

大数据围猎新冠病毒

计算社会经济学：一种定量评估可持续发展目标实现进展的数据驱动研究框架

高见, 周涛, 刘权辉
新经济发展研究院INED

数据采集和处理能力的提升，人工智能和统计机制研究的进展，已经在迅速和深刻地改变着社会和经济研究的方法论。近年来，由大数据和人工智能驱动的社会科学研究范式转变，为定量评估联合国可持续发展目标（Sustainable Development Goals, SDGs）的进展，提供了一系列有潜力和新的数据驱动方法。这种转变涉及到的领域非常广泛：消除贫困到健康和福祉，从性别平等到优质教育，从经济增长到创新和基础设施。国家各级政府部门都能从应用“计算社会经济学”¹（Computational Socioeconomics）框架下的新方法中获益，以更高的效率和更低的成本来评估在时间和空间维度上实现可持续发展目标的进展。

新方法论的转变

社会和经济研究变得越来越依赖于分析真实数据。然而，传统获取真实数据的方法面临很多局限性。例如，更大规模和更精准数据的获取过程，通常会消耗大量的社会资

源且缺乏及时性。幸运的是，近几十年来数字化浪潮席卷全球，社会经济领域的研究人员面临前所未有的机遇，发展出一套基于真实数据的量化研究方法论。特别地，交叉学科研究方向“计算社会经济学”（Computational Socioeconomics）的出现恰逢其时。

社会经济发展和人类活动过程中的数据，被越来越多的传感设备、在线平台和数据采集终端（例如：遥感卫星、手机、社交媒体和在线交易平台等）所记录下来^{2,3}。另一方面，这些数据体量更大、更新几乎实时且分辨率更高，可以减少很多源于小数据规模的稀疏性和偏差，也能减少社会经济发展过程中的不可见部分。因此，基于这些大规模的、新颖的数据，我们原则上可以在感知社会经济状况、评估社会经济发展进度、预测未来的社会和经济发展趋势等方面取得长足的进步⁴。

新数据的体量和多样性的不断增加，给研究方法论带来了两个方面的改变。一方面，简单的统计工具不适用于分析非结构化数据，

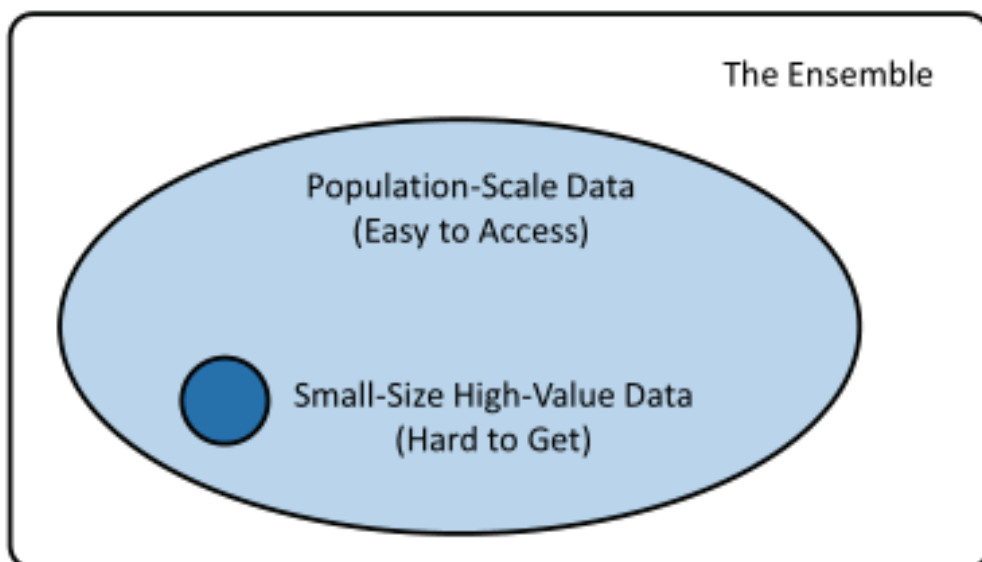


图1. 计算社会经济学框架下的新型数据集示例
来源:Gao, Zhang & Zhou. Physics Reports, 2019.

例如街景图像和文本内容。研究人员迫切地需要掌握更先进的数据挖掘和机器学习技术⁵。另一方面，利用全体尺度数据，可以专注于分析一小部分的抽样子集，然后对数据添加高价值的新维度。利用传统采集方法，例如问卷调查，可以方便获得数据的新维度。基于小样本数据训练得到的模型，能够从原始数据中推断出新维度。进一步把模型应用到整体数据集，可以获得所有个体数据的新维度。

这种新的集成分析框架囊括了一些常规方法，例如采样、标注和问卷，但它在实践应用中的作用更强大。例如，获取移动通信和行为活动相关的全体数据相对比较容易，而如果不进行全体规模的经济普查，就很难知

道每个家庭的家庭收入情况。在新的计算社会经济学框架下（图1），我们首先通过常规问卷调查，获得一些家庭的收入数据。然后，利用已经获得的小规模数据训练机器学习模型，进而基于家庭成员的手机数据预测一个家庭的收入情况。

尽管推断出来的数据并不完美，但如果预测算法设计地非常好，预测结果可以与真实数据非常接近。值得注意的是，新框架具有一个非常显著的优势，即能以很低的成本获得几乎所有个体的高价值数据。结合容易获取的全体尺度数据、高价值但难以获取的小样本数据、以及适当选择和精心设计的算法，就能推断出所有个体的高价值数据。这种研究框架是计算社会经济学研究的一种创新的

和具有代表性的方法论（图1），显示出了社会科学和计算机科学方法的深度融合。

感知贫困和经济增长

以近乎实时的和低成本的方式揭示社会和经济发展的状况，是阻碍实现联合国可持续发展目标（Sustainable Development Goals, SDGs）的长期问题之一。为了实现消除贫困的发展目标，第一步就需要准确地绘制贫困的空间分布地图。计算社会经济学中引入的新数据和新工具，例如卫星遥感数据和手机数据，已被用来更好地揭示、解释和预测全球贫困状况和经济增长情况。

遥感卫星提供的高分辨率数据，例如夜间照明（nighttime lights, NTLs）卫星图像数据，已经被用于推断与经济活动相关的信息，尤其针对传统经济普查数据不足的发展中国家。夜间照明数据可以清晰的指示出经济发展的空间分布。例如，Jean等人⁶应用深度学习算法关联分析了夜间照明数据和日间卫星图像数据。夜间照明数据能够预测财富的空间分布，而日间卫星图像包含更丰富的地表特征信息。利用从日间卫星图像中提取的特征，预测五个非洲国家的平均家庭资产情况，最高可解释75%的经济情况差异性。另外，该方法能够以很高的准确度重建出基于问卷调查得到的区域贫困指标。

手机能够记录大量的信息，这为获取个体层面的数据提供了一种十分有效的方法。基于与消费和生活支出相关的手机记录数据，能利用机器学习方法推断个体的社会经济状况，进而聚合得到国家和地区层面的社会经济状况估计。例如，Blumenstock等人⁷提出了一种新方法来探究发展中国家的手机使用情况与财富之间的关系，通过分析卢旺达的手机数据，例如所联络的不同地区的数量，他们发现家庭的支出与手机的使用情况正相关。Blumenstock等人⁸对部分手机用户进行后续访谈获得追踪数据，发现应用机器学习算法能很好地预测个体财富水平，准确地识别出相对比较贫困的个体。随后，他们对样本外的150万手机用户进行了预测，生成了高分辨率和准确性的卢旺达财富状况分布地图。文中提出的方法非常有应用前景，能用于描绘整个国家人口的财富和社会经济分布情况。

理解如何实现经济繁荣，是经济增长研究面临的长期挑战。Hidalgo和Hausmann⁹提出了一种新的称为经济复杂性（Economic Complexity Index, ECI）的非货币性指标，能定量评估国家未来经济增长的潜力。具体而言，他们提出了一种“反映法”（Method of Reflections, MR）来表征国际贸易中“国家-产品”二部分网络的结构，这种“反映法”产生的变量可以被当做经济复杂性度量指标。实证结果



此外, 社交网络平台上的个人行为数据, 也被用来评估个体的人格特征和心理状态, 如抑郁和自杀倾向。

表明, 国家的经济复杂性与他们的收入水平非常相关, 并且经济复杂性可以预测国家未来经济增长。随后, 统计物理方法也被用来定义一组新的指标, 能够刻画国家的健壮性 (Fitness) 和产品的复杂性 (Complexity)。在预测五年人均GDP上, Tacchella等人¹⁰提出的这种方法的准确率比国际货币基金组织 (IMF) 的现有方法高25%以上, 并且其误差是可预测的。这些复杂性和健壮性指标, 已经被用于定量刻画不同尺度上的经济复杂性和发展状况, 例如中国的区域经济复杂性¹¹。

感知区域和城市

高分辨率的数据和不断改进的方法, 使我们能够揭示区域和城市层面的经济活动和社会经济状况。例如, 源自于夜间照明 (NTLs) 和超高分辨率 (very high resolution, VHR) 图像的指标, 已被用来描绘出具有精细时空粒度的贫困分布地图。特别是, 来自手机和Google街景的新数据, 为感知城市和社区提供了一种非常有前途的方法。

贫民窟在中低收入国家中比较常见, 这些国家的基础服务 (例如供水、电力和卫生设施) 质量往往很差。识别和监测贫民窟地区, 对于实施旨在改善生活条件的政策具有

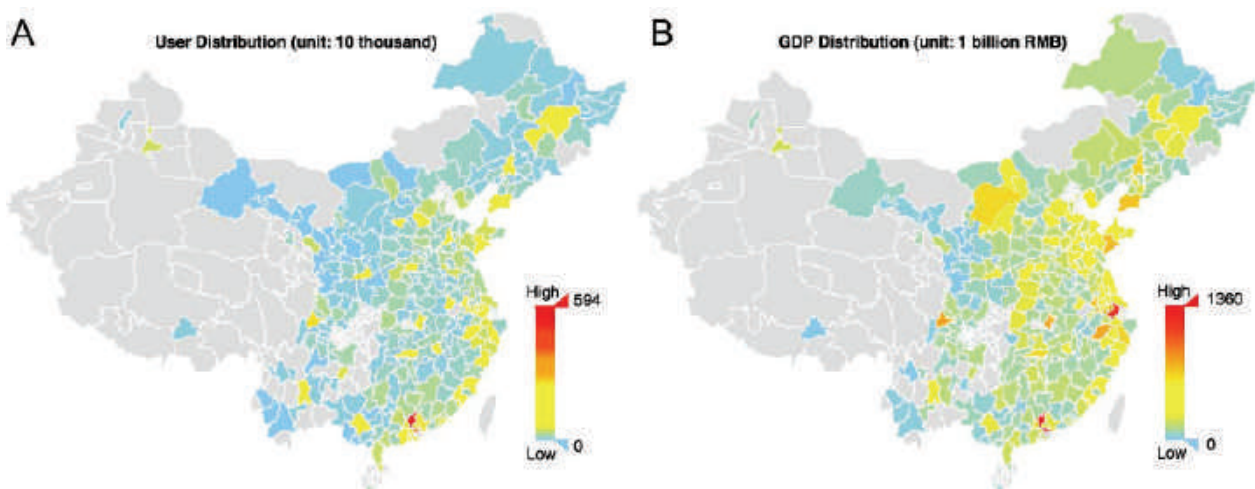


图2.中国地级市2012年(A)在线活动强度和(B)GDP数值的空间分布情况

来源:Liu et al. Physica A, 2016.

重要意义。最近，超高分辨率图像数据已经越来越多地用于清查贫民窟的位置和物理组成。例如，Kit等人¹²提出了一种“隐秘”（lacunarity）的概念，用来识别印度海得拉巴（Hyderabad, India）地区的贫民窟。他们使用的最佳模型能以高达83.33%的准确性识别贫民窟地区，且能追踪2003年至2010年期间贫民窟地区的变化模式。类似地，Kuffer等人¹³利用灰度共生矩阵（gray-level co-occurrence matrix, GLCM）的方差来从超高分辨率图像中识别出贫民窟地区。结果表明，在随机森林分类器（random forest classifier）中，通过向灰度共生矩阵添加谱信息（spectral information），可以将算法整体的准确性提高到90%。

社交媒体（social media）数据已被用于跟踪社会经济发展状况。例如，基于中国近2亿微博用户的注册位置信息，Liu等人¹⁴分

析了在线活动情况与社会经济指标之间的关系（图2）。结果发现，在线活动强度与社会经济指标密切相关，这表明能从城市级别的在线活动中推断出社会经济发展水平。尤为重要的是，他们进一步提出了一种检测异常城市的方法，这些城市的GDP水平远远高于拥有相同注册用户数量的其他城市。类似地，基于来自Gowalla平台的美国朋友关系和地理位置数据，Holzbauer等人¹⁵研究了区域经济发展水平和社会连接量化测度之间的关系。他们发现，美国跨州长程联系与三个经济指标密切相关，包括GDP水平、专利数量和创业公司数量。

众包方法和计算视觉技术已用于衡量城市的宜居性、安全性和不平等性，推断城市生活状况以及量化城市街道景观的变化。例如，基于几千张带有地理标记的街景图像数据，Salesses等人¹⁶提出了一种测量城市感观安

全性、等级性和独特性的方法，并将其应用到两个美国城市和两个奥地利城市。他们发现，两个美国城市的街景在感官上更加不平等。另外，城市街景感观的空间变化，有助于解释纽约地区（邮编分辨率）的暴力犯罪率变化。随后，基于在线问卷调查得到的数据，Naik等人¹⁷提出了一种场景理解模型，称为“街景评分”（Streetscore），利用通用图像特征来预测从街景图像中感知到的安全性。邻里区域的物理外观不是一成不变的，而是随着时间不断变化。Naik等人¹⁸提出了一种计算机视觉方法，基于不同时间的街景图像理解城市的物理外观动态变化。他们发现，能提升邻里区域物理外观状况的因素包括：较大的教育和人口密度、靠近市中心的地理位置以及良好的初始外观风貌。

基于深度学习的计算机视觉技术已应用于分析数字图像，提供了一种更快速、更低成本的社区调查替代方案。例如，Geburu等人¹⁹提出了一种新方法，能从200个美国城市的5000万张街景图像中估算出社会经济发展趋势。他们首先利用目标识别算法，从图像中自动检测出2200万辆不同车辆，然后使用卷积神经网络（convolutional neural network, CNN）确定车辆的特征，并将每辆车归类到2657个详细类别。基于所得到的数据，他们训练逻辑回归模型（logistic regression model）来估计种族和教育水平，采用岭回归模型（ridge regression

model）来估计收入和选民的偏好。与美国社区调查数据（American Community Survey）相比，得到的人口统计学估计在城市水平上具有非常不错的准确性。另外，该方法还能在更细粒度（邮编分辨率）下保持良好的准确性。例如，在估计西雅图的亚洲人占比时，在邮编分辨率级别上的预测结果与社区调查数据表现出高度的一致性。

性别平等和社会隔离

个体的人口属性会对其社会经济水平产生显著的影响，然而基于问卷访谈和人口普查的传统方法，在解析个人状况上成本高昂，且时间非常滞后。最近，来自社交媒体和手机等新型数据源的大规模数据，已逐渐替代性地用于预测个体的人口统计特征以及分析社会和宗教隔离。此外，社交网络平台上的个人行为数据，也被用来评估个体的人格特征和心理状态，如抑郁和自杀倾向。

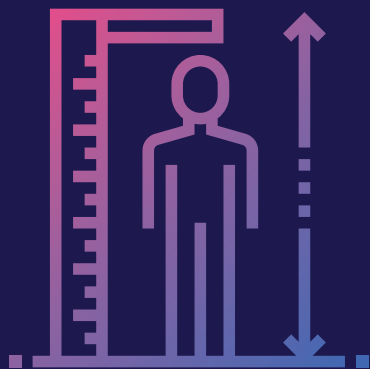
手机数据和在线数据已经用于推断人口统计信息，尤其是性别。Frias-Martinez等人²⁰分析了通话详细记录（call detail records, CDRs），发现男性和女性用户在行为和社交模式方面存在显著差异，例如通话时间和在社交网络中的连接度。他们提出了一种半监督分类算法（semi-supervised classification algorithm），能以高达

80%的精度识别手机用户的性别。Felbo等人²¹应用一种卷积神经网络架构，把每周的手机通讯数据转换为高阶的行为特征，然后应用卷积滤波器将不同周的行为模式聚合起来。基于带有径向基函数核（radial basis function kernel）的支持向量机（SVM），他们设计了一种二步模型，在性别预测上的准确率达到79.7%，效果略优于目前的最新方法。在研究在线平台使用的性别差异上，Mislove等人²²根据推特（Twitter）用户的姓氏来推断性别，其中推特用户量覆盖超过1%的美国人口。他们发现，71.8%的推特用户为男性，显示出推特用户的显著性别差异，更加倾向于男性用户。针对劳动力市场的身高溢价（height premium）问题，Yang等人²³分析了超过14万中国求职者的简历数据，发现身高溢价对女性的影响大于对男性的影响。特别地，随着教育水平的提高，身高溢价的性别差异逐渐减小；在固定所有控制变量之后，性别差异变得不显著。

通过应用新的大规模数据，不同社会经济水平的人口城市隔离也逐渐被展开研究。例如，基于路易斯维尔（Louisville）带有地理标记的推文数据，Shelton等人²⁴提出一种研究邻里内部隔离、流动性和不平等的方法。他们发现，路易斯维尔的邻里状况可以理解为流动性（fluid）、多空度（porous）和积极活动（actively produced）。类似地，基于一款手

机应用跟踪的数据，Yip等人²⁵分析了香港人的移动行为模式。他们发现，人们与其他收入群体的互动是有限的：富人倾向于搬到富人区，而穷人倾向于搬到穷人区。最近，Louf和Barthelemy²⁶更为直接地定义了居民隔离，发现高密度地区的富裕阶层人数过多。由此，他们特别建议将人口密度作为理解城市收入结构、解释城市中观察到的差异的相关因素。

来自社交网络的数据也已用于研究宗教隔离和城市本土化。通过分析基于微博数据构建的宗教社交网络，Hu等人²⁷对宗教隔离进行了量化分析。他们发现，宗教网络呈现出高度隔离，宗教隔离的程度高于种族隔离。此外，有46.7%的跨宗教联系可能与慈善议题有关，这表明慈善活动在促进跨宗教交流中的重要作用。通过分析中国五个大城市大约137万次签到记录（check-ins），Yan等人识别出了本地人和外地人的独特人口流动行为模式。他们发现，本地人访问地理位置频率的空间分布相对均匀，因为他们通常会重复访问对个人来说重要的地理位置。相比之下，外地人访问地理位置频率的分布更趋异质，因为他们倾向于访问热门地区。有了这样的洞见之后，Yang等人²⁸提出了一种称为“本地化系数”（indigenization coefficient）的指标，能完全基于签到行为数据估计一个人是本地人的可能性，或一个人在多大程度上表现得像本地人。该方法能用于估计外地人与本地



针对劳动力市场的身高溢价(height premium)问题, Yang等人分析了超过14万中国求职者的简历数据, 发现身高溢价对女性的影响大于对男性的影响。

人表现出相同的行为模式所需的时间, 并且可以提高人口流动性的预测准确性。

气候行动和救灾

气候变化和灾害监测对社会和经济系统至关重要。随着城市化进程的加快和气候的不断变化, 当前很多地区面临着前所未有的突发事件和自然灾害, 这些灾害对人类的生活和经济发展造成了许多威胁。为了应对这些问题, 亟需迅速的态势感知和有效的管理策略, 以减少人类苦难和经济损失。在农村地区, 对自然灾害的评估通常会很延迟, 给灾害响应和救济带来困难。在城市地区, 检测自然灾害(例如地震、洪水和飓风)不但对政府的快速灾难响应至关重要, 还对深入理解极端情况下的人类行为有意义, 将有助于更好地设计救灾策略。

新数据源已被用于提高灾难感知和应急管理, 例如遥感卫星、手机和社交媒体, 这些数据的显著优势在于获取成本低、更新及时和时空分辨率高。特别地, 深度学习算法已经用于分析卫星遥感数据, 更快速地绘制地震灾情地图。针对2010年海地(Haiti)地震, Cooner等人²⁹评估了几种深度学习算法在检测地震破坏中的有效性。他们发现, 利用多层前馈神经网络框架(multilayer feedforward neural network framework)从卫星图像中提取的空间纹理和结构特征, 能以低于40%错误率识别出受损的建筑物。类似地, Bai等人³⁰提出了一种深度学习算法, 绘制2011年日本东北地方太平洋近海地震(Tohoku Earthquake-Tsunami)造成的破坏地图。利用他们算法分析灾前和灾后图像, 能对地震造成的损害进行分类, 总体准确率为70.9%。

基于手机数据的快速应急事件识别，可以提高人道主义救援的响应能力，减少极端事件造成的损失。结合卢旺达手机活动数据和官方事件记录数据，Dobra等人³¹开发了一套行之有效的系统，能够识别在许多紧急事件和非紧急事件下具有异常行为模式的日期。手机数据也已用于评估大规模灾难中的人口应急安置，提高应急响应速度。例如，Lu等人³²研究了海地（Haiti）地震后人口迁移的可预测性。他们发现，地震发生后的三个月内，由于人口迁移，海地最大城市太子港（PaP）中的人口减少了23%。另外，在前三周内离开太子港的人群的目的地，与他们在正常时间的移动模式密切相关。

社交媒体是非常有价值的信息来源，有助于获取态势感知，检测和定位紧急事件，提高灾难应急响应，并改善救灾效果。社交媒体数据的逐步应用，已经改变了地震检测和预警的方法，可以在几分钟内根据与地震相关的推文绘制地震的分布图³³。例如，Acar等人³⁴研究了推特上的地震信息共享，分析了2011年日本东北地震中来自两个受灾地区附近的推文，他们发现，直接受灾地区的人们在推特上表达自己的不确定、不安全状况，而远离受灾地区的人们在推特上向粉丝们传递自己是安全的信息。社交媒体数据也已越来越多地用于即时性的洪水监控和灾情绘制。例如，Arthur等人³⁵利用推文数据来识别和定位英国的洪水事件。他们收集了

包含洪水相关术语的推文，通过分析许多指标（例如提到的地名和GPS坐标）来定位洪水事件。进而，根据带有地理标记的相关推文，制作了高质量的洪水事件地图，并通过官方数据验证了洪水事件地图的可靠性。

良好健康与福祉

SDG3的目的是使全人类过上健康和幸福的生活。在过去的十年，随着社会经济的发展 and 大数据时代的到来，人们在延长预期寿命、降低婴儿和产妇死亡率等方面取得了阶段性的成果。同时随着可利用的数据资源越来越丰富，许多基于数据的模型被建立，并在疟疾、季节性流感、大流行等传染病的预防等方面得到了广泛的应用，并且取得了很多重大进展。

由美国疾病预防控制中心（CDC）发表的一份报告显示，从2005年到2018年³⁶，美国平均每年大约有2841万个人感染流感、461111个病人需要住院治疗、40500个病人的死亡与流感相关，进而每年造成58亿美元的经济负担³⁷。由于流感病毒的血凝蛋白的进化，季节性疫苗的效力取决于疫苗所含抗原与流感病毒循环株所含抗原的匹配程度。Sah等人³⁸研究假设如果用75%的有效通用疫苗替代10%的典型季节性疫苗，采用疾病预防控制中心的监测数据建立流



行病传播模型，他们发现这种替换每年大约可避免530万例病例、81000例住院病人和6300例与流感相关的死亡。利用美国603个城市每周温度、相对湿度和大气压数据，Dalziel等人建立了一个基于气候的易感染态-暴露态-感染态-移除态-易感染态流行病传播模型。他们发现，城市的发病率与城市人口规模成正相关，小城市的流感季节周期较短。他们的研究进一步还表明城市中心在高峰气候条件外孕育的关键传播链能够改变群体免疫的时空几何学。为了更好的理解和预测流行病的传播，Liu等人³⁹基于意大利和荷兰两个国家高度精细化的社会人口统计数据，分别建立符合两个国家人口社会统计特性人工社会网络的子集，并采用真实流行病传播数据估计参数，发现流行病学经典理论在具有社会接触结构的人工社会网络上会失效。而针对流行病学经典理论的失效，他们基于贝叶斯理论提出一种方法能够较准确地预测流行病有效再生数，从而为流行病传播的演化和预测提供相关依据和支撑。Litvinova等人⁴⁰通过一项基于日记的接触调查实验，收集流感季节期间学校关闭之前和学校关闭之后人群的接触行为数据，同时综合这些微观的接触行为数据和宏观的社会人口统计数据，建立一个创新的混合的流行病传播模型，他们首先用真实数据验证了逐步反应性的学校关闭政策能够有效地缓解流行病的传播。

新发传染病如2003年的SARS、2009年的H1N1以及2019年的新型冠状病毒（COVID-19）的出现和流行严重影响成千上万甚至更多人的正常生活。由于新发传染病疫苗的缺乏和全球化的蔓延，高致病性的新型传染病增加了世界上各城市的入侵风险。Brockmann等人⁴¹根据航空流量数据，通过定义有效距离能够准确地预测疾病达到各个城市的时间，并在2003年的SARS和2009年的H1N1真实数据中得到验证。Zhang等人⁴²建立了一个数据驱动的流行病传播模型，该模型考虑全球各地区的人口数据、地区之间人群的流动性、社会经济、温度和媒介密度等数据，估计得到寨卡病毒进入到巴西的时间并揭示了寨卡病毒在美洲的传播特征。针对2019年新型冠状病毒（COVID-19），Chinazzi等人⁴³建立了一个基于个体的集合种群网络的流行病传播模型，该网络模型将全球190多个国家和地区分割成3300多个区域，各个区域之间由人口流动数据建立联系。他们基于中国境外发现的病例的相关信息，率先估计得到武汉潜在的疫情规模和基本再生数。类似于Chinazzi，Imai等人⁴⁴也采用海外确诊病例人员的信息以及武汉国际机场的交通流量，估算得到新型冠状病毒在中国传播的基本再生数，与Chinazzi⁴³的结果很相近。

愿景和行动

来自社会经济系统的大规模、高分辨率数据的可用性，已经提供了一种改善城市空间公平性的新途径。例如，通过分析两个西班牙城市的银行卡交易数据，Louail等人⁴⁵提出了一种自下而上的方法以改变个人的少部分购物行程来重新分配资金流动，实现更加平等的发展目标。具体而言，他们首先构建了“个人-业务”两个部分的空间网络，其中每条连边代表银行卡交易。然后，他们对“个人”和“业务”的连边进行重新连接，把个人的交易重新分配到位于不同社区的同一业务类别上。这样做的目标是，在社区之间重新平衡商业收入，但保留人口流动的基本特征。他们发现，仅需重新分配大约5%的个人交易记录，就能减少邻里之间80%以上的收入不平等性，甚至还能改善其他可持续发展指标，例如总体旅行距离和空间融合情况。他们的工作呈现了一种出色的众包（crowdsourcing）应用，即“罗宾汉效应”（Robin Hood effect），将资本重新分配以减少不平等。

计算社会经济学中引入的方法和数据源，将有益于开展针对实现可持续发展目标和评估这一进程的行动。特别是，在大数据和人工智能的推动下，本文提到的计算社会经济学中的新颖观点和创新方法，将很有希望成为促进实现可持续发展目标行动的主流研究框架。

参考文献

1. Gao, J., Y.-C. Zhang, and T. Zhou, Computational socioeconomics. *Physics Reports*, 2019. 817: p. 1-104.
2. Gao, J. and T. Zhou, Big data reveal the status of economic development. *Journal of University of Electronic Science and Technology of China*, 2016. 45(4): p. 625-633.
3. Mayer-Schonberger, V. and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. 2013, New York, NY, USA: Houghton Mifflin Harcourt.
4. Gao, J., *Research on the Spatial Structure and Dynamics of Socio-Economic Systems*. 2019, University of Electronic Science and Technology of China.
5. Lecun, Y., Y. Bengio, and G. Hinton, Deep learning. *Nature*, 2015. 521(7553): p. 436-444.
6. Jean, N., et al., Combining satellite imagery and machine learning to predict poverty. *Science*, 2016. 353(6301): p. 790-794.
7. Blumenstock, J., Y. Shen, and N. Eagle. A method for estimating the relationship between phone use and wealth. in *QualMeetsQuant Workshop at the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. 2010. ACM Press.
8. Blumenstock, J., G. Cadamuro, and R. On, Predicting poverty and wealth from mobile phone metadata. *Science*, 2015. 350(6264): p. 1073-1076.
9. Hidalgo, C.A. and R. Hausmann, The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. 106(26): p. 10570-10575.
10. Tacchella, A., D. Mazzilli, and L. Pietronero, A dynamical systems approach to gross domestic product forecasting. *Nature Physics*, 2018. 14(8): p. 861-865.

11. Gao, J. and T. Zhou, Quantifying China's regional economic complexity. *Physica A: Statistical Mechanics and its Applications*, 2018. 492: p. 1591-1603.
12. Kit, O., M. Lüdeke, and D. Reckien, Texture-based identification of urban slums in Hyderabad, India using remote sensing data. *Applied Geography*, 2012. 32(2): p. 660-667.
13. Kuffer, M., et al., Extraction of slum areas from VHR imagery using GLCM variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016. 9(5): p. 1830-1840.
14. Liu, J.-H., et al., Online social activity reflects economic status. *Physica A: Statistical Mechanics and its Applications*, 2016. 457: p. 581-589.
15. Holzbauer, B.O., et al. Social ties as predictors of economic development. in *Proceedings of the 12th International Conference and School on Advances in Network Science*. 2016. Springer.
16. Salesses, P., K. Schechtner, and C.A. Hidalgo, The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE*, 2013. 8(7): p. e68400-e68400.
17. Naik, N., et al. Streetscore--Predicting the perceived safety of one million streetscapes. in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2014. IEEE Press.
18. Naik, N., et al., Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences of the United States of America*, 2017. 114(29): p. 7571-7576.
19. Gebru, T., et al., Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 2017. 114(50): p. 13108-13113.
20. Frias-Martinez, V., E. Frias-Martinez, and N. Oliver. A gender-centric analysis of calling behavior in a developing economy using call detail records. in *Proceedings of the 2010 AAAI Spring Symposium: Artificial Intelligence for Development*. 2010. AAAI Press.
21. Felbo, B., et al. Modeling the temporal nature of human behavior for demographics prediction. in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*. 2017. Springer.
22. Mislove, A., et al. Understanding the demographics of Twitter users. in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 2011. AAAI Press.
23. Yang, X., et al., Height conditions salary expectations: Evidence from large-scale data in China. *Physica A: Statistical Mechanics and its Applications*, 2018. 501: p. 86-97.
24. Shelton, T., A. Poorthuis, and M. Zook, Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 2015. 142: p. 198-211.
25. Yip, N.M., R. Forrest, and X. Shi, Exploring segregation and mobilities: Application of an activity tracking app on mobile phone. *Cities*, 2016. 59: p. 156-163.
26. Louf, R. and M. Barthélemy, Patterns of residential segregation. *PLoS ONE*, 2016. 11(6): p. e0157476-e0157476.

27. Hu, J., Q.-M. Zhang, and T. Zhou, Segregation in religion networks. *EPJ Data Science*, 2019. 8: p. 6-6.
28. Yang, Z., et al., Indigenization of urban mobility. *Physica A: Statistical Mechanics and its Applications*, 2017. 469: p. 232-243.
29. Cooner, A.J., Y. Shao, and J.B. Campbell, Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti Earthquake. *Remote Sensing*, 2016. 8(10): p. 868-868.
30. Bai, Y., E. Mas, and S. Koshimura, Towards operational satellite-based damage-mapping using U-Net Convolutional Network: A case study of 2011 Tohoku Earthquake-Tsunami. *Remote Sensing*, 2018. 10(10): p. 1626-1626.
31. Dobra, A., N.E. Williams, and N. Eagle, Spatiotemporal detection of unusual human population behavior using mobile phone data. *PLoS ONE*, 2015. 10(3): p. e0120449-e0120449.
32. Lu, X., L. Bengtsson, and P. Holme, Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences of the United States of America*, 2012. 109(29): p. 11576-11581.
33. Allen, R.M., Transforming earthquake detection? *Science*, 2012. 335(6066): p. 297-298.
34. Acar, A. and Y. Muraki, Twitter for crisis communication: Lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities*, 2011. 7(3): p. 392-402.
35. Arthur, R., et al., Social sensing of floods in the UK. *PLoS ONE*, 2018. 13(1): p. e0189327-e0189327.
36. Centers for Disease Control and Prevention, Seasonal influenza vaccine effectiveness, 2005–2018. <https://www.cdc.gov/flu/professionals/vaccination/effectiveness-studies.htm>. Accessed 26 July 2018.
37. Ozawa S., et al., Modeling the economic burden of adult vaccine-preventable diseases in the United States. *Health Aff. (Millwood)*, 2016.35, 2124–2132.
38. Sah P., et al. Future epidemiological and economic impacts of universal influenza vaccines. *PNAS*, 2019.116(41), 20786-20792.
39. Liu Q.-H., et al. Measurability of the epidemic reproduction number in data-driven contact networks. *PNAS*, 2018.115(50), 12680-12685.
40. Litvinova M., et al. Reactive school closure weakens the network of social interactions and reduces the spread of influenza. *PNAS*, 2019.116(27), 13174-13181.
41. Brockmann D., et al. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*, 2013. 342, 1337-1342.
42. Zhang Q., et al. Spread of Zika virus in the Americas. *PNAS*, 2017.114, E4334-E4343.
43. Chinazzi M., et al. Series Reports Entitled “Preliminary assessment of the International Spreading Risk Associated with the 2019 novel Coronavirus (2019-nCoV) outbreak in Wuhan City” (unpublished).
44. Imai N., et al. Transmissibility of 2019-nCoV (unpublished).
45. Louail, T., et al., Crowdsourcing the Robin Hood effect in cities. *Applied Network Science*, 2017. 2: p. 11-11.