

洞察力强分析员 可用“小数据”解决问题

陈劲禾 报道
jinhet@sph.com.sg

数据科学要能改善生活，不一定要靠大数据或复杂的算法达成。政府的数据科学专家说，在洞察力强的数据分析员手中，就连“小数据”也可以解决问题。

政府科技局数据科学署代副司长林猷茂昨日在一场数据科学研讨会上，以地铁环线故障为例，说明政府科技局的数据分析员如何以少量的数据查出列车故障背后的蹊跷。

“环线的数据分析案例用的不是大数据，而是小数据。这组数据只不过是一份只有300排资料的Excel文件。”

他说，这组数据只记录了列车故障的日期和时间、列车起点站、列车在哪里故障，以及列车编号，却足以让数据分析师在三小时内发现列车故障的模式。

原来是因为有一列列车会发出破坏性信号，导致反方向列车故障，因此每当那列“捣蛋列车”向一个方向行驶，它越靠近终点站，反方向列车故障的地点就会越靠近起点站。

林猷茂说，数据分析师是参考了美国波斯顿地铁系统用来分析地铁行驶资料的Marey图表，然后把数据转换成类似图表再加以分析才得出结论的。

“这种分析工作用的不是什么花哨的机器学习技术，而是懂得把数据形象化。”

林猷茂是在新加坡国立大学理学院旗下的数据分析咨询中心主办的首届数据科学研讨会上演讲，并在主题演讲中分享了几个政府应用数据科学的案例。

他举的另外一个例子是建屋发展局前线的客服人员如何利用数据说服高管改善预购组屋屋主领取钥匙的流程。

林猷茂说，建屋局的产业行政与房地产部门每年平均会收到10万个电邮，各种类型的询问都有。

他说，虽然客服人员知道屋主要更换领取钥匙的日期是非常普遍的问题，流程迫切需要简化，但高管要操心的事太多，这个看似芝麻绿豆的小事一直都被忽略。

他说，政府科技局的数据分析员用计算机把电邮按内容分类，发现有关改期领钥匙的电邮果然占多数，客服部门最终以这个数据说服了高管正视这个问题，成立网上系统让屋主自己更改日期。

林猷茂说，这名数据分析师因为帮忙解决了问题而获得许多赞美，不久后就跳槽到电商公司Lazada去了。“这说明如果你是一个能够解决现实问题的统计学毕业生，你的技能是非常抢手的。”