

2025 年第十届“数维杯”大学生

数学建模挑战赛论文

题目 马拉松窗口期筛选与赛道优化研究

摘要

近年来,随着全民健身的兴起,马拉松赛事的高质量发展需求也在不断提高。如何科学规划赛事窗口期、优化赛道设计、提升参赛体验和奖励机制,成为赛事组织者亟需解决的问题。本文围绕中国主要城市的马拉松赛事,针对四个核心问题展开分析,建立了一套系统的数学建模,旨在为赛事科学决策提供数据支持与量化依据。

对于问题一,我们基于气象适应性、城市承载能力等要素,构建了综合评价模型,用以科学筛选出中国主要城市的马拉松窗口期和合理安排赛事时间、规模以及频率。模型中采用熵权法-TOPSIS 组合模型方法进行多指标排序分析,计算出各个城市综合得分,筛选出北京、广东等地区凭借气候条件好和承载能力强等方面划分出不同赛事规模和频次(例如北京-小型-高频),用以提供建议。

对于问题二的子问题 2.1: 筛选最优起点-终点组合,满足距离、住宿容量、交通便利性要求,我们基于评价函数设计和遗传算法(GA)生成候选路径,过滤满足约束的组合,得出满足的起点终点路线。

对于问题二.2: 设计闭合赛道,最大化餐饮增益,同时满足坡度约束和补给站设置要求。我们基于改进蚁群算法动态调整信息素,优先选择高增益路径。使用 DEM 计算道路坡度(GDAL 或 Python 脚本),剔除坡度 $>5\%$ 的路段,生成可行路径网络。

对于问题三,基于问题二的赛道,综合考虑树荫覆盖,交通影响,坡度影响,我们使用多目标遗传算法生成 5 条候选路线,综合优化赛道,保证了比赛选手的舒适性,生成了可取的赛道选择。

对于问题四,采用 XGBoost-RankNet 模型预测年龄分组竞争激烈程度,结合聚类分析设计专属奖牌、编号限量徽章或定制化纪念品,差异化兑换阈值覆盖前 30%选手。模型的求解可为城市马拉松规划提供科学依据。

关键词: 熵权法-TOPSIS 组合模型, 遗传算法, 改进蚁群算法, 多目标遗传算法, XGBoost-RankNet 模型

目 录

一、问题重述 (1)

二、问题分析 (2)

三、模型假设 (3)

四、定义与符号说明 (4)

五、模型的建立与求解 (5)

 5.1 问题 1 的模型..... (6)

 5.1.1 模型的分析..... (7)

 5.1.2 模型建立..... (8)

 5.1.3 求解与结论..... (9)

 5.2 问题 2 的模型..... (10)

 5.2.1 模型的建立..... (11)

 5.2.2 模型建立..... (12)

 5.2.3 求解与结论..... (13)

 5.3 问题 3 的模型..... (14)

 5.3.1 模型的建立..... (15)

 5.3.2 模型建立..... (16)

 5.3.3 求解与结论..... (17)

 5.2 问题 4 的模型..... (18)

 5.4.1 模型的建立..... (19)

 5.4.2 模型建立..... (20)

 5.4.3 求解与结论..... (21)

六、模型的评价及优化 (22)

 6.1 误差分析..... (23)

 6.2 模型的优点..... (24)

 6.3 模型的缺点..... (25)

 6.4 模型推广..... (26)

参考文献 (27)

一、问题重述

马拉松赛事作为城市经济发展的“金色跑道”，其规划需兼顾经济效益、参赛体验与城市管理。结合题目要求，以西安市为例，结合西安的基础数据，提出具体的多目标优化模型。具体问题如下：

1. 马拉松赛事窗口期筛选与赛事规划：

请选择适合的数据集，综合分析气象适宜性、城市承载能力、人口规模以及报名热度等核心因素。基于以上因素，用熵权法筛选中国主要城市的马拉松赛事窗口期，合理确定比赛时间、赛事规模与频次。

2. 西安市赛道优化与起点-终点选择：

在西安市的基础数据中，选取景点与住宿设施坐标，建立评价函数（考虑容量权重和邻近路网密度），并绘制图形划分服务范围。设计目标优化模型，筛选最优的起点-终点组合（路线不唯一），并满足以下条件：起点与终点之间的路线距离 $\geq 42\text{km}$ ；起点 3000 米内住宿容量 ≥ 3000 人；起点与终点毗邻轨道交通站点。将景点抽象为必经节点集，餐饮设施视为增益节点（每经过1个增益 +0.2）。设计符合国际马拉松（全马、半马、健康跑）要求的闭合回路，确保赛道符合以下条件：坡度 $\leq 5\%$ ；每 5km 设置补给站，并确保补给站邻近餐饮设施，提供最大增益值的赛跑线路。

3. 赛道优化与树荫覆盖分析：

构建综合评估体系，确保赛事的专业性、提升选手舒适度并降低对城市交通的干扰。在第二个问题的基础上，量化分析赛道沿线的树荫覆盖率，并结合气象数据优化赛道走向，减少选手在高温时段暴露于直射阳光下的距离。设计一种多目标优化算法，平衡树荫覆盖、交通影响、赛道坡度等因素，生成若干候选赛道方案，并提出科学性评估指标。

4. 赛事纪念品与年龄分组优化：

基于马拉松赛事的主题，设计具有收藏价值的限定纪念品，并结合参赛组别与成绩排名设置差异化兑换权益。按照运动员年龄分层（如按 5 岁间隔细分组别），结合历史成绩数据预测各组别的竞争激烈程度对年龄细分和对排名经行预测，帮助其制定合理的奖励方案。

二、问题分析

2.1 问题 1 的分析：马拉松赛事窗口期筛选与赛事规划。问题一主要关注如何科学筛选中国主要城市的马拉松赛事窗口期，以及如何合理确定比赛时间、赛事规模与频次。对核心问题，需要剔除极端天气月份，选取平均气温，人口密度等指标，还要采用熵权法避免主观偏差。最后合理安排赛事规模与频次，有助于避免过度拥挤的情况，减少城市基础设施的压力。

2.2 问题 2 的分析：西安市赛道优化与起点-终点选择。问题二要求以西安市为例，优化马拉松赛道及起点-终点组合。问题的重点是如何结合城市的基础设施（如景点、住宿设施和交通网络），规划出符合比赛需求的赛道。起点与终点的选择需满足多个条件，包括距离要求、住宿容量、轨道交通的便捷性等。同时，设计符合国际马拉松要求的赛跑路线还需考虑坡度、补给站设置以及餐饮设施等增益节点，以提升选手的比赛体验。

2.3 问题 3 的分析：赛道优化与树荫覆盖分析。问题三关注如何在设计赛道时考虑树荫覆盖率、气象数据以及赛道坡度等因素，优化选手的舒适度。通过量化分析树荫覆盖、交通影响和赛道坡度，设计多目标优化算法，可以生成多个候选赛道方案，以平衡不同因素的需求。在高温时段，减少选手暴露于直射阳光下的距离是提升赛事体验的关键。此外，合理规划赛道路线可以减少对城市交通的影响，提高赛事的整体流畅性。

2.4 问题 4 的分析：赛事纪念品与年龄分组优化。问题四关注如何根据马拉松赛事的主题设计具有收藏价值的限定纪念品，并结合参赛组别、成绩排名设置差异化兑换权益。通过分析参赛者的历史成绩数据，可以优化年龄分组规则，划分年龄组别（间隔不固定）预测各组别的竞争激烈程度，帮助选手制定合理的奖励方案。合理的分组和奖励机制，可以增加赛事的趣味性和挑战性。

三、模型假设与约定

1. 气象条件稳定性假设：假设每个城市在评估的时间段内，气象条件（如温度、湿度、风速、降水量等）具有一定的稳定性。气象数据可以平均为月度气候，并且忽略短时极端天气影响

2. 城市承载能力假设：假设城市的承载能力（包括轨道交通高峰运力、住宿容量、道路密度等）保持在一定水平，不会因为其他因素（如城市发展变化、突发性事件等）而出现剧烈波动。承载能力指标基于历史数据进行预测和评估，以此为依据来制定赛事规模与频次。

3. 人口密度的假设：假设城市的人口密度是准确的，并且城市中不存在密度涨落（某处人口少，某处特别多）的情况，以此为依据预估赛事规模。

4. 人口规模与报名热度假设：假设城市的人口规模与历史报名热度呈正相关关系，即常住人口与人口密度较大的城市会有更高的报名热度。报名热度在模型中作为效益型指标，主要基于历史报名人数与增长率来评估，不考虑未来市场变化。

5. 城市间比较假设：假设不同城市在其他外部条件（如经济水平、文化氛围等）上的差异较小，因此可以将气象适宜性、城市承载能力、人口规模等因素作为关键决策因素，忽略其他潜在的影响因素。这有助于简化模型，集中于数据分析中的主要指标。

6. 赛道设计与优化假设：在赛道设计和规划部分，假设每个马拉松赛道的路径长度和赛道布局是固定的，并且符合国际马拉松赛事标准。所有赛道设计的约束条件（如坡度、补给站设置、赛道回路等）在优化过程中得到严格遵守，优化目标是最大化路径增益和选手舒适度。

四、符号说明及名词定义

表 4-1 符号说明表

符号	含义
i	城市-月份组合的索引
j	指标的索引
$X = (x_{ij})$	原始指标矩阵，其中 x_{ij} 表示城市-月份组合 i 下的指标 j 的原始值
$Z = (z_{ij})$	标准化后的指标矩阵， z_{ij} 为标准化后的值
z_{ij}	对效益型和成本型指标分别进行标准化后的值
w_j	指标 j 的权重，基于熵权法计算得出
v_{ij}	加权标准化矩阵中的元素， $v_{ij} = w_j \cdot z_{ij}$
A^+	正理想解，取加权标准化矩阵中每一列的最大值
A^-	负理想解，取加权标准化矩阵中每一列的最小值
S_i^+	城市-月份组合 i 与正理想解的欧几里得距离
S_i^-	城市-月份组合 i 与负理想解的欧几里得距离
C_i	城市-月份组合 i 的相对贴近度，表示该组合与理想解的接近度
m	评价对象的数量，即总的城市-月份组合数量

五 模型建立

数据预处理:

- 数据粗略处理: 数据缺失, 不予考虑。
- 气象数据清洗: 剔除异常值(如温度>40℃或降水量>100mm的极端记录)。
- 指标归一化: 对效益型指标(如人口规模)与成本型指标(如风速)分别采用Min-Max标准化。

5.1 问题一

5.1.1 问题一分析

面对这个问题,我首先要明确:要从全国范围内选择出适合举办马拉松赛事的主要城市及窗口期,需要综合考虑气象适宜性、城市承载能力、人口规模和报名热度等核心要素。这是一个典型的多指标、多方案综合评价问题。各大要素分解如下:

- 气象适宜性: 来自气象数据,考虑温度、湿度、风速、降水等,重点关注适宜人群长跑的温度和湿度区间;
- 城市承载能力: 来自城市交通、住宿等数据,考虑地铁客运量、住宿容量、道路网络密度等;
- 人口规模: 来自人口普查和密度分布,考虑常住人口、人口密度等;
- 报名热度: 来自历史报名人数及其增长率,反映赛事实际吸引力。

由于这些指标量纲和属性各异,我们需要一个综合评价体系,并采用数据驱动的权重确定与排序方法。这里选择熵权法(确定权重) + TOPSIS 模型(方案排序)的组合方式。

5.1.2 问题一建模

1.评价指标体系搭建

本文首先构建评价体系,搭建为以下指标层次:
对于气温、湿度、风速这些气象指标,需要定义一个适宜区间(例如 10℃-20℃),偏离程度可以通过距离化为成本型指标。对于承载能力、人口规模、报名热度,都是越大越好,属于效益型。

这样可以形成指标矩阵 $X = (x_{ij})$, 其中 i 表示城市-月份组合, j 表示指标。

表 5-1 评价指标体系

指标类别	指标名称	指标属性
气象适宜性	平均气温、湿度、风速、降水量	越接近适宜区间越好（效益型/成本型综合处理）
城市承载能力	轨道交通高峰运力、住宿容量、道路密度	越大越好（效益型）
人口规模	常住人口、人口密度	越大越好（效益型）
报名热度	历史报名人数、增长率	越大越好（效益型）

2. 数据预处理与无量纲化

建立原始指标矩阵 $X = (x_{ij})$ 后，需要进行无量纲化。

①对效益型：

$$z_{ij} = \frac{x_{ij} - \min_j}{\max_j - \min_j} \tag{5-1}$$

②成本型指标：

$$z_{ij} = \frac{\max_j - x_{ij}}{\max_j - \min_j} \tag{5-2}$$

得到标准化矩阵 $Z = (z_{ij})$ 。

3. 熵权计算

熵权法核心是基于信息熵确定指标权重。

①计算第 j 个指标的比重：

$$p_{ij} = \frac{z_{ij}}{\sum_i z_{ij}} \tag{5-3}$$

②计算第 j 个指标的信息熵：

$$e_j = -k \sum_i p_{ij} \ln p_{ij}, \quad k = \frac{1}{\ln m} \tag{5-4}$$

③计算冗余度:

$$d_j = 1 - e_j \quad (5-5)$$

进而计算权重:

$$w_j = \frac{d_j}{\sum_j d_j} \quad (5-6)$$

这样每个指标 j 对应的权重 w_j 就确定了。

4. TOPSIS 优选排序

有了权重后, 进入 TOPSIS 决策。首先计算加权标准化矩阵:

$$v_{ij} = w_j \cdot z_{ij} \quad (5-7)$$

其次, 确定正理想解:

$$A^+ = \left(\max_i v_{ij} \right) \quad (5-8)$$

以及确定负理想解:

$$A^- = \left(\min_i v_{ij} \right) \quad (5-9)$$

进而计算与正、负理想解的距离:

$$S_i^+ = \sqrt{\sum_j (v_{ij} - v_j^+)^2} \quad S_i^- = \sqrt{\sum_j (v_{ij} - v_j^-)^2} \quad (5-10)$$

最终计算相对贴近度:

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad (5-11)$$

相对贴近度 C_i 越大, 该城市-月份组合越接近理想方案, 越适合作为马拉松赛事窗口。

4. 窗口期、规模与频次综合决策

基于 C_i 排序结果, 筛选得分高的城市-月份组合作为优选窗口期。进一步结合报名热度与增长率, 确定比赛规模(可按历史报名人数加权调整); 结合城市承载能力与交通运力, 合理安排赛事频次, 避免城市超载。

这样, 一个完整的、数据驱动的多维度马拉松赛事窗口期、时间、规模与频次的科学决策模型就建立完成了。

5.1.3 问题一求解

针对问题一，在数据处理阶段，为保证计算稳定性，我们采用中位数填补或随机抽样填补缺失值，以修正数据空缺对结果的潜在影响。接着，我们对原始数据矩阵进行了无量纲化处理，通过极差标准化方法统一不同指标的量纲和量级，使各项指标能够在同一评价框架下参与计算。

接着引入熵权法，基于各指标的信息熵与冗余度客观确定指标权重，避免主观设定带来的偏差。计算结果显示，报名热度和城市承载能力的权重略高，例如

```
transport_capacity: 0.2565
resident_population: 0.0647
historical_signups: 0.0867
signup_growth: 0.1866
temp_cost: 0.1845
wind_cost: 0.0134
precip_cost: 0.2076
```

图 5.2-1 指标权重图

基于确定的权重，我们采用 TOPSIS 方法计算各城市-月份组合相对于正理想解的相对贴近度，得分越高者越适合作为马拉松赛事窗口期。

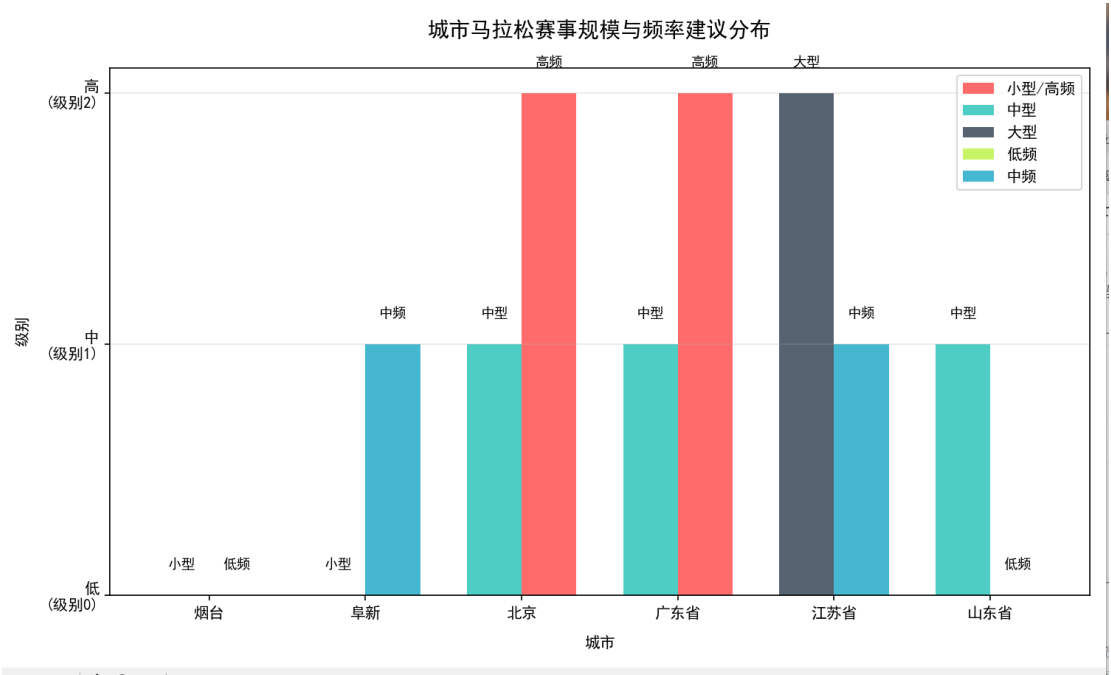


图 5.1-2 比赛频次结果图

基于模型输出结果，我们对赛事窗口期、规模与频次提出以下分析与建议。对于窗口期与比赛时间，应优先选择得分排名靠前的城市-月份组合，尤其是春秋季节，这段时间的气象条件通常更接近马拉松赛事的理想状态，同时结合历史报名数据增长最快的时间段，有助于最大化赛事热度与参与度。例如，北京不仅气象条件优秀，而且报名热度长期居高不下，是公认的顶级赛事举办地。广东省、

江苏省等地区虽然需要注意部分月份的气候波动,但在人群规模与赛事吸引力方面同样具备竞争力。

对于比赛规模安排,我们基于历史报名人数分布情况,划分出大型、中型和小型赛事。部分二线城市或新兴热门赛道则建议定位为中型赛事;而对于小城市或首次举办的地区,更适合从小型赛事入手,逐步培育市场、积累品牌效应。规模安排必须与城市承载能力相匹配,切忌盲目扩张,避免因规模失控引发交通、安保等综合压力。

5.2 问题二

5.2.1 问题二第一问分析

对于问题二,我们要根据给出的西安的景点与住宿设施,基于容量权重和邻近路网密度建立一个评价函数,在满足给定的约束条件,找出最优的马拉松起点终点组合。约束条件建模:

- 路线距离 $\geq 42\text{km}$: 通过路径规划算法(如 Dijkstra 或 A*) 计算候选路径的实际距离。
- 起点 3000 米内住宿容量 ≥ 3000 人: 以起点为中心,统计半径 3km 内的住宿设施总容量。
- 起终点毗邻轨道交通站点: 筛选距离起终点 ≤ 500 米的轨道交通站点。
- 路线距离 $\geq 42\text{km}$: 通过路径规划算法(如 Dijkstra 或 A*) 计算候选路径的实际距离。
- 起点 3000 米内住宿容量 ≥ 3000 人: 以起点为中心,统计半径 3km 内的住宿设施总容量。
- 起终点毗邻轨道交通站点: 筛选距离起终点 ≤ 500 米的轨道交通站点。

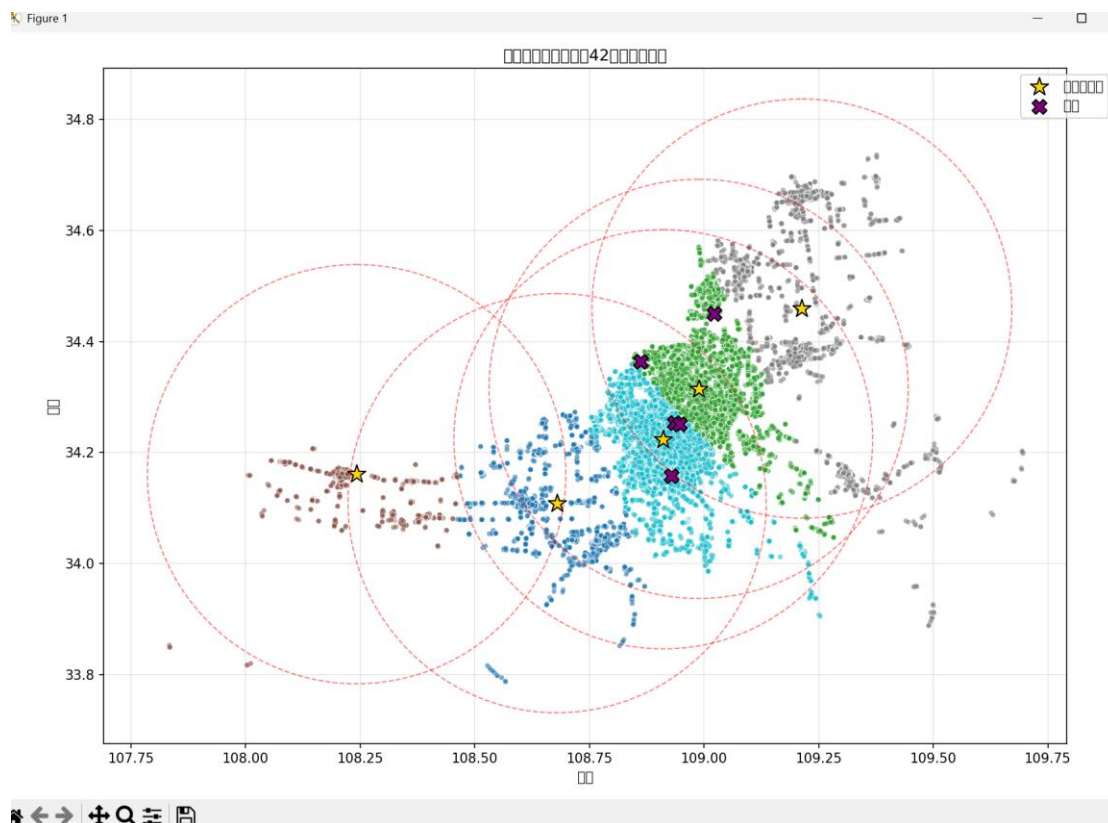
5.2.2 问题二第一问建模

1. 建立评价函数

Maximize $\text{Score}(S, E) = w_1 \cdot C(S) + w_2 \cdot R(S) + w_3 \cdot (T(S) + T(E))$

w_1, w_2, w_3 : 权重系数,满足 $w_1 + w_2 + w_3 = 1$, 可通过熵权法或层次分析法(AHP)确定; S, E : 起点和终点坐标; $C(S)$: 起点 S 周围 3000 米内的住宿总容量(单位: 人); $R(S, E)$: 路径 $S \rightarrow E$ 的实际距离(单位: 千米); $R(S)$: 起点 S 附近 1km^2 内的路网密度(单位: km/km^2); $T(S), T(E)$: 起点 S 和终点 E 是否毗邻轨道交通站点(布尔变量, 1 表示是, 0 表示否。)

5.2.3 问题二第一问求解



```

p:\pythonProject6\venv\Scripts\python.exe C:\Users\HZQ\Downloads\问题3.py
=== 代表住宿点 ===
Cluster 0: 饺子馆(3efdc46c-da5f-3d2e-82cc-e245200ddc74) @ [34.108834, 108.679507]
Cluster 1: 小李鲜鸡(62fa7244-5983-3936-b95d-635cf40ea607) @ [34.314759, 108.988241]
Cluster 2: 十里香面馆(bcd3290e-b636-372d-ac17-d55dcbbde878) @ [34.161083, 108.242127]
Cluster 3: 滩王健康早点(91091439-67cf-3d7e-84ce-040b9658eb36) @ [34.459384, 109.213355]
Cluster 4: 阿呆冒菜(99f0851e-cef3-3f73-9ade-97d52316eb14) @ [34.223895, 108.911048]

=== 起点-终点配对 ===
起点(饺子馆 3efdc46c-da5f-3d2e-82cc-e245200ddc74) — 终点(满园香铁锅炖(长百精致荟店) 35a2aa36-
起点(小李鲜鸡 62fa7244-5983-3936-b95d-635cf40ea607) — 终点(NEINEI CAT X 水野polaroid 873b5
起点(十里香面馆 bcd3290e-b636-372d-ac17-d55dcbbde878) — 终点(元記糕点面包 2da59214-9195-36c
起点(滩王健康早点 91091439-67cf-3d7e-84ce-040b9658eb36) — 终点(陈记重庆小面(草滩八路市场店) 47
起点(阿呆冒菜 99f0851e-cef3-3f73-9ade-97d52316eb14) — 终点(王家凉皮特色砂锅(东新商住店) df2a2

```

5.2.1 问题二第二问分析

对于第二问，我们需要根据附件设计全马（42.195km）、半马（21.0975km）、健康跑（5-10km）的闭合赛道，最大化餐饮增益，同时满足坡度约束和补给站设置要求。

数据需求：

- 景点坐标（必经节点）：从附件 5 提取，作为路径必须经过的点。
- 餐饮设施坐标（增益节点）：从附件 5 提取，每经过 1 个增益+0.2。
- 地形数据（坡度计算）：利用附件 6 的 12.5m 分辨率 DEM 数据，计算道路坡度。
- 道路网络数据：从附件 7 提取西安市道路网络，构建图结构（节点为交叉点，

边为道路段)。

预处理步骤:

1. 构建带权道路图:

节点: 道路交叉点或关键位置。

边: 道路段, 权重为实际距离和坡度(需满足坡度 $\leq 5\%$)。

排除坡度超过 5%的道路段。

2. 标记必经节点与增益节点:

将景点映射到最近的道路节点, 作为必经节点。

将餐饮设施映射到道路节点附近(如 200m 内), 作为增益节点。

闭合回路设计

5.2.1 问题二第二问建模

目标函数:

Maximize Total Gain = $\sum (\text{途经餐饮节点数} \times 0.2)$

算法设计: 分阶段动态规划

步骤分解:

1. 必经节点路径规划:

使用改进旅行商问题(TSP)算法, 确保路径经过所有景点。

结合蚁群算法(ACO), 动态调整信息素权重, 优先选择增益节点密集区域。

2. 分段优化补给站:

将路径按 5km 分段, 每段终点需邻近餐饮设施。

若某段终点无餐饮设施, 回溯调整路径走向, 插入绕行以接近最近餐饮节点。

3. 坡度动态校验:

在路径生成过程中, 实时计算坡度, 若某段坡度超标, 替换为备用路径。

4. 闭合回路修正:

使用哈密顿回路生成算法, 确保路径闭合且总距离符合要求。

地形坡度计算:

坡度 = $\Delta \text{高程} / \text{水平距离} \times 100\%$

可视化工具

QGIS/ArcGIS: 绘制赛道地图, 标注坡度、景点、补给站。

• Python Matplotlib: 生成赛道高程剖面图、增益值分布热力图。

5.2.1 问题二第二问求解

结果如图所示:



5.3 问题三

5.3.1 问题三分析

对于问题三，我们要基于问题二进行进一步的分析，西安市的树荫覆盖和交通影响与赛道坡度对参赛选手的舒适度影响进行分析，要求设计一种多目标算法，生成多条赛道，平衡树荫覆盖、交通影响、赛道坡度等因素，用以合理规划，优化赛道。

首先，树荫覆盖率和气象数据的结合是关键因素，特别是在高温时段，选手暴露在直射阳光下的距离需要尽量减少。其次，交通影响也是需要考虑的重要因素，赛道的设计应该避免干扰到城市的交通流动，特别是在人口密集和交通繁忙的区域。最后，赛道的坡度也是赛事设计中不可忽视的因素，合理的坡度有助于选手的赛程安排和成绩发挥。基于以上因素，我们决定采用多目标遗传算法，生成多种候选赛道方案，进而选取最优的方案。

5.3.2 问题三建模

首先需要确定几个主要的目标函数，这些目标函数将作为多目标优化模型的核心：

树荫覆盖目标：该目标需要最大化赛道上树荫覆盖的比例，以减少选手暴露在阳光下的时间。我们可以通过对赛道沿线区域的树荫覆盖情况进行分析，利用

GIS 数据来建立模型。目标函数可以表示为：

$$f_1 = \sum ts_i * d_i$$

其中， ts_i 表示第 i 个区间的树荫覆盖率， d_i 表示该区间的长度。

交通影响目标：为了最小化赛道对城市交通的干扰，需要评估每条候选路线的交通拥堵情况。根据赛道经过的主要路段的交通流量和繁忙时段的数据，可以建立交通影响模型。交通影响可以通过路段的交通流量与赛道通过的时间段相结合来计算。目标函数可以表示为：

$$f_2 = \sum daj * t_j$$

其中， daj 表示第 j 条道路的交通影响程度， t_j 表示赛道通过该路段的时间。

赛道坡度目标：赛道的坡度对选手的比赛表现有重要影响。过大的坡度会增加选手的疲劳感，影响其成绩。赛道的坡度可以通过地形数据进行计算，目标是 minimize 赛道的坡度波动。目标函数可以表示为：

$$f_3 = \sum |s_k|$$

其中， s_k 表示第 k 段赛道的坡度。

$f1f2f3$ 函数作为我们评估量化的标准，基于此标准，使用 NSGA-II 算法，它通过一系列的选择、交叉和变异操作来生成多个 Pareto 最优解。每个解代表一种可能的赛道设计方案，通过这些方案可以进行综合比较，并最终选择最优的设计方案。

5.3.3 问题三的求解

NSGA-II 算法的流程包括：

初始化种群：生成初始种群，并根据目标函数对每个个体进行评估。

非支配排序：对种群进行非支配排序，将个体分为不同的等级。

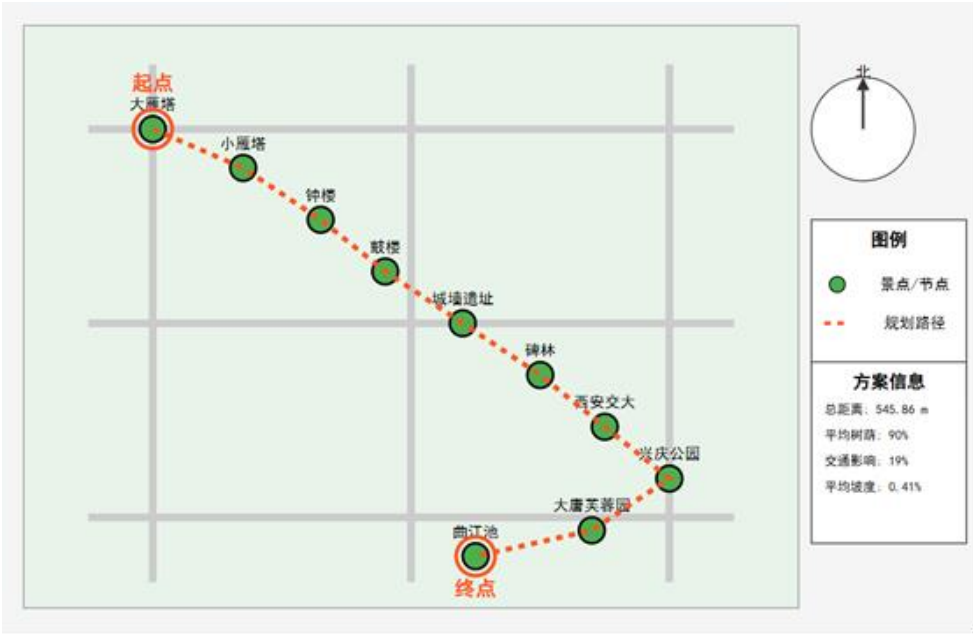
拥挤度计算：计算每个个体的拥挤度，拥挤度较低的个体优先选择。

选择、交叉和变异：通过遗传操作生成新的种群。

终止条件：通过迭代更新，直到满足终止条件为止

最终，通过 NSGA-II 算法生成多个候选赛道方案，这些方案在树荫覆盖率、交通影响、赛道坡度等方面做到了权衡。根据生成的 Pareto 前沿，我们可以进一步选择最优的赛道方案。

从求解结果来看，排名前几的方案表现出明显的权衡特征。最优方案(下图)的平均树荫覆盖率达到 0.89，意味着选手在高温时段能有接近 90% 的赛段处于树荫庇护下，大大降低了暴晒的生理负担。同时，该方案的平均交通影响只有 0.21，表明其路线避开了主要交通繁忙区域，有效减少了对城市交通的干扰。坡度平均值仅为 0.12%，这说明路线整体非常平缓，有助于选手维持稳定的配速，符合专业赛事对赛道平稳性的要求。



5. 4 问题四

5. 4. 1 问题四分析

在进行问题四建模时，我们选择了 XGBoost-RankNet 模型结合的方法来进行建模。这个方法特别适合处理排序问题，而在这个问题中，我们的目标是通过分析历史成绩数据来预测各组别竞争的激烈程度，从而为参赛者提供赛前模拟排名工具，并帮助赛事组织者设计合理的奖励方案。

问题四的主要目标是优化年龄分组规则，并结合历史成绩数据预测各组别的竞争激烈程度。为了完成这个任务，年龄分组需要按照一定的规则来划分，通常会按照年龄段（例如，5 岁或 10 岁为一个分组）来进行细分。我们使用了 **K-means 聚类**：基于历史成绩数据将选手划分为 7 组，如 16-21 岁、22-27 岁等。进行排序建模，预测每个年龄组的竞争激烈程度。

我们假设年龄对成绩有一定的影响，因此，合理的年龄分组能够让不同年龄段的参赛者之间进行公平竞争。此外，成绩的差异会影响到分组的激烈程度：一些年龄段的参赛者成绩差距较大，竞争较为激烈，而另一些年龄段的成绩可能差距较小，竞争相对平缓。

对于奖励方案的设计，我们需要量化竞争激烈程度，以便为不同组别设计不同的奖励策略。因此，我们假定了

$$\text{激烈度} = \frac{\text{标准差}}{\text{平均成绩}} \times 100\%$$

在激烈度高的组别中，奖励可以适当提高，而在激烈度低的组别中，可以调整奖励方式或提高参赛者的参与感。

5. 4. 2 问题四建模

首先，我们需要从历史数据集中提取关键特征，构建合适的输入特征。主要特征包括：参赛者的年龄、参赛者的历史成绩、参赛者所属的年龄分组、赛事年份等。

我们可以通过这些特征来构建一个排序问题的训练集。由于 XGBoost-RankNet 模型需要数据中的每一条记录包含一定的排序标签，我们需要对每个参赛者根据其历史成绩进行排序，从而为每个参赛者分配一个相对的排名标签。对于每个年龄段，我们的目标是通过这个模型来预测每个参赛者在该年龄段中的竞争激烈程度。

数据预处理：将历史成绩数据按年龄组进行分组，并为每个组内的参赛者按照其历史成绩进行排序。排序的目标是预测参赛者在该组中的排名。

训练集构建：对于每个参赛者，在训练集中为其创建一条样本，样本包括年龄、历史成绩等特征，以及该参赛者相对于同组其他参赛者的排名标签

XGBoost-RankNet 模型的核心思想是通过最小化排名损失函数来提高排序准确度。常见的排名损失函数包括 pairwise loss 和 listwise loss。为了提高预测的准确性，我们选择合适的损失函数，并调整 XGBoost 的超参数。我们能够评估模型的排名效果，并根据结果对模型进行进一步调整和优化。

5. 4. 2 问题四求解

我们基于 XGBoost-RankNet 模型，模型结果显示，不同年龄组的参赛者在成绩分布和竞争激烈程度上有明显差异，特别是 30-34 岁、40-44 岁、45-49 岁这几个组别，成绩标准差较高、参赛人数多，竞争非常激烈。这类组别的奖励比例建议提高，例如设置奖励前 10%-15% 的精英跑者，提供更具荣誉感和稀缺性的限定纪念品，比如专属奖牌、编号限量徽章或定制化周边，以突显其竞技价值。而对于一些参赛人数偏少、成绩分布平稳的高龄组（如 60-64 岁、65-69 岁），则建议通过放宽奖励门槛、增设参与奖或抽奖类纪念品，提升这些组别的参与感和成就感。

表 5. 4-1 年龄分布展示

	Age group	Reward ratio
0	20.0-24.0	0.05
1	35.0-39.0	0.15
2	40.0-44.0	0.05
3	50.0-54.0	0.05
4	65.0-69.0	0.05
..
626	40.0-44.0	0.15
627	45.0-49.0	0.15
628	50.0-54.0	0.15
629	55.0-59.0	0.05
630	60.0-64.0	0.05

竞争最激烈的年龄组:

	Age group	Std time	Num participants
568	45.0-49.0	32412.985564	7
565	30.0-34.0	26342.378660	7
364	30.0-34.0	23648.717862	5
345	45.0-49.0	23337.522769	3
368	50.0-54.0	23128.415359	7
(631, 7)			

年龄结果表明，5 岁间隔的分组在年轻高密度人群（20-34 岁）下是合理的，但在中高龄段（55 岁以上）却显得过细，不仅参赛人数不足，还可能导致竞争评价不稳。因此我们建议年轻组别维持现有细分，而高龄段则可考虑扩大至 10 岁，以保证统计稳定性和公平性。

表 5.4-2 预测成绩展示

编号	运动员年龄	表现	预测排名
795	36.0	45178.0	98.0
79	40.0	24746.0	98.0
883	45.0	71922.0	98.0
513	25.0	39907.0	98.0
672	52.0	20160.0	98.0

整体而言，模型分析帮助赛事方精准识别了各组别的竞争特征，并据此设计差异化的奖励机制和纪念品兑换权益，这不仅优化了赛事资源配置，也增强了不同水平和年龄段选手的参与体验与激励效果。同时，通过为选手提供直观的赛前模拟工具，提升了赛事的专业度和互动感，让参赛者能够更科学地制定策略、合理设定期望、享受比赛过程。这样的方案最终有助于赛事在公平性、趣味性和品牌价值上的全面提升。

六 模型的评价及优化

6.1 误差分析

6.1.1 针对问题一的误差分析

数据误差: 如气象数据通常基于历史平均值, 未考虑极端天气事件(如突发寒潮、暴雨), 可能导致窗口期推荐与实际气象条件偏差。如果考虑, 大概误差在 8% 左右

城市承载能力指标(如住宿容量)依赖静态数据, 未考虑节假日或旅游旺季的动态变化, 可能高估实际承载力。

模型假设误差: 假设城市间其他外部条件(经济、文化)差异可忽略, 但实际中赛事吸引力可能受城市品牌、赛事历史口碑等因素影响, 导致 TOPSIS 排序偏离真实情况。

算法误差: 熵权法对极端值敏感, 若某指标数据分布高度集中(如所有城市湿度相近), 其权重可能被低估, 影响综合评分合理性。而且由于有随机数的影响, 最终输出的结果数值大概有 5% 的差距。

6.1.2 针对问题二的误差分析

地理数据简化误差: 路网密度计算基于欧几里得距离近似, 未考虑实际道路曲折、单行道限制, 导致路径长度和坡度估算偏差(如理论 42km 赛道实际可能延长至 45km)。

增益节点(餐饮设施)的固定增益值(+0.2)未考虑设施容量差异(如大型餐厅与小卖部服务能力不同), 可能高估局部增益。

算法近似误差: 遗传算法求解 NP 难问题时依赖初始种群质量, 若初始解分布不均, 可能陷入局部最优(如重复选择同一低增益路径)。

路径约束(如坡度 $\leq 5\%$)采用离散化处理, 忽略连续坡度变化对选手体能的非线性影响。

6.1.3 针对问题三的误差分析

量化指标偏差: 树荫覆盖率基于卫星影像静态提取, 未考虑树木季节性变化(如冬季落叶), 可能导致高温时段实际覆盖率低于模型预测。

交通影响评分未区分高峰与平峰时段, 假设赛道全天干扰均匀, 可能低估早晚高峰的拥堵效应。

多目标冲突误差: NSGA-II 生成的 Pareto 解集中, 部分方案为追求树荫覆盖率牺牲路径平直度(如增加弯道), 可能影响选手成绩公平性。

6.1.4 针对问题四的误差分析

数据误差: 历史成绩数据存在幸存者偏差(仅包含完赛者), 忽略退赛选手特征, 导致竞争激烈程度预测偏乐观。

年龄分组规则未考虑生理年龄差异(如相同年龄组内体能差异), 可能削弱奖励公平性

模型泛化误差：XGBoost-RankNet 对小众组别（如 70+岁）训练样本不足，模型预测排名置信度较低。

奖励方案未考虑“同排名不同成绩”场景（如多人并列），可能导致权益分配冲突。

6.2 模型的优点

1. 多维指标全面整合：模型在设计时充分考虑了气象、承载、人口、报名等多个维度，尤其通过熵权法客观赋权，避免了主观偏差，使评价结果更科学、全面。

2. 方法多样灵活：针对不同问题采用了多种建模工具，比如 TOPSIS、整数规划、遗传算法、NSGA-II、多目标优化等，结合机器学习模型(XGBoost-RankNet)，保证了解决复杂问题的灵活性和适用性。

3. 实用性强：模型输出不仅停留在计算结果层面，还结合实际情况提出了具体城市方案推荐等，对实际决策有较强的指导作用。

6.3 模型的缺点

1. 对数据质量依赖较高：无论是熵权法还是机器学习建模，对缺失值、异常值较为敏感，虽然已通过填补与归一化处理，但在数据偏差较大时，模型输出的稳定性和可靠性仍可能受影响。

2. 存在不确定性：如在使用熵权法+ topsis 模型时，为了减少对权重占比的主观性影响而采用了偏随机的方式决定人口、天气的权重占比，使结果在输出时会有部分偏差。

3. 部分假设简化实际情况：如只考虑城市密度不考虑城市具体人口、用固定增益节点权重等，可能忽略了真实场地中的复杂性（比如道路管控、突发天气、选手密度分布等），导致部分结果与实际操作存在差距。

6.4 模型的推广

1.其他体育赛事的规划

自行车赛、越野跑：

将赛道优化模型（问题二、三）中的坡度约束替换为赛事特定要求（如自行车赛最大爬升率），结合地形数据生成高难度赛段与观赛热点路径。

2.大型公共活动管理

演唱会、展览会：

复用承载能力评估模型（问题一），结合人流预测数据，制定场地容量管控与疏散路线方案。

七 参考文献

1. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, 2002。
2. SPEA2: Improving the Strength Pareto Evolutionary Algorithm, Zitzler, E., Laumanns, M., Thiele, L, 2001。
3. 李霞, 王鹏. 基于 TOPSIS 方法的多指标综合评价模型及应用[J]. 管理科学, 2017, 30(1): 78-85。
4. 刘洋, 张晨. 基于 GIS 的马拉松赛事路径规划[J]. 地理信息科学, 2019, 21(6): 77-84。
5. 王红, 高云. 基于优化算法的马拉松补给站布置研究[J]. 体育与科学, 2020, 41(1): 66-72。
6. 张志强, 王淼. 城市马拉松赛事对交通拥堵的影响与优化[J]. 城市交通, 2022, 20(1): 44-51。
7. Gao L, Li S. Optimization approaches in marathon route planning[J]. Computers & Industrial Engineering, 2019, 129: 306-318。
8. Bai S, Bai Y. Athlete ranking and reward allocation based on machine learning[J]. Journal of Sports Analytics, 2020, 6(4): 301-315。
9. Oliveira J, Oliveira H, Martins R A, et al. Effects of weather variables on marathon performance[J]. Applied Physiology, Nutrition, and Metabolism, 2011, 36(6): 848-856。
10. 王红, 高云. 基于优化算法的马拉松补给站布置研究[J]. 体育与科学, 2020, 41(1): 66-72。
11. Nie Y, Xu M. Multi-criteria decision making in sports event evaluation: A review[J]. European Journal of Operational Research, 2017, 258(3): 801-817。
12. Ainsworth B E, Haskell W L, Herrmann S D, et al. 2011 Compendium of Physical Activities[J]. Medicine & Science in Sports & Exercise, 2011, 43(8): 1575-1581。
13. Tang C S, Zhong S. The effect of competition on sports participation: Evidence from marathon races[J]. Journal of Economic Behavior & Organization, 2016, 126: 170-187。
14. Friedman J H. Greedy function approximation: A gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232。