL ^[36]	VGG19 [#] /ResNet101	60.3/68.2	51.2/69.0	27.5/61.3
CLSWGAN ^[35]	VGG19 [#] /ResNet101	21.1/68.2	22.7/—	18.0/60.8
ViFC _{non-adapt}	VGG19	74.6	75.3	75.5
ViFC _{unsuper-adapt}	VGG19	82.6	81.4	82.0
ViFC _{semi-adapt}	VGG19	94.2	94.8	41.5

#:使用 VGG19 网络提取的特征复现该方法所得正确率

§:Xian 等人^[25]使用 ResNet101 网络提取的特征复现该方法所得正确率^[25]

通过无监督特征域适应,ViFC 方法在 3 个数据集上的分类正确率都有显著的提升,说明产生的未见类样例在经过域迁移后在特征域中的分布更接近于真实未见类样例的分布,所以筛选后的样例能够更好地体现未见类的特点,因此无监督特征域适应方法能够在一定程度上减小域间偏移问题.由于类属性所提供的辅助信息有限,相同类在不同域之间仍存在偏移,因此该问题并未彻底解决.

在构造样例过程中,基准参考样例与目标样例的差异属性所对应的特征维度会被替换,因此差异属性的替换次序对实验结果可能会产生一定的影响.图 3 展示了 ViFC_{non-adapt} 方法采用 5 种随机属性替换次序在数据集 AwA 上的实验结果.实验结果表明,在构造过程中,替换次序的不同,对分类结果不会产生关键性的影响,分类正确率差异不大.随机选择替换维度的次序能够保证构造样例的多样性,从而使构造出来的样例在特征表示的层面上,大概率地覆盖未见类属性所对应的特征表示.ViFC 方法采用了随机次序的属性替换.

**Fig.3** Accuracy comparison of 5 random orders of attribute replacement in ViFC_{non-adapt}**图 3** ViFC_{non-adapt} 方法中 5 个随机属性替换次序的正确率比较

半监督特征域适应的 ViFC 方法(表 1 中标记为 ViFC_{semi-adapt})产生了尤为突出的分类结果.在此实验中,以每类 10 个随机测试样本形成一个批量输入.在 AwA 和 AwA2 数据集上,半监督特征域适应 ViFC 分别获得了 94.2%

和 94.8% 的 Top-1 分类正确率;而在 SUN 数据集上,其只获得了 41.5% 的 Top-1 分类正确率.通过分析此实验过程和实验结果,可以发现不同数据集上分类正确率的巨大差异,主要是由批量测试数据的聚类效果不同导致.在 AwA 和 AwA2 数据集上,批量输入的测试样本能够被 K-means 算法有效聚类,其聚类后产生的簇与其真实类别的样例集合是一致的.因此,在 ViFC 方法的特征域适应阶段中,如果簇中心和类别中心能够正确匹配,ViFC 方法构造的未见类样例将会迁移到真实的特征域位置,这将使得在这些样例上训练的 SVM 模型获得较高的分类正确率.在 SUN 数据集上,情况则截然不同.由于其测试样本难以形成正确的聚类结果,使得特征域适应时,构造的未见类样例错误地迁移,造成分类结果变差.此实验结果表明,半监督特征域适应 ViFC 方法能够适应特定的数据集,产生优良的零样本学习分类结果,但存在通用性差的缺陷.

在半监督特征域适应时,批量输入的测试样本量对分类正确率存在影响.表 2 显示了在批量输入测试样本时,在每类采用不同数量样本的情况下,ViFC 方法获得的分类正确率.从表 2 可以看出,当每类以随机的 50 个样本批量输入时,ViFC 方法在 AwA 和 AwA2 数据集上获得了 95.1% 和 94.9% 的分类正确率.由于 SUN 数据集每个未见类最多含有 20 个样本,其在全部测试样本都参与批量输入时,ViFC 方法获得的分类正确率最高(62.0%).半监督特征域适应的优化效果依赖于数据集样本分布特点.在 AwA 和 AwA2 中,各类的样本数量和空间分布比较类似,样本量足够体现各类别特点,并与类别属性表示的类别空间分布相吻合.因此,在半监督特征域适应时,其聚类中心的空间分布与类别属性表示的空间分布能够相互匹配,从而获得了有效的特征迁移的目标领域.SUN 数据集的情况则恰恰相反,每一类仅有 20 个样本,样本数量少,类内样本特征差异大,无法形成有效的聚类,其聚类的中心构成的类别空间分布与属性表示的类别空间分布不能有效匹配,从而造成半监督特征域适应时,构造样例迁移到错误的目标领域,产生了无效的未见类类别样例,分类结果变差.因此,半监督特征域适应的 ViFC 方法不适合样本类别多、类内样本多样性大且样本数量少的情况.

Table 2 Classification accuracy comparison of ViFC_{semi-adapt} with different category sizes as input

表 2 在每类输入不同样本数量情况下 ViFC_{semi-adapt} 方法的分类正确率的比较

数据集	批量输入时每类含有的样本数量(%)			
	10	20	50	ALL
AwA	94.2	93.3	95.1	94.1
AwA2	94.8	93.1	94.9	94.2
SUN	41.5	62.0	-	62.0

在 ViFC 方法中,阈值的设定直接影响了样例过滤的效果.当阈值 d_{th} 过大时,不符合真实未见类数据分布的样例没有被过滤掉;当阈值 d_{th} 过小时,过滤后的样例不能全面地体现出真实数据的分布特点,使得构造的样例失去了多样性,以上两种情况都无法使特征域适应的效果达到最优.表 3 给出了在不同的阈值 d_{th} 设置下,ViFC 方法在 AwA 上的分类结果.表 3 中,maxValue、minValue、meanValue、mediumValue 表示每一类构造出来的样例在视觉特征空间中相似距离的最大值、最小值、平均值、中间值.表 3 说明,阈值 d_{th} 的选择对 ViFC 方法的有效性有较大的影响,合理的阈值选择策略可以有效地提升特征域适应的效果.本文采用经验值设定的方式,从表 3 所列的 5 种阈值设置方式中选取最优的阈值.

Table 3 Classification accuracy of ViFC on AwA dataset with different threshold d_{th} values

表 3 AwA 数据集中 ViFC 在不同阈值 d_{th} 下的分类正确率

阈值 d_{th}	设置方式 1	设置方式 2	设置方式 3	设置方式 4	设置方式 5
	meanValue	mediumValue	<u>maxValue - meanValue</u>	<u>maxValue - mediumValue</u>	<u>meanValue - minValue</u>
正确率(%)	80.0	80.4	78.2	80.5	82.6

图 4 展示了 ViFC 方法在 AwA 数据集上针对各个未见类的分类结果 ROC 曲线.从图 4 可以看出,ViFC 方法能够平衡而有效地对 10 类未见类样本进行分类.而且对于“黑猩猩”和“座头鲸”两类的分类,ViFC 方法的分类 AUC 值为 1.这充分体现了 ViFC 方法的有效性.另一方面,图 4 显示“猪”类的 AUC 值为 0.93,此类分类正确率低的原因是其基准参考类与其距离值较大,根据基准参考样例不易产生有效的“猪”的样例.具体而言,“猪”类的基

准参考类是“牛”,距离值为 0.099 8,相比于“杀人鲸”与其基准参考类“蓝鲸”的距离 0.044 3,其类别属性上的差异是比较大的,因此利用“牛”的样本构建“猪”的样例,难度较大,造成了较差的 AUC 结果.

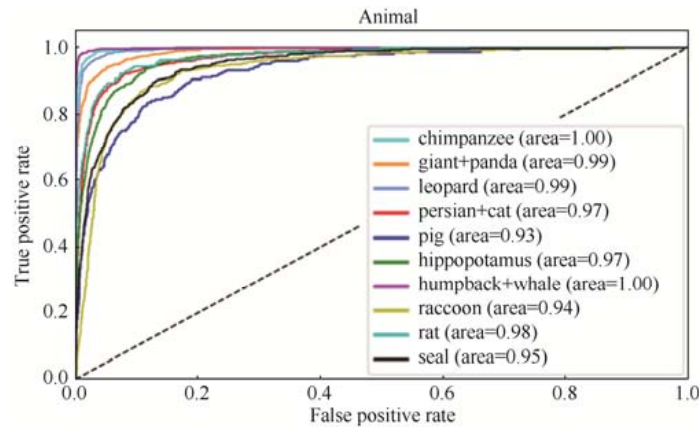


Fig.4 ROC curves of ViFC on the dataset AwA

图4 ViFC 方法在 AwA 数据集上的 ROC 曲线图

为了展现 ViFC 方法产生的未见类样例的分布特点,图 5 通过 t-SNE 可视化^[43]显示了 ViFC 方法构造的未见类样例和其对应的实际未见类测试样本的二维分布图.在图 5 中,构造产生的未见类样例采用星形标示,实际的未见类测试样本采用圆形标示,同种颜色表示同种类别.从图 5 可以看出,经过样例过滤后 ViFC 方法所构造的样例多数位于实际未见类样本的覆盖区域.这些构造样例能够有效的表示未见类,并可以直接训练分类器,形成针对未见类的分类模型.这进一步解释了 ViFC 方法在零样本学习任务上有效的原因.

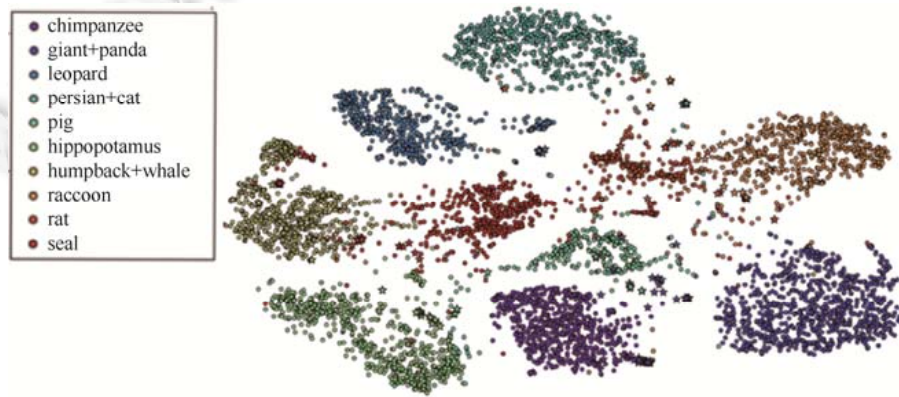


Fig.5 T-SNE visualization of the real test examples (circles) and the generated examples (stars) of ViFC on AwA dataset

图5 AwA 数据集上 ViFC 构造样例和真实测试样例的 t-SNE 可视化分布图

ViFC 方法构造的样例与真实的未见类样例具有一定相似性.表 4 列出了在 AwA 数据集的所有未见类上,ViFC_{non-adapt} 方法(迁移前)与 ViFC_{unsuper-adapt} 方法(迁移后)构造的样例的类中心与实际未见类的类中心的欧式距离(Euclidean)与余弦距离(Cosine).表 4 显示,在半监督特征域适应的作用下,ViFC_{unsuper-adapt} 构造的样例在整体上与真实的未见类的类中心距离更近,尤其是余弦距离,这说明 ViFC_{unsuper-adapt} 构造的样例在空间结构或者数据分布上与真实样例更为相似,同时也说明 ViFC 方法是一种有效的解决零样本学习的样例产生方法.

Table 4 Distance between centers of constructed examples and real centers of unseen classes

表 4 构造样例类中心与真实未见类样例类中心的距离

	迁移前(Euclidean)	迁移后(Euclidean)	迁移前(Cosine)	迁移后(Cosine)
chimpanzee	0.568 1	0.582 9	0.264 1	0.267 0
giant+panda	0.800 9	0.752 4	0.559 6	0.482 4
leopard	0.678 0	0.696 0	0.413 2	0.432 5
persian+cat	0.717 5	0.682 7	0.478 2	0.452 2
pig	0.559 2	0.490 2	0.376 6	0.301 0
hippopotamus	0.479 8	0.484 9	0.232 3	0.227 5
humpback+whale	0.188 9	0.202 2	0.032 6	0.037 9
raccoon	0.690 0	0.615 4	0.399 6	0.349 1
rat	0.484 7	0.505 2	0.260 8	0.255 4
seal	0.489 1	0.484 5	0.265 5	0.252 1

由于 ViFC 方法是在特征层次上构造样例,样例特征难以转化为具体图片,难以直观地显示构造样例.然而,通过特征层次上的相似度判断,可以得到构造样例与实际未见类样本的相似距离,从而判断构造样例的合理性.图 6 显示了在 SUN 数据集的测试图片中,与构造样例相似的 Top-5 图像.图 6 中,位于图片下的实数值是此图片特征与构造样例特征的欧式距离,红色边框的图片表示正确匹配的同类图片,绿色边框的图片表示没有正确匹配的图片.从图 6 可以看出,ViFC 方法产生的样例(如构造样例 x_1, x_3, x_4)与它们所属的未见类中的实际样本是最相似的.对于错误匹配的样例(如构造样例 x_2),其最匹配的实际样本在外观上也会让人们混淆,产生错误分类.这进一步说明了 ViFC 方法能够从模拟人类联想认知的角度出发,构造未见类样例.所以,图 6 从构造图像的视觉合理性上揭示了 ViFC 方法的有效性.



Fig.6 Top-5 most similar test images to an example generated by ViFC

图 6 在测试集中与 ViFC 产生的样例最相似的 Top-5 图片实例

综上,实验验证了 ViFC 方法在零样本学习任务上具有卓越分类效能和可解释性的内部原理,并能有效地模拟人类识别未见类事物的过程.

4 结 论

受到人类具有的对未见类事物联想认知能力的启发,本文提出了一种基于视觉特征组合构造的零样本学习方法(ViFC 方法).在类别属性信息的辅助作用下,ViFC 方法能够获得属性与特征维度之间的关联关系,并

通过特征组合的方式构造未见类样例的特征,将零样本学习问题转化为传统的监督学习分类问题,从而实现未见类样本的有效分类.对于 ViFC 方法,在特定数据集上,无监督特征域适应和半监督特征域适应都能解决零样本学习的特征域迁移问题,并且半监督特征域适应揭示了在零样本学习中未见类样本存在明确的聚类分布的特点,这为将来的研究工作提供了新的启发.在 3 个经典数据集上的实验结果表明,ViFC 方法是一种通过视觉特征组合构造解决零样本学习问题的有效方法,且在 AwA 数据集上取得了当前最优的分类结果.在后续研究工作中,将会引入特征投影,研究更有效的未见类样例过滤策略,产生更加符合实际未见类样本特点的样例特征.

References:

- [1] Jayaraman D, Grauman K. Zero-Shot recognition with unreliable attributes. In: *Advances in Neural Information Processing Systems*. 2014. 3464–3472.
- [2] Bucher M, Herbin S, Jurie F. Improving semantic embedding consistency by metric learning for zeroshot classification. In: *Proc. of the ECCV*. 2016. 730–746.
- [3] Changpinyo S, Chao WL, Gong B, Sha F. Synthesized classifiers for zero-shot learning. In: *Proc. of the CVPR*. 2016. 5327–5336.
- [4] Changpinyo S, Chao WL, Sha F. Predicting visual exemplars of unseen classes for zero-shot learning. *arXiv preprint arXiv:1605.08151*, 2016.
- [5] Farhadi A, Endres I, Hoiem D, Forsyth D. Describing objects by their attributes. In: *Proc. of the CVPR*. 2009. 1778–1785.
- [6] Jern A, Kemp C. A probabilistic account of exemplar and category generation. *Cognitive Psychology*, 2013,66(1):85–125.
- [7] Kodirov E, Xiang T, Fu Z, Gong S. Unsupervised domain adaptation for zero-shot learning. In: *Proc. of the ICCV*. 2015. 2452–2460.
- [8] Lake BM, Salakhutdinov R, Tenenbaum JB. Human-Level concept learning through probabilistic program induction. *Science*, 2015, 350(6266):1332–1338.
- [9] Lampert CH, Nickisch H, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer. In: *Proc. of the CVPR*. 2009. 951–958.
- [10] 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.
- [11] Yang L, Jing LP, Yu J. Heterogeneous transductive transfer learning algorithm. *Ruan Jian Xue Bao/Journal of Software*, 2015, 26(11):2762–2780 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4892.html> [doi: 10.13328/j.cnki.jos.004892]
- [12] Rohrbach M, Ebert S, Schiele B. Transfer learning in a transductive setting. In: *Advances in Neural Information Processing Systems*. 2013. 46–54.
- [13] Wang X, Ji Q. A unified probabilistic approach modeling relationships between attributes and objects. In: *Proc. of the ICCV*. 2013. 2120–2127.
- [14] Yu X, Aloimonos Y. Attribute-Based transfer learning for object categorization with zero/one training example. In: *Proc. of the ECCV*. 2010. 127–140.
- [15] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: *Advances in Neural Information Processing Systems*. 2013. 3111–3119.
- [16] Lei BJ, Swersky K, Fidler S. Predicting deep zero-shot convolutional neural networks using textual descriptions. In: *Proc. of the ICCV*. 2015. 4247–4255.
- [17] Long Y, Liu L, Shao L. Towards finegrained open zero-shot learning: Inferring unseen visual features from attributes. In: *Proc. of the WACV*. 2017. 944–952.
- [18] Lu Y. Unsupervised learning on neural network outputs: With application in zero-shot learning. *arXiv preprint arXiv:1506.00990*, 2015.
- [19] Biederman I. Recognition-by-Components: A theory of human image understanding. *Psychological Review*, 1987,94(2):115–147.
- [20] Papadopoulos C, Hayes BK, Newell BR. Noncategorical approaches to feature prediction with uncertain categories. *Memory & Cognition*, 2011,39(2):304–318.
- [21] Xian Y, Akata Z, Sharma G, Nguyen Q, Hein M, Schiele B. Latent embeddings for zero-shot classification. In: *Proc. of the CVPR*. 2016. 69–77.
- [22] Norouzi M, Mikolov T, Bengio S, Singer Y, Shlens J, Frome A, Corrado GS, Dean J. Zero-Shot learning by convex combination of semantic embeddings. *arXiv preprint arXiv:1312.5650*, 2013.
- [23] Zhang Z, Saligrama V. Zero-Shot learning via semantic similarity embedding. In: *Proc. of the ICCV*. 2015. 4166–4174.

- [24] Zhang Z, Saligrama V. Zero-Shot learning via joint latent similarity embedding. In: Proc. of the CVPR. 2016. 6034–6042.
- [25] Xian Y, Schiele B, Akata Z. Zero-Shot learning—The good, the bad and the ugly. In: Proc. of the CVPR. 2017. 3077–3086.
- [26] Akata Z, Perronnin F, Harchaoui Z, Schmid C. Label-Embedding for image classification. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016,38(7):1425–1438.
- [27] Akata Z, Reed S, Walter D, Lee H, Schiele B. Evaluation of output embeddings for fine-grained image classification. In: Proc. of the CVPR. 2015. 2927–2936.
- [28] Zhang L, Xiang T, Gong S. Learning a deep embedding model for zero-shot learning. In: Proc. of the CVPR. 2017. 3010–3019.
- [29] 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 with English abstract). <http://www.jos.org.cn/1000-9825/5338.html> [doi: 10.13328/j.cnki.jos.005338]
- [30] Guo Y, Ding G, Han J, Gao Y. Synthesizing samples for zero-shot learning. In: Proc. of the IJCAI. 2017. 1774–1780.
- [31] Fu Y, Hospedales TM, Xiang T, Gong S. Transductive multi-view zero-shot learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(11):2332–2345.
- [32] Li X, Guo Y, Schuurmans D. Semi-Supervised zero-shot classification with label representation learning. In: Proc. of the ICCV. 2015. 4211–4219.
- [33] Wang D, Li Y, Lin Y, Zhuang Y. Relational knowledge transfer for zero-shot learning. In: Proc. of the AAAI. 2016. 2145–2151.
- [34] Zhao B, Wu B, Wu T, Wang Y. Zero-Shot learning posed as a missing data problem. In: Proc. of the ICCVW. 2017. 2616–2622.
- [35] Xian Y, Lorenz T, Schiele B, Akata Z. Feature generating networks for zero-shot learning. In: Proc. of the CVPR. 2018.
- [36] Zhu Y, Elhoseiny M, Liu B, Peng X, Elgammal A. A generative adversarial approach for zero-shot learning from noisy texts. In: Proc. of the CVPR. 2018. 1004–1013.
- [37] Duin PD, Elzbieta P. The dissimilarity space: Bridging structural and statistical pattern recognition. Pattern Recognition Letters, 2012,33(7):826–832.
- [38] Xian Y, Lampert CH, Schiele B, Akata Z. Zero-Shot learning—A comprehensive evaluation of the good, the bad and the ugly. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018. [doi: 10.1109/TPAMI.2018.2857768]
- [39] Patterson G, Xu C, Su H, Hays J. The sun attribute database: Beyond categories for deeper scene understanding. International Journal of Computer Vision, 2014,108(1-2):59–81.
- [40] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proc. of the ICLR. 2015.
- [41] Fortin FA, Rainville FM, Gardner MA, Parizeau M, Gagné C. DEAP: Evolutionary algorithms made easy. Journal of Machine Learning Research, 2012,13:2171–2175.
- [42] Romera-Paredes B, Torr P. An embarrassingly simple approach to zero-shot learning. In: Proc. of the ICML. 2015. 2152–2161.
- [43] Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008,9(11):2579–2605.

附中文参考文献:

- [11] 杨柳,景丽萍,于剑.一种异构直推式迁移学习算法.软件学报,2015,26(11):2762–2780. <http://www.jos.org.cn/1000-9825/4892.html> [doi: 10.13328/j.cnki.jos.004892]
- [29] 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.html> [doi: 10.13328/j.cnki.jos.005338]



杨刚(1979—),男,山东青岛人,博士,副教授,CCF 专业会员,主要研究领域为多媒体信息检索,深度学习,神经网络,软计算.



李锡荣(1983—),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为人工智能,媒体计算.



刘金鑫(1995—),女,硕士生,主要研究领域为计算机视觉.



许洁萍(1966—),女,博士,副教授,CCF 专业会员,主要研究领域为多媒体信息处理.