零样本学习研究进展

张鲁宁 [1] 左 信 [1] 刘建伟 [1]

摘 要 近几年来, 深度学习在机器学习研究领域中取得了巨大的突破, 深度学习能够很好地实现复杂问题的学习, 然而, 深度学习最大的弊端之一, 就是需要大量人工标注的训练数据, 而这需要耗费大量的人力成本. 因此, 为了缓解深度学习存在的这一问题, Palatucci 等于 2009 年提出了零样本学习 (Zero-shot learning). 零样本学习是迁移学习的一种特殊场景, 在零样本学习过程中, 训练类集和测试类集之间没有交集, 需要通过训练类与测试类之间的知识迁移来完成学习, 使在训练类上训练得到的模型能够成功识别测试类输入样例的类标签. 零样本学习的意义不仅在于可以对难以标注的样例进行识别, 更在于这一方法模拟了人类对于从未见过的对象的认知过程, 零样本学习方法的研究, 也会在一定程度上促进认知科学的研究. 鉴于零样本学习的应用价值、理论意义和未来的发展潜力, 文中系统综述了零样本学习的研究进展, 首先概述了零样本学习的定义, 介绍了 4 种典型的零样本学习模型, 并对零样本学习存在的关键问题及解决方法进行了介绍, 对零样本学习的多种模型进行了分类和阐述, 并在最后指明了零样本学习进一步研究中需要解决的问题以及未来可能的发展方向.

关键词 零样本学习, 描述, 属性, 训练类, 测试类, 嵌入空间

引用格式 张鲁宁, 左信, 刘建伟. 零样本学习研究进展. 自动化学报, 2020, **46**(1): 1−23

DOI 10.16383/j.aas.c180429

Research and Development on Zero-Shot Learning

ZHANG Lu-Ning[1] ZUO Xin[1] LIU Jian-Wei[1]

Abstract In recent years, deep learning has made great breakthroughs in the field of the machine learning. The application of deep learning can be especially useful for coping with some complicated problems. However, one of the biggest drawbacks of the deep learning is that it requires a great amount of manual data annotation, this issue requires a lot of labor costs. Therefore, in order to alleviate the burden of deep learning, Palatucci et al. proposed zero-shot learning in 2009. Zero-shot learning is a special scenario of transfer learning. In the zero-shot learning progress, the samples of training and test classes do not intersect. It is necessary to complete the training by realizing the knowledge transfer between training class and test class, in order to successfully identify the label of the test class instance. The significance of zero-shot learning lies not only in the identification of difficult-to-annotate, but also in the fact that this method simulates the human cognitive process of objects that have never been seen before, so the study of zero-shot learning will contribute to the advancement of human cognitive science research. In view of the application value, theoretical significance and future development potential of zero-shot learning, this paper systematically reviews the research progress of zero-shot learning. First, the definition of zero-shot learning is summarized, then we introduce four typical zero-shot learning models, and the systematic problems in the zero-shot learning and the solution methods are introduced. We have categorized and elaborated on the multiple models of zero-shot learning. At the end, we pointed out the problems that need to be solved in the further study of zero-shot learning and possible future development directions.

Key words Zero-shot learning, description, attribute, training class, test class, embedded space

Citation Zhang Lu-Ning, Zuo Xin, Liu Jian-Wei. Research and development on zero-shot learning. *Acta Automatica Sinica*, 2020, **46**(1): 1−23

随着机器学习领域的发展, 机器学习在自然图像识别领域也取得了长足的进步, 在对于车辆、人脸等特定对象的识别与分类等方面尤为突出. 因此, 机器学习技术广泛地在这些领域中投入商业使用, 例如支持向量机 (Support vector machine, SVM)[1]、卷积神经网络[2] 和递归神经网络[3] 等. 但是, 现有的识别模型如果想要得到较高的预测准确度, 都需要大量的人工标注样本进行训练, 一般来说, 每一个对象类, 都需要数以千计的标注样本.

随着图像识别技术应用的更加广泛, 以及需要进行识别的对象类不断增加, 未来图像识别领域

收稿日期 2018-06-15 录用日期 2018-08-30
Manuscript received June 15, 2018; accepted August 30, 2019
国家重点研发计划项目 (2016YFC0303703-03), 中国石油大学 (北京) 年度前瞻导向及培育项目 (2462018QZDX02) 资助
Supported by National Key Research and Development Program (2016YFC0303703-03) and China University of Petroleum (Beijing) Prospective Orientation and Cultivation Project (2462018QZDX02)
本文责任编委 张敏灵
Recommended by Associate Editor ZHANG Min-Ling
1. 中国石油大学 (北京) 自动化系 北京 102249
1. Department of Automation, China University of Petroleum (Beijing), Beijing 102249

的发展不应完全寄希望于这种需要大量训练样本的学习方法. 例如, 人类能够识别大约 30 000 个类中所包含的对象, 还可以对这些类中所包含的子类进行辨别, 例如不同款式的汽车[4], 或者不同品种的狗[5]. 甚至, Murphy 认为, 人类可以在无限数目的类中完成分类任务, 因为人类可以随时创造新类[6]. 理论上如果使用现有的机器学习模型实现上述功能, 至少需要数百万, 甚至数亿个高质量标注的训练样本, 而且训练时间也会显著增加.

而且, 对于某些特定的对象类, 训练样本是难以获得的. 以濒危物种为例, 由于处于濒危状态, 其图像资料是难以获得、极为珍贵的, 同时也正因图像资料的重要性, 如果能够实现对于特定对象类不依赖于大规模训练样本 (因为特定类图像资料较少, 无法建立有效的训练样本集) 的野外的濒危物种识别、摄录, 将会带来巨大的商业价值和生态价值.

尽管存在一些减少训练样本和提高训练效率的算法[7−10], 但是, 这些算法仍然需要一定数量的训练样本对模型中的特定类进行训练, 才能实现对测试样本中的测试样例的分类和预测. 人类学习机制与现有的机器学习机制相比具有很大的差异, 人类通常可以在大量的训练样本上很好地进行学习, 但人类也可以在少量或无样本情况下, 通过其他与所要学习的目标相关的辅助信息 (Side information), 完成对特定目标的学习. 在机器学习领域中, 能够对从未见过的对象类中的样例进行识别的能力, 即为零样本学习 (Zero-shot learning).

零样本学习衍生于迁移学习[11], 是迁移学习的变种之一, 零样本学习与其他迁移学习最主要的区别是, 训练类样本集和测试类样本集没有交集. 随着近年来的不断发展, 零样本学习已经逐渐脱离迁移学习, 成为一个独立的机器学习研究方向. 零样本学习方法与现有的分类方法相比, 具有如下三点优势:

1) 对于某些还没有建立样本集的特定类 (例如新定种的生物物种或濒危物种, 最新设计的工业产品等), 通过零样本学习, 可以成功地对这些对象进行识别、分类, 既能满足实际需求, 又可以降低人工和经济成本.

2) 零样本学习的核心机制与人类的学习机制有很多的共通之处, 对于零样本学习进行深入的研究, 会为人类认知科学领域提供强有力的帮助.

3) 零样本学习与深度学习并不矛盾, 两者可以有机结合、博采众长、融合发展, 从而更好地满足未来对象识别领域的需求.

鉴于零样本学习的理论意义, 所蕴含的应用价值以及可观的发展潜力, 本文对零样本学习的研究进展进行了系统性的综述, 为进一步深入研究零样本学习机制、开发零样本学习应用潜力确立良好的基础. 文中首先在第 1 节对零样本学习进行了概述, 阐明零样本学习的发展过程以及定义; 并在第 2 节着重介绍了零样本学习初始阶段具有重大影响力和历史意义的 4 种方法; 第 3 节指出了零样本学习目前仍然面临的三大障碍以及解决思路; 第 4 节对目前的零样本学习模型进行了分类及介绍; 第 5 节首先介绍了零样本学习常用的 4 个数据集, 并分析了目前零样本学习中典型模型的实验结果; 第 6 节介绍了目前零样本学习现有的应用场景; 最后, 在第 7 节指出了零样本学习未来的可能发展方向.

1 零样本学习概述

1.1 零样本学习的形成与发展

零样本学习是在二十一世纪初逐渐发展形成的. 在深度学习发展过程中, 人们期望机器学习能够不再局限大样本、有监督的学习, 希望可以通过某些方法实现无监督、小样本数据学习, 甚至零样本学习. Bakker 等于 2003 年, Bonilla 等于 2007 年都对类的特征进行了研究[12−13], 但他们都没有考虑过在没有训练数据情况下如何学习. 此时, 迁移学习、概念学习等新兴机器学习理论的诞生为零样本学习提供了坚实的基础, 经过近十年从无到有的发展, 2009 年, Lampert 等提出了一种基于属性的类间迁移学习机制[14], 在这篇论文中, 训练集与测试集没有交集, 而且测试集中不包含训练样本, 这在本质上已经符合了零样本学习的定义. 同年, Palatucci 等正式提出了零样本学习 (Zero-shot learning)[15] 概念, 零样本学习这一方向也得以形成, 真正成为机器学习领域中重要的一部分.

深度学习技术的逐渐成熟, 也促进了零样本学习的发展, 举例来说, 目前大部分的零样本学习方法的图像特征提取, 都选择使用预训练后卷积神经网络来处理, 利用成熟的深度卷积神经网络技术, 不仅图像特征的提取工作效率得到了大幅提高, 零样本学习模型的识别准确度也得到了显著的提升. Krizhevsky 等使用了一种面向单词为基础单位的卷积神经网络[2], 对零样本学习中的文本描述进行表示学习, 既拓宽了卷积神经网络在文字处理领域的应用, 也提高了零样本学习的预测准确性[16], 另外, 一般的零样本学习方法, 选择使用递归神经网络[3] 对零样本学习中的文字描述进行特征表示学习, 也取得了较好的识别效果.

Zhang 等使用深度神经网络, 将属性特征映射到图像特征空间, 从而一定程度上规避零样本学习中的枢纽化问题, 提高了零样本学习的模型鲁棒性和预测准确性[17].

由此可以看到, 深度学习领域的研究成果, 推动

了零样本学习发展, 使得零样本学习的识别准确性与鲁棒性得到了大幅的提高.

1.2　零样本学习的定义

以下以图像识别领域作为零样本学习讨论的场景, 图像识别的本质是对测试样例类标签进行准确预测, 通过对有标签的图像训练样例进行训练, 从而得到一个 $f: x_{tr} \rightarrow y_{tr}$ 的映射关系, 一旦学习得到 f, 固定 f 中的模型参数, 利用 f 可以将在图像空间 X 的测试样例准确地映射到类空间中 Y 对应的类标签 $\widehat{y}_{te}$ 上.

零样本学习目前虽然有多种模型, 但在本质上都是相似的. 而且, 零样本学习并不是完全不需要训练样本, 零样本学习专注于研究对于特定的某些类缺失对应的训练样本, 但模型在使用其他类的训练样本训练后仍然可以对这些特定类的输入做出预测的情况. 一般来说, 零样本学习中对象类集合应分为两种: 训练类 (又称已见类) 和测试类 (又称未见类).

首先, 定义一个样例–标签对组成的训练集: $D_{tr} \equiv \{(x_i \in X_{tr} \subseteq \mathbf{R}^p; y_i \in Y_{tr} \equiv \{1, \cdots, c_{tr}\})\}_{i=1}^{m}$, 在训练类集中, 每一个样例 (典型的样例为单张图像提取出的特征表示) 都取自 P 维实空间内, 同时, 样例的类标签共有 c_{tr} 个, 即 $y_i \in Y_{tr} \equiv \{1, \cdots, c_{tr}\}$. 另外, 定义测试集为: $D_{te} \equiv \{(x_j \in X_{te} \subseteq \mathbf{R}^p; y_j \in Y_{te} \equiv \{1, \cdots, c_{te}\})\}_{j=1}^{m'}$, 这里 $x_j \in X_{te} \subseteq \mathbf{R}^p$, 与训练集不同的是, $y_j \in Y_{te} \equiv \{1, \cdots, c_{te}\}$, 测试类包括了 c_{te} 个不同的类标签, 而且训练集与训练集的类标签、样例互不相交, 即 $D_{te} \cap D_{tr} = \emptyset$, 令 $c = c_{tr} + c_{te}$, c 即为训练集和测试集的类标签个数总和.

零样本学习的基本思想是利用训练集中的样本, 和样本对应的辅助信息 (例如文本描述或者属性特征等) 对模型进行训练, 在测试阶段利用在训练过程中得到的信息, 以及模型的测试类辅助信息对模型进行补足, 使得模型能够成功对测试集中的样例进行分类.

目前来看, 零样本学习的研究仍然处于发展阶段, 还没有明确一致的定义, 因此为了更清晰地阐述零样本学习问题, 本文对零样本学习定义如下:

定义 1. 零样本学习如果模型在训练过程中, 只使用训练类的样本进行训练, 且在测试阶段可以识别从未见过的测试类样例, 那么就认为该模型实现了零样本学习.

零样本学习的示意图如图 1 所示:

训练阶段, 利用辅助信息给出的类标签到特征子空间的可逆映射 $Y_{tr} = g(S_{tr}), S_{tr} = g^{-1}(Y_{tr})$, 确定每一个类标签对应的特征表示 S_{tr}, 并利用 S_{tr} 和 X_{tr} 的对应关系, 训练样例 X_{tr} 到特征子空间 S_{tr} 的映射函数. 在训练阶段完成, 得到映射函数 $f(\cdot)$ 后, 测试阶段使用 $f(\cdot)$ 将 X_{te} 映射到同一个特征子空间中, 得到它对应的特征表示估计 $S' = f(X_{te})$, 并利用 Y_{te} 的辅助信息, 同样利用可逆映射得到 S_{te}, 对 S' 和 S_{te} 进行相似比较, 与 S' 最为相似的测试类特征表示 s_{te} 所对应的类标签 y_{te}, 即为测试类的类标签估计 $\hat{Y}_{te}$.

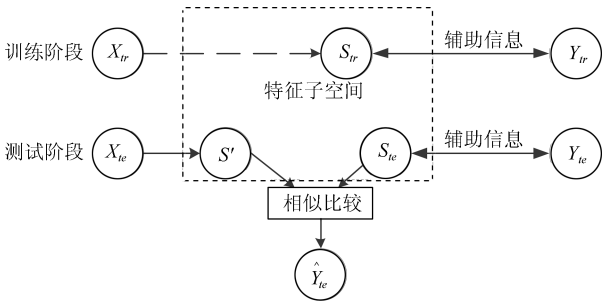

图 1　零样本学习结构示意图

Fig. 1　Zero-shot learning structure

图 1 中 X_{tr} 代表训练类输入样例集合, Y_{tr} 代表训练类的类标签集合, X_{te} 代表测试类输入样例集合, Y_{te} 代表测试类的类标签集合, 虚线方框代表着训练类与测试类共有的特征子空间, 特征子空间包含了测试类和训练类中每一类的特征编码. 右侧的双向箭头代表从类标签到特征编码的双向映射, 这一映射是已知的, 是零样本学习中需要提供的辅助信息, 目前零样本学习中经常使用的辅助信息共有三种: 属性描述、文本描述和类层次结构关系, 这一部分将在第 4.1 节中进行详细的介绍.

在训练阶段, 利用 X_{tr} 和 S_{tr} 对图像空间到特征子空间的映射进行训练, 虚线箭头代表这一映射处于训练过程. 在这一映射训练完成后, 进入测试阶段, 输入 X_{te} 时, 就可以利用已经训练好的映射模型, 将映射到特征子空间中, 从而得到 S', 与 Y_{te} 在特征子空间中的特征编码 S_{te} 进行相似性比较, 就可以确定测试类样例 X_{te} 的类标签估计值 $\hat{Y}_{te}$.

零样本学习的本质, 是通过各种各样的手段, 提高训练后的模型的泛化能力, 使得模型的泛化能力足够强大到识别从未见过的测试类样例, 从而确定测试类样例的类标签, 而要把训练后的模型推广到未见测试类样例, 需要训练样例和测试样例都具有辅助信息, 并在训练时, 学习辅助信息的表示模型, 测试时, 利用训练时学习到的辅助信息模型和测试样例的辅助信息, 预测测试样例的类标签. 给予零样本学习模型充足、有效的辅助信息, 并使得零样本学习模型可将其高效利用, 是实现零样本学习的关键.

1.3 零样本学习的相关领域研究

1.3.1 单样本学习

与零样本学习最为相似的研究领域是单样本学习 (One-shot learning)[8−10], 单样本学习希望识别模型可以实现仅使用某些对象类极为少量 (10^2 以内, 甚至只有 1 个) 的训练样本, 就可以对这些类的样例进行识别, 但是, 单样本学习模型较差的泛化能力是它所具有的致命缺陷, 如果测试阶段的样例与训练样例不是十分相似的话, 那么单样本学习模型很有可能无法准确地识别测试样例的类标签. 因此, 为了回避这一问题, 零样本学习在单样本学习的基础上得以产生.

1.3.2 偶然学习

Zhang 等于 2011 年提出了一种名为 "偶然学习" (Serendipitous learning, SL) 的方法[18], 这一方法希望能够在给定的标签空间之外, 识别未定义类标签的测试样例所属类标签, 这一方法的核心思想是在对见过的训练类的样例进行识别的同时, 也可以对从未见过的未见类样例进行聚类分析. 偶然学习使用最大间隔方法统一权衡已见类的分类损失和未见类的聚类损失, 并基于约束凹凸过程 (The Concave-convex procedure, CCCP) 和捆集方法求解相应的优化问题. 同样, Du 等提出了名为 "预定义标签空间外学习" (Learning beyond predefined label space) 的概念[19], Zhuang 等则提出了一种名为 "半定义分类的概念"[20], 这两种概念与偶然学习的概念有着很多相似之处, 这些模型都在一定程度上具有在对已见类的类标签进行分类的同时, 对未见类样例进行聚类的能力, 可以看出, 偶然学习与零样本学习关系紧密, 类似于传统分类问题与聚类问题之间的关系.

2 零样本学习典型模型

零样本学习在 2009 年正式提出之前, 就已经有多位学者对这一领域进行了一系列深入研究. 我们将在这一节, 具体介绍零样本学习发展过程中, 具有重要节点意义的 4 种模型, 分别是: 新任务的零数据学习[21]、语义输出编码零样本学习[15]、基于属性类间迁移的未见类学习[14]、以及跨模态迁移的零样本学习[22].

2.1 新任务的零数据学习

2008 年, Larochelle 等提出了零数据学习这一概念[21], 零数据学习的目标在于如何构造模型学习没有可用的训练数据、且只有类描述的分类器学习问题. 因此在定义上, 零数据学习与零样本学习本质上是相同的. Larochelle 等[21] 提出了两种零数据学习方法: 输入空间方法和模型空间方法. 这两种方法为之后的零样本学习的发展指出了具有启发意义的方向.

2.1.1 输入空间方法

输入空间方法的核心思想是利用训练类和测试类中对应于每一类的描述信息, 对模型进行信息补充, 使得模型不再是简单的学习由输入样例到类标签的映射关系, 而是把输入样例与对应类描述的信息对作为新的输入, 并学习其到类标签的映射关系. 在训练阶段, 输入空间方法利用训练类的样本以及对应的类描述信息对模型进行训练, 因为是直接在输入阶段, 对模型所需要的信息进行了补充, 所以该方法被命名为输入空间方法.

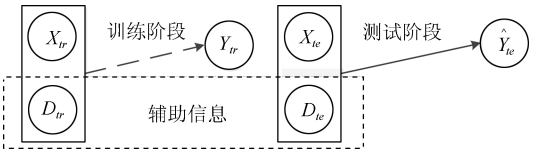

图 2 输入空间方法示意图

Fig. 2 Input space method

如图 2 所示, 输入空间方法将模型输入进行了增补, 在输入训练样例 X_{tr} 的同时输入样例所属的类的描述信息 (即辅助信息) D_{tr}, 以增广输入 $[X_{tr}, D_{tr}]$ 作为模型的输入进行训练, 得到预测模型 $f^*(\cdot)$.

对测试类样例 X_{te} 进行分类时, 此时存在多个不同的候选测试类标签时, 遍历 X_{te} 与全部可能的 D_{te} 的配对组合, 作为已训练预测模型 $f^*(\cdot)$ 的输入, 对所有输出结果进行比较, 最为可信的第 j 个测试类的描述 D_{te}^j 所对应的类标签即为输入样例的类标签估计值 $\hat{Y}_{te}$.

2.1.2 模型空间方法

模型空间方法假设对于每一类 y, 都存在对应的分类函数 $d(y) \in D$. 假定 $d(y)$ 为该类所对应的描述信息, 用把分类函数 $f_y(\cdot)$ 参数化为 $g_{d(y)}(x)$, 即令:

$$f_y(x) = g_{d(y)}(x) \tag{1}$$

零数据学习问题变为用 $d(y)$ 和输入样例学习 $g_{d(y)}(x)$ 的表示问题, 例如, 可以使用以下平均损失最小化来求解模型参数:

$$\frac{1}{|X_{tr}|}\sum_{x_i,y_i} L(y_i, g_{d(y)_{tr}^i}(x_i)) \tag{2}$$

L 为常用的损失函数, x_i、y_i 为对应同一训练类上的样例−标签对.

训练阶段, 如果可以利用给定的 x_i、y_i 以及 $d(y)^i$ 对函数 g 进行充分的训练, 在测试阶段, 假定

每一个测试类都具有给定的 $d(y)$ 时, 就可以利用训练阶段学习到的函数 g, 计算每个测试类样例所对应的分类函数 $g_{d(y)}(x)$, 这样, 就可以利用这一系列测试类函数族, 来判断测试类输入样例的类标签.

例如, 假设存在线性分类器 $y = W(d(y))x$, 在训练阶段, 利用每一个训练类中的样例–标签对 x_i、y_i 以及 $d(y)^i$ 对 $W(d(y))$ 进行训练, 在测试阶段, 用训练阶段学习到的模型 $W(d(y))$, 分别把全部测试类对应的辅助信息 $d(y)^j$ 和测试类样例作为模型 $W(d(y))$ 的输入, 就可以计算得到从未见过的测试类样例的各个线性分类器 $y = W(d(y)^j)x$, 从而确定测试类样例最可能属于的类.

与输入空间方法相比, 模型空间方法具有更强的泛化能力, 输入空间方法相当于令 $g_{d(y)}(x) = g(x, d(y))$, 可看作是模型空间方法的特例. 模型空间方法将 $d(y)$ 作为调整模型 $g_{d(y)}(x)$ 的参数, 所以该方法被命名为模型空间方法, 如图 3 所示.

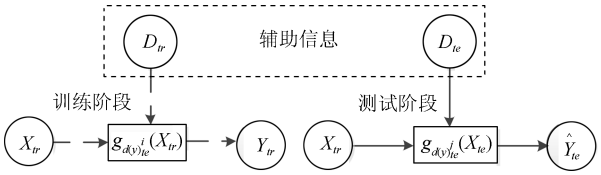

图 3　模型空间方法示意图

Fig. 3　Model space method

第 4.2.1 节中将要介绍的属性描述替换零样本学习[23] 和文本描述零样本学习[24] 都属于典型的模型空间方法.

2.2　语义输出编码零样本学习

2009 年, Palatucci 等首次提出 “零样本学习” 这一概念[15], 并提出利用语义输出编码 (Semantic output code, SOC) 分类器以及包含大量语义知识的标签库, 将训练模型的分类能力推广到测试类上, 实现零样本学习.

首先定义一个维的语义特征空间, 每个维度都对应一个二值编码特征. 例如, 对于 “狗” 这一类标签来说, 在 “多毛”、“有尾巴”、“水下呼吸”、“肉食性” 和 “可以快速移动” 5 条语义特征下的语义空间中, “狗” 这一类可以利用布尔值表示为语义特征向量 [1, 1, 0, 1, 1], 这一表示方法与 Dietterich 等提出的方法在思路上较为相似[25]. 定义语义知识库 $K = \{f : y\}_1^M$ 为具有 M 个编码–标签对的集合, $f \in F^d$ 表示每一类所对应的语义属性特征向量, $y \in Y$ 为类标签, 实验中所有使用的类标签 Y 在知识库 K 中都具有一一对应的编码–标签对. 例如动物知识库中包含许多动物类的语义编码以及标签, 实验中所使用的类标签所对应的语义编码都可以在该知识库中找到.

定义语义输出编码分类器为 $H : X \to Y$, 假定 H 是复合映射函数: $H = L(S(\cdot))$, S 函数负责将图像映射到语义空间: $S : X \to F^d$, L 函数负责将语义编码空间映射到标签集上: $L : F^d \to Y$, L 函数通过查询知识库 K 中语义编码与类标签的对应关系实现映射. 语义输出编码零样本学习过程示意图如图 4 所示:

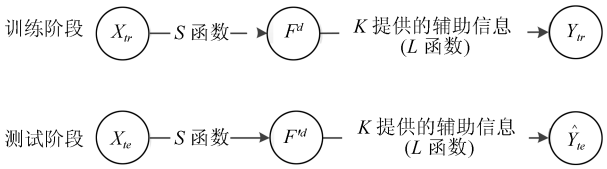

图 4　语义输出编码零样本学习过程示意图

Fig. 4　Semantic output code zero-shot learning process

假定训练类集合为 $D_{tr} = \{x_i \in \mathbf{R}^p, y_i \in Y\}_{i=1}^N$, 通常情况下, M 应该远大于 N, 意味着语义空间中的元素可用于描述更多类标签. 这是因为当知识库 K 与输入数据 D 相比具有更多 Y 的可能值时, 基于语义输出编码构造分类器才是可行的.

训练阶段, 首先需要根据训练类在知识库 K 中对应的语义编码 f, 建立 N 个样例–语义编码对的集合 $\{x, f\}_1^N$, 并使用 $\{x, f\}_1^N$ 对语义编码函数 S 进行训练, 文中的函数 S 为线性函数, $\hat{f} = x \cdot \hat{W}$, 训练后可以得到语义特征表示模型矩阵 $\hat{W}$.

在测试阶段, 输入测试样例时, 分类器将使用训练阶段学习到的 S 函数计算输入的测试样例的语义特征编码. 将测试样例输入 S 函数, 得到对应的语义特征编码估计, 并与各个类标签对应的语义特征编码进行比较. 即使输入的样例属于训练过程模型没有见过的测试类, 如果由 S 函数生成的语义特征编码估计接近于真实类标签的语义编码, 那么 L 函数也将大概率识别出输入的正确标签. 通过使用知识库中每一类对应的语义特征编码, 分类器就能够推断和识别出从未见过的新类. 语义输出编码模型的参数计算公式如 (3) 所示:

$$\begin{aligned} \hat{W} &= (X^{\mathrm{T}}X + \lambda I)^{-1}X^{\mathrm{T}}K_Y \\ \hat{f} &= x \cdot \hat{W} \end{aligned} \tag{3}$$

其中, X 为训练类的输入样例, K_Y 是对应类的知识库中的语义编码集合, I 是单位矩阵, λ 是通过使用 Hastie 等提出的交叉验证评分函数自动选择的正则化参数[26].

2.3　基于属性类间迁移的未见类学习

上一模型中所提到的语义编码, 本质上是与类相对应的属性特征编码. 根据人类认知科学的研究, 人类在遇到从未见过的对象时, 也会首先判断对象的特征属性, 从而与自身的先验知识相结合, 做出其

所属类别的判断[27−28]. 因为受到了这一研究的启发, 在机器学习领域中, 现在也常会考虑使用属性概念辅助解决分类问题. 但在 2009 年之前, 大多数学者所关注的属性仍然较为简单, 仅仅考虑了图像中的几何形状以及颜色等属性[29−31]. Farhadi 等于 2009 也提出了通过语义属性预测对象类标签的模型[32], 实现了对类附加特定的属性描述, 并将属性泛化到不同的类中. 同年, Lampert 等提出了基于属性类间迁移的未见类检测方法[14], 这一模型可谓是零样本学习的奠基之作, 现有的零样本学习方法大多都继承了该模型的思想, Lampert 等建立的数据集 “Animals with attributes” 也成为研究零样本学习所必须使用的数据集之一.

零样本学习最开始所使用的辅助信息, 就是 “属性”. Lampert 等和 Hinton 等不约而同地在 2009 年选择使用属性辅助信息实现零样本学习, 这在某种意义上体现了零样本学习的重要性和出现的必然性. 在 2014 年 3 月, 基于已经提出的属性类间迁移模型, 他们又发表了基于属性的零样本视觉对象分类的论文[33], 两篇论文中的模型基本一致, 在此进行介绍.

文中共提出了两种拓扑结构零样本学习模型, 直接属性预测 (Direct attribute prediction, DAP) 模型以及间接属性预测 (Indirect attribute prediction, IAP) 模型, 它们的具体结构分别如图 5 和图 6 所示:

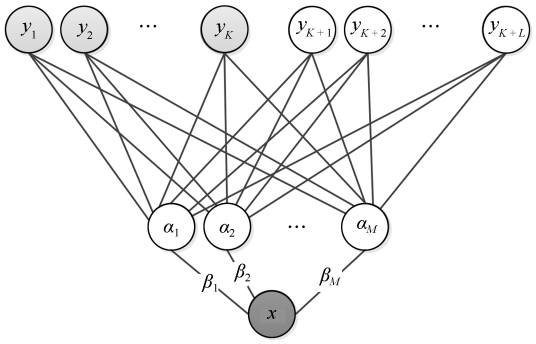

图 5 直接属性预测模型结构示意图

Fig. 5 Direct attribute prediction model

假定 $y_1, \cdots, y_K$ 为训练类, $y_{K+1}, \cdots, y_{K+L}$ 为测试类, $a_1, \cdots, a_M$ 为全部属性特征的集合, 从图 5 中可以看到, 每一类都具有特定的属性指示向量: $y = (a_1^y, \cdots, a_M^y)$, 类的属性指示值可以是布尔值, 也可以是实值. 在训练阶段, 训练类标签 $(y_k)_{k=1,\cdots,K}$ 通过属性指示向量 $(a_m)_{m=1,\cdots,M}$ 对属性分类器 β 进行训练. 在测试阶段时, 尽管模型从未针对测试类进行训练, 但属性层仍然可以利用训练阶段得到的属性分类器 β 对输入的测试类图像所具有的属性进行判断, 得到输入图像的属性特征估计之后, 就可以利用得到测试样例的属性特征估计以及测试类的属性指示对输入样例进行分类判断.

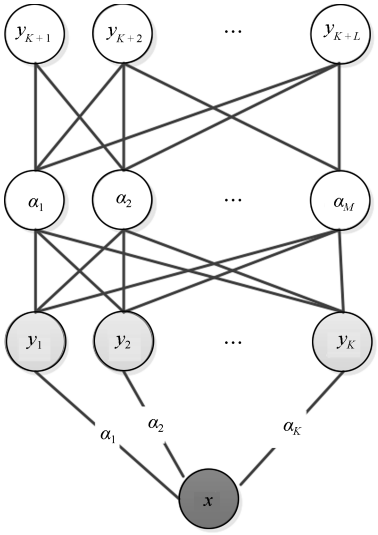

图 6 间接属性预测模型结构示意图

Fig. 6 Indirect attribute prediction model

其分类公式如 (4) 所示:

$$f(x) = \arg\max_{l=1,\cdots,L} p(y|x) = \arg\max_{l=1,\cdots,L} \prod_{m=1}^{M} \frac{p(a_m^l|x)}{P(a_m^l)} \tag{4}$$

这里, M 为属性的总个数, a_m^l 为第 l 类的属性的第 m 个分量, $p(a_m^l|x)$ 是通过分类器判断得到的输入图像具有这一特定属性的概率, $P(a_m^l)$ 是通过计算训练类属性的经验均值得到的先验估计.

对于 IAP 模型, 可以从图 6 中看到, 此时分类器并不能直接得到输入图像的属性估计, 而是利用预测输入图像的类标签以及对应的属性指示向量, 间接得到输入图像的属性特征估计, 公式如式 (5):

$$p(a_m|x) = \sum_{k=1}^{K} p(a_m|y_k)p(y_k|x) \tag{5}$$

这里, K 为训练类的总数, $p(a_m|y_k)$ 是预先定义的训练类属性特征, $p(y_k|x)$ 是通过分类器得到的输入图像属于训练类 k 的概率估计, 在得到 $p(a_m|x)$ 后, 将式 (5) 代入式 (4) 中, 就可以得到 IAP 分类模型的分类公式, 实现 IAP 分类模型下的零样本分类.

属性作为有效的机器学习概念, 在机器学习的多个领域中都具有很大的发展空间, Suzuki 等于 2014 年对于直接属性预测进行了改进, 他们根据属性出现的频率对每一个属性设置了权重, 使得识别效果得到了改善[34]. 也有学者利用属性作为反馈, 在训练阶段以属性作为载体, 向模型返回更多信息, 提高了模型的训练效率[35−36], Kulkarni 等提出了一

种能够利用属性信息理解图像并生成自然语言图像描述的模型[37]. 现如今, 属性识别在计算机视觉领域中的具体应用也已经十分广泛, 例如人脸识别[38]、动作识别[39]、场景识别[40] 和车辆监控[41] 等.

2.4 跨模态迁移的零样本学习

跨模态迁移的零样本学习[22] 由 Socher 等于 2013 年提出, 该方法的意义在于首次将零样本学习问题转化为子空间问题. 从 2013 年开始, 有很多学者延续了将零样本学习问题转化为子空间问题这一思路, 基本上他们都是受到了这篇论文的影响.

跨模态迁移学习的核心思想是将图像和类标签同时映射到相同的子空间内, 目前通常是将图像和类标签同时映射 (或称为嵌入 (Embedding)) 到语义空间中, 并在语义空间内, 利用一定的相似性度量方法, 去确定测试类输入图像的类标签. 跨模态迁移零样本学习示意图如图 7 所示.

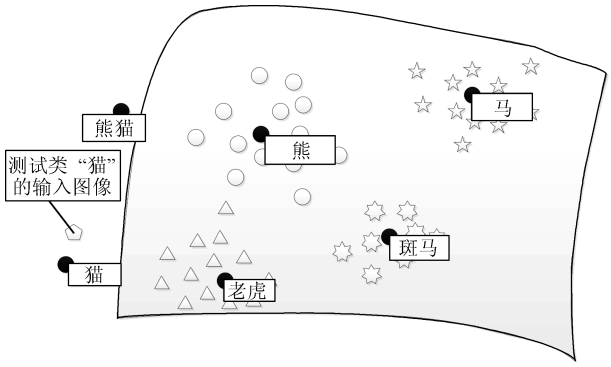

图 7　跨模态迁移零样本学习示意图

Fig. 7　Cross-modal zero-shot learning

图 7 中, 五边形白点代表测试类 “猫” 的输入图像在语义空间中的表示, 除此之外的白点代表训练类的输入图像在语义空间中的映射向量. 在该方法中, 黑点代表所有类标签的语义向量, 所有的语义向量都是预先给定的辅助信息, 这些语义向量可以是属性向量, 也可以是使用文本描述生成的单词语义向量.

训练阶段, 首先利用训练类标签对应的语义向量和训练类的样例对跨模态映射模型进行训练. 训练后, 映射模型可以成功地将输入样例映射到语义空间中, 并准确地映射到输入样例所对应的类标签的语义向量附近.

在测试阶段输入测试类的样例时, 首先利用训练阶段得到的跨模态映射模型将测试类输入样例映射到语义空间中, 并根据相似性判断方法 (如余弦相似性、K 近邻方法等), 将与输入样例的的语义向量估计最为相似的语义向量的类标签, 作为测试类图像的类标签估计.

这一方法使用双层神经网络作为跨模态映射模型:

$$J(W)=\sum_{y\in Y_0}\sum_{x_i\in X_y}\left\|s(y)-W_2f(W_1x_i)\right\|^2 \tag{6}$$

这里, x_i 为输入图像, $s(y)$ 为类标签在语义空间内的映射向量. W_1 和 W_2 分别为每一层神经网络的权重矩阵, $f=\tanh$ 为非线性双曲正切函数, 通过最小化 $J(W)$ 求解跨模态映射模型参数 W_1 和 W_2.

当然, 将类标签的属性特征映射到图像特征空间中也是可以的, Zhang 等提出了一种反向映射 – 将图像和类标签同时映射到图像特征空间中的方法[17], 这一方法将在第 4.2 节部分中进行介绍.

2.5 分析与比较

这一部分中所说明的这 4 种模型在零样本学习的形成阶段奠定了零样本学习的基础, 但由于当时的局限性, 这 4 种模型仍然有不可忽视的缺点.

新任务的零数据学习在当时只给出了 “模型空间方法” 和 “输入空间方法” 的概念, 两种方法的具体细节并没有给出, 这两种方法的意义在于为零样本学习指出了概念上的实现方向.

语义输出编码模型虽然首次提出了零样本学习这一概念, 但这一方法与后来的零样本学习方法相比过于简陋, 只是使用训练后得到的矩阵将图像特征线性映射到语义空间中, 实验效果较差.

基于属性的类间迁移模型首次正式地将 “属性” 这一概念引入零样本学习, 并希望以此实现较为准确的测试类识别, 但由于该方法使用属性分类器对测试类进行识别, 导这致其对于属性的识别准确度较高, 对于测试类的识别准确度却较低.

跨模态迁移模型首次将零样本学习问题考虑为特征子空间的映射问题, 这一思路真正意义上开创了零样本学习的发展空间, 但这一方法也带来了零样本学习中最为关键的问题之一, 映射域偏移问题[42], 在训练类上训练的映射模型, 在映射测试类样例时产生的偏差, 会显著影响零样本学习模型的识别准确度, 这一问题我们将在下一节进行讨论.

3 零样本学习关键问题

虽然目前来说零样本学习仍处于快速发展的阶段, 前景十分可期, 但零样本学习自身方法中存在的问题使得前进道路上横亘着三个无法忽视的障碍. 这三个障碍分别是广义零样本学习 (Generalized zero-shot learning)[43]、枢纽度问题 (Hubness)[44] 以及映射域偏移问题 (The projection domain shift problem)[42], 本节将对这三个问题以及对应的解决思路进行介绍.

3.1 广义零样本学习

在实际应用中, 目前的零样本学习与现实应用的学习环境出现了一定程度的矛盾, 这是因为零样本学习中假设 (为了避免误会, 在第 3.1 节中改称为已见类和未见类), "零样本学习在测试阶段, 只有未见类样例出现. " 这在实际应用中是不现实的, 已见类的对象往往是现实世界中最为常见的对象, 而且, 如果在训练阶段已见类样本容易得到、未见类样本难以获取, 那么在测试阶段就也不应只有未见类样例出现.

所以, 为了让零样本学习真实反应实际应用中的对象识别场景, 零样本学习模型应对所有输入样例进行识别, 即少量的未见类样例夹杂在大量的已见类样例中, 输入样例的可能类标签大概率属于已见类, 但也有可能属于未见类. 因此, 定义广义零样本学习为:

定义 2. 广义零样本学习. 如果模型在训练过程中, 只可以使用训练类的样本进行训练, 并且在测试阶段可以准确识别已见类样例以及从未见过的未见类样例, 那么就认为该模型实现了广义零样本学习.

在零样本学习定义确定之前, 也有学者对这一方面进行过尝试性的研究, 跨模态迁移模型在对未见类和已见类进行识别之前, 加入了已见类和未见类的判别器, 在测试阶段, 首先对输入样例进行新颖检测, 判断它属于已见类还是未见类, 然后使用不同的已见类和未见类分类器对输入样例的类标签进行判断.

语义凸组合模型[45] 和综合零样本学习模型[46] 两种模型思路相似, 利用未见类和已见类辅助信息中的联系, 将未见类标签表示为已见类标签的线性组合, 从而使得模型可以同时识别已见类以及未见类, 但在实验结果中, 这两种方法对于未见类的识别准确度仍然不是十分理想.

虽然广义零样本学习定义十分简单, 但传统零样本学习方法仍会出现上述未见类识别准确率较低的问题, 这是因为模型在训练过程中只使用了已见类样本进行训练, 已见类的先验知识也更为丰富, 从而使得已见类模型占主导地位. 所以在输入测试样例时, 模型会更加倾向于对未见类样例标注为已见类的标签, 从而造成识别准确率与传统零样本学习相比的大幅度下跌, 实际实验结果也证明了这一点[43]. 对于这一问题, 一般有两种解决方法, 一种是在输入样例时, 判断其属于已见类或未见类, 这种方法称为新颖检测, 另一种方法是在模型判断输入样例为已见类时, 叠加一个校准因子以平衡模型对已见类识别的倾向性, 这一种方法称为叠加校准, 两种方法具体内容如下:

1) 新颖检测

首先, 定义得分函数 $N(x)$, 对于每一个输入样例 x, 都可以输入得分函数, 得到其分值, 利用判断公式如式 (7):

$$\hat{y}=\begin{cases}\arg\max\limits_{c\in S} f_c(x), & N(x)\leq -\gamma\\ \arg\max\limits_{c\in U} f_c(x), & N(x)>-\gamma\end{cases} \tag{7}$$

这里 S,U 分别代表已见类和未见类, γ 为人工给定的超参数, 用于判断输入样例属于测试类或训练类, 当得分值 $N(x)$ 小于等于 $-\gamma$ 时, 我们认为该输入样例属于已见类, 反之, 我们认为新输入的样例, 属于测试类.

2) 叠加校准

$$\hat{y}=\arg\max_{c\in T}[f_c(x)-\gamma \boldsymbol{I}(c\in S)] \tag{8}$$

$\boldsymbol{I}[\cdot]\in\{0,1\}$ 为指示函数、当 c 为已见类时, $\boldsymbol{I}[\cdot]=1$. 其他情况下 $\boldsymbol{I}[\cdot]=0$. 表示测试类, 测试类包含了所有已见类和未见类, γ 为校准因子, 该方法的意义在于模型 f 判断输入样例属于已见类时, 通过减去 γ 来平衡模型倾向于判断输入样例为已见类的倾向性.

我们考虑两种极端情况, $\gamma\rightarrow+\infty$ 时, 分类规则将会忽视所有已见类, 并将所有输入归为未见类, 该方法就转化为传统零样本学习. 相反, $\gamma\rightarrow-\infty$ 时, 模型只考虑所有已见类标签, 该模型转化为传统多分类模型.

3.2 枢纽化问题

枢纽化问题 (Hubness) 并不是零样本学习所特有的问题, 所有利用特征子空间的机器学习模型, 都在实验中发现了这一现象, 维度越高, 这一现象愈发明显, 在嵌入空间中, 这一问题尤为严重. 而且由于目前零样本学习最为流行的方法就是将输入样例嵌入到特征子空间中, 这也导致了零样本学习中的枢纽化问题尤为突出.

枢纽化问题最早是在 2010 年由 Radovanovi 等[47−48] 提出, 这一问题是指将原始空间 (例如图像特征空间或类标签空间) 中的某个元素映射到特征子空间中, 得到原始空间中某个元素的在特征子空间中的新表示, 这时如果使用 K 近邻方法进行相似性度量时, 可能会有某些原始空间中无关元素映射到多个测试样本特征空间中表示最近的几个近邻中, 而这些无关向量, 就称为 "枢纽 (hub)", 枢纽的出现, 污染了测试类的近邻列表, 对零样本学习产生了较大的影响. 枢纽化现象的示意图如图 8 所示.

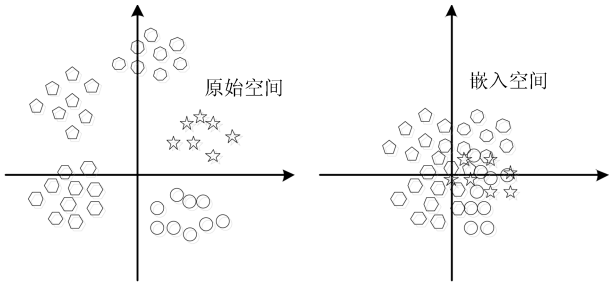

图 8　枢纽化问题示意图

Fig. 8　Hubness

从图 8 中我们可以看到, 在嵌入空间中, 各个输入样例的嵌入向量全部向子空间的原点聚集. 不同类的输入样例之间的距离在有些时候甚至要小于同类输入样例之间的距离, 因此零样本学习的识别效果不可避免地会受到这一问题的负面影响.

Lazaridou 等推断产生这一现象是由于用正则化最小二乘目标函数求得映射模型, 使得求解的映射模型计算得到的映射向量的方差远远小于原始输入向量的方差, 训练类中的映射向量也更接近于空间原点以及其他向量, 使得其测试类输入映射的近邻向量中的列表中出现这些枢纽向量的概率也大为增加[44]. 因此, 不选择使用岭回归这种带有正则化项的最小二乘目标函数, 例如将岭回归函数替换为铰链损失函数, 可以在一定程度上缓解枢纽化问题.

Dinu 等则提出一种全局矫正方法 (Globally corrected) 来应对枢纽化问题, 在该方法中, 与多个标签相似的向量, 将会受到惩罚, 并使用整个映射数据集的近邻统计量, 对目标元素进行排名[49]:

$$GC_1(x,T) = \arg\min_{y\in T} Rank_{\,x,T}(y) \tag{9}$$

简单来说, 该方法与过去以往的分类方法相反, 这一方法不选择输入样例在语义空间映射向量的最近邻点作为其类标签的估计, 而是去查找类标签对应的语义向量附近的输入样例的映射向量, 这一方法只有在类标签的语义向量附近没有比较准确地输入样例映射向量时, 才会出现枢纽化现象.

Zhang 等反其道行之, 选择将语义向量映射到特征空间中[17], 这一方法基本规避了枢纽化问题, 这是因为输入样例映射到语义空间中的枢纽化问题与类标签映射到图像特征空间中的枢纽化相比, 由于输入样例图像的个数远远多于类标签的个数, 前者对于识别准确性的影响更加严重.

3.3　映射域偏移问题

映射域偏移问题的根源在于映射模型较差的泛化能力: 模型使用了训练类样本学习由样例特征空间到类标签语义空间的映射, 由于没有测试类样本可以用于训练, 因此, 在映射测试类的输入样例时, 就会产生一定的偏差, 示意图如图 9 所示:

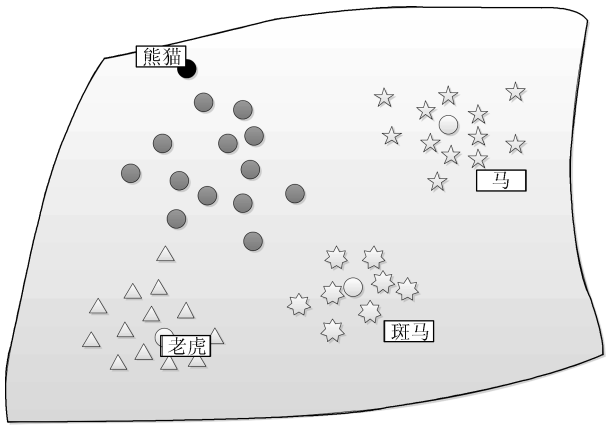

图 9　映射域偏移问题示意图

Fig. 9　The projection domain shift problem

图 9 中白色圆点为训练类标签的语义向量, 黑色圆点为测试类标签的语义向量, 灰色圆点为测试类输入在语义空间中的映射向量, 其余为训练类在语义空间的映射向量, 可以看出, 老虎、马和斑马为训练类, 熊猫为测试类, 训练类的样本都很好地映射到了训练类语义向量的附近, 而测试类输入样例映射到语义空间后, 却与它所对应的类标签距离较远. 因此, 如果想要解决映射域偏移问题, 就需要提高映射模型的泛化能力.

Fu 等提出了一种转导性多视角嵌入零样本学习框架[42], 利用多个视图: 属性视图、单词向量视图和图像特征视图, 利用正规相关分析, 让多个视图进行互相约束, 使得映射偏移误差尽可能达到最小.

Kodirov 等将投影偏移问题转换为域自适应学习问题[50], 并提出了一种全新的稀疏编码框架, 该框架将训练类视为源域, 测试类视为目标域, 利用线性映射将语义向量映射到特征空间, 以及引入一系列自适应正则化约束, 通过优化如式 (10) 的目标函数, 使得测试类的语义嵌入向量能够映射出较为准确的视觉特征向量, 从而进一步实现分类:

$$\begin{aligned}\{D_t, Y_t\} = \arg\min_{D_t,Y_t} \Big(& \|X_t - D_t S(Y_t)\|_F^2 + \\ & \lambda_1 \|D_t - D_s\|_F^2 + \\ & \lambda_2 \sum_{i,j} w_{ij} \left\|s(y_i) - p_j^t\right\|_2^2 + \lambda_3 \|Y_t\|_1 \Big) \\ \text{s.t. } & \|d_i\|_2^2 \le 1 \end{aligned} \tag{10}$$

这里 D_t 为测试类的映射矩阵, 负责将测试类的语义向量映射至图像特征空间中, D_t 为训练类的映射矩阵, 第一个正则化项用于限制 D_t 和 D_s 之间的距离. $i, j \in \{1, \cdots, c_{te}\}$, c_{te} 是测试类的个数, w_{ij}

是判断输入样例 i 属于类 j 的概率, p_j 是第 j 个类真实标签的语义向量原型, 对于测试类, w_{ij} 是无法获得的, 因此需要利用 IAP 模型[14], 对 w_{ij} 进行估计. 同过引入这样的正则化函数进行约束之后, 测试类的映射模型得以使用训练类中的信息进行校正, 从而缓解了映射域偏移问题.

4 零样本学习方法研究进展

本节将具体介绍多个代表性的零样本学习方法, 并把目前的零样本学习分为两大类, 阐明不同方法之间的关系. 目前的零样本学习方法一般都沿袭了将输入和输出映射到子空间的思路, 在这一思路下的零样本学习方法可以分为两类: "相容性模型" 和 "混合模型". 以 DAP 模型为例, DAP 模型是一种典型的相容性模型. 首先将类标签转化为属性特征表示的向量、并对输入图像包含的属性进行识别得到输入样例的属性特征表示向量, 进而将输入样例的属性特征与各个类标签的属性特征表示进行比较, 是一种将输入输出映射到属性子空间中比较两者相容性的方法. 相容性模型的示意图如图 10 所示:

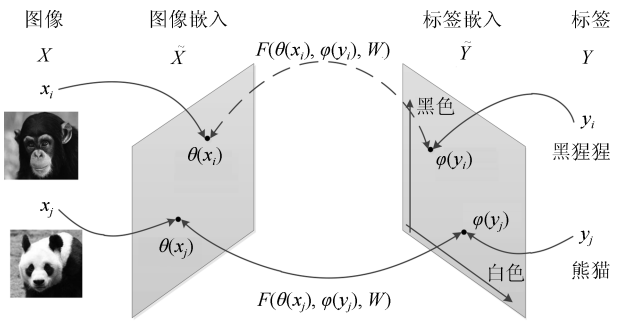

图 10 相容性模型示意图

Fig. 10 Compatibility model

图 10 中 $\theta(\cdot)$ 为图像嵌入函数, $\varphi(\cdot)$ 为类标签嵌入函数, $\varphi(\cdot)$ 一般提前给定, 代表零样本学习中的辅助信息, $\theta(\cdot)$ 为常用的图像特征提取方法, 如卷积神经网络等. 图 10 中黑猩猩为训练类, 大熊猫为测试类. 在训练阶段, 利用嵌入模型分别将图像和类标签嵌入到子空间中, 并对相容性判断函数 $F(\theta(x), \varphi(y), W)$ 进行训练. 在训练完成后, 输入测试样例时, 使用相同的嵌入模型, 将类标签和图像嵌入到子空间中, 即可利用已经训练完成的相容性判断函数来判断测试类输入的类标签.

零样本学习中的相容性模型定义如下:

定义 3. 分别将输入输出映射 (或称为嵌入) 到子空间, 并在子空间内判断输入输出映射向量的相容性, 并确定类标签的模型, 称为相容性模型.

有一点值得注意, 相容性模型不一定将输入输出嵌入到相同的空间中, 例如类标签的嵌入空间维度, 可能会远远低于图像特征嵌入空间的维度.

根据相容性函数是线性函数或是非线性函数, 相容性判别模型又可以分为两类: 线性相容性模型和非线性相容性模型. 与相容模型相对的是混合模型, 典型的混合模型为 IAP 模型. 混合模型的特点是使用训练类的类标签, 来估计测试类输入样例的类标签. 这一目标是通过测试类与训练类之间的关联关系, 也就是辅助信息实现的.

以 IAP 模型为例, 在测试阶段输入训练类的测试样例时, 各个训练类分类器通过判断输入与各个训练类的相似性, 进而利用各个训练类的属性表示, 得到输入样例的属性表示的估计, 并将输入样例的属性估计向量与测试类标签所对应的属性指示向量进行比较, 最为相似属性向量对应的类标签, 即为输入样例的测试类标签估计. 混合模型定义如下:

定义 4. 利用训练类类标签所对应的特征子空间映射的混合组合来表示测试类输入样例在特征子空间中的映射, 进而判断输入样例的映射与测试类类标签映射之间的相似性, 得到输入样例类标签估计的模型, 称为混合模型.

混合模型示意图如图 11 所示, 混合模型首先在训练阶段利用训练集中的样本对典型分类器进行训练, 使得该分类器可以准确识别训练类的样例, 并且在测试阶段, 首先利用 X_{tr} 分类器判断输入样例与各个训练类之间的关系 $P(X_{tr}|X_{te})$, 进而利用各个训练类的子空间特征和输入样例与训练类之间的关系, 使用训练类子空间特征的组合来估计输入样例的子空间特征, 从而完成混合模型零样本学习下的识别.

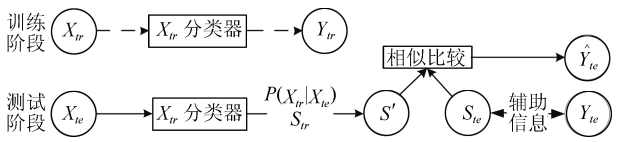

图 11 混合模型示意图

Fig. 11 Hybrid model

相容性模型与混合模型的区别, 在于所侧重的信息不同, 相容性模型原理简单, 易于实现, 侧重于使用高质量的辅助信息对类标签嵌入函数进行训练, 使得嵌入函数能够准确地将类标签映射到特征空间中.

相容性模型又分为线性相容性模型和非线性相容性模型. 非线性相容性模型使用非线性相容函数, 与线性相容性模型相比具有着更强的表达能力, 因此, 目前的零样本学习研究中, 非线性相容性模型在相容性模型中占据主流地位.

而混合模型则侧重于利用与测试类具有较高相似度的训练类, 这一模型具有很强的泛化能力, 可以

同时对测试类和训练类样例进行识别, 适用于广义零样本学习, 但是这一方法对于测试类训练类之间的相似性依赖程度较高, 鲁棒性较差, 如果在训练中使用的训练类样例与测试类样例相似性较低, 那么测试类样例的识别效果将会受到较大的影响.

零样本学习中各个子类的分类图如图 12～14 所示, 图中从模型引出到下一模型的箭头, 代表了下一模型是在上一模型的基础上发展演变得来.

在介绍这两类模型之前, 首先介绍一下零样本学习目前训练类和测试类中常用的辅助信息, 以及获得辅助信息的方法.

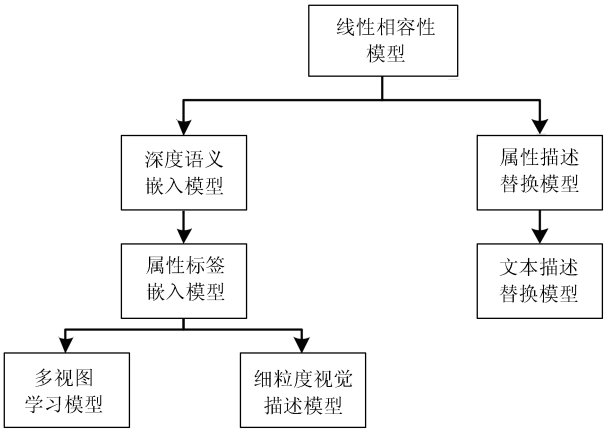

图 12　线性相容性模型分类示意图

Fig. 12　Linear compatibility model classification

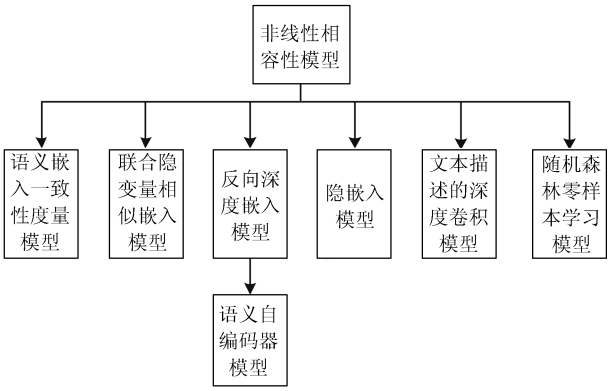

图 13　非线性相容性模型分类示意图

Fig. 13　Nonlinear compatibility model classification

4.1　辅助信息介绍

辅助信息用于描述训练类和测试类, 并将训练类和测试类关联起来, 使得训练类和测试类之间可以共享某些特征, 是成功实现零样本学习的关键, 目前来说, 有三种形式的辅助信息成功应用于零样本学习中, 分别为: 手工标注的属性、文本学习模型、以及层次结构.

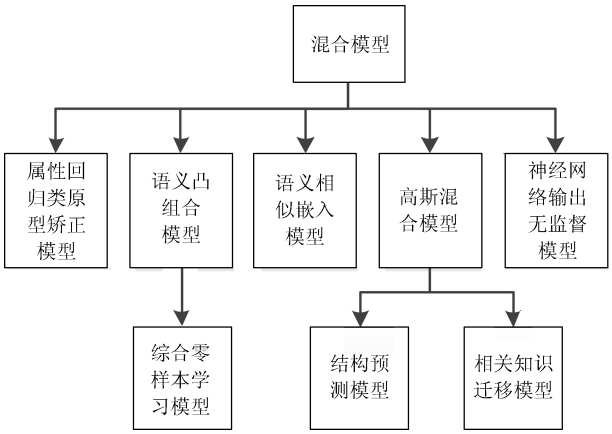

图 14　混合模型分类示意图

Fig. 14　Hybrid model classification

4.1.1　手工标注的属性

属性特征是非常有效的辅助信息, 可以直接应用于零样本学习中, 其属性特征向量表示形式为 $a^l = \{a_1^l, a_2^l, \cdots, a_M^l\}$.

$l \in \{1, 2, \cdots, L\}$ 代表类标签索引值, M 为第 l 个类标签中所有可能出现的属性个数. 在类标签个数逐渐增大时, 或是需要对分类对象进行细粒度分类时, 类标签个数 L 或属性个数 M 也会增大, 从而很大程度上增加了人力损耗. 因此为了规避这一问题, 可以使用文本模型学习来代替手工标注的属性.

4.1.2　文本学习

文本学习方法希望能够利用网络中现有的文本资料 (例如每一个类标签对应的维基百科文章), 以及现有的自然语言处理模型, 得到每个类的特征表示, 一般来说, 得到的特征表示为类标签单词在语义空间中的特征向量. 目前, 应用在零样本学习中的自然语言处理工具共有三种: Bag-of-Words[51]、Word2Vec[52] 和 GloVe[53].

这一方法的优点在于网络中的文本资料是易于获得的, 如果能够实现通过文本描述的语义空间中的特征向量表示学习得到类标签的准确辅助信息, 零样本学习的建模成本会显著下降, 增加了零样本学习的可实现性.

此外, 也可以将文本学习得到的类特征与人工标注的属性特征组合后形成新的辅助信息, 从而使得零样本学习模型获得更好的性能.

4.1.3　类层次结构关系

WordNet 为包含 100 000 个英语单词的数据库, 可以提供大量的类层次结构关系, 可以使用多个相似性度量函数对各个类标签的相似度进行度量[54], 但目前来看, 层次结构信息在零样本学习中应用仍然较少.

4.2 相容性判断模型

4.2.1 线性相容性模型

线性相容性模型与非线性相容性模型的主要区别在于 $F(\theta(x),\varphi(y),w)$ 是否为线性函数. 在线性相容性模型中, 相容性函数定义为:

$$F(\theta(x),\varphi(y),w)=\theta^{\mathrm{T}}(x)W\varphi(y) \tag{11}$$

在本节中, 我们将要介绍 6 种线性相容性零样本学习模型: 分别是深度语义嵌入模型[55]、属性标签嵌入模型[56]、属性描述替换模型[23]、文本描述替换模型[24]、多视图零样本学习模型[57] 和细粒度数据视觉描述的深度表示模型[58].

1) 深度语义嵌入模型

深度语义嵌入模型是第一种使用线性相容性函数的模型[55]. 在训练阶段, 深度语义嵌入模型使用 (12) 的目标函数对模型进行训练:

$$\sum_{y\in Y_{tr}}\max\left[0,(\varepsilon-\theta^{\mathrm{T}}(x)W\varphi(y_n)+\theta^{\mathrm{T}}(x)W\varphi(y))\right] \tag{12}$$

其中, ε 为阈值常数, 实验中设为 0.1, y_n 代表正确的类标签, W 为需要学习的模型参数, 该模型所使用的类标签在特征子空间中使用的映射, 为第 4.1.2 节中介绍的文本学习后得到的词汇向量. 通过对目标函数使用随机梯度下降方法求解模型参数矩阵 W.

2) 属性标签嵌入模型

属性标签嵌入模型[56] 是在深度语义嵌入模型的基础上, 使用属性标签嵌入替代了词汇向量嵌入, 此时的类标签的嵌入向量为 $\varphi(y)=[a_{y,1},\cdots,a_{y,M}]$, 这里 M 为类标签 y 中包含的属性个数.

训练阶段, 该模型选择使用等级目标函数[56]:

$$l(x_n,y_n,y)=\boldsymbol{I}(y_n=y)+\theta^{\mathrm{T}}(x)W[\varphi(y)-\varphi(y_n)] \tag{13}$$

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\beta_{r_{\Delta(x_n,y_n)}}}{r_{\Delta(x_n,y_n)}}\sum_{y\in Y_{tr}}\max\{0,l(x_n,y_n,y)\} \tag{14}$$

求解属性标签嵌入模型参数 W. 其中 Y_{tr} 代表训练类的类标签集合. 其中 $r_{\Delta(x_n,y_n)}=\sum_{y\in Y}\boldsymbol{I}(l(x_n,y_n,y)>0)$, $\beta_k=\sum_{i=1}^{k}\frac{1}{k}$. $\{x_n,y_n\}$ 为训练类样例–标签对, y 为训练类样例–标签对数据集中与 x_n 不同的类标签, 上述的加权排序目标函数将会保证正确标签的排名高于错误标签的排名, 在测试阶段, 该模型同样利用线性相容性函数 (11) 实现零样本学习下的测试类标签识别.

3) 属性描述替换零样本学习模型

属性描述替换模型是一种简单有效、易于理解的零样本学习模型[23]. 属性描述替换模型分类问题的目标函数为: $\min L(X^{\mathrm{T}}W,Y)+\Omega(W)$.

为了适用于零样本学习, Paredes 等将 W 矩阵分解为子权值矩阵 V 和属性信息矩阵 S 相乘的形式, 这样, 就得到了属性描述替换零样本学习模型的目标函数: $\min L(X^{\mathrm{T}}VS,Y)+\Omega(V)$

这一模型实质上是第 2.1 节中所介绍的模型空间方法的扩展. 在训练阶段, 使用训练类的辅助信息矩阵 S, 求解子权值矩阵 V. 之所以说这一方法是模型空间方法的延伸, 是因为模型空间方法中的 $d(y)$, 对应着这一方法中的 S. 在完成训练后, 使用训练得到的子权值矩阵 V, 在测试阶段把 S 替换为测试类的属性信息 S', 从而可以计算出测试类样例的类标签的估计值 S', 实现零样本学习. 这里 S'_j 表示测试类属性信息矩阵 S' 的第 j 列, 对应着第 j 个测试类的属性信息, 事实也对应了线性相容性模型式 (12) 中的 $\varphi(y)$. 这样, 测试阶段分类公式也变为 $\mathrm{argmax}_j\, x^{\mathrm{T}}VS'_j$.

损失函数形式为 $L(P,Y)=\|P-Y\|_F^2$, 且正则化项取式 (15) 形式时:

$$\Omega(V;S,X)=\gamma\|VS\|_F^2+\lambda\|X^{\mathrm{T}}V\|_F^2+\beta\|V\|_F^2,\quad \beta=\gamma\lambda \tag{15}$$

属性描述替换模型 V 具有解析解为:

$$V=(XX^{\mathrm{T}}+\gamma I)^{-1}XYS^{\mathrm{T}}(SS^{\mathrm{T}}+\lambda I)^{-1} \tag{16}$$

4) 基于文本描述替换的零样本学习模型

Qiao 等对属性描述替换模型进行了改进, 将模型中的属性信息替换为易于获得的、取自类描述文本得到的类描述信息, 并基于这一思想提出了基于文本描述的零样本学习模型[24], 该方法将属性描述替换模型中的矩阵 V 再次分解为 W_x 和 W_z 两个部分, 即令 $V=W_x^{\mathrm{T}}W_z$.

在训练过程中, W_x 和 W_z 同时进行训练, 训练目标函数如式 (17) 所示:

$$\min\Big(\left\|X^{\mathrm{T}}W_X^{\mathrm{T}}W_ZZ-Y\right\|_F^2+\lambda_1\left\|W_X^{\mathrm{T}}W_ZZ\right\|_F^2+\lambda_2\sum_{i=1}^{d}\left\|w_Z^i\right\|_2\Big) \tag{17}$$

w_Z^i 为 W_z 矩阵的第 i 个列向量, 共有 d 个. 第三项将会使得 W_z 矩阵中每个列向量的无关分量趋近于零, 用于对文本 Z 所引入的噪声进行滤波, 由每一类的文本描述使用第 4.1.2 节中方法提取得到的列向量所组成, 对应于线性相容性模型式 (12) 中的 $\varphi(y)$.

此时的 Z 为训练类对应的文本描述, 并在测试阶段替换为测试类对应的文本描述.

W_x 则是基于文本描述的零样本学习模型的模型参数, 负责对输入的图像进行分类. 此时分类公式如式 (18):

$$y* = \max_{y_{te}} x^{\mathrm{T}} W_x^{\mathrm{T}} W_z z_{y_{te}} \tag{18}$$

5) 多视图零样本学习模型

Akata 等基于自己所研究的属性嵌入模型的基础上, 利用了目前可用的三种辅助信息, 引入多个视图对应的多个相容性函数, 建立了多视图下的零样本学习模型[57]:

$$\begin{aligned} F(x, y; \{W\}_{1,\cdots,K}) &= \sum_k \alpha_k \theta_k^{\mathrm{T}}(x) W_k \varphi_k(y) \\ \text{s.t.} \sum_k \alpha_k &= 1 \end{aligned} \tag{19}$$

其中, $k \leq 3$, α_k 为超参数, 需要事先给定, 该模型针对 k 个视图, 建立了 k 个相容性函数, 每一个相容性函数的模型参数 W_k 都在训练过程中单独训练, 与其他视图上的模型参数无关, 在训练完成后, 使用凸组合将各个相容性函数组合在一起, 利用多个视图综合判断输入样例的类标签, 进而提高预测类标签的置信度.

6) 细粒度数据视觉描述的深度表示模型

Reed 等基于文本描述的辅助信息建立了一种端到端的零样本学习模型, 与其他模型不同, 该模型选择卷积神经网络分别对图片和文字进行处理[58], 提取其中的特征, 并利用式 (25) 的相容函数完成分类:

$$F(x, t) = \theta^{\mathrm{T}}(x) \varphi(t) \tag{20}$$

其中, x 代表视觉信息, t 代表文本描述, 在训练阶段, 对图像、文本描述这两个卷积神经网络学习 $\theta(x)$ 和 $\varphi(t)$ 函数. 在测试阶段, 该模型不仅可以在输入文本描述后对图片进行分类, 也可以输入图片对给定的文字描述进行分类, 分类函数分别为:

$$\begin{aligned} f_x(x) &= \arg\max_{y \in Y_{te}} \mathrm{E}_{t \in T_{te}(y)}[F(x, t)] \\ f_t(t) &= \arg\max_{y \in Y_{te}} \mathrm{E}_{x \in X_{te}(y)}[F(x, t)] \end{aligned} \tag{21}$$

第一行公式代表在输入测试类 x 的情况下, 文本描述取自与全部可能的测试类的文本描述 $T_{te}(y)$ (此时测试类全部的文本描述对应哪个类标签是已知的), 哪个分布使得 $F(x, t)$ 的期望最大, 那么这一分布 t 对应的类标签, 即为输入 x 的类标签预测值.

同理, 对于第二行公式, 以文本描述作为输入的测试样例 t 时, 使得该期望最大的测试类图像 x 对应的类标签 (此时测试类全部图像对应哪个类标签是已知的) 即为输入的文本描述 t 的类标签预测值.

7) 线性相容性模型分析和比较

深度语义嵌入模型并不是针对零样本学习所提出的, 也没使用零样本学习常用的数据集进行验证, 在 ImageNet 中泛化到从未见过的测试类时, 实验结果并不是十分理想, 但该模型为零样本学习提供了线性相容性函数这一方法意义重大, 属性嵌入模型、组合多个视图进行相容性判断的多视图模型, 以及细粒度数据视觉描述的深度表示模型都传承于深度语义嵌入模型的思路.

属性嵌入模型与多视图模型相比各有优劣, 人工标注的属性信息的准确度和训练效率要明显高于通过文本描述得到的词汇向量以及类层次结构, 但文本描述以及类层次结构的易获得性仍然让使用非属性辅助信息的模型, 如多视图模型、文本描述替换模型等模型在实际商业应用中有着巨大的潜力.

文本描述替换模型与属性描述替换模型受到了模型空间方法的启发, 在测试阶段将训练类的描述矩阵替换为测试类的描述矩阵, 这两种方法都具有较高的运算效率以及可操作性, 但将图像特征线性映射到特征空间的效果与其他非线性映射函数相比较差.

对于线性相容性模型来说, 由于受限于线性函数表达能力的欠缺, 其准确率与非线性相容性模型相比仍然处于劣势, 因此目前来说, 在相容性模型中, 线性相容模型并不是主流的研究方向.

4.2.2 非线性相容性模型

顾名思义, 非线性相容性模型的相容性函数是非线性的, 一般来说有两种常见形式: 距离函数形式和概率形式. 在本节中, 我们将介绍 7 种典型非线性相容性模型: 语义自编码器模型[59]、语义嵌入一致性度量模型[60]、联合隐变量相似嵌入模型[61]、隐嵌入模型[62]、反向深度嵌入模型[17]、基于文本描述的深度零样本卷积神经网络[16] 和基于随机森林的零样本学习模型[63].

1) 语义自编码器模型

Kodirov 等开创性的将自编码器引入到零样本学习中[59], 该模型假定将输入的图像信号通过编码器编码可以得到辅助信息 S (例如属性、语义向量), 并且可以通过解码器将辅助信息还原图像信号, 自编码器零样本学习示意图如图 15 所示.

在训练阶段, 该模型利用训练类的图像和辅助信息, 对编码器 W 和解码器 W^{T} 进行训练, 该方法的价值在于可以在测试阶段既可以利用 W 也可以利用 W^{T} 实现对测试类的分类.

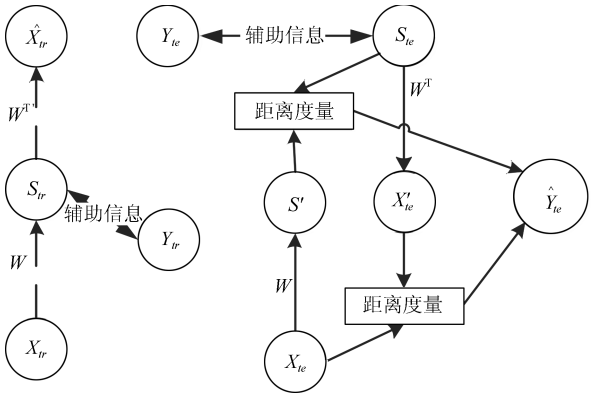

图 15 语义自编码器零样本学习示意图

Fig. 15 Semantic autoencoder zero-shot learning

语义自编码器零样本学习的目标函数为:

$$\min_{W,W*} \left\|x - W^{\mathrm{T}}Wx\right\|_F^2 \tag{22}$$

这里 W 为编码器, W^{T} 为解码器. 因为 $Wx = s$, 所以加入额外的正则化项, 学习目标函数改写为:

$$\min_{w} \left\|x - W^{\mathrm{T}}Wx\right\|_F^2 + \lambda \left\|Wx - s\right\|_F^2 \tag{23}$$

在分类阶段, 已经通过训练得到编码器 W 和解码器 W^{T} 后, 测试类的识别可以通过如下两种方法实现:

1) 可利用编码器 W 将测试类输入样例嵌入到语义空间中, 得到 $\hat{s}_i = Wx_i$ 并比较 $\hat{s}_i$ 与测试类辅助信息 s_j 之间的距离, 距离最小的辅助信息对应的类标签即为输入样例的类标签:

$$y = \arg\min_j D(\hat{s}_i, s_j) \tag{24}$$

2) 同样, 也可以利用解码器 W^{T}, 将各测试类辅助信息嵌入到图像特征空间中, 得到 $\hat{x}_i = W^{\mathrm{T}}s_i$, 比较测试类输入样例与各个辅助信息嵌入向量之间的距离, 与输入测试样例距离最小的辅助信息嵌入向量的类标签, 就是测试类样例的类标签的估计值:

$$y = \arg\min_j D(x_i, \hat{x}_j) \tag{25}$$

2) 语义嵌入一致性度量模型

Bucher 等首次引入了度量学习模型来提高语义嵌入的一致性[60], 一般情况下, 相容性度量函数形式为:

$$F(\theta(x),\varphi(y),w) = d_A(\theta(x),\varphi(y)) = \left\|(\theta(x) - \varphi(y))^{\mathrm{T}}W\right\| \tag{26}$$

语义嵌入一致性度量模型中考虑使用单层神经网络作为图像特征的嵌入函数: $\theta(x) = \max(0, x^{\mathrm{T}}W_x + b_x)$, 对输入进行属性识别, 并将其嵌入到语义空间, 并度量嵌入向量与类标签语义向量之间的距离.

此时度量函数变为:

$$F(\theta(x),\varphi(y),W) = d = \left\|(\theta(x) - \varphi(y))^{\mathrm{T}}W\right\| \tag{27}$$

在训练阶段利用训练类求解语义嵌入一致性度量模型参数 W, 测试阶段时对于输入的测试样例, 分类公式为:

$$k^* = \arg\min_{k\in\{1,\cdots,C\}} F(\theta(x),\varphi(y_k)) \tag{28}$$

$k \in \{1,\cdots,C\}$ 为可能的类标签索引值, 使得度量函数 $F(\theta(x),\varphi(y),W)$ 最小的类标签值, 就是输入样例的类标签估计值.

3) 联合隐变量相似嵌入模型

联合隐变量相似嵌入模型利用训练类对嵌入模型和判别函数同时进行训练[61], 用嵌入模型得到文字与图像的隐表示. 并使用判别函数估计输入输出隐变量的联合后验概率, 从而确定输入输出之间的相容性关系, 完成对输入测试样例类标签的估计.

设图像和文字的隐表示分别为 $v_i = \theta(x_i)$ 和 $t_j = \varphi(y_j)$. 在训练隐表示函数 $\theta(x_i)$ 和 $\varphi(y_j)$ 的同时, 也学习判别函数 (29).

在测试阶段输入测试类样例后, 测试类样例分类规则为:

$$j = \arg\max\{\log p(i = j|v_i, t_j)\}\} \tag{29}$$

其中, $j \in \{1,\cdots,c_{te}\}$, 代表类标签的索引值, 在给定测试类输入样例的隐变量后, 如果第 j 类标签对应的文字描述的隐变量 t_j 使得式 (29) 的值最大, 那么输入样例的类标签估计值即为 j.

4) 隐嵌入零样本学习

Xian 等在多视图零样本学习模型的研究基础上, 构造了一个分段线性兼容性函数[62]:

$$F(x,y) = \max_{1\le i\le K} x^{\mathrm{T}}W_i y \tag{30}$$

这里 K 为预先给定的超参数, 需要通过交叉校验过程确定其取值. 之所以会使用多个模型参数 W_i, 是因为这些 W_i 可以最大化所有训练样例的输入嵌入和输出嵌入之间的兼容性. 不同的模型参数 W_i 可以针对性地识别对象的不同视觉特征的隐嵌入, 即颜色和形状等, 并允许在它们之间分配权重, 使得模型能够做出更好的分类预测. 隐嵌入零样本学习经验损失函数为:

$$\frac{1}{N}\sum_{n=1}^{C_{tr}} L(x_n, y_n) \tag{31}$$

其中, 损失函数:

$$L(x_n, y_n) = \sum\nolimits_{y \in Y_{tr}} \max\{0, I[y_n = y] + F(x_n, y) - F(x_n, y_n)\}$$

为大间隔损失函数.

5) 反向深度嵌入模型

Zhang 等提出了一种与之前将图像特征应用到语义空间相反的思路[17], 他们使用双层神经网络将语义向量映射到图像特征空间, 即 $\varphi(y) = f_2(W_2 f_1(W_1 s(y)))$ 这一模型与之前的模型相比在一定程度上缓解了枢纽化问题, 这是因为类标签的输入明显少于输入图像的个数, 尽管将类标签映射到图像特征空间仍然会有枢纽化现象发生, 但此时的枢纽化现象对于类标签的映射向量造成的污染, 要明显小于将输入图像特征映射到语义空间对输入图像特征映射向量的污染.

在训练阶段, 利用已经提取得到的图像特征, 以及类标签对应的语义向量, 求解 $\varphi(y) = f_2(W_2 f_1(W_1 s(y)))$, 在测试阶段, 则使用式 (32) 的分类函数:

$$y = \arg\min_{y \in Y} D(\theta(x), f_2(W_2 f_1(W_1 s(y)))) \quad (32)$$

式 (32) 中的 D 为距离度量函数, 零样本学习的深度嵌入模型不需要学习相容性参数矩阵 W, 直接使用 F 度量 $\theta(x)$ 和 $f_2(W_2 f_1(W_1 s(y))$ 之间的距离, 使 F 距离最小的类标签, 就是输入图像的类标签的估计值.

6) 基于文本描述的深度零样本卷积神经网络

Ba 等提出了一种直接实现文本描述的新模型, 该模型可以利用文本描述对未曾见过的对象类进行分类[16]. 该模型的核心思想是通过使用文本特征来预测深度卷积神经网络 (Convolutional neural network, CNN) 中的卷积层和全连接层的输出权值来实现零样本学习.

在训练阶段, 首先利用训练集中的样本对 CNN 的模型参数进行训练, 并利用第 2.1.2 节中的模型空间方法的思路确定训练集样本所对应的文本描述与此时的 CNN 参数之间的关系, 即通过训练得到文本描述对应于 CNN 参数的映射. 在测试阶段, 就可以利用测试类对应的文本描述确定此时的 CNN 参数, 从而实现测试类样例的识别.

该模型利用了 CNN 的神经网络结构, 并可以在不同的抽象层次上学习特征, 这一点与之前提到的同时学习将图像特征和类标签嵌入到特征子空间两个嵌入模型的方法完全不同.

7) 基于随机森林的零样本学习模型

与其他基于相容性函数的方法不同, Jayaraman 等提出用随机森林来判断输入样例与类标签的属性特征之间的相容性[63]. 而且, 这一方法同时也考虑了零样本学习中的属性不可靠性 (属性分类器的不可靠性、相同属性在不同环境下的视觉特征也可能明显不同), 并设计了鲁棒随机森林, 校正类标签属性不可靠性对分类器预测结果的负面影响.

在训练阶段, 该模型利用训练类训练每一个属性分类器, 并留出训练类中的一部分样本作为交叉验证集判断每一个属性分类器的准确度. 以每一个属性分类器的准确度作参考, 对测试类的随机森林中人为的给定的属性向量, 即测试类的属性辅助信息进行数值调整, 使得每一个属性的在属性分类器下的可靠性可以传递到每一个叶子结点, 提高测试类随机森林的鲁棒性.

测试阶段利用属性分类器得到输入样例的属性指示向量后, 并使用在训练类调整后得到的属性随机森林, 对测试类输入样例的类标签进行判断. 该方法在属性可靠性较低的情况下, 具有着较强的鲁棒性.

8) 非线性相容性模型分析与比较

反向深度嵌入模型只考虑将语义向量映射至图像空间, 这一方法虽然规避了图像特征空间映射到语义特征空间的映射域偏移问题, 但在语义特征空间到图像特种空间中的映射, 仍然有轻微的映射域偏移问题存在.

与反向深度嵌入模型相比, 语义自编码模型首次同时考虑将图像特征映射至语义特征空间和将语义特征映射至图像特征空间, 两种映射互相约束下, 模型的准确度也得到了提高.

隐嵌入模型是在多视图模型的基础上发展而来, 虽然将加权下的线性相容函数替换为分段线性函数, 但这仍然没有摆脱线性函数的本质.

随机森林模型利用一部分数据对测试类的辅助信息进行了矫正, 提高了模型的鲁棒性, 但这一模型训练过程十分复杂, 应用价值较低.

联合隐变量相似嵌入模型首次将隐变量概念引入零样本学习中, 通过学习图像特征与辅助信息的隐表示, 并比较其相似性实现零样本学习下的测试类识别, 这一方法在训练过程中需要同时优化两个映射函数和相容性函数, 不仅计算量巨大, 而且各个参数之前的互相关联也影响了识别准确性.

4.3 混合模型

1) 语义凸组合模型

语义凸组合模型是最为经典的零样本学习混合模型[45]. 首先, 该模型使用分类器在训练类样本中进行训练, 使得训练得到的分类器能够对训练类进行准确的识别.

测试阶段, 输入测试类样例到训练类分类器时, 每一个训练类分类器则会判断测试类输入与哪些训练类更为相似, 计算度量测试类样例与每个训练类相似程度的后验概率, 之后利用相似程度的后验概率以及训练类标签的语义嵌入向量进行凸组合, 就可以得到测试类样例的语义嵌入向量[45].

首先定义 $\hat{y}_0(x,1)$ 为分类器 p_0 计算得到的输入样例 x 最有可能的训练类标签:

$$\hat{y}_0(x,1) = \arg\max_{y\in Y_0} p_0(y|x) \tag{33}$$

同理, 定义在 $\{p_0(y|x); y \in Y_{tr}\}$ 上, p_0 给出样例属于第 t 个可能的训练类标签 $\hat{y}_0(x,t)$ 的条件后验概率, 给定 p_0 输出的前 T 个 x 的训练类标签的预测, 模型就可以确定 x 的预测语义嵌入向量 $f(x)$, 为训练类语义嵌入向量 $s(\hat{y}_0(x,t))$ 的凸组合, 凸组合权重由分类器判断样例属于每一个类标签的条件后验概率:

$$f(x) = \frac{1}{Z}\sum_{t=1}^{T} p(\hat{y}_0(x,t)|x)\cdot s(\hat{y}_0(x,t)) \tag{34}$$

其中, Z 为归一化常数, $Z = \sum_{t=1}^{T} p(\hat{y}_0(x,t)|x)$, T 为超参数, 用于控制需要考虑在内的最大类标签语义嵌入向量的个数.

在得到测试类输入样例的凸组合嵌入向量之后, 使用余弦相似性度量函数, 比较嵌入向量与测试类类标签的相似度, 就可以确定输入测试样例的类标签值.

2) 综合零样本学习模型

综合零样本学习[46] 方法针对上述凸组合模型中的缺点, 进行了改进, 综合零样本学习模型的分类规则为:

$$\hat{y} = \arg\max_{c} W_c^{\mathrm{T}} x \tag{35}$$

W_c 为对应每一类的分类器模型向量, 因此在零样本学习中, 只要成功学习了测试类的分类器模型向量, 就可以实现零样本学习, 此外, 每一个类都有对应的坐标 a_c, a_c 可以是人为给定的属性坐标[14, 26], 也可以是类标签通过文本描述学习后得到的语义向量[52].

综合零样本学习模型的核心概念是引入了虚拟类 r, 以及对应的虚拟分类器 v_r 和对应的虚拟坐标 b_r $(r = 1,\cdots,R)$. 真实类和虚拟类之间使用权值 s_{cr} 关联起来:

$$s_{cr} = \frac{\exp\{-d(a_c,b_r)\}}{\sum\limits_{r=1}^{R}\exp\{-d(a_c,b_r)\}} \tag{36}$$

$d(a_c,b_r)$ 为距离度量函数, s_{cr} 可以看作是在类 c 邻域中观察到类 r 的条件概率. 为了使嵌入误差最小, 用经验损失函数:

$$\min \left\| W_c - \sum_{r=1}^{R} s_{cr} v_r \right\|_2^2 \tag{37}$$

求解 (37) 就可以得到每一类分类器模型向量的解析解:

$$W_c = \sum_{r=1}^{R} s_{cr} v_r, \qquad \forall c \in \{1,2,\cdots,N_0+N_1\} \tag{38}$$

从训练类中随机抽出 r 类作为虚拟类, 即 $r \in c_{tr}$, 此时 b_r 值即为对应的 a_{tr}. 在训练阶段对每一个训练类分类器进行训练, 测试阶段输入测试类样例时, 利用每一个测试类的坐标 a_{te}, 就可以得到每一个测试类的分类器 W_{te}, 这样, 利用式 (35) 的判别规则, 就可以实现零样本学习下对测试类样例的识别.

3) 语义相似嵌入模型

语义相似嵌入模型分类规则为[64]:

$$y = \arg\max_{y'} \pi^{\mathrm{T}}(\theta(x))\psi(\phi(y')) \tag{39}$$

该模型在图像特征空间以及语义空间中同时利用不同类之间的相似性, 使用稀疏编码 $\psi(\cdot)$ 将测试类的语义向量 $\phi(y)$ 表示为训练类语义向量的组合, 并利用每一测试类所对应的特定相似性变换 $\pi(\cdot)$ 将输入样例的图像特征 $\theta(x)$ 嵌入到与 $\psi(\cdot)$ 相同的空间中, 最终利用测试类标签与训练类标签的相似性和测试类样例与训练类样例的相似性来判断测试类样例的类标签, 实现零样本学习. 示意图如图 16 所示.

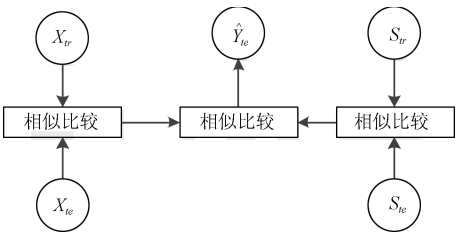

图 16 语义相似嵌入模型零样本学习示意图

Fig. 16 Semantic similarity embedding zero-shot learning model

4) 高斯混合模型

高斯混合模型首先构建了训练类上的输入高斯分布模型[65], 并利用给定的辅助知识将测试类和训练类关联起来, 使用训练类上的高斯模型的混合模

型, 作为测试类的高斯模型, 并使用该模型实现了对测试类样例的分类.

5) 结构预测模型

Zhang 等提出的结构预测模型与高斯混合模型有着一定的相似之处[66], 该模型使用最大后验估计来确定测试类输入的类标签, 并将最大后验概率参数化为测试类样例的高斯分布, 利用训练类中的辅助知识对测试类样例的高斯分布进行估计, 测试类样例的高斯分布表示为训练类样例高斯分布的结构化组合形式, 进而确定测试类的高斯分布模型参数, 实现对测试类样例所属类别的识别.

6) 相关知识迁移模型

这一方法与反向深度嵌入方法十分类似, 是一种利用相关知识迁移实现的零样本学习[67], 相关知识迁移, 是指利用测试类与训练类之间的相关知识 (即辅助信息) 之间的关系来估计测试类的高斯分布. 得到测试类的高斯分布后, 利用它随机生成测试类的虚拟输入样例, 与测试类标签组成测试类样本, 并对图像特征空间的流形完成补全, 相关知识迁移模型使用测试类的 "样本" 进行训练, 实现零样本学习.

7) 属性回归类原型矫正模型

属性回归类原型矫正模型将零样本学习问题转化为属性回归问题[68]. 但是, 与一般的回归问题不同, 该模型在提出了新的观点, 它将测试类与训练类联系起来, 在使用训练类对模型进行训练的同时, 也对测试类的向量表示 (属性向量或语义向量) 进行了矫正. 这一措施有效地缓解了映射域偏移问题和枢纽化问题.

8) 基于神经网络输出无监督学习应用的零样本学习

Hinton 等在 2014 年指出, 经过训练的神经网络的软最大输出包含着比单独的分类器输出更丰富的信息[69]. 举例来说, 分类器可能会对狗有关的图像输入判断其类标签为猫的概率为 0.01, 但判断汽车图像为猫的概率应该比 0.01 更小, 因此, 为了揭示这些隐结构关系, Lu 基于这一理论提出了一种无监督零样本学习方法[70], 将两种无监督学习算法: 主成分分析 (Principal component analysis, PCA)[71] 和独立成分分析 (Independent component analysis, ICA)[72] 应用到具有 1 000 个类的 ImageNet[73] 训练样本的深度卷积神经网络的输出上. 并将这一方法应用在零样本学习中, 通过使用在训练类上训练过的神经网络, 对于测试类输入样例进行无监督学习, 从而判断测试类输入的类标签.

9) 混合模型分析与比较

语义凸组合模型的主要缺点在于直接利用测试类和训练类的语义向量, 并利用训练类语义向量的凸组合来表示测试类输入的语义向量表示, 没有充分利用测试类与训练类语义向量之间的关系.

综合零样本学习模型针对这一缺点进行了优化, 选择利用训练类与测试类之间的关系构建了测试类的分类器, 完成对测试类样例的识别.

相关知识迁移模型, 使用混合训练类高斯分布得到的测试类高斯分布生成测试类 "样例" 对模型进行训练, 这一方法会将一定的噪声引入模型中, 影响训练效果.

基于神经网络输出的无监督学习模型同样依赖于训练类与测试类的相似性, 而且该方法是一种对于深度神经网络输出加以应用的一种尝试, 不仅训练过程十分复杂, 得到的结果也与其他模型相比较差.

但是对于所有混合模型, 尤其是语义相似嵌入模型、相关知识迁移模型和高斯混合模型和结构预测模型, 它们都过于依赖测试类与训练类之间的相似关系, 如果测试类与训练类之间的相似性很高 (例如斑马和马), 那么混合模型很可能会达到非常好的识别, 但如果测试类与训练类之间相似性较差 (例如汽车和猫), 那么, 混合模型的识别效果将会受到很大的影响.

4.4 直推模型和归纳模型

另一种分类方法, 是将零样本学习分为直推模型和归纳模型两类, 这一分类方法是由 Song 等提出[74], 其中最为典型的直推方法为 Fu 等提出的多视图直推模型[42], 直推式模型的特点为, 在测试开始阶段时没有可参考的测试样例或最新输入的测试类样例与其他可参考的测试样例相关性较低时, 利用训练类的类标签与测试类的辅助信息, 确定输入测试样例的类标签, 之后把已经确定类标签的测试类样加入到训练样例集中, 在增广样例集上学习新的判断规则, 进而判断后输入的测试类样例类标签, 这个过程反复迭代, 直到所有测试样例均被标注.

除多视图直推模型外, 直推模型还有无偏嵌入直推模型[74]、模型空间共享直推模型[75] 和无监督域自适应模型[50] 三种. 直推式模型属于一种在测试阶段也会不断学习的在线学习模型, 但这一方法需要较大的计算量, 且十分依赖于初始阶段对于测试类样例分类的准确率, 一旦初始阶段测试类样例分类错误率较高, 那么后续的识别效果将会受到较大的影响, 因此直推模型这一方向目前仍处于探索阶段.

与直推模型相对应的是归纳模型, 归纳模型只使用训练类中的数据, 并从其中归纳出某些特定的规律 (例如相似性或属性) 应用于测试类样例的识别. 除了上述三种直推模型以外, 常见的零样本学习

模型都是归纳模型, 例如深度语义嵌入模型[55]、多视图零样本学习模型[57] 和综合零样本学习模型[46]等.

值得注意的是, 直推式模型的学习方式与人类不断吸收新知识, 不断提高识别准确性的过程十分相似, 因而未来零样本学习的研究方向将必然聚焦在直推模型这一领域之中.

5 实验分析

5.1 实验数据集介绍

图像分类领域最为常用的数据集是 ImageNet, 该数据集包含高达 14 197 122 张图片, 分别分布于 21 841 个类别中, 而且, 其中的类标签都利用 WordNet 中的层次关系, 将类标签在 WordNet 的层次上互相连接.

但是在对象类上, 只使用了单独的类标签进行标注, 没有类标签对应的属性信息, 因此目前来说, 大部分零样本学习并没有在实验中使用 ImageNet 数据集.

在零样本学习中, 最为常用的数据集是 Animals with Attributes (AwA) 数据集[14]、Lampert、Nickisch 等以 50 个动物类作为关键词, 利用 4 个搜索引擎 Google、Bing、Yahoo 以及 Flickr 搜索它们的图片, 得到的 180 000 张图片后, 消除其中的异常图片和重复图片, 通过这一系列预处理后, 得到了 30 475 张图片, 共有 50 个类, 而且对于图片最少的类也仍然有 92 张图片. 同时将每一个动物类通过其属性特征化, 每一类共有 85 属性特征. 属性特征既可以是布尔值, 也可以为实值, 这样就得到了 AwA 数据集. 通过将采集到的图像与语义属性表相结合, AwA 数据集可以应用于任何利用属性知识对模型进行学习, 以及利用属性知识进行分类的实验中.

在 2017 年, Xian 等在 AwA 数据集的基础上, 创立了 AwA2 数据集[76]. AwA2 数据集有 37 322 张图像, 类与属性的数目与内容不变, 样例最多的类 "马" 有 1 645 张图片, 最少的类 "鼹鼠" 也有 100 张图片.

此外, 在零样本学习中广泛应用的属性数据集还有三个: Caltech-UCSD-Birds200-2111 (CUB)[77]、Attribute Pascal and Yahoo (aPY) 数据集[32] 以及 SUN attribute 数据集[40].

CUB 数据集为加州理工大学建立的鸟类数据集, 共有 11 788 张图片, 并使用 312 条属性特征对 200 种不同的鸟类加以描述.

aPY 数据集由 Farhadi 等整理, 共有两部分, 一部分是 PASCAL VOC 2008 数据集的子集, 共有 12 695 张, 另一部分是通过 Yahoo 搜索引擎收集的 2 644 张图片. PASCAL 部分作为训练集, Yahoo 部分作为测试集, 而且两部分的类互不重叠, 因此, 该数据集完全满足了零样本学习的实验要求.

SUN attribute 数据集由 Patterson 和 Hays 整理, 适用于场景识别实验的数据集. SUN attribute 数据集是 SUN 数据集[78] 的子集, 为了简化, 下文中将 SUN attribute 数据集简称为 SUN. SUN 数据集中包括了细粒度的场景类, 以及详细的属性注释, 同样可以应用于场景识别领域的零样本学习中.

这 5 种数据集的主要特征如表 1 所示:

5.2 实验结果分析

我们选取 8 个具有代表性的模型在 4 个数据集中的实验结果. AwA2 数据集为 2017 年由 Xian 等提出目前还没有在实验种大规模中投入使用, 现有零样本学习方法所使用的实验数据集基本仍然以 AwA、CUB、aPY、SUN 4 种数据集为主, 因此文中在此介绍多个模型在该 4 种数据集下的实验结果.

为了保证在同一基准下比较, 这些模型在 AwA 数据集下的实验都选择了同样的 10 个测试类: 黑猩猩、大熊猫、河马、座头鲸、豹、猪、浣熊、老鼠和海豹. 这些类的图像作为测试数据, 其余 40 个类别的 24 295 个图像用于训练.

在 CUB 数据集下测试类为 50 个, 训练类为 150 个. 在 SUN 数据集中, 带星号的准确率表示实验中, 测试类个数为 10, 训练类个数为 707, 不带星号的准确率表示训练类个数为 645, 测试类个数为 72. aPY 数据集下, 训练类为 20 种, 测试类 12 种.

表 1 5 种数据集属性介绍

Table 1 Introduction to the attributes of the five datasets

数据集	AWA	CUB	aPY	SUN	AwA2
图像个数	30 475	11 788	15 539	14 340	37 322
类个数	50	200	32	17	50
属性个数	85	312	64	102	85
注释水平	每一类	每张图片	每张图片	每张图片	每一类
注释类型 (实值或布尔值)	兼有	兼有	兼有	布尔	兼有

表 2 多个模型在 4 个数据集下的实验结果

Table 2 Experimental results of the models under four data sets

模型 — 数据集 (%)	AWA	CUB	aPY	SUN
DAP[14]	41.4	28.3	\	19.1
IAP[14]	42.2	24.4	\	16.9
ESZSL[23]	49.3	\	65.8*\18.7	15.1
SYNC[46]	69.7	53.4	62.8	\
SSE[64]	76.3	30.4	82.5*	46. 2
LATEM[62]	71.9	45. 5	\	\
SJE[57]	66.7	50.1	56.1	\
SAE[59]	84.7	61.4	91.0*\65.2	54.8

从实验结果中我们可以看到随着零样本学习方法的发展, 各种模型的识别准确率逐渐提高, 目前来看, 语义自编码模型是准确度最高的模型, 更为重要的是, 该模型所使用的嵌入函数为线性嵌入函数, 证明零样本学习的准确率仍然有着很大的提升空间.

此外, 不同模型与不同数据集之间的契合度存在着很大的差异, 例如 LATEM[62] 模型对于 CUB 数据集更为敏感, 而 SSE[64] 型对于 AwA 中的样例识别准确度较高.

有一点非常值得我们注意, 4 个数据集中测试类与训练类个数的比值, 测试类样本与训练类样例的比值是有着明显不同的, 以 aPY 数据集比值最大, 准确率在 4 种模型中也相对最低, SUN 数据集中, 比值分别为 10/707、72/645 明显可见测试集与训练集比值越小, 准确率相对越高, 表明模型训练完成度越好.

这一结果是符合我们直观感受的, 训练类样本越多, 模型在训练类中所学习到的知识也就越多, 在测试类种可以使用的辅助信息也就越有效, 而且目前的零样本学习模型, 并没有因为在训练类上具有较大的样本集合而导致过拟合现象, 却反而拥有较强的测试类泛化能力. 因此, 在目前的零样本学习模型下, 对于某些难以获得训练样本的测试类来说, 如果想要较为方便地得到更好的零样本学习识别效果, 直接增多易于获得的训练类个数以及训练类的样本, 应该会是一个简单有效的方法.

6 零样本学习应用

6.1 图像处理

零样本学习中常用的 4 个实验数据集, 两个数据集为动物图像数据集, 因此, 在动物识别方面, 零样本学习有着得天独厚的优势, 属性等辅助信息对于动物的描述, 在提高识别准确率上的效果也是十分明显的, 以 CUB 数据集为例[77], 该数据集为加州理工大学所建立的鸟类图像数据集. 如果在野外使用该数据集下训练过的零样本学习鸟类识别系统, 就可以通过对某种珍稀鸟类的特征描述, 对野外生存的珍稀鸟类进行识别, 这种能力将会为人类带来优秀的生态效益和经济效益.

同样, 零样本学习也可以将 SUN 数据集下训练的优势发挥在场景识别领域中[40], “能够识别从未见过的场景” 这一强大的能力, 在机器人寻路、交通领域、增强现实领域都蕴含着巨大的应用潜力.

Antol 等提出了一种基于零样本学习的人类姿态识别模型[79], 通过输入对某些特定动作的抽象描述后, 模型会将抽象描述转化为具体的参数描述, 准确地识别图片中人类的姿态.

Pieter 等提出了基于零样本学习的指纹识别方法[80], 该方法不再依赖局部或者低维特征, 而是使用整个信号的特征, 实验结果表明使用不同芯片组成的芯片组可以实现精度 99 % 以上的指纹识别, 而且只需使用低廉的商用设备就可以实现电子指纹识别, 且采样速率低于 1 MB 每秒.

Yang 等提出了[81] 一种基于零样本学习的哈希编码算法, 规避了哈希编码学习中的昂贵的人工标签成本. 该算法将未见类压缩为二进制编码, 并利用可见类学习哈希函数, 将每个数据标签投影到语义嵌入空间中, 将可见类中知识迁移到未见类中, 该方法与其他哈希编码方法相比也更为先进.

6.2 自然语言处理

Johnson 等[82] 提出一种基于零样本学习的单个神经机器翻译 (Neural machine Translation, NMT) 模型来翻译多种语言的方法. 该方法并不需要改变基础系统的模型架构, 包括编码器, 解码器和注意力等模型的其余部分都保持不变, 并在所有语言中互相迁移. 通过使用迁移的模型部分, 该方法可以使用单一模型实现多个语言 NMT, 而且不增加任何参数, 这明显简化了以前的多语种 NMT 方法.

Sappadla 等[83] 提出了一种简单的零样本多标签文本分类方法, 该方法使用标签和文档单词的语义嵌入, 并根据标签和文档单词之间的相似性对以前未曾见过的标签进行预测.

7 零样本学习的未来研究方向

作为机器学习领域中的一个新兴方向, 零样本学习近几年来取得了飞速的发展. 零样本学习是一种衍生于深度学习, 与深度学习相对应, 却又紧密相连的一种学习方法. 对于某些测试类样本难以得到的情况, 可以通过零样本学习实现测试类的预测, 但是, 零样本学习仍然需要使用较多的训练类样本进行训练, 并通过一系列算法将训练类样本与测试类共享的辅助信息从训练类迁移到测试类中, 从而完

成对测试类样例分类的任务.

目前来看, 零样本学习在未来的研究中有如下 4 个潜在的方向:

1) 目前来说, 零样本学习如果想要实现高精度的识别, 仍然需要使用属性标注进行学习, 但是训练类和测试类的增加, 也会增加大量的属性标注的工作量. 如果可以使用更好的算法利用网络上现有的文本内容 (例如各个类标签的维基百科), 因为网络中的文本内容都是唾手可得的, 可以大大减少零样本学习的工作成本, 使得零样本学习推广到更多方面.

2) 图像特征映射函数以及语义向量映射函数是零样本学习的核心, 目前来看, 零样本学习如果想要提高识别准确率, 需要将两种映射函数进行改进, 例如不再拘泥于将图像特征映射至语义空间, 沿着深度嵌入模型[17] 以及语义自编码模型[59] 的思路继续深入研究, 考虑将语义向量映射至图像特征空间中或是同时引入这两种映射, 可能会得到更好的实验结果.

3) 单样本学习是与零样本学习相似的一个概念, 单样本学习是指在学习过程中, 对于特定任务或者特定类, 只有一个或者少量几个样本, 通过一系列的方法使得单样本学习模型能够完成特定的识别或者其他任务. 现有的零样本学习与单样本学习虽然概念和方法有一定程度上的相似, 但在具体实现机制上仍然有着较大的区别, 虽然有一些零样本学习模型虽然对输入少量测试类样本用于训练的情况加以考虑, 但仍不成熟. 希望未来研究中能够对这一方面进行深入研究, 实现零样本学习和单样本学习的有机结合, 得到一些较为有效的统一模型或是切换模型.

4) 零样本学习未来可以应用于故障诊断方面, 以设备平时工作状态为训练集, 设备故障工况作为测试集, 并对零样本学习模型给予辅助信息, 使得零样本学习可以在未曾见过故障工况的情况下确定故障原因, 这一应用若能成功实现, 将会为工业领域带来巨大的经济效益.

8 结论

零样本学习是近年来机器学习领域中的新生方向, 这一方向与传统的机器学习方法的不同在于零样本学习能够识别从未见过的类别中的测试样例, 这一方向具有可期的研究前景, 蕴含着巨大的潜在效益.

零样本学习的学习过程包括两大部分: 训练类中的训练过程和测试类中判别过程, 测试类与训练类之间没有交集. 文章首先指出了零样本学习的发展过程, 以及零样本学习的具体定义, 并对 4 种具有历史意义的与零样本学习相关的学习模型: 新任务的零数据学习[21]、语义输出编码零样本学习[15]、基于属性类间迁移的未见类学习[14] 以及跨模态迁移的零样本学习[22] 进行了介绍.

在这之后, 我们对目前零样本学习领域所存在的问题: 广义零样本学习、枢纽化问题、映射域偏移问题进行了介绍并对这三个问题的解决思路进行了说明.

在第 4 节和第 5 节部分, 我们详细介绍了零样本学习的现状发展, 目前的零样本学习模型可以分为两大类: 相容性模型和混合模型, 相容性模型又分为线性相容性模型和非线性相容性模型两个子类, 我们对每一类中所包括的模型进行了介绍, 并对其中较为典型的模型的实验结果进行了分析.

最后, 我们对零样本学习目前的应用场景: 图像处理和自然语言处理中的发展进行了介绍, 对其未来可能的发展方向进行展望, 进一步说明了零样本学习的巨大潜力.

随着零样本学习理论与方法研究的深入, 零样本学习将会更为成熟, 并应用于更多的机器学习场景, 终将为机器学习领域做出更大的贡献.

References

1 Schölkopf B, Smola A J. *Learning with Kernels*. Cambridge: MIT Press, 2001.

2 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe, USA: MIT Press, 2012. 1097−1105

3 Gregor K, Danihelka I, Graves A, Rezende D J, Wierstra D. DRAW: a recurrent neural network for image generation. arXiv preprint arXiv: 1502.04623, 2015.

4 Biederman I. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 1987, **94**(2): 115−147

5 Yao B P, Khosla A, Li F F. Combining randomization and discrimination for fine-grained image categorization. In: Proceedings of the Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2011. 1577−1584

6 Murphy G L. *The Big Book of Concepts*. Cambridge: MIT Press, 2004.

7 Koggalage R, Halgamuge S K. Reducing the number of training samples for fast support vector machine classification. *Neural Information Processing*, 2004, **2**(3): 57−65

8 Li F F, Fergus R, Perona P. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4): 594−611

9 Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap T. One-shot learning with memory-augmented neural networks. arXiv preprint arXiv: 1605.06065, 2016.

10 Fanello S R, Gori I, Metta G, Odone F. One-shot learning for real-time action recognition. In: Proceedings of Pattern Recognition and Image Analysis. Berlin, Heidelberg: Springer, 2013. 31−40

11 Pan S J, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, **22**(10): 1345−1359

12 Bakker B, Heskes T. Task clustering and gating for Bayesian multitask learning. *Journal of Machine Learning Research*, 2003, **4**(12): 83−99

13 Bonilla E V, Agakov F V, Williams C K I. Kernel multitask learning using task-specific features. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics, Atherton, USA: PMLR, 2007. 43−50

14 Lampert C H, Nickisch H, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 951−958

15 Palatucci M, Pomerleau D, Hinton G, Mitchell T M. Zero-shot learning with semantic output codes. In: Proceedings of the 22nd International Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: Curran Associates Inc., 2009. 1410−1418

16 Ba J L, Swersky K, Fidler S, Salakhutdinov R. Predicting deep zero-shot convolutional neural networks using textual descriptions. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 4247−4255

17 Zhang L, Xiang T, Gong S G. Learning a deep embedding model for zero-shot learning. arXiv preprint arXiv: 1611.05088, 2016.

18 Zhang D, Liu Y, Si L. Serendipitous learning: learning beyond the predefined label space. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, USA: ACM, 2011. 1343−1351

19 Du C, Zhuang F, He J, He Q, Long G. Learning beyond predefined label space via bayesian nonparametric topic modelling. In: Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham, Riva del Garda, Italy: Springer, 2016. 148−164

20 Zhuang F Z, Luo P, Shen Z Y, He Q, Xiong Y H, Shi Z Z. D-LDA: a topic modeling approach without constraint generation for semi-defined classification. In: Proceedings of the 2010 IEEE International Conference on Data Mining. Sydney, Australia: IEEE, 2010. 709−718

21 Larochelle H, Erhan D, Bengio Y. Zero-data learning of new tasks. In: Proceedings of the 23rd AAAI Conference on Artificial Intelligence. Chicago, USA: AAAI, 2013. 646−651

22 Socher R, Ganjoo M, Sridhar H, Bastani O, Manning C D, Ng A Y. Zero-shot learning through cross-modal transfer. In: Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA: MIT Press, 2013. 935−943

23 Romera-Paredes B, Torr P H S. An embarrassingly simple approach to zero-shot learning. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: ACM, 2015. 2152−2161

24 Qiao R Z, Liu L Q, Shen C H, van den Hengel A. Less is more: zero-shot learning from online textual documents with noise suppression. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 2249−2257

25 Dietterich T G, Bakiri G. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 1994, **2**: 263−286

26 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York: Springer, 2001.

27 Sloman S A. Feature-based induction. *Cognitive Psychology*, 1993, **25**(2): 231−280

28 Osherson D, Smith E E, Myers T S, Shafir E, Stob M. Extrapolating human probability judgment. *Theory & Decision*, 1994, **36**(2): 103−129

29 Ferrari V, Zisserman A. Learning visual attributes. In: Proceedings of the 21st Annual Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: Curran Associates Inc., 2007. 433−440

30 van de Weijer J, Schmid C, Verbeek J. Learning color names from real-world images. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, Minnesota, USA: IEEE, 2007. 1−8

31 Yanai K, Barnard K. Image region entropy: a measure of "visualness" of web images associated with one concept. In: Proceedings of the 13th Annual ACM International Conference on Multimedia. New York, USA: ACM, 2005. 419−422

32 Farhadi A, Endres I, Hoiem D, Forsyth D. Describing objects by their attributes. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). 2009. Miami Beach, USA: IEEE, 2009. 1778−1785

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

34 Suzuki M, Sato H, Oyama S, Kurihara M. Transfer learning based on the observation probability of each attribute. In: Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics. San Diego, USA: IEEE, 2014. 3627−3631

35 Kovashka A, Parikh D, Grauman K. WhittleSearch: image search with relative attribute feedback. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 2973−2980

36 Parkash A, Parikh D. Attributes for classifier feedback. In: Proceedings of the European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012. 354−368

37 Kulkarni G, Premraj V, Dhar S, Li S M, Choi Y J, Berg A C, et al. Baby talk: understanding and generating simple image descriptions. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE, 2011. 1601−1608

38 Kumar N, Berg A, Belhumeur P N, Nayar S. Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2011, **33**(10): 1962−1977

39 Liu J, Kuipers B, Savarese S. Recognizing human actions by attributes. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE, 2011. 3337−3344

40 Patterson G, Hays J. SUN attribute database: discovering, annotating, and recognizing scene attributes. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 2751−2758

41 Feris R, Siddiquie B, Zhai Y, Petterson J, Brown L, Pankanti S. Attribute-based vehicle search in crowded surveillance videos. In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval. Trento, Italy: ACM, 2011. Article No. 18

42 Fu Y W, Hospedales T M, Xiang T, Fu Z Y, Gong S G. Transductive multi-view embedding for zero-shot recognition and annotation. In: Proceedings of European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 584−599

43 Chao W L, Changpinyo S, Gong B, Sha F. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In: Proceedings of European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 52−68

44 Lazaridou A, Dinu G, Baroni M. Hubness and pollution: delving into cross-space mapping for zero-shot learning. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China: ACL, 2015. 270−280

45 Norouzi M, Mikolov T, Bengio S, Singer Y, Shlens J, Frome A, et al. Zero-shot learning by convex combination of semantic embeddings. arXiv preprint arXiv: 1312.5650, 2013

46 Changpinyo S, Chao W L, Gong B Q, Sha F. Synthesized classifiers for zero-shot learning. In: Proceedings of the 2016 IEEE Conference on Computer vision and pattern recognition. Las Vegas, USA: IEEE, 2016. 5327−5336

47 Radovanović M, Nanopoulos A, Ivanović M. Hubs in space: popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research*, 2010, **11**: 2487−2531

48 Radovanović M, Nanopoulos A, Ivanović M. On the existence of obstinate results in vector space models. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Geneva, Switzerland: ACM, 2010. 186−193

49 Dinu G, Lazaridou A, Baroni M. Improving zero-shot learning by mitigating the hubness problem. arXiv preprint arXiv: 1412.6568, 2014

50 Kodirov E, Xiang T, Fu Z Y, Gong S G. Unsupervised domain adaptation for zero-shot learning. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 2452−2460

51 Harris Z S. Distributional structure. *Word*, 1954, **10**(2−3): 146−162

52 Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe, USA: MIT Press, 2013. 3111−3119

53 Pennington J, Socher R, Manning C D. GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: ACL, 2014. 1532−1543

54 Blanchard E, Harzallah M, Briand H, Kuntz P. A typology of ontology-based semantic measures. In: EMOI-INTEROP. Portugal: Springer, 2005. 160

55 Frome A, Corrado G S, Shlens J, Bengio S, Dean J, Ranzato M, et al. DeViSE: a deep visual-semantic embedding model. In: Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe, USA: MIT Press, 2013. 2121−2129

56 Akata Z, Perronnin F, Harchaoui Z, Schmid C. Label-embedding for attribute-based classification. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE, 2013. 819−826

57 Akata Z, Reed S, Walter D, Lee H, Schiele B. Evaluation of output embeddings for fine-grained image classification. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA: IEEE, 2015. 2927−2936

58 Reed S, Akata Z, Lee H, Schiele B. Learning deep representations of fine-grained visual descriptions. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 49−58

59 Kodirov E, Xiang T, Gong S G. Semantic autoencoder for zero-shot learning. arXiv preprint arXiv: 1704.08345, 2017

60 Bucher M, Herbin S, Jurie F. Improving semantic embedding consistency by metric learning for zero-shot classiffication. In: Proceedings of European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 730−746

61 Zhang Z M, Saligrama V. Zero-shot learning via joint latent similarity embedding. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 6034−6042

62 Xian Y Q, Akata Z, Sharma G, Nguyen Q, Hein M, Schiele B. Latent embeddings for zero-shot classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 69−77

63 Jayaraman D, Grauman K. Zero-shot recognition with unreliable attributes. In: Proceedings of the International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014. 3464−3472

64 Zhang Z M, Saligrama V. Zero-shot learning via semantic similarity embedding. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE 2015. 4166−4174

65 Zhao B, Wu B T, Wu T F, Wang Y Z. Zero-shot learning posed as a missing data problem. arXiv preprint arXiv: 1612.00560, 2016

66 Zhang Z M, Saligrama V. Zero-shot recognition via structured prediction. In: European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 533−548

67 Wang D H, Li Y, Lin Y T, Zhuang Y T. Relational knowledge transfer for zero-shot learning. In: Proceedings of the 13th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI, 2016. 2145−2151

68 Luo C Z, Li Z T, Huang K Z, Feng J S, Wang M. Zero-shot learning via attribute regression and class prototype rectification. *IEEE Transactions on Image Processing*, 2018, **27**(2): 637−648

69 Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. In: Advances in Neural Information Processing Systems 27. Montreal, Canada: MIT Press, 2014. 1−9

70 Lu Y. Unsupervised learning on neural network outputs: with application in zero-shot learning. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA: AAAI, 2016. 3432−3438

71 Baldi P, Hornik K. Neural networks and principal component analysis: learning from examples without local minima. *Neural Networks*, 1989, **2**(1): 53−58

72 Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 1999, **10**(3): 626−634

73 Deng J, Dong W, Socher R, Li L J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami Beach, USA: IEEE, 2009. 248−255

74 Song J, Shen C C, Yang Y Z, Liu Y, Song M L. Transductive unbiased embedding for zero-shot learning. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1024−1033

75 Guo Y C, Ding G G, Jin X M, Wang J M. Transductive zero-shot recognition via shared model space learning. In: Proceedings of the 13th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI Press, 2016. 3434−3500

76 Xian Y Q, Lampert C H, Schiele B, Akata Z. Zero-shot learning-A comprehensive evaluation of the good, the bad and the ugly. arXiv preprint arXiv: 1707.00600, 2017

77 Wah C, Branson S, Welinder P, Perona P, Belongie S. The caltech-UCSD birds-200-2011 dataset, Computation & Neural Systems Technical Report, California Institute of Technology, Pasadena, CA, 2011.

78 Xiao J X, Hays J, Ehinger K A, Oliva A, Torralba A. Sun database: large-scale scene recognition from abbey to zoo. In: Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 3485−3492

79 Antol S, Zitnick C L, Parikh D. Zero-shot learning via visual abstraction. In: Proceedings of the European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 401−416

80 Robyns P, Marin E, Lamotte W, Quax P, Singelée D, Preneel B. Physical-layer fingerprinting of LoRa devices using supervised and zero-shot learning. In: Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks. Boston, Massachusetts: ACM, 2017. 58−63

81 Yang Y, Luo Y D, Chen W L, Shen F M, Shao J, Shen H T. Zero-shot hashing via transferring supervised knowledge. In: Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, The Netherlands: ACM, 2016. 1286−1295

82 Johnson M, Schuster M, Le Q V, Krikun M, Wu Y H, Chen Z F, et al. Google′s multilingual neural machine translation system: enabling zero-shot translation. arXiv preprint arXiv: 1611.04558, 2016.

83 Veeranna S P, Nam J, Mencía E L, Furnkranz J. Using semantic similarity for multi-label zero-shot classification of text documents. In: Proceeding of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium: Elsevier, 2016. 423−428

张鲁宁 中国石油大学 (北京) 自动化系博士研究生. 2016 年获得中国石油大学 (北京) 自动化系学士学位. 主要研究方向为零样本学习与点过程学习.
E-mail: zhang.luning@163.com
(**ZHANG Lu-Ning** Ph. D. candidate in the Department of Automation, China University of Petroleum (Beijing). He received his bachelor degree from the Department of Automation, China University of Petroleum (Beijing) in 2016. His research interest covers zero-shot learning and point-process learning.)

左 信 中国石油大学 (北京) 自动化系教授. 主要研究方向为智能控制, 安全仪表系统的分析和设计, 先进过程控制. 本文通信作者.
E-mail: zuox@cup.edu.cn
(**ZUO Xin** Professor in the Department of Automation, China University of Petroleum (Beijing). His research interest covers intelligent control, analysis and design of safety instumented system, and advanced process control. Corresponding author of this paper.)

刘建伟 中国石油大学 (北京) 自动化系副研究员. 主要研究方向为模式识别与智能系统, 先进控制.
E-mail: liujw@cup.edu.cn
(**LIU Jian-Wei** Associate researcher in the Department of Automation, China University of Petroleum (Beijing). His research interest covers pattern recognition and intelligent system, and advanced control.)