

# 美赛经验分享

周宇洋



## 一、关于美赛题型

MCM :

A: 连续型 (官方说A题是连续型, B型是离散型, 但是其实A、B题的类型非常模糊, 每年出的类型都不一样, 变数非常大, 拿奖也需要较强的建模基础)

B: 离散型

C: 大数据 (数据量最少3000+条, 前几年已经摸索出大概的做题模板, 参与本题的队伍也越来越多, 认真准备可以保M争F)

ICM :

D: 运筹学/网络科学 (C和D都是计算机同学常选的两个题目, 尤其是在深度学习的大背景下, 有机器学习or深度学习基础的同学都是稳M奖的)

E: 环境科学

F: 政策 (如果对自己英文写作水平比较自信, 或者对美国环境/世界局势比较自信可以选E/F, E题也是有做题模板的。很多经管专业的同学做E/F题都能拿很好的奖)

如果有充分的准备时间建议好好准备, 优先选C/D题, 其次A/B/E, 最后考虑F。 (当然要具体情况具体分析, 结合题目选择最适合队伍的题目)





## 二、获奖情况

### 2020 MCM Statistics

**13749 Teams Participated**

**3851 Problem A (28%)**

**2453 Problem B (18%)**

**7445 Problem C (54%)**

**19 Outstanding Winners (<1%)**

**180 Finalist Winners (1%)**

**839 Meritorious Winners (6%)**

**3522 Honorable Mentions (26%)**

**8941 Successful Participants (65%)**

**49 Unsuccessful Participants (<1%)**

**198 Disqualified (1%)**

**1 Not Judged (<1%)**

### 2020 ICM Statistics

- **7199 Teams participated**

- **2089 Problem D (29%)**

- **2526 Problem E (35%)**

- **2584 Problem F (36%)**

- **18 Outstanding Winners (<1%)**

- **234 Finalist Winners (3%)**

- **564 Meritorious Winners (8%)**

- **1565 Honorable Mentions (22%)**

- **4612 Successful Participants (64%)**

- **34 Unsuccessful Participants (<1%)**

- **170 Disqualified Teams (2%)**





保研加分：

O（特等奖）：1.5（还有机会直推）

F（特等奖候选）：1

M（一等奖）：0.5（性价比高的奖，不用花费太多时间就可以加分）

H（二等奖）：0.25

S（成功参与奖）及以下不加分（花钱买奖加分已经不存在了，所以好好准备）

PS：美赛参赛是以团队形式加分，加分政策上也是按团队竞赛加分，但是实际加分其实是按照单人竞赛加分的，所以不用考虑分数分配的问题。并且最好能3人参赛（多个人多份力）





### 三、评分标准（近几年没有具体标准流出，这个评分标准有好久的历史了，仅供参考）

<u>RUBRIC</u>	
10	Summary
30	Modeling Assumptions metrics utilities
15	Science
15	Data/Validity/Sensitivity
10	Strengths/Weaknesses
10	Communication Visuals/charts
10	Recommendations
<u>100</u>	



## 四、赛前准备

### 1、关于论文写作

务必LaTex模板！赛前准备最优先的，比word节省非常多时间！掌握基础语法就可以，套用模板很简单。

<https://github.com/xjtu-blacksmith/easymcm>

👉我用的上面这个模板，普适性比较高，推荐一下

如果队伍三人全员英语很ok，可以直接从建模开始就全程英文写作；

如果队伍里有英语写作不那么熟练的，建议中文写作，通过机翻+人工修改或者直接人工翻，方便队伍沟通。

论文的内容一定要回答题干中的所有问题，否则会非常影响成绩。

论文引用最好不要出现中文文献，在比赛之前接触的所有常见模型，都可以提前找一下它们的出处论文，比赛的时候用到这个模型或者算法就直接把参考文献加进去，提升论文的水平。





## 2、制图制表——比赛拿奖的决定性因素之一

美赛M、H奖公式：创新的思路or做题模板 + 优秀的表达和制图

美赛O、F奖公式：创新的思路or做题模板 + 优秀的表达和制图 + 严谨的数据和代码

**美赛不要求提交源码。**如果比赛的三天半时间比较紧张，代码和数据结果可以没那么严谨，但最后的图标一定要非常华丽，展示得非常好。

首推echart，非常建议赛前学一学，做出来的图片格调非常高，对以后也非常有帮助。

如果时间不够了，根据自己的基础去熟悉相关语言的制图。matlab或者python。比如python的话，就要提前去了解作图的各种库，要能达到熟练调库的水平，比如matplotlib。

要提前了解各种数据图，比如堆叠柱状图，比如这几年o奖论文中非常常见的桑基图。

**务必要注释制图指表！比赛过程中现学来不及的！**



可以去echart官网了解一下各种图，在拜读其他优秀论文的时候思考，他们是怎么展示数据的

折线图

柱状图

饼图

散点图

地理坐标/地图

K线图

雷达图

盒须图

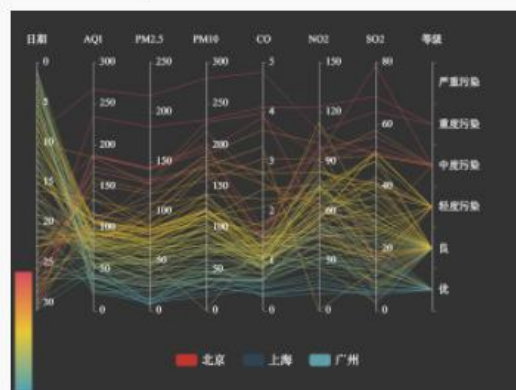
热力图

关系图

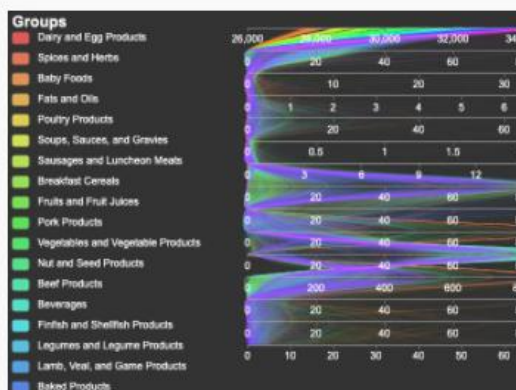
路径图

树图

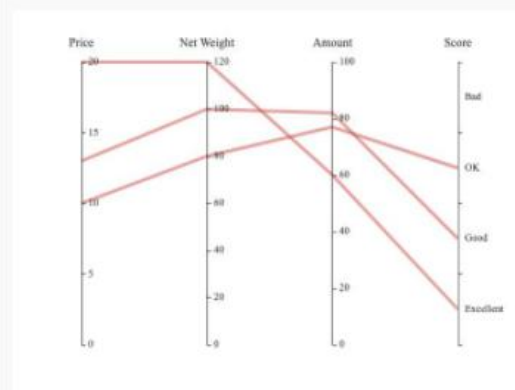
Parallel Aqi



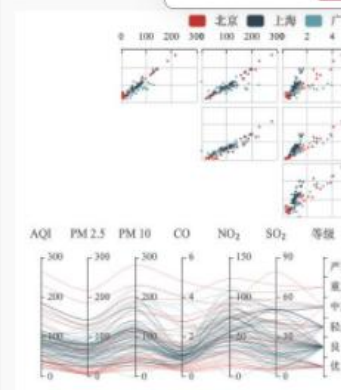
Parallel Nutrients



Basic Parallel

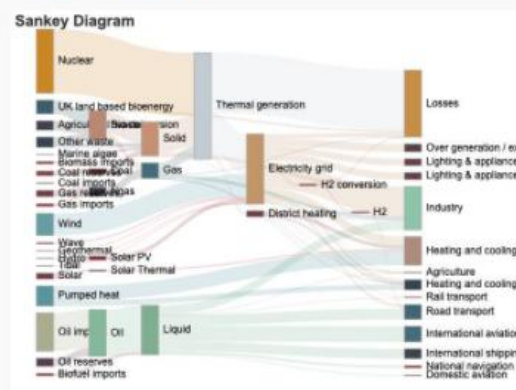


Scatter Ma

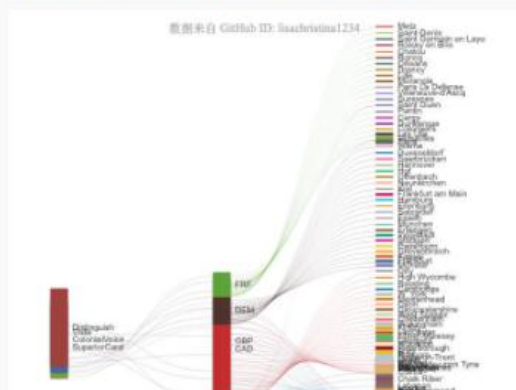


桑基图 Sankey

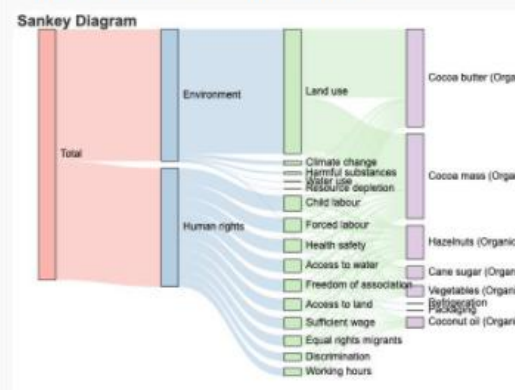
Sankey Diagram



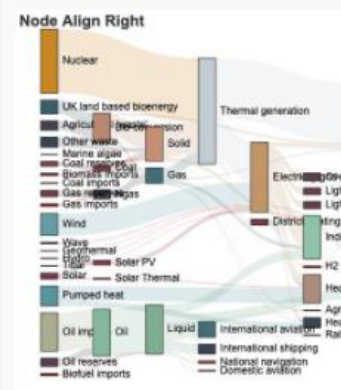
Specify ItemStyle for Each Node in ...



Sankey with Levels Setting



Node Align Left in Sankey





### 3、阅读优秀论文

团队至少精读20篇以上优秀论文，可以三个人分工读不同的论文，然后互相讲一下大概的思路。

#### 优秀论文的选择

- 1) 6个体型选择6篇0奖论文大概阅读一下，决定自己团队的选择范围，选题范围至少要缩小到1~4类题型。
- 2) 在自己的选题范围中，再阅读历年的论文

阅读论文最好能写一下文章的大概思路，用了哪些模型，用了哪些图，怎么用的表格。都记录下来，相当于列个提纲方便考试的时候查找。

ps. 可以注意下优秀论文的参考文献都是怎么组织出来的，对论文写作的思路也有帮助的。



#### 4、模拟赛

正式比赛前建议团队先做一次模拟训练，和能力没啥关系，主要是磨合团队和训练论文写作。

关于团队分工：

最标准的分工是：一人建模、一人码代码出数据（作图）、一人论文写作（作图）  
作图让码代码的来或者让写论文的来都可以

时间紧效率高的分工：先一起讨论大致思路和用到的模型，每个人负责一部分建模和代码加写作，最后汇总到一个人翻译。

建议写作or翻译由一个人来做，这样写作风格可以比较确定。





## 五、C题做题模板

我们队伍连续参加了两年的4美赛，在我加入大数据专业之后对C题有更多的了解，所以我们在比赛前就选定了C题并进行了一次模拟赛。

队伍分工主要是我建模码代码出图，队友负责翻译、写作和其他需要注意的事项（比如论文提交、检查论文是否符合要求等等）

由于我对其他题目没有经验，并且数学建模的模型了解的并不是很多（我的技术栈主要是在机器学习），所以仅能分享C题的做题经验和做题模板，希望能够提供帮助。

ps. 我matlab并不熟练，主要python用得更多。打比赛之前需要根据队伍里大家的情况选择合适的语言。推荐python，上手非常快，而且有很多现成的库可以调用。我之后的介绍也都以python为例。

关于python的学习就不赘述了。



# 1、做题模板总览

## 数据预处理

- 数据清洗
- 数据集成
- 数据规约
- 数据变换

## 数据分析

- 回归
- 分类
- 聚类
- 神经网络
- 集成学习

## 数据可视化

ps 仅仅靠上面的模板是没法写出一篇很好的论文的，所以需要更多地阅读往年优秀论文，学习他们的模型和算法





## 2、数据预处理

无论是数据预处理还是数据分析过程，都可以在Kaggle上找到相关的理论知识和实战。

数据预处理概述：需要解决的三个主要问题：不完整、噪音和不一致。

论文的数据分析可以先围绕这三个点去做讨论，这里讨论的好的话，可以作为论文的一个亮点。

比如去年的C题的数据是关于亚马逊网站商品的评论和星级。在数据处理的时候我们观察数据发现可能存在刷单行为，于是我们对此进行数据清洗。我觉得是论文的一个小亮点。

### 数据可视化

在这里说道数据可视化，意思是在不通过任何算法的情况下，对数据进行展示。比如数据量、每个分类的数据量，数据缺失值，数据的平均值、众数等等特征——直方图、散点图、四分位图在这里都可以很好地使用。



## 2.1、缺失值处理

- 忽略元组
- 手工填入缺失值
- 填充全局常数Unknown
- 填充属性的中心度量（均值或者中位数）
- 填充最可能的值（可以通过回归、贝叶斯等方法或决策树归纳决定）





## 2.2、噪声光滑

这一步并不是必须的，如果题目中的是自然数据的话可以考虑一下这一步，但是做不做这里可以再斟酌一下

噪声：在可测量变量中的随机错误或偏差

常用的噪声光滑技术：

- 分箱（通过考察邻居的值来平滑存储数据的值）首先将数据分割到一些相等深度的桶，然后根据桶均值、桶中间值、桶边界值进行平滑
- 回归
- 离群点分析



## 2.3、检测和处理异常值

- 简单统计
- 3sigma原则
- 箱型图
- 基于数据模型：首先建立一个数据模型，不能和模型完美拟合的对象就是异常值
- 基于距离：可以在对象之间定义邻近性度量，异常对象是那些远离其他对象的对象。
- 基于聚类





## 2.3、检测和处理异常值

处理方法：

- 删除含有异常值的记录
- 将异常值视为缺失值，用缺失值 处理方法来处理
- 用平均值来修正
- 不处理——直接在具有异常值的数据集上进行挖掘建模



### 3、数据集成

数据集成对于多特征的数据非常重要，它直接决定了你挖掘模型的效果。

数据集成能够处理冗余数据

### 4、数据规约

获得数据集的一个简约表示，使得在容量上大大减小，但仍接近于保持原数据的完整性，并产生相同或基本相同的分析结果。

比赛中基本用不上数据规约的方式，因为不要求提交源码，所以最后的数据结果无需那么严谨。

数据规约可以作为论文的一个点。

数据规约的策略：维规约、数量规约、数据压缩





## 5、数据变换

属性构造：（和数据集成的方法有些相似）

规范化：将属性数据按照比例映射到一个小的特定范围。

离散化：数值属性的原始值用区间标签或概念标签替换。

规范化和离散化不一定非要代码实现，可以在论文里提到这个数据处理的过程



## 6、数据分析

准备C题的话下面的这些模型都要了解一下。

回归

- 线性回归
- Logistic回归

分类

- 决策树 (ID3算法, C4.5算法, 信息增益的概念)
- KNN
- 朴素贝叶斯
- 支持向量机

聚类

- K-Means

神经网络

集成学习

上面这些算法的代码都要熟悉，到时候可以直接拿来用，机器学习的算法可能有些太简单，在论文里会不太够看，所以建议还是了解一些深度学习和集成学习的内容，提升论文水平





## 7、其他提示

因为我提出的一些建议都是站在C题的基础上提出来的，难免有些 偏颇。

所以我整理了之前其他人的一些意见，从选题开始，我觉得有参考价值的内容都会添加进来。

### # 评分标准的详细解释

**摘要** 包含问题概述和全文概述：模型、方法和基本结果及模型的优点的概述，对它们有机联系的叙述将得高分

**建模** 叙述建模所需假设，模型对提供定量解答的重要性，好的论文讨论了关键假设及其对建模的重要影响，模型应是数学和文字均衡的表达而非仅仅由几个未经解释的方程和参数

**科学性** 问题牵涉许多科技领域，注意这些科技及其进步对建模的影响对建模是重要的



**数据/验证/敏感性** 建模后选择输入数据，**验证解的精度和鲁棒性**有助于模型和算法的可信度，**用敏感性分析**决定相对变化率，有时比具体结果还重要。

**优缺点 优缺点分析**可体现学生对其建立模型的理解深度，简单的理解透彻的模型远优于从文献中搬来的复杂方程

**表达/可视性/图表** 单纯数学不易被外界理解，图、表等多种模式可清楚地描述所得结果，结果不能被很好理解的不可能进入最后一轮。

**建议**





美赛评审针对中国学生的一些建议：

- 1、“建模的**创造性**”，表现出很好创造性的论文即使出现比较大的错误亦有可能获得一等奖。
- 2、注意离散和随机模型
- 3、“**模型的检验**越多越好”。可以采用证明的方法，但更多的是对某些感兴趣的情形进行计算并分析结果、对重要参数的高中低水平进行计算并分析，考虑放松某些假设等。



## 其他赛前准备

- 1、尽早完成组队和队员磨合工作；
- 2、访问官方网站，仔细研读参赛规则：

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

<http://www.madio.net/thread-168106-1-1.html>含中文参考翻译；

- 3、尽可能多的研读和实践历年获奖论文及其中的模型和求解算法，如有条件，每周都抽出一定时间进行组内队员的研讨，以有助于队员之间的磨合；

- 4、注册成为数学中国论坛<http://www.madio.net/forum.php>的会员并通过各种手段获取尽可能多的体力值以保证赛前和比赛期间下载到所需资料





5、收集可用的外文期刊数据库网址及所需密码以及熟练使用google等搜索引擎的高级搜索方法；

6、了解并熟悉建模竞赛中常用的算法：如蒙特卡罗算法，数据拟合、参数估计、插值等数据处理算法，线性规划、整数规划、多元规划、二次规划等规划类算法，图论算法，动态规划、回溯搜索、分支定界等计算机算法，最优化理论的三大经典算法：模拟退火算法、神经网络算法、遗传算法，网格算法和穷举法，一些连续数据离散化方法，数值分析算法，图像处理算法等。



先合作后分工，以确立模型为分界点

合作阶段：分析问题，选题，集中搜集资料，分组查看筛选资料，讨论和确立模型

分工阶段：建模、数据/代码可能要花费更多时间，撰写论文的同学不要干等着，在数据和图标出来之前有很多可以写的东西

## Highlights

模型：改进，检验，敏感性分析，多模型比较

算法：改进，多种算法求解和比较，鲁棒性分析

论文：摘要，图表结合，多种图并用

计算机模拟 (simulation)





## 关于文献查找

- 1) 中外文期刊数据库+Google引擎+书
- 2) 中文文献：CNKI（中国知网）和万方
- 3) 英文文献：google学术搜索
- 4) 专利
- 5) 文章数量和质量：不要盲目下，不要盲目加入到参考文献中去，参考文献最好不要放中文内容。当如如果是题目需要是可以的。



<http://www.shumo.com/home/>中国数学建模网  
<http://www.ilovematlab.cn/> MATLAB中文论坛  
<http://www.madio.net/>数学中国

其实还有非常多的参考网站，但我觉得上面这三个是最常见的三个，讨论的人数也比较多。

正式比赛过程中可以在这些网站寻找一些思路。  
但是最终的论文大思路还是要自己确定的。





# MATLAB中文论坛推出问题悬赏功能

< >> 点击查看论坛更新

内容详情 >

热门内容

最新悬赏

最新问答

实用教程

文件交换

技术博客

热门问答

更多

**已解决** 二元函数拟合怎样效果最好?



jack350556319

2020-09-30 2:46:45

34

2037

MATLAB 数学、统计与优化

**已解决** 关于matlab tcp接收数据处理的问题



Yan9

2020-10-01 11:29:10

20

1686

MATLAB 基础讨论

**已解决** 如何删除结构体中的某个元素



misillas

2020-10-22 12:15:46

4

863

MATLAB 基础讨论

**已解决** 警告: 在  $t=1.134739e+00$  处失败。在时间  $t$  处, 步长大小必须降至所允许的最小值( $3.552714e-15$ )以下, 才能达到积分容差要求。该怎么解决



wx\_N9TnlWmG

2020-10-11 9:39:19

1

857

MATLAB 基础讨论

站长推荐



MATLAB 和 Simulink 入门必备 >>

MATLAB 和 Simulink 基础入门教程、免费  
正版软件申请, 还有更多实用在线技术资源  
>>

# 建模竞赛中常用方法（针对所有题目，不仅是C题）

## 1. Monte-Carlo方法

计算机随机性模拟方法，也称统计试验方法。MC方法是一种基于“随机数”的计算方法，能够比较逼真地描述事物的特点及物理实验过程，解决一些数值方法难以解决的问题。

## 2. 数据拟合、参数估计、插值等数据处理算法

大量的数据需要处理，而处理数据的关键就在于这些算法，通常使用MATLAB或Python作为工具。与图形处理有关的问题很多与拟合有关系。





# 建模竞赛中常用方法（针对所有题目，不仅是C题）

## 3. 规划类问题算法

此类问题主要有线性规划、整数规划、多元规划、二次规划等。竞赛中很多问题都和数学规划有关，可以说不少的模型都可以归结为一组不等式作为约束条件、几个函数表达式作为目标函数的问题，遇到这类问题，求解就是关键了。

## 4. 图论问题

Dijkstra、Floyd、Prim、Bellman-Ford，最大流，二分匹配

图论问题主要还是出现在D题



# 建模竞赛中常用方法（针对所有题目，不仅是C题）

## 5. 算法设计的思想

动态规划、回溯搜索、分治算法等

## 6. 最优化理论的三大非经典算法

模拟退火算法

神经网络

遗传算法

## 7. 连续问题离散化的方法





# 建模竞赛中常用方法（针对所有题目，不仅是C题）

## 8、数值分析的算法

常体现在A、B题目中

包括函数的数值逼近、数值微分与数值积分、非线性方程的数值解法、数值代数、常微分方程数值解



# 关于建模软件

数学专业的各种大佬一般都直接推荐美赛三件套：Matlab, Lingo, SPSS

但是我感觉python真的万能了，也不用再刻意花时间去熟悉matlab的语法（

关于书籍推荐，首推姜启源老师的《数学模型》。但是准备时间一般没有充足到可以看完这本书的地步。所以建议先了解我PPT中提到的各种算法，找这些算法的学习资料。用到啥学啥，不用太多系统的学习。





# 比赛时间分配

day1: 搞清题目每个题目尝试思路分析，第一天一定要把选题确定下来，上网搜索相关论文。

day2&day3: 模型建立，相关数学公式推导以及计算机模拟，模型可以不复杂，但是一定要描述清楚。

day4: 论文写作，最好花一天时间（24小时左右），最重要的部分。



**预祝美赛顺利！**

