# Vast Challenge 2019 总结感想

2019年7月13日14:57，我们终于结束了为期两个半月的Vast Challenge MC2相关工作，完成了初步调研、数据分析、系统开发、答案撰写、制图与视频剪辑的工作。下面我将从工作内容、注意事项、个人感悟三个方面进行总结。

## 工作内容

**初步调研：**调研工作奠定你对题目内容与故事发展的基本认知，是你在后续工作中进行决策判断的重要依据之一。

**初步调研占比：10%**

**初步调研分类：**

**1）背景调研**：提取背景中的关键词，通过Google学术、百度学术、中国知网等平台搜索相关论文；通过书籍、视频、博客、各类问答平台搜集相关资料，将调研到的信息与背景相匹配，提取和预测该背景下的事件。例如，我们的题目与“核电站”、“地震”、“冷却剂污染”、“辐射监测”等关键词有关，我们需要了解的知识初步有a)核电站的结构、工作原理、题目中描述的核电站属于什么类型的核电站、这类核电站有什么优缺点、有什么相关的辐射泄漏事件等；b)地震的基本类型、地震的震级与对应自然灾害、次生灾害、地震引起的核辐射事件……诸如此类。

**2）技术调研：**结合所给数据对题目内容进行抽象，通过调研论文、书籍、博客等，得到解答题目所必需的技术支持。例如，我们的题目与“数据的不确定性”、“不确定性可视化”、“时序数据处理”等技术有关，通过调研，我们需要知道：什么是数据的不确定性、不确定性的分类、不确定性表示方法、哪些方法适合我们的题目等问题。

**数据分析：**数据分析是所有工作的重中之重，是正确解题的行动指南，是决定解题质量的关键前提。

**数据分析占比：35%**

**数据分析分类：**

**1）数据咀嚼：**通过Excel、Tableau对每一个题目的数据进行一个初步的“阅读”与分析，建立数据间的关联，以更好的为某个题目答题提供佐证。“咀嚼”的过程是繁琐与枯燥的，需要广泛的阅读、归纳、分类、整理等，需要组员共同参与，同步撰写数据分析文档。

**2）数据预处理：**包括去噪、异常检测、空值处理、模式提取等。

**3）数据可视化：**a)对每个题目的数据，利用Tableau进行初步的数据可视化分析；b)对所要答题的数据，利用Tableau进行详细的数据可视化分析，尝试在不写代码的情况下，尽可能多的答题。

**4）数据二次分析：**该分析过程在系统开发之后，利用系统对数据进行二次分析，用以验证前期分析的正确性，发现分析的不一致性。

**系统开发：**你需要知道，除去数据分析步骤中已经解决的，你还需要解决什么。

**系统开发占比：20%**

* **选定可视化方案：**切忌花里胡哨，要选择最具有说服力和解释性的可视化方案。一切以目标驱动，在准确把握目标的基础上再去谈“酷”与“美”的问题。
* **算法设计：**对比赛的问题进行抽象，明确需要解决的问题，为具体问题提供适合的算法以解决问题。大多数情况下，需要对现有算法进行适当改进。例如，数据聚类的距离度量函数，需要根据具体问题进行更换或重写。另外，算法的准确度与速度之间的取舍要根据具体问题而定，注意计算过程中的剪枝与聚合处理。
* **系统界面设计：**注意每一个功能模块的位置的合理性。系统的原始配色以灰色（#ccc）为主、注意每一处可视化颜色编码的含义，整体系统配色尽量不超过5种。
* **系统交互方案：**复杂的联动交互在实现上虽然煞费苦心，但是能成功提升用户的友好服务。良好的系统交互方案，有助于理清答题思路。
* **系统优化：**二次数据分析后的查缺补漏。

**答案撰写：（略）**

**详情参考秦雅琪的总结感想**

**答案撰写占比：30%**

**制图与视频剪辑：**提升美感的加分步骤，显示答题态度的一项依据。

**制图与视频剪辑占比：5%**

**制图与视频剪辑：**

* 制图与视频剪辑建议两个人负责
* 制图需避免色差问题，注意新建文件的颜色编码选项与图片保存格式（建议.jpg）
* 分辨率要求：保存为质量中等的144pi即可
* 制图与视频剪辑避免花里胡哨，以简约风格为主，颜色选择不宜过多

## 注意事项

* 定期与老师沟通交流，在老师意见为主的情况下自由发挥。老师能够把握问题的大方向，提出重要见解。所以对于老师的意见，一定要先执行一遍，不要先自己判断对错。如果有条件的话，紧跟老师的思路。
* 可视化方法千千万万，到底哪个方法才是最适合放到你的系统中的。这需要做一个权衡，这也是一个不断尝试和认知的过程。“适合”这个词意味深长，需要综合很多方面考虑。最后留在界面上的，往往不是最“新潮”的，最“酷”的，最“动感”的方法，但确是最“适合”的。
* 不是所有的代码都能用上，也不要想方设法的硬要把某些代码用在哪些方面。我们最后的系统，只用了一半的代码。剩下一半的代码都在分析与讨论中注释掉了，这些代码不够具有解释性，或不够说明问题，或速度太慢，或偏差太大。
* 注意组内沟通，关键性问题一定要一起探讨，一致通过，减少答题偏差。尤其注意系统开发与答案撰写的协调与统一。
* 从历届作品的回答方式、技术路线、系统设计、制图与视频制作风格等方面汲取灵感，在符合规范的前提下进行创新。
* 团队会讨论出很多解题的方面，很多不同角度的内容。所以，你们需要决定哪些内容需要写在答卷上，哪些需要舍弃。这个工作至关重要，因为很有可能，你们决定舍弃的一些方面才是真正的答案。

## 个人感悟

* 这次比赛全程由赵老师指导，非常感谢赵老师能在繁忙的工作中抽出很多时间来跟我们讨论。这些讨论都是很珍贵且有价值的。在每次的讨论中，我都对比赛的问题有了更多的不同方面的思考。这次比赛对我来说是一个全新的体验。
* 团队成员之间相互帮助，共同奋斗。两个半月以来，我们一遍遍的梳理问题，一遍遍的尝试解题，一遍遍的核查解题步骤，相互监督，相互促进。最后一个星期，大家一同熬夜，一同工作，一同吃饭，忙碌并快乐着。
* 比赛的最大收获就是我能清楚的知道自己的不足是什么。每一个阶段的不足，长期的不足，需要改进的地方，在这次比赛中浮现了出来。
* 这次比赛是对能力的一次考验，通过这几个月的不懈努力，我个人在代码方面、解决问题方面、团队协作方面、沟通方面都取得了一定的进步，希望自己在今后能更好。

蒋昊瑾

2019/7/19