

CPIVis:会场人员信息可视分析系统

李文涛, 付佳, 焦旭峰, 徐劭斌, 张慧杰 (指导老师)



图 1 a:日程安排堆叠图, b:人员密度热力图, c:人员流出量热力图, d:人员类型散点图, e:人员空间特征雷达图, f:人员特征平行坐标图

摘要—本文针对 ChinaVis2019 的 CICSC 会议监控数据, 利用 LDA 文本主题模型和 t-SNE 降维算法对参会人员进行分类, 同时设计并实现了一个分别从空间和人员不同模式探索的交互式可视分析系统 CPIVis。系统可以从时间、空间等多个角度来提取会场日程安排, 同时可以帮助用户提取各类人员特征, 从而探索其移动规律并快速发现异常行为。

关键字—人员分类、异常检测、可视分析、会场管理

1 引言

本文将从四个方面介绍我们的可视分析系统: 第一部分介绍数据的处理; 第二部分结合案例详细介绍可视化设计; 第三部分从多角度来评价我们的可视分析系统; 第四部分将对本文主要内容进行总结。

2 人员类型挖掘

2.1 人员类型划分

首先我们根据会场地图提供的所有场地, 剔除对人员分类影响作用较小的出入口、扶梯与厕所后, 计算出会议三天期间所有参会人员在各个场地的总停留时长, 得到 5256 名参会人员三天内分别在主会场、分会场 ABCD、展厅等 18 个场地总停留时间。

人员划分主要采用两种方法。一是利用 Latent Dirichlet Allocation (LDA) 文档主题生成模型^[1], 将每位参会人员视为一个文档, 每在一个场地停留 60s 记一个该场地的单词, 计算出每个参会人员在各个主题下的概率大小, 取概率值最大的主题作为其类别标签。另外, 我们的系统支持 K-means 聚类来初步划分人员。本文中, 经过多次测试并结合可视化结果, 我们发现当采用 LDA 文档主题生成模型, 且主题设置为 6 时, 可以较好地区分不

同的人员类型。提供多种人员划分方法可以在探索某些大类类别的进一步划分提供辅助建议。我们的系统可以让用户自行选择人员划分个数。

2.2 各类型人员行为模式提取

由于会议包括学术研讨、成果展览、黑客大赛三大部分, 因此参会人员涉及到相关领域专家、学者、学生, 参展单位, 媒体记者以及工作人员等。我们认为相关领域专家、学者和学生主要日程安排包含作报告、听报告以及参加比赛。故我们提取所有参会人员在主会场、分会场 A、B、C、D 所停留的总时间五个特征; 同时工作人员会长时间停留在展厅、走廊、签到处、楼梯等地点, 因此我们计算所有参会人员三天内最大停留时间的地点, 我们将凡是最大停留时间为楼梯、签到处、海报区、展厅和走廊的人员, 设置第六个特征为 1, 其他为 0。此外, 我们发现部分参会人员未进行签到, 我们考虑到签到与否也是区分人员类型的特征之一, 因此我们设置第七个特征, 将未签到的人员设置为 1, 签到的设置为 0。此七个特征归一化后进行 t-SNE^[2]降维, 嵌入维度为二。

为了更好的总结人员移动规律, 我们分三天计算每位参会人员的主要停留场地、停留时间 (起始和持续时间)、跨越的传感器数量以及行走速度等。

- 李文涛 东北师范大学 研究生 E-mail: liwt665@nenu.edu.cn
- 付佳 东北师范大学 研究生 E-mail: fuj905@nenu.edu.cn
- 焦旭峰 东北师范大学 研究生 E-mail: jiaoxf163@nenu.edu.cn
- 徐劭斌 东北师范大学 研究生 E-mail: 2233935216@qq.com
- 张慧杰 东北师范大学 教授 E-mail: zhanghj167@nenu.edu.cn

3 可视化设计

基于挑战赛提出的问题，我们设计的 CPISVis 可视分析系统从会场和参会人员两种角度出发，遵循先整体概览再进行细节探索的设计原则，构建了一个完整的分析流程。

3.1 会场日程安排分析

首先我们从时间角度出发，用堆叠图展示主会场及各分会场的参会人员密度时序分布（如图 1-a）。其中横轴表示时刻，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。从而我们可以推断出各会场的日程安排，例如会议第一天主会场在上午开放，而各分会场在下午开放，且各分会场开放时间极为类似。

同时，我们从空间角度出发，绘制人员密度热力图（如图 1-b）和人员流出量热力图（如图 1-c），展示不同楼层各个时段的人员空间分布规律从而对日程安排进一步探索。我们使用由黄到红的颜色映射密度或流量的大小。用户可以从时间轴定位具体的时间，例如堆叠图可以发现粗略发现主会场十二点左右人员开始减少直至为零，移动热力图时间轴可以清晰确定十二时人员开始流动，主会场会议结束。

3.2 各类型人员移动规律探索

我们首先从全局出发，总体展示所有的人员类型，然后进一步在细节方面展示各类型人员的不同特征，进而探索不同类型人员的移动规律。

(a) 我们利用 LDA 和 t-SNE 算法提取人员特征并将其投影到可视空间（如图 1-d）。图中每个点代表一位参会人员，颜色映射人员类型划分结果，提示工具显示该人员 id 和所属类别。系统支持对参会人员的交互式圈选，雷达图（如图 1-e）和平行坐标图（如图 1-f）与之联动，总结圈选人员的总体特征。

(b) 人员空间特征雷达图（如图 1-e）计算散点图中所有圈选人员在各个场地平均停留时间（红色实线），以及在人员名单中选择的特定 ID 人员在各个场地停留时间（黄色实线）。当鼠标悬停于线段之上时可查看在各个场地停留时长。

(c) 人员时序特征数据雷达图（如图 2）主要分三天展示圈选人员的进场时间、退场时间和在主要停留地的持续时间。数据取自该部分所有人员对应时间的众数。上述两个雷达图可以任意切换，便于用户探索不同人群的移动规律。

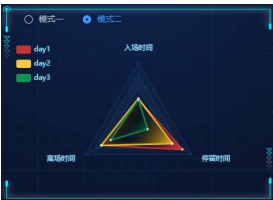


图 2：时序特征数据雷达图

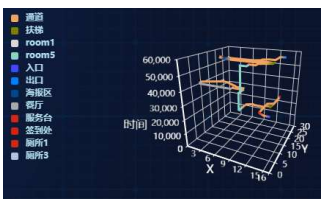


图 3：3D 时空立方体

(d) 平行坐标图（如图 1-f）主要按天展示散点图中圈选部分人员参会过程中的各种特征。分别为所属类别，当天去往主要场地、停留总时长，跨越传感器数量。同时我们每隔十秒钟计算一次所选人员的行走速度，分别标记为 0m/s, 0-1m/s, 1-2m/s 以及大于 2m/s，数值表示出现该速度的次数。其中停留时间以及入场离场时间均以秒为单位。

(e) 为了能更好的探究人员类型的移动规律和异常现象，我们绘制 3D 时空立方体（如图 3），其中 X 轴和 Y 轴分别对应传感器

坐标中的 X、Y 轴，Z 轴以秒为单位编码停留时间，颜色编码不同场地。同时我们可以在轨迹热力图（如图 4）中查看圈选人员在不同传感器经过的次数，进一步探索该人员的详细移动规律。



图 4：轨迹热力图

4 讨论

4.1 可交互性

CPISVis 可视分析系统提供丰富的交互操作。在散点图中，缩放功能可以查看遮挡的节点，圈选功能可以进一步联动探索某类人员的特征；同时平行坐标、雷达图和热力图均提供鼠标悬停提示工具功能，在可视化图表展示的基础上加上数据的显示，使用户更加清楚人员特征。系统提供多个视图可按照用户需求任意切换。而在人员轨迹时空立方体图中可以 360° 旋转，方便用户从多个角度查看人员轨迹。

4.2 实用性

我们的可视分析系统由探索日程安排视图和人员类型移动规律视图两个层级构成，可以让用户对系统的架构有清晰的了解。同时以人员类型散点图作为主视图，用户可以任意圈选人员进行多级联动，简洁有效的对人员类型进行移动规律探索。

4.3 新颖性

我们在人员聚类时一方面使用了传统的 K-means 聚类，另一方面使用了 LDA 文本主题模型。在特征提取使不同类型人员成簇时，我们选择了能最大化展示不同类别的人员特征进行 t-SNE 降维。同时，我们绘制了 3D 参会人员轨迹图，在人员轨迹基础之上增加时间轴，可以更清楚地展示人员移动规律。

4.4 可扩展性

我们的可视化方案不仅能解决挑战一的所有问题，同时还具有良好的可扩展性，可用于其他无监督高维数据进行聚类与特征提取，同时也可针对其他时序数据进行可视化探索。

5 总结

本文针对挑战赛题目一的内容，设计了全面直观的可视分析系统 CPISVis。既可以从时间和空间角度出发探索会场日程安排，又可以针对我们提出的人员划分依据，由表及里进行人员移动规律探索以及发现异常行为。CPISVis 能够有效的加强会场管理，处理各种突发状况，为会议各项议题的顺利进行保驾护航。

参考文献

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," the Journal of machine Learning research, vol. 3, pp. 993-1022, 2003.
- [2] Wong J A H A . Algorithm AS 136: A K-Means Clustering Algorithm[J]. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979, 28(1):100-108.
- [3] Laurens V D M , Hinton G . Visualizing Data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(2605):2579-2605.