

大数据时代的空间数据价值

——《空间数据挖掘理论与应用》评介

周成虎

(中国科学院地理科学与资源研究所,北京 100101)

今天,进入一个新的时代:互联网将一切的人、事、物都链接在一起,社交网、物联网、传感网等各类基础网络与应用网络系统不断涌现,一切都可以数据化,一切都被数据化,一个数据的世界正在飞速发展。正如维克托·迈尔-舍恩伯格在《大数据时代》一书中所说:大数据正在改变我们的生活以及理解世界的方式,成为新发明和新服务的源泉,而更多的改变正蓄势待发。一个以数据分析为核心的科学范式得到科学界的认同。我们也清楚地认识到:一切事物都在一定的空间中存在,一切事物也都在一定的时间内发生。时间和空间是世上万事万物的基本属性,因此,从空间视角认识数据世界,用空间数据构建世界,挖掘空间数据中所隐含的知识,是大数据研究的重要内容。

李德仁先生高瞻远瞩,深刻把握时代的脉搏,率先在国际上带领团队开展空间大数据知识挖掘研究,并坚持不懈,取得了创新性的研究成果。《空间数据挖掘理论与应用》专著的出版,不仅反映了李先生团队的研究成果,更指明了学科研究方向和科学上的突破点。

新版专著分为十二章,系统地阐述了空间数据挖掘的理论、方法与应用。前三章从概念和实际应用出发,叙述了学科发展的态势,论述了空间数据挖掘内涵,将空间尺度、空间知识类型等核心概念有机联系在一起,并从数据流的角度对空间数据获取、处理与分析、组织与管理等过程进行了剖析,为深入认识空间数据的内涵奠定基础,特别是对非地学领域专家认识和理解空间数据挖掘研究,提供了入门的基础。第四章和第五章转换视角,从数据挖掘的学科出发,系统阐述空间数据清理和空间数据挖掘的理论,拓展了空间数据分析的视角,特别是为地学分析专家提供了理论与方法的指点。

数据场、云模型、粗空间、概念格等四大概念与方法构成了该专著的核心理论体系与创新之处。该专著借助场理论,结合空间数据的特点,提出了场与场强、势与势场等概念的数据场理论,发展了相应的数据分析与可视化的方法;运用粗糙集理论,构建了由粗实体、粗关系和粗算子构成的地学粗空间;基于空间不确定性理论,提出了包含概率论的正态分布和模糊集中的隶属函数的云模型,实现了用不确定性来描述不确定性;针对空间数据挖掘的一般性理论研究,构建了形式化表达空间概念的形成、空间概念之间的结构关系与层次关系等空间概念格方法,为空间关联与空间概念聚类奠定了基础。这四个核心,相互融合,互成一体,为空间模式的发现、空间关联关系的探测等奠定了较完备的理论方法基础。

为了更好地展现空间数据挖掘理论方法的应用前景,该专著最后三章选择了GIS数据、遥感图像和滑坡过程数据作为案例,开展深入研究。这些案例也为相关领域研究人员提供了学习的典型案例、思考的典范。空间数据挖掘是一个崭新的研究领域,尚处于发展的中初期,尚有许多科学问题需要探索,尚有许多方法需要建立,更需要跨学科合作研究,该专著则是一个多学科团队的长期研究结晶。李德仁院士在测绘科学有着卓越的建树,李德毅院士在人工智能领域有着深厚的科学积累,他们二位先生联合培养的王树良教授融二老之大成而成,实现了多学科的交叉融合,并进一步发展,开创了一个崭新的空间数据挖掘的新局面。

科学研究是一项令人着迷的工作,也是一份需要耐得起寂寞的职业。当代大数据轰轰烈烈,大数据商业热热闹闹,但作为一个科学领域,则需要喧闹的环境下,静心思考,潜心研究。《空间数据挖掘理论与应用》则是两代科学家二十多年科研成果的总结晶,在这璀璨光辉的背后则是点点滴滴汗水与勤劳,这也正是我国科技强国之梦实现需要的精神。

《空间数据挖掘理论与应用》的第一版问世于2006年,一经出版,即在数据挖掘与知识发现等研究领域引起共鸣,国内外学界给予了高度评价,有力地推动了空间数据挖掘研究的大发展。写此评介,一是表达对作者的深深敬意,也敦促自己需要奋发努力;二是向行业学者推介,期望更多学者参与推动该领域的研究。