

地理信息系统在传染病疫情分析中的应用

谢学勤

(北京市公共卫生信息中心 北京 100050)

〔摘要〕 简要介绍地理信息系统 (GIS) 基本概念及其在流行病学研究中的应用, 以北京市传染病信息系统为例, 具体阐述系统设计、主要功能及存在问题, 建立传染病流行的 GIS 数据平台, 在此基础上充分利用 GIS 系统为有效控制传染病提供决策依据。

〔关键词〕 地理信息系统; 传染病; 疫情分析

The Application of Geographic Information System for Epidemic Situation Analysis of Infectious Diseases XIE Xue-qing, Beijing Public Health Information Center, Beijing 100050, China

〔Abstract〕 The paper briefly introduces the basic concepts of geographic information system (GIS) as well as its application in epidemiology research, taking Beijing infectious disease information system as example, it concretely describes system design, main functions as well as the existing problems, constructs GIS data platform for infectious diseases. Based on this platform, it fully utilize GIS to provide decision-making support for controlling infectious diseases effectively.

〔Keywords〕 Geographic information system (GIS); Infections diseases; Epidemic situation analysis

1 引言

1.1 地理信息系统基本概念

地理信息系统 (Geographic Information System, GIS) 是以地理空间数据库为基础, 在计算机软硬件的支持下, 对空间相关数据进行采集、管理、操作、分析、模拟和显示, 并采用地理模型分析方法, 适时提供多种空间和动态的地理信息, 为地理研究和地理决策服务而建立起来的计算机技术系统。GIS 区别于一般信息系统的关键在于其处理空间相关数据的能力, 而大量医疗卫生信息, 如传染

病的发生与流行、地方病的分布及病因、医疗卫生机构的分布等都与空间信息密切相关, 这种空间相关特点成为在公共卫生领域应用 GIS 的前提。20 世纪 60 年代兴起的遥感技术 (Remote Sensing, RS) 是指在高空和外层空间的各种平台上, 运用各种传感器 (如摄影仪、扫描仪和雷达等) 获取地表的信息, 通过数据的传输和处理, 从而研究地面物体形状、大小、位置、性质及其环境的相互关系的一门现代化应用技术科学^[2]。随着计算机技术和空间技术的进步, 卫星遥感遥测技术已经进入一个能够快速、及时提供多种对地观测数据的新阶段。GIS、RS 和全球定位系统相结合, 简称 “3S 技术”, 可以实现对地面环境的综合分析和动态监测。利用遥感观测数据及时更新 GIS 中的数据, 可以建立动态 (或四维) 的 GIS; 另一方面, GIS 中的数据也可以辅助遥感图像的判断与分类。

〔收稿日期〕 2011-02-15

〔作者简介〕 谢学勤, 北京市公共卫生信息中心副主任, 主要研究方向为流行病学、卫生统计等, 发表论文多篇。

1.2 地理信息系统在流行病学中的应用

流行病学研究的重要内容之一是研究疾病的分布,以确定高危人群和高发地区,以及相关的危险因素。近年来随着地理信息系统迅速发展和广泛使用,为流行病学研究提供了新的机遇和挑战,目前GIS系统已成为很多公共卫生和流行病学项目的一个重要组成部分。分析传染病发生、发展中所涉及的诸多影响因素,建立传染病流行的GIS数据平台,在此基础上充分利用GIS系统为研究和决策提供依据。到目前为止,地理信息系统与遥感遥测技术在流行病学中的应用主要包括:在媒介传播疾病的流行病学研究中,如果明确了与疾病传播和媒介生物有关的环境因素,并且这些环境因素能够通过遥感资料获得,那么应用地理信息系统、遥感遥测技术和全球定位系统就能为这些媒介传播疾病的研究、监测和控制提供有力的工具。本研究采纳的主要技术已经在国内外得到广泛应用,城市传染病可视化分析及决策支持系统,以计算机网络技术为支撑,疫情数据库为基础,地理信息为平台,融合完善的区域经济统计、分析信息,建立和执行规范化的疫情信息分类标准和统一的地理空间拓扑关系,对各项疫情基础空间地理数据进行科学的存储和管理,实现快速采集、建库、查询、检索、显示、输出、更新、统计、分析及信息的网上发布和资源共享,为国民经济的良好发展保驾护航。

1.3 研究目的

本文利用案例研究,通过采用地理信息系统技术,对疫情数据进行空间表达,探讨传染病病例的空间分布,有助于有目标、有针对性地采取正确的策略,以控制疫情的继续蔓延。同时,该系统充分利用了GIS的空间分析功能,从时间和空间上对确诊病例/疑似病例进行查询分析,按照不同指标和空间位置进行分类汇总和统计分析。通过图文一体化的表现方式对潜在的传染区进行预测,辅助有关部门制定相应的预防策略。研究目标在于通过收集和提供及时、详细、准确疾病的检疫、防治、监控、评估、治疗等综合信息,为社会各部门、单位

实施区域经济的宏观管理和调控提供权威依据,提高领导层管理和决策的科学水平与调控能力。

2 系统设计

2.1 系统结构

北京市传染病信息系统包括:传染病疫情报送子系统、流行病调查子系统和GIS分析决策支持子系统,3个子系统共享统一的数据库。该系统是一个包括数据、图形、图像,集数据处理和传输为一体的综合数字业务网,旨在建立一个以数字化电子地图作为基础的疫情地图数据库,通过在先进成熟的地理信息系统(MapInfo)平台上开发的疫情分析应用软件,建立地图数据与疫情和流调数据库的关联,将疫情数据与其空间位置相联系,实现疫情数据的地图化、可视化,结合地图数据进行疫情数据的检索、查询、分析及网络发布。

2.2 系统平台

本研究在空间数据库管理上基于Oracle Spatial的方式实现了跨平台的地理信息数据管理。在位置信息服务引擎上采用了基于J2EE技术的MapXtreme for Java,在地理编码搜索上采用MapSearch这一完全适应国内特殊需求的地址引擎,企业级中间件采用Oracle9iAS或Weblogic。

2.3 多进程多线程的体系结构

系统采用多线程多进程体系结构,多个数据库请求(地图数据、属性数据)可以由一个服务进程处理。同时,系统根据当前的负载情况动态分派服务进程的数量和使用。从而充分利用多用户操作系统“多用户多任务”的工作方式,并有效地减少系统的资源消耗,达到最大的处理速度及吞吐量。

2.4 可伸缩的体系结构

系统具有良好的可伸缩性,从客户端应用到企业级应用,都可以在相同的体系结构上支持并有较高的性能价格比。同时,无论是单CPU的计算机,还是多CPU的计算机,系统均有成熟的技术支持,

这是由于系统具有先进的体系结构。地理信息平台体系结构为在线地图化应用系统提供的可伸缩性,能够支持大用户量以及大事务量的工作负荷。通过将分布式任务分布到多个处理器或机器上,譬如在一个集群环境中,可以大大提高单个事务的响应速度和整个系统的吞吐能力。提供的一种自动的、动态的自调整能力甚至能够跨越多种硬件和操作系统平台对处理负荷进行均衡处理。

2.5 数据库行级锁技术

在多用户同时工作的时候,操作竞争及资源竞争经常会成为影响系统运行性能的一个关键因素。地理信息服务平台在支持多个用户并发工作的时候,利用了数据库行级封锁、读一致快照等技术,在最少的系统资源占用的前提下,能达到系统的最大吞吐量,并达到最佳的处理性能。

2.6 融入了适合中国国情的地理编码技术

在数据搜集、应急系统的实际应用过程中,用户经常需要在已知地址的情况下,在电子地图中找出该地址对应的相应位置(这个过程就称为地址编码)。中国大多城市规划比较乱,地址填写也不规则,因此,目前在国内并没有非常完善的解决办法。为了更好、更快地回复用户位置信息的服务需求,Mapsearch 的出现解决了适合中国国情的地理编码应用难点,基本功能是将输入的“门牌、地址、单位字符串”更快、更准确地输出为代表其位置的经纬度值,并在地图上进行定位显示。

3 主要功能

3.1 GIS 技术在流行病防治和决策支持系统中的主要应用

- 疫情空间分布: 基于空间位置直观显示各种疫情分布情况。

- 疫情空间查询与专题分析: 根据病人的医院、地址、报告日期等综合查询条件,组合各种疫情专题类别,利用多种空间查询模式,实时展现全市各地区的多种疫情分布。

- 专题类别: 以各级行政区划为单位来展现各种疫情信息的分布和进行空间统计;可以精确到街道等方面的空间展现和统计。

- 查询条件: 可以按地点、医院、单位、住址等进行查询。

- 疫情趋势分析: 动态渲染地图,按时间段显示各级行政区的疫情变化情况。

3.2 区域疫情

按 19 个区或所选区的街道办事处在地图上以颜色渐变对人数变化情况进行渲染(色彩愈深,人数愈多,疫情越严重),可以使用“点击查询”工具查询各区、街道疫情详细信息,并统计出北京市总体情况或所选区县的疫情状况,同时按职业类型进行分类渲染,进行区域疫情分析时可以选择最新上报时间、全部时间、前 1 天、后 1 天进行疫情统计分析(可按发病数、发病率进行统计),见图 1。

3.3 病例分布

在地图上定位显示病例不同状态的分布情况;可以按区、职业类型进行区分定位显示病例分布情况;也可以按时间选择,如最新上报时间、前 1 天、后 1 天、全部时间定位显示病例分布情况。可以将结果在远程打印输出。

3.4 病例转归分布

在北京市地图上定位显示全部病例的转归情况,包括病例痊愈、病例死亡的分布情况,同时亦可以按职业类型、时间定位显示病例转归分布情况。可以将病例转归分布查询结果在远程浏览器所在终端打印输出。

3.5 病例聚集密度分析

系统能将不同聚集密度的行政区疫情情况及附近标志性建筑列出,用户使用鼠标点击附近标志性建筑,系统将对其高亮定位。系统缺省使用 900 米网格进行分析,也可在查询栏上部选按 300 米网格进行分析。

3.6 名称或地址定位查询

• 名称定位查询: 选择疾病控制中心/疾病控制医院, 输入查询信息 (如胸科), 点击开始查询。系统将包含有输入信息 (如胸科) 的所有疾病控制中心/疾病控制医院列出, 然后选择某项 (如北京市胸科医院) 后, 系统将对其高亮定位并列出其详细信息。

• 地址定位查询: 在地址输入处输入地址信息 (如西三环北路 27 号), 在地图上对病人、医院、

控制中心进行定位, 输出时与名称定位相似。

3.7 传染链查询

由于该系统实现了疫情、流行病调查合一, 疫情报送子系统、流行病调查子系统和 GIS 分析决策支持子系统有统一的共享数据库。因此流调中发现的传播链在地理信息系统中得到了很好的空间显示, 以便有关部门有针对性地在该区域内做好预防和控制工作, 防止疫情进一步扩散。

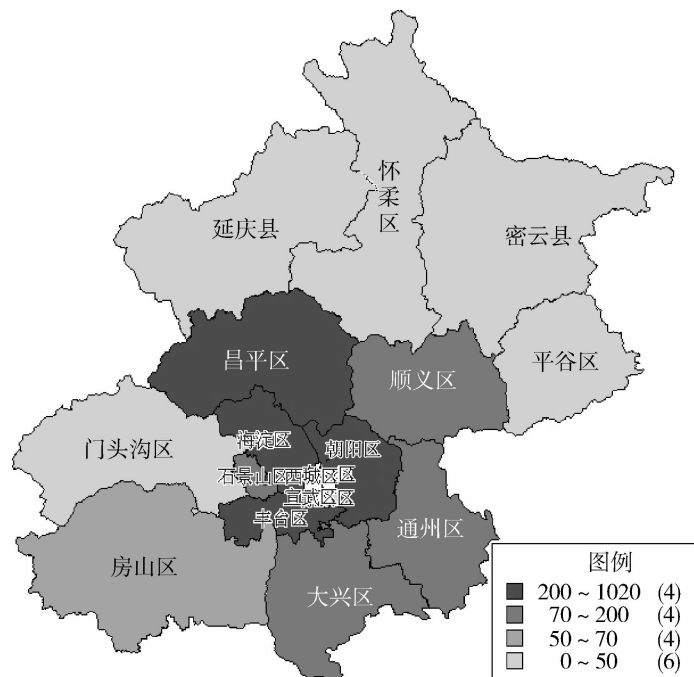


图 1 北京市 2005 年麻疹区县发病数

4 存在问题

4.1 病人地址匹配问题

由于地理编码的不规范, 以及各单位报送上来的病人地址不够详细, 为地理信息系统的地址匹配定位工作带来难度。

4.2 病人信息的频繁变化对系统运行构成影响

由于许多病人先后报来的信息 (地址、年龄、单位等) 不一致, 造成基础数据不停变化。尤其是病人地址和单位的变化, 使得原先的地址匹配工作

失效, 造成对该病人的定位错误。鉴于这种情况, 专有数据处理人员每天定时整理出这类信息发生变化的患者信息, 进行新的地址匹配工作, 以防发生定位错误。

4.3 密切接触者的数据采集加大工作强度

流调信息也需要进行数据采集, 获得这些人群密切接触者的详细地址, 这样就加大了流调人员和地理信息系统的数据处理人员的工作量, 并增强了工作难度。同时这也存在地址匹配精确度问题。

(下转第 21 页)