

面向统计数据的三位一体可视化关联分析

张金区^{1,2}, 王开泳², 王云鹏³

(1. 华南师范大学计算机学院, 广州 510631; 2. 中国科学院地理科学与资源研究所, 北京 100101;
3. 中国科学院广州地球化学研究所, 广州 510640)

摘要: 根据社会经济统计数据的特点, 借鉴地理信息系统的空间表达方法, 利用 Flash 技术提出属性、图形和图表三位一体可视化关联分析系统。设计行政区空间图形库和社会经济数据库, 以组件方式构建图表对组件库和模型库进行分析, 实现系统在线即时可视化表达与分析功能。应用结果表明, 该系统能提高对社会经济等相关数据的分析效率。

关键词: Flash 技术; 统计数据; 可视化关联分析; 地理信息系统

Visual Associated Analysis with Three Module in One View for Statistical Data

ZHANG Jin-qu^{1,2}, WANG Kai-yong², WANG Yun-peng³

(1. Computer School, South China Normal University, Guangzhou 510631, China;
2. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
3. Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China)

【Abstract】 According to the analysis to the characteristics of socio-economic data, referring to the information expressing method in Geographic Information System(GIS), this paper designs a visualization analysis system to the socio-economic data with attribute, graphics and charts associated style based on Flash technology. According to the requirement, administrative unit spatial graphics database, socio-economic database are designed and the chart component library and model library are constructed. Analysis result shows that the system can improve the efficiency in socio-economic and other relevant data analysis.

【Key words】 Flash technology; statistical data; visual associated analysis; Geographic Information System(GIS)

DOI: 10.3969/j.issn.1000-3428.2012.03.085

1 概述

社会经济数据是反映一个国家或地区国民经济和社会形势的重要指标数据, 每个国家或地区都会定期发布本地区最近的社会经济数据统计信息。如今, 各统计部门已经积累了非常丰富的社会经济数据, 对社会经济数据的信息发现和知识挖掘变得越来越重要, 不仅用于系统研究和对比分析地区差异, 而且为地方政府动态性的区域监测和政策调控以及未来的经济发展规划提供参考。对社会经济数据的深入分析还是实施国家宏观战略、制定区域发展政策的重要依据。如何从丰富的社会经济数据中快速挖掘信息和知识是当前经济快速发展背景下最紧迫的需求^[1]。

对研究人员来说, 要获取自己需要的统计数据, 一种方法是自己购买统计年鉴然后手工输入数字化, 另一种方法是通过网络途径获取已经数字化好的年鉴数据, 然后再运用地理信息系统(Geographic Information System, GIS)或其他统计软件进行分析整理。这个过程耗时费力, 难以适应快速发展的经济形势。从对社会经济数据的分析方法来看, 主要包括统计图表分析和空间分布专题分析 2 种类型。统计图表分析主要借助于统计软件进行数理统计, 空间分布专题分析主要借助于 GIS 软件将属性统计数据和行政区空间矢量图形进行关联, 然后根据属性字段生成专题图。这 2 种分析方法都是数据与软件紧密结合, 属于研究者的个体行为, 导致了不同研究人员都要重复安装软件、重复组织数据、重复分析数据^[2]。另外, 这 2 种软件方法都还缺少属性、图形和图表同步动态

可视化关联的分析功能, 在可视化效果上不能给研究人员以整体动态的表达和思考。

鉴于以上分析, 本文采用 Web 的方式, 构建基于 Flash 技术的社会经济数据属性、图形和图表三位一体的动态可视化关联分析系统。通过该系统能够方便研究人员对社会经济数据的在线发布、获取、组织和可视化分析, 并为地学信息化科研环境的构建奠定数据在线分析的技术基础^[3]。

2 社会经济数据特点分析

社会经济数据具有以下特点:

(1) 以各级行政区为基本统计单元。这个特点使得社会经济数据组织较为规范, 内容便于查找。

(2) 时间序列性好。由于社会经济数据定期每年出版, 形成了指标数据的良好序列性。

(3) 统计类型和指标多样。既有统计数据, 又有抽样调查数据, 并且每年都有新变化。

(4) 仅以列表形式出现, 枯燥单调。没有图形数据, 让使用者难以形成空间分布的概念。

基金项目: 国家自然科学基金资助项目(40901088); 国家“863”计划基金资助项目(2006AA06A306); 中国博士后科学基金资助面上项目(20090460505)

作者简介: 张金区(1980—), 男, 副教授、博士, 主研方向: 空间信息可视化; 王开泳, 助理研究员、博士后; 王云鹏, 研究员、博士

收稿日期: 2011-08-15 E-mail: zjq@scnu.edu.cn

统计数据有统计指标(体系)、时间、空间等属性,可以概况为统计数据的指标维、时间维、空间维 3 个维度^[4]。总体上看,社会经济数据具有一定的规范性和历史继承性,但它同时又具有复杂性的特征^[5]。复杂性主要表现在统计指标的年度变化,使得在研究分析过程中很难设计统一的表格结构进行管理,这个特点给社会经济数据的统一组织和有序管理提出了挑战。同时,这些特点也使得对社会经济数据的信息挖掘变的复杂,难以快速挖掘和提取有效信息。

随着地理信息系统和其他统计分析软件的出现,使得社会经济数据的信息挖掘和可视化表达越来越容易实现,提高了数据使用的直观性,不断改善数据的可视化表达方法。但已有的研究基本上以单体研究为主,在数据的获得、输入、分析、处理环节都需要亲自动手操作。针对社会经济数据的使用难题,本文立足于社会经济数据的快捷的可视化表达和方便的网络共享,开发了基于 Flash 的社会经济数据属性、图形和图表三位一体可视化分析系统,完善和提升社会经济数据的可视化表达方法和途径,丰富了数字地图的内涵和表现形式,实现了社会经济数据的深度挖掘。

3 系统框架设计

3.1 技术分析

根据社会经济数据的特点,要实现友好的可视化关联分析,需要:(1)设计良好数据库表结构,便于数据的检索和提取;(2)选择优秀的可视化表达分析技术,便于信息的多样化展示;(3)确定网络功能强大的技术方案,便于功能拓展和网络发布。从方便可视化表达和满足对网络支撑的强大需求这 2 个角度来考虑,Flash 技术无疑是最佳的选择。Flash 本身是制作网络动画的软件,即具有超强的图形可视化渲染能力,同时本身就是为支持网络而诞生。随着丰富互联网应用程序(Rich Internet Applications, RIA)的兴起,Flash/Flex RIA 技术框架已经成为互联网应用程序开发的主流^[6-7]。尤其是在 Flash ActionScript3.0 脚本编程语言推出后,更是大大增强了其网络应用功能的支持^[8]。Adobe Flex SDK 本身还提供了丰富的图表控件,可以直接应用于对二维表格属性数据的分析,简化了 Flash 编程的工作量。

为使社会经济数据属性、图形和图表三位一体可视化关联分析系统具有灵活性和通用性,而不是像一般的可视化分析系统那样只是适用于特定的数据库。本文以 Flash 技术为基础,集成数据库技术和数据库访问技术,将整个系统划分为 4 个模块,分别为行政区空间图形数据库、社会经济数据库、图表分析组件库和模型库。各模块之间采用事件触发的方式进行关联,以达到数据同步分析的目的。系统各功能模块之间的关系如图 1 所示。

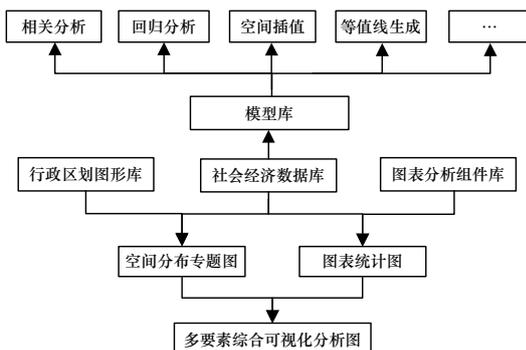


图 1 系统功能模块

3.2 社会经济数据库

社会经济数据库是各功能模块的核心,其他功能模块的展示都要直接或者间接依托社会经济数据库中的数据来完成。根据前面描述的社会经济数据的特点,为了更好地适用于不同的数据库结构和不同的数据库服务器,本文设计的社会经济数据库主要在已有的数据库基础上添加 2 个表:(1)主题分类表,用于对属性表进行分类,便于按主题组快速检索;(2)表格索引表,用于详细记录每一个属性表的元数据信息,便于用户精确检索。

对数据库的访问主要涉及数据的读取和查询,常用 ASP、JSP 或 PHP 技术均能实现,本系统采用了 ASP 技术。由于 Flash 本身具有处理 XML 的功能类包,为了便于将 ASP 从数据库中读取的社会经济数据传递到 Flash 中,并进一步和行政区空间图形关联以及做相应的图表分析处理,将 ASP 读取的数据最终以 XML 的格式进行输出。

3.3 行政区空间图形数据库

Flash 本身是制作矢量动画的软件,本身具有良好的空间图形支持功能,在构建矢量的 WebGIS 系统方面有其独特优势。行政区空间图形数据主要来自地球系统科学数据共享网提供的基础地理数据,包含全国 33 个省级行政单元的、340 个地级市和 2 400 个县的行政边界。原始数据格式为 ESRI Shape 文件,为了实现基于 Flash 的空间图形表达,必须将 ESRI Shape 文件转换为 Flash 的 SWF 文件,供网络直接调用和表达。为此,本文在参考 Adobe 公司提供的 Flex SDK 开源框架的基础上,开发了 Shape 文件向 SWF 文件的转换工具,该工具已上传至相关网站(<http://www.gdrsgis.cn/zjqweb/index.asp>),可以供用户免费下载。

在转化成的 Flash SWF 矢量图形文件中,都完整保留了原先 Shape 文件所具有的属性信息,其中每一个图斑都具有地区名称和地区代码 2 个属性值,以便于和社会经济数据库中的属性数据相关联。建立行政区空间图形数据库的目的在于实现社会经济数据的空间化处理和空间分布专题图制作。行政区空间图形库以文件的形式进行组织和管理,每一个文件都用该地区的行政代码进行命名。

3.4 图表分析组件库

图表分析组件库主要用于对从社会经济数据库中提取的数据进行图表分析,图表类型包括柱状图、折线图、散点图、条形图、饼状图、面积图等。在 Adobe 提供的开源 Flex SDK 中本身已经包含了这些组件,但要实现这些图表控件的自由调用和展示,还需要结合 Flash Actioscript3.0 进行编程实现。

3.5 模型库

模型库的建设更多的需要专业用户的参与,为了给用户更大的自由度,本系统在设计模型库时,只是将通用的相关分析模型、回归模型、反距离加权倒数插值模型、等值线生成模型等集成到系统中,其他的模型则提供模型设计接口供用户自行设计,即时计算。

上述 4 个模块之间的协调和关联由 Flash 脚本编程语言 ActionScript 3.0 编写的虚拟控制与计算中心统一处理,并将处理结果直接以数据、图表和地图一体化的方式展示给用户。

4 系统实现及应用

根据上文设计的系统框架,在应用过程中采用“模型-视图-控制中心”的组织架构,模型和控制中心主要采用 Actionscript3.0 进行编程实现。视图则以网页嵌入 Flash SWF 对象的方式,SWF 视图对象根据用户的操作,向社会经济数

数据库和空间图形库发送请求，并将结果以属性数据、图表和地图的形式进行关联展示。社会经济统计数据可视化分析流程如图 2 所示。

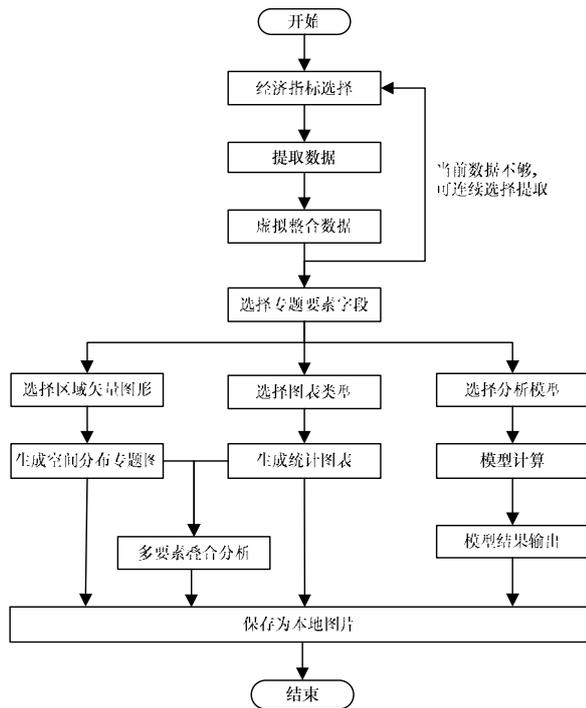


图 2 社会经济数据可视化分析流程

为加强系统应用的灵活性，在实现过程中采用配置文件的方式，即功能开发完成后，只需要通过修改配置文件，即可实现不同功能的应用组合。系统配置信息则采用 XML 文件的方式进行存储。

通过编写 XML 文件对系统初始化配置完成后，即可放到服务器上应用发布。目前，属性、图形和图表三位一体可视化关联分析系统已成功应用到中科院 E-GeoScience 信息化科研环境建设和中国区域发展经济监测指标和评估指数的可视化表达分析中，其展示效果如图 3 所示。



图 3 系统可视化关联展示效果

5 系统特点

本文设计的属性、图形和图表三位一体可视化关联分析系统具有以下特点：

(1)可快速获得某一区域某一经济要素的空间分布图。

实时在线可视化系统的开发与应用，实现了社会经济数据的快速获取可视化展示，用户通过选择区域、时间和相应指标，即可实时生成区域空间分布图以及柱状图、曲线图等统计图表。这样既节省了收集整理数据的工作量，又避免了研究者对地理信息系统的复杂操作和分析，方便了政府部门与研究人员对特定区域的分析与把握。

(2)可方便获得某一区域某个发展要素的动态变化图。

掌握某个区域或城市的时空发展演化是准确把握和预测该地区未来发展态势的必要条件。本文所构建的在线可视化系统，只需要选择所需的时间段和要表达的要素，就可以在线生成不同年份的空间分布图，并可以连续播放进行对比分析，较直观地满足了研究者和政府决策部门对发展演化态势的把握。系统不仅可以实现单个地区的空间图表表达，还可以进行不同区域的对比和分析。

(3)可快速实现同一经济数据的多样化展示。

基于 Flash 的社会经济数据属性、图形和图表三位一体可视化关联分析系统较好地实现了即时性的查询与数据图表显示。具有灵活多样的可视化手段，如运用本系统开发的区域监测评价系统，用户可以自行指定空间分布专题图的图例颜色，可以根据需求选择折线图、饼状图、柱状图、条形图等图表样式。有助于使用者从多个角度对经济发展进行动态把握和空间分析决策。

(4)可自定义区域发展评价指数体系。

为综合地评估某个地区的发展概况，可以在数据库中挑选指标建立评价指标体系，由系统自动计算指数。在创建评价指标时，既可以直接用原始指标，也可以选择原始指标经过运算得到评价指标，或通过为每个评价指标赋予权重进行计算(权重之和必须为 1)。选择要计算的时间、地区和指数后，系统会自动从监测数据库中调出相应原始数据得到评价指标，标准化后代入综合评价模型得到需要的指数结果，具有较强的智能性和便捷性。

6 结束语

地理信息技术的数据获取、存储、空间分析和信息查询为地球信息的模拟、分析和预测奠定了基础^[9]。地理空间数据的可视化在一定意义上，可以看作数字时代的地图学^[10]。本文设计了用于社会经济数据、图表和地图一体化表达的分析系统，并将其应用于中国区域发展监测指标和评估指数的可视化表达分析上，实现了基于地理空间位置的中国分省、分地区、分县以及重点区域的国民经济、人口、就业、教育、科技、卫生等区域发展指标和经济、社会发展指数、创新能力指数和区域综合发展指数的快速多样化表达，为研究分析中国区域发展的空间分异规律和动态变化趋势提供了较为直观和便捷的技术手段。

目前，空间数据挖掘已成为国际研究的热门课题^[11]，利用 GIS 的可视化技术对统计数据进行可视化表达，是进行统计分析、数据挖掘的重要方法。下一步将对高分辨率可视化图表的生成和调用、可视化方式的多样化与数据适宜性方面进行研究。