

# 2019 年第六届中国可视化与可视分析大会

## 数据可视分析挑战赛-挑战 1

(ChinaVis Data Challenge 2019 - mini challenge 1)

### 答 卷

参赛队名称：东北师范大学—李文涛—挑战 1

团队成员： 李文涛，东北师范大学，liwt665@nenu.edu.cn，队长

付 佳，东北师范大学，fuj905@nenu.edu.cn

焦旭峰，东北师范大学，jiaoxf163@nenu.edu.cn

徐劭斌，东北师范大学，2233935216@qq.com

张慧杰，东北师范大学，zhanghj167@nenu.edu.cn，指导老师

团队成员是否与报名表一致（是或否）： 是

是否学生队（是或否）： 是

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）： D3, Echarts, Excel,

Python, Jupyter Notebook

共计耗费时间（人天）： 30 人天

本次比赛结束后，我们是否可以在网络上公布该答卷与视频（是或否）： 是

(灰色字为参赛信息填写模板，请参赛者在提交时参照模板填写)

**挑战 1.1：请您通过分析数据，推测会议期间主会场和各分会场的日程安排。（建议参赛者回答此题文字不多于 800 字，图片不多于 5 张）**

会议期间，各会场内布置的无线传感器实时检测并记录着附近的参会人员，因此我们可以分析各会场内人员密度的时序变化，推断会议期间的日程安排。

### （一）人员密度统计

- 基于传感器布置表和传感器地图，筛选出主会场与各分会场的传感器。
- 根据传感器数据，以分钟为粒度，统计三天中不同时刻主会场和各个分会场的人员密度。

### （二）探索会议期间日程安排



图 1-1 第一天人员密度时序分布堆叠图

根据得到的人员密度统计数据，我们绘制不同时刻各会场的人员密度时序分布堆叠图（如图 1-1）和人员空间分布热力图（如图 1-2），从时间和空间两个角度来进行会场日程安排的探索。图 1-1 展示了第一天的人员密度情况，堆叠图中横轴表示时刻，时间粒度为一分钟，纵轴表示该时刻对应场馆的人员密度。不同颜色的河流线标识不同的会场，面积越大则表示该段时间人员密度越大。热力图对应会场地图，每个格子对应一个传感器。图 1-2-a 表示人员密度热力图，图 1-2-b 表示传感器在 100 秒之内人员流出量热力图。我们使用由黄到红的颜色映射密度或流量的大小。用户可以经由时间轴定位具体的时间。



图 1-2 第一天 12 时 0 分 34 秒各区域人员空间分布热力图

以第一天为例，我们对人员密度时序分布堆叠图和人员空间分布热力图进行交互式探索，确定各会场时间安排。

**主会场：**从图 1-1 可以看出，第一天上午仅主会场有活动。主会场在上午 8:30 左右人数开始增加，当我们使用鼠标悬停在 9:00 时，发现人员达到 1500 人左右，到 9:10 人员开始稳定在 2000 人左右。故推测主会场在第一天九点开始举行开幕式，九点十分正式进入会议报告阶段。9:10-9:40 期间，会场人数变化平稳，推测此时正在举行会议，9:40-9:50 期间人员有所减少，推测有一个短暂的休息或进行下一个报告。9:50-10:30 主会场人员再次保持平稳，推测该阶段继续举行会议。10:30-11:00 期间人数有一个骤减为零的大幅度下降，同时 room3 人数开始增加，我们推测该阶段有一个长时间的茶歇休息。11:00 时主会场人数又开始增加，持续至约 12:10 时人数开始减少直至为零。我们继续移动时间轴，通过人员流出量热力图(图 1-2-b)可以清晰确定十二点时人员开始移动，故我们推测 12:00 主会场会议结束。主会场上午活动结束，下午没有会议安排。

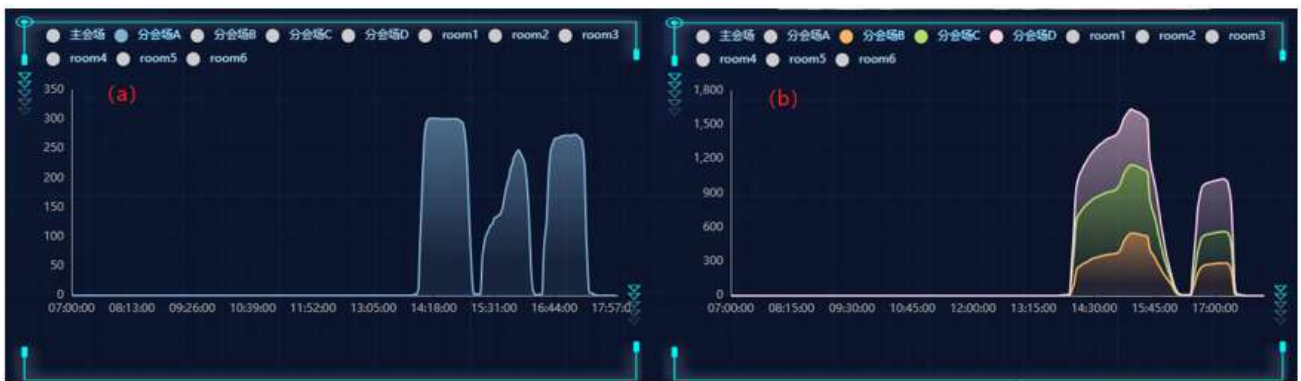


图 1-3 分会场 A 与分会场 B、C、D 人员密度时序分布堆叠图

**分会场：**当我们选择查看各分会场的人员密度变化时，发现分会场 A（如图 1-3-a）的变化模式不同于分会场 B、C、D（如图 1-3-b）。第一天下午阶段分会场 A 监控到人员密度主要持续在 14:00-17:20，其中 14:00-15:00、15:20-16:10、16:30-17:20 该三段时间人数变化基本平稳，推测为正在举办报告或者比赛。其中间隔阶段均有小段时间的人数骤降为零的休息安排。而分会场 B、C、D 随时间变化的人员密度模式基本一致。14:00-15:30 期间人数变化平稳，推测为报告或者比赛，15:30-16:30 期间人数减为零，推测为休息时间。16:30-17:15 期间人数重新增加并保持平稳，推测该时间段内继续安排报告或者比赛。至此整个第一天日程结束。

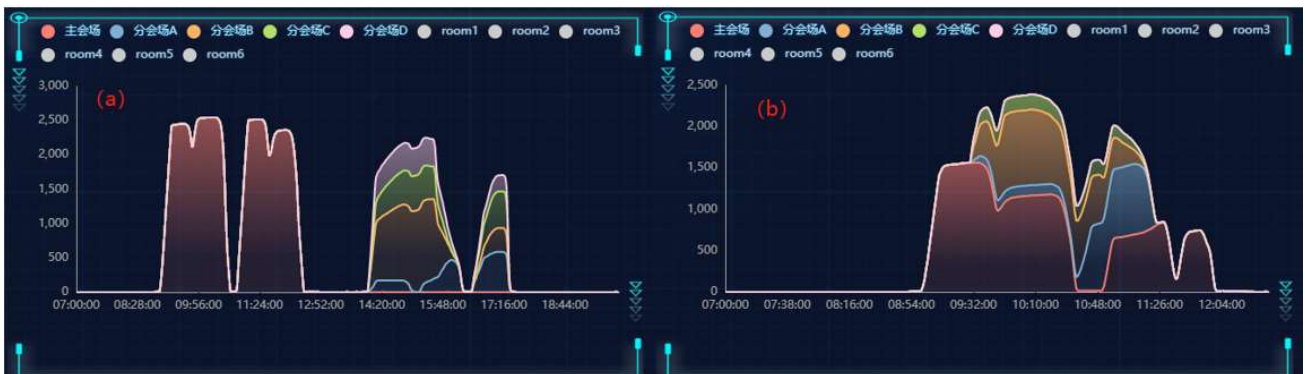


图 1-4 第二天和第三天人员密度时序分布堆叠图

按照上述分析，我们分别计算并绘制了第二天（如图 1-4-a）和第三天（如图 1-4-b）的人员密度时序分布堆叠图。总结日程安排如图 1-5 所示：

| 会议期间主会场和各分会场的日程安排 |     |             |             |                              |
|-------------------|-----|-------------|-------------|------------------------------|
| 工作日               | 时间段 | 时间          | 场馆          | 备注                           |
| day1              | 上午  | 9:00-9:10   | 主会场开幕式      | 主会场第一段主题<br>推测为两位专家作报告，中间含换场 |
|                   |     | 9:10-9:40   | 主会场1.1      |                              |
|                   |     | 9:50-10:30  | 主会场1.2      |                              |
|                   |     | 10:30-11:00 | 茶歇休息        | 主会场第二段主题                     |
|                   |     | 11:00-12:00 | 主会场2        |                              |
|                   | 下午  | 14:00-15:00 | 分会场A1       | 分会场A的三段报告或比赛<br>中间均夹杂小段休息    |
|                   |     | 15:20-16:10 | 分会场A2       |                              |
|                   |     | 16:30-17:20 | 分会场A3       |                              |
|                   |     | 14:00-15:30 | 分会场B1、C1、D1 | 分会场BCD的第一段报告或比赛              |
|                   |     | 15:30-16:30 | 茶歇休息        |                              |
|                   |     | 16:30-17:15 | 分会场B2、C2、D2 |                              |
| day2              | 上午  | 9:00-9:40   | 主会场3.1      | 主会场第三段主题<br>推测为两位专家作报告，中间含换场 |
|                   |     | 9:50-10:30  | 主会场3.2      |                              |
|                   |     | 10:30-11:00 | 茶歇休息        | 主会场第四段主题<br>推测为两位专家作报告，中间含换场 |
|                   |     | 11:00-11:30 | 主会场4.1      |                              |
|                   |     | 11:40-12:00 | 主会场4.2      | 分会场A的三段报告或比赛<br>中间均夹杂小段休息    |
|                   | 下午  | 14:00-15:00 | 分会场A4       |                              |
|                   |     | 15:20-16:10 | 分会场A5       |                              |
|                   |     | 16:30-17:20 | 分会场A6       |                              |
|                   |     | 14:00-15:30 | 分会场B3、C3、D3 | 分会场BCD的第三段报告或比赛              |
|                   |     | 15:30-16:30 | 茶歇休息        |                              |
|                   |     | 16:30-17:15 | 分会场B4、C4、D4 |                              |
| day3              | 仅上午 | 9:00-9:40   | 主会场5.1      | 主会场第五段主题<br>推测为两位专家作报告，中间含换场 |
|                   |     | 9:50-10:30  | 主会场5.2      |                              |
|                   |     | 10:30-11:00 | 茶歇休息        | 主会场第六段主题<br>推测为两位专家作报告，中间含换场 |
|                   |     | 11:00-11:30 | 主会场6.1      |                              |
|                   |     | 11:40-12:00 | 主会场6.2      | 分会场A的两段报告或比赛<br>中间均夹杂小段休息    |
|                   |     | 9:30-10:20  | 分会场A7.1     |                              |
|                   |     | 10:30-11:00 | 分会场A7.2     |                              |
|                   |     | 9:30-11:00  | 分会场B5、C5    | 分会场BC的第五段报告或比赛，D结束           |

图 1-5 主会场及各分会场日程安排





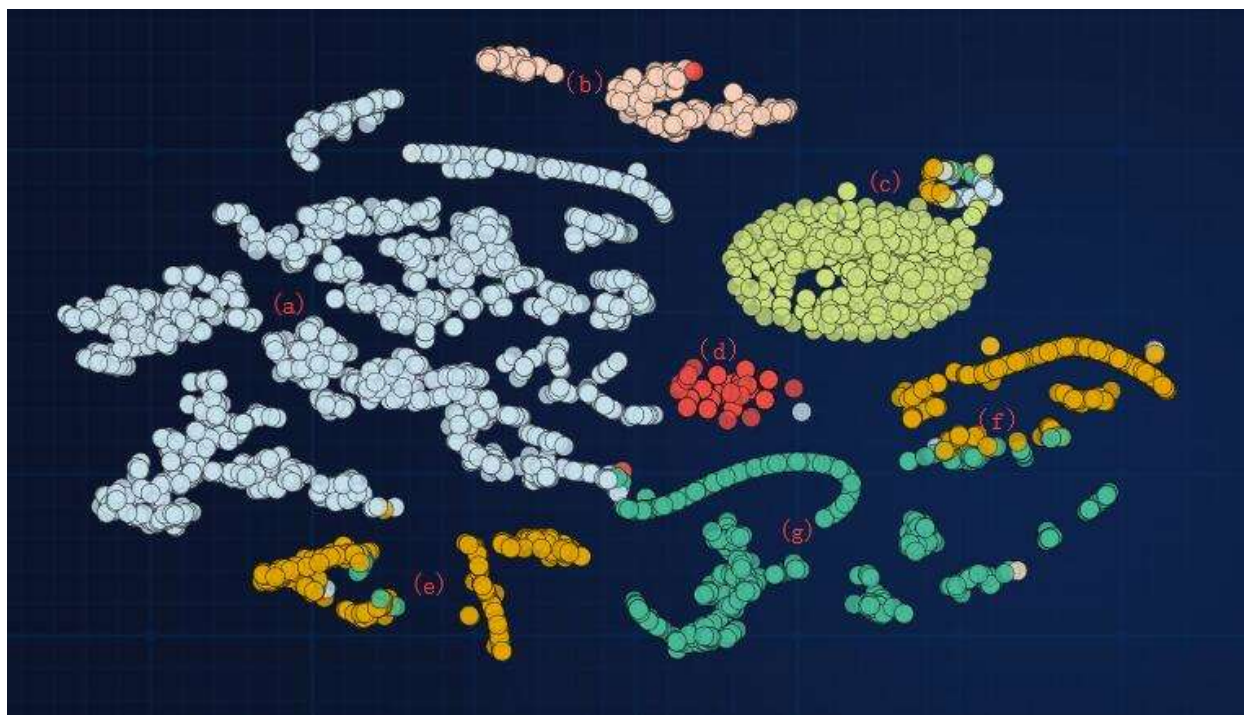


图 2-1 所有参会人员散点图

系统支持对参会人员的交互式圈选，雷达图和平行坐标图、热力图与之联动，进而总结圈选人员的总体特征。雷达图模式 1（如图 2-2-a）和模式 2（如图 2-2-b）分别从空间和时间角度展示圈选人员的参会特征。平行坐标图（如图 2-2-c）进一步细化并展示参会人员移动特征，主要包括参会期间的步数和速度等。

（1）我们使用刷选工具圈选图 2-1-a 区域，分别从空间和时间角度对该类人员的移动规律进行总结。

**空间特征：**计算该类人员在各个场地的平均停留时间，可以发现该部分人员最长停留时间为主会场，同时在分会场 A、B、C 也停留半小时到一小时之间，如图 2-2-a 所示。

**时间特征：**在雷达图模式 2 中我们发现该类人员会议三天参会时间基本一致（如图 2-2-b 所示），均在主会场会议开始（合 32400 秒）之前到场。我们进一步在平行坐标（如图 2-2-c 所示）中发现该类人员离场时间明显分为两类，一类在 45000 秒（12 点半）左右即主会场会议结束，另一类在 65000 秒（18 点）左右，即分会场会议结束。我们分别刷选两类人员，人员名单列表实时展示相关人员 id，点击后在时空立方体中进一步探索该人员的详细移动轨迹，其中 x 轴和 y 轴对应会场布置图坐标，z 轴以秒为单位编码时间长度。发现 40000 秒离场人员仅参加了主会场会议（如图 2-2-d 所示），而 65000 秒离场人员在上午参加完主会场会议之后下午又参加分会场的会议（如图 2-2-e 所示）。

因此我们推测该类人员主要为参会学者或学生，他们更倾向于听主会场的报告，同时有部分人对分会场的报告也感兴趣。

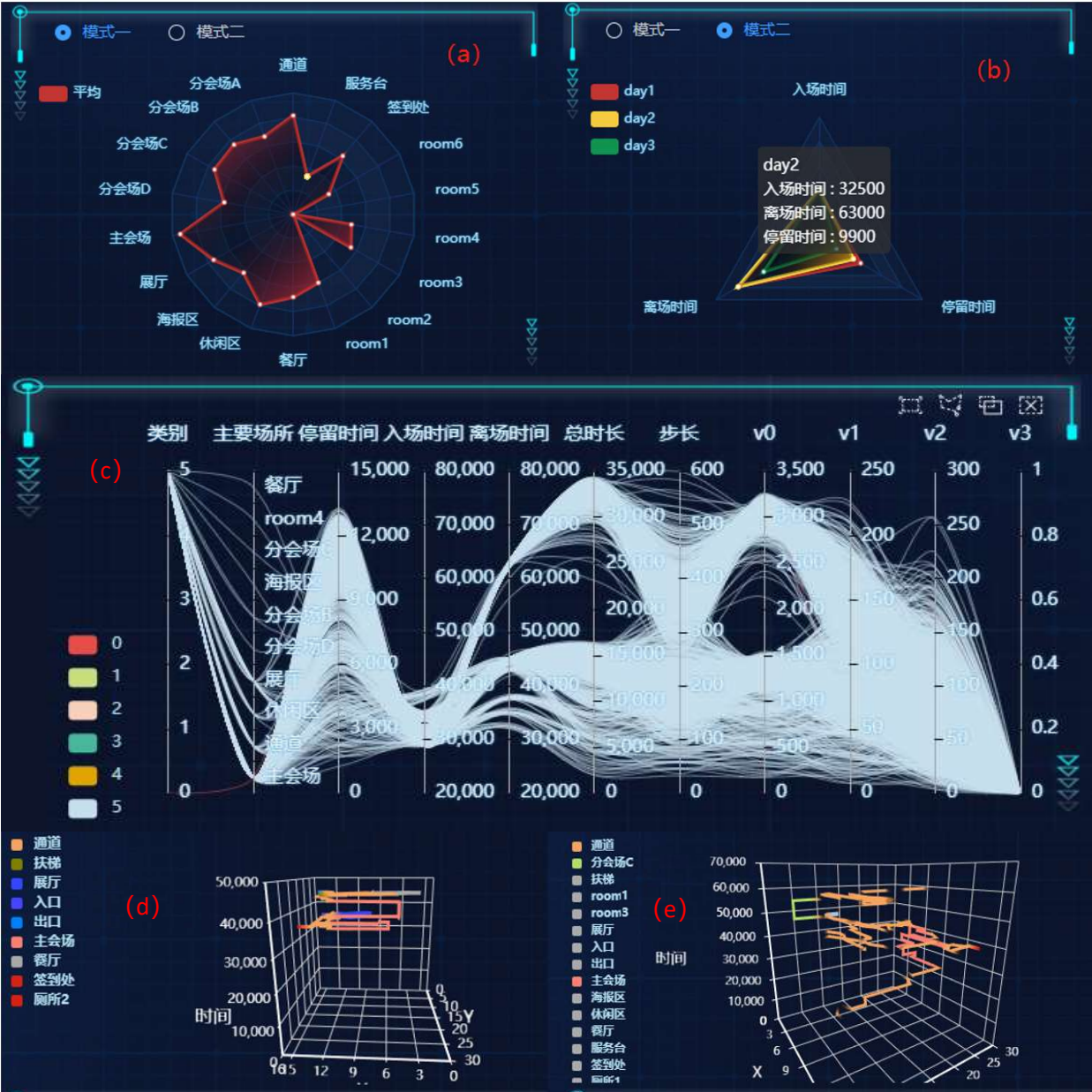


图 2-2 2-1-a 区域参会人员特征

(2) 我们发现图 2-1-b 区域附近有一簇浅粉色标签人员，我们观察其分布特征。

**时间特征：**该类人员主要入场时间为 31100 秒（8 点 40 分），退场时间为 42800 秒（11 点 50 分），恰为会议日程时间（图 2-3-b）。

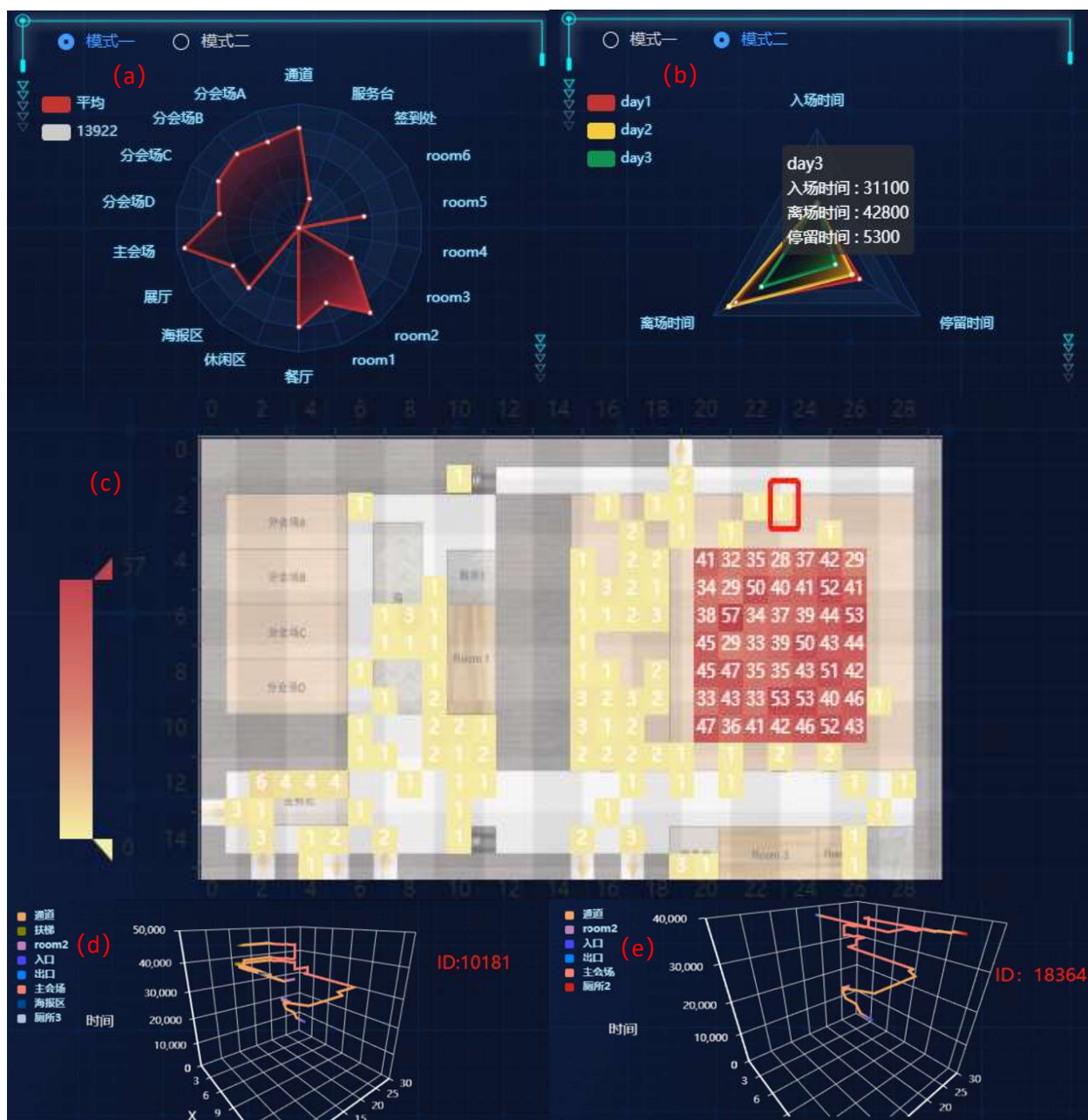


图 2-3 2-1-b 区域参会人员特征

**空间特征:** 主要停留场地为主会场与分会场 ABCD，与图 2-1-a 区域的不同之处在于该类人员均未在签到处停留（如图 2-3-a）。我们在该类人员中随机选取 ID 为 10181 的人员，观察其第一天的轨迹（如图 2-3-d）。发现该人员从入口进入后，直接经走廊进入主会场，在主会场停留三小时左右后离开场地。同时我们选择在主会场会议阶段查看人员密度热力图（如图 2-3-c），查看主会场中心偏里位置（即报告人员）人员 id 及所属类别，发现该人员类别为 2，与 2-1-g 区域人员类别一致，id 为 18364，且轨迹图(如图 2-3-e) 极为相似。

因此我们推断该类人员为相关领域专家或商业大咖，他们的参会时间与一般参会人员大致相似，只是不需要签到。



(3) 当我们圈选图 2-1-c 区域时，观察其分布特征。

**空间特征：**该类人员主要停留场地为通道、服务台、签到处、海报区和展厅,但未在主会场及各分会场停留（如图 2-4-b）。

**时间特征：**入场时间为早上八点（合 30000 秒左右）左右（如图 2-4-c），退场时间为下午五点（合 60000 秒左右），最大停留时间超过八小时（合 28800 秒）。

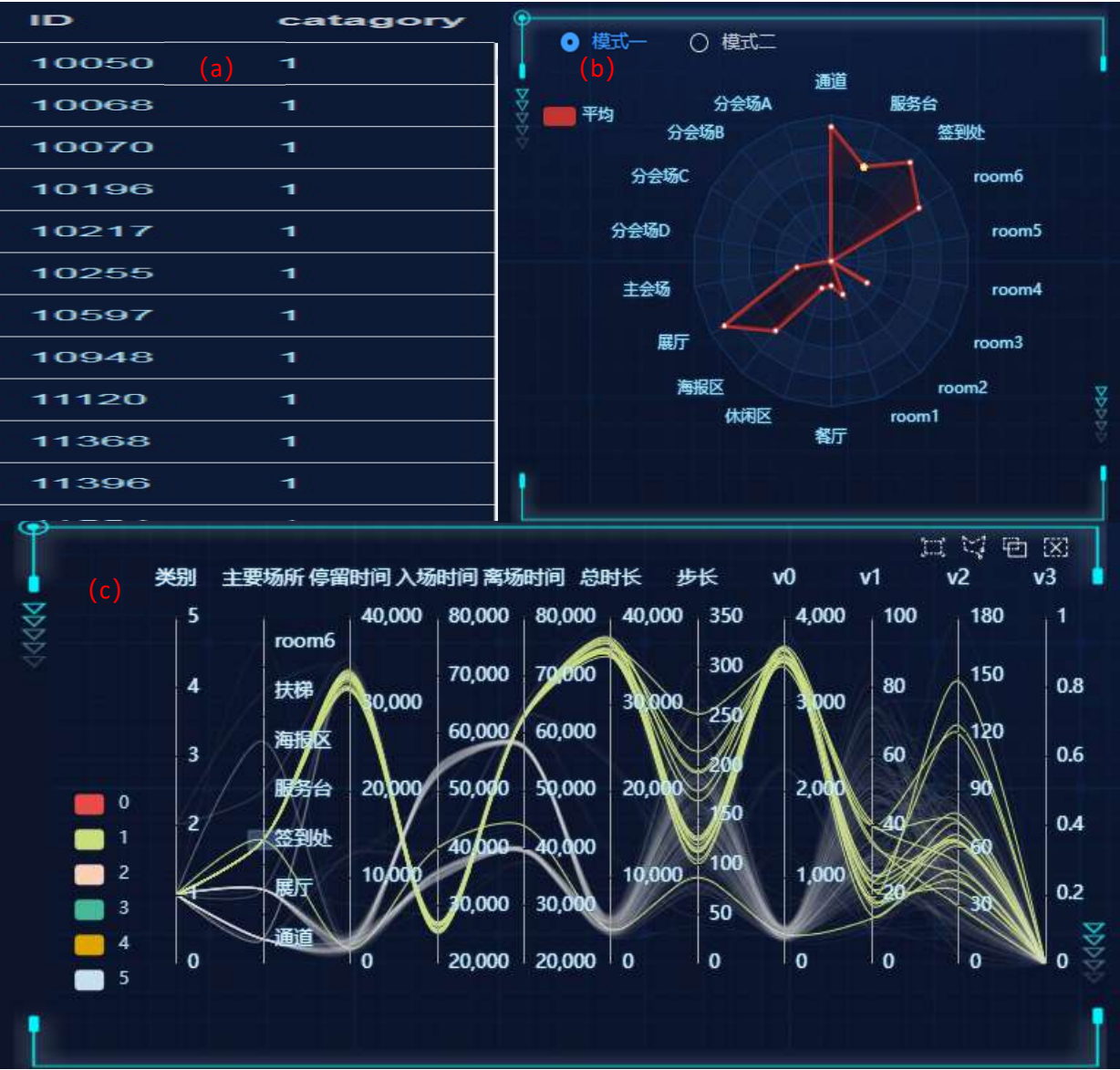


图 2-4 2-1-c 区域参会人员特征

因此我们判断该类人员主要为工作人员，例如负责签到人员、保安、礼仪或展厅单位工作人员等。主要移动规律为在会议开始之前到场，会议结束之后离场，且长时间停留在自己的工作岗位上。

(4) 当我们圈选图 2-1-d 区域人员时，观察其分布特征。

**空间特征：**该类人员主要停留场地为通道和各个房间（如图 2-5-a），其中以 room5 为主，且同样未在主会场和各分会场出现。

**时间特征：**入场时间为早上八点（合 30000 秒左右）左右（如图 2-5-b），退场时间为中午十一点半（合 41400 秒左右）。其中在 room5 停留时间将近十个小时。

因此我们推断该类人员同样为工作人员，但工作区域与图 2-1-c 区域工作人员有所不同，如 room5 的工作人员。主要移动规律为在会议开始之前到场，会议结束之后离场，且长时间停留在自己的工作岗位上。

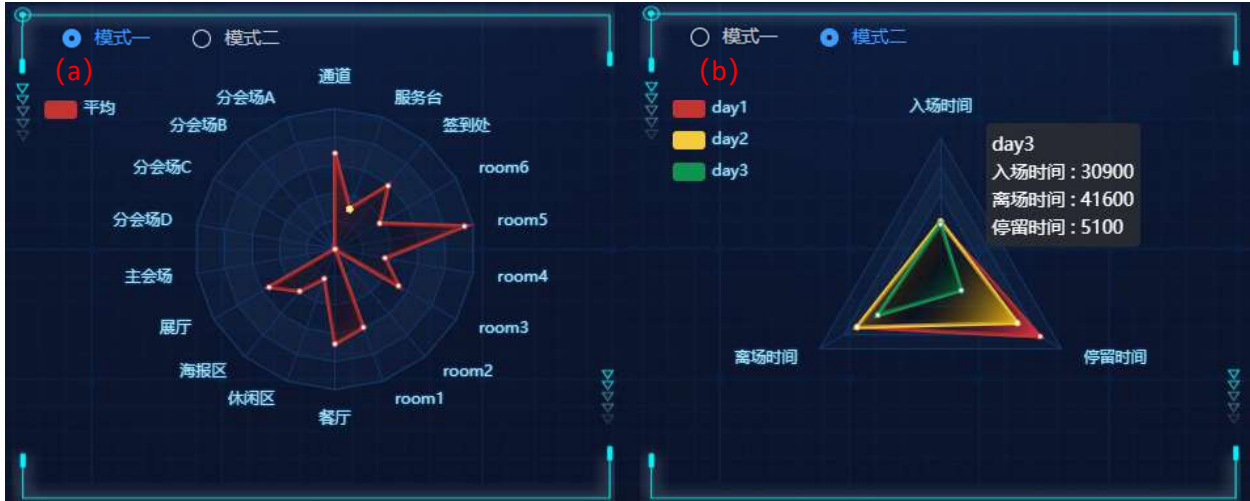


图 2-5 2-1-d 区域参会人员特征

(5) 当我们圈选图 2-1-e 区域人员时，观察其分布特征。

**空间特征：**发现最大停留时间主要在分会场 D（如图 2-6-a）。

**时间特征：**入场时间为下午两点（合 50400 秒左右）左右（如图 2-6-b），退场时间为下午五点半（合 63000 秒左右），恰为分会场会议时间段。

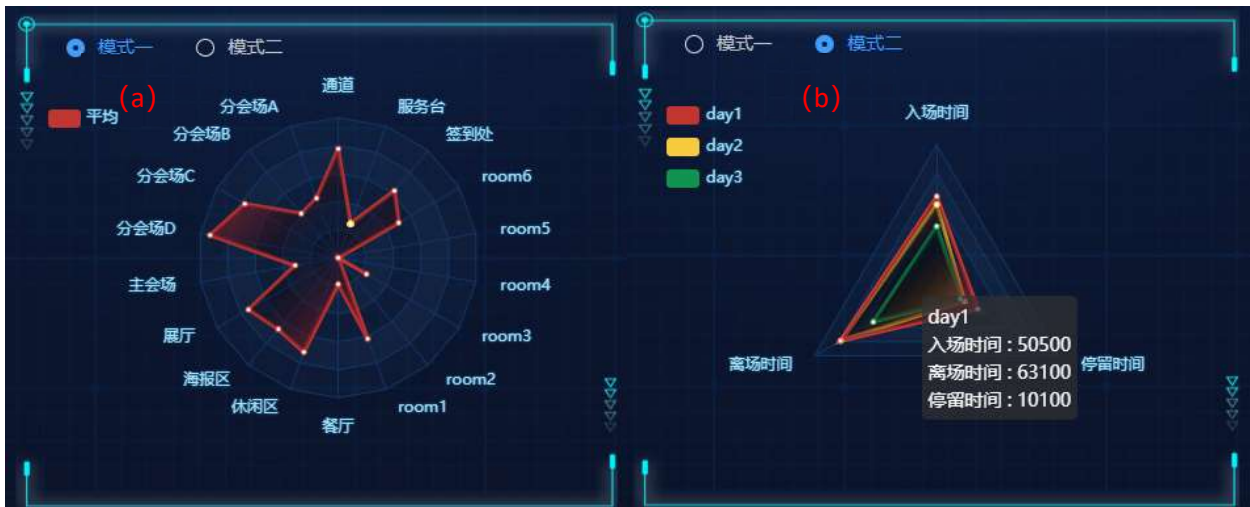


图 2-6 2-1-e 区域参会人员特征

(6) 刷选图 2-1-f 区域人员，观察其分布特征。

**空间特征：**最大停留时间主要为分会场 C（如图 2-7-a）。

**时间特征：**入场时间为下午两点（合 50400 秒左右）左右（如图 2-7-b），退场时间为下午五点半（合 63000 秒左右），恰为分会场会议时间段。

故我们推测该类人员主要参与分会场 C、D 的讲座或比赛。

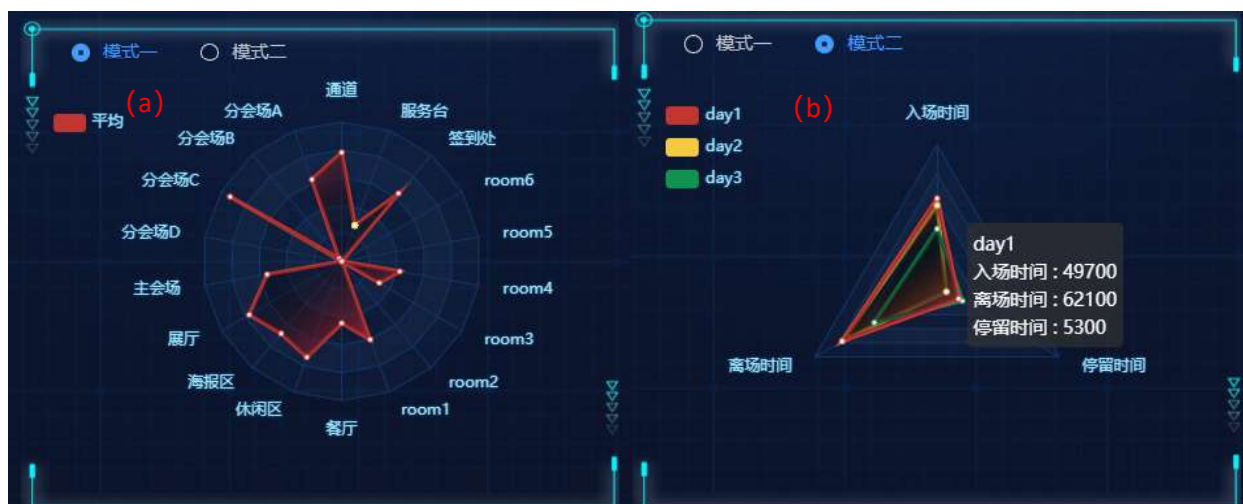


图 2-7 2-1-f 区域参会人员特征

(7) 圈选图 2-1-g 区域人员时，观察其分布特征。

**空间特征：**该部分人员主要停留在分会场 A、B（如图 2-8-a）。

**时间特征：**入场时间为下午两点（合 50400 秒左右）左右（如图 2-8-b），退场时间为下午五点半（合 63000 秒左右），恰为分会场会议时间段。

因此我们推测该部分人员主要参与分会场 A、B 的讲座或比赛。

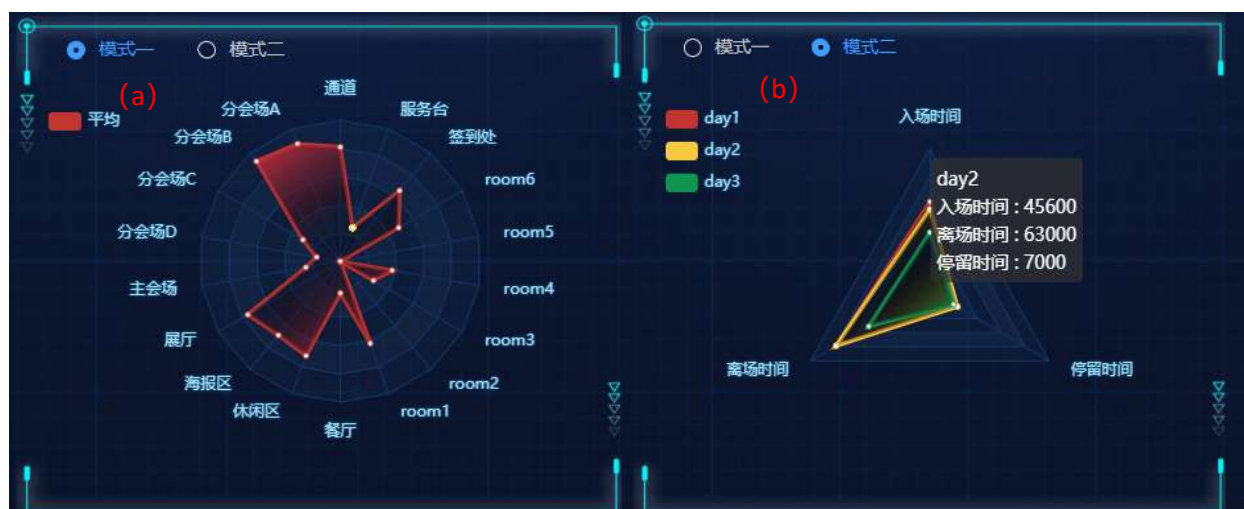


图 2-8 2-1-g 区域参会人员特征

因此经过我们的系统，可以划分出相关领域专家、学者、学生，参展单位，媒体记者，保安、礼仪以及其他普通工作人员等。



**挑战 1.3: 请您找出至少 5 个会议期间值得关注的异常事件。（建议参赛者回答此题文字不多于 1000 字，图片不多于 10 张）**

### （一）参会人员 ID 冒用

我们计算了参会人员会议三天之内经过的传感器数量，发现 ID 为 16632 的人员经过数量大于 800，远远超过其他人员。我们每隔十秒钟计算一次参会人员的速度，分别标记为 0m/s, 0-1m/s, 1-2m/s 以及大于 2m/s，同时我们发现该人员行进速度有四十多次超过两米每秒（如图 3-1）。进而查看 16632 号人员的轨迹数据（如图 3-2）我们发现该人员从入口开始，同时有两条轨迹出现在整个会场内，一条轨迹去往主会场，另一条去往 room2。因此我们怀疑出现人员 ID 冒用现象。



图 3-1 参会人员移动数据平行坐标图

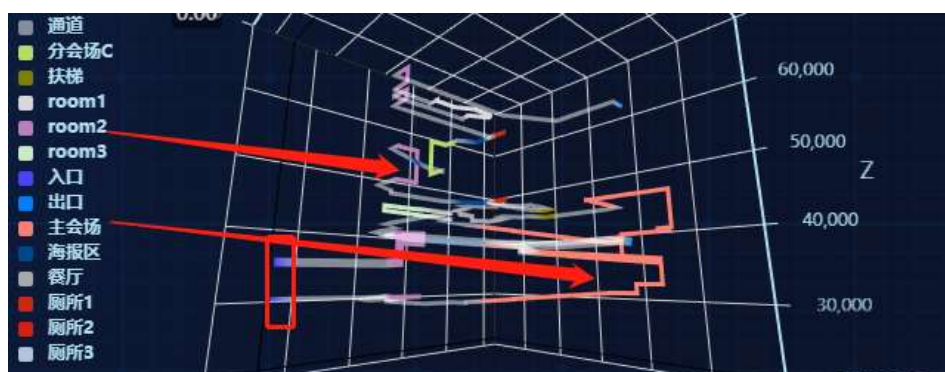


图 3-2 人员轨迹时空立方体

（二） 部分传感器失灵现象

当我们计算会议第二天的人员速度时，发现有部分人员曾出现 1-2 次速度大于两米每秒（如图 3-3），由于正常步速为每秒一米左右，因此我们怀疑此为异常现象。进一步计算每个传感器在 100s 之内的人员流出量时，我们发现在会议第二天 13:35-13: 54 这将近二十分钟之内，展厅部分区域没有任何人员流动现象（如图 3-4），因此我们怀疑该区域传感器此阶段失灵没有记录到人员流动数据，进而导致经过该区域的人群速度过快。

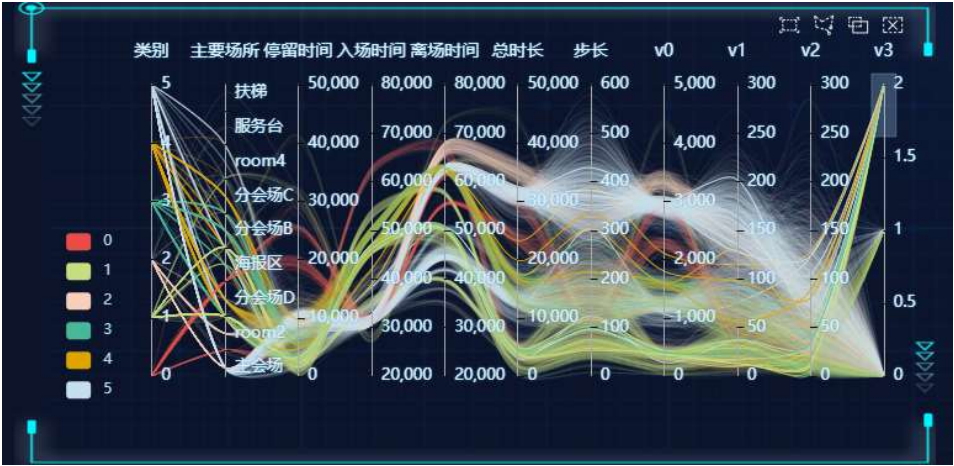


图 3-3 参会人员移动数据平行坐标图



图 3-4 人员流出量热力图

（三） 在某一场地停留时间过久现象

我们发现在会议第二天，有七名人员速度为零次数高达 4000 多次即超过十一个小时（如图 3-5）。我们查看此类人员轨迹（如图 3-6）发现，id 为 19027 的人员进入会场之后，长时间静止在通道处（对应地图为签到处附近），期间没有任何移动现象如去餐厅、厕所等，观察其他六位人员也有类似现象。因此我们怀疑此七名工作人员极有可能存在消极怠工现象，即传感器在而人不在。

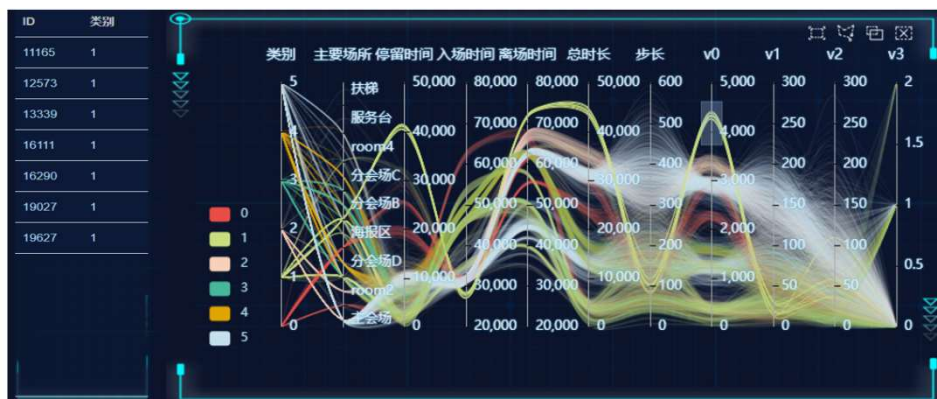


图 3-5 参会人员移动数据平行坐标图

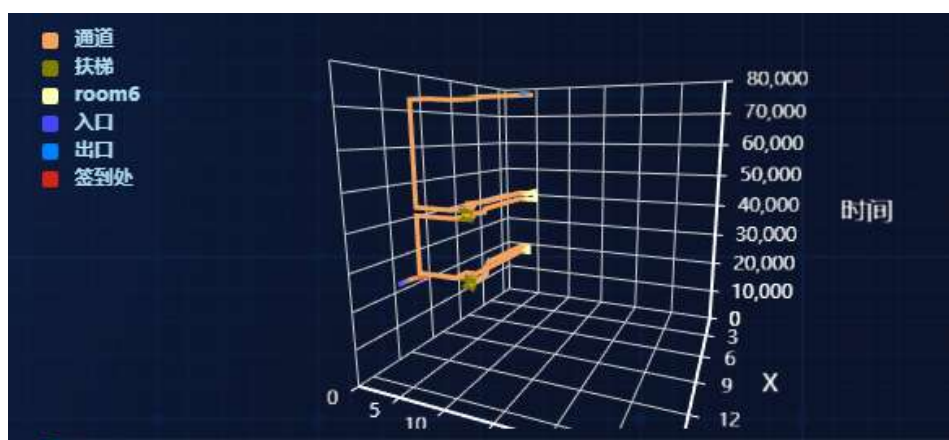


图 3-6 人员轨迹时空立方体

#### (四) 会场出现持续一个小时的无人移动现象

我们发现在会议第二天 18:54-19:55 期间，整个会场传感器没有任何移动记录（如图 3-7-b），但一直有且仅有七人站在出入口附近（如图 3-7-a）。该段时间会议已早早结束，且现场并无其他工作人员，因此该七人此行为无法解释，较为异常。



图 3-7 各区域人员空间分布热力图



（五）签到即走现象

我们注意到有部分参会人员离场时间过早，例如 32000 秒左右即八点五十左右（如图 3-8 所示）。我们随机选择一名 id 为 14180 的该类人员查看其轨迹发现，其在第一天早上 8:30 左右到达入口，在签到处停留几分钟后经过走廊在海报处停留约二十五分钟后在 8:55 即主会场开幕式之前直接离场。因此我们怀疑有部分人签到即走（如图 3-9 所示）。

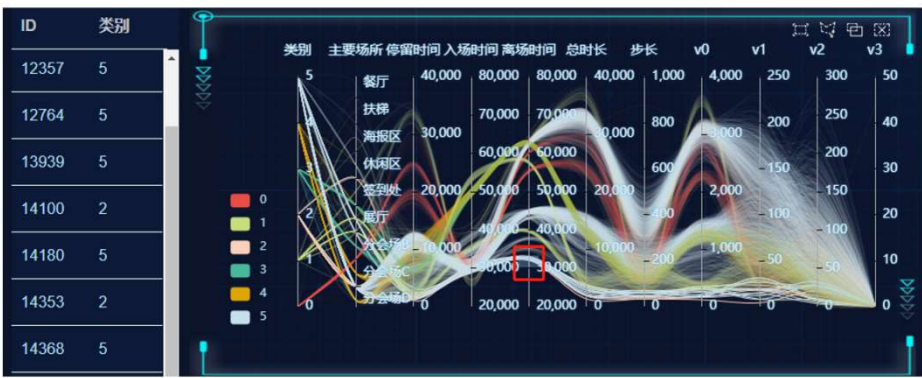


图 3-8 参会人员移动数据平行坐标图

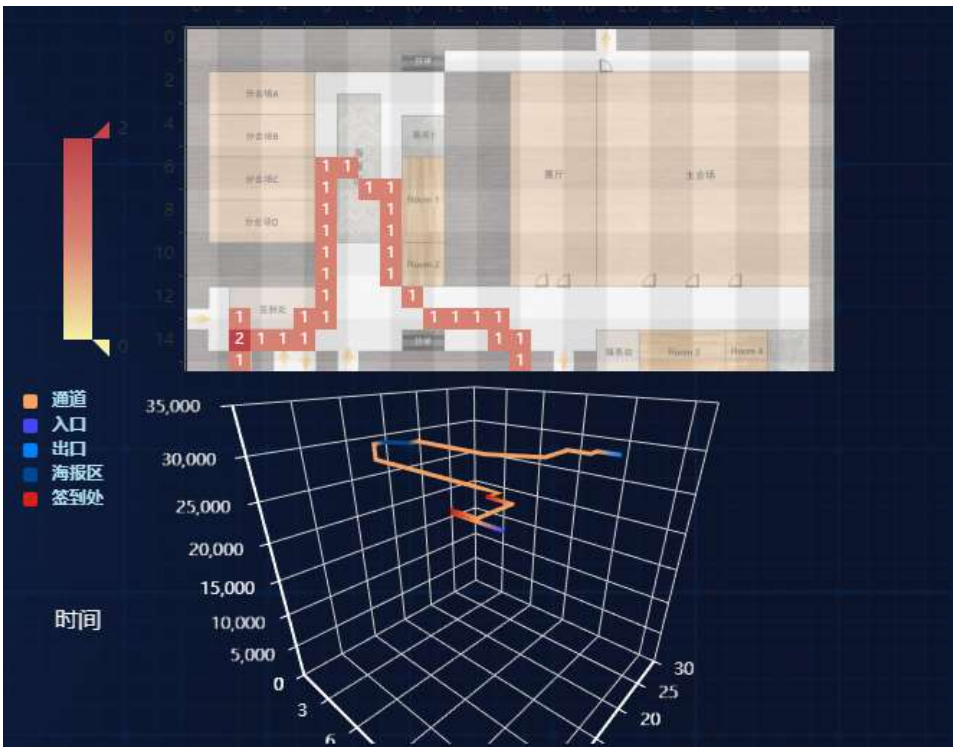


图 3-9 参会人员时空轨迹图



**挑战 1.4：您认为这次会议在组织和管理方面有哪些不足？（建议参赛者回答此题文字不多于 500 字，图片不多于 3 张）**

### （一）加强身份识别管理工作

由异常 1 我们发现出现身份冒用现象，因此建议工作人员应该加强参会人群身份识别管理工作，例如在签到时应核对其身份信息。

### （二）时刻注意设备管理工作

由异常 2 我们注意到部分传感器出现失灵现象，因此建议现场工作技术人员应该时刻监管传感器数据信息，以便能够及时修复。

### （三）加强工作人员管理工作

由异常 3 我们注意到，存在部分工作人员持续工作时间较为异常，疑似存在消极怠工现象，因此我们建议加强工作人员的管理工作，避免出现此类异常行为。

### （四）加强安全管理工作

我们注意到会议结束期间，主会场退场时曾出现拥堵现象。我们从会场人员密度热力图（如图 4-1）发现，主会场有四道门，但是会议结束时，极大部分参会人员均从其中一道门流出，造成 8m\*8m 区域内高达 200 多人，因此我们建议加强工作人员引领疏导管理，组织参会人员有秩序的迅速退场，避免发生安全事故。

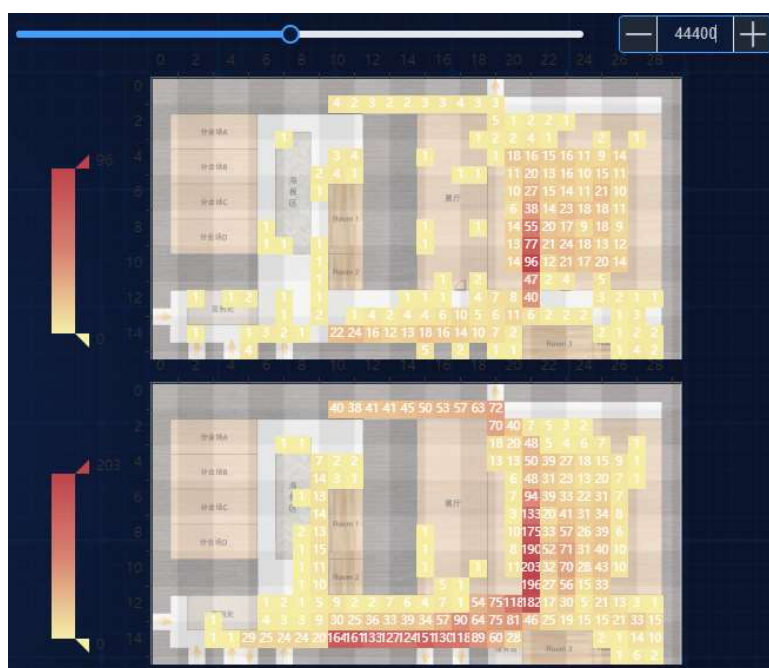


图 4-1 各区域人员空间分布热力图

## （五）提高各会场利用率

我们在第一题中发现，会议期间主会场只在上午开放，而各个分会场只在下午开放。这样安排容易造成空间的浪费，因此我们建议修改各会场日程安排，合理提高会场利用率。

## （六）加强会场公共场所提示功能

我们发现主会场会议茶歇阶段，主会场附近的厕所 2 人员密度较大，而距离分会场较近的厕所 1 人员密度较少（如图 4-2-a）。分会场会议茶歇阶段反之（如图 4-2-b）。同时，距离签到处较近的扶梯始终比位于会场内部的扶梯人员密度更多（如图 4-3）。因此我们建议加强公共场所引导和指示功能，避免参会人员的盲目等待。



图 4-2 人员密度热力图



图 4-3 人员密度热力图