# 数学建模题型解析举例及软件的推荐和使用

## 常见数学建模题型

### 1.1优化问题

#### 1.1.1 数学规划模型

线性规划、整数线性规划、非线性规划、多目标规划、动态规划。

#### 1.1.2微分方程组模型

阻滞增长模型、SARS传播模型。

#### 1.1.3 图论与网络优化问题

最短路径问题、网络最大流问题、最小费用最大流问题、最小生成树问题(MST)、旅行商问题(TSP)、图的着色问题。

#### 1.1.4 概率模型

决策模型、随机存储模型、随机人口模型、报童问题、Markov链模型。

#### 1.1.5 组合优化经典问题

##### (1) 多维背包问题(MKP)

背包问题： 个物品，对物品 ，体积为 ，背包容量为 。如何将尽可能多的物品装入背包。

多维背包问题： 个物品，对物品 ，价值为 ，体积为 ，背包容量为 。如何选取物品装入背包，是背包中物品的总价值最大。

多维背包问题在实际中的应用有：资源分配、货物装载和存储分配等问题。该问题属于 难问题。

##### (2) 二维指派问题(QAP)

工作指派问题： 个工作可以由 个工人分别完成。工人 完成工作 的时间为 。如何安排使总工作时间最小。

二维指派问题（常以机器布局问题为例）： 台机器要布置在 个地方，机器 与 之间的物流量为 ，位置 与 之间的距离为 ，如何布置使费用最小。

二维指派问题在实际中的应用有：校园建筑物的布局、医院科室的安排、成组技术中加工中心的组成问题等。

##### (3) 旅行商问题(TSP)

旅行商问题：有 个城市，城市 与 之间的距离为 ，找一条经过 个城市的巡回（每个城市经过且只经过一次，最后回到出发点），使得总路程最小。

##### (4) 车辆路径问题(VRP)

车辆路径问题（也称车辆计划）：已知 个客户的位置坐标和货物需求，在可供使用车辆数量及运载能力条件的约束下，每辆车都从起点出发，完成若干客户点的运送任务后再回到起点，要求以最少的车辆数、最小的车辆总行程完成货物的派送任务。

TSP问题是VRP问题的特例。

##### (5) 车间作业调度问题(JSP)

车间调度问题：存在 个工作和 台机器，每个工作由一系列操作组成，操作的执行次序遵循严格的串行顺序，在特定的时间每个操作需要一台特定的机器完成，每台机器在同一时刻不能同时完成不同的工作，同一时刻同一工作的各个操作不能并发执行。如何求得从第一个操作开始到最后一个操作结束的最小时间间隔。

### 1.2分类问题

#### 1.2.1多类分类问题

对于任何一部电影，电影的中央委员会会根据电影的内容颁发证书。例如，如果你看上面的图片，这部电影被评为“UA”(意思是“12岁以下儿童需在父母陪同下观看”)。还有其他类型的证书类，如“A”(仅限于成人)或“U”(不受限制的公开放映)，但可以肯定的是，每部电影只能在这三种类型的证书中进行分类。简而言之，有多个类别，但每个实例只分配一个，因此这些问题被称为多类分类问题。

### ****1.2.2**多标签分类问题**

每部电影都有可能被分成一个或多个不同的类别。所以每个实例都可以使用多个类别进行分配。因此，这些类型的问题被称为多标签分类问题。

### 1.3 回归预测问题

比如预测股票价格问题，就有单因子回归，多因子回归，时间序列回归，截面回归，多是产量，价格相关的问题。

## 数学建模的几种方法

### 对于优化问题

（1.）对于数学相关的优化问题，更多的会用到数值分析的方法（我会直接放弃）

（2.）对于其他几种非数学相关问题，大部分与编程有关，比如旅行商问题用动态规划可以解决，多目标优化，如机场停机位问题，可以使用遗传算法也可以使用粒子群算法，也可以使用图匹配算法。区别在于是否是全局最优，也有可能是局部最优。另外时间复杂度和空间复杂度也是一个问题。

### 2.2 对于分类问题

传统的机器学习方法，如贝叶斯，决策树，支持向量机，包括稍微复杂一点的随机森林，lighGBM集成算法都可以解决分类问题。这个时候就要来比较他们的优劣。比如在什么数据规模下使用某个模型效果较好。同时他们之间的差别有哪些，都需要有所了解。当这些模型都无法解决，或者说是评估太次，可以考虑使用神经网络。

### 2.3对于回归问题

常见的回归模型有线性回归模型，非线性回归模型。线性回归又包括一元线性回归，多元线性回归，广义线性等。非线性回归模型，大部分机器学习分类算法都有其回归形式。这非常适合拟合高维数据。目前最强大的回归方式应该是深度学习。

## 3 数学建模实例分析

实例：泰迪杯数据挖掘大赛-C题：运输车辆安全驾驶行为的分析。

车联网是物联网技术在交通领域的典型应用，随着市场成熟的提高，其内涵

也被进一步拓展。本文通过对某运输企业车辆自动采取的当前驾驶行为下的车辆

状态信息以及环境信息进行处理并建立数学模型进行分析，对车辆形势过程中的

驾驶行为对行车安全、运输效率与节能情况进行了评估。

研究问题

(1) 利用附件 1 所给数据，提取并分析车辆的运输路线以及其在运输过程中的速度、加 速度等行车状态。提交附表中 10 辆车每辆车每条线路在经纬度坐标系下的运输线路图及对 应的行车里程、平均行车速度、急加速急减速情况。

(2) 利用附件 1 所给数据，挖掘每辆运输车辆的不良驾驶行为，建立行车安全的评价模 型，并给出评价结果。

(3) 综合考虑运输车辆的安全、效率和节能，并结合自然气象条件与道路状况等情况， 为运输车辆管理部门建立行车安全的综合评价指标体系与综合评价模型。

### 3.1 寻找问题类型

通过关键字查找相关文献，确定相关研究方向，确定问题类型。

研究领域：城市计算

问题类型：轨迹数据挖掘

### 3.2 查找相关解决方案

对每一个问题，查找是否已经有解决方案。如有,加入实施方法名单；如没有，开启头脑风暴，自定义方法（即使是最简单的方法也可以）。

对于本次比赛相关的研究方案：

1. 轨迹分段-聚类算法
2. 地图可视化-google Api可视化
3. 自定义安全等级
4. 天气数据-数据增强
5. 总和评价指标-模型分类

### 3.3 着手实施

着手实施这一块儿，需要开发人员，建立一个问题大纲，在项目中简历逻辑清晰的模块。方便管理。动手越快越好，数据挖掘类，一般是一环扣一环。所以，尽快实施。

#### 3.4 数据挖掘类问题一般解决过程

（1）数据预处理

对于数据的预处理，我们一般需要做空值处理，数据去重，数据纠偏，数据去噪，数据归一化。其中空值处理包括（均值或中值替代，如果较少也可以直接去掉，或者使用相关算法进行生成。）。对于数据纠偏常用的是卡尔曼滤波，均值滤波等。数据去噪一般一般使用基于距离的方式。对于离散性数据我们还需要热编码或者其他编码方式转换成数值形式。在自然语言处理任务中对于句子或者词，可以使用向量表示。

对于本次竞赛，我们主要使用了卡尔曼滤波对速度进行了纠偏（文献中有提到，所以用了。），对数据进行了去重，去噪采用的方式是聚类。在生成路线的同时，去掉噪声。（距离也要归一化到同一尺度。）

（2）数据可视化

数据可视化一般用来观察数据的分布，可以使用seaborn，seaborn包含降维算法，能够筛选出数据的重要维度。如果分析两个特征之间的相关性可以使用matplot等工具包。

（3）统计学分析

数据挖掘的统计学分析一般不等同于实证中的假设检验。在数据挖掘中，只要存在相关性，就可以拿来作为特征，我们只用筛选出权值较大的特征列。比如某两列是否高度相关，增添一些数据分析的成分，方便之后模型的解释性。

（5）特征选择

特征选择通常用于那些大规模高维数据，对于这些高维数据，含有较多冗余特征列，所以我们要剔除哪些相关性较小的特征列。常见算法，PCA,SVD，流形降维算法（tsne）等。

这次比赛的特征较少，所以没有采用降维。

（6）模型构建

模型构建这一块儿，针对特征列较少的数据，我们需要数据增强，这里的天气数据就可以拿来作为特征增强。

（7）模型选择

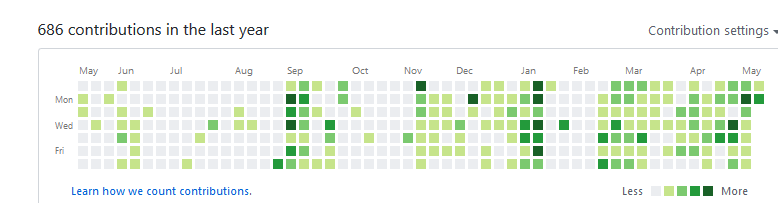
对于分类的模型，一般是非线性的，可以使用一个较为简单构造的基模型作为对照。然后用较复杂的模型与之对比。最后选择一个合适的，能够解释的模型。

对于回归模型，一般使用线性回归作为基模型，在金融领域一般只使用线性，数据科学可以尝试使用稍微复杂的模型。

总结：对于模型选择这一块儿，有一个奥尔姆剃刀原则，越简单，越可解释越好。

这个比赛的项目地址：<https://github.com/BigBigRadish/Taidi_dataMing_C>

## 4如何学习编程



Python数据处理：pandas，numpy工具包；可视化工具包：seaborn，matplotlib;机器学习算法工具包：Sklearn，keras，imblearn（常用来进行数据采样，非平衡数据处理）。

自然语言处理：pyltp，nltk。

如果需要创新，就关注其他数据科学或者计算机科学专栏。

看论坛比如Overstack，CSDN,Medium,Quora。

快速编程平台：Github,Gitlab。

数据科学竞赛平台：Kaggle,天池。

策略研究平台：Jointquant，BigQuant

总之就是不断的编程。

## 软件的推荐

excel

可视化工具-Tableau

思维导图-Xmind

markdown-vscode

notpad，myeclipse，pycharm

## 6．建模的感悟

团队可以规避个人风险