编者按 物联网、云计算、社交网络等新兴技术的蓬勃发展，使得人类社会从以信息为中心的 IT 时代走向了以数据流为中心的 DT 时代。DT 时代的核心是基于大数据的巨大价值的挖掘。但在挖掘的过程中，大数据创新应用还面临着数据共享、数据安全、隐私保护、人工智能应用等的机会与挑战。

数据挖掘：隐私保护与价值挖掘的度如何把握？

作者 | 吴俊宇

2016年，iPhone用户数据泄露、LinkedIn超1.67亿个账户被销售等事件，敲响了大数据安全隐私保护的警钟。当时，LinkedIn1.67亿个账户中有1.17亿个账户信息同时包含电子邮件和密码。而且更值得关注的是，互联网公司也在深入挖掘用户数据背后的价值。

大数据价值显现 数据安全存隐忧

大数据的价值是显而易见的。当样本的数据量“达到某个拐点时，一切都变了”，统计学上的意义便凸显出来。但是，对于更多数据的获取，尤其是像基因等私人敏感的数据，由于各种原因，对于这些被研究者来说已经成为一个巨大的挑战。知识挖掘、机器学习、人工智能等技术的研究和应用使得大数据分析的能力越来越强大，同时也对个人隐私的保护带来了更加严峻的挑战。

一些大型的互联网公司能够将大量数据结合在一起，从而构造出某个人清晰的行为图谱，进而预测他们的偏好与行为。这些数据在消费者市场上非常有价值，能够精确地向确定的人群主动推送某些产品或者服务。

不幸的是，进入移动互联网时代，更多的个人数据每天在产生和曝光，但数据安全性却无从保障。

中国互联网协会发布的《2015中国网民权益保护调查报告》显示，78.2%的网民个人身份信息被泄露过，63.4%的网民个人网上活动信息被泄露过，网民因个人信息泄露、垃圾信息、诈骗信息等现象导致的总体损失约为805亿元。

2016年据媒体报道称，个人乘坐飞机的记录，还有通信运营商的数据、银联的数据，大都可以通过黑市买到。而且电商平台购物导致个人信息被贩卖，如何追究平台责任和卖家责任，法律并不清晰，而且很难取证。

除此之外，企业对于数据的使用目前也很难有所约束。因为用户的身份、地址、联系方式等信息目前都掌握在互联网公司手中，但与这些数据每天打交道的人到底要如何使用数据以避免隐私泄露，是一个非常值得思考的问题。

保护数据安全 需技术与法律并行

针对上述数据泄露导致的安全问题，目前业内已经出现了一些解决方案。尤其是在约束数据使用上，已经有不少互联网公司正在采用数据“脱敏”技术。

数据“脱敏”是指对某些敏感信息通过“脱敏”规则进行数据的变形，实现敏感隐私数据的可靠保护。在涉及客户安全或者商业机密的情况下，在不违反系统规则的条件下，对真实数据进行改造并提供测试使用，如身份证号、手机号、卡号、客户号等个人信息都需要进行数据“脱敏”。

数据“脱敏”的确可以让数据使用更为健康，即当个人信息与某个具体的人或者设备相关联时，一些隐私保护技术可以设法去除数据与个人身份之间的连接，同时，另外一些技术在努力把这些断开的连接复原。当知道一个人所关联的一些信息，就可以从不包括其个人识别信息的数据中推断出这个人的身份标志。

当然，除了上述技术外，法律层面上也需要对数据的使用进行规范。

2012年3月，欧盟就提出了相应的法规，《数据保护法规》(TheDataProtectionRegulation)。违反数据保护条例处罚最高可达公司全球营业额的4%，即对于谷歌这样的科技巨头而言，一旦处罚将会是几十亿美元。这一法规在去年已经审议通过，而且在两年内将GDPR 条款转置成欧盟成员国法律，并将于 2018 年生效。

不过，有关互联网隐私数据的相关法律目前在国内还处于模糊状态。相比欧美在法律上对于数据与隐私的保护，在我国的法律中仅提及“公民的个人数据不得非法搜集、传输、处理和利用”，但我国《民法通则》并未将隐私权作为一项独立的人格权加以保护，所以隐私权方面，我国的立法暂时较为不明确。不过，随着互联网公司与用户信息的碰撞，未来还会出现更多现实案例，而国内法律也会在这种碰撞中逐渐完善。

编辑 | 孙永杰 sunyongjie@bjxintong.com.cn