**石 勇： 数据海洋里的淘金者**

大数据开启了一个大规模生产、分享和应用数据的时代，它给技术和商业带来了巨大的变化，并迅速渗透到各个领域。然而调查显示，未被使用的信息比例高达99．4％，存在于未被使用部分的高信息价值无法得到有效地开发、利用。在大数据时代背景下，如何从大数据中挖掘出有用的信息将是未来大数据发展的关键。

    石勇，中国科学院虚拟经济与数据科学研究中心常务副主任，长期从事多目标线性规划为基础的数据挖掘方法理论和应用研究，最擅长为海量数据信息编制“增值链”，从大数据中挖掘大价值。

    7月19日，中国经济时报记者在北京见到了他。

**一**

    与石勇的对话是从一位已故英国人开始的，这位英国人影响他至深。

    理查德·普赖斯（RichardPrice），英国著名经济学家、政论家和哲学家，十八世纪英国定量思想发展方面的主要人物。也是贝叶斯统计学派开山鼻祖——托马斯·贝叶斯的好朋友。石勇说，普莱斯做了一件很重要的事，这件事现在很多人可能都不知道。在统计学中有一个著名的贝叶斯决策理论，贝叶斯生前并没有发表相关文章，在他去世以后，普莱斯帮助整理并发表了，并告诉世人‘这不是我的成就，是我好朋友的成就’。”

    石勇认为普莱斯不仅学术造诣高，人品也很高尚，并由衷敬佩。

    最近一次去英国，石勇还专门去伦敦寻找普莱斯的墓地，“我找了很长时间，最后终于找到了，他的坟墓已经不是很好了，毕竟已经几百年了，他没有儿女，我想有一天有机会能把他的坟墓重建一下。”

    石勇说，普莱斯是最早用统计学方法研究英国人的缴税情况并以此预测英国国债的人。他当时做的事情就是现在我们做的大数据，普莱斯在1783年发表这篇文章的时候做了一个表格，叫多维数据表，每一行记录个人缴税的情况，每一列记录一个人什么时候以什么方式缴税，现在所有的数据分析都和这个表格有关，只不过现在的表格是由成千上万的行和列组成。

    “我希望我所做的事情在几百年以后，也能像普莱斯一样让后人记得。这是很崇高的境界，也是我的梦想。”这是石勇对自己在数据挖掘领域所追求的终极目标，“就是将来我不在的时候人们怎么看我。”

**二**

    石勇的学术之路，不仅起步早，而且过程颇为亮眼。

    从1978年在西南石油学院读数学专业开始，石勇就一直在和数字、数据打交道。大三那年，石勇发表了两篇有关模糊群和模糊积分的论文。与此同时，他还与老师合作，用模糊数学的办法分析河南油田的油层孔隙结构。这一项目后来获得石油工业部科技进步二等奖，还在核心期刊上发表了论文。2009年石勇因在这领域的杰出成就，获得康托学术奖，到目前为止他仍是唯一获得该奖项的中国大陆学者。

    1985年，在大连的中国工业科技管理培训中心完成MBA课程后，他远渡重洋，攻读美国堪萨斯大学管理科学博士。在此期间，他把模糊数学和多目标决策结合起来，做出了开拓性的成果。

    多年来他一直从事管理科学和信息技术的理论和应用方面的研究，取得了一系列有影响的成果。其代表性的成果包括：多目标多资源线性规划；最优线性系统设计和多目标线性规划数据挖掘与知识管理法等。

    1986年，由石勇为主提出的多目标多资源线性规划的观点，使他从助理教授一跃成为讲座教授。

    在美国期间，石勇获奖很多。1993年他获得内布拉斯加州立大学院长卓越研究奖；1997年至2000年获美国电子电气工程师协会（IEEE）卓越演讲者；1997年获内布拉斯加州立大学信息科学优秀教授奖。

    此外，石勇还先后获得来自美国政府、基金和公司150多万美金的科研经费。

    令人意外的是，在美国发展得顺风顺水，并拥有稳定事业和生活的石勇很快做出一个决定——回国。“在美国是一种比较平稳的生活方式，挑战性比较小，中国快速发展，机会很多，挑战性很大。”石勇更喜欢挑战。

**三**

    国家973项目子课题 “抗体分子结构与功能进化研究”，澳大利亚必和必拓公司国际合作项目 “地质数据挖掘项目”，网易公司“VIP邮箱数据挖掘项目”……石勇不但人回来了，还把先进的数据挖掘技术带回国内，带领研究团队进行各种实际应用研究。

    数据挖掘技术作为当代信息服务业的核心技术和新知识经济时代的支撑技术，正日益体现出现实的科学意义和广泛的商业应用价值。科学是没有国界的，但科学家有祖国。

    石勇在看到国外在这一领域飞速发展的同时也对中国数据挖掘的未来充满了紧迫感。

    所谓数据挖掘就是运用各种数学工具，从海量的数据库中寻找隐含的、潜在的、新颖的、有用的数据关系。简言之，就是在海量的数据信息里“淘金”。

    石勇涉足数据挖掘领域，始于他1998年夏天在美国第一数据公司（FirstData）的应用研究。在对信用卡数据处理上，该公司以往一向是从统计学的方法入手，而石勇则力主使用“最优化理论”切入。

    回国之后，石勇一个很重要的工作是开发全国个人信用评分系统。“中国人民银行个人征信评分系统”是石勇和他的科研团队历时3年开发成功的 “基于数据挖掘方法”的全国个人信用评分系统。这个系统的评分模型包括500多个衍生变量，通过基于多目标线性规划的数据挖掘技术，对数据库内的个体进行评分。这个被命名为“ChinaScore”的评分系统，在挖掘方法上国际领先且具有独创性知识产权。

    石勇告诉本报记者，目前该系统已经在部分商业银行试用，并将逐步服务全国13亿人口的经济信贷活动，成为中国金融信息化的基础工程。石勇非常看重这个项目的社会意义，因为“这是管理科学界的方法在中国第一次影响到每一个人。”

**四**

    回顾自己的学术路程，从模糊数学基础理论到多目标决策，从数据挖掘到知识管理，看似很多工作，但他认为归根结底都是一件事：用数学模型反映现实生活中的问题，并通过实际数据对模型进行检验、修正和完善。

    大数据作为一种重要的战略资产，已经不同程度地渗透到每个行业领域和部门，其深度应用不仅有助于企业经营活动，还有利于推动国民经济发展。石勇说，人们只看到了大数据的好处，没有看到大数据背后数据科学怎样支持大数据的发展，石勇说，中国的科研水平和国际上几乎是同步的，只是在政策上滞后一些，政府要在相关政策制定上加快步伐。

    采访中，石勇并未透露他的另一个身份——民建第九、十届中央委员会委员。作为民建中央委员，石勇曾参与起草民建中央在2009年全国两会上提出的《关于加强社会信用体系建设，增强抵御经济风险能力》的提案。提案认为，信用风险防范失控是酿成当前国际金融危机的根源之一，我国只有加快社会信用体系建设，才能安全有效地发挥虚拟经济对经济发展的促进作用。这项提案被列入当年全国政协的100件重点提案，并进一步选定为需要实地调研的五件提案之首。

    记者很好奇，是什么原因让他对数据挖掘产生如此浓厚兴趣？石勇的回答意味深长：“当你发现你所做的一些理论工作立刻被现实生活所验证、所应用，而且产生经济价值，这个感觉就不一样了。”他说，虽然现在社会上普遍认识到大数据的好处，但至于怎样推进大数据的技术进步和发展，没有一个结构性的或者普世性的规划。“这是我们学术研究者所关心的问题。这一领域也是更容易与现实结合起来和超越现实的研究领域。”

    石勇有一个嗜好，每到一个国家就会去拜访当地已故的科学家、音乐家、文人等的墓地。“他们是我从小就崇拜的人，站在他们的墓前，用手摸着他们的墓碑，就觉得与他们零距离了。”石勇说，人类的生命是很短暂的，一个人一生平均可以活3万多天，而真正可以工作的时间很短，我们需要做的事很多，所以应至少做一件可以让人们记住你的事。

    对于人生的价值，他似乎有着特别的感悟，“我上次去英国牛津图书馆，看到里面有我写的书，忽然觉得这才是最大的价值实现。它不在今生，而在来世。”

（来源：中国经济时报-中国经济新闻网 2013-10-21  记者/张娜）