A COLLABORATIVE EPIDEMICAL SURVEILLANCE AND RESPONSE SYSTEM BASED ON GIS AND INFORMATION TECHNIQUE

HU Bisong ^{a,*}, GONG Jianhua ^{a,*}, SUN Jia ^a, CAO Wuchun ^b, FANG Liqun ^b

a Institute of Remote Sensing Applications, Chinese Academy of Sciences, Datun Road 3, Chaoyang District, Beijing, 100101, P.R. China ^b Institute of Microbiology Epidemiology, Academy of Military Medical Sciences, Beijing, 100071, P.R. China

Commission I, ICWG-I/V

KEY WORDS: Bayesian Analysis, Collaborative, C/S, Epidemiology, GIS, Response, Surveillance, Scan Statistics,

ABSTRACT:

Since the manual operations of traditional epidemical surveillance and response are excessive, there are obvious drawbacks of sensitiveness and effectiveness during the process of surveillance and response. GIS and information technique could help lessen those disadvantages and promote the preciseness and efficiency of epidemical surveillance and response. Existed epidemical surveillance and emergency response systems are completely separate on the whole. Due to the time difference caused by the manual operations, the persons working for public health cannot deal with the bursting diseases in time and effectively. A collaborative epidemical surveillance and response system, which is designed with C/S structure and based on internet and mobile network, was studied and designed, and its collaborative prototype system was implemented. A complete work flow of epidemical surveillance, fieldwork and emergency response was built up in this paper. The surveillance algorithm applied in this system is spatial and space-time scan statistics, and Bayesian analysis is used in the intelligent disease diagnosis module of the response system.

1. INTRODUCTION

GIS could provide geographic data related to diseases and analysis methods for spatial data. Information technique could provide effective technologic methods for the epidemical modeling. Therefore, GIS and information technique are increasingly applied and playing an important role in the epidemical researches. Many epidemical researches focus on the distribution and determinants of diseases and injuries in human populations, and the transmission factors and characteristics. The combination of spatial data provided by GIS and disease data can help researchers analyze and understand the relationship of epidemical rules and circumstance factors. Furthermore, GIS could answer the questions related to the space and time of disease burst, since infectious diseases have both spatial and temporal characteristics while transmitting.

The bursting of various infectious diseases on different regions of the world has done much harm to the safety of population life and property, causes the social panic and turmoil easily, and influences the social and economic development. Many researches focus on the development of convenient and useful epidemical surveillance information systems. And after the significant infectious disease bursting, to reduce the loss and damage to a minimum, the response and treatments to disease bursting and the managements of response resource, which are the main contents of emergency response information system, are extremely required.

Traditional epidemical surveillance relied on the clinical physicians and laboratory reports. It has obvious disadvantages

in sensitivity and efficiency due to the complexity of surveillance persons and programs. Current epidemical surveillance includes collecting and sorting of disease data, automatic data analysis, result report and detection, abnormal signal feedback and validity evaluation. It adopts new methods, such as spatial data analysis and spatial statistics, and has prominent promotion in precision and efficiency.

Emergency response in public health is adopting emergent measures and lightening the injury to citizen health and life due to the bursting epidemical events, and controlling the damage into minimum area by applying most effective measures and least resource waste. In general, an emergent treatment system should comprise a series of functions, such as assistant query, event treatment, and assistant disease diagnosis. Because the bursting events are abrupt and unexpected, the response personnel have especial demands on emergency response systems in mobile equipments.

Many researchers studied and developed GIS-based epidemical surveillance information systems (Tsui, et al, 2003; Kelly, et al, 2004; CAO, et al, 2006; Drake, 2005; Jacquez, et al, 2005), and received nice effects in practical applications. Existed epidemical response information systems (Piot, et al, 2001; Kaplan, et al, 2002; Li, et al, 2004; WANG, et al, 2005; Chai, et al, 2007) mainly focus on the response methods and functions after disease bursting, and have not reached a collaborative and integrated work mode with the surveillance systems. In other words, most epidemical surveillance systems and emergency response systems are completely separate. There are still manual operations between the surveillance and response. Due to the time difference caused by the manual operations, the

^{*} Corresponding author. Email: hubisong624@126.com; Phone: +86-10-64849299 (O); Fax: +86-10-64849299.

^{*} Corresponding author. Email: jianhuagong@sina.com; Phone: +86-10-64849299 (O); Fax: +86-10-64849299.

persons working for public health cannot treat the bursting events in time and effectively.

The thoughts and methods of collaboration have been frequently applied to the fields of computer science, geography, epidemiology and so on in recent years. This paper designed a collaborative epidemical surveillance and response system, which is based on internet and mobile network and under C/S (Client/Server) mode, built up a whole work flow which is fit to epidemical surveillance and detection, event investigation, and emergency response, experimented in practical work, and reached a satisfying result in precision and efficiency.

In the collaborative epidemical surveillance and response system designed in this paper, the broadly applied spatial and space-time scan statistic method (Kulldorff, 1997, 2001; Kulldorff, et al, 2005) was adopted in the disease detection module. This method analyzes the existed and current disease data, finds out the disease abnormal clusters in space and time, and finishes the epidemical space-time surveillance and detection. Bayesian analysis method was used for the function of assistant disease diagnosis in emergency response module. This function could finish the intelligent disease diagnosis according to the disease symptoms, patient characteristics, and clinical analysis information, and provides a rapid and scientific reference for the working people of event response. These methods above have reached a nice effect in practical experiments.

2. EPIDEMICAL SURVEILLANCE AND RESPONSE

2.1 Surveillance

The detailed contents of epidemical surveillance include data collecting, data analysis, early detection, et al. The collecting data comprise disease reports, symptom information, medical cases, etc. Data analysis is classifying and analyzing the surveillance data statistically using spatial analysis methods, providing various ways (text, figure, electronic map, etc) for the epidemical investigation and inquiry, and predicting the probable hotspot regions of disease bursting. Early detection is analyzing the statistical data, publishing the detection signals automatically, and predicting the epidemical hotspot regions and spreading by the means of statistical models.

GIS and information technique could provide strong help to epidemical surveillance. Visualization information technique can help researchers describe the complicated spatial and attribute data as visual geographical maps, accomplish the collaborative working mode of text and graph information, and study the epidemiology visually and conveniently. Furthermore, GIS could provide various topological structures and visual spatial analysis methods for epidemical researches, determine the geographical characteristics of infectious diseases, analyze the surveillance and investigation data comprehensively, and furnish emergency response with helpful suggestions and directions. Meanwhile, GIS could distinguish multiple data types and structures, and help the collecting of disease data.

2.2 Response

Epidemical response is to determine the response methods and resource managements after infectious diseases burst. The response system can use the diagnosis algorithms, give a quick judge to the bursting cases, and complete the intelligent disease diagnosis on the basis of epidemical investigation, clinical examination of patients and examination results of laboratory. In addition, the response system should provide introduction of infectious disease characteristics and direction of fieldwork. Generally, epidemical response system could include the following modules: intelligent disease diagnosis, fieldwork treatment, response recommendation, and so on.

3. ALGORITHMS OF SURVEILLANCE AND RESPONSE

3.1 Spatial and space-time scan statistics

Spatial and space-time scan statistics methods were proposed by Kulldorff of Harvard Medical School in 1997 and 2001, and are widely applied in the fields of medicine, biology and so on.

Log likelihood ratio (LLR) is used as the evaluation norm in spatial scan statistics (Kulldorff, 1997). Null hypothesis (e.g. the possibility of persons in the studied regions is identical) is required before the spatial scan statistics. Firstly, a series of scan clusters are generated. Generally, these clusters are circles whose centre points are geographical centres of studied regions. The radiuses of these scan circles vary from zero to a specified maximum value. Secondly, LLR values of all scan circles are calculated. Several (4 to 5) regions with the maximum LLR values are chosen as the available hotspot regions of disease bursting. Finally, Monte Carlo hypothesis testing is used for the statistical significance evaluation of those available hotspot regions. Regions which have passed the testing are the detected hotspot regions.

Models used in spatial scan statistics include Bernoulli model, Poisson model, Ordinal model, Exponential model, etc. For instance, LLR value of Poisson model is calculated as the following expression:

$$LLR = \frac{\left(\frac{n_Z}{\mu(Z)}\right)^{n_Z} \left(\frac{n_G - n_Z}{\mu(G) - \mu(Z)}\right)^{n_G - n_Z}}{\left(\frac{n_G}{\mu(G)}\right)^{n_G}}$$
(1)
$$= \left(\frac{n_Z}{\mu(Z)}\right)^{n_Z} \left(\frac{\mu(G) - n_Z}{\mu(G) - \mu(Z)}\right)^{n_G - n_Z}$$

where Z is a 3-dimensional vector, including the coordinates of the centre point and the radius of the scan circle, n_Z is the real disease case amount of the scan circle region, $\mu(Z)$ is the population amount of the scan circle region, n_G is the total disease case amount of the studied regions, and $\mu(G)$ is the population amount of the studied regions.

The main model of space-time scan statistics is space-time permutation model. The processing method used in space-time scan statistics is similar to spatial scan statistics. The scan cluster used in space-time permutation model is not circle but cylinder. The height of scan cylinder stands for time value (e.g. day amount), and the bottom surface of scan cylinder has the same meaning as the scan circle in spatial scan statistics. The height and radius values of scan cylinder also have specified maximum values.

Suppose C_{zd} is the observed number of cases in region z during day d. The total number of observed cases (C) is:

$$C = \sum_{z} \sum_{d} C_{zd} \tag{2}$$

For each region and each day, the expected number of cases $\mu_{i,j}$ is calculated as follows:

$$\mu_{zd} = \frac{1}{C} \left(\sum_{z} C_{zd} \right) \left(\sum_{d} C_{zd} \right)$$
(3)

Under the null hypothesis, the expected case number of particular scan cylinder A could be expressed as follows:

$$\mu_A = \sum_{(z,d)\in A} \mu_{zd} \tag{4}$$

Suppose C_A is the observed number of cases in a particular cylinder A. When $\sum_{z \in A} C_{zd}$ and $\sum_{d \in A} C_{zd}$ are small compared to C, C_A is approximately Poisson distributed with mean μ_A (Evans, 2000). Poisson generalized likelihood ratio (GLR) is used as a measure of the evidence that cylinder A contains a disease outbreak (Kulldorff, et al, 2005). GLR value could be calculated as follows:

$$GLR = \left(\frac{C_A}{\mu_A}\right)^{C_A} \left(\frac{C - C_A}{C - \mu_A}\right)^{(C - C_A)}$$
(5)

3.2 Bayesian analysis

One important module of epidemical response system is intelligent disease diagnosis. According to the epidemical fieldwork investigation, clinical examination of patients and examination results of laboratory, intelligent disease diagnosis module uses specified algorithm and program to complete quick judgement to the bursting case and diagnosis to unknown diseases. Epidemical response personnel can choose specified disease from the diagnosis result on the basis of their experience or expert suggestions, and the system could provide relevant response suggestions according to the diagnosed disease.

Intelligent disease diagnosis uses possibility as the judgement measure, and proposes the diagnosis result according to the information of symptoms, disease characteristics, laboratory examination, etc. The relationship between disease classes and judgement information should be determined from analysis results of numerous disease cases. The diagnosis result is more precise with larger amount of sample cases.

Two basic discriminatory analysis methods used in intelligent disease diagnosis are Fisher and Bayesian analysis. Bayesian analysis method which is based on probability and statistics is of higher preciseness and effectiveness. According to the prior probability of some certain event, its corresponding posterior probability could be calculated by Bayesian analysis.

Suppose events $A_1, A_2, A_3, \dots, A_n$ comprise a complete event group. For each event A_i , its occurrence probability is defined as

 $P(A_i)$ ($P(A_i) \ge 0, i = 1, 2, ..., n$). Under the condition of event *B* occurring, for each event A_i , its occurrence probability $P(A_i | B)$ could be calculated as the following expression:

$$P(A_{i} | B) = \frac{P(A_{i}) * P(B | A_{i})}{P(B)}$$

$$= \frac{P(A_{i}) * P(B | A_{i})}{\sum_{i=1}^{n} P(A_{i}) * P(B | A_{i})}$$
(6)

For the application in intelligent disease diagnosis, suppose the patient has *n* kinds of clinical symptoms corresponding to *m* kinds of diseases. For disease *j* (j = 1, 2, ..., m) corresponding to symptom *i* (i = 1, 2, ..., n), the probability is defined as $P(B_i, A_j)$. Meanwhile, the incidence rate of certain disease *j* is defined

as $P(A_j)$. Therefore, the occurrence probability of disease A_j

corresponding to all the *n* kinds of clinical symptoms could be calculated as follows:

$$P(A_{j} | B) = \frac{P(A_{j}) * P(B | A_{j})}{P(B)} = \frac{P(A_{j}) * P(B | A_{j})}{\sum_{j=1}^{m} P(A_{j}) * P(B | A_{j})}$$
(7)
$$= \frac{P(A_{j}) * \prod_{i=1}^{n} P(B_{i}, A_{j})}{\sum_{j=1}^{m} (P(A_{j}) * \prod_{i=1}^{n} P(B_{i}, A_{j}))}$$

4. SYSTEM STRUCTURE DESIGN

This collaborative system is designed in C/S architecture mode, and consists of collaborative server, PC client, field/PDA (Pocket Digital Assistant) client, and communication network. The people working in the system include server managers, PC client operators, and filed working people. The communication network comprises internet and mobile network.

The structure design of collaborative system is shown in Figure 1. The surveillance subsystem includes server and PC client, and the response subsystem includes server, PC client and PDA client. These two subsystems have the same server, and complete their collaborative work through PC client.

The work flow is as follows: The server runs a full-time epidemical surveillance and detection program. Once an

abnormal detection appears, the server sends out the detection signal. And the PC client downloads the newest detection information from the server through the internet, communicates with the field client, and disposes the event investigations and measures. The field client uses the portable digital equipment (PDA or intelligent cell phone), communicates with the server and PC client through mobile network, sends back the investigation data and information to the server, and completes the emergency response work.



Figure 1. Structure design of collaborative surveillance and response system

The collaborative server is an internet work station with powerful computing and storage abilities, employs large database systems (e.g. Oracle, SQL Server) for the storage of disease data, runs a full-time epidemical surveillance program, and with the multi-thread mode and the network protocol, implements the communication and information transmission between the server, the PC client and the field client. PC client has a collaborative operation program and a 3-D visual surveillance program, downloads the detection data from the server through the internet, completes the visualization of surveillance and detection, and provides directly visible effects about the detection which could help the operators make the correct and exact decisions. The portable digital equipment of the field client has a communication program which could connect to the server and PC client through mobile network, completes the data collecting and feedback in event investigation, and meanwhile has a program for event treatment, which could help the working people finish the emergency response conveniently and effectively.

5. IMPLEMENTATION OF PROTOTYPE SYSTEM

Prototype system of this paper is developed based on standard C/C++ programming language. Programs of server and PC client are developed by Microsoft Visual C++ IDE, and programs of PDA client are developed by Microsoft eMbedded C++ IDE. Collaborative server, whose operation system is Windows 2003 Server Edition, has CPU of Intel Core 2 T5600, 2G memory and 1000M network card. Database system used in server is Microsoft SQL Server 2000. PC client is common microcomputer. PDA client is PDA with Windows Mobile 5.

Collaborative server, which is the main part of epidemical surveillance subsystem, assumes the responsibility of wholetime surveillance, and meanwhile provides collaborative service for PC client and PDA client. The prototype system uses a powerful computing and storage server in internet as the collaborative service, and disposes two services: PC service and PDA service which are responsible to the PC and PDA clients respectively. PDA service is a multi-thread program for data exchange based on WinInet structure, and is in charge of the data exchange between collaborative service and PDA client. PC Service is a multi-thread service program developed with Socket protocol, and monitors and responds the requests of computer terminals in specified port. According to the function demands, PC Service could include terminal computing service, database service and information service of sending and receiving. Furthermore, according to the information types during the information sending and receiving, information service could be classified into three different types: message service, text service and file service. The service types and functions of collaborative server are shown in Table 1.

PC client is the information bridge of collaborative work between surveillance and response subsystems. When the surveillance computing results of server terminal appear exceptions, collaborative server will send exception information to PC client. And then PC client directs PDA client to perform emergency response work in exception regions. PC client builds up the communication connection to collaborative server and PDA client by Socket and WinInet respectively.

PDA client accepts the directions from collaborative server and PDA, provides various functions of response subsystem for the response personnel to accomplish the epidemical fieldwork investigation and emergency response, and sends the response data back. The communication connection between PDA client and collaborative server is built up on the basis of Http protocol and GPRS mobile network.

Service Name		Service Function		
PDA Service		Collaboration and data exchange		
		between Server and PC Client		
	Computing	Terminal surveillance computing		
PC Service	Service	and information exchange of the		
		computing results		
	Database	Database server and storage of		
	Service	geographical and epidemical data		
		Message	Information exchange	
	Information	Service	of messages	
	Service	Text	Information exchange	
	(Send and	Service	of texts	
	Receive)	File	Information exchange	
		Service	of binary data files	

Table 1. Service types and functions of collaborative server

6. APPLICATION OF PROTOTYPE SYSTEM

The terminal computing program of collaborative server uses spatial and space-time scan statistics methods, and implements the real-time epidemical surveillance according to the existed geographical and disease data. When the exceptions appears in both spatial and space-time scan statistics, collaborative server would send the surveillance exception information to PC client, and then PC client would direct PDA client to carry out the work of fieldwork investigation and emergency response. Meanwhile, the database service of collaborative server is providing continuous data and information exchange service for server, PC client and PDA client. Figure 2 shows the collaborative working flow of prototype system.



Figure 2. Collaborative working flow of prototype system

Real geographical data and simulating disease data are used in this experiment of prototype system, and spatial scan statistics results of surveillance module in different scales are shown in Table 2. Moreover, Figure 3 shows the epidemical surveillance computing program of collaborative server, including multiscale surveillance computing results. Intelligent disease diagnosis module and fieldwork response module of PDA client are shown in Figure 4 and Figure 5 respectively.



Figure 3. a) Epidemical surveillance computing program of collaborative server (scale: state)



Figure 3. b) Epidemical surveillance computing program of collaborative server (scale: province)



Figure 3. c) Epidemical surveillance computing program of collaborative server (scale: county)

Hotspot Cluster: 1	
Circle	
Guangxi Province	
450000	
108.312	
22.8061	
0.382493	
117635283	
695	
286.895	
21.134766	
7.892134	
0.001	

Table 2. a) Spatial scan statistics results (scale: state)

	Hotspot Cluster: 1	Hotspot Cluster: 2
Cluster type	Circle	Circle
Centre Name	Youping County	Longtou County
Centre Code	45102808	45142101
Centre Longitude	106.540440	107.961732

Centre Latitude	24.928055	22.692492
Radius	1.204827	0.481931
Population Number	1669635	1901586
Observed Cases	49	52
Expected Cases	16.766081	19.095279
LLR	27.824853	9.668875
Relative Risk	3.341264	3.140871
Р	0.001	0.001

Table 2. b) Spatial scan statistics results (scale: province)

	Hotspot Cluster: 1	
Cluster type	Circle	
Centre Name	Shuikou Town	
Centre Code	45142309	
Centre Longitude	106.586283	
Centre Latitude	22.478791	
Radius	0.156412	
Population Number	41485	
Observed Cases	2	
Expected Cases	0.704587	
LLR	2.967778	
Relative Risk	5.676992	
Р	0.001	

Table 2. c) Spatial scan statistics results (scale: county)



Figure 4. Intelligent disease diagnosis module of PDA client: a) choose the symptoms; b) choose the districts and the exposed time; c) reconfirm the input items; d) display the results.



Figure 5. Fieldwork response module of PDA client

In the experiment of prototype system, the surveillance computing program could deal with multi-scale geographical and disease data, and provide corresponding computing results. Intelligent disease diagnosis program of PDA client could accomplish the intelligent diagnosis of several hundred kinds of diseases. Meanwhile, on the condition of nice internet and mobile network, collaborative server, PC client and PDA client could implement the one-to-one, one-to-many, many-to-one, and many-to-many collaborative work rapidly and accurately.

7. CONCLUSION

The structure design and implementation of a collaborative epidemical surveillance and response system was discussed in this paper. This system combined epidemical surveillance system and response system using collaborative working mode, and accomplished the surveillance computing program and intelligent disease diagnosis program based on scan statistics and Bayesian analysis methods. The prototype system had completed the architecture design, basic program developing, equipment testing and practical experiment, and reached a satisfying result. Before applied in practical applications, the system has some possible problems to solve or improve, such as:

a) Due to the full-time and long-period system running demand, tests for the architecture mode and the carrying capacity of the server are needed.

b) Further improvement to the architecture mode is required for the optimization of the P2P (Person to Person), P2C (Person to Computer) and C2C (Computer to Computer) collaborations.

c) The disease models in the surveillance and response modules are possibly required to be improved and optimized.

d) A great deal of real disease data are needed for the precision and validity tests to the system and models.

ACKNOWLEDGEMENTS

This research is partially supported by the National High-tech R&D Program (863 Program) 2007AA12Z240, the Beijing's Natural Science Key Foundation Project No.7061005, and the National High-tech R&D Program (863 Program) 2006AA12Z109.

REFERENCES

Cao Ming-hua, Li Qun, Hu Chuan-lai. Overview on early outbreak detection system. *Chin J Epidemiol*, Vol.27, No.11, November 2006, 1005-1007.

Chai Wei-tao, Song Shu-jun. Design and implementation for response decision system of Chengdu public health emergency based on WebGIS - a case study of Chengdu. *Resource Development & Market*, 2007, 23(9): 821-824.

Edward H. Kaplan, David L. Craft, Lawrence M. Wein. Emergency response to a smallpox attack: the case for mass vaccination. *PNAS*, 2002, 99(16):10935-10940.

Evans M, et al, 2000. *Statistical distributions*, 3rd edition. New York: Wiley. 221 p.

Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. *MMWR*, 2004, 53 RR05: 1-11.

Fu-Chiang Tsui, Jeremy U, Espino, et al. Technical description of RODS: a real-time public health surveillance system. *J of American Med Infor Association*, 2003, 10: 399-408.

Geoffrey M. Jacquez, Dunrie A. Greiling, Andrew M. Kaufmann. Design and implementation of a space-time intelligence system for disease surveillance. *Journal of Geographical Systems*, vol.7, No.1, 2005: 7-23.

Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, et al. (2004). Syndromic surveillance in public health practice: The New York City emergency department system. *Emerg Infect Dis*, 10: 858-864.

John M Drake. Fundamental limits to the precision of early warning systems for epidemics of infectious diseases. *Public Library of Science Medicine*, 2005, 2: e144.

Kelly J, Henning. Overview of syndromic surveillance: what is syndromic surveillance? *MMWR*, 2004, 53 suppl: s1-s11.

Kulldorff M (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26, 1481-1496.

Kulldorff M (2001). Prospective time-periodic geographical disease surveillance using a scan statistics. *J R Stat Soc A Stat Soc*, 164: 61-72.

Li Qi, Liu Chunbo, Li Bin. A spatial data model for disease emergency response and management - on the bases of the Hefei SARS-fighting decision support system. *Computer Engineering and Applications*, 2004, 40(1): 1-6.

Liu Fang, Ye Ying, Wei Sheng, Li Qiong. Bayesian inference for stochastic epidemic model with varied latent periods. *Mathematical Theory and Applications*, Vol.26 No.2, Jun. 2006, 74-77.

LIU Tao, CHEN Zhi-ping. Implementation of unknown virus detection on Bayesian algorithm. *Journal of Hunan Institute of Science and Technology (Natural Sciences)*, Vol.19 No.2, Jun. 2006, 65-68.

Maria-Gloria Basanez, Clare Marshall, Helene Carabin. Bayesian statistics for parasitologists. *Trends in Parasitology*, 2004, 20(2):85-91.

Martin Kulldorff, Richard Hefferman, Jessica Hartman, Renato Assuncao, Farzad Mostashari. A space-time permutation scan statistic for disease outbreak detection. *PLoS Medicine*, vol.2, 2005, 216-224.

O'Neill,P.D., Roberts,G.O. Bayesian inference for partially observed stochastic epidemics. *Journal of Royal Statistical Society A*, 1999, 162: 121-129.

Peter Piot, Awa Marie Coll Seck. International response to the HIV/AIDS epidemic: planning for success. *Bull World Health Organ*, vol.79, no.12, 2001.

WANG Zheng, LI Hua-qun, CHEN Jian-guo, et al. Emergency management system based on GIS for SARS prevention in China. *Journal of Safety and Environment*, 2005, 5(3): 82-86.