

## 大数据真的是“未来的新石油”？

孔建会

2013-05-03 10:40:09

孔建会

当前IT业最热门的话题应该就是大数据了，这个两年前才被炒作起来的概念，现在已无处不在，占据了互联网的半壁江山，它与云计算一起成为未来信息技术发展最重要的内容。处于IT行业的人们，如果不知道大数据，不了解大数据所创造的那些神话般的故事，就好像已经跟不上信息技术的发展，成为时代的落伍者了。

那么什么是大数据？现在还没人能给出一个确切的答案，它同云计算一样，是一个形象但却笼统的概念。大家一致的理解是从4V的角度来描述这一概念的内涵。所谓4V就是说大数据具有体量巨大（Volume）、种类繁多（Variety）、实时处理（Velocity）、数据真实（Veracity）的特性。信息技术的发展，使互联网遍布世界的每个角落，各种各样的信息都可通过网络进行传播。世上所发生的每件事情，人类所进行的每项思考，社会所进行的每点进步，知识所增加的每种内容都可通过互联网进行汇集。网络技术、传感技术与射频技术使各种事物都可产生大量数据。事物的内容、时空、方位、特性都可转化为数据，这些经济的、社会的、规范的、杂乱的、简单的、复杂的等各式各样的数据代表了不同事物的内涵。代表不同事物的海量数据可以通过网络进行汇集，汇集后的整体数据反应了事物的本质特性或发展趋势。而这种反应事物本质特性或发展趋势的数据具有一定的价值，通过高性能计算机的数据提取、价值分析、数据表达等大数据技术，可以从大数据中把这些价值提取出来。不同的需要可以提取出不同的价值，所有的价值都可以为人类服务，这就是大数据的本质。

大数据，这一鲜明而又形象的概念之所以如此引人注目，是因为许多像亚马逊、谷歌这样的大公司为它杜撰了无数神话般的故事。每一个故事都告诉人们，大数据具有无所不能的洞察力。它可以象伟大的先知一样预测未来，可以利用玄妙的相关关系推测人的喜好，可以利用宠大的数据窥探人的内心，可以利用趋势分析引导人的行为，甚至可以用事物发展曲线来控制人的思维，改变人类社会的发展方向。

大数据利用最多的地方是商业领域。大家都知道尿布与啤酒的故事，通过大数据分析，找出这两个毫不相干的事物之间的关联，利用推荐系统，向不同用户提供不同的产品推销方案，提高产品的销售量；另一个众所周知的故事就是通过购物记录来分析少女怀孕的案例，通过分析用户的购买行为，推测用户的现在状态与未来需求。尽管这一案例因为侵犯别人隐私而广受批评，但却说明了大数据分析在商业领域的价值。还有无数这样或那样的故事，都说明大数据在分析人们的喜好与购买习惯上的巨大作用，利用这种分析，商家可以更好地进行营销。权威的市场调查机构说，大数据将会改变未来的商业规则，实时竞价系统（RTB）、更加智能的推荐系统以及基于大数据的商业智能分析（BA），使企业能够生产更加适合用户喜好的产品，能够找到向用户推销产品更好的方法，能够为用户提供更好的服务，能够利用有限的资源产生更多的效益，大数据的这些功能将会使经济社会更快更好地发展。

大数据的另一个应用领域就是社会公共领域。利用大数据，政府可以更好地管理社会，为社会提供更好的公共服务；教育部门可以更好地获取与传播知识，提高民众的素质；卫生部门可以预测重大流行疾病的发生，提前作出防范预警；安全部门可以监控社会状况，确保国家不受破坏。谷歌公司前几年利用大数据绘制了“世界流感地图”<sup>[1]</sup>用来帮助人们了解世界各地流感的发病情况；利用大数据预测禽流感在各地的爆发趋势，为民众提供预防建议，向卫生部门提供监控资料。在公共安全领域，大数据也有巨大的作用，如美国就建立了基于各种传感技术与大数据技术的网络监控系统，预测重大社会公共事件的发生，比如爆发大规模游行活动、可能遭受的恐怖袭击等。

随着科技的发展，大数据的应用也延伸到个人服务领域。穿戴式设备的产生标志着信息技术的发展到了以人为中心的阶段。像谷歌眼镜、苹果手表这样的穿戴式设备把技术与人体结合起来，使人类可以更加便捷地获取信息。这些穿戴式设备利用大数据为个人提供学习、生活、健康等方面的数据服务。大数据通过穿戴式设备发挥其作用，同时穿戴式设备本身也成为收集数据、构建大数据的来源。穿戴式设备作为一种工具，大数据是其服务的内容。

大数据有如此巨大的作用，让一些信息技术发达的国家倍加关注，很多大型信息公司也把大数据当作公司未来发展的重

点。美国是当今世界最为重视大数据发展的国家，也是利用大数据较好的国家。美国政府把大数据作为未来信息技术发展的重中之重，认为大数据将引起信息技术发展的革命，改变未来社会的发展模式。2012年3月美国政府发布了《大数据研究和发展倡议》<sup>[2]</sup>，把大数据称作“未来社会发展的新石油”，把发展大数据研究提到国家战略的高度上来。为掌控这一“未来的新石油”，引领世界技术发展方向，联邦政府制定了大数据研究发展计划，推进相关研究机构进行科技研究和创新发展。改进从海量数据中提取价值的方法，提高从数据中获取知识的能力，加速美国在科技领域中的前进步伐，确保美国在信息技术领域的领先地位。美国政府指定了国家科学基金（NSF）、国家卫生研究院（NIH）、能源部（DOE）、国防部（DOD）、国防部高级研究计划局（DARPA）、地质勘探局（USGS）6个联邦部门来负责推进大数据的研究，并计划投资2亿美元用来研究收集、储存、保留、管理、分析和共享海量数据的技术。

政府的投入与政策的支持，使那些本来就注重大数据研究的公司更加主动，纷纷加大对大数据研发与利用的投入，取得了不菲的成绩。如Apache基金会开发的Hadoop分布式处理系统，经过不断发展完善，现已成为大数据处理的标准结构；微软开发的Windows Azure平台使大数据的处理更加人性化、简化；还有其它公司，经过不断的创新研究让大数据技术更加完善、更加实用。技术的发展使大数据的价值更加突显，人们对它的期望也不断提高。据说谷歌公司利用大数据不但可以预测流感的发生，也可预测美国的经济复苏情况；不但可以预测明年服装的流行色，也可预测某位娱乐明星是否可以红遍世界。谷歌公司最近一个广为流传的大数据案例就是，在2013年的奥斯卡电影节上，它预测准了大部分的奥斯卡奖项得主<sup>[3]</sup>。Facebook也是利用大数据的典型公司，大数据是这家社交网站最大的价值所在。通过对大量社交信息的分析，Facebook可以了解用户的生活习惯与购买喜好，利用定向推荐广告，为这家年轻的公司每年创造上百亿的产值。

还有许多真实的或杜撰的有关大数据的美妙故事，每一个都反应了大数据在某一领域的巨大作用，这些故事的汇编，给人们描绘了一个大数据主宰的奇幻世界。在那里，人们利用大数据可以预测任何一件事情的发展趋势，可以完成一切不可能完成的事情。每一个人都是先知、都是智者，在大数据的支持下，每一个人都通天文、晓地理，前知一千后知八百，如齐天大圣般十八般武艺样样精通。然而真的是这样吗？

可惜的是，理想很丰满，实现很骨感。至少在当前大数据还没有那么大神奇，就连谷歌公司所说的利用人们的检索词可以用来预测流感的发展这种人人都相信的故事也没能发生。当H7N9在中国出现的时候，谷歌没能给出任何有关这一流感发生与发展的预测，更不用说事先给民众发出预警，提醒人们防范了；欧洲的金融危机持续了五六年，至今没有一家大数据公司能预测这场金融危机何时结束，就连未来是向好还是向坏发展的方向都不能预测；在商业领域，个性化广告推荐这种理可以做好的事，大数据也没能表现出其超强的功用。最近经过大数据改造的淘宝网个性推荐系统没比五六年前好多少；号称利用大数据使图书销售量大幅增加的亚马逊网站上推荐的书，没有一本是我喜欢的。

2013年4月15号，波士顿马拉松比赛现场发生了震惊世界的恐怖袭击<sup>[4]</sup>。美国人悲伤地发现，那些基于现代信息技术与大数据分析的网络监控系统，并没能确保国家的安全，大数据分析在防范恐怖袭击方面，并不象人们所期望的那样有效。看来这个被称为未来技术竞争至高点的大数据，在当下还不能如人们所期望的那样有用。

产生如此悲观想法的原因，一是虽然大数据可能代表了未来信息技术发展的方向，具有光明的前景。但是在人们的意识与相关的配套技术还没有发展起来的情况下，大数据还不能发挥其巨大作用。二是大数据还没有一个确切的定义，它所具有的内涵还不能被清晰地诠释。美国把大数据被称作“未来的新石油”，然而石油是不可再生的，数据却是随时随地都在产生，越是往后产生的数据越多，数据的准确性越高，对数据的处理能力越强，这让人们对现在就收集并占有大量数据的必要性产生怀疑；还有从大数据中提取价值首先要有需求，没有需求就不知道从大数据中提取什么，当人们的思维还没有发展到可以完全掌控数据，还不知道大数据到底有什么用处时，很难对其产生需求；没有需求就没有数据收集的方向，没有方向的收集就会造成巨大浪费；在方向不明的时候，唯一可做的就是收集所有可能收集到的数据，然而这样不但会给数据的存储、管理造成巨大的成本，也会产生一些法律的问题，例如是否会侵犯他人隐私等。再有，随着科技的发展，穿戴式设备会不断智能化并在生活中得以普及，这种以人为中心的智能设备本身就会不断分析所取得的数据，提取个人的喜好等特性，学习与适应个人的品性，具有人的品性特征，这样就使得从大数据中分析人的品性失去了意义。

以上这一切都让人们们对大数据的发展未来产生了怀疑。因此，当美国制定大数据发展战略，把它看成“未来的新石油”时，世界其它国家虽然也十分重视这一技术的发展，却没有制定相关发展战略。这有可能是其它国家在大数据技术方面落后于美国，还不能制定明确的发展目标，也可能是人们对这一技术的发展前景还没有一个清楚的认识，无法确定大数据是否真的可以引领未来技术的发展方向。因此，我们还是暂时忘了大数据吧！别再关注那些炙热的概念炒作，静下心来认认真真地想一想大数据的真正内涵。

#### 参考文献：

- [1] 《谷歌新工具能绘“流感传播地图”》，站长之家网站，<http://www.chinaz.com/news/2008/1113/44673.shtml>
- [2] 《美国推动大数据技术发展的战略价值及启示》，新华网，[http://news.xinhuanet.com/theory/2012-06/20/c\\_123311339.htm](http://news.xinhuanet.com/theory/2012-06/20/c_123311339.htm)
- [3] 《谷歌搜索猜中四项奥斯卡大奖获选名单》，新浪网，<http://tech.sina.com.cn/s/2013-02-26/08038090001.shtml>
- [4] 《美国波士顿马拉松赛发生爆炸》，新华网，[http://news.xinhuanet.com/photo/2013-04/16/c\\_124584570.htm](http://news.xinhuanet.com/photo/2013-04/16/c_124584570.htm)

文档附件:

---

隐藏评论

用户昵称:  (您填写的昵称将出现在评论列表中)  匿名

请遵纪守法并注意语言文明。发言最多为2000字符（每个汉字相当于两个字符）

9018

发表

中国社会科学院电话: 010-85195999    中国社会科学网电话: 010-84177865; 84177869    Email: [skw01@cass.org.cn](mailto:skw01@cass.org.cn)

投稿邮箱: [skw01@cass.org.cn](mailto:skw01@cass.org.cn)    网友之声信箱: [skw02@cass.org.cn](mailto:skw02@cass.org.cn)    地址: 中国北京建国门内大街5号

版权所有: 中国社会科学院 [版权声明](#)    京ICP备05072735号