2020年美赛竞赛心得

崔正达 张文心 张若愚

今年的美赛，我们小组获得了美赛的F奖，在此记录一下建模比赛过程之中的一些心得，希望能够帮助到之后的同学。

在准备阶段（比赛开始前一周），我们主要参考了往届同学的建议，学习了之前几届的O奖文章，搜集了一些常用的建模方法。根据我们自己的分工，在准备上还会有一些侧重：比如负责写作的同学提前熟悉了Latex，负责建模的同学熟悉了美赛可能是用到的一些模型和算法，负责编程的同学熟悉了一些package的使用要领。在这一阶段，我们的建议是，要根据自己的情况做出合理的分工，把每个人的优势都发挥出来；同时也不能掉以轻心，不做准备，这样开始之后可能陷入慌张与忙乱。

在比赛开始后，我们首先进行选题。选题的原则有两个：第一是看团队的兴趣在什么地方，没有兴趣的议题不建议去做；第二看数据的可获得性，比赛时间并不长，全部花在找数据上会很吃亏。在经过短暂的讨论后，我们在拿到题目2个小时之内就确定了题目，进入建模阶段。因为大家对于网络模型有基本了解，我们选择的是D题，没有选择对专业背景和物理相关知识要求较高的A题。总结一下我们使用的编程工具是python和matlab，用到的知识：最小支撑树，泊松回归，最短路径，图的相似度，pagerank，optimization model，Markov process，simulated annealing.因为疫情的关系，大家都待在自己家里。于是我们开了腾讯会议，从早上九点到晚上十点左右随时沟通进展和想法。我们组不算很肝，前几天基本11点收工，最后一天改论文改到早上5点左右，写文章的同学共享屏幕，大家一起挑错（typo，缺少空格等）。

建模阶段绝不是一帆风顺的，我们其实走了一些弯路。第一天我们确定了一个思路，但没有理清整个模型，只是对几个因素跑了线性回归。但在第二天的时候，我们觉得目前的模型太单薄了，就停下手来认真地讨论了一个下午和半个晚上，最后讨论出了一个自觉比较完整的模型。此时离比赛结束只有两天了，时间相当紧迫。在这里我觉得有一个教训就是：一定要确定好模型思路再开工，边做边想不太现实，宁愿之前多花一些时间，也不要之后回过头再讨论，不然时间会来不及。如果不确定自己的想法是否能实现，可以简单的做一些探索性数据分析，看看这样想的效果好不好，之后再正式确定方向。

在建模完成之后，我们开始了编程与论文撰写。编程的主要工作其实是数据的可视化，负责编程的同学可以多熟悉熟悉可视化的绘图包（如seaborn）等。我们这道题的数据量不算特别大，因此感觉更看中小组的想法是否新颖、可视化的效果是否出众。论文撰写方面，我们的安排是两位同学只用负责自己的部分，另外一位同学负责整合与格式。使用Latex协作是很繁琐的，因此负责写作的同学最好不要换，由自己负责到底，不然会拖慢进度。在全部写作完成后，我们3个人将论文从头到尾全部过了一遍，修改了小错误与笔误。后半程的过程就比较顺利了，我们一直写到最后一天早上六点，完成了报告。在这里还有一个建议：三个人有明确的分工：写作、建模、编程是很好的，不过最好大家对自己负责的方面以外的内容也有一些了解，这样讨论起来会更舒服，大家也可以互相帮助。

总的来说，我们组三个人都是第一次参加美赛，在比赛之前也没有做太多的准备。在参赛之前拜读了一些历年O奖的作品，参考其他同学的参赛建议，我们认为好看的图、自洽的数学模型以及清楚合理的概念定义或指标设计较为重要。好看的图不单单指报告中出现的数据分析、结果展示，也包括在介绍的部分放一个总结性质的框架图。

除此之外，还有几点具体的建议：

1、成果可视化！！美赛的重点是数据分析以及结果呈现，复杂的建模和算法分析可以起到加分，但是对大多数同学可能不是必备的。论文的完成度可能是影响得奖的更关键因素。

2、拿到题目多讨论，确保组内同学理解一致。

3、初期多查一些相关研究，并brainstorm。

4、不要太纠结科学性严谨性，言之有理可。（个人看法，毕竟只有4天的时间）

5、推荐知识、概念迁移，例如我们使用pagerank这个定义网页重要度的概念去描述球队中各个球员的重要度。

6、在总结模型优缺点的时候，可以参考一下历年优秀文章中使用过的评估角度，例如鲁棒性，某些概念定义的主观性，使用的方法的潜在缺陷（例如遗传算法仅得出近似解）等等。

写文章的同学可以早一点开始写，这样可以预留较多的时间修改润色，内心也比较笃定。