

模型总结与案例分析

郁孜欣

目录

- 题型分类
- 基本模型简介
- 例题与对应模型整理
- 题目形式介绍
- 具体案例分析

* 具体介绍题型与对应模型的原因

1. 建模比赛的题型固定
2. 对应题型有“固定”解法

HiMCM题型分类



基础模型整理

预测	评价	优化	统计
微分/差分方程	AHP 层次分析法	线性规划	主成分分析
时间序列模型	TOPSIS 优劣解距离法	图论	方差分析
回归分析预测	熵权法	排队论	k均值聚类
马尔可夫链	模糊综合评价	现代优化算法	回归模型

* 为什么不推荐神经网络

1. 数据量限制，基本100%过拟合
2. 不符合建模比赛的原意

预测模型

2022 HiMCM Problem B: CO₂ and Global Warming

1. Do you agree with CO₂ level claims? Use **CO₂ Data Set 1** to analyze CO₂ changes.

- Do you agree that the March 2004 increase of CO₂ resulted in a larger increase than observed over any previous 10-year period? Why or why not?
- Fit various (more than one) mathematical models to the data to describe past, and predict future, concentration levels of CO₂ in the atmosphere.
- Use each of your models to predict the CO₂ concentrations in the atmosphere in the year 2100. Do any of your models agree with claims and predictions that the CO₂ concentration level will reach 685 ppm by 2050? If not by 2050, when do your models predict the concentration of CO₂ reaching 685 ppm?
- Which model do you consider most accurate? Why?

2. What's the relationship between temperature and CO₂? Many scientists think that there is a relationship between warming global temperatures and the concentration of CO₂ in the atmosphere. Use your work in part 1 and **Temps Data Set 2** to assist in your comparison of land-ocean temperatures and CO₂ concentration levels.

- Build a model to predict future land-ocean temperatures changes. When does your model predict the average land-ocean temperature will change by 1.25°C, 1.50°C, and 2°C compared to the base period of 1951-1980?
- Build a model to analyze the relationship (if any) between CO₂ concentrations and land-ocean temperatures since 1959. Explain the relationship or justify that there is no relationship.
- Extend your model from part 2.b. into the future. How far into the future is your model reliable? What concerns, if any, do you have with your model's ability to predict future CO₂ concentration levels and/or land-ocean temperatures?

1. [CO₂与年份]

分析CO₂的变化，预测未来的CO₂浓度，应用不同的模型对比结果

2. [CO₂与温度]

CO₂的增长趋势（预测）

CO₂与温度的相关性分析

3. 敏感性分析，模型优劣势

一道十分简明，简单，直白的预测题

时间序列预测：ARIMA，ARMA

对比模型：应用回归模型

相关性分析：Pearson系数（数据正态分布），Spearman系数（数据非正态分布）

敏感性分析：调整ARIMA中的参数，绘图比较；分析模型的优劣势

(如果你想让内容更丰满)

另一种对比的**Approach**: 考虑非机器学习的模型, 查论文, 建模

$$\Delta T = 1.66 \ln\left(\frac{C}{C_0}\right)$$

$$T_n = T_{n-1} + 1.66 \ln \frac{C_n}{C_{n-1}}$$

结合参数的调整, 更方便做对比分析, *** Holistic considerations

十分简单的公式与步骤,
都可以包含在**solution paper**的内容中

评价模型

2020 HiMCM Problem A: The Best Summer Job

1. What factors should high school students who are looking for a summer job consider? List and describe the various factors your team identifies. Note that factors may be quantitative or qualitative, constant or variable, and *deterministic* or *probabilistic*. Be sure to include units as appropriate.
2. Use your factors to develop a model or algorithm (or set of models/algorithms) for a high school student to use to evaluate their summer job options based on their own situation and preferences as inputs to your model.
3. Test your model with at least ten *fictional* persons that you create with reasonable data. Explain your development of these fictional persons and the data you chose. Analyze the results of the application of your model on these persons.

1. 识别影响summer job选择的因素

是deterministic (processes that have only one outcome) 还是probabilistic (processes based on the theory of probability)?

识别所有可以量化的指标，建立变量

每个变量不是单独的字母，需要经过一些考虑与计算

2. 利用所有的变量，建立模型

最简单的方案TOPSIS + AHP，可以利用刚刚介绍过的别的评价模型适当对比

3. 建立十个虚拟的人物测试模型 = 文字描述 + 每个变量输入适当的数据 + justification

* AHP的common drawback

过于主观!!!

因为重要性矩阵是人为评价的

优化模型

2023 IMMC International: Using Land: A valuable resource

1. Determine a **quantitative** decision **metric** that defines “best” so the decision makers can feel confident in their final use of the land. The metric should consider short- and long-term benefits and costs.
2. Choose at least two of the options listed above and determine the values of those options in your “best” metric. You may need to find data. You will need to decide which factors to focus and why and make assumptions including on values in order to use your metric. Explain and defend your values or use a range of values to better understand the effects and sensitivities of your assumptions.
3. In October 2022, it was announced that Micron Technology, Inc. will build a very large semiconductor fabrication facility (fab) in Clay, NY, USA, a town just north of Syracuse, NY. Soon after the announcement, it was reported by news outlets that “If fully built, the fabs could employ up to 9,000 people making an average of \$100,000 each year. They would create some 40,000 other jobs among suppliers, construction firms and other businesses. the new plant will directly support 9,000 jobs and create nearly 40,000 additional jobs.”^[2] How will the new fab impact your metric? Re-evaluate the options you identified in the previous question using your “best” metric.
You may *also* consider alternative options for using the land (either an additional item from the initial list or any other option not listed above) and evaluate that option using your “best” metric; justify your decision to consider and evaluate another option.

1. [Determining decision metrics]

什么定义了好的land use?

参考： 建立三个大的门类

Societal, Economic factors and Sustainability

如何为你建立的factor提供justification?

比如经济学中外部性的概念， 经济并不是评估发展的唯一因素。

Societal

Recreational and Touristic
Provision of Work, Human Health and
Recreation, Cultural
Cultural, Self-development
Recreational, Political Sovereignty
Leisure and Recreational

Economic

Residential, Production
Residential and Land-independent
Production, Land Based Production,
Transport
Production, Consumption,
Distribution
Primary Production, Manufacturing
Activities, Life
Economic, Residential

Environmental

Environmental
Provision of Abiotic Resources,
Support and Provision of Biotic
Resources, Maintenance of
Ecosystem Processes
Fundamental, Regulation
Strategic Resources, Ecosystem
Protection

2. 细化你建立的指标，提供计算公式

找数据，带入数据计算

敏感性分析：例如，随着时间的变化

3. 一个新建立的工厂会如何影响你的评估？

很简单，同样在你建立指标的维度上分析影响

比如它带来的经济效益，以及跟更多job opportunities减少了失业率，但它同时也污染了环境， etc

一个小总结

1. 为什么不介绍统计模型: statistics is everywhere!

2. 为你做的所有的工作提供充分的 justification

你的工作的前提是什么?

为什么要这样设置变量, 基于什么考量

3. 充分利用互联网上的source

Source, Data, Formulas

4. 建立由基础模型到推广的逻辑

一道建模题目会包含什么

- Background information and Requirements
- 第一道题：通常是一些简单的变量/指标建立
- 第二道题：模型的细化与推广，数值带入，敏感性分析
- 第三道题：一般会介绍一个新的影响因素，或者是提供一个更细致的背景。基于已经建立的指标分析对模型产生的影响即可
- 第四道题：Non-technical Article/Poster
- 发挥你的美术功底！
- 数据题会有data，新奇的名词会给definition

一个小建议

不要在比赛过程中一直看官方提供的问题描述

太冗长，太累赘，容易漏掉信息

第一天可以花一整天仔细审题，规划后面要做的工作的方向
(漏题是很致)

题目形式介绍

Requirements

Your team decides to use Lake Mead to investigate and learn more about drought impact on water reservoirs, and to consider the recycling of wastewater as a solution to water shortages. As you begin your investigation, we provide the following information:

- The level of Lake Mead is measured by *elevation* of the water surface in feet above *mean sea level*.
- The area of Lake Mead is measured in acres.
- The volume of water in Lake Mead is measured in acre-feet.

Table 1 shows the relationship of elevation, area, and volume of Lake Mead as calculated by the Bureau of Reclamation in 2010^[4].

Elevation (feet)	Area of Lake (acres)	Volume of Lake (acre-feet)
1229.0	159,866	29,686,054
1219.6	152,828	28,229,730
1050.0	73,615	10,217,399
895.0	30,084	2,576,395

Table 1. Area and Volume of Lake Mead by Elevation Level

1. Lake Mead Volume. The amount of water in Lake Mead is a function of *inflow*, *outflow*, and loss. The Colorado River provides over 96% of the inflow with additional water coming from three other tributaries, as well as direct precipitation on the lake surface. Outflow occurs through releasing water (for example, through a dam) and consumption directly from the lake. Loss occurs through evaporation.

2021 HiMCM Problem B: Tackling the Drought

a. Identify and describe factors that impact inflow, outflow, and loss in Lake Mead. Discuss the relationship of these factors and their relative influences on the volume and water level of Lake Mead.

b. Lake Mead has a very irregular shape and varying depths throughout. Consider how you would verify the elevation, area, and volume relationships shown in Table 1. Describe what information and data you would need and discuss how you would use mathematics to calculate these measures. Note: You are NOT required to obtain these data or calculate the measures.

2. Lake Mead Water Level. The two problem data files provide information about Lake Mead water levels given by elevation in feet above mean sea level.

a. Consider the data provided. Briefly discuss overall patterns in the historical data for Lake Mead water levels. Define your criteria for drought periods and identify the beginnings and ends of periods of drought. Comment on how the most recent drought period compares to earlier ones.

b. Develop two models for the water level in Lake Mead as a function of the year. Use each model to predict the water level in Lake Mead in the years 2025, 2030, and 2050. Compare and evaluate your two models and their predictions.

- Model 1: Consider data from only the most recent drought period and assume the most recent drought period’s pattern continues.
- Model 2: Use water level data from 2005 - 2020 and assume this period’s pattern continues.

3. Based on your models and water level predictions in Part 2, address the impact on future water usage demands and consider if the recycling of wastewater could make up all or a part of any future shortfalls.

a. Identify and describe the factors you would include in a plan to recycle wastewater. Consider the decisions local leaders would need to make and the priorities they might set that would impact your plan.

b. Describe your plan and how you would measure the impact of implementing your plan.

4. Write a one-page non-technical news article reporting the key takeaways and recommendations from your investigation.

Your PDF solution of no more than 25 total pages should include:

- One-page Summary Sheet.
- Table of Contents.
- Your complete solution.
- One-page Article.
- References list.

Note: The HiMCM Contest now has a 25-page limit. All aspects of your submission count toward the 25-page limit (Summary Sheet, Table of Contents, Reference List, and any Appendices).

完整流程 Tackling the Drought

1. Lake Mead 的体积估量，分析模型里的factor对于体积影响的程度

- a. $V_{total}(t) = V_{total}(t - 1) + V_{input}(t) - V_{output}(t) - V_{loss}(t)$
- b. 识别具体影响公式中factor的因素
- c. 相关性系数分析/层次分析法评估每个factor的权重

2. 分析elevation， area of lake 以及 Lake Mead 提及的关系

Elevation (feet)	Area of Lake (acres)	Volume of Lake (acre-feet)
1229.0	159,866	29,686,054
1219.6	152,828	28,229,730
1050.0	73,615	10,217,399
895.0	30,084	2,576,395

Table 1. Area and Volume of Lake Mead by Elevation Level

$$V(L) = \int_{L_0}^L A(l)dl$$

$$a(L) = Al^2 + BL + C$$

解一下这个十分简单的微分方程，得到V关于elevation的表达式

四个数据点带入， 最小二乘法估计a, b, c参数 （最小二乘法的公式也建议放在paper里）

第二题： 不详细分析了，就是一个**时间序列+敏感性分析**

第三题： 建立一个模型分析 Recycling Plan

基于water treatment system中的facility，根据具体条件做**线性规划**

第四题： 好看。

谢谢大家！