Artificial Intelligence in Medicine (2006) **38**, 157—170

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

Spatiotemporal reasoning about epidemiological data

Peter Revesz [a,*,1], **Shasha Wu** [b,2]

[a] *Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA*
[b] *Computer Science Department, Spring Arbor University, Spring Arbor, MI 49283, USA*

Received 23 December 2004; received in revised form 5 May 2006; accepted 8 May 2006

KEYWORDS
Epidemiology;
Knowledge-base;
Recursive definition;
Spatiotemporal data;
Visualization;
West Nile Virus

Summary

Objective: In this article, we propose new methods to visualize and reason about spatiotemporal epidemiological data.
Background: Efficient computerized reasoning about epidemics is important to public health and national security, but it is a difficult task because epidemiological data are usually spatiotemporal, recursive, and fast changing hence hard to handle in traditional relational databases and geographic information systems.
Methodology: We describe the general methods of how to (1) store epidemiological data in constraint databases, (2) handle recursive epidemiological definitions, and (3) efficiently reason about epidemiological data based on recursive and non-recursive Structured Query Language (SQL) queries.
Results: We implement a particular epidemiological system called West Nile Virus Information System (WeNiVIS) that enables the visual tracking of and reasoning about the spread of the West Nile Virus (WNV) epidemic in Pennsylvania. In the system, users can do many interesting reasonings based on the spatiotemporal dataset and the recursively defined risk evaluation function through the SQL query interfaces.
Conclusions: In this article, the WeNiVIS system is used to visualize and reason about the spread of West Nile Virus in Pennsylvania as a sample application. Beside this particular case, the general methodology used in the implementation of the system is also appropriate for many other applications. Our general solution for reasoning about epidemics and related spatiotemporal phenomena enables one to solve many problems similar to WNV without much modification.

* Corresponding author. Tel.: +1 402 472 3488; Fax: +1 402 472 7767.
E-mail addresses: revesz@cse.unl.edu (P. Revesz), swu@arbor.edu (S. Wu).
[1] Part of this work was done while the author was visiting the Max Planck Institut für Informatik, Saarbrücken, Germany.
[2] Part of the work presented here was done while the author was at UNL.

0933-3657/$ — see front matter
doi:10.1016/j.artmed.2006.05.001

1. Introduction

Infectious disease outbreaks are critical threats to public health and national security [1]. With greatly expanded travel and trade, infectious diseases can quickly spread across large areas causing major epidemics.

Efficient computerized reasoning about epidemics is essential to detect their outbreak and nature, to provide fast medical aid to affected people and animals, to prevent their further spread, and to manage them in other ways.

Several characteristics of epidemics make them special in terms of computer reasoning needs. First, epidemiological data are usually some kind of spatiotemporal data, that is, they have a spatial distribution that changes over time. Second, epidemiological data are recursive in nature. This means that the best predictions of the spread of infections are based on earlier situations. Third, we need a fast response from any knowledge-base that contains epidemiological data. A flexible information system that can be easily modified to model new epidemics is critical in assisting people to handle the outbreaks of new diseases.

The above three characteristics in combination pose a difficult problem. Geographic information systems generally can represent only static objects that do not change over time, or if they change, then they change only slowly, for example, the population density of counties. Such a slow change may be represented in a geographic information system by a limited number of separate maps. However, continuous change over time is not easy to represent and is hard to reason about in geographic information systems.

We propose new methods to visualize and reason about epidemiological data. The major contributions and novel features of our article are the following:

- *General method for recursively defined spatiotemporal models*:

  We propose a new general method to model a class of recursively defined spatiotemporal concepts, which appear in many research areas including epidemiology. In this article, we extend the definition in [2] to allow linear combinations of the measurements of the indicators and a different time delay for each indicator.
- *Recursive epidemiological definitions*:

  We apply this new method to express the recursive epidemiological definitions and predictions about the spread of infectious diseases.
- *Implementation using recursive SQL*:

  The Prolog language is the choice for recursive definitions in many knowledge-base systems. However, Prolog is not good for querying spatiotemporal data. It is also less well-known than the widely-used Structured Query Language (SQL), which is the standard query language for both relational and constraint databases. The latest SQL standard added to the SQL language a form of recursion, enabling the expression of the needed recursive definitions. It is expected that the latest SQL standard will be implemented in all major relational database products. As part of our contributions, we also implemented for the first time in the MLPQ [3] constraint database system, which is one of the most sophisticated constraint database systems, the recursive SQL queries.
- *Epidemiological data stored in constraint databases*:

  Relational databases and geographic information systems cannot easily manage epidemiological data because of their inherently spatiotemporal nature. Constraint databases [4—6], which are very suitable for spatiotemporal data, were proposed as extensions of both relational databases and geographic information systems. There are software tools that can export any relational database or geographic information system data into a constraint database [7,8].
- *WeNiVIS—The West Nile Virus Information System*:

  We developed an example epidemic information system for reasoning about West Nile Virus infections. This system can show visually the spread of the epidemic and any other spatiotemporal data that may be generated by the system. We chose this example, because it has a typical infection pattern, it is currently still spreading through the North America, and data for it was readily available from Pennsylvania's West Nile Virus Control Program [9].

The rest of the article is organized as follows: Section 2 describes some basic concepts and related work. Section 3 describes the new general method for modeling recursively defined spatiotemporal concepts. Section 3.1 proposes a general recursive definition for spatiotemporal concepts. Section 3.2 describes the solution and optimization for the recursive definition using recursive SQL query language. Section 4 describes the source data we use for the West Nile Virus analysis (in Section 4.1), their interpolation and storage in a constraint database (in Section 4.2), and the West Nile Virus Information System we developed for the WNV analysis (in Section 4.3). Section 5 presents major results and benefits of this project. Section 6 discusses some specific issues about our method and system. Finally, Section 7 gives some conclusions and directions for future work.