# 题目 1: 交通拥堵预测

## 背景

交通拥挤是交通运输中最严重的问题之一。尽可能早地预测道路拥挤程度是有 价值的，这样司机和行人就可以通过预报来规避拥堵。

## 数据描述

The GCM (Gary-Chicago-Milwaukie) Corridor 包含 16 座城镇之间的所有主干 道。在所有路口共放置 855 个传感器，来收集一天中的交通数据流，其格式为每

5 分钟记录一个拥堵状态，一天有 288 条数据。每个传感器都实时收集交通状 态，并通过无线通讯将路口位置和拥堵信息传输到中央服务器。每一条流数据包 含下列属性（date, time, direction, type, linkID, length, travelTime, volumn, speed, occupancy, and congestionLevel）。拥堵状态共有四种，non, light, medium, and heavy（通畅、轻微拥堵、中度拥堵、重度拥堵）。以下是一条数 据流的样例:

0707,0000,NORTH\_BOUND,FREEWAY,WI-MNT\_XML\_V001-

21012,1268,40,218,31.292915,2.4,NON\_CONGESTION

## 问题描述

我们下载了 5 天的交通数据，其中 4 天数据给大家作为模型训练。请基于训练 集建立模型来预测未来的交通拥堵状态。

对数据进行预处理操作是必要的。可以使用各种数据挖掘算法和机器学习方法 建立预测模型。

评估:

**提交模型预测**。将第 5 天的数据作为测试集进行模型评估。助教将提供几个时

间区段内所有传感器收集的实际的交通拥堵状态。请预测接下来 30 分钟内所有传感器的拥堵状态。评估指标为全部30分钟时间段的准确率。请提交实验报告来阐述你的实验方案、方法和性能分析。

为了便于测试，拥堵预测的输出格式如下：

WI-MNT\_XML\_V001-21012(传感器 ID):0,1,2,3,3,2(0 表示通畅，1 表示轻微拥

堵、2 表示中度拥堵、3 表示重度拥堵，连续 6 个数字代表 30 分钟内的拥堵状 态)

训练集、测试集网盘地址：

<https://pan.baidu.com/s/1GcIzkOylHtMYMdOVilZeCg>

提取码：6xz8

请加入对应qq群：816888547

# 题目 2: 领域术语抽取

背景：

领域术语是指特定领域的理论概念词汇，是在特定学科领域用来表示概念的称谓的集合，人们可以通过术语来快速了解学科的发展，这对科学研究具有重要意义。然而，从海量的文本集合中手工提取领域术语是很费时费力的，而一些传统的分词工具也不能很好地将术语进行分词。因此，在不同的领域中，领域术语的自动提取是很有必要的。

## 数据描述

提供管理领域的书籍文本txt格式，以及每本书对应的由专家总结的专业术语列表。

例如，《人力资源管理概论》的报酬形式、变化幅度、标杆基准法等都是领域专业术语。

## 问题描述

我们提供了 三本书的数据，以及对应的专业术语列表。请利用这些数据建立术语抽取模型。建模方法不限，可以使用传统语言学方法，也可以使用深度学习方法等。

对文本数据进行预处理操作是必要的。可以使用各种文本处理的开源工具包进行分句分词等预处理。

评估:

**提交模型预测**。助教提供一本新的“管理”书籍，运行模型后，生成这本书中的领域术语列表。

请提交实验报告来阐述你的实验方案、方法和性能分析。为了便于测试，输出的格式应和提供的术语列表一样，生成txt文件，每一行代表一个专业术语。

评估指标为模型抽取术语和专家术语的准确率和召回率。

训练文本网盘地址：

<https://pan.baidu.com/s/1TluXPbLQ5dDdRj21-XJylA>

提取码：4a1q

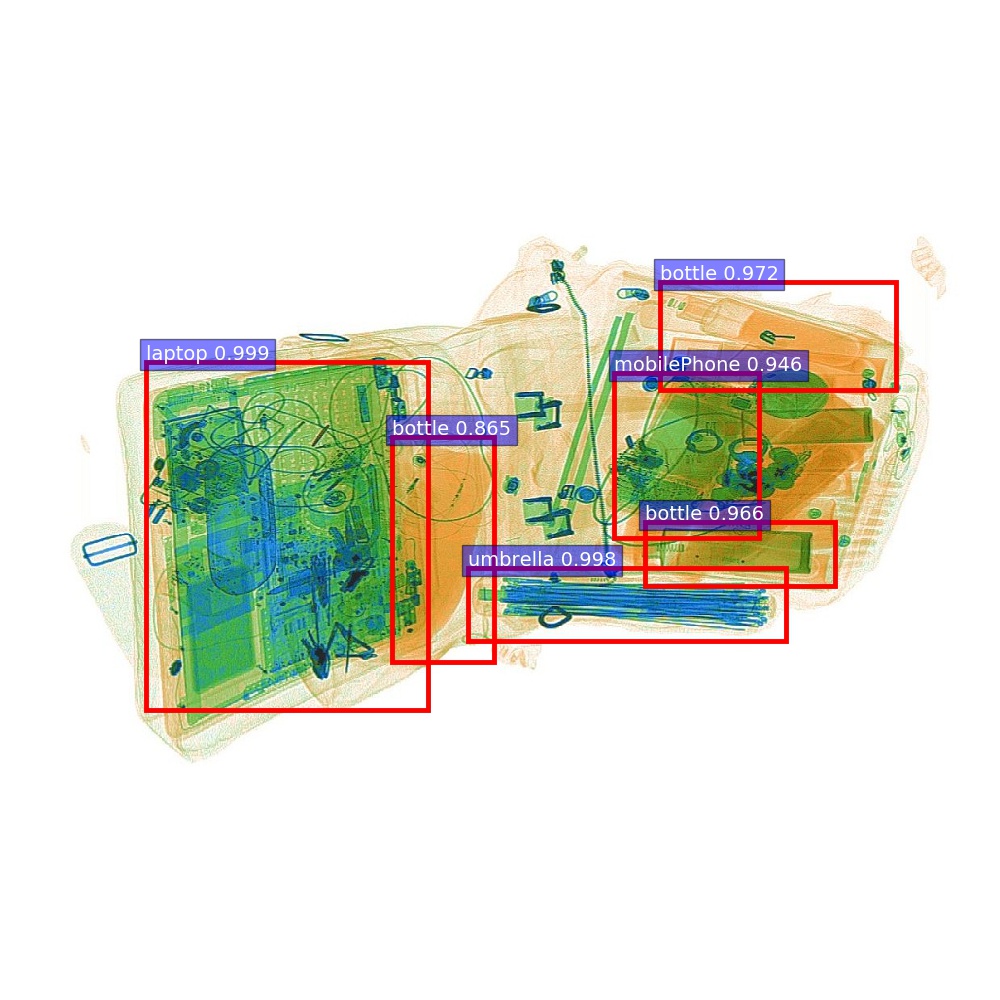
请加入对应qq群：878563259

题目 3: 安检机 X 光图像物体检测识别

* 任务:

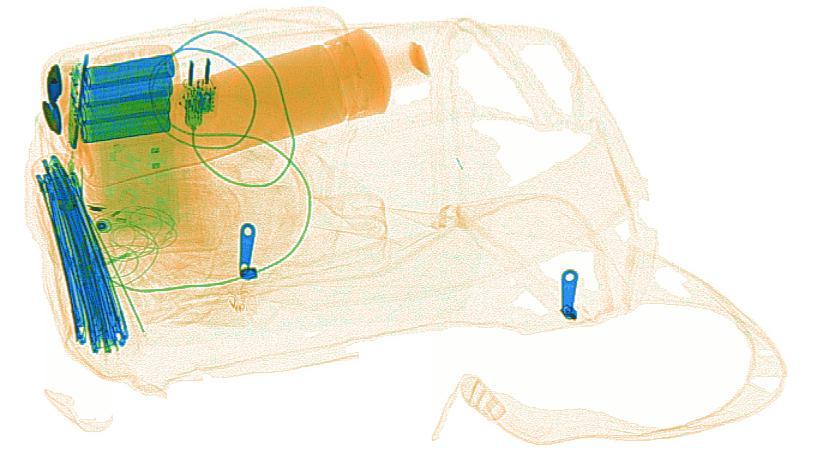
检测出x光图像中的电脑、充电宝、雨伞、瓶子、手机和 钥匙共 6 类物体的检测与识别。参考方法：SSD, YOLO, Mask R-cnn, CornerNet和centerNet等

检测结果如下：



* 数据介绍:

本题提供的图片拍摄于真实安检机。图片中包括电脑（laptop）、充电宝(power bank)、雨伞(umbrella)、瓶子(bottle)、 手机(mobilePhone)和钥匙(key)共 6 类物体。示例图如下：



* 训练集: 共计3000张带标注文件的图片，数据集格式为VOC格式
* 测试集: 共计1000张带标注文件的图片，数据集格式为VOC格式。
* 评价方法：mean Average precision (mAP)
* 得分评估

1. 项目报告。算法描述、实现方法及和在测试集上的实验结果分析，格式为一般论文格式，中英文均可。（0~20）
2. 现场demo展示。助教提供10张未标注图片，现场检测结果并可视化展示。自行携带笔记本, 运行检测程序,展示异物检测结果。 （0~5）
3. 创新性。在现有流行模型的基础上，提出改进策略。（0~5）

项目交流和数据下载:

