# NEIL: Extracting Visual Knowledge from Web Data学习笔记

## 概述

NEIL的目的是自动地从互联网图片中提取语义信息，以这些信息（带标签的实体和常识关系）扩充知识数据库，利用这些数据帮助建立更好的分类器和检测器。

#### 提取的信息

* 带有盒边界的对象类别的标签样本
* 带标记的场景
* 带标记的样本属性
* 对象类别的可视子类
* 关于场景、对象、属性的常识关系

## 技术

#### 通过Google Image Search播种分类器

半监督学习建立视觉分类分类器。建立初始分类器的一种方法是通过少量手动标记的图片训练。

这里作者选择一种替代方法。作者使用基于文字的图像检索引擎来索引提供最初的种子图像，训练最初的检测器，对于场景和属性分类器，可以直接让搜索引擎的图片作为正数据。

然而对于目标和属性检测，这个方法不奏效，原因有四：

1. Outliers异常值：搜索引擎的结果可能存在一些不相关的内容。
2. Polysemy歧义性：同一个词可能具有不同的意义，比如Apple既可以指水果也可以指苹果公司。
3. Visual Diversity视觉差异：搜索引擎的图片可能因为视角、光照等原因具有较大的类内变化。
4. Localization局部化：搜索的图片往往是一个没有边界框的场景，为了训练检测器，需要人为添加局部的边界框。

通过聚类算法（K-means），可以有效解决歧义、视觉差异和一些异常值。不过K-means存在两个问题：

1. 高维度：作者使用Color HOG后得到了高维度的表示，而高维度并不适合用欧式距离衡量差异

伸缩性：聚类算法倾向于划分完整的特征空间，由于没有目标的边界线，所以图像中的背景像素在举类时都是噪音像素。

作者提出了两步策略进行聚类：

1. 从搜索引擎下载的图片中建立候选目标窗口，每张图片通过exemplar-LDA训练检测器，这些检测器用于对同一集合的图片做目标检测，选择前个具有较高分数的窗口。这种方法有效避免了图片中无关内容的干扰。
2. 作者利用每个窗口的检测签名（每个窗口表示为一个在窗口上的ELDA分数种子向量）建立一个的affinity matrix，每一个是窗口的向量和窗口向量的点积。随后使用AP聚类算法进行聚类操作，AP聚类还能从图像中提取出针对某一类别的符号化的表示。

聚类后，对每一类别，使用其中的图像训练检测器，剩余用于验证。

#### 提取关系

在得到了目标、属性检测器和数学、场景分类器后，我们可以自动地从数据中提取关系。

###### Object-Object关系

形如眼睛是身体的一部分、BMW 320是一款汽车、天鹅和鹅看起来很像。

为此，我们构建一个co-detection矩阵，其中的元素表示类别被同时检测到的概率，更准确一些，当检测器的边框盒在检测器的内部，那么矩阵对应元素具有较高的值。

获取矩阵后，对于任意两类别之间的关系，可以通过它们在相对位置、相对长宽比、相对分数、相对尺寸的均值和方差信息学习关系的性质。用函数评估两个类别的检测匹配程度。

关系被分成两个语义类别（part of, taxonomy/similar）。

###### Object-Attribute关系

形如香蕉是黄色的，披萨是圆形的。

属性通过样本的标签分类表示，其原理和上文一样，通过利用检测得分构建co-detection矩阵，以寻找最佳的属性对象关系。

###### Scene-Object关系

形如公交车在公交站台、监视器在控制室。

通过在随机选取场景的图片中做目标检测，构建co-presence矩阵，其中表示在场景中出现物体的似然值。

###### Scene-Attribute关系

形如海洋是蓝色的、小径是狭窄的。

计算co-classification矩阵，其中表示属性出现在场景中的分类得分，最高得分元组被提取为一条关系。

#### 通过标记新的实例以重训练模型

一开始我们初始了一系列分类器和检测器，通过他们寻找关于已有对象和类别的新实例，这一系列实例可以添加到标记数据集中，用于重训练新的分类器和检测器，进而提取更多的关系以获得更多的数据。

直接使用这一循环可能引发semantic drift（语义偏移，指词的含义随历史发展而产生变化），为了避免语义偏移，使用已有的关系约束找到的实例，换言之实例需要满足关系库约束（比如车必须包含轮子，所处的场景是车道）。

用表示object-object, object-attribute and scene-object三种关系的集合。

表示第个检测器的潜力，表示场景潜力，评估对象分类的匹配程度。

对于一个新实例其属于对象类别的评分可以有以下公式得出

其中是被评估的窗口，是object-attribute关系的顶级窗口。

第一项评估的是实例分类的分数，用SVM检测器在窗口上评估。第二项是评估实例所具有的两种关系的评分，就是关系中另一方的顶级窗口。第三项是对对象所处场景的评分。

对于新场景实例(属于分类)的评分可以用以下函数

第一项是对场景本身的评估项使用场景分类器。第二项是scene-attribute关系，对于和有关的关系对应的属性，使用属性分类器进行评估。第三项则利用scene-object关系，对场景对应的对象用相应的检测器进行检测。

疑问：如何从AP聚类后的簇中知道哪些是object\attribute\scene