

大数据揭示经济发展状况

高 见^{1,2}, 周 涛^{1,2,3}

(1. 电子科技大学互联网科学中心Complex Lab 成都 611731; 2. 电子科技大学大数据研究中心 成都 611731;
3. 成都数联铭品科技有限公司 成都 610041)

【摘要】随着大数据时代的到来,与经济活动有关的数据数量和质量都得到了极大的丰富和提高。通过分析这些来源于社会经济系统中的大规模数据,人们有机会在几乎不花费调查成本的情况下对经济发展状况进行精准和实时的测量。该文关注大数据对于经济发展状况的刻画,简述了不同类型的数据在揭示宏观经济结构和微观社会状况方面的具体应用,并进一步分析了大数据助力解决区域经济发展战略和宏观产业结构升级问题的可能途径。

关键词 大数据; 经济复杂性; 经济发展; 产业结构; 社会经济系统

中图分类号 F062.5; F222 **文献标志码** A **doi**:10.3969/j.issn.1001-0548.2016.04.015

Big Data Reveal the Status of Economic Development

GAO Jian^{1,2} and ZHOU Tao^{1,2,3}

(1. Complex Lab, Web Sciences Center, University of Electronic Science and Technology of China Chengdu 611731;
2. Big Data Research Center, University of Electronic Science and Technology of China Chengdu 611731;
3. BusinessBigData (Chengdu) Co., Ltd Chengdu 610041)

Abstract With the advent of the era of big data, both the quantity and quality of economic activity related data have been enormously enriched and improved. By analyzing these large-scale data from socio-economic systems, we have the opportunity to quantify the status of economic development instantaneously and accurately with nearly no cost. In this paper, focusing on how big data reveal the status of economic development, we briefly summary the applications of different types of big data on quantifying macro-economic structures and micro-social status. Further, we discuss and provide some promising ways to apply big data to improve regional economic development strategies and upgrade macro industrial structures.

Key words big data; economic complexity; economic development; industrial structure; socio-economic systems

经济状况不仅影响着国防预算和产业投资等国家宏观战略决策,也关系到个人的生活标准和消费策略^[1]。如何实时和准确地掌握当前经济发展状况,一直是困扰着经济学家和统计学家的难题^[2]。传统经济学研究中,最为直接和普遍的方法是进行经济普查,再基于普查数据计算得到相应的宏观经济指标,如国民生产总值(GDP)^[3]。然而,宏观经济指标计算所牵扯到的很多数据需要从各级政府收集和汇总,整个过程会耗费大量的时间、人力、物力和财力^[2]。随着普查技术的提高,虽然经济统计速度加快,但仍然难以满足经济决策实时性的需求。

为了快速和有效地刻画当前经济发展状况,一些非直接经济指标已经被广泛采用,例如克强指数^[4]、

消费者价格指数(CPI)^[5],生产者物价指数(PPI)^[6],采购经理人指数(PMI)^[7]和网络零售价格指数(淘宝CPI)^[8]等。这些指标虽然行之有效,但也不可避免地存在各自的缺陷:克强指数只包含3个主要经济指标,忽略了农业和服务业,缺少对经济发展状况的全方位把握;CPI和PPI指数的计算过程非常复杂,并且依赖于长时间的数据收集;PMI指数主要关注制造业,计算结果与抽样方法密切相关;淘宝CPI指数仅仅反映中国网络消费状况,缺乏完整稳定的商品目录。在洞悉经济发展状况方面,迫切需要全面、快速和精准的新策略。

随着信息技术的革新,一方面经济发展促使人们使用互联网和电子信息产品在全球范围内获取和

收稿日期: 2016-05-15

基金项目: 国家自然科学基金(61433014)

作者简介: 高见(1989-),男,博士生,主要从事复杂网络科学和社会经济系统等方面的研究。

分享信息^[9]，另一方面高科技产品也忠实地记录下人们在社会经济系统中的大量行为数据^[10]。这些海量非干预数据的开放和使用，可能会对社会经济研究产生深远影响^[11]。事实上，已经有一些开创性的工作利用国际贸易、手机记录、社交媒体、互联网检索、银行转账等数据揭示区域经济发展状况，甚至提前预测一些关键经济指标^[12]。与传统的经济普查相比，这些数据所涵盖社会经济系统的范围更广。全新的研究策略和方法，不仅极大节约了统计成本，而且可以支撑经济决策的及时性。

本文主要介绍社会经济系统中的大数据在揭示经济发展状况方面的具体应用，重点关注宏观经济结构的刻画和微观经济指标的关联分析。第1节简要介绍数据驱动的经济研究背景以及社会经济系统大数据。第2节主要介绍大数据在刻画宏观经济方面的相关工作，包括经济状况、产业结构、经济复杂性和新经济指数等。第3节主要介绍大数据用于分析和预测微观经济状况方面，包括财富状况、不平等性、物价指数和失业率等。第4节简述大数据在理解城市发展规律，改善经济发展平衡性和产业转型升级方面的可能帮助。最后，展望大数据在揭示经济发展状况方面的整个研究图景并勾勒未来的发展方向。

1 数据驱动经济研究

1.1 实证经济研究

经过十几年的发展，经济学研究已经逐渐朝着更加注重实证工作的方向迈进。文献[13]回顾了近五十年在顶级经济学杂志上发表的论文，结果发现当今基于实证数据的论文数量已经超过了70%。回想二十年前，感兴趣的第一手数据资源还相当紧缺。对于某些产业，还必须跑遍图书馆，从行业出版物中手动提取和统计数据。现如今情况已极大改观，研究人员可以从不同渠道获得丰富的社会经济数据，不仅有政府部门掌管的行政记录数据，还有私营部门积累的大量数据^[2]。得益于几十年的数据革命，经济学家广泛使用新数据开展研究工作，这为经济学的发展带来了新的机遇和挑战^[14]。

社会经济系统大数据在很多特征上都区别于传统的实验和普查数据^[15]。一方面，大数据具有及时性。相比于调查数据滞后数月甚至数年，社会经济系统的数据实时持续更新，有利于及时指导经济政策，如互联网检索数据可以用来实时预测通货膨胀和劳动力市场活动。另一方面，大数据可以拓宽数据记录类型。有些之前很难测量但经济学家又感兴

趣的数据已经可以被广泛采集，如个人通讯和社会网络数据。基于这些数据，可以分析社会连接和地理接近性对消费者消费行为和求职的影响，根据个体通讯多样性来预测社会经济状况。值得注意的是，社会经济系统的数据一般是非结构化的，数据记录形式多样，并且维度比较高。要从众多的信息中提取感兴趣的特征并加以分析，这也给实证经济研究带来了新挑战。

1.2 政府部门数据

在工商、税务、教育和征信等一系列国家和地方管理系统中，各级政府部门收集和详细记录了个人和企业的信息数据^[16]。这些行政记录数据所蕴含的价值远远超过传统的门户普查数据：门户普查数据难免存在数据缺失，而且样本数量有限；行政记录数据由政府部门维护，是有时间序列和截面两个维度的面板数据，因而数据质量更高。一旦政府部门数据得以完全开放，将会给从事劳动经济学、公共财政、健康和教育领域的科研工作者带来机遇。

政府部门掌握的数据不仅总量大，而且涵盖的时间长，在刻画人口变化、收入、支出、生产力和工资等方面有独到的优势^[17]。例如，应用税务记录数据计算收入和财富份额^[18]，使用政务数据揭示区域经济流动性的差距^[19]，识别同类企业工资和生产率的差异^[20]等。在政府部门数据开放化方面，文献[21]提出了扩大行政记录数据访问的基本原则，挪威、瑞典和丹麦等欧洲国家也已经开始汇集行政记录来方便科学研究。在不久的将来，政府部门数据在经济学研究中将扮演越来越重要的角色。

1.3 私营部门数据

随着信息产业的发展，私营部门也逐渐注重社会经济系统数据的收集。除了大众熟知的Google、Amazon和Twitter等互联网公司进行海量的信息收集和存储以外，任何一个经济领域内的公司都在收集和汇总客户和业务数据，如各大银行、信用卡公司、零售商、手机运营商等^[2]。私营部门所掌握的每家每户个体层面的数据，不仅包含家庭金融活动，还包含消费支出和个人行为轨迹等。事实上，这些数据不仅能用于商业活动，还能促进经济研究。

私营部门的数据信息量更为丰富，计算资源更强大，数据使用的灵活度更高，能够被用来刻画和预测经济指标，实现对经济状况的实时追踪：麻省理工学院实施的“十亿价格项目”^[22]，基于互联网零售记录所构建的每日价格指数与劳工局发布的消费者价格指数非常接近；文献[23]在分析10种主流报

纸的文本数据的基础上构建了能够刻画经济政策不确定性的新指标;文献[24]发现利用网页检索数据能提前预测用户在未来的集体行为,如预测视频游戏首月销量。随着企业存储数据量的增加,科研工作者与私营部门之间的合作将逐渐深化,在隐私和保密机制下获取更多数据来服务经济研究。

2 宏观社会经济结构

2.1 宏观经济指标

社会经济的繁荣来源于人类活动的聚集,手机通讯和社会网络等与人相关的社会系统理应与经济发展状况密切相关。在工业化发达国家,比较容易通过收集社会经济系统数据来分析人口和经济统计特征。然而,在相对落后的发展中国家,不仅数据记录少,数据收集本身也很困难。幸运的是,即便是在贫穷的国家,手机的使用也越来越普及。手机数据不仅记录着通讯的频率、时间和收发的信息,还实时传递着GPS坐标、旅行记录、数据使用、消费和支出等其他信息^[25]。

不再依赖于政府调查数据,手机数据的分析逐步揭开了社会网络结构与经济发展之间的联系,并由此可以建立刻画经济状况的新指标。文献[26]分析了科特迪瓦的手机用户数据,发现用户的移动模式和打电话模式与经济活动、贫富程度、发电和能源等普查数据等都有很强的关联性,可以通过一天中不同时间的通话频率推断出用户的家庭和工作区域。文献[27]分析了英国手机通讯网络数据,发现个体的社交网络多样性与国家的经济发展状况的关联性很强。这些研究表明,仅仅使用手机数据就能推断出社会经济系统的多样性,为评估贫困国家的宏观经济状况提供了可行办法。

在线社会网络已经深入到社会和经济生活的方方面面,这为实现零成本实时估计经济发展状况提供了可能性。文献[28]分析了带有位置标记的社会网络数据,发现使用社会网络的结构特征可以预测美国各州的经济发展状况。特别的,不同州人们之间的长程社会连接与国民生产总值、专利和初创公司数量等经济指标有很强的关联性。文献[29]分析了微博2亿用户连续4年的在线社交网络数据,发现在线社交活跃度与区域经济发展水平之间的关联性很强,并且社交活跃度还能反映区域的宏观产业结构。这种基于数据挖掘的分析框架,不仅能简单地预测经济发展状况,还为实时感知区域经济结构的健康状况提供了新途径。

预测金融和经济系统中关键指标并不容易,已有一些开创性工作使用来自网页的文本资源来分析和预测金融和经济变量。文献[30]提出了一种利用复合异质网络数据流来预测经济变量的新方法,采用贝叶斯分类器组合模型对非农就业指数进行了高精度预测。文献[12]在分析西班牙个人银行转账记录数据的基础上,提出了35种个体经济行为量化指标。结果发现,个体消费行为与社会经济指标之间存在显著关联性。通过所提出的量化指标,能够预测GDP、房屋价格、失业率、犯罪率、高等教育比例、生活成本和预期寿命等社会经济统计指标。以上工作为实现采用互联网大数据来预测宏观经济状况提供了新方向。

2.2 世界贫困地图

贫穷是21世纪所面临的慢性难题之一。据《世界发展指标》(WDI)发布的数据显示,全球仍有42%的人口生活在贫困中^[31]。在研究区域经济活动和贫困程度的空间分布上,海量廉价的手机数据可以大显身手。文献[32]结合卢旺达手机网络数据和问卷数据,发现可以通过手机历史使用记录来推测个体的社会经济状况。在此基础上,可以反过来重建出整个国家的财富状况,绘制出高分辨率的国家财富分布地图,所得到的结果与政府调查数据高度一致。在人口普查和家庭调查数据紧缺的情况下,可以通过该方法及时掌握全球的贫困状况。

除了手机数据,卫星拍摄的夜间灯光亮度数据也被用来评估区域的贫富状况。基于夜间灯光亮度图像数据和人口数据,文献[33]在国家 and 区域层面对人均收入的空间分布情况进行建模,所得到的结果可以作为区域财富状况的近似。文献[31]使用估计方法得到的贫穷指数与家层面统计得到的贫穷指数相一致,并绘制了全球贫穷程度地图。文献[34]则将夜间灯光亮度数据与经济生产率数据一起分析,针对欧盟和美国的发现夜间灯光亮度与区域生产总值在不同时空尺度上都相关。文献[35]还分析了21个国家的明亮地区、GDP和电力消耗总水平之间的关系,结果发现夜晚灯光亮度与财富状况之间非常相关。这些研究表明,通过夜间灯光亮度数据可以对经济活动状况进行估计,为区域和政治地理研究提供了新见解。

2.3 产业结构空间

如何定量刻画经济发展状况一直是困扰着经济学家的问题。传统方法中所使用的一系列综合指标,不但计算复杂,而且很难用一个平均值来刻画经济

发展中的多样性。例如,以GDP来衡量经济的总体发展状况,只能粗略地展现当前经济处在阶梯的什么位置,却无法进一步给出当前经济状况的全貌。

在分析国际贸易数据的基础上,文献[36]提出了全面分析国家经济发展状况的新框架,根据贸易数据构建和分析了国家出口产品之间的关系网络,即“产品空间”。如果两个产品经常被一起出口,也就意味着产品被一起生产,有相似的制度、基础设施、物质要素、技术等。分析结果发现,技术复杂和简单的产品分别占据“产品空间”的中心和边缘位置;富裕的国家占据“产品空间”的中心位置,而贫穷的国家则只能占据边缘位置。

值得注意的是,并非所有的国家在面对经济发展时都有相同的机会,这取决于国家所处在“产品空间”中的位置以及当前产业发展的策略^[36]。贫穷的国家处在“产品空间”的边缘,想要发展新产品很困难;富裕的国家却可以快速地攀升到某个行业的领导地位。产业发展过程中,需要优先发展与当前产业相近的产业,因为跨越“产品空间”的发展往往很困难。换句话说,一个国家当前的“产品空间”决定这个国家未来的经济发展状况。这一研究把带动经济整体发展的过程模型化,在实证数据的基础上,给出了一种国家经济发展模式的全新解读。

文献[37]以沪深A股上市企业数据为基础计算了不同产业的接近性,并以此建模刻画了中国区域产业结构,即“产业空间”。结果发现,“产业空间”具有“哑铃型”结构和“中心-边缘”结构:劳动密集型和技术密集型产业分别占据两个核心位置;复杂程度高和低的产业分别占据“产业空间”的中心和边缘位置。“产业空间”随时间缓慢演化:以劳动/资本密集型产业为支柱的相邻省份,产业结构逐渐趋同:距离越近,产业结构越相似;以技术密集型产业为支柱的相邻省份,产业结构逐渐分化:距离太远,产业结构反而不相似。这些初步结果对理解区域宏观产业布局和调整有启发意义。

2.4 经济复杂性

经济的增长不仅与个体活动总量有关,还与个体在相互作用中涌现的复杂性密切相关。在进一步分析国际贸易数据的基础上,文献[38]提出了解释经济增长和发展的新观点,发现所构建的“国家-产品”二部分网络的结构特征可以定量刻画国家经济的复杂性;经济复杂性与国家的收入水平非常相关,甚至可以用来预测国家未来的经济发展水平。类似地,文献[39]基于企业注册信息数据分析和定量刻画了

中国区域经济复杂性,结果发现经济复杂性与人均GDP有很强正相关。这些结果暗示,未来经济的发展状况,至少在短期内,是主要由国家产业结构复杂性所决定的。所以,想要实现持续的经济增长和繁荣,就应该把力气花在满足自身经济复杂性涌现的条件上。

尽管对于经济发展的预测本身就十分困难,经济复杂性这一全新分支已经为解决这一难题提供了思路。在研究国际和本土经济中产业的兴衰时,文献[40]发现“国家-产品”二部分网络的嵌套结构不仅随时间保持稳定,还能用来预测产业在某个区域的出现和消失。事实上,非货币的经济多样性与国家经济的发展潜力密切相关。文献[41]提出了一种基于“国家-产品”矩阵的非线性迭代算法,同时刻画国家发展潜力和产品复杂性,所得结果比文献[38]更好地解释不同国家的经济竞争力。进一步,文献[42]通过对比国家发展潜力和人均GDP来量化国家的潜在经济增长力。结果发现,不同国家的经济复杂性在“发展潜力-人均GDP”相图中的演化模式差异性很大:在“层流”区域的国家,经济发展已演化出清晰的轨迹;在“混沌”区域的国家,经济发展模式则比较混乱。鉴于此,文献[42]进一步提出了一种数据驱动的方法来预测国家经济发展趋势。

2.5 新经济指数

大数据可以捕捉到新金融和其他新业态的早期发展趋势,再结合更多传统普查和统计数据,就能勾画出一幅更加完整和动态的新经济发展图景。在进行浩繁的大数据挖掘基础上,财新智库和数联铭品(BBD)联合打造和推出了新经济指数(NEI)^[43],尝试量化中国新经济的发展现状以及新经济在整个经济中占比的变化,从而填补经济转型过程中新经济度量的空白。

新经济指数是一个详实的指标体系,细项指标代表了新经济活动的侧面,包括了1个全国指数,4个维度,15个三级指标及每月8 700万个事件的原始数据,共涉及9大行业类别,111个4位数代码行业,覆盖全国294个地级以上城市。构建NEI的基础数据是网络公开的大数据,包括企业网络公开招聘信息、新成立企业工商登记信息、风险投资数据、招投标数据、三板上市数据、各类专利及专利转化数据等。

新经济指数首次将大数据用于度量中国经济转型中新经济相对于传统经济或旧经济的活跃程度,为即时、动态统计、刻画和预测我国新经济的发展变化提供了崭新视角^[43]。新经济指数还可以展示经

经济发展的地域差异, 有利于政策的差异化和资源的优化配置。新经济发展是人口跨地区流动的重要方向标, 结合人口出行大数据, 可以刻画人口流动热力图, 更好引导消费和服务业资源的跨区配置。

3 微观社会经济状况

3.1 个体经济水平

大数据及其分析技术的发展, 让人们有机会快速、全面和细致地理解和描述个体和群体行为。与此同时, 大数据所记录的精细的人类活动数据, 如手机数据, 也能带来对于社会经济发展刻画的高保真度的近似指标。文献[44]设计了一套数据驱动的分析框架, 通过手机数据来抽取人们移动模式和社会行为模式, 以此估计社会经济系统的发展指标。结果显示, 人类移动的多样性与人均收入和剥削指数有显著的关联性。这暗示人类移动的多样性对于解释社会经济状况很关键, 甚至比社会接触和人口特征的多样性还重要^[45]。

通过分析手机记录数据, 文献[46]发现手机用户集群行为中的固有模式可以在时空粒度上很好的计算出最新的贫困指数。基于两个发展中国家数据的实证研究, 发现手机用户的记录和特征与普查数据中得到的用户贫困指数非常相关。类似地, 文献[47]发现利用手机数据能够判断出群体的社会经济水平。如果以通讯行为总量为特征构建预测模型, 则可以对社会经济水平进行预测, 对于城市群体的预测准确率超过80%。

3.2 经济不平等性

产业结构不仅可以预测经济增长, 还能用来预测区域收入的不平等性。仍然基于国际贸易数据, 文献[48]还分析了国家经济多样性和收入不平等性之间的关系。发现国家的经济复杂性与收入不平等性呈现很强的负相关, 即出口高复杂度产品的国家收入不平等性较低, 反过来, 出口低复杂度产品的国家收入不平等性较高。进一步, 通过引入“产品基尼系数”的概念, 估计出国家出口产品所带来的收入不平等程度, 并展示了产业结构的变化如何影响收入的不平等性。

在判断一个城市的经济不平等性时, 人们甚至不需要翻阅相关的报告和经济数据, 只要乘坐出租车到处转转就知道了, 因为大多数城市的不平等性很显眼。文献[49]通过大数据将城市感受的不平等性与城市景观联系起来。基于从互联网收集的成千上万张带有地理标记的图片, 量化了城市街景所带给

人们的感受。结果发现, 纽约和波士顿的街景所引发的感受范围比林茨和萨尔茨堡更宽广, 这说明纽约和波士顿的城市景观更加不平等。进一步, 发现所测量出的城市安全和阶级感受与凶杀案有显著的关系。这些结果表明, 街景图片所包含的信息能够揭示城市经济发展的不平等性。

3.3 经济与物价指数

预测经济和物价指数是非常重要的, 华尔街将零售活动和物业销售的政府报告作为整体经济的指导性指标。然而, 这些传统的数据源由于时间的滞后性很难满足经济政策及时性的需求, 因为政府报告都是几周或几月发布一次。谷歌的经济学家们已经开始着手解决这一问题, 核心思想是用户对特定关键词的搜索量与相应的经济指标具有关联性。

文献[50]提出的谷歌趋势(Google Trends)已经能够追踪32种类型搜索量的上升或下降, 包括航空旅行、房地产和零售业等, 并且能以此预测汽车销量、失业人数、旅行规划和消费信息等经济指标的近期数值。文献[51]分析了以色列的谷歌搜索查询数据, 发现在人力资源等六个类别里都有很好的预测能力, 并能反映出2008年以色列的经济衰退。另外, 谷歌趋势还被用来解释哥伦比亚特定产业的经济活跃性变化, 定量刻画金融市场的交易行为^[52]。随着互联网在未来的深入, 谷歌趋势数据的代表性将更高, 与经济发展状况先行指标的关联性也会更强。

联合国Global Pulse实验室^[53]分析了Twitter上提及食品和油价的推文与官方公布指标之间的关系, 发现印尼提及大米和油价的推文数量变化与测量得到的食品价格通胀指标密切相关。文献[54]则分析了新闻、油价和国际金融市场发展对于俄罗斯债券和股票市场日收益的影响。文献[55]分析了金融新闻发布20分钟后的股票报价, 发现新闻的发布内容和当前股价能很好预测未来股价。这些研究结果表明, 社交媒体数据与官方统计数据之间存在某种关联关系, 从而验证了利用社交媒体来分析经济发展状况的可行性。

3.4 失业率

失业率是体现经济发展状况的重要指标, 也是国家重大经济决策的重要依据之一, 因此, 及时掌握失业率是极其有价值的。人们已经能利用手机数据在前所未有的层面理解人类活动、移动和交互模式。那么, 手机数据能否从多尺度上观测经济冲击及其后果? 事实上, 如果你失业了, 你使用手机的模式也会发生改变。

基于手机的呼叫信息数据,文献[56]提出了检测大规模裁员、识别出受影响人员和预测总体失业率变化的新方法。在小区层面,通过分析信号塔通讯总量来估计企业大规模裁员的时间。在个体层面,发现失业员工在后续几个月里的呼叫总量下降51%。由此,可以比官方提早2~8周预测当前失业率,提早4个月预测未来失业率。基于公开的社交媒体数据,文献[57]分析了西班牙不同经济区域内千万条带地理标记的信息,并以此定量刻画了个体的行为特征。结果发现,有更多移动流量多样性、更早期的昼夜节律和更准确的语法规则的区域会表现出更低的失业率。进一步,提出了从社交媒体数据中精确推断区域内失业率的数学模型。

互联网数据也可以用来提高对于经济指标的及时预报,尤其是失业率。文献[58]利用谷歌趋势数据来改善对于西班牙失业率数据的预测。文献[59]利用谷歌检索数据预测了挪威的失业率,得到的结果优于目前主流指标。文献[60]基于Twitter的数据构建了失业、求职和职位招聘的指标,比官方数据预测的更好。文献[61]分析德国互联网活动数据时发现,搜索关键词和失业率之间存在很强关联性。爱尔兰社交平台上与就业有关的迷茫情绪,也被发现可以比官方提早三个月预测失业率的增加^[62]。文献[63]通过互联网搜索数据预测了美国员工的具体失业数,所得结果与官方数据有很强的正相关。文献[64]基于企业社会化平台数据分析了雇员网络特征,发现员工所处网络位置与职业发展关联性很大。事实上,员工的网络特征可以用来预测员工的离职和升职^[65-66]。文献[24]则进一步指出了利用互联网数据预测失业率一些局限性。

4 经济发展路线

4.1 企业发展规律

通过大数据来实时刻画和揭示经济发展状况,其核心目的是促进经济向更好的方向发展。在努力实现这一目标的同时,首先要认识和理解经济发展的客观规律,如城市经济的多样性、企业兴衰和分布规律。随着大数据资源的丰富,针对城市经济活动具体结构和运行规律的研究已经逐步展开。

通过分析美国366个大都市统计区域内不同企业类别数据,文献[67]提出了一种测量经济活动内在多样性的分析框架,并发现了一种所有城市都共有的普适结构,体现在自相似的内部经济结构和综合指标,如GDP、专利和犯罪率。随着城市人口规模

的扩大,独立的商业类别会发生系统性的改变,新商业类型的引进率会逐渐放缓,这一结果体现出不同尺度的经济分化过程。

企业是当今人类社会最基本的经济单元,理解企业的生命周期和死亡率对于经济的发展也有重要意义。文献[68]分析了北美超过25 000家上市公司的全方位数据集,发现上市公司在很长观测期内都存在近似恒定的危险率——公司的死亡率独立于公司年龄;任何商业类型公司的生命半衰期大约都为10年。不仅是企业,城市也同样存在生死兴衰问题。文献[69]分析了意大利六大城市的手机数据、普查数据和开放地图数据,从实证角度验证了城市生存的4个理论条件。

在城市的发展中,人口和设施之间也存在普适的规律。当新设施出现时,应当由周边人口的分布决定选址,例如杂货店、学校和消防站等。考虑经济机制主导微观经济活动的框架,文献[70]在理想状态下研究了人口和设施密度之间的关系。结果发现,商业设施密度与人口密度呈现指数为1的幂律关系,而对于公共设施来说幂律指数为2/3。

4.2 城市经济调整

丰富的大数据资源既可以用来实时掌握经济发展状况,又能用来设计更好的城市发展规划,实现商业设施的合理布局。不仅如此,利用大数据及其分析技术,还有机会设计行之有效的策略,改善区域经济发展的不平衡性。

在城市设施的邻里范围内,一般都分布了餐馆、咖啡厅和图书馆等便利设施。文献[71]通过分析包含百万设施精准位置的数据集,研究主导这些设施共同出现的规律,帮助人们理解城市内自然出现的在邻里尺度内设施的聚集现象。在分析数据的基础上,构建了不同类型设施共同出现可能性的网络,即“*Amenity Space*”。进一步,构建推荐系统,根据当前专业化模式来为邻里推荐缺少设施,以实现更好的城市设施布局。

城市中社会经济系统的发展不平衡嵌入到空间中,导致邻里效应的出现,这种危害的后果很难通过单独的机制规划来消除。文献[72]在分析匿名信用卡消费数据的基础上,提出了一种自底向上的方法,通过微调人们购物路径来实现对资金流的重新分配,以此改善空间的经济发展不平衡。令人兴奋的是,在保持总路程不变的条件下,只需要改变大约5%的购物行程,就能达到整个城市的经济发展平衡性,并且还能改善城市的其他可持续发展指标。

4.3 产业结构升级

区域的经济增长路径是如何涌现的一直是主流经济地理学家关注的问题。在依靠大数据揭示经济发展状况和理解经济运行规律的同时,更为紧要的是探寻一条能够在遵循产业发展规律的前提下快速实现国家和地区产业转型和升级的道路。

从发展经济学的角度来看,区域更有可能发展与与已有产业技术相接近的产业分支。在研究产业接近性和区域经济发展路径之间的关系上,文献[73]分析了瑞典70个地区超过三十年的工厂级别经济演化数据。结果显示,瑞典长时间的经济版图演化有很强的路径依赖:与之前产业在技术上接近的产业更容易进入这个区域;与区域内产业在技术上关联性小的产业将逐渐被淘汰;区域内的产业具有高度的技术聚集现象。这些结果表明,区域层面的产业兴衰与产业技术的关联性非常相关,在产业优化升级过程中应当优先发展与当前技术相接近的产业,以此逐步实现产业的转型和升级。

事实上,经济发展路径依赖于产业技术接近性的发现与林毅夫教授提出的新结构经济学理论不谋而合。新结构经济学^[74]指出,一个经济体的禀赋及其结构在每个发展阶段都是给定的,并随发展阶段不同而不同。因而,经济体的最优产业结构也会随发展阶段的不同而不同。换句话说,区域在产业发展过程中要选择与基础设施相适应的产业结构。另外,经济发展阶段是一条从低收入农业经济到高收入工业化经济的连续频谱。也就是说,处于特定发展阶段的经济体,应当以与之相适应的产业结构为升级目标,逐步向更高发展程度的经济体的产业结构转变。最后,在每个发展阶段,市场都是配置资源的最有效的根本机制。除市场以外,政府也要发挥积极作用,协调和促进产业的多样化和升级。

5 结束语

随着大数据时代的来临,数据驱动的研究范式正在逐步影响和拓宽社会经济学的研究范畴。本文简述了实证经济学崛起的背景,回顾了大数据在揭示宏观经济结构和微观社会状况方面的具体应用,并对大数据解决区域经济发展平衡性、产业转型和升级问题进行了展望。

尽管很多社会经济系统的研究问题是对历史的回顾,数据的详实和准确比获取的及时性更为重要,但是基于政府部门和私营部门大数据的分析仍然能为指导经济决策提供很大帮助。虽然目前大数据还

没有颠覆传统的经济学理论,但毫无疑问,经济学研究的图景正在悄然发生改变,这也让人们备受数据驱动的新经济理论的诞生多了一些期待。

参 考 文 献

- [1] LEWIS W A. Theory of economic growth[M]. London: Routledge, 2013.
- [2] EINAV L, LEVIN J. Economics in the age of big data[J]. Science, 2014, 346(6210): 1243089.
- [3] BALASUBRAMANIAN N, SIVADASAN J. What happens when firms patent? New evidence from US economic census data[J]. The Review of Economics and Statistics, 2011, 93(1): 126-146.
- [4] Economist. Keqiang ker-ching: How China's next prime minister keeps tabs on its economy[EB/OL]. (2010-12-09). <http://www.economist.com/node/17681868>.
- [5] BOSKIN M J, DULBERGER E R, GORDON R J, et al. Consumer prices, the consumer price index, and the cost of living[J]. The Journal of Economic Perspectives, 1998, 12(1): 3-26.
- [6] BERNDT E R, GRILICHES Z, ROSETT J G. Auditing the producer price index: Micro evidence from prescription pharmaceutical preparations[J]. Journal of Business & Economic Statistics, 1993, 11(3): 251-264.
- [7] KOENIG E F. Using the purchasing managers' index to assess the economy's strength and the likely direction of monetary policy[J]. Federal Reserve Bank of Dallas Economic and Financial Policy Review, 2002, 1(6): 1-14.
- [8] LIAO J, WEI K, CHEN G. Relationship between pricing and customer's perception C2C commerce--Basing on study of the channel of C2C in mainland China[C]//2010 International Conference on E-Product E-Service and E-Entertainment (ICEEE). Henan, China: IEEE, 2010: 1-4.
- [9] SCHWEITZER F, FAGIOLO G, SORNETTE D, et al. Economic networks: The new challenges[J]. Science, 2009, 325(5939): 422.
- [10] MAYER-SCHÖNBERGER V, CUKIER K. Big data: a revolution that will transform how we live, work, and think[M]. Boston: Houghton Mifflin Harcourt, 2013.
- [11] MOKYR J. Intellectuals and the rise of the modern economy[J]. Science, 2015, 349(6244): 141-142.
- [12] SOBOLEVSKY S, MASSARO E, BOJIC I, et al. Predicting regional economic indices using big data of individual bank card transactions[C]//Proceedings of the 6th ASE International Conference on Data Science. Stanford, USA: ASE, 2015: 1-2.
- [13] HAMERMESH D S. Six decades of top economics publishing: Who and how?[R]. Cambridge, USA: National Bureau of Economic Research, 2012.
- [14] EINAV L, LEVIN J D. The data revolution and economic analysis[R]. Cambridge, USA: National Bureau of Economic Research, 2013.
- [15] VARIAN H R. Big data: New tricks for econometrics[J]. The Journal of Economic Perspectives, 2014, 28(2): 3-27.
- [16] JUTTE D P, ROOS L L, BROWNELL M D. Administrative record linkage as a tool for public health research[J]. Annual Review of Public Health, 2011, 32:

- 91-108.
- [17] JENKINS S P, LYNN P, JÄCKLE A, et al. The feasibility of linking household survey and administrative record data: New evidence for Britain[J]. *International Journal of Social Research Methodology*, 2008, 11(1): 29-43.
- [18] PIKETTY T, SAEZ E. Inequality in the long run[J]. *Science*, 2014, 344(6186): 838-843.
- [19] CHETTY R, HENDREN N, KLINE P, et al. Where is the land of opportunity? The geography of intergenerational mobility in the United States[R]. Cambridge, USA: National Bureau of Economic Research, 2014.
- [20] SYVERSON C. What determines productivity?[R]. Cambridge, USA: National Bureau of Economic Research, 2010.
- [21] CARD D, CHETTY R, FELDSTEIN M S, et al. Expanding access to administrative data for research in the United States[R]. Arlington, USA: National Science Foundation white paper, 2010.
- [22] CAVALLO A. Scraped data and sticky prices[R]. Cambridge, USA: National Bureau of Economic Research, 2015.
- [23] BAKER S R, BLOOM N, DAVIS S J. Measuring economic policy uncertainty[R]. Cambridge, USA: National Bureau of Economic Research, 2015.
- [24] GOEL S, HOFMAN J M, LAHAIE S, et al. Predicting consumer behavior with Web search[J]. *Proceedings of the National Academy of Sciences, USA*, 2010, 107(41): 17486-17490.
- [25] LANE N D, MILUZZO E, LU H, et al. A survey of mobile phone sensing[J]. *Communications Magazine*, 2010, 48(9): 140-150.
- [26] ŠĆEPANOVIĆ S, MISHKOVSKI I, HUI P, et al. Mobile phone call data as a regional socio-economic proxy indicator[J]. *PLoS ONE*, 2015, 10(4): e0124160.
- [27] EAGLE N, MACY M, CLAXTON R. Network diversity and economic development[J]. *Science*, 2010, 328(5981): 1029-1031.
- [28] HOLZBAUER B O, SZYMANSKI B K, NGUYEN T, et al. Social ties as predictors of economic development[M]// WIERZBICKI A, BRANDES U, SCHWEITZER F, et al. *Advances in Network Science*. Switzerland: Springer International Publishing, 2016: 178-185.
- [29] LIU J H, WANG J, SHAO J, et al. Online social activity reflects economic status[J]. *Physica A*, 2016, 457: 581-589.
- [30] LEVENBERG A, SIMPSON E, ROBERTS S, et al. Economic prediction using heterogeneous data streams from the World Wide Web[C]//*Proceedings of ECML/PKDD 2013 Workshop on Scalable Methods in Decision Making*. Prague, Czech Republic: [s. n.] 2013.
- [31] ELVIDGE C D, SUTTON P C, GHOSH T, et al. A global poverty map derived from satellite data[J]. *Computers & Geosciences*, 2009, 35(8): 1652-1660.
- [32] BLUMENSTOCK J, CADAMURO G, ON R. Predicting poverty and wealth from mobile phone metadata[J]. *Science*, 2015, 350(6264): 1073-1076.
- [33] EBENER S, MURRAY C, TANDON A, et al. From wealth to health: Modelling the distribution of income per capita at the sub-national level using night-time light imagery[J]. *International Journal of Health Geographics*, 2005, 4(1): 1.
- [34] DOLL C N H, MULLER J P, MORLEY J G. Mapping regional economic activity from night-time light satellite imagery[J]. *Ecological Economics*, 2006, 57(1): 75-92.
- [35] ELVIDGE C D, BAUGH K E, KIHN E A, et al. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption[J]. *International Journal of Remote Sensing*, 1997, 18(6): 1373-1379.
- [36] HIDALGO C A, KLINGER B, BARABÁSI A L, et al. The product space conditions the development of nations[J]. *Science*, 2007, 317(5837): 482-487.
- [37] GAO Jian. Quantifying local industry structure of China[EB/OL]. (2015-11-18). <http://gaocn.net/project.html#industry>.
- [38] HIDALGO C A, HAUSMANN R. The building blocks of economic complexity[J]. *Proceedings of the National Academy of Sciences, USA*, 2009, 106(26): 10570-10575.
- [39] GAO Jian. Modeling local economy complexity [EB/OL]. (2015-11-18). <http://gaocn.net/project.html#complexity>.
- [40] BUSTOS S, GOMEZ C, HAUSMANN R, et al. The dynamics of nestedness predicts the evolution of industrial ecosystems[J]. *PLoS ONE*, 2012, 7(11): e49393.
- [41] TACHELLA A, CRISTELLI M, CALDARELLI G, et al. A new metrics for countries' fitness and products' complexity[J]. *Scientific Reports*, 2012(2): 00723.
- [42] CRISTELLI M, TACHELLA A, PIETRONERO L. The heterogeneous dynamics of economic complexity[J]. *PLoS ONE*, 2015, 10(2): e0117174.
- [43] 陈沁, 沈明高, 沈艳. 财智BBD中国新经济指数技术报告[EB/OL]. (2016-03-04). <http://www.nsd.edu.cn/teachers/professorNews/2016/0304/25596.html>.
CHEN Qin, SHEN Ming-gao, SHEN Yan. BBD think tank: New economy index of China[EB/OL]. (2016-03-04) <http://www.nsd.edu.cn/teachers/professorNews/2016/0304/25596.html>.
- [44] PAPPALARDO L, VANHOOF M, GABRIELLI L, et al. Estimating economic development with mobile phone data[EB/OL]. (2016-05-30). http://www.cisstat.com/BigData/CIS-BigData_08_Eng%20%20IT%20Luca%20Pappalardo%20Et%20Al%20Estimating%20Economic%20Development.pdf.
- [45] YAN X Y, ZHAO C, FAN Y, et al. Universal predictability of mobility patterns in cities[J]. *Journal of The Royal Society Interface*, 2014, 11(100): 20140834.
- [46] SMITH-CLARKE C, MASHHADI A, CAPRA L. Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks[C]//*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, USA: ACM, 2014: 511-520.
- [47] SOTO V, FRIAS-MARTINEZ V, VIRSEDA J, et al. Prediction of socioeconomic levels using cell phone records[M]//KONSTAN J A, CONEJO R, MARZO J L, et al. *User Modeling, Adaption and Personalization*. Berlin: Springer Berlin Heidelberg, 2011: 377-388.
- [48] HARTMANN D, GUEVARA M, JARA-FIGUEROA C, et al. Linking economic complexity, institutions and income inequality[EB/OL]. (2015-08-13). <http://arxiv.org/abs/1505.07907>.
- [49] SALESSES P, SCHECHTNER K, HIDALGO C A. The

- collaborative image of the city: Mapping the inequality of urban perception[J]. *PLoS ONE*, 2013, 8(7): e68400.
- [50] CHOI H, VARIAN H. Predicting the present with Google Trends[J]. *Economic Record*, 2012, 88(s1): 2-9.
- [51] SUCHOY T. Query indices and a 2008 downturn: Israeli data[R]. Jerusalem, Israel: Bank of Israel, 2009.
- [52] PREIS T, MOAT H S, STANLEY H E. Quantifying trading behavior in financial markets using Google Trends[J]. *Scientific Reports*, 2013, 3: 01684.
- [53] Global Pulse. Mining Indonesian Tweets to understand food price crises[EB/OL]. (2014-08-07). <http://www.slideshare.net/unglobalpulse/global-pulse-mining-indonesian-tweetsfoodpricecrises-copy>.
- [54] HAYO B, KUTAN A M. The impact of news, oil prices, and global market developments on Russian financial markets[J]. *Economics of Transition*, 2005, 13(2): 373-393.
- [55] SCHUMAKER R P, CHEN H. Textual analysis of stock market prediction using breaking financial news: the AZFin text system[J]. *ACM Transactions on Information Systems (TOIS)*, 2009, 27(2): 12.
- [56] TOOLE J L, LIN Y R, MUEHLEGGGER E, et al. Tracking employment shocks using mobile phone data[J]. *Journal of The Royal Society Interface*, 2015, 12(107): 20150185.
- [57] LLORENTE A, GARCIA-HERRANZ M, CEBRIAN M, et al. Social media fingerprints of unemployment[J]. *PLoS ONE*, 2015, 10(5): e0128692.
- [58] VICENTE M R, LÓPEZ-MENÉNDEZ A J, PÉREZ R. Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing?[J]. *Technological Forecasting and Social Change*, 2015, 92: 132-139.
- [59] ANVIK C, GJELSTAD K. "Just Google it": Forecasting Norwegian unemployment figures with web queries[R]. Oslo, Norway: Center for Research in Economics and Management, 2010.
- [60] ANTENUCCI D, CAFARELLA M, LEVENSTEIN M, et al. Using social media to measure labor market flows[R]. Cambridge, USA: National Bureau of Economic Research, 2014.
- [61] ASKITAS N, ZIMMERMANN K F. Google econometrics and unemployment forecasting[J]. *Applied Economics Quarterly*, 2009, 55(2): 107-120.
- [62] Global Pulse. Big data for development: Challenges & opportunities[EB/OL]. (2013-10-20). <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobaIPulseJune2012.pdf>.
- [63] ETTREDGE M, GERDES J, KARUGA G. Using web-based search data to predict macroeconomic statistics[J]. *Communications of the ACM*, 2005, 48(11): 87-92.
- [64] YUAN J, ZHANG Q M, GAO J, et al. Promotion and resignation in employee networks[J]. *Physica A*, 2016, 444: 442-447.
- [65] 高见, 张琳艳, 张千明, 等. 大数据人力资源: 基于雇员网络的绩效分析与升离职预测[M]//刘怡君. 社会物理学: 社会治理, 北京: 科学出版社, 2014: 38-56.
- GAO Jian, ZHANG Lin-yan, ZHANG Qian-ming, et al. Big data human resources: Performance analysis and promotion resignation in employee networks[M]//LIU Yi-jun. *Social Physics: Social Governance*, Beijing: Science Press, 2014: 38-56.
- [66] 张琳艳, 高见, 洪翔, 等. 大数据导航人力资源管理[J]. *大数据*, 2015(1): 2015012.
- ZHANG Linyan, GAO Jian, HONG Xiang, et al. Human resource management based on big data[J]. *Big Data Research*, 2015(1): 2015012.
- [67] YOUN H, BETTENCOURT L M A, LOBO J, et al. Scaling and universality in urban economic diversification[J]. *Journal of The Royal Society Interface*, 2016, 13(114): 20150937.
- [68] DAEPF M I G, HAMILTON M J, WEST G B, et al. The mortality of companies[J]. *Journal of The Royal Society Interface*, 2015, 12(106): 20150120.
- [69] DE NADAI M, STAIANO J, LARCHER R, et al. The death and life of great Italian cities: a mobile phone data perspective[C]//Proceedings of the 25th International Conference on World Wide Web. Montreal, Canada: IW3C2, 2016: 413-423.
- [70] UM J, SON S W, LEE S I, et al. Scaling laws between population and facility densities[J]. *Proceedings of the National Academy of Sciences, USA*, 2009, 106(34): 14236-14240.
- [71] HIDALGO C A, CASTAÑER E E. Do we need another coffee house? The amenity space and the evolution of neighborhoods[EB/OL]. (2015-09-09). <http://arxiv.org/abs/1509.02868>.
- [72] LOUAIL T, LENORMAND M, ARIAS J M, et al. Crowdsourcing the Robin Hood effect in cities[EB/OL]. (2016-04-28). <http://arxiv.org/abs/1604.08394>.
- [73] NEFFKE F, HENNING M, BOSCHMA R. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions[J]. *Economic Geography*, 2011, 87(3): 237-265.
- [74] LIN J Y. New structural economics: a framework for rethinking development and policy[M]. Washington D C: The World Bank, 2012.

编辑 蒋晓

高见(1989—), 博士生。主要从事复杂网络科学、社会经济系统、数据建模与分析等交叉学科方面的研究。获2011年“成电杰出学生”称号, 2012年美国大学生交叉学科建模竞赛特等奖。近年来在 *Scientific Reports*, *Europhysics Letters* 等国内外期刊发表论文十余篇。

