

# 数字化、广义统计与数族协同<sup>\*</sup>

赵彦云

**内容提要:** 互联网技术革命已经收敛到数据资源、数据生产要素、数据资产,正在通过新科学、新技术、新产业推动人类社会的巨大进步。本文针对互联网技术革命作用的基本架构做出系统研究,提出数字化、全面量化、广义统计、数族协同一系列关键领域的分析研究,追求解析互联网技术革命的重大发展趋势、重要途径和主要科学手段,以及走向智能化的设施支撑和社会生态系统平台的优化及演化作用,特别阐述统计学科在互联网技术革命中基础地位和巨大发展空间,为统计实践工作和战略定位提供理论依据。

**关键词:** 互联网数字化;全面量化;广义统计;数族协同

**DOI:** 10.19343/j.cnki.11-1302/c.2020.05.010

**中图分类号:** C829.29      **文献标识码:** A      **文章编号:** 1002-4565(2020)05-0117-12

## Digitalization, Generalized Statistics and Digital Synergism

Zhao Yanyun

**Abstract:** Internet technology revolution has converged to data resources, data production factors and data capitalization (data assets), and is promoting the great progress of human society through new science, new technology and new industries. This paper makes a systematic research on the basic framework of Internet technology revolution, presents the analysis of a series of core factors like digitization, comprehensive quantification, generalized statistics and digital synergism, and pursues to analyze their development trend, important approaches and main scientific means in Internet technology revolution, as well as the facility support to go intelligent, and the optimization and evolution of social ecosystem platform, especially the fundamental position and huge development potential of statistics in the Internet technology revolution, thus providing the theoretical basis for statistical practice and strategic positioning in China.

**Key words:** Internet Digitization; Comprehensive Quantification; Generalized Statistics; Digital Synergism

党的十九届四中全会《中共中央关于坚持和完善中国特色社会主义制度、推进国家治理体系和治理能力现代化若干重大问题的决定》明确了数据作为生产要素参与分配,加速互联网数字化技术的创新扩散,实现数字经济的快速发展,具有重要的战略意义。数据资源、数据生产要素、数据资产或资本化,是互联网技术革命过程的系统收敛。对于这个枢纽中心性的过程收敛,到底蕴藏着什么样的科学内涵?什么样的技术变革?什么样的平台经济?什么样的社会组织生态系统及运行机制的变革演化?以及怎样通过新科学、新技术、新产业的互联网技术革命推动人类社会的巨大进步?本文试图探索其中的关键理论和科学内涵。

<sup>\*</sup> 基金项目:中国人民大学科研基金重大项目“互联网统计学”(17XNLG09)。本文是在2019年10月26-27日中国统计学会第20次全国统计科学讨论会特邀报告“广义统计与数族协同”基础上成文的,感谢中国统计学会第20次全国统计科学讨论会的邀请。

### 一、互联网技术革命作用力解析

互联网技术推动知识谱系的科学解析,强化交叉学科的应用研究,推动新理科、新工科、新文科的创新发展,互联网技术革命的作用和影响如日中天(施巍松等,2017)。本文研究认为,互联网技术革命正在推动新科学、新技术、新产业发展,见图1。

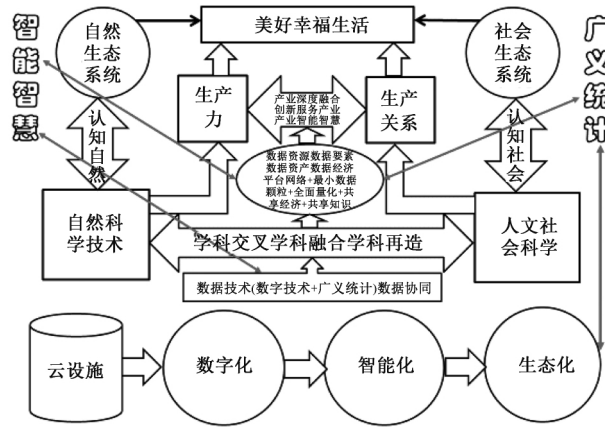


图1 互联网技术革命结构力解析

人类社会发展的根本动力是追求美好的幸福生活,互联网技术革命作用比任何时候都更加突出,实现这一目标就是要发展更加强大的生产力。支持生产力发展的是人类发明和累积的知识,包括自然科学技术和人文社会科学,前者科学技术直接支持生产力发展,后者人文社会科学直接支持生产关系发展。互联网技术革命创造了全社会的互联互通互动网络通信技术,实现了全社会前所未有的网络连接,形成了巨大的社会复杂生态系统和网络平台运行生态机制,并逐步走向智能化。这不仅创造了互联网数字化智能化等新科学知识和新技术并直接转变为新的生产力,而且生产力与生产关系无缝对接走向更加广泛深入的自然与社会组织生态系统,扁平化、个性化、公平均等、共享经济、共享知识等智能化生产、服务和运行机制,使人类社会走向前所未有的互联网时代。

互联网技术革命使自然科学技术与人文社会科学之间的学科交叉、学科融合、学科再造,在收敛于数字化、数据资源、数据生产要素、数据资本化、智能化方向上,正在催生新科学、新技术、新产业、新生态机制。其中,升级智能制造、智慧农业是新科学、新技术的直接转化运用,更大规模更加细分的新服务业是在云设施、智慧城市、共享经济、共享知识平台上通过自然科学技术与人文社会科学深度融合发展的结晶,创造了云平台智能服务业产业链,也是新产业体系中服务业比重大幅提升的根本性原因。

钱学森提出的人类社会复杂巨系统理论<sup>①</sup>,在互联网数字技术支撑下变成现实。互联网技术创造了数字技术条件,数字技术+广义统计创造了数据技术最广泛的应用,全面量化提出了数族协同的普遍性科学问题,因此,数族协同将是互联网技术革命下新科学的一个核心,统计学习、深度学习、数字孪生、边缘计算等新技术(赵晟和姜进磊,2016;毕马威和阿里研究院,2019)也将从此起航,新产业是在新科学和新技术推动下的智能化巨轮,将乘风破浪,为人类造福。

统计思维、统计分布、统计模拟、统计网络、统计学习(李舰和海恩,2019; Goodfellow等,2016),把最大统计总体与最小统计样本数据颗粒,以及空间、时间、维度细分最大系统和最小系统,构成新

<sup>①</sup> 钱学森1980年代初提出中医整体观复杂巨系统等一系列系统科学理论,是他对数学、应用力学、空气动力学、工程控制论等研究基础上向医学和社会科学等方面发展的研究,今天看来非常有科学意义。

的广义统计(赵彦云 2018),发展了互联网大数据资源,不仅是生产力,而且是智能化、智慧化的科学技术发展的基础。全面量化+数据资源,成为深入智能制造,开创基于数据生产要素、数据资产的新兴服务业,全面升级互联网时代新产业体系的强大驱动力,创造平台经济、共享经济、数字经济,推动新经济全面发展。

## 二、互联网数字化

### (一) 数字化的技术革命作用

没有互联网数字化、没有数据技术,就不可能有人工智能和智能化(GB/T 36625.2-2018,智慧城市数据融合 2018)。概括当今时代发展,已经非常明确的是一切离不开互联网、物联网,大千世界一切互联、互通、互动的数字化、数值化成为第四次技术革命变化的实质特征,数据技术基础设施、数据技术生态圈,云数据、云计算、互联网大数据成为重要的科学技术领域,因此,适应和满足互联网技术要求的标准化、数字化、互联网统计工作是互联网技术革命的基础,是人工智能和智能化不可或缺的重要基础环节。

信息化、数字化成为互联网时代发展过程的重要概念。一种误解认为数字化就是信息化,但是,大家都能看到数字化的内容已经发生很大变化,相关的数字技术不断发展,专门的物联网、传感器、穿戴设备、专用的芯片,以及遥感和数字地球技术等,已经应用到国民经济各个行业和社会的各个领域。如同蒸汽机、电力和半导体等为通用目的技术在工业技术革命中的作用一样,互联网数字化对人工智能和智能化具有决定性的通用目的技术的作用,在第四次工业革命时期的重要意义越来越被社会认可,没有数字化,人工智能和智能化不可能实现。

从信息化到互联网,再到物联网、智联网的第四次工业革命,数字化应用的数据技术(DT)成为信息技术(IT)发展后的一个更高级的阶段,其中统计学发挥重要作用,包括对全部社会信息和自然信息的数字化或数值化技术;不产生数据垃圾的大数据统计设计;微观到宏观的系统一体化统计分类;统计优化计算和监测、风险评估、决策等大系统统计理论方法等内容。因此,统计思维、统计思想、统计理论、统计方法、统计计算、统计分析、统计学习、统计预测、统计决策等,在数字化上发挥着重要的作用(宗成庆 2013; UNECE 2018)。

### (二) 数字化推动数字经济

数字化推动数字经济是互联网时代的重要特征。1995年美国学术界首先提出数字经济概念,1998年和1999年美国商务部电子商务秘书处(Secretariat on Electronic Commerce)相继发布了《新兴数字经济》(Emerging Digital Economy)之一和二,这两份报告都是与电子商务有关,跟目前的数字经济有很大差距。2017年,《中国信息经济发展报告》改为《中国数字经济发展报告》体现了中国向数字经济的重点转变。2019年,联合国贸易和发展会议发布了《2019年数字经济报告》(Digital Economy Report 2019)。德国经济和能源部2009-2011年开始发布《德国数字化监测报告》(Monitoring-Report Deutschland Digital) 2012年开始,转为以《数字经济监测报告》(Monitoring-Report Digital Wirtschaft)发布,涉及到德国16个州的情况,成为德国数字经济发展指南。数字经济是利用互联网数字化技术推动产业升级和新产业、新业态、新商业模式的快速发展,智能制造、平台经济、共享经济、智慧金融、智慧交通、智慧城市等等,已经呈现出来数字化技术的产业经济转化,数字经济成为互联网时代生产力作用经济的集中表现。

以利用数字化知识、信息、数据为重要推动力是数字化推动数字经济的集中表现,包括信息产业的数字产业化,以及产业数字化,增强数字经济与传统产业深度融合,大力提升产业的生产率和经济效率。数字经济是一种新的经济社会发展形态,日益成为经济发展的新动能。这是因为数据

成为新的重要生产要素,数字技术在创新推动发展中形成新的动力,信息技术的产业融合推动数字经济快速发展,平台设施推动产业组织的创新发展,深入运用大数据成为产业发展的新方向,系统全面治理成为数字经济的新治理方式。总之,数字化、数据技术应用深入发展成为数字经济发展的巨大推动力。

### (三) 数字化内涵是统计

数字化概念在具体业务领域应用体现的工作就是全面量化,为什么要量化?一是为了增强概念和事物关系的精准理解,二是为了时间和空间上的精确比较,三是为了认知的深入运用数理逻辑进行统计学习的需要,四是归纳到统计分布和统计预测的客观规律。显然,量化就是实现统计思维和统计逻辑及数学计算、统计计算等的需要。因此,全面量化就是广义统计,数字化就是互联网统计。

相对来讲,数字化、全面量化、广义统计都是新概念,实际上,这三个概念的基础条件都是互联网技术。数字化是利用互联网技术,包括物联网、传感器、穿戴设备等技术设备,以及政务网、企业网、社会网等产生的结构化数据和非结构化数据的总称,实际上是在互联网技术网络平台上,建立一种依赖全面细分有效数据从事生产、生活和社会活动的全新生产方式和社会组织机制。全面量化、广义统计是保障数字化应用的科学技术理论和方法。

社会活动数字化及互联网技术的全部联系,发展的是全面量化和广义统计。全面量化为传统思维下的微观、中观、宏观世界形成最大可能的一致性统一,活动社交关联量化最简约的一致性,是在计算机网络技术、云存储、云计算和统计、数学相结合的科学标准编码下实现的,可统计、可归纳、可计算、可增维、可降维、可组合、可优化,全面量化推动广义统计全面发展。从结构化数据到非结构化数据,事后统计到事先统计、潜在统计,追逐人们思维形成初期到社会生产、生活行为的全面统计解析理论方法的科学发展。

### (四) 数字化与社会组织生态系统

传统社会组织存在很多问题,如信息失灵、组织僵化、层级过多、资源整合能力弱、个体潜力未能充分发挥。因为组织是科层制、金字塔式,正被迫走向网络化、扁平化、平台化和协同化。数字化对组织和个人的关系转变也产生了深度影响,包括创客、自我雇佣、合伙制、社群等。在共生时代,组织与个人的关系将从单一到多维、固定到松散、被动到主动、雇佣到合作、“管”到“帮”、命令控制型向赋能赋权型等方向变化,两者的关系也将更为紧密和多样,组织将服务于人、成就于人,一起实现共赢、共进化。

数字化是运用互联网平台的个人赋能、企业赋能、政府赋能,实现互联网数字化赋能就要社会组织生态系统的创新发展。对组织和个人进行全方位一站式数字化赋能<sup>①</sup>,包括智慧协同办公、创建协作组织、评估数据资产、商机智能匹配、行为绩效激励、辅助经营决策等统计应用。建立以人为本、数据赋能一个核心;做好智慧组织新载体、数据智能母平台的两个定位;解决数据共享、业务协同、组织进化三类问题;追求数据化、协作化、定制化、可视化、生态化五大理念;发挥有助于工作、学习、内部管理和协作、实现内外资源做最优化匹配、引导个人和组织价值实现与财富增长的六大功能。

互联网数字化的社会变革作用是通过数据融合解决组织新生发展问题。数据价值的特殊性表现在,数据越用越多、越用越好。数据是黏合剂,可慢慢突破组织的风险、僵化问题;数据是融合剂,

<sup>①</sup> 杨冰之,《人类组织协作的数据实践》,中国社会科学院信息化研究中心、北京国脉互联信息顾问有限公司、清华大学国家治理研究院联合主办“2019智慧中国年会”,2019年11月28日-29日。

助力跨部门协作;数据是催化剂,使我们具有更多创新能力;数据具有穿透性,可打破制度壁垒。数字化强调数据资产清单,推动社会组织转型;数字化强调高效协作;数字化强调机会均等与公平;数字化增强数据驱动创新;数字化可让每个成员获取其应该获取的数据,让数据更好地支撑业务工作,增强网络化组织;数字化可实现弹性化快速组合组织,基于项目任务组合后解散,增强利益分配和组织架构动态化调整。所有的这些数字化社会变革作用都是在统计科学理论方法下实现的,互联网数字化也大大推动了广义统计的深入应用发展。

### 三、广义统计及其作用

#### (一) 广义统计概念

广义统计是对人类活动的全面量化、全面统计的基本概念(赵彦云,2016;赵彦云,2018)。相对于统计而言,广义统计是放宽了统计条件。在广义统计概念下社会经济统计不仅是统计报表、普查、抽样调查等统计,而且还包括政府监管的行政记录、文本、图像、音频等数据的统计化,以及利用卫星、遥感、地理信息系统的数字地球空间标准数据、物联网、传感器、穿戴设备、无线射频标签(RFID)、云设施等统计。广义统计理论不仅包括概率论与数理统计的随机性统计对象的理论方法研究,而且还包括确定性统计对象的理论方法研究,以及二者相结合的统计研究和面向互联网智能化的数字孪生的统计复杂系统等理论方法研究。广义统计肩负着利用互联网技术消除所有信息孤岛的伟大使命,发展人类活动统计量化的科学基础、设施平台基础,为科学、生产、生活、文化、社会组织活动提供完备的统计技术方法。

#### (二) 最大统计总体设计

在人类认知自然规律和利用自然规律的科学技术发展史上,最大统计总体设计和最小统计样本颗粒标准编码体系可能是广义统计发展最有意义的两个侧面,在互联网时代表现得最为突出。最大统计总体设计不等于任一相对统计总体设计没有意义,而是如果我们掌握了最大统计总体设计,那么可能为所有的相对统计总体设计的统计研究建立了最简洁的统计标准体系,也为问题的认知方向提供最大概率的可靠性。以天文学发展史为例,一直以积累了数个世纪的太阳、月球和行星运动及日月食的观测数据开展科学研究,统计分析非常突出,也提出了模式识别的各种方法,但是在哥白尼日心说之前,虽然做了许多统计数据分析和得到各种学说,结论结果显然都是错误的,日心说的统计总体设计根本地扶正了问题的研究和统计数据的运用,才有了天文学的科学发展。在互联网时代,数字地球实际上是继承了经纬度科学观,但是在卫星、遥感、航拍、地球物理等科学技术作用下,人类才进入现代数字地球的时代,这是今天人类社会相对最大的统计总体,包括人类所有的经济社会活动所产生的统计数据体系,追求全面量化自然生态和社会生态大系统发展特征,可以容纳所有科学知识包括自然科学技术和人文社会科学的探索和研究。

从经济社会统计的发展史来看,如何把全世界作为最大的统计总体设计,如何看待生产最大统计总体设计,如何把经济和社会统一成最大统计总体设计,都是一直追求的统计科学问题。生产最大统计总体设计曾经出现国民经济平衡表体系(MPS)和国民经济核算体系(SNA)两个设计,前者包括五大物质生产部门,后者采取综合全面生产的设计,即所有社会劳动为最大统计总体设计,家庭劳动被排除在外。最大统计总体设计,要考虑最小统计样本颗粒的假设条件,就经济社会统计最大统计总体设计而言,最小统计样本单位的设定有两种标准,一个是以家庭为最小统计样本颗粒,一个是以个人为最小统计样本颗粒,社会劳动是以社会分工为标准,社会分工又以最小统计样本颗粒为起点的统计原则。互联网时代,增强了许多统计技术条件,可以探索基于个人的最大统计总体设计的研究。

### (三) 最小统计样本颗粒标准体系

最小统计样本颗粒标准体系,自联合国1947年成立以来,联合国统计局一直组织推进,到1968年联合国国民经济核算体系(SNA)提出了比较完整的经济社会统计系统的最小统计样本颗粒标准体系,包括机构单位、基层单位、国民经济行业、产品和服务。其中机构单位是最小法人统计单位的概念,基层单位是最小产业活动统计单位的概念,国民经济行业可以细分到最小统计单位,产品和服务可以细分到最小统计单位。考虑到当时的操作能力,对上述四个最小统计单位建立分级的统计标准,分别是全世界统一的统计标准、国际区域如北美地区、欧盟区的统计标准、国家级标准。由上述4类最小统计单位分别建立了3个统计分类标准体系,国民经济机构部门分类标准体系、国民经济行业部门分类标准体系、产品和服务分类标准体系,其中基层单位与特征产品和服务一起约束国民经济行业最小统计单位,因此基层单位没有单独建立国民经济分类标准体系。这三个最小统计单位的分类标准体系是按照分类标准编码,不仅可以看到最小统计单位标准编码,还可以形成有序的分层统计分类,例如产品和服务10位码,每一位编码可以看成是一个层次,即上一个层次编码可以包含所有下层的分类编码。显然,这些最小统计单位并没有充分覆盖社会活动的最小统计单位。尽管社会统计中也是用各种统计分类,但是,社会统计分类不是从最小统计单位出发的。在互联网时代,社交网络统计成为最为重要的特征,因此,以个人身份证号码为最小统计单位唯一标准编码成为互联网统计体系的一个基础。

互联网技术为什么要求最大统计总体设计和最小统计样本单位,主要是两个目标,一是消除统计数据孤岛,建立全面系统的互联互通互动联系;二是建设完整的数据资源体系,让互联网统计数据发挥生产要素的重要作用,全面推动新产业和新产业体系的发展。统计总体设计是最小统计颗粒标准编码的坐标系。实际上,数字化与智能化都要求矩阵化思维、生态化系统思维,创新有了统计标准化编码上的约束,实现互联网云平台的有效协同,使数字模拟推动局部创新与整体创新的一体化,更深层次上的重要科学意义是推动多学科在数字化应用上形成数族协同的创新推动。

### (四) 统计表与系统网表

统计表是统计量化认知的重要方式方法,包括从原始统计数据信息到指标数据的统计及数据加工整理,以及针对一个具体目标的统计指标体系上的数据与信息加工整理。统计表和统计报表制度,在长期的政府统计工作中发挥着重要的作用。但是,随着互联网技术的深入应用,统计表也表现出不适应社交网络系统量化及分析研究的问题,统计表是形成数据信息相对孤岛的源头,表现出相对封闭的数据关系限制,而且是主观性的人为所致,当然与传统技术方法的有限能力相对应,是历史发展过程的必然。互联网就是要建立数据之间无限的联系,为数据分析应用创造无限可能,也是保证认知自然生态系统和社会生态系统的充要条件。

从统计表向业务客观系统网表与系统网图转变,成为广义统计应用的一个重要方面。基于互联网平台上最小颗粒的业务逻辑、多维数据联通的一张系统网表和一张系统网图,可以适应社交网络等大数据分析、统计学习、深度学习、数字孪生、可视化分析应用。魔方网表是目前应用的一种,采用纯B/S架构、在线填报、即时共享。它采用主流关系数据库,可以灵活地数据迁移,可用于各种管理系统的广泛应用,实现多单位或集团公司的综合信息化平台,可以细化到个人时间管理应用,能够充分适应企业的管理战略,保障人财物企业经营的高效率、高成功率。

### (五) 广义统计与人工智能

自然与人可持续发展关系是中国哲学史中的老子学说。互联网大数据、云技术、智能化的强劲发展说明所有的人类社会认知都可以统一到“量”“定量”“量化”,因为只有这一步才能实现物联网智能化和机器人的智能化。

人们认识自然、归纳与演绎社会活动及其属性,是从“质”的概念定义开始,并在知识体系的学科划分基础上发展其理论,其中“数”与“量”是结合“质”并形成独立的数学和统计学,人们在“质”“数”与“量”基础上发展物理、化学等实验科学,从而进化成学科知识体系的不断分工和发展。哲学是最早的科学,然后划分为社会科学和自然科学,一直进化到科学门类如哲学、经济学、管理学、人文社会科学、理学、工程、农学、医学等。“质”与“量”的关系,随着科学知识的进步,也在发生变化。“数”到“量”和“质”是一个人类认知自然与社会的知识发现与累积过程,“质”到“量”和“数”是另一个人类认知自然与社会的知识发现与累积过程。随着互联网大数据的文本、音频、视频、图像等信息的非结构化数据定义,以及在“量化”中转化统计数据,进入更加广泛深入的数学和统计学的理论方法研究之中,加速人类认知自然与社会的知识发现与累积过程的速度,“数”与“量”的广度、宽度、密度、深度在计算机、互联网、云技术实现的互联、互通、互动基础上,形成前所未有的生产力和进步的巨大能量(可以成为新动能)。

在“数”与“量”的认知能力上,数学还是最基础的科学力量,因为实践上可操作的“数”与“量”集中在统计学,社会经济统计学的确定性问题,概率论与数理统计的不确定性即随机性问题,都是在数学的“数”所涵盖下进化发展的。

人工智能要应用于智能城市、大数据、智能制造、创新设计,以及知识中心的发展之中,从自动采集信息、数据处理、数据中心、云计算等问题,到人工智能,将会有一次质的飞跃。人工智能的必要条件是计算机网络技术,包括通信技术和移动通信,但不是充要条件。人工智能的充要条件是人类知识体系和自学习累积的进化优化过程,最后通过全面量化过程及统计分析转到计算机联网技术实现的。知识经济并通过量化统计实现人工智能化是这场信息技术革命背后的最大推动力。计算机可以处理语言信息,处理视觉信息,处理图形信息,处理听觉信息,但是计算机很难把若干种媒体(全部社会活动的定量)打通,广义统计可以支持互联网统计的互联统计、互通统计、互动统计的全面信息量化的打通。因此,具有全面量化的广义统计是当今互联网大数据、云技术、智能化发展的最大新动能。

## 四、数族协同及发展

### (一) 解析数族协同

数据科学与大数据技术已经列入人才培养的新专业,非常鲜明地强调多学科的交叉,包括数学、统计学、计算机科学、计算机工程技术、数据科学和应用领域。图2给了一个简要的解析。显然,这些科学和技术都与“数”相关,而且可以讲“数”是各个学科出发点,最中心的交集是数族协同的根基,当互联网技术革命到来,通过互联互通互动等技术条件改变,将最大可能推动这个中心交集的数族协同根基的发展,由此,通过向各个学科的数据族协同科学扩散,创造新的科学知识协同体系。

但是,在互联网技术革命之前,因为缺乏互联网技术下的数族协同条件,所以各个学科所深入发展“数”的方向是不同的,主要反映了人们长期以来在认知客观发展规律上的自然科学分工。数学在“数”的抽象空间做了深入的发展,它支持物理学、天文学、计算机科学、统计学等一切学科的发展,在数理逻辑、数学理论方法、数学计算和服务其他科学和工程技术等,做出了巨大的贡献。统计学以“数”的应用为核心,对非随机的确定性统计和随机性统计问题,发展了系统的统计方法论,作为量化认知自然和量化认知社会及提高定量精确性与科学性上,发挥了重要的手段工具作用。计算机科学和计算机技术,特别是计算机网络技术,针对数学和统计,以及一切科学认知研究需要,运用物理学原理方法和技术,发展了针对“数”的存储、传输、计算的科学技术,全面支撑人类社会走向互联网技术时代。数据科学虽然比较年轻,但是它提出了统计发展的新问题、新理论、新方法,

而且紧密与计算机科学技术相结合,全面推动互联网大数据技术的发展。因此,我们提出来数族协同的研究问题,目的就是为了强化对数族协同的科学探索,发现“数”之科学之源,协同数族各学科之间的分工协同与融合发展,消除人类知识发展中的“知识孤岛”,顺应互联网技术革命要求,做好“数”的知识生产力的全面发展。

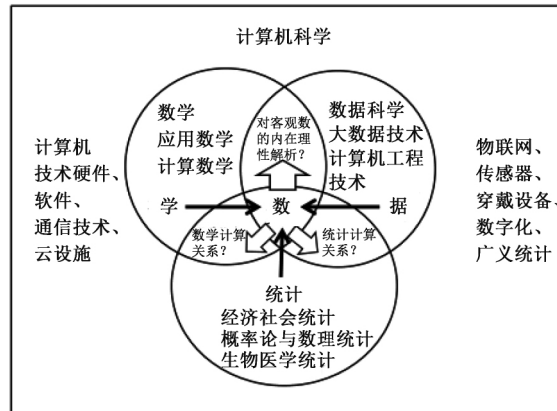


图2 数族协同学科关系

超低代码平台<sup>①</sup>的发展,全面推动 IT(信息技术)、OT(操作技术)与业务领域知识者的交叉,分工合作为 IT 编码、OT 设备(计算)、业务知识的一体化生态功能,目标是专业平台。这与计算机、数学、统计的交叉一体化生态功能一样,不断实现专业智能化的应用目标。这里背后的科学原理是什么?实际是三者最小颗粒上求得内容的统一,每个颗粒都内涵计算机、数学、统计学三种知识的无限能量,但是由分工不同的知识体系分别工作,实际应用成为一个无限组合中一个个事例的选取。统计在 IT、OT、业务知识领域都发挥着重要的作用,IT 编码如果走向超低代码平台,那么如何细分到最小颗粒的编码成为关键,这需要统计量化来处理企业(或政府机构部门、非营利组织机构)业务知识领域与 IT 之间的协同关系,因为未来二者结合需要统计数据分析和统计学习的支持。OT 的计算同样需要联系业务知识领域的统计的全面量化及量化上的生态逻辑关系,因为未来的应用主要是在回归分析和模拟预测或深度学习等统计逻辑思维之上。

数族协同是针对互联网技术革命提出来的新概念和新理论,之前主要是学科协同,例如计量经济学主要是经济学、数学、统计学之间的协同,后来还包括计算机软件技术的协同。学科协同本质上还是各学科知识孤岛基础上的桥连接的协同作用,例如计量经济学首先是经济学知识包括提出经济问题,以及运用经济学建立起来解决问题的经济理论,然后转化成数理经济学包括数学与经济概念的变量对接,并引入数理逻辑和证明,把经济问题的数学方程及其经济社会与政策因素之间的数理关系从数学应用角度确立下来,统计学把数学方程中的经济社会与政策变量经济社会统计化,即完成实际统计数据与数学方程变量的统一并完成变量统计样本数据序列或截面数据的整理,之后运用数理统计理论方法完成数学方程变量的参数估计和假设检验,最后运用经济学理论阐释统计参数估计与假设检验的变量结果的实际意义,并运用于经济分析、经济预测、政策模拟及效应分析等。数族协同是伴随互联网技术革命提出的知识创新新方向,即消除学科知识孤岛,开创人类知识的生态系统协同。消除学科知识孤岛,直接意义上是把所有知识能够根据人类社会需要建立起互联的联系并形成协同作用与协同效果,更深入的意义是缩短各学科知识的垂直层次,建立更多人类社会服务目标的多学科知识网络生态协同系统网络关系,大力推进学科知识应用的扁平化和个

① 林雪萍. 超低代码拓荒记:工业互联网新边疆. 知识自动化 2019年12月19日。



性化,支撑互联网智能化智慧化应用。数族协同包含“数”“族”“协同”三个含义的组合。在互联网技术革命下,大数据驱动、大数据作用具有更广泛深刻的科学意义,“数”内涵数据是资源、数据是资产、数据是生产要素的社会普遍性价值作用,在人类科学知识意义上,体现定量、量化、全面量化的认知世界的更大作用。“族”是“数”的自然与社会生态网络,包含人类认知自然和社交的网络、业务<sup>①</sup>网络、产业链网络、社会分工网络、技术网络、知识网络,而这些无数的网络都可以用全面量化的“数”网络统一起来。“族”是建立在“数”的基础上,整理所有知识背后业务的自然和社会生态逻辑。所谓“数”的基础根本的是统计的作用,包括最大总体、最小样本、最小数据颗粒,以及统计分布、概率转移、网络结构变化等内容。“协同”包含业务、技术、经济、社会、文化、环境等客体对象和人类社会需求目标之间上的各个学科知识的生态系统耦合关系。

### (二) 数学的作用

数学是一门“研究数量关系与空间形式”(即“数”与“形”)的学科<sup>②</sup>,分为纯粹数学与应用数学。纯粹数学研究其自身提出的问题,应用数学研究来自现实中的数学问题。纯粹数学追求解决外部问题如天文学、力学、物理学等需要提出来的问题,为了要研究从外部世界提出的数学问题如分子运动、网络、动力系统、信息传输等,往往从更抽象、更纯粹的角度来研究才有可能解决问题。数学也是广泛深入应用的理论方法,现代物理中已很难分清物理的数学化还是数学的物理化。化学在物质测定、化学平衡、电离层预测,生物在遗传、生态种群,天文、地学、通讯都深入运用数学。建立数学“模型”,使数学研究对象在“数”与“形”的基础之上又有扩充,各种“关系”,如“语言”“程序”“DNA 排序”“选举”“动物行为”等都能作为数学研究的对象。

数学是各门科学和技术的语言和工具,具有抽象性、逻辑性严密、广泛应用、精确性描述,以及研究对象的多样性与内部的统一性等特点。数学像一个庞大、多层次、不断生长、无限延伸的网络的一个“有机的”整体,各种概念、命题和定理,各层次的网络和结点之间是用严密的逻辑连接起来的,高层次的网络是由低层次网络和结点组成的,这种连接是客观事物内在逻辑的反映。建立新的结点,寻找新的连接,清理和整合众多的连接,并从客观世界吸取营养来丰富、延伸这个网络,一旦建立的数学模型和我们已有的结点或者低层次的网络相关,所有建立起来的连接都可能发挥作用,从而为现实世界提供解决问题的思路、理论和方法。

### (三) 计算机科学的作用

计算机科学和计算机工程技术的发展(宗成庆,2013),不仅解决了庞大的数学计算、统计计算、量化模拟等功效问题,而且在服务人类社会自然认知、社会认知和科学认知上发展了专门的互联网、宽带通信、物联网、传感器、穿戴设备、无线射频标签、云设施、机器学习、深度学习、边缘计算、知识图谱、人工智能等理论方法、芯片及工程技术(赵晟和姜进磊,2016)。大系统复杂网络数字孪生及智能化应用成为强劲的发展目标,例如神经形态计算。随着深度神经网络应用的成功发展,制造企业努力追求计算的替代模型——主要借鉴利用生物神经网络理论方法——不断提高硬件的效率和计算系统的稳定性相关模型和应用技术。

互联网技术革命经过了信息高速公路、数字地球、大数据、云计算,正在走向深入各个领域的智能化过程(Jun,2016)。回顾19世纪末内燃机替代蒸汽机、20世纪初汽车替代马车、1970年代电话替代电报,新旧技术的交替发展往往以拐点作为重要的分水岭出现,2019年阿里提出将云计算全

<sup>①</sup> 包括各个学科领域认知基础上的内容,例如物理学观察的内容,农学、植物学、林学、生态学、生物学、病理学等观察的内容,地球物理、大气物理、天文学、气象学、卫星遥感地理信息系统等观察的内容,等等。都有各自的内容,我们统称为“业务”。业务网络实质是所有专业业务内容的互联互通互动的网络系统。

<sup>②</sup> 张恭庆. 数学与国家实力. 北京大学数学科学学院网站。

面替代传统IT的拐点,所谓拐点是指全面上云时代必然已经到来。对传统IT的全面超越,云、大数据、AIoT(物联网)和移动化技术引领时代发展。根据阿里的研究开发,政府和企业全面上云将经过基础设施上云、大数据上云、云上中台和云上智能四个阶段。总之,计算机科学和网络技术已经为科学研究和产业应用发展做好了基础,为数学和统计提供复杂大系统量化和分析计算,以及数字孪生模拟的存储、计算、平台服务等强大的能力(艾瑞咨询 2019;毕马威和阿里研究院 2019)。

#### (四) 统计的作用

数学是数字化的抽象理论方法,统计是数字化的现实理论方法,计算机科学、通信技术、网络技术、软件技术,以及支持这些科学技术的物理学、化学都为统计奠定了扎实的科学依据和技术工具。传统统计被哲学社会科学、自然科学、工程技术、农学、医学等分解,成为彼此独立相互交叉有限的分散性学科。我们需要探索统计实践和统计学应用的统计思想发展过程。(1)统计实践从计数的汇总得到统计总体的数据,以达到用统计认知样本单位和总体范围及规模的方法。(2)从样本单位数统计发展到样本单位活动或属性特征的量化统计,实现样本之间的统计比较认知,再进一步到统计分布、统计分类、统计分组的统计分析,以及样本单位不同统计特征(统计变量)之间的相关关系、回归关系等统计分析方法、预测方法、目标优化方法。(3)统计分析的发展,一些来自其他学科的量化分析方法的统计学大家族归属方法的发展,例如主成分分析、因子分析等等。(4)互联网大数据,引来了社交网络关系的统计分析、数据挖掘、机器学习、统计学习、深度学习。文本量化和自然语义数据处理,使统计非常强地依赖计算机的存储和计算。互联网技术革命的冲击波有许多,但是,消除数据孤岛可能是社会网络化技术首先要解决的统计问题。

在数族协同之中,统计发挥着重要的数字化和全面量化及解析社会生态系统演化发展的基础作用,统计元数据标准,把统计数据按照计算机网络技术和数学计算、统计计算标准要求规范,包括统计数据背后的生态网络系统逻辑和全面量化计量单位。统计核算、统计回归包括分层分位回归、周期分析、协整分析、因果研究、结构方程、数据挖掘、统计学习、统计模拟等统计理论方法在经济、社会、工程技术、人文社会科学与自然科学研究等绝大多数领域都有广泛的应用(艾瑞咨询 2019)。

这里的统计主要讲的是随机性统计,还应有确定性统计的协同研究,包括关联可解释性和稳定性、高效的总体设计,以及关联实际对象的互联、互通、互动基础上的元数据标准编码。实际上,确定性统计可以把统计基础与计算机、数学很好地衔接,还是智能化的社会化因素衔接到最优的过程上的基础。数族协同需要在互联网技术基础上的统计思维逻辑上的细节研究,重新确定统计学知识体系和科学分工合作,统计学内部协同是做好外部协同的必要条件。

#### (五) 数族协同与人工智能

人工智能研究已经深入许多基础性探索研究<sup>①</sup>,从数学、统计和计算的角度,设立“人工智能的数理基础”重大研究方向,并在该方向开展三方面研究:可解释性的新型人工智能模型、新型的机器学习算法、深度学习的基础理论,以期打破基于计算机实验和神经科学的人工智能的惯用建模范式,建立以数学与统计理论为第一原理的新一代人工智能方法论。

1. 从应用或计算数学的角度来理解深度学习,研究架起这个桥梁之后可以给网络构架的设计,以及一些理论相关的研究提供一些新的指导思想。化学的前身就是炼金术,在没有系统理论指导的情况下是炼金术阶段,但被赋予系统的理论指导之后就从一个看似纯实验的学科变成有一定理论体系的科学,这也是所有想做基础理论学者的目标——赋予深度学习一套科学理论,做到

<sup>①</sup> 北京智源人工智能研究院(Beijing Academy of Artificial Intelligence, BAAI). 打破惯用范式,以数学与统计构建新一代人工智能方法论. 2019年5月14日。

Mathematical/Theoretical Deep Learning。之所以要这么做,是因为如今看起来发展甚好的深度学习仍然面临着各种各样的挑战:标注缺乏(或者说非常高质量的标注相对缺乏);如何在多样的数据上学习;使深度学习模型更加透明。

2. 破解机器学习中的维数灾难,加强可辨识性研究。近年来,以深度学习为代表的“黑箱”机器学习算法在应用中获得了巨大成功,但缺乏可解释性和严格的理论基础。而可辨识性和维数灾难的概念是从统计学角度理解深度学习与机器学习算法的关键。

3. 探讨在脑神经网络上建立数学模型来剖析由物质世界的物理定律演变成思想世界的意识智能的过程。离散数学更多是在计算机技术关注,连续数学里面有很多很强大的工具可以使用,应该有一个更大的框架,把离散数学和连续数学放在一起考虑来解决既利用计算机优势又利用数学优势的脑神经网络上的人工智能问题。

4. 基于深度学习的分子动力学数值方法。发展基于深度学习的原子间相互作用建模方法——深度势能,通过模型一致性估计深度势能模型误差的方法,在本质上改善原子间相互作用的建模精度。

5. 从“数学工程”的角度来理解机器学习。机器学习发展经历了基于规则的学习、基于统计推理的学习、基于深度表示的学习等重要阶段。机器学习的四个基础原则:泛化性、稳定性、可计算性和可解释性等。机器学习的关键在于表示学习,表示需要适合预测和适合计算。而深度表示所面临的挑战包括:由于大数据的需要可能导致过参数化;由于多层的表示,导致问题高度非凸化。统计为解决问题提供了最佳建模途径;概率论、随机分析等工具可以用来研究AI的数学机理。

人工智能的实践性问题如何依靠各学科的高深发展的知识体系之间的协同需要一个深入研究的发展过程,针对互联网数字化的局部问题,以单学科创新发展还是目前阶段的主要特征。互联网技术革命,已经提出网络为人类服务的更大更宽更深的大网络应用问题,显然数字化和智能化不仅是局部性的应用,而且更具价值的是全网络数字化和智能化应用的挑战,因此,各学科现有知识体系针对互联网数字化与智能化求解出一套直接为互联网技术平台新创造发展的社会组织生态系统使用的基础理论方法,实质就是数族协同。

#### (六) 数族协同与社会生态系统量化

探索数族协同作为一个完整的基础理论方法,目前还面临许多挑战。因此,我们还要顺从实际发展,主要是两种方式,一是如同前面所述,针对人工智能应用方向,从数学、生物模型、概率统计等出发,积极推进机器学习、统计学习、深度学习应用上的交叉学科协同发展解决智能化提出的基础性问题,由此归结到数族协同;二是直接利用数字化,以建立互联网平台技术所需要的社会组织生态系统及优化自律智能化机制,探索广义统计基础上的数族协同理论方法及其应用。把自然科学、工程技术、工业制造等领域应用成熟的理论方法应用到经济社会领域,因为忽视社会生态系统的专门规律,经常遇到许多挑战性问题。因此,推进数族协同应用,在自然生态系统和社会生态系统两大方向上分别发展、互动发展可能是一种积极有效的发展态度。我们提出社会生态系统全面量化的广义统计思维,属于带电作业的研究方式,因为互联网技术革命的影响,社会组织生态系统正在适应新生产力发展需要做出重大变革发展的举措,例如政府部门间数据共享平台将支持背后的广义统计和数族协同理论方法的创新。广东省大数据管理局全面整合省、市、县、镇(乡)、社区(村)一体化的政务大数据新社会组织生态系统,其中背后也有广义统计和数族协同理论方法的创新需求。国家电网互联网大数据明确互联网技术平台的新发展和新应用,背后也是广义统计和数族协同理论方法的创新。从这个意义基础上,我们要从互联网技术平台+社会生态系统全面量化出发的广义统计和数族协同应用开发其技术和理论方法的研究。

## 五、基本结论

综上所述,我们得到以下结论:

1. 互联网技术归于数字化,在理解互联网技术革命的科学技术体系上,具有重要的科学意义,也是关联技术革命创新和理解数字经济的“枢纽”。

2. 广义统计是落实数字化连接自然与社会生态系统量化的理论方法,数字化体现为数字技术,数字技术+广义统计是数据技术广泛应用的充要条件,是走向数字经济的基础。

3. 全面量化是广义统计和数字化的数理哲学的理论方法基础,顺应互联网技术革命,我们要发展哲学社会科学理论方法,尤其需要研究量与质和互联网技术相结合的新理论方法及生态逻辑理论,以及探索科学技术与人文社会科学之间的学科交叉、学科融合、学科再造的理论,指导互联网技术革命中新科学、新技术、新产业、新生态机制的发展,全面推动数字经济的科学发展。

4. 数族协同是互联网数字化提出来的重要科学问题,直接意义是探索数学、统计学、计算机科学技术,以及生物学、物理学、化学、农学、医学,从学科交叉向数族协同的学科融合再造发展,更好指导云平台+社会组织生态机制+智能化的新服务业产业链的创新发展,推动互联网时代的新产业体系快速形成。

5. 从互联网技术到数字经济形成的社会巨大复杂生态系统优化机制,互联网技术革命收敛于数字化、数据资源、数据生产要素、数据资本化(数据资产)、云平台、智能化,从而形成互联网技术革命的中心枢纽,由此向前向后的社会链条、产业链条、技术谱系、知识图谱,以及向新科学、新技术、新人文社会科学和它们之间的融合发展,追溯网络互联互通互动的广义统计与数族协同的全面发展,为人类社会发展的巨大进步做出我们统计科学的重要贡献。

### 参考文献

- [1] 艾瑞咨询. 中国制造业企业智能化路径研究报告: 中国智能制造之路[R]. 2019.
- [2] 毕马威, 阿里研究院. 从工具革命到决策革命——通向智能制造的转型之路[R]. 2019.
- [3] GB/T 36625. 2—2018, 智慧城市 数据融合[S]. 北京: 中国标准出版社, 2018.
- [4] 李舰, 海恩. 统计之美: 人工智能时代的科学思维[M]. 北京: 电子工业出版社, 2019.
- [5] 施巍松, 孙辉, 曹杰, 等. 边缘计算: 万物互联时代新型计算模型[J]. 计算机研究与发展, 2017(5): 907-924.
- [6] 赵晟, 姜进磊. 典型大数据计算框架分析[J]. 中兴通讯技术, 2016, 22(2): 14-18.
- [7] 赵彦云. 互联网统计研究[J]. 统计研究, 2016(12): 3-10.
- [8] 赵彦云. 互联网统计与广义统计学[J]. 统计研究, 2018(6): 3-10.
- [9] 宗成庆. 统计自然语言处理[M]. 北京: 清华大学出版社, 2013.
- [10] UNECE. Generic Statistical Business Process Model (v5.0) [EB/OL]. <https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model>, 2018-7-3.
- [11] Goodfellow I, Bengio Y, Courville A. Deep Learning[M]. Massachusetts: MIT Press, 2016.
- [12] Jun S. Patent Big Data Analysis by R Data Language for Technology Management[J]. International Journal of Software Engineering and Its Applications, 2016, 10(1).

### 作者简介

赵彦云,中国人民大学杰出学者特聘教授,应用统计科学研究中心、统计学院博士生导师。研究方向为经济统计和分析、国民经济核算、国际竞争力与创新指数、互联网统计及应用。电子邮箱: cas-kriu@ruc.edu.cn。

(责任编辑: 刘 鸾)