VAST Challenge 2010

Mini Challenge 2: Hospitalization Records - Characterization of Pandemic Spread

Robert Howe - phalanui@gmail.com

Domain, Task, and Dataset

The task being taken on in this project is that of solving mini challenge number 2 from the VAST Challenge for 2010. In this challenge there has been a major pandemic outbreak of an unspecified disease, and the task is then to help health officials to analyse the illness across several different countries in order to help characterise the spread of the disease. The data that is available consists of hospitalisation admittance and death records for cities and countries involved in the pandemic. It should be noted that all of this data is synthetic, it has been constructed based on real data, but it is not real data itself.

The domain being explored here is that of visualising the spread of disease. There have been many major disease outbreaks throughout the course of human history, with some having devastating effects on populations around the world. Understanding the spread of these diseases is key to preventing and controlling any major future outbreaks, as well as ensuring the survival of the human species.

This proposal is for an analysis project, so the task will mainly focus on previous work in the area to find out what has been the most effective in this domain and why, as well as a comparison of currently available software packages in terms of their applicability to this domain.

Personal Expertise

I completed my undergraduate degree in Medical Science, with some courses taken from Virology, Immunology, and more general “Disease in Society” subjects. I haven’t explicitly studied any epidemiology, however hopefully some of the disease-related subjects that I have taken will be able to provide a somewhat deeper insight into the spread of this disease, and the spread of epidemic and pandemic diseases in general.

Proposed Solution

The solution to this task will consist of two parts due to the nature of the problem as an analysis project. Firstly, an extensive survey will be carried out on both the previous VAST Challenge 2010 entries, and on the general literature available for visualising disease spread.

Once this has been completed, a visualisation based on the available data will be constructed using Tableau[1] and another software package (yet to be decided). The visualisations will include features deemed effective based on the available literature in terms of past VAST Challenge solutions, and general literature in the domain of visualising epidemic and pandemic diseases. The two software packages will then be compared and contrasted to determine the strengths and weaknesses of each, and their applicability to this domain.

Scenario of Use and Possible Interface

As this is an analysis project, the scenario of use and possible interface are somewhat less applicable. For the tools developed in this domain, they would typically involve users loading relevant data into the tool, and then choosing different filters to study such as age, symptoms, hospitalisation date etc, as well as potentially geographic data to better understand the physical spread of disease. The interface here will be the standard given by Tableau and the other software package (yet to be decided), and the suitability of the interface for this domain will be analysed.

Implementation
The visualisations for this task will be implemented using Tableau[1] and another software package (yet to be decided).

Milestones and Schedule

14 November - Draft write-up of current literature available in the area along with past VAST Challenge solutions

21 November - Basic implementation in Tableau complete

24 November - More refined Tableau implementation

28 November - Basic implementation in other software package complete

30 November - More refined other software package implementation

1 December - Comparison of software packages complete, along with a more refined version of the original analysis of current literature and past VAST solutions

5 December - Presentation slides finished and ready

12 December - Presentation slides and talk finished and ready

15 December - Final report finished

Previous Work

Previous work in the area falls into two main categories. Firstly, as this is a VAST Challenge that was run in 2010, a number of individuals and organisations have already competed and published their attempts at a solution:

- The giCentre at the City University London specialises in developing techniques for visualising data, and their entry into the VAST Challenge was awarded with “Good overall design and analysis”. For their solution they developed a tool called PandemView which allowed them to visualise the data in several different ways, such as: examining symptom frequencies as an alphabetical cloud tag; exploring spatio-temporal patterns by filtering by time and place with geographic data; and using stacked bar charts and histograms to analyse hospitalisation and death trends over time.[2]
- Palantir is a commercial company that specialises in production software to analyse “big data”, and as such they completed this challenge using one of their own products (Horizon). Their solution primarily consisted of histograms of filtered results (with filters on age, symptoms, hospitalisation date etc), as well as geographic data overlaid with number of infected persons for the affected areas. This solution was very effective at correctly determining when and where the disease originated, as well as characterising the spread of the disease.

The second main category is that of more general visualisation approaches for disease spread found in the literature. As this domain is such a critically important one to human health, there have been some significant attempts to help characterise disease spread through data visualisation:

- Maciejewski et al. (2011) developed the PanViz visual analytics toolkit using mainly choropleth maps in a main view with hospitalisation and death records over time placed in smaller side views in order to characterise the spread of influenza viruses. This approach is similar to a few found in the VAST Challenge entries.[4]

- Brigantic et al. (2010) created a “Quick Look Tool” (QLT) based around modelling pandemic influenza infections, which again uses similar approaches to those found in the VAST Challenge, mainly looking at hospitalisation and death records over time to characterise the onset of the pandemic. This tool is compared to the PanViz tool and makes the observation that PanViz can be considered an interactive graphical front-end for the QLT.[5]

Bibliography

[1] - http://www.tableausoftware.com/
[2] - http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202010/challenges/MC2%20-%20Characterization%20of%20Pandemic%20Spread/entries/giCentre,%20City%20University%20London/
[3] - https://www.palantir.com/2010/12/vast2010/
[4] - R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller and S. W. Sanders, (2011) “A pandemic influenza modeling and visualization tool” in *Journal of Visual Languages & Computing*, 22(4), pp 268-278
[5] - R. T. Brigantic, D. S. Ebert, C. D. Corley, R. Maciejewski, G. A. Muller, and A. E. Taylor (2010) “Development of a quick look pandemic influenza modeling and visualization tool” in *Proceedings of the 7th International ISCRAM Conference–Seattle* (Vol. 1)