**问题的提出：**

我们先了解一下机器学习，或者说数据分析中最基本的步骤：

1. **探索性分析**：举个例子，我们应该都用过excel，通过excel中的一些图表，例如直方图，折线图，饼图，可以反应数据背后的内容，让数据自身“说话”。

**2）**如果探索性分析效果很差的话，很可能数据本身就存在问题。因为如果我们用的是大数据的话，这种”大”不单单是数据量大，还包括种类多，价值密度低，数据增长速度快，以及数据存在可信赖度问题。所以我们需要进行数据清洗。**数据清洗也是个细活**，就好比三星堆黄金面具出土时，需要进行特定的几道工序，我们才能看到它本来的样貌，如果换成我们普通人，保不齐就弄坏了。

**3）**数据清洗完后，由于不同类型的数据可能会存在冗余，因此我们需要进行**特征提取**。举个例子，如果我们知道一个圆的半径，面积和周长，我们当然知道面积和周长相对于半径来说是个冗余特征。你也许觉得“特征”这个概念还是很抽象，但如果把它理解成一个事物的客观属性，我们就好理解了。如果你看到一个穿着黑衣服的移动的物体，他身高1.7，能直立行走，我们就可以大概率判断他是个黑衣人，而不是金刚。

特征提取是一个工程，通过特征工程，我们提取到的特征越来越具代表性。数据中包含特征，特征中包含特征，是一个套娃的关系，就好比分子包含原子，原子包含中子和质子，中子和质子还能继续往下分成夸克…

4）接下来就是各种**算法**的选择，比如分类算法，回归算法，它们分别对应**分类任务和回归任务**。

这两个任务是机器学习中两个基本任务。举个例子，要想让计算机识别出某个物体，分类任务需要对这个物体有个准确的分类，是猫，是狗还是小脑斧，然后有了类别之后，我们需要对这个物体进行定位，指出这个物体在图片或者在三维场景中准确的位置，也就是回归任务。**回归和分类的主要区别就是**，分类的输出结果是由输出空间上的离散取值决定的，输出空间如果只有{1,…,10}，那么它的输出结果就不存在11或0.5；而回归输出的结果是由输出空间上的取值范围决定的，输出值是连续的。分类和回归任务的输出结果需要通过不断优化损失函数得到，这个损失函数是算法中最核心的一个部分之一，目的是衡量当前的预测值和真实值的误差，误差越小，预测越准确。通常情况下，分类和回归的损失函数是不一样的。

5）有了算法，往算法里填数据就变成了模型。完成模型的训练，我们需要额外地对模型好坏进行评估。因为一个模型，或者说算法，它是根据某个实际问题提出来的，因此要想适应这个问题，模型往往包含很多可以调节的参数。不同参数会得到不同的模型，得到不同的效果

顺便说一下，机器学习和数据分析它们侧重点略有不同：前者侧重模型的预测准确性，而后者侧重于模型的可解释性，前者输出的是模型的预测结果，后者输出的是报表等决策信息。对于上面这幅图，把这两个概念看成是一样的并不妨碍我们对这个流程的理解。

这篇论文的切入点就是在算法的优化上，明确本文的定位之后，我们再接着讲下面的内容。

多任务联合学习的特点是不同子任务**共享低层特征**，并通过不同子任务的协同学习进行信息的交互，在相互促进中**完成整个任务**。这里以Faster R-CNN网络为例，输入的是一张图片，输出的结果是判断一张图片中是否包含人脸，以及人脸的位置，即分类和回归任务。这两个任务共同完成对人脸的定位，如果分类结果不包含人脸，就没有后续的人脸定位的操作了，因此这**两个任务是相互促进**的。

以前关于多任务联合学习的总损失函数通常使用**线性加权**的方式进行求解。然而单纯地使用这种方法，却会存在些问题：例如模型对于权重参数很敏感，如该图所示

。

对于（a）图，作者通过调整语义分类和深度回归两个任务的损失函数的权重，分析这两个任务的联合总损失函数在优化过程中的最终**输出误差**，这个误差包括两个部分，一个是语义分类的IoU，而另一个是深度回归的误差。

其实我们不用关心这两个任务究竟用来干什么，说句实话，我也不了解，但是不妨碍我们理解这两幅图。

在图（a）中红色的这一块，表示分类任务，纵轴表示递增的IoU，IoU越大表示误差越小，分类效果越好;

而横轴表示权重的衰减过程，从1衰减到0；

蓝色部分表示回归任务，纵轴表示递减的误差（或者），越往上，误差越小；而横轴表示权重的增加过程，从0增加到1；蓝色这一块可以理解成关于红色这一块的一个镜像。

如果要让全局损失最小，我们发现折中的方法是找到两个曲线相交的位置，但是交点可能会存在多个（比如图b），如果只通过交点位置的权重来计算全局的最小损失，误差还是蛮大的，因为可以看到曲线起伏很大，权重的大小很难把握，模型对于参数存在敏感性问题。

**解决方案**

**实验结果**

如果只进行任务1，IoU损失值为59.4%，如果只进行任务2，误差为4.61，如果只进行任务3，误差为0.64

如果等权重分配给三个损失函数，任务1的IoU损失值为50.1%，任务2的误差为3.79，任务3的误差为0.592，发现任务1受影响很大

如果使用之前的多个损失函数线性加权的方法，计算得到最优的权重分配策略，任务1的IoU损失值为62.8%，任务2的误差为3.61，任务3的误差为0.549。

然后作者对文章中的三个任务进行两两组合，进行不确定性损失加权，效果和线性加权的差不多

实验结果可以看出基于三个任务的不确定性损失加权的效果最好

**线性噪声**：

我们可以发现，线性随机噪声的方差越大，信息的不确定性越大，子任务的损失函数被赋予的权值也就越大。

**结论**

1、在多任务联合学习中，基于不确定性的损失函数加权方法在总体上优于之前的损失函数线性加权的方法

2、由于该算法是基于两个假设的，即回归任务的误差满足正态分布，分类任务的输出服从统计物理学中的玻尔兹曼分布（也就是之前说的函数），因此在实际中如果模型不满足该假设，可能会存在效果变差，性能不稳定的问题。

有没有印象，我在开头的时候介绍了探索性分析，而且**在数据分析流程中还有一个other**？本本文的分析方法采用的是验证性分析，需要在满足该理论下模型效果才会好，这种分析方法就包含在之前的other里。