中国常业信息

第32卷第4期2020年8月

China Agricultural Informatics

2020, 32 (4): 32-41

Vol.32, No.4 Aug., 2020

Hadoop 支持下的土地利用变化可视化方法探讨*

张定祥¹,安 琼²,季宏伟¹,武建佳²,李士江¹ (1.中国国土勘测规划院,北京100035; 2.北京博信高科信息技术有限公司,北京100029)

摘要:【目的】掌握土地利用数据量大、数据模型复杂的特点,研究土地利用变化数据可视化表达模型,有助于探索直观有效的土地利用现状与变化数据的可视化方法。【方法】针对土地利用现状与变化数据分析展示过程中的时序展示、空间展示、流向变化展示三方面的需求,创新设计出多年土地利用流向图、多年地类图斑变化视频、三维立体图 3 种展示方法,研发了相关组件、实例以及 API 接口,并利用 Hadoop 并行技术完成了海量数据快速预处理。【结果/结论】从适用数据、处理流程、展示效果、交互、接口、多图表交互等多个层面结合研发的土地利用变化可视化成果,较好地实现了土地利用数据的交互式可视化展示,可为今后海量土地基础数据分析挖掘提供技术支撑。

关键词: 土地利用; 基础数据; 可视化; Hadoop

DOI: 10.12105/j.issn.1672-0423.20200404

0 引言

通过多年土地调查工作,全国范围的土地利用及变化数据资源不断积累,土地数据的挖掘分析需求日益旺盛。各地针对土地利用变化数据的挖掘分析工作不断深入,借助专业软件已经积累一些土地利用数据可视化展示方式。但是,目前土地利用现状及变化数据的展现形式大多停留在传统的专题图(专题统计地图)、统计表(表格)、统计图(饼图、折线图)方式,这些展示方式可以直观反映出土地资源的现状和变化情况,但是展示交互性、直观性仍有一定局限。特别是对于多年土地利用数据的时间特征、空间特征,传统的专题图、统计图展示方式均不能动态展示,无法直观地展现土地利用数据的变化特点。土地利用现状及变化数据在空间维、时间维、动态流媒体可交互展示方面研究和应用仍然较少。

随着 Hadoop 大数据并行技术的推进,一系列新型展示方式逐步出现,并在其他领域进行了应用,取得了良好效果,创新地展示方式极大的提升了数据的易读性,使得数据的价值能够快速被读者接受。基于此,土地利用数据的直观表达需要借鉴其他领域数据的可视化表达方式,研究筛选适合土地利用现状及变化数据可视化表达的形式,为海量数据资源分析挖掘结果的直观展示积累技术。文章将结合土地利用变化数据的特点和应

收稿日期: 2020-05-04

第一作者简介: 张定祥(1969—),研究员。研究方向:遥感应用与土地调查数据集成。Email:dingxiangzhang@163.com*基金项目:"十二五"国土资源调查评价土地基础数据整合集成与共享平台建设(DCPJ121707-01)

2020年8月-

用需求,以土地利用现状及变化直观表达为目标,通过借鉴其他领域数据的可视化表达 方式,研究适合土地利用数据可视化表达的形式,进而提升土地基础数据的决策支撑能力,带动新的展示技术在土地行业中的广泛应用。

1 系统模型

1.1 总体流程

根据土地利用变化数据特点以及可视化需求,该文提出了土地利用数据可视化展示的总体设计流程,如图 1 所示。土地利用变化数据的展示总体流程可以概括为数据标准化处理、数据分析处理和可视化展示 3 个步骤。

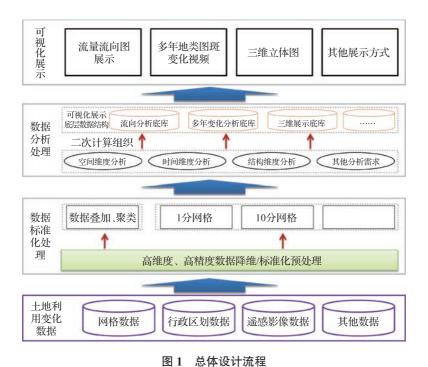


Fig.1 Overall design flow

数据标准化是指将现有的各类土地利用变化数据和相关的基础地理数据,根据分析需求进行处理,将高维度、高空间精度、高时间精度的数据进行降维和标准化处理。通过预处理,可以将海量数据汇总为数量较小但不影响宏观分析展示的数据集合。例如,在预处理过程中,可以将全国土地利用图数据进行汇总,形成全国范围的1分网格数据(长宽均为1分的正方形);或者按照用地类型,汇总为不同时间维度、不同地类的总量数据。

数据分析处理是指将标准化数据,根据展示需求进行二次计算,生成可以快速实现 数据可视化的底层数据结构,如将多年地类面积数据转换为按年度关联的流向数据模型, 或将不同网格内新增加建设用地面积数据转换为以网格为中心点、以用地面为高度的三 维数据模型。

数据可视化展示是指将计算的数据进行渲染,并通过程序语言实现可以交互操作的、 直观化的图像表达方式。该文围绕土地利用数据的特点,针对土地数据分析展示过程中 的时序展示、空间展示、流向变化展示三方面的需求,设计了多年土地利用流向图、多 年地类图斑变化视频、三维立体图 3 种有代表性的展示方法,并针对土地利用数据量大, 处理时间长的问题,设计了基于 Hadoop 并行计算技术的快速数据预处理方法。

1.2 土地变化流向数据可视化展示模型

流向数据可视化展示模型基于土地利用变化指标动态变化的特点,实现对土地利用变化情况直观展示分析。变化情况流量流向分析主要适用于土地利用现状调查数据、土地利用规划数据以及耕地数据等年度变更数据^[1]。进行多年数据变化流量流向分析,首先对任意两年的现状数据进行叠加、计算,形成新的年度变化数据;其次按照三大类、一级类、二级类或其他分析分类要求进行汇总,循环叠加处理后的数据通过可视化预处理,形成数据文件,通过前台可视化展示组件进行展示。具体技术路线如图 2 所示。土地利用变化流向图,可量化展示不同时间的土地利用数据在不同类型间的变化情况^[2]。

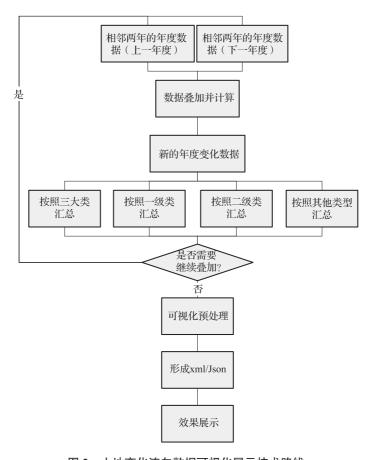


图 2 土地变化流向数据可视化展示技术路线

Fig.2 Technical route of visualization of land change flow data

2020年8月-

1.3 土地利用空间变化数据视频化表达模型

土地利用空间变化数据视频化表达模型主要针对土地利用变化数据多年度时序变化的特点,利用 Hadoop 大数据并行处理技术,结合空间数据计算,快速提取多年度土地变化图形,将变化情况平均等分成多桢,并利用 GIS 要素渲染技术,渲染为图片,继而利用视频技术,以视频的形式展示多年空间数据时序动态变化^[3]。多年土地利用空间数据时序变化视频展示将多年的地图文件按照国家、省、市县 3 种类型进行数据处理,其中全国范围的按行政区聚合并计算各地类面积及百分比,再与遥感图层叠加并绘制地图,转换为数据文件后进行并行绘制,最终形成视频展示文件。具体技术路线如图 3 所示。

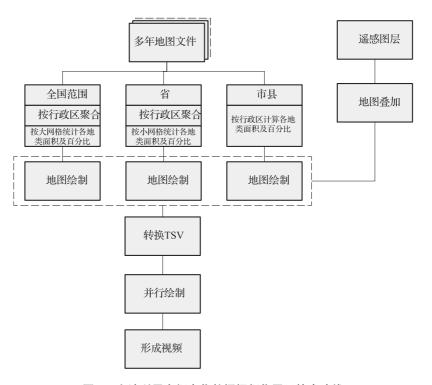


图 3 土地利用空间变化数据视频化展示技术路线 Fig.3 Technical route for video presentation of land-use spatial change

1.4 空间网格指标数据三维立体化表达模型设计

指标数据三维立体化表达模型主要是基于土地利用变化统计数据产品数据量大、数据格式较为规范统一、便于计算等优势,针对数据指标空间化展现、动态变化等特点,并借鉴其他领域的三维立体化展示方法,采用"三维地球"模型进行量化展示^[4]。从空间数据维度而言,三维立体化展示方法可用于分析行政区空间统计数据、空间网格统计数据以及空间点位统计数据;从数据类型而言,可分析土地管理类统计数据,以及与土地相关的人口、GDP等社会经济数据;从数据指标而言,可分析例如农用地面积、建设用地面积、未利用地面积、国有土地面积、集体土地面积、土地供应面积、土地储备面积、地类分布面积等土地核心指标,以及与人口、时序值、社会经济数据等叠加计算形

第 32 卷第 4 期

成的人均值、变化量、变化幅度等衍生指标。

对于大范围的地图文件,按照行政区进行聚合,划分网格,按网格进行统计数据的 均一化处理;对于市县等小范围地图文件,则直接进行数据均一化处理、格式转换,由 此形成三维可视化展示效果。具体技术路线如图 4 所示。三维立体展示中以虚拟的三维 地球为载体,以柱状图展示某一主题的量度,各类统计数据可按照行政区划、空间网格 或点位信息进行三维空间的数据展示^[5]。

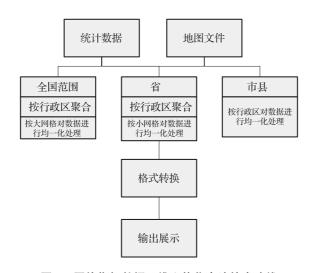


图 4 网格指标数据三维立体化表达技术路线

Fig.4 Technical route for three-dimensional representation of grid index data

2 系统实现

基于上述提出的系统模型和数据可视化表达技术路线,本文实现了相关的数据处理模型、算法,研发了可视化展示工具,下面对各个可视化展示方法的实现及其效果进行介绍。

2.1 土地利用变化流量流向展示

土地利用变化流量流向图以用地类型为分析维度,通过不同的色块表示各地类,以连线表示地类间的流向,即某一地类分别流向其他地类的情况;以连线的宽度代表地类的流量,即线条越宽,地类流量越大,线条越窄,地类流量越小^[6]。流量流向图为动态的展示图,当鼠标移动到某地类流量流向图上时,会突出显示该地类具体流向线以及实际的流量数据,同时,利用数据列表,明确地展示地类的流向与流量。以选取的某试点区 2013 年初至 2014 年末土地利用现状数据一级地类、二级地类流向图为例展示土地利用变化情况流向图。具体如图 5~6 所示。2013 年、2014 年该区土地利用变化流量流向图中,左侧为 2013 年初各地类面积,中间代表 2013 年末各地类面积,右侧为 2014 年末各地类面积,从线条宽度变化可以看出此阶段各地类的变化情况,同时通过线条的连接方向,可以直观地展示各地类的流向情况。

2020年8月-

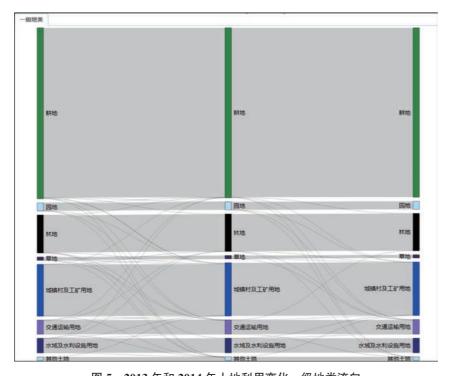


图 5 2013 年和 2014 年土地利用变化一级地类流向 Fig.5 Flow maps of land-use change of first-level types in 2013 and 2014

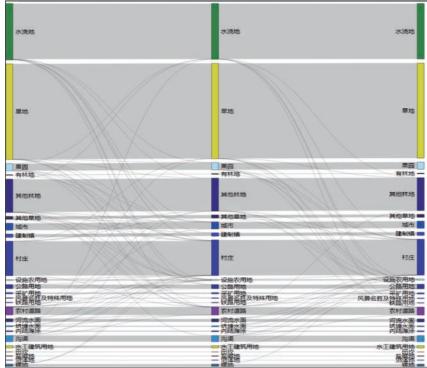


图 6 2013 年和 2014 年土地利用变化二级地类流向

 $Fig. 6 \quad Flow \ maps \ of \ land-use \ change \ of \ second-level \ types \ in \ 2013 \ and \ 2014$

2.2 多年土地利用变化

以 2007—2013 年北京市土地利用现状数据为例,利用 Hadoop 并行运算环境对多年土地利用变化数据进行多级栅格化、图像视频合成等处理步骤,生动展示了北京市土地利用现状建设用地、农用地、未利用地三大类数据空间分布变化情况,具体如图 7 所示。2007—2013 年北京市土地利用现状变化视频中,用不同的色块代表不同的地类,当视频播放时,代表建设用地、农用地、未利用地的色块在不停地变换位置,直观的展现了全市多年间建设用地、农用地、未利用地的连续变化情况,从图 7 中可以看出,多年间建设用地持续增加的趋势与农用地、未利用地不同阶段的变化情况。

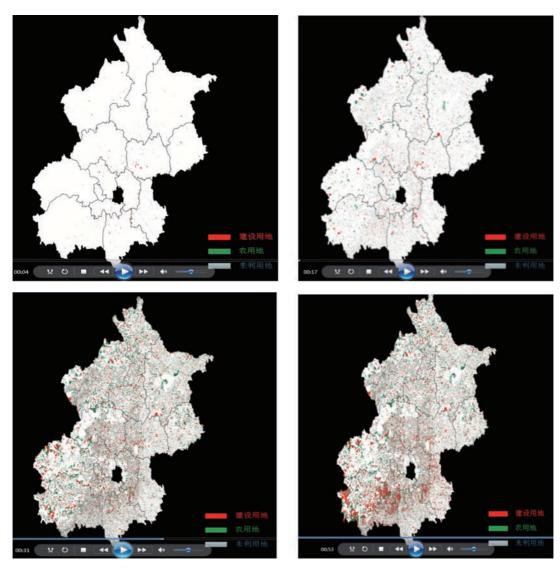


图 7 2007—2013 年北京市土地利用变化 Fig7 Video Map of Land Use Change in Beijing from 2007 to 2013

2020年8月—

2.3 土地利用程度三维地球立体展示

借助 Hadoop 并行运算环境,对经过网格化处理的空间统计数据进行多年度、多行政区快速叠加聚合,再通过计算网格土地利用面积变化比和网格土地利用程度,以三维地球为载体展现各个网格的用地情况^[7],具体公式为:

①网格土地利用面积变化比

网格土地面积变化比 =
$$\frac{\sum_{i=1}^{n} |A_{i}[y] - A_{i}[y-1]|}{\text{网格面积}}$$
 (1)

式(1)中,表示网格中第个地类在年的土地面积,是网格内地类数,网格面积是该网格的土地总面积。

②网格土地利用程度

提前设定好各地类的用地程度系数,具体地类的土地利用程度计算公式为:

网格土地用地程度 =
$$\frac{\sum_{i=1}^{n} (A_i[y] \times W_i)}{\text{网格面积}}$$
 (2)

式(2)中,表示网格中第个地类在年的土地面积,是网格内地类数,是第个地类的 用地程度权重,网格面积是本网格的土地总面积。其计算结果如表1所示。

表 1 不同类型土地用地程度分级指数表

Table 1 Classification index for different types of land use

类型	未利用土地级	林草水用地	农业级用地	城镇聚落用地级
土地利用类型	未利用土地或难利用 土地	林地、草地、水利用 地	耕地、园地、人工草 地和其他农用地	城镇、居民、工矿用 地、交通用地
分级指数	1	2	3	4

土地利用变化比和土地用地程度均可以三维地球为载体展示。这里以 2010 年全国 10 分网格的土地用地程度为例,按照以上技术路线进行数据处理分析,展示全国网格的 土地利用程度,具体如图 8 所示。

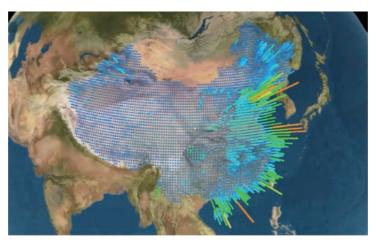


图 8 2010 年全国土地利用程度 3D 可视化 Fig.8 3D visualizations of Chinese land use levels in 2010

2010年全国 10 分网格的土地用地程度三维展示图中,通过柱状图来表示土地利用程度,柱子越高代表土地利用程度越大,柱子越密集代表土地开发程度越大。用户可以通过参数选择实时调整柱子高度,并可以进行单个立方柱的数值查询。

3 结论

该研究基于土地利用管理实际业务,探索了多种新的土地基础数据展示方式,证明可以将主流热门技术与现有土地基础数据相结合,满足大数据挖掘分析的直观可视化表达的需求。同时,该项目形成的可视化控件成果为下一步开展深入分析研究和技术扩展提供了良好的基础。Hadoop并行运算环境为进一步丰富了土地数据可视化分析展示手段准备了条件。但总体来看,目前土地利用变化可视化模型设计、组件研发还是初级的,下一步将扩展可视化展示模型,同时提升数据处理效率,进一步优化组件展示效果,实现基于海量数据的实时可视化展示。

参考文献

- [1] Anand Rajaraman, Jeffrey David Ullman. 大数据•互联网大规模数据挖掘与分布式处理. 人民邮电出版社, 2012(22):
- [2] 周宁, "信息可视化及应用"专题序. 现代图书情报技术, 2010(7/8): 1-2.
- [3] 杨峰,李蔚.信息可视化显示效果的优化方法研究.现代情报,2012,32(5):34-36,47.
- [4]任东怀, 胡俊. 多维数据可视化技术综述. 工程地质计算机应用, 2006, 44(4): 4-9, 34.
- [5] 李霖,许铭,尹章才,等.基于地图的地理信息可视化现状与发展.测绘工程,2006,15(5):11-14.
- [6] 周宁, 张李义. 信息资源可视化模型方法. 北京: 科学出版社, 2008: 244-250.
- [7] 刘波,徐学文.可视化分类方法对比研究.情报杂志,2008,27(2):28-30.

Discussion on visualization of land use change based on Hadoop framework

Zhang Dingxiang¹, An Qiong², Ji Hongwei¹, Wu Jianjia², Li Shijiang¹

(1.Chinese Land surveying and Planning Institute, Ministry of Land Resources, Beijing 100035, China; 2.Beijing Paramount Information Technology co.Ltd., Beijing 100029, China)

Abstract: [Purpose] This paper focuses on the characteristics of large amount of land use data and complex data model, studies the visual expression model of land use change data, and explores an intuitive and effective visualization method of land use status and change data. [Method] To meet the needs of time series display, space display and flow map display in the process of land use/change data analysis and display, three display methods, such as multi-year land use flow chart, multi-year pattern change video and three-dimensional stereoscopic map, are designed, relevant components, examples and API interfaces are developed,

2020年8月

and mass data rapid pre-processing is completed by using Hadoop parallel technology. [Result/Conclusion] In this paper, the interactive visualization of land use data is realized from the application data, processing flow, display effect, interaction, interface, multi-chart interaction and so on. It can provide technical support for the analysis and mining of massive land basic data in the future.

Key words: land use; basic data; visualization; big data

欢迎订阅《中国农业信息》

《中国农业信息》(双月刊)由农业农村部主管,中国农学会农业信息分会、中国农业科学院农业资源与农业区划研究所共同主办,是我国目前全方位传播和刊载国内外农业遥感/农业信息科学领域的信息获取、处理、分析和应用服务的理论、技术、系统集成、标准规范等方面最新进展和成果,促进学术交流以及农业信息学科关键技术与产品的创新研发、集成推广和应用示范的综合性科学技术期刊。

主要刊登农业遥感、农业传感器、农业信息智能处理、精准农业/智慧农业、农业监测预警与信息服务系统、农业物联网、智能装备与控制、虚拟农业、人工智能、信息技术标准等方向学科热点领域的最新、最重要的理论研究和应用成果。主要栏目有:农业遥感、智慧农业、综合研究、农业信息技术、农业物联网、专题报道等。目前被中国知网(CNKI)、万方数据、中文科技期刊数据库、中国核心期刊(遴选)数据库等多家数据库收录。

《中国农业信息》为国内外公开发行的刊物,开本为16开,彩色四封,读者范围广,影响面大,全国各地邮局均有订阅。每双月25号出版,定价为25.00元/册,150元/年。

邮局汇款

收款人:《中国农业信息》编辑部

地 址:北京市海淀区中关村南大街 12 号中国农科院资源所区划楼 319

邮 编: 100081

银行汇款

开户行:农行北京北下关支行

行 号: 103100005063

账 号: 11050601040011896

单位名称:中国农业科学院农业资源与农业区划研究所

电 话: (010) 82109632 传 真: (010) 82109632 E m a i l: nyxxbjb@caas.cn

邮发代号: 2-733

投稿网址: www.cjarrp.com