A Visual Analytics Toolkit for Evaluating Potential Syndromic Outbreaks

Ross Maciejewski, Stephen Rudolph, Shaun J. Grannis, David S. Ebert
Purdue University Regional Visualization and Analytics Center

Objective

This paper presents a toolkit designed to aid in the assessment of disease outbreak by visualizing spatio-temporal trends and interactively displaying detailed statistical data.

Background

Our toolkit adds statistical trend analysis, interactive plots, and kernel density estimation to an existing spatio-temporal visualization platform. The goal of these tools is to provide both a quick assessment of the current syndromic levels across a large area and then allow the analyst to view the actual data for a specific region or hospital over a period of time along with an indication as to whether or not a given data point is statistically significant. The sample data used for this toolkit come from over 70 emergency rooms throughout the state of Indiana.

Methods

In order to present a succinct overall picture, our toolkit overlays the result of a kernel density estimate [1] for the current day's syndromic data as a gradated color portrait. Areas colored with a dark red have the highest proportion of syndromic patients (patients with a chief complaint matching the currently selected disease or disease type). These areas offer a natural place to start searching for outbreaks.

Once a hospital or region is selected, more detailed information is displayed in a plot on the right. This plot contains the recent history for the selection for the currently specified syndrome. A cumulative summation of syndromic data for a 28-day window, with a 3-day lag is also plotted on the same chart. Whenever the historical data reaches a point higher than the summation, an alert is generated, highlighting an unexpected event and prompting the analyst to investigate.

Results

A screenshot of our toolkit is shown in Figure 1. The kernel density estimate is shown as a bi-color scale moving from blue to red. There are several patches of a darker red throughout the state of Indiana, indicating a high estimated percentage of syndromic occurrence in that area. The analyst has selected one of these peaks just north and east of Indianapolis as well as a hospital in that same region. Both the selected region and the hospital's data is shown for the past week (the blue lines). Both plots contain several data points above the expected values (shown in green). These higher-than-expected points are marked with a red disc. At this point, the analyst may decide to take action or to simply monitor the area more closely over the next week to see if the situation worsens.

The analyst has also selected an alert generated by the hospital data. If the alert had been for a day in the past, the current day would move to the day of the alert. The small red dots that represent the homes of syndromic patients are only shown if they contribute to the selected alert, allowing the analyst to ascertain the geographic distribution of the outbreak. In this case, there are at least four or five red dots further away from the hospital, suggesting a wider distribution.

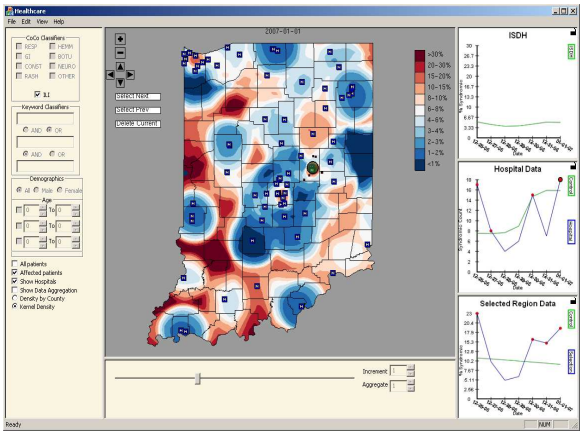

Figure 1 – Syndromic outbreak screenshot showing the kernel density estimate and an analysis plot for the selected region.

Conclusions

The toolkit provides an analyst the ability to quickly surmise the current conditions for a given disease or type of disease over a large region. When highlighting areas of interest, the analyst is shown the raw data and statistical analysis over time for that region. Using this procedure, the analyst will is able to form and investigate hypotheses about potential disease outbreaks quickly and accurately.

References

[1] B. W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC, 1986

Further Information:
Ross Maciejewski, rmacieje@purdue.edu
www.purvac.org