！先上美赛官网溜一圈，熟悉他的页面内容，并且仔细阅读相关规则、提交说明。

<https://www.comap.com/undergraduate/contests/mcm/instructions.php>

相关规则说明，一定要在赛前弄懂。

[队内分工]

下列所有内容，最好有两个队友会，至少有一个队友会

matlab，sps，lingo，写论文，画图，做表，查找文献，看文献，处理数据，查找数据，翻墙，论文排版，latex，word，excel，英语

下列内容，最好有一个队友会

python，ps，编程画图

[备赛学习内容]

（1）看比赛经验贴、历年赛题和优秀论文

（2）正式比赛前准备两次模拟实战，所以实际上你只有半个月的时间准备

（3）知道基本的建模方法， 并知道他们是做什么用的，深入搞懂两个复杂的建模方法

（4）设计自己的latex模板，写论文、排版等问题；如果实在没时间就用word。

（5）excel，python，spss等数据处理要会一个

（6）编程手，matlab，lingo，python，C/C++、spsspro等建模语言，至少学会一个，学到可以建模的程度。

（7）研究选择赛道的o奖论文，学会他的建模方法和论文技巧、流程

[选题]

六点放题 赛前可以随意一点 不用太在意 睡得着就睡睡不着就拉倒 看看书看看题复习一下，反正放平心态就好。

[翻译]

！可以先用翻译软件先翻译一下 看一下大致的方向

！选择适合的赛题

看起来比较简单，但是在做的过程中可能会遇到各种问题

！如果找不到适合的题目

对于自己擅长的赛道，去查相关文献最多的题目，根据文献来找思路

！使用最多半天确定选题，并且完成查找文献的工作

[论文]

赛前准备：

写论文的同学

1、研究一下历年优秀论文的论文样式

2、选择比较好看的论文作为自己的论文模板

3、做一些修改，使模板更好看，用word或latex存储好

赛时：

写论文的同学

1. 写背景（在比赛第一天就可以写，一边查文献一边总结出这个问题或类似问题的背景，即目前的解决方案、他们的优缺点）
2. 主要Task部分的方法、结果
3. 所有Task部分的方法、结果。1和2步骤的同时，将建模中所制作的图表插入到论文中
4. 写优缺点，写总结
5. 写介绍、写summary
6. 排版、润色、微调

其他

1. 将自己做的事情记录下来，以文字和图表的形式记录下来，方便写论文。
2. 由对建模最了解的队友去梳理整篇论文的流程、优化论文的描述。写摘要、总结和引言。
3. 检查

[读题]

1. 探究出赛题的老师是如何想的，从中找到建模的思路（解决方法），在题目中会有提示，这些提示的关键必须要在论文中体现出来。
2. 提取关键字
3. 查找相关文献，整理思路、拓展思路以及找到解决方法。对于每个问题的解决方法也有清晰的思路。
4. 整理出解决问题的思路（包括数据查找、方法的解决）【赛题不一定提供数据】
5. 列出时间表，解题框架，实时更新

！读懂题目

1. pdf转word
   1. 直接用word方式打开pdf，但是可能会有许多错误
   2. 在线文件格式转换 <https://smallpdf.com/>
   3. win的一款格式转换工具：格式工厂
2. 英译中用谷歌翻译
3. 中译英用百度翻译，可以翻过来看一下是不是想要的意思，也方便了解关键语词的英语表达
4. 专业的英语单词、专有名词需要单独查询。
5. 微软爱写作解决语法问题

[小技巧]

1. 可以调整文档的宽度增加页数
2. 可以将一些问卷、调查、图之类的作为附录写进论文，也可以增加页数
3. 最好贴上代码，但是代码一定要跑通
4. 参考文献一定得是英文

[查找文献]

1. 直接在搜索引擎中搜索
   1. 下载文献后，重命名加上标号
   2. 建立一个文档，专门记录文献索引
2. 谷歌学术、百度学术、SCI、知网、WOS，用好学校的数据库

非英文文献要翻译成英文

1. wiki搜索关键词，看介绍、引用的文献、去梳理他的引用和被引线路

[比赛时间安排]

第一天：上午九点、十点左右确定选题，下午查找相关文献、数据等，晚上时对赛题有大致的建模思路，确定主要的Task采取什么方法。我们已经确定类型的可以将这个时间缩短。

第二天：完成建模思路的设计，完成主要的任务Task，还有时间就写论文

第三天：一个人负责写主要任务的论文部分，其他队友解决其他问题

第四天：完成其他部分的论文，润色、修改、排版

\*论文中需要的参考文献及时记录，需要制作的图表在建模中能做的就及时做了。

！比赛中会发生很多情况：

1、因为各种原因换题：截止到第一天的下午三点，再后面的时间内就不要考虑换题了

2、数据长时间找不到，也爬取不到（不是自身数据查找能力的问题）

考虑转换思维，用一些方法巧妙的转换所需的数据类型或领域

3、没有思路，看文献也没有思路；遇到某个问题，想不通；遇到某个bug很长时间没解决掉：出去冷静一下，或睡一觉，冥想一下，放松一下

4、除了最后一天晚上，三个人不要同一晚上熬夜

[数据来源]

【常用公开数据】

1、kaggle：有很多公开免费的数据 https://www.kaggle.com/

为开发商和数据科学家提供托管数据库、举办机器学习竞赛、编写和分享代码的平台。

类似的竞赛网站都有数据，DataCastle，阿里天池，Datafountain

2、国家数据库 https://data.stats.gov.cn/

3、Github-Awesome Public Datasets

https://github.com/awesomedata/awesome-public-datasets

一个Github项目维护了全球40多个国家涵盖科技、文化、农业、天气、经济等等领域的数据，当然他仅提供数据源的链接，不提供数据源。

4、和鲸：中国版kaggle

5、UCI：经典的机器学习、数据挖掘的数据集

6、CEIC：涵盖超过195个国家400多万个时间序列的数据源，最完整的一套超过128个国家的经济数据，能够精确查找GDP、CPI、进口、出口、外资直接投资、零售、销售以及国际利率等深度数据。

7、万得：被誉为中国的Bloomberg，在金融业有着全面的数据覆盖，金融数据的类目更新非常快，据说很受国内的商业分析者和投资人的亲睐。

8、搜数网：已加载到搜数网站的统计资料达到7,874本,涵盖1,761,009张统计表格和364,580,479个统计数据，汇集了中国资讯行自92年以来收集的44个行业所有统计和调查数据。

9、亚马逊：来自亚马逊的跨科学云数据平台，拥有免费且开源的数据库，包含化学、生物、经济等多个领域的数据集。

！细分数据源

金融财经数据

同花顺数据中心：

提供股票债券等金融数据

和讯数据：

提供各种股票、基金、外汇、债券等实时数据，有付费，有免费

零壹财经：

大大小小的网贷数据、排行榜等

金融数据网：

提供每日的黄金、汇率、农产品、汽油价格的相关数据

萝卜投研：

提供股市、证券等数据，多以数据研究报告形式展示

金融界：

提供股市、融资等各种资金流向数据，以及丰富的财报和研报

东方财富网：

提供多国的股票、财税、行业、消费等大量丰富的经济数据

吉林金融网：

提供吉林的融资、市场数据

搜狐证券：

提供货币、外汇、行业、市场等宏观数据

CCER经济金融数据库：

提供企业财务年度数据、股票收益数据等数据库下载

香港金融管理局：

提供香港的各种经济及金融数据，以宏观数据为主

世纪未来：

数据服务平台，提供银行业务研究、金融大数据等各种付费数据

新浪财经：

提供国内国民经济、行业信息、对外经贸、居民收入等各种宏观的经济数据

司尔亚司数据信息有限公司：

覆盖超过195+个国家的经济数据库，数据全面但收费

INSEE数据：

法国统计与经济研究院公开数据

投中研究院：

提供投资领域的分析报告，网站更新频率较高

互联网数据

营销查询：

5118 / chinaz：

支持查询网站排名及发展趋势、百度收录情况等信息

百度指数：

百度旗下/以百度网民行为数据为基础的数据分享平台，支持查询需求图谱

微信指数：

微信旗下/基于微信大数据分析的移动端指数，打开微信→搜索“微信指数”可用

移动观象台：

基于移动设备用户的操作行为，提供应用、公众号排行等，以及大量数据报告

新榜/微小宝/易赞：

几款工具比较类似，支持查看公众号排行、看公众号人群画像等功能

阿里指数：

阿里旗下/反映淘宝平台市场动向的数据分析平台

报告分析：

易观智库：

199IT导航的母网站，权威的互联网数据平台，涵盖战略新兴产业、电商互联网、共享经济、社交营销、移动互联网以及其他互联网服务的数据及分析报告

艾瑞网：

高频率更新互联网前沿的新闻资讯、分析报告，旗下包括艾瑞指数、艾瑞咨询等

艾媒网：

艾媒与艾瑞无关，艾媒网发布的研究报告多偏向移动互联网方向

CBNDdata：

以阿里巴巴的商业数据库为基础，输出产业经济分析报告

QuestMobile：

周期性地发布一些关于APP的研究报告

阿里研究院：

阿里旗下/发布研究电商等方向趋势的数据报告，内容多与阿里相关

360研究报告：

360旗下/移动、PC、网站、企业、诈骗等安全领域的研究，以及大数据等非安全领域的研究

中国互联网信息研究中心：

经国家主管部门批准组建的管理和服务机构，经常发布一些有价值的互联网信息报告

中国信通院：

针对互联网多个行业的发展趋势，发布白皮书，角度较为宏观

中国互联网数据平台：

提供全国各地区互联网发展报告、基础数据及研究分析报告，数据及报告偏向学术性研究

清博大数据：

提供微信、微博、头条榜单数据，各种研究报告、数据报告、舆情报告及行业热点

数据观：

丰富的前沿行业资讯，及部分研究报告下载

腾讯大数据：

数据服务平台，提供调查研究、移动互联网及特色数据报告，质量较高

大数据世界：

提供大数据相关资讯、应用案例、技术方案及资料下载

地理数据

水土气候数据：

水土保持生态建设网http://www.swcc.org.cn/

黄河风情http://www.yellowriver.org/

黄河流域数据中心http://henu.geodata.cn/index.jsp

黄土高原水土保护数据库http://www.loess.csdb.cn/

数字黑河http://heihe.westgis.ac.cn/Default.aspx?tabid=106

大气科学数据库http://data.iap.ac.cn/

中国气象科学数据http://cdc.cma.gov.cn/

中国气象科学数据共享服务网http://cdc.cma.gov.cn/

全球森林土地覆盖动态http://www.fao.org/gtos/gofc-gold/

FRA 2000http://edcsns17.cr.usgs.gov/glcc/fao/index.html全球森林资源

NOAAhttp://www.noaa.gov/美国国家海洋和大气管理局

NCDC http://lwf.ncdc.noaa.gov/oa/ncdc.html美国国家气候数据中心

国际水管理研究所http://www.iwmidsp.org/iwmi/info/main.asp

气候研究所http://www.cru.uea.ac.uk/

植被覆盖数据http://www.vgt.vito.be/

环境数据：

黄土区农业与生态环境数据分中心http://loess.geodata.cn/

东北黑土农业生态数据库http://www.blackland.csdb.cn/default.asp

中国西部环境与生态科学数据中心http://westdc.westgis.ac.cn/

西部数据中心http://westdc.westgis.ac.cn/data

黄土高原区域数据共享运行服务中心http://loess.geodata.cn/Portal/?isCookieChecked=true

遥感数据：

地面站RS地图查询http://cs.rsgs.ac.cn/cs\_cn/query/query\_map.asp

甘肃省基础地理信息中心http://www.cehuiju.gansu.gov.cn/5chcg/giscenter.asp

遥感数据网络资源http://ecology.bnu.edu.cn/gejp/student/zhj/webresource.htm

中国遥感数据网http://ids.ceode.ac.cn/

GLCF http://glcf.geodata.cn/etm/Welcome.do

GIS FORUM http://www.gisforum.net/全球最大的GIS中文门户网站

遥感数据共享http://ids.ceode.ac.cn/

其他自然人文数据：

地理国情监测云平台 http://www.dsac.cn/

中国资源卫星应用中心 http://www.cresda.com/cn/profile.htm

中国自然资源数据库 http://www.data.ac.cn/index.asp

中国科学院资源环境科学数据中心 http://www.resdc.cn/resdc/first.asp

中国地球系统科学数据共享网 http://www.geodata.cn:8086/Metadata/index.jsp

国家地球系统科学数据共享服务平台 http://www.geodata.cn/

PM25.in | PM2.5(细颗粒物)及空气质量指数(AQI)实时查询！：

中国空气质量监测实时数据http://pm25.in/

国际科学数据服务平台 http://datamirror.csdb.cn/index.jsp

原始数据搜索 https://wist.echo.nasa.gov/api

中国科学院资源环境科学数据中心 http://www.resdc.cn/resdc/first.asp

世界聚集人口数据 http://beta.sedac.ciesin.columbia.edu/gpw/index.jsp

光谱数据分析网站 http://modis.gsfc.nasa.gov/

美国航空航天局NASA http://daac.gsfc.nasa.gov/

城市轨道交通http://www.ntsg.umt.edu/

人地系统主题数据库http://webpanda.iis.u-tokyo.ac.jp/

影像数据：

USGS Earth Explorer：

提供Landsat、MODIS等数据

ESA’s Sentinel Mission：

提供Sentinel系列卫星数据，可以下载欧洲航天局所属的影像数据

NOAA CLASS：

提供高品质大气数据集，包括美国国防部（DoD）、极地作战环境卫星（POES）、NOAA的地球静止环境卫星（GOES）以及其他派生数据

NASA Reverb：

提供多种类数据，包括：

Aqua, Terra, Aura, TRMM, Calipso, NASA DC, JASON, ENVISAT等

EOLi：

欧洲航天局的地球观测目录和订购服务的客户，通过Java 程序下载数据，支持的数据包括：

Envisat, ERS, IKONOS, DMC, ALOS, SPOT, Kompsat, Proba, IRS, SCISAT.

INPE：

提供中巴资源卫星数据

Bhuvan Indian Geo-Platform of ISRO：

提供印度境内的MS-1， Cartosat,，OceanSat and ResourceSat数据以及全球的NVDI数据、南亚国家的CartoDem Version -3R1数据和北印度洋的气候数据

政府数据

新加坡政府公开数据：

提供新加坡各类数据

美国政府公开数据：

提供美国全国各类数据

法国政府公开数据：

法国政府开放数据平台

英国政府公开数据：

提供英国各类数据

国家数据：

中国国家统计局权威数据

中国统计年鉴：

1999年至今中国统计年鉴，国家统计局官网，提供单页Excel文件下载

中国统计信息网：

全国及各级政府各年度统计公报、年鉴等，为收费数据

年鉴汪：

全国城市统计数据搜索引擎，浏览免费，下载收费

伦敦市公开数据：

提供伦敦人口、就业、环境等各类数据

国土资源部：

国土资源部对外公开的信息报告

其他细分行业数据

中研网数据：

提供医疗、房产、制造业、服务业、零售消费、车辆等全行业数据

中国报告大厅：

提供各行各业的基础数据、调查报告、分析报告、预测报告，种类丰富

CADMAPPER：

世界各大城市的DXF文件，原始数据源为OSM，NASA，USGS

亚马逊网络服务公共数据集：

跨科学云数据平台，包含化学、生物、经济等多个领域的数据集

Awesome Public Datasets：

Github公共数据搜集项目，自然科学、社会科学覆盖较面，包含各个细分领域的数据库资源。

figshare：

数据分析与研究成果共享平台。

英国公开数据浏览工具：

James Trimble制作的英国数据可视化浏览工具集；

数据法国：

法国各类数据的可视化呈现；

DataEye：

基于国内游戏、汽车行业数据，进行多角度的行业调查报告撰写发布

CBO中国票房：

基于国内票房数据，统计票房排行、上座率等信息。

易车指数：

反映国内汽车销售市场，为购车者或汽车从业者提供参考的数据指数

高德地图：

支持实时查看国内交通情况，此外高德周期性提供一系列数据报告

房天下：

提供中国指数研究院和CREIS中指数据，展示国内房地产数据情况

艺恩：

CBO中国票房数据的提供方，提供一些行业数据报告，如动漫IP价值研究报告

[流程分析]

！数据处理

1、确保数据的完整性、唯一性、权威性、合法性、一致性

2、让数据更适合做挖掘或展示

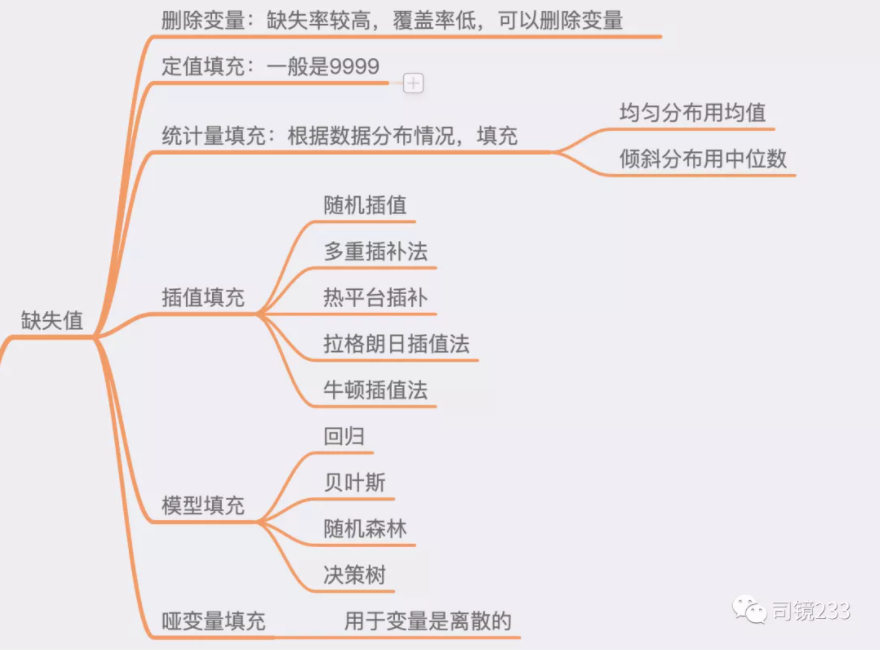
适合的维度，有效的信息，相关的数据

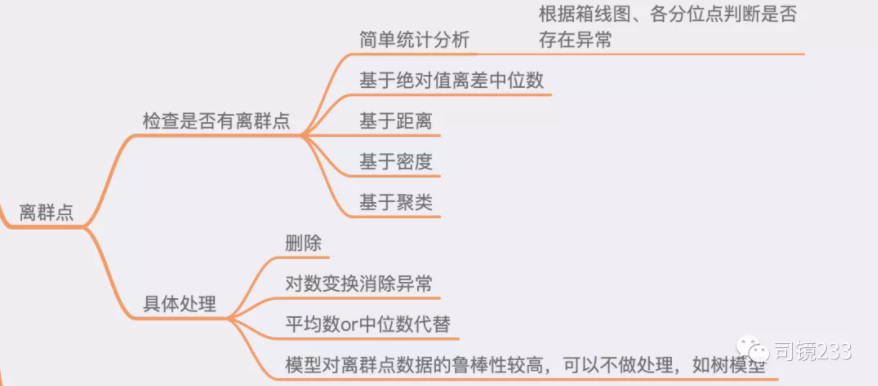
**【基本步骤】数据清理--数据集成--数据规约--数据变换**

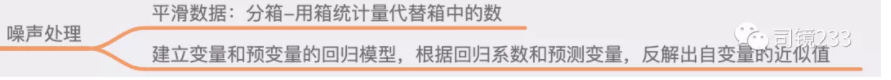


数据中有类型、格式、存储错误的需要提前预处理

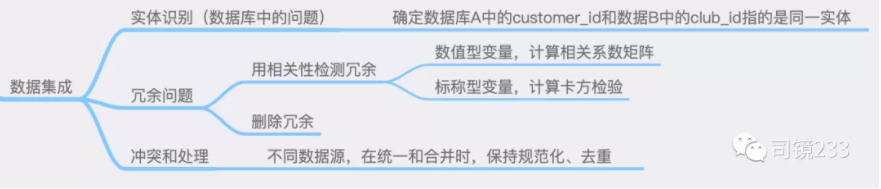
1. 数据清理



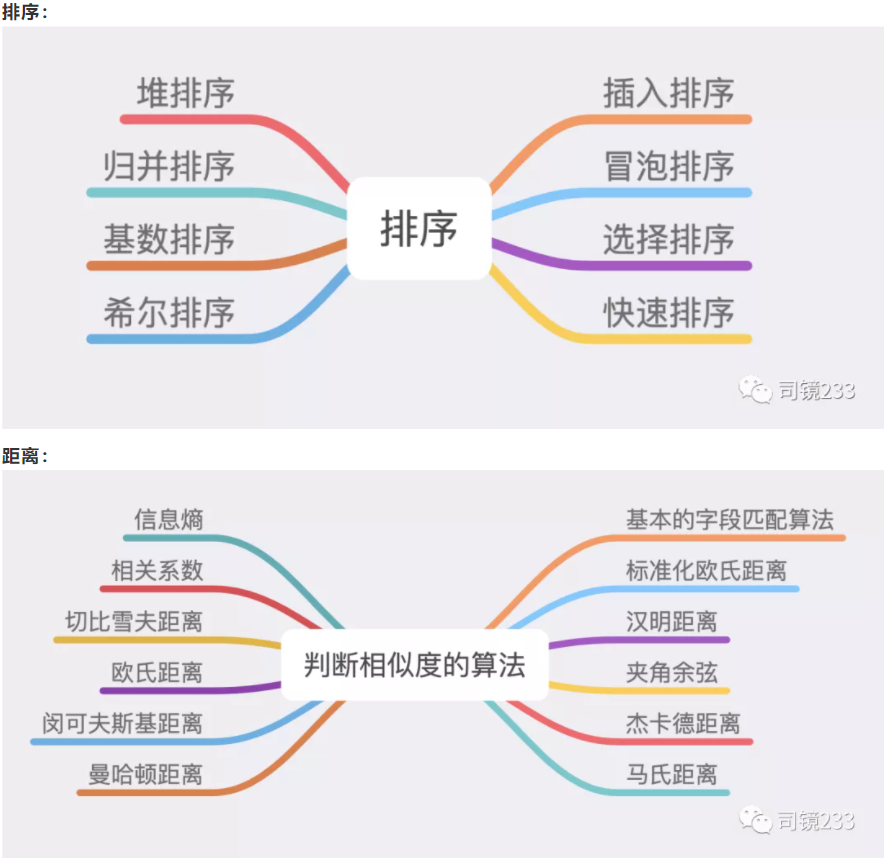




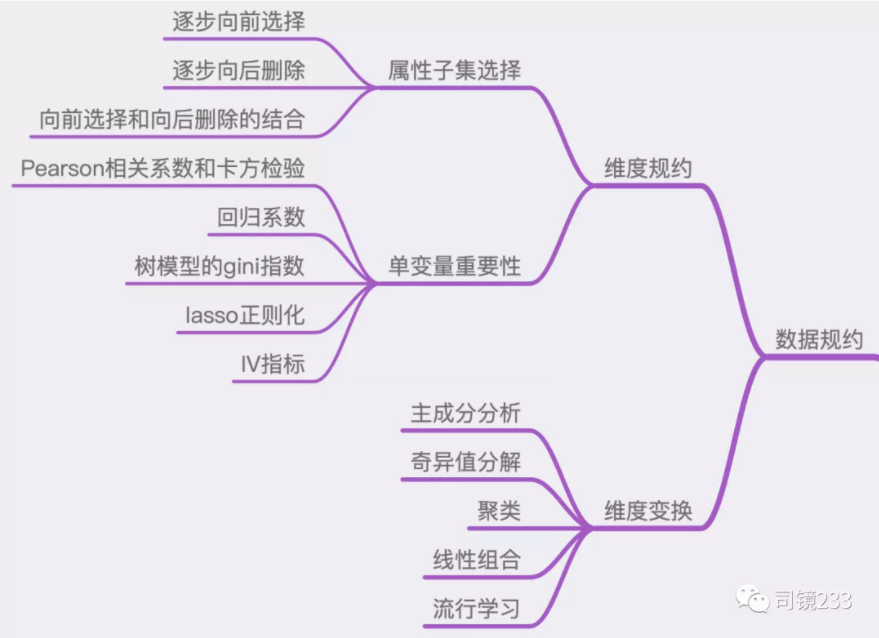
1. 数据集成



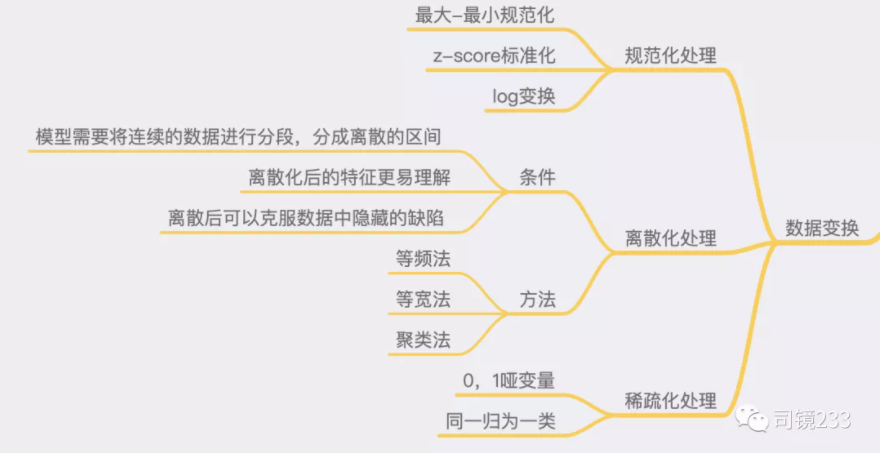
删除冗余数据，基本思想是排序、判断相似性、合并或删除



1. 数据规约



1. 数据变换



！模型

1. 分析问题

！根据已知信息

* 1. 白箱模型：内部规律比较清楚

知道模型的信息，不同变量之间的关系。对于未知参数，可以通过某种方式估算这些参数

* 1. 黑箱模型：内部规律基本不为人所知

神经网络：模型不透明

非线性系统识别NARMAX（带有异质输入的非线性自回归移动平均模型）：模型透明

/\*

参数问题：任何非纯白箱模型，都有参数

参数的优化称为训练，神经网络、机器学习中常用

模型超参数的优化称为调整，通常会使用交叉验证

常规的的建模，参数通常由曲线拟合确定

\*/

！加上主观分析（先假设后检验）

1. 简化问题
   1. 判断哪些核心部件必须保留，哪些可以简化！！！！！

由于模型是将现实世界抽象

* 1. 假设非常有必要

实际的情况非常复杂，所有的模型都不是准确的，只是近似。比如族群遗传学模型，假设族群固定大小。

* 1. 拟合

模型过于适合数据，就会失去了将其推广到一个极端，缺乏观察到新事件的能力，可以放到检验的时候去研究，如果出现这种问题一定要试一下其他数据

1. 创建模型

（1）实际因素转化为变量

（2）分析变量的关系，用合理的数学式表达

（3）根据实际问题，选取合适的数学框架：算法，建模方法

（4）计算机模拟编程，求出结果

4、分析结果

计算结果解释实际问题，并分析结果的可靠性。使用信息可视化

1. 模型评估

A:将数据，分为不相交的两个子集：训练数据、验证数据

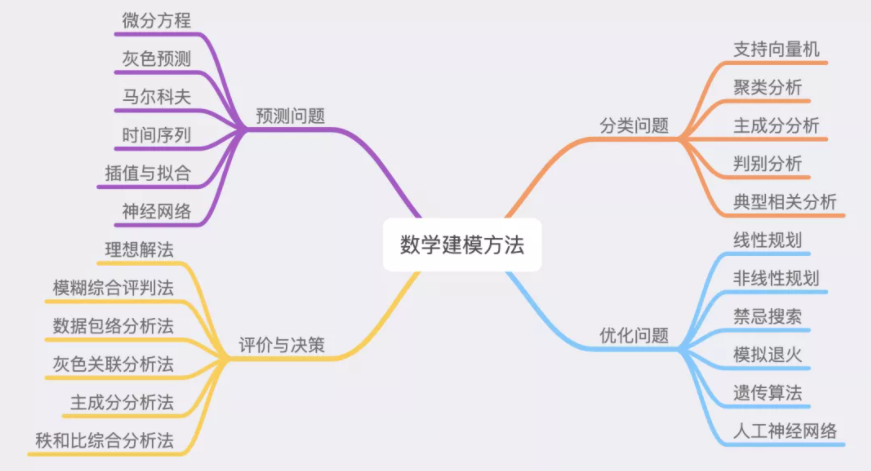
训练数据用于估计模型参数，验证数据用于检验模型——交叉验证

B：定义度量，计算实际数据与预测数据之间的距离

（决策理论，经济模型中常用损失函数）

C：测试参数的有效性。比较简单，测试模型的有效性会比较困难

微分方程模型，统计模型，非参数统计工具，最小假设





！图表制作

Excel、python、matlab、R、PPT

原则：

简单明了

符合规范

表明数据关系：构成、比较、趋势、分布、练习

选择合适的图标种类

<https://cloud.tencent.c1om/developer/article/1052457>

！论文书写基本流程，除上述板块外

摘要，每个任务的解决方法，核心问题多写点（方法、解决什么问题、效果、稳定性）

介绍

分问题 介绍方法、效果、优缺点

附录 正文中没放入的内容、更多的支撑内容、图表、代码

[赛题补充与更新]

在整个竞赛期间，可能在MCM/ICM问题网页上发现有竞赛题的说明与更新。如果你们还有网站上未涵盖的问题，你们可以发送邮件到mcm@comap.com，邮件中一定要包含你们的参赛队号。

在竞赛期间，一定要注意查看网站上的更新还有自己的邮箱。如果有竞赛信息的任何更新（可能包括特定问题的更新），COMAP都将会更新这个网站。

[论文书写注意事项]

！论文必须用英文书写，采用一种不小于12P的易读字体，页面尺寸不大于A4幅面。解答必须包括完整的书面文本，以及必要的图形、图表，或其他的书面材料。但不接受非论文的支撑材料，例如计算机程序文件或工具软件文件等。

！在解答论文每一页的页眉上必须显示正确的参赛队控制号和页码数。例如，在每一页上使用下面的页标题：

Team # 54321 Page 6 of 20

所以，务必要准备好模板，开赛后能直接更改相关信息就可使用

！论文中不能包含任何形式的参赛队身份信息，如学生姓名、学校名称或地理区域等。如果要求你们提交一封信，一定不要在信上签写你们的真实名字。如果你们认为在信的最后需要有一个正规的结束语，建议使用：Sincerely, Team#54321.

！要求每个参赛队提交一份自己的论文电子版（PDF格式或Word格式文件），发送邮件到solutions@comap.com。任何一个参赛队员或指导教师都可以发送邮件，但不要多次重复发送。

在邮件的主题行写：

COMAP 你们队的控制号

例如: COMAP 54321

邮件的附件名要用你们队的控制号命名

例如：54321.doc，或54321.pdf

！一定要仔细检查队号

！COMAP只接受你们解答论文的一个PDF 文件或 Word文件。在你们的邮件中不能包括你们的控制页（Control Sheet）、计算程序、或工具软件，这些在评判过程中是没有用的。要求每封邮件只能包括一个参赛队的论文。你们队的摘要（summary）必须是论文文件的第一页，不能分为两个文件。

！邮件的附件必须小于17MB！一定不要使用云附件或超大附件功能，例如Google Docs、QQ的超大附件和网易的超大附件等，否则COMAP将不能收到你们的附件文件。你们的邮件必须包含一个PDF文件或Word文件。

！控制页/提交

每一个参赛队员必须在控制页上签上自己的名字，以表明他或她遵守了竞赛规则和要求。

将签了名的控制页（Control Sheet）发送邮件到COMAP：

在签名的控制页准备好之后，发送邮件到：[forms@comap.com](mailto:forms@comap.com)。

邮件主题行写：COMAP 你们的控制号，例如：COMAP 54321。

COMAP只收受一个你们队控制页的PDF文件或照片图像文件。要求每封邮件只发送一个控制页。