



## 驾驭大数据（下）

文 / 徐端

大数据能在各行各业发挥其他工具完全无法代替的作用，但大数据并不是万能的，并不是任何时候、任何场合都适用的。大数据本身也有局限性，在大数据成为一个热门话题的今天，我们不能迷信大数据，而是需要弄清楚状况，知道什么时候需要使用大数据，什么时候需要使用其他工具。

几年前，世界爆发金融危机时，一家大银行的CEO做出一个让很多人都觉得不符合常规的决定。考虑到经济的疲软以及未来欧元危机的前景，很多人认为他应该会退出意大利市场，可是他最终决定留在意大利挺过任何潜在的危机。做决定前，这位CEO让手下的智囊团预测出可能会发生的一系列不利情况，计算出这些情况对于公司意味着什么。但是最终，他还是根据价值判断做出了决定。他的银行已经在意大利经营几十年，他不想让意大利人觉得他的公司是一个不可以共患难的朋友，他也想让公司里的员工觉得时局艰难时公司不会转移，即便这样做会有一些短期的成本损失。他在做决定时没有忘记数据，但最终，他遵循了另外一条思路。结果表明，这条思路无疑是正确的。商业有赖于信任，信任是带有感情的互惠行为。在艰难时期仍然坚守诚信的公司和人会赢得别人的好感和自尊，即便这些不易通过数据来衡量也是极有价值的。

这个故事里面暗藏了大数据分析的优点和局限。在当今这一历史性时刻，用于数据收集的计算机正调节着我们的生活。在这个世界，数据可以用于理解令人难以置信的复杂情况，可以帮助我们弥补自己直觉上的过度自信，帮助我们减轻因为情

感、观念、经验等主观因素导致的对事实的扭曲。但是，还有很多事情，大数据是无能为力的。

比如，大数据对准确描述社会活动是无能为力的。人的大脑在数学方面很差，但是在社会认知上很优秀。我们总能从一个人的面部表情的微弱变化捕捉到其很细微的情绪，从一个微小的动作判断对方的心理状态。同时，我们很多时候需要用情感来对一些事物进行价值判断。这些方面，大数据并不擅长。大数据分析本身是由计算机来进行的，它善于衡量社会交往的数量而非质量。比如，一个社交网络专家或许可以通过大数据分析绘制出你在平时80%的时间里与常见的10名同事或朋友的交往情况，但他没办法通过大数据分析捕捉到你对在某个很遥远的地方生活的近些年从来没有见面的前女友的复杂情感。因此，在做有关社会关系的决策时，要想用办公桌上的粗糙机器替代神奇大脑的想法是很浅薄和愚蠢的。

大数据在解决很多领域的重大问题方面也有局限。一个公司可以做一个随机对照试验来判断到底是哪一封促销邮件勾起了用户的购买欲，但一个政府不能用同样的办法来刺激萧条的经济，因为没有另外一个社会作对照。怎样能够刺激经济增长，这个问题经济学家和政府官员都很关心，也引发过很多争论。关于这个问题，我们有堆积如山的数据可用，但是没有哪位参与争论的人会被数据说服。

而且，大数据分析更偏向分析潮流和趋势，对一些突出的、特异的却没有办法。当大量个体对某种文化产品迅速产生兴趣



时,大数据分析可以敏锐地侦测到这种趋势,但其中一些可能非常杰出的东西从一开始就被数据摒弃了,因为它们的特异之处并不为人所知。

另外,数据本身也有局限。纽约大学教授丽莎·吉特曼有一本学术著作叫作《原始数据只是一种修辞》,书中指出,数据从来都不可能是原始存在的,因为它不是自然的产物,而是依照一个人的倾向和价值观念而被构建出来的。我们最初定下的采集数据的办法已经决定数据以何种面貌呈现出来。数据分析的结果看似客观公正,但其实价值选择贯穿了从构建到解读的全过程。数据会掩盖价值。没有任何数据是原始的,往往是根据人的倾向和价值观构建起来的。最终的结果看起来很无私,但实际上从构建到演绎一直伴随着价值选择。

这并不是说大数据就没什么了不起的,而是说数据和其他工具一样,在一些方面有优势,而在另一方面则有缺陷。

### 个人隐私的战争

有一段时间,谷歌执行董事长埃里克·施密特成了网络红人,谷歌的不少用户都将自己的网络账户头像更换成了埃里克·施密特的照片。这不是因为施密特是他们的偶像,而是用户对谷歌新广告的抗议。这次大规模的抗议让施密特陷入一个尴尬境地。这是怎么回事呢?

原来,谷歌推出了一个名为“共同代言”的广告政策,这个政策放宽了对个人资料的限制,它将允许广告商使用谷歌用户在Google+社交网站以及谷歌其他服务上的姓名、照片、评论等信息,以便为自己的广告产品背书。这一举措引起很多用户的强烈反感,他们认为这侵犯了个人隐私,于是他们“以牙还牙”,将“无辜”的施密特作为头像,让施密特也感受一下到处都是自己做广告的心情。

这一次,谷歌只是一时处在了“风口浪尖”。其实这背后暗含的是一场无法避免的大数据和个人隐私之间的“战争”。

谷歌新广告政策所谓的“共同代言”,究竟是怎么回事?

举个例子,你用谷歌搜索某一个餐厅,网页上可能会出现你的朋友“张某某”、“李某某”等在这一餐厅用过餐后点了“赞”或是作出其他评价,而且“张某某”、“李某某”都是实名制的,他们的头像也赫然在目。然后,你会对这个餐厅很有感觉,你可能也会依照朋友们的评价来做出一些消费选择。这样一来,“张某某”、“李某某”无形中就为这家餐厅做了广告,成为其形象代言人,甚至有一天,他们还可能得到一些广告费的分成。

如果你是“张某某”或“李某某”,你是否真的愿意做这样的广告,把自己的行迹暴露给其他人?即使广告的对象是自

己的朋友,你可能也不是每一次都愿意无条件分享的。更何况,对你来说,你可能根本记不得以往在哪些地方留下过“到此一游”的墨宝,天知道这种广告会出现在哪里,又会被哪些人看到。

不过,谷歌也做出解释:是否参与“共同代言”,其实是用户自主选择的。只不过,用户如果不修改自己共享的批准设置,就是默许谷歌在广告中使用用户的各类信息。

即使如此,谷歌的这一新广告政策还是遭到“侵犯用户隐私”的质疑。美国电子隐私信息中心主管马克·罗特伯格也指责谷歌新广告将网络用户照片等信息商品化有失公允,他同时呼吁美国联邦贸易委员会介入调查评估。谷歌用户们更是纷纷表示抗议,据外媒报道,短短时间里,许多Google+用户已经把自己的个人简介照片替换成施密特的照片,一旦谷歌决定在广告中使用用户的照片,这些照片显示都将是施密特本人。

如果说谷歌新广告政策是将用户信息直接拿来为己所用,侵犯了用户的个人隐私,那么,在当下这一个大数据时代,几乎所有线上的商业行为本质上都是通过对用户个人隐私的洞察或利用来获取商业收益的。

我们每一个人在网络上的一切行为,都可以被服务方知晓。当我们浏览某一个网页、发了条微博、逛了逛某个社交网站抑或是进行了一次网络购物,等等,所有的举动实际上都被网络以数据的形式记录下来。而基于对这些数据的综合分析,谷歌可以清晰地掌握你的网页浏览习惯,而亚马逊、淘宝等电商十分了解你的购物习惯,微博、微信等则似乎什么都知道,包括你的社交关系网络。除此之外,还有网络地图和定位,更是能够实时监控你的行踪。

这些所有的“你”的信息,原本都是个人隐私,如今却成为互联网商家们的“库存”数据。商家们可以随时拿来识别“你”,并对你进行商业价值的挖掘。谷歌新广告的做法只是其中之一,商家们基于大数据的所谓精准营销则是更普遍存在的隐私利用形式。

比如,你在某个社交网站上偶然提及某种产品或服务,这类型的产品或服务就能主动找到你;又如,你想要网购一双雪地靴,在某个电商网站上浏览过该类产品,那么相关的雪地靴介绍或者购买的链接就会在未来一段时间内不断推荐给你;再比如,你仅仅是通过网络地图搜索过某一个餐厅的地址,你可能都没有想过去消费,但关于餐厅的介绍或是其各个时段的促销优惠信息就会定期推送给你,这其中自然还包括谷歌新广告所涉及的好朋友代言。

这些商家的做法,看似更懂消费者,能够直达消费者,但这何尝不是对消费者隐私的一种侵犯呢?“大数据时代,几乎没什么隐私可言”,说得一点都不为过。你是谁,你在哪,你的喜好,你的消费习惯,你此刻想要做什么……大数据都能给出



准确的或接近准确的答案。对于互联网企业来说，基于大数据资源去掌握更多的用户隐私，才能迅速占领更多的市场份额。甚至有业内人士将互联网的下一轮竞争直白地定义为“隐私之战”，用户的隐私就是核心竞争力。

业内人士坦言，大数据时代要保护个人隐私，让用户个体的隐私不泄露，在技术上很难做到。我们唯一能够而且必须要采取的措施，就是让这些用户的个人隐私不被无良商家、非法机构恶意地使用。一方面，这需要政府和行业监管机构加强监管，但凡发现有恶意使用用户隐私并且给用户造成伤害的商家，就要给予严厉的惩罚。

### 未来的福尔摩斯

小说里的神探，不管是福尔摩斯、波洛还是狄仁杰、柯南，都有一个共同的特点，就是有一个具备强大分析能力的大脑。他们能够观察到细小的证据，并把这些证据关联，分析出犯罪事实。一个比较残酷的现实是，神探几十年出一位，罪犯却天天在行动。幸运的是，最新的大数据分析工具正在将每一位普通警员都变成神探，甚至能预测犯罪。最新的案例显示，大数据在警务中的应用已经收到了良好的效果。

作为美国警界最早的大数据预测分析试点单位，圣克鲁斯警察局通过城市大数据预测犯罪地点和时间。大数据分析可以帮助警察分析历史案件，发现犯罪趋势和犯罪模式，找出共同点和相关性，通过分析城市数据源和社交网络数据，甚至能预测犯罪。过去需要几天、几周甚至几个月的数据资料分析，在最新的警用Hadoop大数据分析系统中几个小时内就能完成，从而大大提高了警察办案的效率。

警方使用的大数据分析工具，可以帮助分析人员采集和分析文本、图像以及其他信息，还可以进行文本提取、案例组织等工作。此外，它的协同工具还可以避免重复的工作以提高效率。它的联合查询和模糊查询的功能可以让用户在即便拼错单词的情况下，以一个查询语句访问不同的数据库。

据国外媒体报道，美国孟菲斯市警察局采用BlueCRUSH预测型分析系统后，过去5年暴力犯罪率大幅下降。最近美国马里兰州和宾夕法尼亚州也采用了一种能极大降低凶杀犯罪率的犯罪预测软件，不但能预测罪犯假释或者缓刑期间的犯罪可能性，还能成为法庭假释条款和审判的参考依据。例如，用软件分析发现14岁第一次杀人比30岁第一次杀人的罪犯更容易再次行凶，因而调整假释条款。

大数据在查案办案方面将能够起到非常重要的作用。一个能够对犯罪案件相关大数据进行数据分析的人，便是未来的福尔摩斯。

每个人都知道，一个人在犯罪前，无论大脑里有过多少

有关犯罪的预谋，都不构成犯罪事实。而在斯皮尔伯格导演、汤姆·克鲁斯主演的电影《少数派报告》里描述了这么一个时代：随着科技的高度发展，人类发明了能侦察每个人的脑电波的机器人“先知”。“先知”能侦察出人的犯罪企图，所以罪犯在犯罪之前就已经被犯罪预防组织的警察逮捕并获刑。这一点与大数据时代的预防犯罪有共通之处。这不禁让我们思考：在大数据时代，究竟怎样才是真正的犯罪？

IBM的一则广告片讲述数据分析如何帮助警察在罪犯作案前赶到现场预防犯罪，可以说电影《少数派报告》的现实版正在上演。作为警用大数据市场的重要厂商，IBM的SPSS预测分析软件以及i2COPLINK数据库应用软件可以提供数据整合数据分析，以及数据可视化功能。通过这些软件，可以进行数据集中管理，帮助警方不同部门协同工作，并可以帮助发现犯罪线索。在美国南卡罗来纳州的查尔斯顿，警方利用IBM的数据分析工具帮助当地的400多名警察更加准确地进行犯罪模式的分析。根据IBM的说法，警方利用分析预测工具进行警力调配，发现犯罪热点地区提前预防犯罪发生，从而降低了当地的发案率。在美国，从纽约到洛杉矶，有越来越多的警察局在计划或者开始部署大数据分析工具。很多警察局认为，大数据分析工具从长期来说可以加速办案效率，优化警力资源分配，从而提高社会和公众安全水平。

在这两个案例里，警察还是在遵照现实生活中的逻辑：没有行动的犯罪不能定罪。如果大数据分析到非常精准的程度，甚至可以算出某人在某个时间犯罪的概率有多大，事情会不会发生变化呢？比如大数据分析显示，某个人某天深夜可能在家谋杀他的妻子，可能性高达99%。警察该如何去预防犯罪呢？如果他们相信大数据分析的结果，他们就要高度监视这个人或者在危险将要发生的时候破门而入，这都是有风险的。前者是没有证据而假定一个人会犯罪而去监视，侵犯了个人隐私；后者则难度很大，破门而入的时机早了，犯罪还没有任何迹象，则是警察侵犯了个人安全，如果时机晚了，犯罪已经发生，警察的责任则更为严重。

未来到底会怎么发展呢？谁也不知道。我们还身在大数据时代的前奏曲里，只能慢慢随着这个壮丽的曲子走下去，看看之后这些到底会怎么发展。

### 算法不能代替判断

有个笑话是这么说的：3位统计学家去郊野打猎。他们发现了一只兔子。第一位统计学家率先开枪，结果没打着，子弹大约往左偏移了一米远；第二位统计学家开枪射击，结果还是没打着，子弹大约往右偏移了一米远；第三位统计学家大喊：道：“我们逮住它了！”



你发现这个笑话的笑点了吗？这个笑话的笑点在于嘲讽“平均数”，一个往左偏移一米，另一个往右偏移一米，平均算起来就是正好打中了。这个笑话的荒谬之处如同“你家有100平方米的房子，邻居家有1000平方米的房子，你们两家平均住着550平方米的房子”。也许你不觉得这个笑话可笑，但是你可能会发现，现实生活中许多的人就是在这么计算的。你不能说他们的计算是错误的，但显然得出的结论是毫无意义的，或者说得出的结论是用来蒙蔽真相的，就像乔布斯曾和比尔·盖茨说“比尔，我们两个人统治了全部的电脑操作系统”。比尔·盖茨微微一笑，什么都没说。要知道，全世界90%的电脑都使用微软的windows系统啊。因此，乔布斯这句话在知道底细的人看来是非常“自抬身价”的。

大数据的世界里，计算机要处理和分析相当庞大的数据，这些数据如何被处理、分析，用着怎样的算法，最终会得出怎样的结论，这一点是可疑的。如果在大数据的处理分析中，算法师使用的是一个如同上文中统计学家的算法一样的算法，那么大数据呈现给我们的数据可能会让我们对现实情况产生误判。

不仅是算法，数据本身的采集可能也是有问题的。就像新闻报道中文字游戏非常普遍，这也可能对大数据本身产生影响。例如，某厂的利润第一年为0，第二年为100万，第三年为200万，今年（第四年）利润为250万。我们可以有很多方式来报道这个新闻：某某厂今年利润增长25%；某某厂今年利润增速降低50%；某某厂平均每年利润为137.5万元；某某厂利润增长渐趋稳定……这么多种表述方式没有一个是错误的，可是带给我们的感觉却完全不同。

我们或许可以在内特·希尔的著作《信号与噪音》中找到这些问题的精彩答案。希尔引用了菲利普·泰洛克对专家意见所进行的经典研究。这项研究显示，数量多得令人不安的专业领域的“专家”在预测可能结果方面的表现往往差得离谱。此外，专家们往往对其预测的质量过度自信，简言之，专家意见时常获得两个世界的最差结果：以妄自尊大的态度给出了错误答案。这不是成功的秘诀。

从IBM的超级电脑“沃森”、谷歌的搜索算法到亚马逊网站的推荐引擎，数据驱动的计算系统无疑能够获得非凡的成功，特别是当它们专注于现实生活测试而不是抽象理论的时候。希尔说：“真正像谷歌一样懂得大数据的公司并没有将大量时间花在构建模型上。这些公司每年从事数十万次实验，在真实的顾客身上测试自己的想法。”

从希尔的书里我们可以得出一个颇具讽刺意味的结论：一个人获得的数据和事实越多，预测就越有意义，人的判断也就显得愈发重要。这好像不是我们想要的答案。我们希望能够通过大数据分析直接告诉我们结果，告诉我们该怎么做，尽量减

少我们的判断，而事实上，我们的判断变得越来越不可或缺。

## 没有隐私的世界

2013年6月，一本名为《1984》的书突然间火了起来，亚马逊网上书城记录，这本书的销量增长了70多倍，在24小时里甚至涨了500多倍。这是怎样的一本书？推动销量增长的原因又是什么呢？

《1984》是英国作家乔治·奥威尔创作的一部政治讽刺小说，小说创作于1948年，出版于1949年，已经被翻译成至少62种语言。它与1932年英国赫胥黎的《美丽新世界》以及俄国尤金·扎米亚金的《我们》并称“反乌托邦的3部代表作”，通常也被认为是政治小说文学的代表作。书中讲述了一个令人感到窒息和恐怖的以追逐权力为最终目标的假想的未来极权主义社会，通过对这个社会中的一个普通人温斯顿·史密斯的生活描写和对抗无时无刻不在监视着他的“老大哥”最后完全失败的命运，投射出现实生活中极权主义的本质。《1984》曾在某些时期内被视为危险和具有煽动性的，并因此被许多国家（不单是有时被视为采取“极权主义”的国家）列为禁书。本书被美国《时代》杂志评为“1923年至今最好的100本英文小说之一”，此外还在1956年、1984年改编成电影上映。

《1984》对英语世界产生了意义深远的影响，书中创造的一些术语和小说作者已经成为讨论隐私和国家安全问题时的常用语。比如“奥威尔式的”形容一个令人想到小说中的极权主义社会的行为或组织，而“老大哥在看着你”（BIGBROTHERISWATCHINGYOU，小说中随处可见的标语）意指任何被认为是侵犯隐私的监视行为。这一次，这本离第一次出版已经经过50多年的旧作，是因为一个叫作爱德华·斯诺登的人而火起来的。

爱德华·斯诺登是一个美国人，生于1983年，曾是美国中央情报局技术分析员，后供职于一个国防项目承包商。2013年6月，斯诺登把美国国家安全局关于“棱镜”监听项目的秘密文档披露给《卫报》和《华盛顿邮报》，很快，他被美国政府通缉。斯诺登早就预料到会是这样，所以事发时人并不在美国，而是在香港，随后他飞往俄罗斯并获得为期1年的临时避难申请。后来，斯诺登又向全世界披露了很多轰动性的信息。2013年12月11日，美国《外交政策》杂志评选美国国家安全局承包商前雇员斯诺登为2013年全球百名思想家榜首。

斯诺登第一次披露美国国家安全局如何从技术公司获取电话记录和数据以后，奥威尔的《1984》销量就开始不断上升。很明显，人们由这次事件联想到《1984》里描写的世界。这个世界里，美国国家安全局就像书中的老大哥一样盯着你，毫无隐私可言。无论是美国人还是其他各个国家的人，很多人都像





看影视剧一样关注着这件事情的发展，却不知道有一个比美国国家安全局更为强大更为隐秘的存在在窥探着我们的隐私。这就是大数据。相比起美国国家安全局，大数据更像老大哥。

一些公司正凭借手中规模迅速增长的个人信息，利用各种新型数据分析方法和人工智能来进行产品和服务决策，乃至预测客户的需求。谷歌首席执行官拉里·佩奇表示，他眼中的理想技术就像“一名高度智能化的助手，为你做各种事情，免得你自己操心”。是的，技术可以为我们做很多事情，比如在不久的将来我们可能看到这样的场景：在我们起床前就有智能电饭煲已经为我们做好饭，洗漱完毕后打开手机看看它为我们怎么规划好的日程安排，坐在驾驶位置喝茶任凭谷歌无人汽车自动驾驶带我们去目的地……这种生活确实让人向往。但是，有多少人知道我们可能为这样的生活承担了多大的风险，付出了多大的代价呢？

多数人可能没有意识到，自己每天制造了多少数据，以及一些大数据企业用以利用这些数据的科技已经发展到怎样的地步。技术发展如此迅速，两年前还不可想象的事情如今已变得稀松平常。当你使用错误的关键词在搜索引擎上搜索时，搜索引擎会猜测到你实际想搜索的关键词，这不就是一种预测吗？实际上，大数据分析的预测功能比这个要强得多。哈佛大学教授拉塔尼娅·斯威尼的研究显示，只要知道一个人的年龄、性别和邮编，并与公开的数据库交叉对比，便可识别出87%的人的身份。而在大数据时代，通过分析用户4个曾经到过的位置点，就可以识别出95%的用户。社交网络和互联网公司收集的数据呈现出很强的身份特征。大数据公司之所以非常强大，是因为它们能够将客户的个人信息与他们的行为特征结合起来，从他们购买了哪些商品到他们身在何处。这可以帮助企业非常精准地猜测到用户可能存在的需求。你在那些公司的分析下完全没有任何隐私，就像裸体站在他们面前。

世界经济论坛的一份关于个人数据的报告里说道：“推测

数据可能像一个无所不知、盯着监控摄像头的‘老大哥’。”大数据分析本身没有任何危险和过错，但如果一些数据被用于特定的、不怀好意的目的，我们如何防范？**大数据时代才刚刚开始，未来，谁也说不好会不会出现一次因为大数据分析导致的信任危机。**一旦有因为大数据分析造成的严重的隐私泄露事件爆发，大数据分析可能就像今天的美国国家安全局一样，失去人们的信任。大数据有许多值得我们惊叹的地方，但我们要学习掌握好它，利用好它，不让它成为“隐私杀手”，可能还需要很长的时间。

### 计算机的危机

高登·摩尔，1929年1月3日出生于美国加利福尼亚州旧金山，他是仙童公司的创始人之一，也是英特尔公司的创立人之一，同时还是大名鼎鼎的“摩尔定律”的提出者。这几十年来，计算机计算能力的增长一直遵循着摩尔定律，而这样的增长给大数据的发展带来了机遇。

1965年的一天，《电子杂志》写了一封信给摩尔向他约稿，希望摩尔能为他们写一篇关于半导体工业未来发展的文章。1965年4月19日，《电子杂志》上刊发了一篇名为《Cramming More Components Onto Integrated Circuits》的文章，在这篇文章中摩尔总结了过去几年集成电路的发展情况，然后大胆预测了半导体工业的发展速度：由于晶体管工艺的提高，晶体管体积逐年减小，集成电路的整合度每年都会翻一番，也就是说半导体元件的性能和功能将会以几何数字逐年递增并且一直持续下去。这就是在以后的40年中神奇地支持着半导体工业发展的摩尔定律。后来，广为人知的18个月的说法是时任英特尔公司主管的大卫·豪斯的说法。

1975年，摩尔对摩尔定律进行了更加精确的修正：半导体集成电路的密度以及性能，每两年翻一番。同牛顿定律不同，



摩尔30年后补充说道：“我并不希望这个推论有多么精确，我只是试图让人们知道未来半导体技术是什么样子的，希望我们能向着这个目标而努力。”摩尔定律不是一个自然科学定律，而是一条融自然科学、高技术、经济学、社会学等学科为一体的多学科、开放性的规律。摩尔定律所带来的经济学效益，已经完全成为英特尔公司发展的潜在规则。

半导体总是讲究越小越好，在摩尔定律的激励下，英特尔已经投入量产应用的制程技术可以印制出比病毒还小的电路——比人类的头发细1000倍，同5个原子的高度差不多。半导体总是讲究越多越好，英特尔半导体芯片的复杂程度持续攀升：摩尔发表摩尔定律的时候每个集成电路中包含大约60个电子元件，而现在最新的英特尔安腾处理器中集成了17亿个晶体管。半导体总是讲究越便宜越好，半导体的世界中，越便宜的就是越好的。在1968年，集成电路中平均每个晶体管的成本是1美元，截止到2004年，每个晶体管的成本不超过0.000001美元，而这个价格同在报纸上印制一个单词的成本差不多。性能的提升、功能的增加、成本的下降使得半导体芯片应用到广泛的领域中：从数字电话到个人电脑到股票交易所到太空飞船。而驱动这一切的都是40年前发表在《电子杂志》第114页的那篇文章。

在摩尔定律的神奇激励下，现在计算机的计算能力已经相当强大，足以支撑我们这个大数据时代的数据分析工作。而另外两个问题则给大数据分析的未来带来了不确定因素。一个是摩尔定律的失效，另一个则是安迪·比尔定律。

2013年12月，美国通信芯片厂商高通创始人亨利·萨姆利表示，对于提升速度、降低功耗以及降低成本，芯片厂商目前不能三者兼顾，而只能专注于其中两方面。推动摩尔定律向前发展需要更复杂的制造技术，这样的技术本身成本昂贵，因此削弱了芯片换代带来的成本优势。现在的芯片制造工艺虽然本身仍有提升空间，但在未来15年中也将面临瓶颈。在进行3次换代后，芯片制造工艺将达到5纳米。在这样的情况下，每个晶体管栅极从头至尾的长度仅为10个原子。在此基础上，进一步发展是不可能的。他的观点代表了很多业内人士的看法。

这对大数据来说不是一个好消息。大数据正以不可思议的速度在迅速增加着，未来需要处理的数据将越来越庞大，如果计算机的计算能力增长变慢，开始面临摩尔定律失效的问题，那么大数据分析也同样会受到影响。这个不好的消息之后是另一个不好的消息，这就是安迪·比尔定律。

安迪·比尔定律也是广为人知的一条定律。安迪·比尔不是一人名，而是指英特尔的前任CEO安迪·鲁夫和众所周知的微软前CEO、现任董事长比尔·盖茨。安迪·比尔定律具体指什么？用句俏皮话来说就是安迪给我多少，比尔就拿走多少。无论计算机的计算能力提升有多强，无论英特尔生产出多么厉害的

CPU，微软总会用越来越臃肿的系统和软件统统吃掉增长的硬件性能，不管硬件性能提升多少，吃性能的新软件总能够化神奇为腐朽，导致在使用过程中新的计算机未必比旧的计算机更快。这就是安迪·比尔定律的本质。

无论摩尔定律拔高了多少硬件性能，人们都不得不年年掏钱升级新设备，这就是安迪·比尔定律的威力所在。

随着功能的加入和交互的革新，软件做得越来越大是很正常的事情。然而，软件臃肿化并不是单指软件体积膨胀，重点是软件增加的功能和变大的体积不成比例。安迪·比尔定律的主角是英特尔和微软，但其实在整个业界中软件臃肿化已是尾大不掉。无论是桌面平台还是移动平台，软件臃肿化的迹象都随处可见。当年微软推出windows vista时，硬件厂商特别是内存厂商都喜笑颜开，因为vista的前任XP仅仅依靠256M的内存就能很流畅，而vista在1G内存下才勉强达到；苹果的移动操作系统IOS4推出时只有600M不到，现在的IOS7已经1.3G；腾讯QQ当年只占几M内存，安装包大约10M，现在什么都不做也占近40M……整个大环境下，数据分析软件也免不了受到影响。

两条定律、3个名人，呈现出大数据时代的特殊烦恼。摩尔定律的失效和安迪·比尔定律的威力，都将使大数据分析面临一个问题：数据越来越庞大，越来越复杂，需要计算能力更加强化的计算机和功能更加强化的软件。而计算机硬件发展开始停滞不前，软件又变得更加臃肿和吃性能，计算机将面临计算危机。这无形中给未来大数据分析的发展蒙上了一层阴影。好在目前软件工程师和硬件工程师们都在极力寻找解决这些问题的办法，大数据分析的未来才显得依然光明。（全文完）BR



#### 作者简介

徐端：湖北孝感人。关注互联网十五年，见证了中国互联网从出现到蓬勃发展的全过程。兴趣广泛、涉猎颇广，勤于思考、善于研究，然经常停于设想、怠于行动，人称“一只脚在IT界，一只脚在文化界的树懒。”