

文/杨书卷

“大数据”魅力正现

“我女儿才 17 岁,你们竟然给她发婴儿尿片和童车的优惠券!”美国一名男子闯到他家附近的 Target 店铺大声抗议,店铺经理立刻前来致歉。一个月后,这位粗心的父亲却惊呆了:他的女儿的确怀孕了。

这是一个在微博上流传甚广的真实案例,它的始作俑者,正是近一年来被炒得沸沸扬扬的“大数据”。

Target 是美国第二大的百货超市,它的市场经理一直在绞尽脑汁寻找把孕妇“截留”在自己商店中的方式,因为孕妇属于“含金量”极高的顾客群体,但一般会去专门的孕妇商店采购而不是 Target 这样的普通超市。“你怎么在孕妇第 2 个妊娠期就把她们给确认出来?”顾客数据分析部高级经理 Andrew Pole 接到了这样的棘手指示。因为等孩子出生后,新生儿母亲就会被铺天盖地的产品优惠广告包围,那时候 Target 再行动可就晚了。

可怎么才能得到“怀孕”如此私密的消息呢?Pole 开始建模分析。他从一个“迎娶聚会”的登记表中发现了一些极为有趣的数据模式:许多孕妇在第 2 个妊娠期开始买大包装的无香味护手霜,以及大量购买补充钙、镁、锌之类的保健品。最后, Pole 选出了 25 种典型商品的消费数据,构建了“怀孕预测指数”,通过这个指数, Target 能够在很小的误差范围内预测到顾客的怀孕情况,并早早把对孕妇的优惠广告加在其他广告内,装作“不知情”地发给顾客。这一策略显现出惊人的效果:孕期用品销售呈现了爆炸性的增长,而上述故事则是其中的一个意外小插曲。

这就是大数据的实质:从各种各样类型的数据中,快速获得有价值的信息。 Pole 的大数据分析技术从孕妇中尝到甜头后,又推广细分到其他客户群,这一举措使 Target 的销售额在 8 年间从 440 亿美元增长到了 670 亿美元,大数据的威力由此可见一斑,而这仅仅只是冰山一角,

随着各类数据量爆炸性地增长,大数据技术几乎在所有的社会领域都有着极强的应用潜力。

其实,从计算机诞生伊始,对数据进行分析的软件工具就层出不穷,大数据技术能从传统的数据分析工具中脱颖而出,成为“风口浪尖”的最前沿技术,是因为它是一种新的数据处理模式,具有更强的决策力、洞察发现力以及更快的处理速度。在它面前,以前常用的处理简单数据的“关系数据库”完完全全变成了“小儿科”。

例如,环境是一个极其复杂的系统,治理污染就需要大数据技术的帮助。如果

合欧盟各国的资源和信息,进行数据分析,支持犯罪调查。

欧洲网络犯罪中心将重点打击一些有组织犯罪集团实施的非法网上活动,尤其是攻击网络银行、网上儿童性虐待、损害网络重要基础设施和信息系统的行为。由于网络犯罪的特殊性,在调查时往往涉及数以百计的受害者,加上犯罪嫌疑人又在全球不同地区,单靠一个国家的警力难以胜任。欧洲网络犯罪中心就可以整合欧盟各国的资源和信息,使用计算化的地图,跟踪犯罪的各种蛛丝马迹,以找到最终的解决方案。

也许,有些天才凭着敏锐的直觉也能走上成功的巅峰,但对于绝大多数人来说,还是值得一提统计学界的一句名言:“我们信靠上帝。除了上帝,任何人都必须用数据来说话。”

我们要处理湖泊的污染,首先要通过传感器捕获水质情况,即溶解氧、水温、电导率、氨氮、pH 值等参数。然后,结合视频图像、人工采集和卫星遥感的数据一起分析。因为大数据的特点就是数据来源的“多源异构”,这样的数据才能保证判断的可靠性。

收集到各类数据之后,后台再进行分析挖掘,建立模型,跟历史数据进行对比,同时利用遥感技术、地理信息系统、和全球定位系统和云计算等,最终实现对污染的处理方案。在这个过程中,不仅要进行复杂的数据分析,还需要建立模型进行仿真,最后得到的也不是一大堆不知所云的数据,而是直观的可视化结果,而这一切所蕴含的技术复杂度,已远远超出了我们的想象。

令人欣喜的是,“大数据技术”也能在人们现在最关注的“网络犯罪”问题上大显身手。欧盟委员会有统计显示,12%的互联网使用者曾遭受过网络欺诈,全球每天约有 100 万人成为各种网络犯罪的受害者,而网络犯罪每年导致受害者损失大约 2900 亿欧元,欧洲网络犯罪中心已于 1 月 11 日在荷兰海牙正式成立,它将集

早在 2012 年 2 月,《纽约时报》就刊载文章认为,“大数据时代”已经降临,正在对每个领域都造成影响。例如,在商业、经济

及其他领域中,决策行为将日益基于数据和分析而作出,而并非基于经验和直觉;而在公共卫生、经济发展和经济预测等领域中,“大数据”的预见能力也已经崭露头角。而谷歌搜索、Facebook 的帖子和微博消息也使得人们的行为和情绪的细节化测量成为可能。例如,搜索引擎 Google 就有个著名的大数据应用——心情分析工具,它能从网民的上千万条留言中分类出 6 种心情,进而对道琼斯工业指数的变化进行预测,准确率竟达到 87%。通过不同维度的数据分析和服

务,新的社会检测模式正在源源不断地被创造出来。虽然我们的社会已如此依赖数据,但也有例外产生。众所周知,苹果创始人乔布斯就不喜欢搞市场调研这一套。记得当初有人问乔布斯:“如何通过市场调查了解大众需求,才能让产品如此成功?”乔布斯的回答是:“不用做调查,消费者并不知道他们需要的是什么,而苹果会告诉他们什么才是潮流!”也许,有些天才凭着敏锐的直觉也能走上成功的巅峰,但对于绝大多数人来说,还是值得一提统计学界的一句名言:“我们信靠上帝。除了上帝,任何人都必须用数据来说话。”