

自研分析工具深挖“地球大数据”

■本报记者 胡琅琦

2015 年,联合国正式通过《改变我们的世界——2030 年可持续发展议程》,提出了包含一系列涉及 17 个领域 169 个具体问题的可持续发展目标(SDGs)。但是,由于指标数据和方法的缺失,可持续发展目标实施进程监测始终面临艰巨的挑战。

地球大数据具有宏观动态多尺度监测能力,是实现 SDGs 的重要手段。然而,科学家们要如何对那些海量、多源、多时相、高度集成的大数据进行挖掘分析,进而开展 SDGs 评估?

在近日举行的 2021 年可持续发展大数据国际论坛上,中国科学院软件研究所(以下简称软件所)研究人员做了题为《面向 SDGs 的交互式在线分析技术与系统》的技术报告,对这一问题进行了解答。

数据、分析工具缺一不可

承诺消除一切形式和表现的贫困,包括到 2030 年时消除极端贫困,是 SDGs 一项关键目标。然而,要摸清贫困国家的贫困状况本身就很困难。

传统的人口调查方式成本极高,且严重依赖稳定的国内环境,许多发展中国家根本负担不起。而利用遥感技术,捕捉夜间光亮的卫星图像数据,结合全球人口动态统计分析数据库的人口数据,使用统计方法估计国家和区域的社会经济状况,提高对全球贫困问题的及时感知,已经不是什么新闻。

从理论上讲,更高的亮度意味着更多基础设施、更高的发展水平。但是,很多模型无法区分极端低光照水平和零光照水平。而且,在人口密集的地方,仅凭夜间光照水平评估贫困状况也会有很大偏差。

2016 年,斯坦福大学研究人员在《科学》上发表了一项研究成果。他们把高精度卫星图像和深度学习、迁移学习等技术结合起来,从而预测了非洲五国尼日利亚、乌干达、坦桑尼亚、卢旺达和马拉维的贫困状况。

这项研究的创新之处在于,科学家用高精度、高质量的卫星遥感图像数据来训练一个深度卷积神经网络,以预测同区域的夜光数值大小。但这并非最终目的,而是为了提取遥感图像数据中的特征,比如道路、房屋、汽车等,然后把这些特征和少数采样点的贫困调查数据作为标签,训练一个普通的神经网络。这一计算工具最终可以用来高精度地预测一个地区的贫困程度。

“这项研究给我们的启示是,要想实现 SDGs 指标评估,数据和有效的挖掘分析工具是缺一不可的。”软件所副研究员刘杰说道。他同时指出,SDGs 指标量化评估涉及地



地球大数据挖掘分析系统 EarthDataMiner 作为“可持续发展大数据平台系统”重要组成部分正式公开发布。

“我国科研团队自主研发了一套地球大数据挖掘分析云服务系统(EarthDataMiner),支持科学家在线开展遥感影像及其他科学数据的智能分析处理,支持 SDGs 指标全流程在线计算,从而降低科学家利用云平台、大数据、前沿人工智能算法开展大规模数据分析的难度。

球大数据分析处理的全流程,包括遥感影像的访问与语义分析、各种数据产品的解析和预处理、多源数据的融合计算与可视化等,需要采用大数据与人工智能等大量前沿技术,这也给开展 SDGs 评估的科学家们带来了一系列技术挑战。

“虽然针对大数据挖掘分析已经有一系列通用软件,包括 Matlab、Python 等单机版软件,还有分布式环境软件和云服务模式的分析软件,但对领域科学家而言,这些软件存在巨大的应用开发鸿沟。”刘杰表示,面对海量、多源、多时相、高度集成的大数据,科学家需要软件提供数据处理功能的服务、分析模型的服务,以及成果应用的服务。

自主研发地球大数据挖掘分析系统

9 月 6 日,全球首个以大数据服务联合

国 2030 年可持续发展议程的国际科研机构——可持续发展大数据国际研究中心正式成立,由其开发的“可持续发展大数据平台系统”也正式发布。作为它的重要组成部分,依托平台系统的栅格数据引擎 Databox、广目云平台 CASEarthCloud,科研团队自主研发了一套地球大数据挖掘分析云服务系统(EarthDataMiner),支持科学家在线开展遥感影像及其他科学数据的智能分析处理,支持 SDGs 指标全流程在线计算,从而降低科学家利用云平台、大数据、前沿人工智能算法开展大规模数据分析的难度。EarthDataMiner 支持科学家访问海量数据,提供算法 API,支持在线编写代码开发 SDG 算法,并将算法成果发布为 Web App 工具,支持全球用户访问使用。

据介绍,基于 EarthDataMiner 实现的 SDGs 指标在线计算通用流程具体包括:将指标计算相关数据上传到系统;编写 Python

代码进行数据处理;遥感影像数据检索;遥感数据信息提取;部署指标计算算法与模型;采用 EarthDataMiner 提供的 Web App 工具的开放框架,将 SDGs 评估算法发布为 Web 页面工具。

刘杰解释,EarthDataMiner 最重要的就是要给科学家们提供一个分布式遥感影像计算函数库,平台预计研制百余项 SDGs 在线计算的函数及配套文档和案例,涵盖遥感影像计算与机器学习算法。“对科学家而言调用的只是简单的函数,而平台能把它‘翻译’成高效的分布式任务去执行。至于如何‘翻译’得又准确又高效,取决于算法库的不断优化。”

提供交互式在线分析工具

目前,在中国科学院战略性先导科技专项“地球大数据科学工程”专项的统一部署下,EarthDataMiner 研发团队已和相关科学家团队合作,基于 EarthDataMiner 开发了 4 个 SDGs 指标在线评估工具,即 SDG6.6.1 地表水随时间变化评估工具、SDG11.3.1 城镇化进程评估工具、SDG15.1.1 森林覆盖率评估工具、SDG15.3.1 土地退化零增长评估工具。

以 SDG15.3.1 全球土地退化指标监测为例。作为“2030 年可持续发展议程”的一部分,SDG 第 15 项目标是:保护、恢复和促进可持续利用陆地生态系统,可持续管理森林,防治荒漠化,制止和扭转土地退化,遏制生物多样性的丧失。每个特别发展目标都有具体目标,涉及陆地上生活的不同组成部分。其中 SDG15.3 的目标是:到 2030 年,防治荒漠化,恢复退化的土地和土壤,包括受荒漠化、干旱和洪水影响的土地,并努力实现零土地退化世界。

“作为实现 SDG 15.3 的主要抓手,土地退化零增长(LDN)的执行涉及基线确定、土地利用规划、进展评估 3 个关键工作。”中科院空天信息创新研究院研究员李晓松在《中国科学院院刊》发表的文章《地球大数据促进土地退化零增长目标实现:实践与展望》中解释,“其中退化土地(SDG 15.3.1)基线与动态监测是核心,不仅可以直接评估 LDN 进展情况,也能为土地利用规划提供信息支撑。”

李晓松表示,发展基于云计算的 SDG 15.3.1 在线计算工具是地球大数据支撑土地退化零增长的一项重要贡献。这一工具通过对接 Databox、EarthDataMiner 等多个数据计算引擎和数据环境,可为用户在线提供基于感兴趣区域的 SDG 15.3.1 单指标计算评估、综合计算评估,包括土地覆盖、土地生产力与土壤碳三个方面。在他看来,这对能力相对落后国家的 SDG15.3.1 监测评估具有重要意义。

下一代域名系统:虚拟世界“导航”建设新思路

■本报记者 倪思洁

正如拜访朋友前要先知道朋友家怎么走一样,访问虚拟世界里的网站同样也得知网站在哪里、如何到达,要做到这一点,就需要一个被称为“域名系统”的“导航系统”。

近日,记者对互联网域名系统国家工程研究中心(以下简称域名国家工程研究中心)了解到,该中心于近期获得了近亿元的 B 轮融资,主要用于下一代域名系统(NG-DNS)技术的研究和开发。

那么,“下一代域名系统”这个概念是怎么来的,有什么意义,又将如何实现?对此《中国科学报》采访了域名国家工程研究中心的相关负责人。

互联网发展需要更高效的“导航”

域名系统是互联网服务的入口。然而,在万物互联的时代,如何在有海量需求的情况下,快速、准确地完成“导航”,成为现实对域名系统的考验。

“海量边缘和移动设备对寻址解析架构提出了新要求,传统中心化域名系统架构需要向去中心化、轻量化架构发展。”域名国家工程研究中心主任毛伟告诉《中国科学报》。

他表示,海量互联互通的背后,域名解析量呈几何级爆发式增长,互联网域名系统在性能、安全、扩展、可靠性等方面都需要全面升级。与此同时,在过去数十年里,计算、存储和网络传输能力都在不断升级、融合、叠加,域名系统同样要从基础解析寻址向“全面感知、可靠传输、智能处理、精准决策”的新模式升级。

不仅如此,“全球互联网核心资源,如顶级域名、互联网地址空间等,正处于新一轮的分配和关键部署阶段,域名系统的重构,也将关系到我国在未来全球互联网基础技术领域的话语权。”毛伟说。

在毛伟看来,时至今日,域名系统的发展早已不再满足于做好简单的“导航”,而是要

推动互联网治理、网络资源、软硬件系统的全面升级。

国家信息化需要更安全的“导航”

毛伟等人提出“下一代域名系统”的新理念,不仅是因为来自新基建、互联网应用、技术变革、互联网核心资源再分配的需求驱动,还因为国家信息化发展对互联网根基自主可控提出了新的要求。

回溯传统的互联网技术产业格局,互联网在诞生之初就已确立了“三层架构”——物理设施层、基础资源层、应用层。

毛伟介绍,物理设施层是由计算、存储、网络构建的有线和无线网络,好比信息高速公路;应用层是金融、政务、购物、娱乐、搜索、社交等基于互联网的各种应用,就像跑在高速公路上的汽车;两层中间还有一层基础资源层,由域名系统和 IP 地址组成的寻址解析系统,好比导航,导航一旦失效就会断网。

自我国于 1994 年加入国际互联网至今,中国互联网经历了近 30 年的飞速发展。然而,“与蓬勃发展的互联网应用和逐步走向国产化替代的物理设施的状况完全不同,中国仍然面临网络根基‘断网’风险。以域名系统为例,包括根、顶级域在内的互联网核心基础资源的自主率极低,承担基础服务的域名系统解析设备 98%使用的是国外软件。”毛伟说。

当前,国家高度重视互联网域名系统的发展,《“十三五”国家信息化规划》指出要建设互联网域名安全保障系统;《推进互联网协议第六版(IPv6)规模部署行动计划》提出要加快互联网域名系统的全面改造。

“域名系统的英文简称是 DNS,对 DNS 我们有新的诠释。D 是‘域’(Domain),域名系统是全球最重要的公共基础设施之一,我们需要在标准建设、治理规范、技术革新等层面有参与、有担当、有作为,为构建网络空间命运共同体贡献力量。N 是‘名’(Name),新顶级域名的开放申请,开启了全球互联网空间



图片来源: unsplash

资源发展的新契机,我们要在规则制定、平台建设、技术标准等方面做好充足准备。S 是‘系统’(System),域名系统是重要的核心业务控制枢纽,是支撑网络安全稳定运行的关键基础设施,我们需要打造自主可控的网络枢纽。”毛伟说。

下一代“导航”,兼顾共享与自主

“下一代域名系统是传统域名系统的延伸,除了寻址解析之外,还涵盖了互联网根基、顶级域市场和企业级市场,是从互联网人口到互联网根基的全栈域名系统。”毛伟说。关于“下一代域名系统该如何实现”,域名国家工程研究中心总经理邢志杰介绍,应当从新体系、新技术、新产业三个维度,打造下一代域名系统。

“新体系即以互联网根基、顶级域和企业级的全栈技术,参与全球互联网体系治理;新技术即以自主可控技术为支撑,满足对海量的解析和智能调度需求;新产业即以数字化

转型和新基建为契机,推动传统域名行业和产业升级。”邢志杰说。

在毛伟和邢志杰等人看来,下一代域名系统的建构将是一个既助力全球互联网共享共治,又实现我国域名系统自主可控的过程。

毛伟介绍,目前,在构建网络空间命运共同体方面,我国正在积极参与网络根基治理,域名国家工程研究中心参与起草了多项 IP 根(RPKI)相关国际标准以及多份互联网名称与数字地址分配机构(ICANN)根服务器治理报告。在实现软件自主可控方面,域名国家工程研究中心推出了自主研发的域名系统基础软件“红枫”,相比国外同类软件获得数十倍性能提升,且全面适配国产 CPU 和操作系统。

邢志杰表示,域名国家工程研究中心近期获得的融资中,80%以上将投入到下一代域名系统技术的研究和开发之中。“需求引领研发。把中国的网络问题解决了,就能够引领国际,也就自然而然地能够为国际互联网大家庭作出应有的贡献。”邢志杰说。

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

智能与「云」技术登陆冬奥赛事直播

可预测新冠患者用氧需求的人工智能算法问世

本报讯 近日,英国剑桥大学科学家与来自世界各地的 20 家医院及医药公司联手,首次利用人工智能与机器学习技术,在全球范围内预测新冠肺炎患者的氧气需求,准确率高达 95%。这项名为 EXAM 的研究于发表在近日的《自然—医学》期刊上。

研究人员利用来自全球的数据,通过人工智能工具,分析具有新冠肺炎症状的医院患者的胸部 X 光片和电子健康数据,进而预测 COVID-19 患者在医院治疗的最初几天可能需要多少额外氧气。而且,经过深度学习后,该算法会将分析整合在一起,构建能够预测世界各地医院 COVID-19 患者的氧气需求的人工智能工具。

此外,为了严格保护患者隐私,所有数据均为匿名,并向每家医院发送了一个算法,因此数据不会被共享或移位。

相关论文信息:
<https://doi.org/10.1038/s41591-021-01506-3>

日研究人员发明更轻更长新型机械臂

本报讯 日本科学与技术高级研究所和东京大学的研究人员最近开发出一种机械肢体,取名为 AugLimb。与其他大多数被安装在上半身的可穿戴机械手臂相比,AugLimb 折叠后的体积小、重量轻,不影响佩戴者的日常活动,适用于长时间佩戴。而且,研究人员增加了 AugLimb 的长度,是人类平均前臂的大约 2.5 倍长。它可以支持人类完成各种任务,如烹饪或清洁。

目前,AugLimb 仍处于原型阶段,该团队计划继续进行测试并改进其能力,实现商业化。该研究论文已发表在 arXiv 上。

瑞莱智慧发布升级版隐私保护平台

本报讯 当前,数据安全、数据孤岛等问题让释放数据价值掣肘颇多,如何充分保证数据安全并提升数据融合能力尤为重要。日前,清华大学人工智能研究院孵化的产学研技术企业——北京瑞莱智慧科技有限公司(以下简称瑞莱智慧,RealAI)发布了全新升级版的隐私保护计算平台 RealSecure。

此次发布的 RealSecure 平台能够在确保数据隐私保护的前提下促进数据流通,实现“数据不出库、模型多跑路”,以保护数据隐私和数据安全,为人工智能等创新应用提供更好的数据环境。

其中,“数据不出库,模型多跑路”是该平台的核心功能,在数据不出本地、无中间方作为协调者的前提下,完成多方参与的模型训练并获得可投入生产的模型,既保障数据隐私安全,又以人工智能的方式驱动整个数据处理闭环。

“传统的做法可理解为‘盗版印刷’,我们实现的是‘活字印刷’,兼容主流机器学习算法,无需改写,只要调用函数,在编译器里重新编译一遍就可以。”瑞莱智慧合伙人朱萌萌解释,“RealSecure 平台可以以数据科学家以最熟悉的方式使用隐私计算,大幅提升易用性,这也是加速隐私计算落地的最优路径。”