



New Model New World

2023 MCM/ICM

#MCM

卷首语

请记住，相比于将美赛称为“数学竞赛”，“数学作文竞赛”这个名字应当更恰当一些。参加美赛的过程和写高考作文一样，都是“戴着镣铐跳舞”，不能跑题、偏题，还必须有模型、有逻辑、有文采。

对于解题步骤/速度的要求

Day 1 上午

- 题目翻译：发题后，一人选择两道题，使用机器翻译+人工调整的方式快速译成中文，并依次向队友简述题目的内容是什么，三个人都简述完之后，进入下一步。这一步的时间不应超过 0.5 小时
- 选题：大方向是 A/C 任选一个。发题以后首先看 A，再看 C。如果 A 题太难以至于不可接受（出现高难度的最优化问题），或者队伍之中有一个人非常熟悉 C 题或 D 题的背景，则可以考虑换 C/D，其余时候，为了保证最大化获奖率，先从 A 开始做。此外，在这一步中，可以考虑先查一查文献，看看能搜到的文献数量。这一步的时间不应该超过 1.0 小时
- 初次读题：选题后，三个人应当一起读题，并且整理以下内容：本题到底有多少个大的模块；每个模块有多少个小问，每个小问的具体要求是什么。三个人必须独立地做这一步。做完后，三人应对照答案，防止有一个人理解出错。这一步的时间不应该超过 0.5 小时

*** 在初步阅读完题目后，可以开始解答第一问。第一问可能分为两个情况：（1）相对简单的一小问，作为后面所有问题的铺垫（例如，预测海水温度；建立功率曲线；预测大黄蜂的筑巢情况，等等）；（2）直奔主题，揭示题目的主旨（例如，设计量化策略、预测阿片类药物传播，等等）。但是，无论是哪种情况，作为题目的开端（在选手尚未进入状态时），第一问必然相当难做。此时，我们需要：

- 首先自己思考，对比数学建模书的目录和自己曾经在论文里面见过的模型，想想如果仅仅让自己来做，这一问**属于经典数学建模课程中的哪一类模型**，自己学过这个领域的哪些背景知识，自己会如何解决这个问题，以及**问题中需要考虑的最关键因素是什么**，此后每个人依次简述，回答这几个问题。**这一步的时间不应该超过 1 小时**
- 其次，开始查阅文献，主要由队伍中的两个人查阅文献，另一个人此时可以手写模型的方程试一下，或者继续思考该问题的关键因素。查阅文献**必须同时兼顾中文和英文（当然，大部分应该是英文）**，**必须使用从每篇文章后的参考文献中再找文献的方法**，数量要求充足，每个人必须搜到 3-5 篇可能有用的文献，并且，每篇文章需要用 30-50 字的一句话解释主要讲了什么。**这一步的时间不应超过 1.5 小时**

Day 1 下午

- 接第一天上午，如果没有搜完论文，就继续查阅论文（**最多再看半小时**）。此后一起看所有论文（这一步先看看论文中使用的模型名字），将所有论文中共有的、重要的模型抽出来（**到底哪些模型是重要的、是切合题意的、是表达了题目的核心要求和核心思想的，需要临场判断**），作为重点观察的模型。接下来，每个人开始研读重点模型，研读一段时间后，**将论文中的模型和自己思维中的模型结合起来**，三个人依次说明自己所建立的模型（至少说明自己想要在模型中考虑哪些因素，如何考虑），由一人汇总后作为第一问的总模型**这一步的时间不应超过 2 小时**
- 再次对照论文、数学建模教材，对建立的模型从**可行性、逻辑性、是否考虑了关键因素、是否全面**等方向进行检验，检验完成后，找相应数据、所有人一起处理相应数据，整合成可以直接使用的形式。**这一步的时间不应该超过 1 小时**
- 一个人对总模型编程实现。**同时，其他人必须开始重复***处的步骤，开始试图解决第二问**总的编程时间不能超过 2.5 小时

Day 1 晚上, Day 2, Day 3 上午

- 如果程序没有写好，继续写程序
- 写完程序后，**以上解决问题一的步骤不断重复，从而不断滚雪球一般地解决所有问题。****解决题目全部问题的时间限制是第三天上午**
- 具体的时间不再限制。因为每一问的难度都不同。

Day 3 下午, Day 4, Day 5 凌晨

现在开始写论文

- 首先，需要共享一份中文 LATEX 文件，一个人专门负责书写中文论文，并且敲出公式。但是这份中文文件不需要套用美赛模板，不需要插图（图直接通过微信群传递即可）。**论文写作期间，第一个人将只能在这篇中文论文上操作，不能看英语论文。**

- 其次，另外两个人之间需要共享一份英文文档，这份文档将作为英文论文的正式版。另外两个人需要首先将论文进行快速的翻译（推荐机器翻译），然后其中一人需要理解模型建立的逻辑，并对逻辑进行检查（论文中是否清晰地介绍了模型建立的过程？是否对公式的含义、公式中的变量和参数进行了详细的阐释？），并且修改文段中逻辑不合理的地方、解释不完整的地方；另一人需要对文段中的单词、语法进行检查，确保单词使用统一、语法正确，同时，这名队员应该检查论文中对于模型背景信息的介绍是否完整，可以适当地在论文中补充与背景有关的句子。
- 当正文部分完成后，所有人应该聚在一起，完成敏感性分析和模型优点、缺点和扩展的成分。
- 此后，英文论文应当被所有人一起检查，确保所有地方的逻辑链通顺、模型合理、解释详实、背景丰富。

现在开始写摘要（这时差不多就是第四天晚上上了）

- 三个人必须同时开始针对论文书写摘要，如果你可以直接写英文，就写英文摘要；否则写中文摘要。个人摘要时间不得超过 1 小时
- 三个人的摘要将汇集在一起，并且此时三个人一起来评判这份摘要，三份摘要中各取精华部分，汇集成一份摘要，翻译成英语，并且对语法进行检查！集体摘要时间不得超过 1.5 小时
- 三个人一起写小文章，和摘要一样，最后一起审核排版。小文章所用时间不得超过 2 小时

最终排版（包括小文章排版），检查，交卷

建模和论文书写中的注意事项

关于建模

- 【建模】建立的模型不能太复杂，尤其需要避免的情况是：假如问题的第一问只是一种引入，那么我们绝对不要在第一问就列出大量的方程/使用复杂的算法，这会使得问题难以求解
- 【建模】每一道题目都具有一些不能被忽略的核心内容，例如，在【2022 A 功率曲线】中，选手必须考虑各种阻力的影响，并且用合适的方式对赛道进行离散化；在【2022 C 交易策略】中，选手必须考虑如何控制投资中的风险（裁判并不会在乎你是否真正做出了在实盘中可以使用的投资策略，因为仅凭几个大学生，是无法开发出这样的策略的，但凡是很好地处理了投资风险的队伍，不管是使用标准差、协方差甚至是预测值的置信区间，都获得了不错的成绩）因此，对于一般的题目，我们必须在基本模型正确的情况下全面考虑细节，对于相对复杂（甚至业界难以处理）的题目，我们仅仅需要从一个简单的角度切入，但是必须抓住题目中的关键因素
- 【建模】模型中必须包含自己的创意，这可能是把传统模型（如元胞自动机等等）与题目结合；可能是提出全新的评价指标；可能是原来两个模型的组合，等等。不求每个模型都创新，但求每个模型用得合适，此外，所有人应该积极地将生活常识、（自己已经掌握的）各个不同领域和学科的知识与正在解决的问题想结合，只有如此，才会有创新的火花。

- **【建模，论文】**在建立模型的过程中，应该不断地对模型进行**反省**：这个模型对不对，是不是符合现实情况；如果不对，应该如何修改。这个过程中，必须使用一些手段来**验证模型的正误**：例如，将模型求解的结果与现实世界对比（例如，男性运动员的爆发力通常好于女性运动员，同样的赛道上，男性运动员的骑行时间也通常少于女性运动员，两种鱼的生活适宜温度不同，因此分布区域也应该不同，等等）；使用历史的数据/题目之外的数据（交易策略一题中，取熊市中的一段时期，观察策略在熊市中能否挣到钱）。
- **【建模】**模型要符合题目的精神。例如，在 B 题（离散型）中，评委更希望看到离散型的方法，因此元胞自动机相对于流体力学的模型是更好的做法；在 E 题（运筹学/网络科学）中，尽管题目是对于某些公司数据系统的评价，但是我们仍然可以使用网络科学的方法（搭建该公司的数据流网络），使用运筹学的方法（如何帮助该公司提升业务水平）
- **【建模】**建模型时一定要多角度地考虑：种树不能只种白桦；挖矿不能只挖铂金。分成多个类别讨论，将模型做的细致一些，更有利于获奖。
- **【建模】**在使用预测模型或者带有随机性的模型时，一定要好好处理模型中的不确定性：给出置信区间或者研究最好和最差的情形。

关于论文

- **【论文】**文章必须要对题目的每一个问题、每一个要求、每一句话（尤其是 Requirements 里面书写的全部内容）悉数做出完整的回应，只要落下一条，直接 S 奖。并且，这种回应绝对不能是“语文建模”的回应，必须有对应的数学模型和算法（除了某些题的最后一问，例如多人骑车、如何改变看法策略等历史遗留问题，目前没有队伍对这些问题做出完美的解答）
- **【论文】**应该详细地写明模型的假设，假设后应该附上验证。必要的时候，可以在验证中加上参考资料。写假设不是为了糊弄裁判，所有的假设必须要在后文中被使用。今年规定：假设从 Assumption 1 开始编号，后文中引用时直接引用假设的编号！这样可以保证每一条假设真的用到了！**如何在 LATEX 中给文字加上指向特定页面的超链接？**
- **【论文】**必须写 Literature Review，这个过程中必须充分参照前人文献，要图文并茂地写（具体写法模仿浩然即可）
- **【论文】**对于黑箱模型，最好不要直接写，而是需要做以下的事让评委认为你真的理解了这个模型：（1）通过探索性的数据分析，说明为什么你要选择这个算法（主要是机器学习算法）；（2）介绍这个算法的流程（画简要流程图/写伪代码），根据算法的流程来说明这个算法为什么合适？（3）如果使用机器学习算法，解释超参数的选择；如果使用统计学方法，请充分对前提条件进行验证（例如，线性回归模型里的异方差性）；在求解方程时，请写明你是使用哪个包求解的，用什么方法求解的（例如 ode45，使用了 Runge-Kutta 方法）
- **【论文】**文章开头说过，MCM 也是一种作文比赛，我们应当很好地将自己的模型传递给他人，**使得他人可以根据我们论文的思路推导模型、复现模型**。在每个模型中，解释你为何要选择这个方法？（这里允许跟国赛一样，对多种方法进行横向的比较，从而发现他们的优缺点；允许模型进行“进化”，即一开始先建立一个简单的模型，此后发现这个模型有缺点，于是再将模型改进为一个更新的模型）。还需要解释：公式的含义（为什么这个方程要这样列），变量的含义（这个字母表示什么）、参数是如何确定的（最难的一个点）。
- **【论文】**进行合适的可视化，而不只是漂亮的可视化。所有的可视化都是为了文章的内容服务，切勿因为追求美观而丢掉内容。此外，注意可视化过程中的基本准则，例如，条形的长

度比正方形/圆的面积更加容易识别！不要使用很多小图，而要思考：如何将很多有用的信息集成在一张大图上呈现给裁判。

- 【论文】最后的敏感性分析是文章的一部分，不是做给裁判看的。这一部分中，应当调整所有自定义的参数，并且调整的范围应该足够大，直到结果出现变化为止。此后，应该基于敏感性分析的结果，重新表达模型的结论，并进一步分析（例如，分析模型在多种环境中的适用性，等等）（建议这一段至少 1 页纸）
- 【论文】美赛不避讳优缺点的分析，优点和缺点都要认真写，以及可以在模型的发展方向这一节中写下前面因为过于复杂而放弃的因素。
- 【论文】小文章和摘要同等重要。在写这两个文章的时候，除了把文章要求的各部分写好之外，还应该注意题目的**一个总要求**是什么，比如亚马逊的数据分析题目的总要求是给公司提供销售建议，那么小文章种不仅要包含题目每一步的结果，还要提供相应的建议。最好在写小文章时将自己代入现实世界中的一个身份。
- 【论文】图片下面必须三步走：**按图索骥**：图上的曲线、点是什么样子，就描述什么，描述它们的分布规律、变化趋势等等；**由果溯因**：描述为什么会这样，是什么特性、什么机制使得图像呈现了这样的形状；**联系现实，交叉验证**：从图像间接地验证这个模型是正确的。

做题的意识：从以往的论文中，我们到底学到了什么

首先，我们应该明确，近年来最卷的两份论文是：2021C, 2021D

- 2019C：处理地理区域数据时，如果要处理的区域太多，或者每个区域上有太多缺失值，可以考虑将区域聚类成更大的区域再处理；此题中的预测是一个非常罕见的预测问题——历史数据少，且阿片类药物的传播与地理位置有关。因此，在预测时，大部分队伍充分利用了地区之间的相似性。此外，还有一队给出了一个药物溯源算法。
- 2020 A：本题也涉及到了地图，而且和上面那个题不太一样。上面的题给出了行政区划，而本题中需要自行将地图离散化。海温的预测最好用的还是统计学方法。后面的鱼类迁徙部分展示了元胞自动机的活用——鱼类向着水温梯度的反方向迁徙。此外，本题也可以使用基于多智能体的模型。对于收益的估算，一些人选择了每个格子的收益随着距离指数递减，一些人选择了画出渔船可以到达的范围，等等。
- 2020B：作为一个离散化的题目，使用元胞自动机模型研究沙堡的倒塌是合适的方法。此外，一些论文中提出的“优胜劣汰”的沙堡进化方法也是很好的尝试。
- 2020C：一道简单的数据科学题目。对于文本的处理不必繁杂，只要能从中识别出情感/抽出重要的情感词，并依据此对评论进行打分即可（只是为了抽出一个特征）。之后，使用多个信息综合，给商品进行打分，打分时，要注意一段时间前的评论也能影响商品当前的声誉。在预测商品未来的变化趋势时，大部分队伍使用了相当传统的预测方法。在评价商品的成功和失败时，有些队伍选择了模糊的评价方法，有些队伍使用了使用 Logistic 函数映射后设置一个阈值（硬阈值或者软阈值）
- 2020D：此题非常灵活，这体现在几个方面：各个队伍对于“相似度”的定义上、衡量流派间相互影响所使用的数学模型上，以及后面识别音乐突变点、识别各个流派领军人物的过程上。本题大幅度体现了数学建模所要求的原创性。
- 2021A：典型的先建模再实验的问题。在建模阶段，需要清晰地考虑某种真菌与其他真菌竞争、某种真菌对树木的分解过程，使用恰当的微分方程描述，以及，描述某种真菌生长的速率是如何受到外界环境限制的（这个限制可以加在内禀增长率上，也可以加在 Logistic 函数

的衰减项上)。在实验中,可以考虑使用函数模拟温度和湿度的变化,也可以拿到实际数据。

- 2022A: 本题比较考验查阅文献的能力。首先必须查到 3 参数功率曲线,之后建立一个合理的模型描述能量的消耗与补充(不适合直接对总能量消耗进行硬性约束),之后建立物理模型和最优化模型即可。本题自由度很高,但是也很难。
- 2022C: 控制风险,控制风险,还是控制风险!

数据查阅/文章翻译等等各种网站

卷尾语