

Research on Natural Focal Diseases Based on 3S/DM Integration

Sun Jiuyun, Qiu Yinguo

Dept. school of environment and science and spatial informatics, china university of mining and technology, Xuzhou, China
wdjysun@tom.com

Abstract: Natural focal disease is a kind of epidemic that human being and beasts (or birds) sicken together and infected mutual. The disease even if didn't epidemic in human at one period of time, but it can exists in nature by infection among beasts (or birds), and may arose in human being when people neglect. Because of the fearful fatalness, now days it gradually arouse people pay attention to this kind of diseases. The disease has strong zone characteristic, and with human being activity, resource and environment exists correlation. Previously because of the ability shortage in data acquisition, management and spatial analysis, that restricted people to open out the interactional inherence mechanism and disciplinarian. Along with the development of 3S(GIS, GPS and RS) and spatial data mining, their powerful profundity analysis function fetched a chance to research distributing rules and developmental trends. This paper take 3S as data collection, administrator and visualization plat form, and using spatial data mining as analytical method, to discussed the construction of decision supported system in epidemic prevention. Based on the data of mainly natural focal diseases in Ji-angsu province, bring forth a new geographical cellular automata based on irregular plynon network. Then using the model to forecast the development of nature focus, and it was validated by compared with actual information. Thereby an effective scheme of decision supported system in nature focal diseases prevention was explored.

Keywords: natural focus; natural focal diseases; 3S integration; data mining; cellular automata

基于 3S/DM 集成的自然疫源性疾病研究

孙久运, 邱银国

中国矿业大学环境与测绘学院, 徐州, 中国, 221116
wdjysun@tom.com

摘要: 自然疫源性疾病是一种重要疫灾, 其产生、发展、变异、流行、蔓延与人类活动和所处环境密切相关。虽然这已早为医学地理学家所注意, 但之前由于受分析手段、方法与信息量的限制, 并没有取得显著的研究成果。近年来发展起来的数据挖掘与 3S 集成在数据深层次分析与空间特性分析方面有着强大的功能, 对于深入研究疫源的分布传播规律、自然疫源性疾病与人类活动之间作用带来了契机。本文探讨了 3S/DM 集成在自然疫源性疾病研究中作用、方法与措施, 提出了应用的三个层次, 并以江苏省主要疫源地为例, 应用所提出的基于不规则多边形的地理元胞自动机对疫源地的发展进行预测, 从而探索出一条行之有效的自然疫源性疾病防治的决策支持系统建设方案。

关键词: 自然疫源地; 自然疫源性疾病; 3S 集成; 数据挖掘; 元胞自动机

引言

地球上自然和人类活动诱发的自然变异, 孕育于地表环境中, 依时渐进, 当变异给人类

社会带来危害时, 即构成自然灾害。与地质、气象等自然灾害相比, 对人类影响最持久、危害最严重的是“疫灾”。疫灾是瘟疫流行所致的灾害, 历史上对人类健康和生命构成极大的危害。抗生素与西方医学的发展使人类一度认为

资助信息: 中国矿业大学科技基金资助项目(2006A031)。

可以远离疫灾的危害,但近年来 SARS、禽流感和猪流感的流行,使人类重新认识到了其极大的危害性。从灾害影响看,疫灾是人类灾害链网中的顶级灾害。地球表层系统存在一张对人类生存与发展构成危害的灾害链网,疫灾处在这张灾害链网的纲位,直接威胁着处于生物链网顶端的人类的健康与生命安全。从灾害历史看,疫灾是与人类共始终的永恒灾害。生命具有不断进化的特质,病原体也是如此。人类好不容易征服一种疫病,可随时又会产生新的更难征服的疫病。病原体与人类同进化,疫灾与人类相始终^[1]。生存于自然界的人类不可能完全脱离疾病的危害,在提高人类战胜疾病能力的同时,通过人与自然的和谐发展,降低负面因素的影响,监测、减少或限制疫灾的发生与影响,从源头上保护人类健康是关键所在。

自然疫源地或自然疫源性疾病属于疫灾中最典型的一类,它是潜在于自然生态环境中的隐患,是一种“特殊生态系统”,这种“特殊生态系统”与人类活动、资源和环境密切相关^[2]。因此,揭示它们之间相互影响的内在机制与规律对于疫灾防治至关重要,但是早期数据获取、管理和空间分析能力的不足,制约着深入研究的展开。随着信息技术,尤其以 3S(GIS: Geography Information System, GPS: Globe Position System, and RS: Remote Sensing 统称为 3S)为代表的空间信息技术,和数据挖掘(DM: Data Mining)为代表的数据库分析技术的发展,为自然疫源性疾病的研究提供了契机。本文对 3S 和 DM 在自然疫源性疾病研究中的应用方法进行探讨,并以江苏省肾综合症出血热为例,应用地理元胞自动机对疫源地发展趋势进行预测,探索 3S/DM 集成在自然疫源性疾病中的研究。

2 基于空间分析的自然疫源性疾病研究概况

自然疫源性疾病是人畜(或禽)共患且相互传播、区域性、季节性明显,即使某一段时间内人间不流行,也可能通过动物间的传染在自然界中循环延续,在人类疏于防范、环境允许的情况下,造成人间流行的一种传染病,如大灾之后常有大疫是自然环境变化导致疫源性疾病集中爆发的体现。自然疫源性疾病的产生、

发展、变异、流行、蔓延与人类活动和所处环境存在着密切的关系。研究自然疫源性疾病的产生机理、与周围环境的关系、传播路线、人类活动造成的影响、长期潜伏的可能性、分布规律等一系列相关问题对这种疾病防治非常重要。

关于疫灾的记录中国已有 3000 多年的历史,欧洲也对中世纪欧洲流行的烈性传染病——鼠疫,有着详尽的记录。但是,这些记录还只是停留在资料收集和管理上,所做的分析有限。最早有记录的把空间分析手段引入疫灾研究的是 19 世纪对伦敦霍乱分布与水井关系的分析。在 20 世纪 80 年代 GIS 出现后美国对基于 GIS 的流行病研究做了大量工作。我国虽然利用空间技术对自然疫病的研究开展的较晚,但在 2000 年前后随着 GIS 技术的深入应用,基于空间技术的自然疫源性疾病的研究获得快速发展。军事医学科学院微生物流行病学研究所的韩光红开展了我国重要疫源地地理信息系统的建立与应用^[3];曹务春、方立群等探讨了 3S 技术在自然疫源性疾病中的应用及发展前景^[4];虞以新提出统筹区域发展中不应忽视自然疫源地与自然疫源性疾病的观点;江苏省血吸虫病研究所的周晓农等对全国血吸虫病流行状况的地理信息系统空间分析应用,和基于遥感的主要媒介钉螺分布状况进行了研究^[5]。

虽然目前国内外研究已经取得了显著的研究成果,但是随着世界范围内人类活动范围的扩大、交流的密切和自然资源的开发,自然疫源性疾病的发展近年呈现出加快的趋势。SARS、禽流感、猪流感给人类带来了极大的危害和恐慌,但是人类对其自然疫源性尚不明确,表现出了自然疫源性疾病研究亟待深入的研究。

3 3S/DM 集成及其在自然疫源性疾病研究中的应用

3.1 3S 集成

3S 中 RS 侧重于对地观测,获取地表环境信息;GPS 侧重于定位,实现全球范围内的快速、高精度位置测定;GIS 则是对空间信息的管理、存储与分析工具。各技术都有其特长和不足,3S 结合应用、取长补短是发展趋势,三者之间的相互作用形成“一个大脑,两只眼睛”的框架,即 RS 和 GPS 向 GIS 提供或更新区域信

息以及空间定位, GIS 进行相应的空间分析, 以从 RS 和 GPS 提供海量数据中提取有用信息, 并进行综合集成, 使之成为决策的科学依据。

3.2 3S/DM 集成

数据挖掘 (Data Mining) 是从存放在数据库、数据仓库或其他信息库中的海量数据中获取有效的、新颖的、潜在有用的、最终可理解的模式、规律的非平凡过程。一般的查询、检索、统计分析并不能称为数据挖掘, 数据挖掘是深层次的数据分析。近年来数据挖掘的研究与应用已从属性数据的挖掘延伸到对空间数据的挖掘, 从关系型和事务型数据库扩展到空间数据库。对空间数据所进行的数据挖掘称为空间数据挖掘 (SDM: Spatial Data Mining), 它是数据挖掘在空间信息中的延伸与应用。

数据挖掘技术作为数据分析的一种, 不是孤立的, 必然要与其它的工具或系统发生联系, Forrester 和 Gartner 对此有形象的描述: “Stand-alone data mining is dead. The demise ‘stand alone’ data mining.”^[6]。另一方面 3S, 尤其 GIS 对所管理的海量数据空间分析能力不足, 限制了 3S 集成的高效应用, 需要结合数据挖掘提高其空间分析功能。3S/DM 集成可以达到倍增效应, 其意义可以概括为: 首先体现在提高 GIS 空间分析和 RS 影像解译能力上, 发掘数据潜力, 增强空间分析功能, 使有限数据变为丰富的知识; 其次 DM 在 3S 平台上的应用丰富了空间数据库, 使原来难以关联的、分散的数据联合分析成为可能; 第三, 3S 为 DM 提供了稳定的数据挖掘平台, 使算法得以实现、表达方式更加多样。

3.3 3S/DM 集成在自然疫源性疾病研究中的应用

3S/DM 集成在自然疫源性疾病研究中的应用可以概括为: 监测、管理、分析、跟踪。

3.3.1 3S 技术应用于疫情监测

GPS 可以准确定位发病地点、疾病蔓延路线、调查点的海拔高度及调查区域范围, 为后续分析提供精确的空间位置坐标信息。通过对地观测所获取的 RS 影像解译可以直接或间接确定媒介变化情况, 从而达到监测目的, 比如通过遥感监测水质情况确定钉螺、按蚊分布与

数量, 从而监测血吸虫、疟疾的可能爆发区域; RS 还可以长期的监测地表环境变化信息, 为疫源性疾病相关性分析提供详实的数据。应用 GIS 的数据累积, 可以在时间序列、空间分布上监测疫情发展情况。

3.3.2 3S 技术应用于疫情管理

自然疫源性疾病种类多、媒介多、受周围环境影响复杂、发展情况各异, 要成功分析必须建立在对多源数据的有效管理上。比如需要建立疾病库、媒介库、环境库、影像库, 和空间信息基础设施, 以及为进行数据挖掘建立数据仓库。并利用 GIS 建立疫源地地理信息系统、媒介地理信息系统等, 医学地图编制、发布等。

3.3.3 3S/DM 集成应用于疫情分析

3S/DM 集成应用于疫情分析分为三个层次。第一层次是依靠 GIS 的空间分析功能实现邻近分析、叠置分析、缓冲区分析、地理统计等, 确定影响范围、相关影响因素、高危暴露人群、量统计等。第二层次是由于自然疫源性疾病影响因素复杂, 要获得有效分析成果必须集成 DM, 实现聚类分析、典型相关分析、预测分析和构建流行病模型等。第三个层次是揭示流行病和自然环境、人类活动相互影响的内在机制与规律, 达到防疫决策支持的要求。

4 基于 Polygon 元胞自动机的疫源地发展趋势预测

自然疫源性疾病的传播和扩散在时间上具有一定的连续性, 在空间上靠近疫源地的区域受到疫情影响的可能性比较大。所以说它的扩散是一个复杂的空间动力学问题, 是一个具有时空特性的空间复杂系统。元胞自动机是一个天然的空间动力学模型, 在时间上元胞自动机是一个离散的无限集, 它能够模拟系统的动态行为过程; 在空间上, 元胞自动机建立在元胞空间上, 元胞空间可以看作是对现实空间的离散化划分。因此, 元胞自动机用于空间复杂系统研究具有其内在的优越性。另外, 元胞自动机自上而下的构模方式、强大的复杂性计算能力和灵活性, 也使其应用在疫情模拟方面具有很大的可能。本文以 2000 年以来江苏省省肾综合症出血热为例, 应用元胞自动机对其发展进行模拟预测。

4.1 地理元胞改进

数据以江苏的各县域为单元，因此用县域作为自动机中的“元胞”，利用各个县区域邻接长度来确定影响疫情传播的权重。使用县域作为地理元胞自动机中元胞，改变了元胞自动机元胞的基本构型，因其基本构型都是规则的，如四边形、三角形或者六边形。本文把这种非规则的构型称为 polygon 元胞，元胞精度是与行政区划的详细程度相关。这样做的原因有二：其一，研究数据是以各个县域为单元的形式给出，如强行划分成格网形式，在县域内的格网的状态值就只能人为的赋予一种值，无意义；其二，所进行的研究是基于统计信息，划分成格网后对应于格网的数据在获取上存在很大的困难。

4.2 元胞空间及元胞邻居的定义

对于 Polygon 元胞，其大小不是定值，而是和空间县域联系起来。在研究中，没有考虑不同区域的问题，比如说，在不同的地貌区域有不同的空间尺度，甚至转换规则也会有所不同。有待在进一步的研究中改进。

将与中心的“元胞”（某个县域）具有接边的各个县域单元来作为研究“元胞”的邻居关系。相应地，和其有接边的各县域单元的状态值就成为了该“元胞”下一个时刻状态改变的主要影响因素。

4.3 转换规则的构建

本文的转换规则和通用的地理元胞自动机的转换规则的确定是一个相反的思路。在通用模型中，是通过元胞的状态及邻居状态来定义

该元胞对周围邻居元胞的影响概率，从而确定下一个时刻邻居元胞状态的变化情况；本研究中，是通过所有的邻居元胞的状态来定义该时刻他们对该元胞的综合影响，从而确定该元胞下一个时刻状态的变化。其中综合影响是各个接边县域对该地理元胞的影响概率之和，影响概率通过邻居县域与该县域的接边长度与该县域的周长的比值和邻居的状态参数来确定。即邻居地理元胞与该地理元胞地空间公共边的长度的大小决定了邻居的状态对该地理元胞影响的大小。

4.4 离散时间

研究中，一次循环在实际的时间尺度上相当于一年。离散时间的对应关系主要是通过模拟结果和实际的疫情分布进行比较来确定。由于实际的疫情数据是以年为单位来统计的，所以在时间尺度上目前只能精确到年，如果疫情数据统计的更精确些，即统计到月或日，那么离散时间可以跟着作进一步的调整。

4.5 模拟预测结果

本文使用可视化编程语言 C#和组件式 GIS 开发，实现在没有任何人为疫情控制因素的影响下，疫情按照设计的预测模型进行扩散的过程，预测结果见图 1。图 1 (a) 是 2000 年自然疫源地聚类分析结果，图 1 (b) 是在 2000 年聚类分析结果的基础上，模拟预测的 2003 年自然疫源地发展情况。由于模型设计中没有考虑其他的对疫情的影响因素，如地貌条件、水系条件、交通条件和人为干预等，预测结果和实际的疫情状况存在一定的偏差，但仍然有一定的可信度。

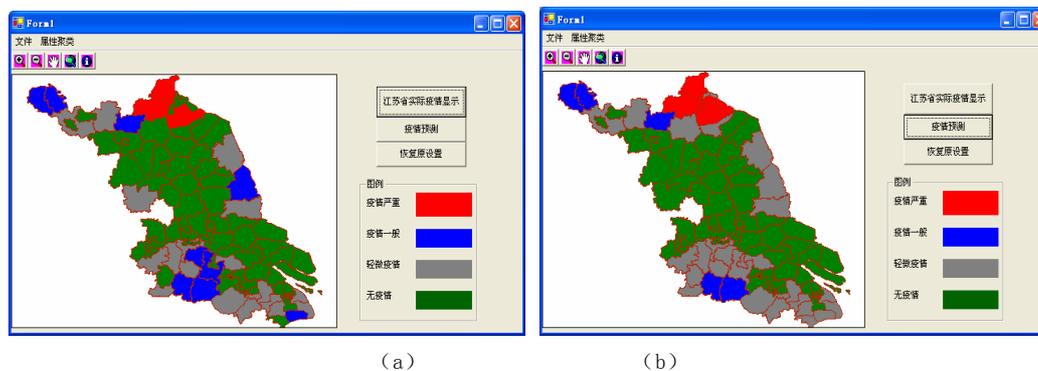


Figure 1. Simulative forecast of hemorrhagic fever with renal syndrome in Jiangsu province

图 1. 江苏省肾综合症出血热自然疫源地的模拟预测

5 小结

自然疫源性疾病的类型多，大部分自然疫源性疾病的存在、发展、流行和蔓延原因复杂。本文以 3S 作为数据获取、管理与可视化平台，以数据挖掘为分析手段，探讨了自然疫源性疾病研究中 3S/DM 集成的应用方法，并构建了防疫决策支持系统，对江苏某一种疫源地发展进行了预测分析，证明其可行性。但是要完全摸清其规律与机制尚存在很多问题，要取得有意义的成果，应在以下几个方面开展进一步的工作：

(1) 加强 3S 数据采集与管理的工作，采集更详实的疫情信息，建立疫情数据库；

(2) 加强分析预测模型的研究，使空间数据挖掘与病理模型结合更加紧密；

(3) 增强疫情信息、预测信息的表达、发布功能，使分析结果服务大众，切实提高公众健康。

References (参考文献)

- [1] GONG Shengsheng. Changes of the Temporal-spatial Distribution of Epidemic Disasters in 770BC-AD1911 China[J]. *Acta Geographica Sinica*, 2003, 58(6), P870-878(ch).
龚胜生, 中国疫灾的时空分布变迁规律[J], 地理学报, 2003, 58 (6), P870-878
- [2] HAN Guang-Hong, ZHANG Xi-Tan, FANG Li-Qun, et al. Establishment of Geographic Information System of Major Natural Foci in China[J]. *Bull Acad Mil Med Sci*, 2004, 28(2), P123-125(ch).
韩光红, 张习坦, 方立群 等, 我国重要自然疫源地地理信息系统的建立[J], 军事医学科学院院刊, 2004, 28 (2), P123-125
- [3] FANG Li-Qun, CAO Wu-Chun. Application and Development Foreground of 3S Technology in Natural Focal Diseases[J]. *Chinese Journal of Zoonoses*, 2004, 20(9), P807-811(ch).
方立群, 曹务春, 3S 技术在自然疫源性疾病中的应用及发展前景[J], 中国人兽共患病杂志, 2004, 20(9), P807-811
- [4] ZHENG Ke-chun, SONG Zhi-Zhong. Impact of Natural Disasters on Natural endemic Diseases[J]. *China Tropical Medicine*, 2009, 9(7), P123-125(ch).
郑克纯, 宋志忠, 自然灾害对自然疫源性疾病的影响[J], 中国热带医学, 2009, 9 (7), P123-125
- [5] ZHOU Xiaonong, SUN Leping, JIANG Qingwu, et al. Geographic Information Systems Spatial Analysis on Transmission of Schistosomiasis in China[J]. *Chinese Journal of Epidemiology*, 2000, 21(4), P261-263(ch).
周晓农, 孙乐平, 姜庆五 等, 全国血吸虫病流行状况的地理信息系统空间分析[J], 中国流行病学杂志, 2000, 21(4), P261-263
- [6] Lu W., J. Han and B. C. Ooi. Discovery of General Knowledge in Large Spatial Databases. Proc. of 1993 Far East Workshop on Geographic Information System(FEGIS'93), Singapore, June 1993, pp. 275-289