



山东大学

為天下儲人材 為國家圖富強

问题驱动的数学建模及其竞赛

刘保东

山东大学 计算机科学与技术学院

2020年7月9日

个人简介

- 1982-1986, 山东大学计算数学专业, 本科;
- 1986-1989, 山东大学应用数学, 硕士研究生
- 1994-1998, 山东大学计算数学, 在职博士研究生
- 2003-2004, 加拿大约克大学, 访问学者

历任: 山东大学环境科学中心, 助教、讲师、副教授;
数学学院应用数学研究所所长、硕士生导师

现任: 山东大学计算机科学与技术学院教授、硕士生导师;
山东大学“系统分析与软件设计”校级创新平台负责人;
山东大学(青岛)公共教学管理服务中心主任;
全国大学生数学建模竞赛山东赛区专家组组长;
深圳杯数学建模挑战赛专家组成员;
中国工业与应用数学学会数学模型专业委员会委员;
中国研究生数学建模竞赛评审专家;



目 录

一



理解数学建模与数学科学

二



如何学习数学建模

三



组队与赛前准备

四



参赛注意事项

第一部分 理解数学建模与数学

问题·一、何为数学？

- 数学是对现实世界的数量关系与空间形式的描述
- 关于数学科学：
 - 研究对象：现实世界
 - 研究内容：数量关系（量、量的关系、量的变化、量的变化与变化之间的关系、量的关系的变化等等）与形态
 - 语言表达：数字、符号、关系（等式、不等式等）、图形、表格等---数学是语言，是能被世界科学家认可并能读懂的科学语言
 - 数学的结构：量化关系，空间形式
- 数学的地位：数学是所有科学的基础和研究工具，所有科学研究的最终表达往往也是数学

理解描述与量化-----从物理到数学

- 回顾中学物理学过的物理量，体会数学的本质：
 - 力、重力、浮力、摩擦力、拉力、牵引力、动力、阻力等
 - 温度、热流密度、热通量、热量
 - 速度、加速度、位移、时间、通量、流量
 - 动量、冲量、动能、势能等等
- 给定一个情景，分析情景状态，给出状态描述的方式，即量的发现与量的定义；
- 备注：与文学描述不同，数学上量的定义必须是精准的、可度量的或者可计算的

- 如：
 - 文学描述：“借问酒家何处有？牧童遥指杏花村”
 - 数学描述：坐标、方向、距离
- 量化方法：把问题视为研究系统是一个好的做法，研究系统的特征，把系统的状态视为果、引起或导致系统状态发生变化的因素视为因，找出因、果之间的内在机理或依存关系，构建其量化关系表达
- 量化及量化关系的表达----回归问题的刻画与情景描述

通过案例理解数学建模-----CUMCM-2013B

- 背景 (略)
- 问题一：对于给定的来自同一页印刷文字文件的碎纸机破碎纸片（仅纵切），建立碎纸片拼接复原模型和算法，并针对附件1、附件2给出的中、英文各一页文件的碎片数据进行拼接复原。如果复原过程需要人工干预，请写出干预方式及干预的时间节点。复原结果以图片形式及表格形式表达
- 附件1，如右图，共18个图像文件，中文
- 附件2，同样为18个图像文件，纯英文

	文件类
000.bmp	143,638 21,166 BMP 文件
001.bmp	143,638 20,684 BMP 文件
002.bmp	143,638 21,336 BMP 文件
003.bmp	143,638 23,111 BMP 文件
004.bmp	143,638 20,930 BMP 文件
005.bmp	143,638 22,993 BMP 文件
006.bmp	143,638 16,720 BMP 文件
007.bmp	143,638 20,472 BMP 文件
008.bmp	143,638 18,973 BMP 文件
009.bmp	143,638 21,897 BMP 文件
010.bmp	143,638 22,342 BMP 文件
011.bmp	143,638 20,178 BMP 文件
012.bmp	143,638 22,241 BMP 文件
013.bmp	143,638 21,972 BMP 文件
014.bmp	143,638 22,475 BMP 文件
015.bmp	143,638 25,714 BMP 文件
016.bmp	143,638 21,679 BMP 文件
017.bmp	143,638 20,687 BMP 文件
018.bmp	143,638 22,662 BMP 文件

魂
柳
卖
小
郑
从
休
双
鬃
寒

北
奇
无
萧
萧
梅
笛
飞
留
国
似
道
待
扶
仙
时

问
扬
作
恨
相
非
到
更
把
新
苞
渐
消
凤
楼
斗

昼
嶙
二
月
一
重
衣
旧
如
吟
黛
流
日
去
阻
痴
纤
重
处
树
竹
说
俯
毛
真
过
谁
枕
会
京
余
茅
帽
范
无

.wa
skit
. N
wy
ilk]
the
ian
ot v
s m
eryt
l. A
to i
o liv
a v.
o h
ind
out
reas
tea
d s
anc
bo
life
lace
s cl
May
ntu
s. A
n. .

right
s. T
y a
ias
se f
an.
ord
D
de
g in
en i
give
n g
um
who
han
ee's
Ple
you
l di
ater

e fi
ce h
ns
ou
not
tch
l a

模型结构发现：分析、抽象、概化、假设、建模

• 问题的数学本质识别

- 图片---数字矩阵，均匀切割等同于矩阵阶数相同
- 由题意，碎纸图片是印刷文字，因此行间距、字间距应比较规范；其二，文字笔画大多是连续的、被切割开后，相邻两个碎片的边缘信息应比较接近，反映到图片格式，即相邻两个图片的边缘向量的灰度值应比较接近
- 图片拼接—左右边缘拼接
- 两张图片是否可以拼接---边缘灰度值向量相似程度，等价于两个数字矩阵的左右列向量的相似度问题

分析、抽象、概化



模型结构与表达----引入量、定义量及关系构建

记第 i 张碎纸片像素点矩阵为 $A^{(i)} = (a_{kl}^{(i)})_{K \times L}$

相似度定义:

空间距离
$$d_{ij} = \sum_{k=1}^K |a_{kL}^{(i)} - a_{k1}^{(j)}|, \quad d_{ii} = \infty$$

Jaccard距离
$$d_{ij} = \frac{m_2}{m_1 + m_2}$$

其中 m_1 为对应像素点均为 1 的配对数, m_2 为不配对数.

其它定义: 欧式距离, 夹角余弦, 相关系数等

模型构建与求解----TSP模型

0-1整数规划模型

$$\begin{aligned} \min \quad & \sum_{i,j=1}^n d_{ij}x_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^n x_{ij} = 1, \quad j = 1, 2, \dots, n \\ & \sum_{j=1}^n x_{ij} = 1, \quad i = 1, 2, \dots, n \\ & x_{ij} = 0 \text{ 或 } 1, \quad i, j = 1, 2, \dots, n \end{aligned}$$

TSP模型

$$\begin{aligned} \min \quad & \sum_{i,j=1}^n d_{ij}x_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^n x_{ij} = 1, \quad j = 1, 2, \dots, n \\ & \sum_{j=1}^n x_{ij} = 1, \quad i = 1, 2, \dots, n \\ & u_i - u_j + nx_{ij} \leq n - 1, \quad 2 \leq i \neq j \leq n \\ & x_{ij} = 0 \text{ 或 } 1, \quad i, j = 1, 2, \dots, n \\ & u_i \geq 0, \quad i = 2, 3, \dots, n \end{aligned}$$

求解：算法设计

- 数学软件的实现---Lingo
- 自行设计算法：打印版纸张有什么特点？如何识别边界？
- 算法的描述？
- 中文与英文字体有什么差异性？
- 如何理解人工干预？条件？方式？

求解结果的表达？

结果的正确性检验：

如何检验拼接是否正确？
如果不正确？如何调整？

更深入的研究（问题二、问题三）

问题二： 同时存在纵切、横切情形单面印刷中英文纸张如何拼接？模型、算法？

问题三： 同时存在纵切、横切的正反面印刷中英文纸张如何拼接？模型、算法？

思考： 中英文字体印刷的差异？

其他字体印刷是否模型也适用？

会不会出现无法拼接或拼接错误的情形？

印刷纸排版中存在图形、公式等情形会影响拼接吗？

为什么要学习数学建模？创新意识、研究能力训练

- 理解背景--- 文献检索与查询；
- 提出问题--- 问题的发现意识与能力；
- 如何发现问题涉及的量，如何找出量与量之间的关系（模型构建）
---文献检索、阅读理解能力、建模能力、模型的精准表达能力
- 模型求解---算法设计、编程计算、数学实验、模拟或仿真能力
- 假设、模型、结果检验---验证意识与检验能力
- 结果应用---可行性、可操作性
- 研究成果的展示---研究论文或研究报告（论文撰写能力）

- 创新人才的标志：发现问题、提出问题、解决问题；
- 数学建模教育的根本就在于强化问题的发现意识和解决问题的能力
- 数学建模没有标准答案，只有是否更可行，有利于改进、创新

三、数学建模问题类型



评价问题：如人才吸引力、雾霾、拥堵、幸福、污染、经济形势、疫情状态、军情、政情、国情等等



优化、决策问题：控制变量、目标函数、约束条件、求解算法



预测、控制问题：预测发展趋势、评价趋势状态；依据预测评价结果，控制发展态势

第二部分 如何学习数学建模



从赛题看常用的建模方法



- 空间与解析理论、线性代数、微积分
- 概率、统计（方差、回归、时间序列、相关分析、聚类或判别）分析
- 数学规划：线性、整数、0-1、非线性、多目标规划
- 图论与网络优化(最短路径、最小费用流等)
- 多因素综合评价
- 插值与拟合
- 差分与微分方程等基于机理分析的动力学问题建模方法
- 排队论、对策论、决策论、博弈论
- 其他：计算机智能算法（遗传算法、神经网络、模拟退火、蚁群算法、支持向量机等）、计算机仿真算法（如元胞自动机）

常用数学软件或编程语言

- **综合数学软件：**至少熟练掌握一种数学软件，如MATLAB, Mathematica, Maple等
- **专业软件：**
 - 运筹与优化软件 Lingo/Lindo
 - 统计软件：SAS,SPLUS,SPSS,R等
- **计算机编程语言：**C 或 Java, Python
- **其它文字编辑类：**
 - word\excel, WPS, Openoffice,Ctex等

如何学习数学建模？

- 尽可能**广、浅、新**（数学学习要**专、深**）
- 注重**思想为先、方法为后**，兼顾**算法与实验**
- 注重**问题导向**，抛开问题谈方法，是典型的建模误区
- 注重强化**模型结构化思维和问题思维、验证思维**
- 注重培养**文献意识**，以及科研规范（包括文献引用、论文写作规范等）
- 创新型人才的标志 --- 遇到问题，能想出解决办法。
- 没有问题，就没有创新；解决了问题但无法实施，等于没解决。建了模型，但没有验证，不能说你做对了

几点建议

- **参赛不是学习建模唯一目的，只是检验自己建模能力的一种手段。**
- **数学建模的学习是一个长期思考与学习的过程，建议提早学习相关选修课程，同时用建模的思维，思考与建模相关的专业课程，以及大学数学课程**
- **提早组队，并长期磨合**
- **有意识地加强科技论文写作的训练。**

第三部分 组队与赛前准备

一、组队要求



组队：3人一队，来自同一所学校，在校大学生

- **要求：**知识、能力相对全面
- **参赛能力要求：**
 - 建模能力
 - 算法设计、编程计算或数学实验能力
 - 论文写作能力

- **建议：**学科、专业、能力、性别 交叉

- **成功的关键：**
 - 力求3个人组成一个最强的队, 知识全面、能力互补;
 - 同心协力、共同攻关

二、组队并成功参赛关键



- 提前组队(建议7月底前完成);
- 适当分工 (建模、算法、写作), 有计划、针对性训练;
- 注意合作, 强化补位、配合、协作
- 有意识地撰写1-2篇建模论文, 训练论文写作能力和写作规范

- 了解竞赛规则 (违规内涵: 抄袭、雷同、相似度检测、泄露信息等), 以及参赛注册、报名方式,
- 论文提交方式、时间节点要求,
- 除学校官方群外, 赛前应退出所有与建模相关的互联网交流平台

- 适当分工, 充分讨论;
- 理解问题要透彻, 假设、建模方法、结果要论证
- 时间与进度安排要紧凑, 留出足够的时间处理论文提交事宜;

- 分析总结自己在竞赛过程中的得失, 找出自己的优势和缺失或不足
- 在后期的大学学习中放大自己的优势, 补上自己的短板,
- 让自己的科研人生更完美。

三、组队中常见的问题

- 为了参赛而参赛，临时组队
 - 提示一：一定**确认**你的队友**真实存在**，
 - 提示二：一定要有一段时间的**磨合**，确认你们性格、脾气、秉性、追求等可以合作
- 对竞赛规则缺乏必要的了解，导致违规而不自知。
 - 提示一：数学建模竞赛是学术类竞赛，竞赛规则是底线和不可逾越的红线
 - 提示二：明确哪些行为是违规行为，学会养成规则意识
- 依赖心理：侥幸、抱大腿思想，把成功寄予他人身上

三、组队中常见的问题

- 赛前磨合不够，彼此缺乏了解：
 - 缺乏默契，少了凝聚力、向心力，无法形成合力
 - 没有必要的分工与合作，竞赛时容易各行其是，互不认同，最终导致分裂；
- 分工过于明确，缺乏合作和配合意识
 - 参赛是团队的事，不是某一人做的事！
- 最好有一个队长，好协调、分工、写作进程安排。
 - 队长最好是在训练中形成的，可以自荐，也可以三个人共同认可

四、参赛队注册

- 注册信息填报要求:

姓名1 性别1 专业1 学号1 入学年份1 电话1 Email1

姓名2 性别2 专业2 学号2 入学年份2 电话2 Email2

姓名3 性别3 专业3 学号3 入学年份3 电话3 Email3

教师姓名 教师性别 教师职称 教师电话 教师Email

- 按规范填写好注册组队信息，统一用 Excel 表的形式发给学校指定建模负责人，汇总后，上传全国注册报名及论文提交系统

(<http://cumcm.cnki.net>)

- 个人信息在竞赛开始前可以修改

五、赛前训练

- **知识（方法、算法）学习：**数学建模课程、课外数学建模专题培训
- **数学模型求解算法与数学实验**
- **历年赛题研究**
- **建模优秀论文、命题人或专家组长撰写的论文研读**
- **竞赛论文写作**
- **竞赛规范**
- **队友分工、讨论、交叉合作，针对性训练**

第四部分 参赛注意事项

一、选题、审题

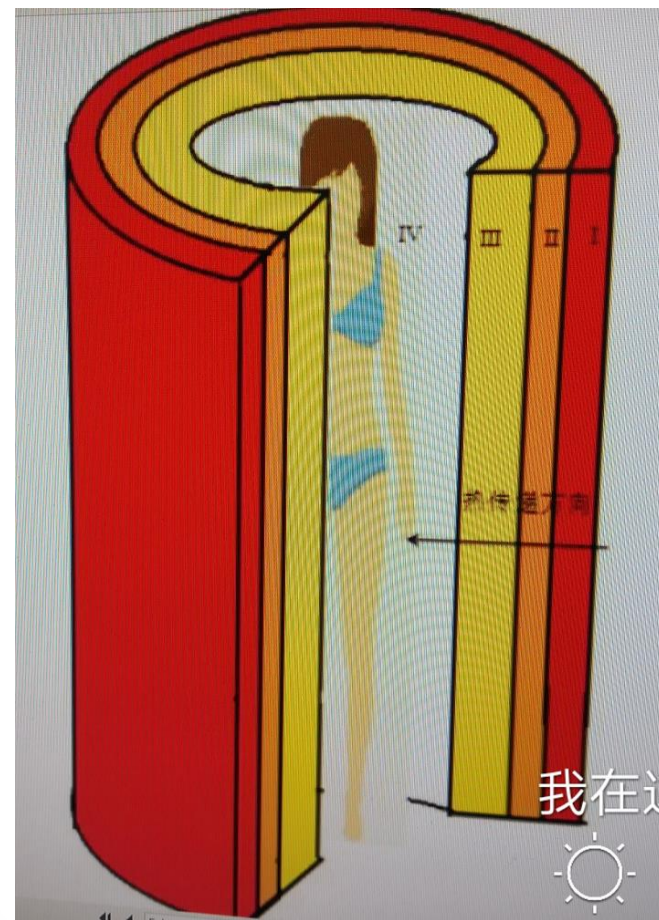
- 竞赛时间只有几天，时间有限，不宜在赛题选题上浪费太多的时间。
- **建议：**
 - 拿到赛题后，小组成员先分别独立阅读赛题，以2个小时左右阅读理解题意，提出选题意向
 - 队友讨论确定，尊重多数人的意见，或听取想法比较合理的一方。
 - 一旦确定就不再更改
- **仔细阅读赛题，确认理解透彻，重点把握：**
 - 问题的研究背景、建模目标或期望解决的问题
 - 注意把握关键词，不放过可能引导建模方向与建模目的的一切词汇。
 - 明确问题涉及的量及量化方式，分析量之间的相关或因果关系---确定建模方向和模型结构
 - 问题已具备的条件，包括数据、文献等
 - 建模结果的表达方式
 - 可能的验证问题和验证方式

以2018CUMCM-A题为例—问题理解

- 为什么会有热传导？机理是什么？空间、时间维数如何界定？
- 从单一介质热传导模型很容易得到热传导方程
- 该问题的瓶颈在于四个介质，介质跟介质之间联通，如何表达模型？
 - 四个耦合偏微分方程，形式结构一样，但空间位置和介质参数不一样
- 相邻两个介质之间存在一个交界面，如何定义传递条件？
- 外边界及初始条件如何定义？
- 介质参数如何识别？---参数辨识
- 复杂系统无法给出解的解析表达式，只能用数值解，只要查阅一下文献即可知道显式格式稳定性不好，应选隐式格式

从评审看学生理解的误区

- 过多地引用文献，把时间花在无用的工作上
- 不理解问题的共性和个性差异，无法针对不同的个性问题构建模型
- 把建模当美术，如右图
- 问题的机理是传热和吸热两个过程，不考虑热辐射等因素影响，没有理解热传导与热辐射差异
- 热传导方程大学物理基本都介绍过，但真正理解寥寥无几----- 大学物理教学存在问题
- 这类问题属于反问题，建议引起重视



二、时间、节点把握

- 参赛报名时间控制
- 开赛时间
- 竞赛结束时间
- MD5码提交时间
- 参赛论文及支撑材料提交时间

二、关于参赛作品与MD5码

- 参赛作品：
 - 参赛论文
 - 支撑材料
- MD5码是保存文件**创建信息**和文件**保存结构信息**的**检验信息**
 - 参赛队**赛前**可以通过报名注册系统 (<http://cumcm.cnki.net>) 下载MD5码生成客户端。
 - 下载完成后，绑定一台计算机（最好是自己的笔记本电脑）
 - 竞赛期间，可以测试上传MD5码
- 生成MD5码后的文件**不得再打开**
 - 防止因MD5异常无法上传作品

三、参赛作品上传

竞赛结束前，整理好论文和支撑材料



生成MD5码:参赛论文和支撑材料



在规定时间内，在全国系统上传MD5码



竞赛结束后，在规定时间内，上传与MD5码匹配的论文和支撑材料

备注

- 今年开始，参赛论文和支撑材料的**文件名不再统一要求**
- **参赛论文电子版：**
 - 不能包含**承诺书和编号专用页**在内
 - 文件格式只能用PDF或Word格式之一（建议用PDF格式），**不要压缩。**
 - 论文研究中所编的源程序代码应列在附录中（如程序过多，至少要附主要源程序代码，如确实没有编程，也应该在附录中明确注明“本文没有源程序”）
 - **备注：承诺书和编号专用页是推荐参加全国奖评审时要求的（全国为纸质版评审）；山东赛区推荐参加全国奖评审时，会由赛区组委会负责打印添加。**
- **支撑材料：**
 - 用于支撑参赛论文中模型、结果和结论的所有必要材料，通常应包含所有可运行的**源程序**、自己查阅并使用的数据和难以从公开渠道找到的相关资料等。
 - 使用WinRAR压缩为一个文件

四、论文上传常见问题

- 侥幸心理，能拖就拖；MD5码不能在规定时间内生成并上传，导致参赛失败。
- 生成MD5码后，不放心，又打开了参赛作品文件，导致文件存储物理信息异动，上传参赛作品时与前期上传的MD5码不匹配，导致传不上去。
- 支撑材料压缩包内的文件夹或文件名含有参赛编码信息
- 论文转存时（WORD—PDF），太捉急，导致公式、图片丢失，影响评审。
- 参赛作品内含承诺书、编号页，被判违规。

五、关于竞赛规范



二个明确

- 竞赛期间不得与除队友以外的任何人交流，尤其是指导教师！
- 不得与利用QQ空间、公众号、电话、邮件等社会、社交网络平台或个人进行与赛题相关内容的讨论



论文写作

- 在论文正文或附件中不得泄露参赛学校、队员姓名等信息
- 不得抄袭文献或网页等有关内容，如果需要引用，可在正文中标注引用文献编号，并在参考文献中列出。

六、主要存在的违规问题



论文上传及写作问题：

- 未能及时生成或上传MD5码，或上传后又打开了论文，导致MD5与上传文件不匹配，失去评奖机会
- 相似度过高问题：有可能被判抄袭或降低获奖等级



涉嫌信息泄露：论文含有承诺书等泄露学校和队员信息的内容



严重违规之一：网购、部分网购；雷同；抄袭（即使引用规范，只要抄袭过多亦构成违规）



严重违规之二：参与社会网络，包括商业网站、公司、个人等QQ群、微信公众号、直播、个人空间等赛题分析讨论

论文写作中的问题



- 论文写作规范：文字叙述、版面排版、公式符号、图形、表格不规范
- 参赛论文电子版**附录**中应附**源程序**代码（至少主程序代码），如确实没有，应在附录中注明“本文没有源程序”



文献引用不规范（应该引用没引，或列了参考文献但没有引用），论文中需要引用的文献内容包括：图形、数据、网页、图书、期刊、会议论文等等



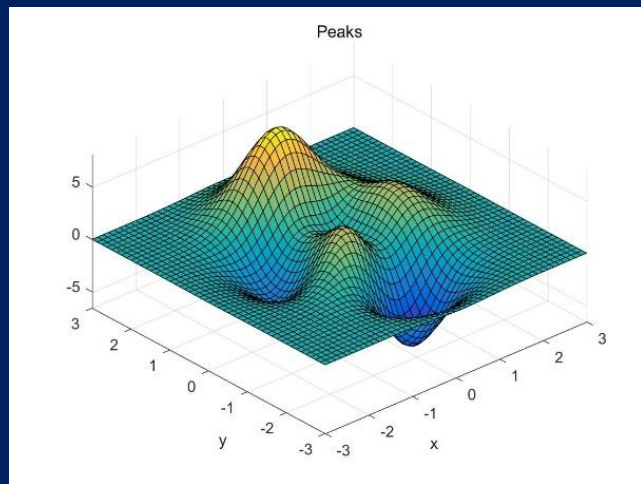
过多地抄袭文献或往年论文内容，造成**相似度检测**结果过高



摘要写作不符合评阅要求，失分过多

从评审看数学建模常见的误区

- 问题理解不透彻，方法导向，抛开问题谈方法或算法，失去对象
- 方法适用性理解把握不足
- 智能算法万能
- 全局优化意识缺失
- 算法意识与能力不足
- 三把斧现象：层次分析、模糊综合评价、灰色预测与评价
- 验证意识匮乏
- 文献引用意识与规范
- 功利意识过强，不惜违规，失去参赛意义。



常用网址



全国大学生数学建模竞赛组委会

• <http://www.mcm.edu.cn>



山东赛区官网

<http://www.dxsjs.sdu.edu.cn>



全国建模竞赛报名及论文提交平台

• <http://cumcm.cnki.net>



中国大学生在线—数学建模

• <http://www.univs.cn/zx/qkt/sxjm/>

预祝取得好成绩！