



朱小黄：数据重构与数字化进程



文/意见领袖专栏作家 朱小黄

在电脑、互联网、通讯技术等科技进步的过程中，人类生产和服务的发展也突飞猛进。20 世纪初以来，我们大致可以观察到几个不同的阶段：从信息化与自动化，到数据化与智能化，再到目前的到大数据与数字化，都是信息技术发展的不同阶段，核心还是数据和数据运用。

在数字化转型的今天，如何认识和理解数字化，是向数字经济和数字化社会转型的关键。在专业认识层面上，各种观点和议论很多，笔者总觉得意犹未尽，希望再做一些深入的讨论。



一、数字化的含义

在数字化热潮中，我们听到的设想大都是 20 世纪六七十年代发达国家已经实现的信息化与自动化。无非把固化的生产流程和生产要素的配置通过后台电脑和软件进行信息处理，机器自动操作。随着互联网技术和 IT 技术的提升，这种信息处理的能力越来越强大，越来越具有智能的算法，但仍然长期处于自动化的阶段。

那么什么是数字化呢？随着大数据技术的发展，对生产和生活环境中的边际（实时）数据的处理能力，即算法和捕捉数据的能力也越来越强大，智能化程度（即机器模拟人脑的能力，机器运用知识和快速处理数据的能力）也大幅提高，那么社会就会逐步进入数字化阶段。到了这个阶段，人们对数据质量的要求（即科学性和运用的准确性、有效性的要求）也越来越高，传统的数据观和数据运用方式已经不能满足数字化时代机器对人脑模拟的要求。

数字化阶段的本质特征就是具备能够动态、实时、智能地捕捉相关数据并及时处理生产与生活场景中的微观变化，指挥机器进行同步调适，保持生产与生活场景良好状态的能力。包括相应的计算机技术、网络技术、通讯技术和一整套组织系统和运行规则。如果元宇宙是未来那数字化就是必由之路。

所以既不要把自动化误以为是数字化，也不要急于构建元宇宙，数字化阶段要做好这个阶段的事情，失去数字化的基础，会迷失未来的方向。

二、不确定性数据重构与数据治理

不确定性研究认为，世界的本质是不确定性的，而人类社会是一个偶然性社会。按照热力学熵增定律，宇宙客观世界是充满无序动力的，由此决定了不确定性的客观性和人类社会发展的偶然性特征。为了对抗无序和不确定性，人类发展了自己的文明，不确定性的分布状态也是多样化的，在变化比较缓慢，时间跨度比较长的领域（如风化和氧化），人类建立了科学知识和社会秩序等文明成果，以此减少熵增，建立相对稳定的环境和预期，以求生存发展。从这样的认识出发，我们发现到目前为止，所有的数据如同一团乱麻，还需要按照不确定性原理进行分类，才能更合理、有效地加以运用。

1

按照不确定性原理将数据分类重构大致可以做以下分类：

历史数据与边际数据。前者主要用于分析过往和预测未来，后者主要用于大数据技术条件下，实时现象的快速准确处置与反应以及智能化支撑。

自然数据和行为数据。前者是客观世界运行状况的纪录，后者是人类行为活动的纪录。前者中有的数据对人的行为有影响，有的没有影响，需要进行选择运用。后者中有的数据对未来有影响，有的没影响，同样需要选择运用。

可重复数据和不可重复数据，也称之为必然性数据和偶然性数据。前者来源于人类测量计算的科学成果，如物理、化学定律和地理测绘成果，

这些数据都是可验证可重复利用的，另外也来源于人类社会秩序规范下的人类行为数据，由于秩序对人的行为的约束，大量的行为数据也是可重复，可预测，有一定必然性的数据。可重复数据往往可以预测未来可能发生类似事件的概率。后者是一次性发生的偶然事物的纪录，是不可重复不可验证的数据。可以运用于分析过往，但很难用于预测未来。有些在数学上称为离散分布的数据，即离散值，就是偶然性数据，很多计算场景下需要剔除。

原生数据衍生数据。前者直接反映了事物的原貌特点，后者是在前者基础上主观分析的产物。例如各种指数，各种对客户的评价评分，是经过抽象与人的主观判断的数据。在不同的运用场景中对这两类不同的数据需要赋予不同的权重。

2

模型变量因子的样本选择对模型的影响

任何经济预测模型或者计算函数的确定，都是在大量历史数据和边际数据中提取样本分析确定变量、因子，确定模型需要的变量。在数据未能按不确定性原理进行分类之前，因为数据中混杂大量不可重复的偶然性数据，对这些数据直接进行分析极有可能发生偏差和错误，影响模型的科学性和计量的准确性，这样的现状需要做根本性的改变。所以不确定性原理下的数据重构对经济模型的改进和完善也具有重大意义。

数据治理首先是数据的科学分类

数据治理更多地被赋予了安全性方面的意义，这是不全面的，更底层的问题还是数据的真实性、完整性、有效性问题。

历史数据浩如烟海，在社会、经济、法律、道德等方面如何有效运用，是数字化社会的重要问题。从法律上看，有数据产权制度、个人隐私信息保护等问题；从道德上看，有商业模式的选择如何有助于科技向善、守护伦理边界的问题；在经济上，数据治理主要是发挥数据生产要素效率的问题。科学的分类和符合客观属性的数据运用是数据治理的核心问题。

三、边际数据与智能化

实时发生并被捕捉收集的数据是智能化的基础。大数据技术被广泛应用以来，人类海量数据的收集能力和计算能力越来越强大，在不确定性原理中，边际数据与历史数据一样，也有大量偶然性数据应该被忽略，以节省算力，这就有赖于算法上智能程度的提升。真正的智能化，即机器模拟

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_49331

