**姓名:刘鑫杰 班级:04011806**

**Zero-shot Learning**

***The first week***

**前言**:

选择这个方向后,自己也比较迷茫,不知从何下手.老师说用MATLAB编程,于是自己学了两天的MATLAB的基础语法.然后从老师和自己在网上收集的一些资料后,大概对ZSL有了初步的了解,不过脑中还是有很多疑问,毕竟还没有什么经验,网上资料太杂,而且自己也还没有看相关的文献资料,理论基础还不完全.之前一直想去复现一下别人的代码,也出现了很多问题.但反思一下自己连自己在做什么都不知道,谈什么复现.于是又去把线头找到,从最基本的做起.

**目录:**

**1.ZSL的背景和概念**

**2.ZSL的理解和实现思路**

**3.ZSL需要解决的问题**

**4.ZSL常用数据集介绍**

**5.数据集处理**

**6.ZSL基础算法介绍**

**7.ZSL中存在的问题以及解决方法**

**8.有关ZSL的一些其他的概念**

**9.总结**

**1.ZSL的背景和概念及应用场景**

1.1背景

早期的零样本学习研究可以追溯到2008 年，Larochelle 等人针对字符分类问题提出了零样本学习方法，并且识别准确率达到了60%。2009年Lampert 等人提出了Animals with Attributes(简称AWA数据集)数据集和经典的基于属性学习的算法，才真正打开零样本学习的关注度。

在传统图像识别任务中，训练阶段和测试阶段的类别是相同的，但每次为了识别新类别的样本需要在训练集中加入这种类别的数据。一些类别的样本收集代价大，即使收集到足够的训练样本，也需要对整个模型进行重新训练。这都会加大识别系统的成本，零样本学习方法便能很好的解决这个问题。

1.2概念

传统的“零样本学习”方法首先是让智能体（Agent）对类别进行语义理解。将类别标签利用辅助知识（如属性）嵌入到语义空间中，再利用训练集中的数据学习这种从图像到语义的映射关系。此后，即使遇到新的类别，只要提供了该类别的语义知识，模型即可识别该类别，这就是零样本学习。即模型对其从没有见过的类别进行分类,让机器具备一定的推理能力.但是要求就是对于要分类的类别对象一次也不学习.

定义:利用训练集数据训练模型，使得模型能够对测试集的对象进行分类，但是训练集类别和测试集类别之间没有交集；期间需要借助类别的描述，来建立训练集和测试集之间的联系，从而使得模型有效。

1.3应用场景

未知物体识别,例如由马,老虎,熊猫🡪斑马

未知语言翻译——比如说要进行三种语言之间的翻译，按照传统的方法需要分别训练六个网络，在日语和韩语之间没有那么多样本的情况下，训练英语→特征空间→日语，韩语→特征空间→英语这两个网络，那么就可以自动学会韩语→特征空间→日语这个翻译过程。

未知类别图像合成

图像哈希:图像的自动标注、处理

视频识别,即挖掘视频,音频,字幕,弹幕,评论等多模态信息之间的相关性.

**2.ZSL的理解及实现思路**

2.1理解

对于要分类对象,一次也不学习.形象解释:马的轮廓+老虎的条纹+熊猫黑白色🡪斑马

2.2实现思路(以马,老虎,熊猫🡪斑马为例)

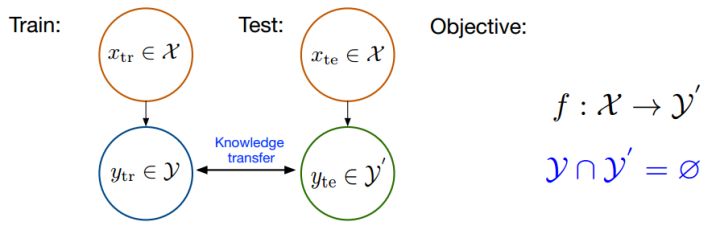
训练集数据X\_tr(马,老虎,熊猫图片) 训练集标签Y\_tr(马,老虎,熊猫标签)

测试集数据X\_te(斑马图片) 测试集标签Y\_te(斑马标签)

训练集类别描述A\_tr 测试集类别描述A\_te

我们将每一个类别 y\_i∈ Y ，都表示成一个语义向量 a\_i∈ A 的形式，而这个语义向量的每一个维度都表示一种高级的属性，比如“黑白色”、“有尾巴”、“有羽毛”等等，当这个类别包含这种属性时，那在其维度上被设置为非零值。对于一个数据集来说，语义向量的维度是固定的，它包含了能够较充分描述数据集中类别的属性。

在ZSL中，我们希望利用X\_tr和Y\_tr来训练模型，而模型能具有识别X\_te的能力，因此模型需要知道所有类别的描述A\_tr和A\_te



**3.ZSL需要解决的问题**

3.1类别的描述A如何获取?

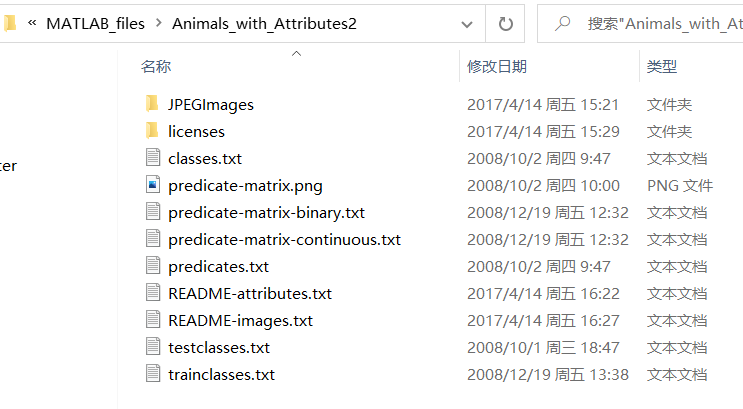
通过人工专家定义的, 通过海量的附加数据集自动学习出来的，但前者的效果目前要好很多。

3.2如何建立一个合适的分类模型?

目前大部分工作都集中在这个问题上,而第一问题的研究进展缓慢. 因为A的获取主要集中于一些NLP的方法,难度比较大;第二个问题能用的方法比较多,容易出成果.之后介绍的算法部分也围绕这个问题展开.

**4.ZSL常用数据集介绍(主要介绍AWA2数据集, 其他暂略)**

*4.1Animal with Attributes(AWA2)*:都是动物的图片,包括50个类别,其中40个类别作为训练集,10个类别作为测试集,每个类别的语义为85维,总共有30322张图片.



文件介绍:

JPEGImages：动物图片，子文件夹名为动物名称

licenses：包含每一张图片的授权,是些用处不大的文件

classes.txt：动物列表，顺序很重要！！！

predicates.txt：属性列表，共85种属性, 顺序很重要！！！

predicate-matrix-binary.txt：该文件记录了50种动物，每一种动物的85种属性特征情况，是一个50x85的矩阵，1表示有该特征，0表示无

class.txt文件，每列对应的属性参考predicates.txt文件，二值属性

predicate-matrix-continuous.txt：每行对应每个动物的属性，每行对应的动物类型参考classes.txt文件，每列对应的属性参考predicates.txt文件，连续属性，用于后续的unseen样本的分类，值为每个属性的置信度，可能在实际应用中需要归一化,属性描述为连续数字

testclasses.txt：用于测试的类别, 共10个类别

trainclasses.txt：用于训练的类别, 共40个类别

predict-matrix.png: 文件predicate-matrix-binary.txt的图形化

README-attributes.txt和README-images.txt: 这两个说明文件对我们也是没有用的

二值属性:有与没有

连续属性:属性存在的一个分值,可以理解为当前属性对于动物类别判定的一个重要程度,比如斑马,有没有花纹属性相对于有没有毛发属性,要重要的多.

4.2 Caltech-UCSD-Birds-200-2011（CUB）

4.3 Sun database（SUN）

4.4 Attribute Pascal and Yahoo dataset（aPY）

4.5 ILSVRC2012/ILSVRC2010（ImNet-2）🡪大型数据集

**5.数据集处理**

5.1图片读取

5.2准备属性标签

5.3使用预训练的resnet101提取图片特征

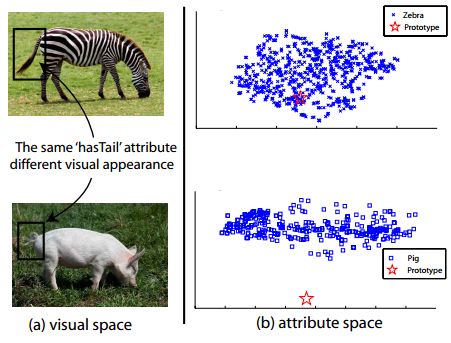
**6.ZSL基础算法介绍**

DAP和IAP

**7.ZSL中存在的问题以及解决方法**

7.1领域漂移问题（domain shift problem）

简单来说，就是同一种属性，在不同的类别中，视觉特征的表现可能很大。如下图所示，斑马和猪都有尾巴，因此在它的类别语义表示中，“有尾巴”这一项都是非0值，但是两者尾巴的视觉特征却相差很远。如果斑马是训练集，而猪是测试集，那么利用斑马训练出来的模型，则很难正确地对猪进行分类。

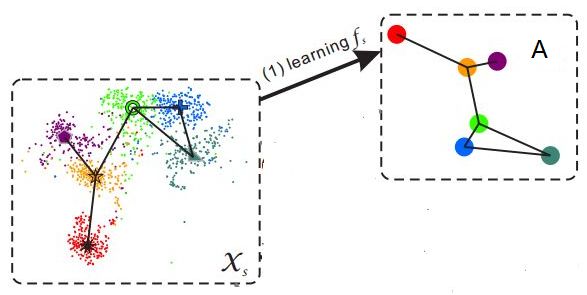


7.2枢纽点问题（Hubness problem）

这其实是高维空间中固有的问题：在高维空间中，某些点会成为大多数点的最近邻点。由于ZSL在计算最终的正确率时，使用的是K-NN，所以会受到hubness problem的影响，有人也证明了基于岭回归的方法会加重hubness problem问题。

7.3语义间隔（semantic gap）

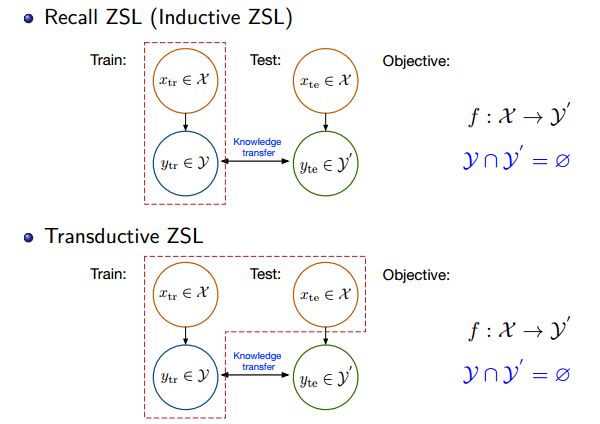
样本的特征往往是视觉特征，比如用深度网络提取到的特征，而语义表示却是非视觉的，这直接反应到数据上其实就是：样本在特征空间中所构成的流型与语义空间中类别构成的流型是不一致的。



**8.有关ZSL的一些其他的概念**

8.1 直推式学习(transductive setting)

这里的直推式学习其实是指在训练模型的时候，我们可以拿到测试集的数据，只是不能拿到测试集的样本的标签，因此我们可以利用测试集数据，得到一些测试集类别的先验知识。这种设置在迁移学习中很常见。



8.2 泛化的ZSL（generalized ZSL）

上文中提到的ZSL，在测试时使用K-NN进行正确率的评估时，只在测试类别中找最近邻的类别，但是在现实的问题中，拿到的样本也可能属于训练集类别，因此在测试时，同时加入训练集类别。现在的很多方法都开始测试模型在这种设置下的能力。

**9.总结**

第一周在网上看了一些资料,大概对ZSL有了初步的了解, 后面看会一些相关的论文. 从最早到最近的, 加深自己对ZSL的发展有一定理解.

zero-shot learning的一个重要理论基础就是利用高维语义特征代替样本的低维特征，使得训练出来的模型具有迁移性。语义向量就是高维语义特征，比如一个物体的高维语义为“四条腿，有尾巴，会汪汪叫，宠物的一种”，那我们就可以判断它是狗，高维语义对它没有细节描述，但是能够很好的对其分类，分类是我们的目的，所以可以舍去低维特征，不需要“全面”。

（1）Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer

ZSL问题的开创性文章，当然是必读的喽，而且可以顺便看看别人是如何阐述一个新问题（挖坑）的。

（2）An embarrassingly simple approach to zero-shot learning

有着很强的理论基础，算法简单、有效，虽然已经过去很多年了，但还是目前新工作需要进行对比的方法之一。

（3）Transductive Multi-View Zero-Shot Learning

第一次定义了domain shift问题。

（4）Zero-shot recognition using dual visualsemantic mapping paths

解决semantic gap问题的简单做法。

（5）Predicting visual exemplars of unseen classes for zero-shot learning

从本质的角度出发，将ZSL问题，看作聚类问题，用最简单的方法直接建立映射。

（6）Semantic Autoencoder for Zero-Shot Learning

引入自编码器结构的第一篇文章，直接导致现在出现的新方法大都具有这种结构。

（7）Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly

综述性的文章，总结了17年底以前的方法，提出了新的评价标准，对当时领域发展比较混乱的地方做出了一些更标准的评估。

（8）Zero-Shot Learning via Class-Conditioned Deep Generative Models

将改造为深度模型，并加上一些其他的约束。

（9）Preserving Semantic Relations for Zero-Shot Learning

在自编码器结构的基础上，显示地加入语义类别之间的关系约束。

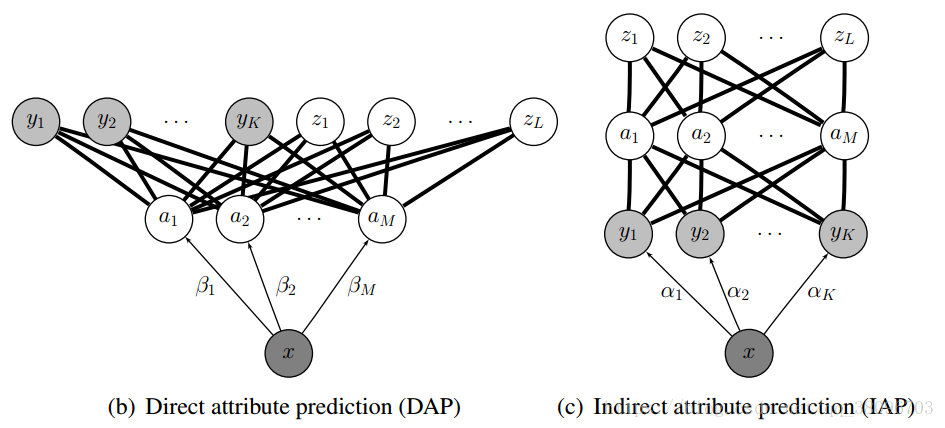
（10）Recent Advances in Zero-shot Recognition

《Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer》论文阅读报告

1.前言：本paper为ZSL的开山之作，主要介绍了一个问题：检测未在训练集标签中出现的类标签的实例检测，即ZSL，同时介绍了两种算法DAP以及IAP。

2.定义：x为实例集合，y为可见的标签集合（训练集），z为不可见的标签集合（测试集），y与z的交为空。a为属性向量，M维，每个维度代表一个类的一种属性且为{0，1}取值，对于每个标签，皆对应一个M维向量a作为其的属性向量

3.传统的多分类器无法胜任ZSL（因为其仅仅建立可见标签集与实例的联系），故介绍两种新算法来解决，DAP和IAP都属于基于类属性的分类，通过加入人类先验高级描述，达到零样本学习的目的，而有别于传统的分类方法。DAP与IAP的区别在于训练类与测试类之间的关系。在DAP中，基于属性层中的属性结果，对所有类一视同仁。相反，通过间接学习IAP中的属性，可以从作为中间特征层的训练类中归纳出属性的值。



3.1 DAP (directed attribute prediction)

（1）添加了一个属性层在图像与类标签中用于解耦, DAP可以理解为一个三层模型：第一层是原始输入层，例如一张电子图片（可以用像素的方式进行描述）；第二层是p维特征空间，每一维代表一个特征（例如是否有尾巴、是否有毛等等）；第三层是输出层，输出模型对输出样本的类别判断。在第一层和第二层中间，训练p个分类器，用于对一张图片判断是否符合p维特征空间各个维度所对应的特征；在第二层和第三层间，有一个语料知识库，用于保存p维特征空间和输出y的对应关系。

简单来讲，就是对输入的每一个属性训练一个分类器(每个分类器判别该样本是否有该属性,有为1,无为0)，然后将训练得出的模型用于属性的预测，测试时，对测试样本的属性进行预测，再从属性向量空间里面找到和测试样本最接近的类别。

（2）自己通俗的理解：以AWA2数据集为例, x表示训练集的图片，经过数据预处理之后，每张图片变成了1x85的矩阵A，每列位置上的数是连续的，通过每个属性分类器将其变为01，每列位置上对应0（无）或者1（有）表示属性的有无，即对应于上图的a1~am,然后属性层和标签层之间有语料知识库🡪这是一个50x85的矩阵B，每一行代表一类动物，每一列代表一种属性，每个位置的值表示属性的有无或者对评估的重要性。通过在训练集上训练可以得到一个模型，这个模型可以通过输入的矩阵A对训练样本的类别进行预测（基于遍历矩阵B进行对比选择近邻），从而对未知类别Z进行预测（同样基于矩阵B）。

（3）形象举例：假设我们已经训练好了一个DAP模型，第一层和第二层间的分类器可以判断 是否黑眼圈、是否喜欢吃竹子 之类的特征，然后在语料知识库里面包含一个映射：黑眼圈 喜欢吃竹子--> 熊猫，那么即使我们的模型在训练时没有见过熊猫的图片，在遇到熊猫的图片时，我们可以直接通过对图片的特征进行分析，然后结合知识语料库判断出这张图片是熊猫。假设即使语料知识库里面不包含 黑眼圈 喜欢吃竹子--> 熊猫 的映射，我们也可以通过计算熊猫图片的特征与其他训练样本的特征的汉明距离度量，得到熊猫和什么动物比较类似的信息。整个DAP的运作思想就是类似于上述过程。

（4）数学推导（后面再补充）

（5）DAP的缺点：

1.无法克服邻域漂移问题；

2.训练很多学习器也是一件费时的事；

3.无法利用额外的属性信息；

4.无法利用新的样本逐步改善分类器的功能；

5.算法引入了中间层，核心在于尽可能得判定好每幅图像所对应的特征，而不是直接去预测出类别；因此DAP模型在判定属性时可能会做得很好，但是在预测类别时却不一定；

3.2 IAP (indirect attribute prediction)

这一方法使用了两层标签,属性层作为中间层，但是属性在两层标签之间形成了一个连接层，一层用于训练时已知的类，另一层用于训练时未知的类。IAP的训练阶段是普通的多类分类。在测试时，对所有训练类的预测都会引导属性层的标记，从而可以推断出测试类上的标记。

（后面补充）