专题引子

关于分类与预测

徐华

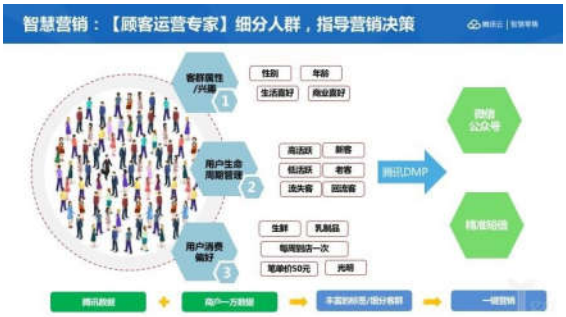
清华大学 计算机系 智能技术与系统国家重点实验室

xuhua@tsinghua.edu.cn

1

关于分类

- **现实生活中我们常常需要给不同的数据对象贴上一定的标签，以区别于其他数据对象**
  - 数据标签：颜色、地域、性别、语言等等
  - 某些标签来自于数据对象的其他属性（对应于 数据预处理中缺失值的填充，消费能力排行-视频1，消费人群定位 视频2 ）
  - **静态标签**：数据对象的固有属性决定的（人的肤色、年龄等）v.s. **动态标签**



2

关于预测的例子1

- **关于新冠人数感染的预测**

研究：保守估计武汉已有 5.4 万新冠感染者

2020 年 02 月 13 日 23:44 来源于 财新网

研究认为，至 2 月 9 日，武汉城内的新冠感染者达到 54000—90000 人；湖北省除武汉外的其他城市保守估计有 21000 人。湖北省对感染病人的收治措施还需升级。

2 月 10 日在 medRxiv 平台发布的论文“武汉市冠状病毒感染患者的统计推断”认为，至 2 月 9 日，武汉城内的新冠感染者保守估计达到 54000 人，多则达到 90000 人；湖北省除武汉外的其他城市，保守估计有 21000 人感染新冠病毒。该论文作者为南开大学统计与数据科学学院教授周永道、美国内布拉斯加州大学医学中心公共卫生学院生物统计学系的 Jianghu （ James ） Dong。medRxiv 线上平台 2019 年 6 月创立 ,由美国的研究机构冷泉港实验室(CHSL)、耶鲁大学和一家全球健康知识提供商 BMJ 共同创办，可以分享未经同行评议的研究。

该研究试图通过对武汉市部分人群的抽样调查 ,预估武汉市感染人群的总体数量。研究者选取了 3.3 万名从武汉返回温州的人员，以及 1 万余名从武汉赴新加坡旅行的人员为样本，以其新冠肺炎感染情况，来估算武汉及其周边区域的整体感染率。

根据 1 月 27 日至 2 月 9 日公开的新冠疫情确诊信息，温州市 448 名新冠确诊患者中，有 202 人有武汉及其周边区域接触史，而从武汉及周边区域返回温州的总人数为 3.3 万，他们均在 1 月 29 日前返回温州。因此，研究者测算，截至 1 月 29 日，这一人群的感染率约为 0.61%。但研究者同时表示，由于温州人大多在武汉经商，比普通人相比有更多机会接触他人，以此推测的武汉整体感染率可能不够准确。因此，研究者决定根据在新加坡旅行的武汉游客的发病率，对这一结果进行修正。

从 2019 年 12 月 30 日到 2020 年 1 月 22 日，共有 10680 名武汉人员旅行至新加坡，他们中至少有 33 人确诊为新冠病毒感染。研究人员表示，这一人群截至 1 月 23 日的新冠肺炎感染率不低于 0.3%，截至 1 月 29 日的感染率则应该更高。

综合这两个样本群体的感染率，研究人员认为，截至 1 月 29 日，武汉的新冠肺炎感染率在 0.3%-0.6%之间。根据武汉市政府在 1 月 26 日新闻发布会的数据，春节期间，共有 500 万人离开武汉，900 多万人仍留在武汉，武汉原有 1400 万人。以 0.3%的感染率计算，武汉有约 42000 人感染新冠肺炎，其中 27000 人在武汉市区，15000 人在武汉周边区域；以 0.6%的感染率计算，则有 84000 人感染，其中 54000 人住在武汉，30000 人住在周边区域。

从 2020 年 1 月 29 日之后的十多天中，温州市采取了严格的控制措施，包括集中隔离疑似病例和密切接触者等 ,但温州市的新冠确诊病例数依然翻倍不止。财新记者查询数据显示，1 月 29 日，温州市累计确诊病例 172 例。至 2 月 9 日已增至 448 例，是 172 例的 2.5 倍。

但考虑到武汉市有居家隔离的疑似病例和轻症确诊病例 ,加大了这些家庭内部传染的概率。所以研究者估计 ,同期武汉市内的感染数将达到 54000—90000 例。其中，下限 54000 例是上述 27000 例的两倍；上限 90000 例则是考虑了家庭内部传染概率增大的结果。

研究者还依此推算，全国感染数在 84000—140000 例之间。由于离开武汉的 500 万人中有 70%回到了湖北省其他地市，可以推知湖北省内武汉以外的感

3

关于预测的例子2

- **中医护心预警手表**
  - **安顿护心预警手表（视频3）**

4

关于预测的例子3

- 有些场景下，依靠常识不如去作一下测算，也许更准确
  - 不如去做一下测算（视频4）

5

关于分类与预测

- 在分类之前都有标签吗？
- 如何做分类（贴标签）或者预测（连续型变化的数值）？
- 如何评估分类和预测的效果？
- 分类和预测模型完成后，是否说明了原因和结果的关系？

6

Thanks !