

大数据时代：从 1.0 到 3.0

——访清华大学社会学系教授、清华-青岛数据科学研究院 RONG 兼职教授罗家德

◆ 本刊记者 关悦

记者：能不能请您简要介绍一下您所做的研究项目？

罗家德：分享一下针对社会、商业、营销大数据的社会计算（social computing）的三个阶段，我简称其为大数据 1.0、2.0 和 3.0 阶段（更准确地说，这是社会计算区分出来的，可简称其为社会计算 1.0、2.0 和 3.0 阶段）。其他大数据应用，如健康、规划、医学等等是不是如此分，我非专业，不敢说，所以这只对与社会科学及管理理论有关的社会计算而言。



清华大学社会学系教授、清华-青岛数据科学研究院 RONG 兼职教授罗家德

1.0 阶段是数据驱动（data driven），2.0 阶段是理论驱动（theory driven），3.0 阶段是 data-driving-theory 数据驱动理论（的改变）。

我们核心要做的东西就是理论驱动型的，往往需要将大数据与结构化数据如既有资料库和 survey 数据相结合，把定性研究融入大数据分析之中。这就需要企业的后台数据，而这个一般企业不太愿意给你。第二个，这是纯学术研究型项目，而现在一般企业要的是数据驱动的实际应用。所以我们这个研究项目，基本上就是不断在跟某些对未来知识有一点点好奇的、有数据的公司合作。说句实在话，这个过程非常辛苦。比如我们跟某些公司合作，但双方为了保护隐私，我们拿不到他们的大数据，他们也不能拿我们的 survey 数据，工作速度就会相对慢很多。因为我们要想尽办法猜想数据挖掘的方法送给对方，对方运行后如果出来的结果不对，我们再重新设计模型，再送过去，两个礼拜往返一次。我们称这种是社会科学理论指导，有抽样的，建构因果关联，能作推论的，不再是简单的数据驱动。

像《大数据时代》《删除》这些维科多的书，它等于是建构了大数据 1.0 时代的思维，但它只能做数据挖掘、数据驱动和相关分析，它做出来的预测模型没有理论基础、没有推论能力，预测边界也是未知的状态，是纯粹实用导向的。比如，要在超市摆尿布的地方也摆上啤酒。隔了半个月之后，我再采集一次，数据还有效，那就继续摆。而我们可能会马上问：那除了摆啤酒之外，还可以再摆什么？这就要推论。因为我们会发觉，理论上证明了买尿片的男性大概是处于 35-45 岁之间的中产阶级。简单来讲，就是推论型统计和描述型统计是不一样的。推论型统计一定是

理论驱动，不会是数据驱动。就是怎么用理论去整合数据，让个人画像越来越清楚。

大数据的未来是会改变社会科学研究典范的。未来的画像，将不止是个人画像，还有集体画像，动态的画像，没有大数据这是很难画出来的。这是现在的理论前沿，未来会因为大数据的发展而大幅推进，会改造整个学术界的理论发展。所以我称它为 3.0 时代，因为大数据而改变了社会科学理论建构的方式，所以是数据驱动理论的改变。

记者：您觉得与 1.0 和 2.0 时代相比，未来的大数据 3.0 时代最大的特点是什么？

罗家德：简单说的话，1.0 时代，就是数据挖掘加实务应用。2.0 时代，就社会科学而言就是 social computing（社会计算或 computational social science 计算社会科学），社科理论驱动，由社科理论找到扎根真相，在扎根真相的座标下作数据挖掘和理论驱动的预测模型。再到未来，一定是数据反过来又可以把理论前沿不断往前推，我觉得这也是未来数据院努力的方向。

未来对很多东西的预测会发生重大变化。比如说股票什么时候崩盘，世界什么时候会爆发重大运动，经济什么时候会转型，乃至预测一个重大系统的制度变革能不能成功。那就是 3.0 时代。将来可以用大数据去做更多更精准的预测。当然有多精准，我们也不知道，它永远是一个改良的过程。但是现在还远远没有成熟，连理论都还比较粗糙。但是你已经可以看到一些文章在努力往这方面走。所以简单来讲，动态系统的图像——不再是静态系统和个人图像，而是一个动态系统的图像——将会在未来的 3.0 时代成为一个最重要的发展。

我们团队出了一本书《Social Network Analysis—Interdisciplinary Approaches and Case Studies》，书里提出了一些方法的想法，但多数文章远远不到“2.0”的程度。所以我们准备要出第二部，要做各式各样的社会科学理论，如人脉、社经地位、社会资本等的大数据资料分

析，会更加细致地把我刚才所讲的东西全部用案例来展现。这就是我们在努力的方向。我们说 2.0 时代已经开始了，因为已经有非常成功的案例：竞选团队利用大数据进行人格分析，推断你对哪一个政治议题是什么样的兴趣，从而进行精准的政治营销。

你可以想象得到，随着个人图像越画越全，包括生活风格、社会经济背景、社会地位、人脉、社会资本、人格，等等，精准营销也将越来越精准。当然，这一定要保护隐私，是在客户自愿接受精准营销之下才能去做。

记者：咱们现在有一个中德大数据研究学生交换项目，能不能请您简要地介绍一下？

罗家德：这个项目主要是给那些对大数据有兴趣的优秀学生提供实际参与大数据研究的机会。我们一年会选派 5 名学生带着我们这边的研究议题去德国哥廷根大学计算机系进行半年到一年的交流。

基本上我们这边的学生都是学社会科学出身，而到那边是去计算机系。当然这些文科学生大都有些理科背景，或者具有很强的数理能力。我们想要建立一个文理兼备型的研究团队，希望他们能在混合学科训练的背景下，学习和提高大数据分析技术与学术能力。

记者：数据院成立四周年，您对数据院有什么期望？或者对未来合作有什么展望？

罗家德：首先要恭喜数据院成立四周年。能够建成今天这个样子，非常不容易。我觉得作为学者和业者之间的一个平台，作为很多方面的一个促成者，这是非常重要的功绩。对未来，我认为大数据的日子才刚开始，如果大数据未来想要不断推动社会理论和社会计算的前沿发展，还需要更多的学者来努力。同时，也希望能有更多的业者认识到支持纯学术性的工作对他们的未来是有非常好的影响的。有了这两方面的促进，我觉得我们的大数据不止是能够照应到今天的需求，也能够开辟对未来的展望。☞