Visual and Decision Informatics (CVDI)

University of Louisiana at Lafayette, Vijay V Raghavan, 337.482.6603, raghavan@louisiana.edu

Drexel University, Xiaohua (Tony) Hu, 215.895.0551, xh29@drexel.edu

Tampere University (Finland), Moncef Gabbouj, +358.400.736613, moncef.gabbouj@tut.fi

MapReduce-Based Spatio-Temporal Hotspots Detection and Prediction

Analysis of hotspots, referred to as spatial/temporal concentrations of abnormal activity, has broad applications in many areas important to daily living. These include epidemiology, disease surveillance, crime prevention, and environmental monitoring, to name a few. Understanding such critically important abnormalities helps identify the underlying causes of and appropriate steps for necessary action and possible remediation.

Researchers at the Center for Visual and Decision Informatics (CVDI) have developed a MapReduce-based framework spatio-temporal hotspots detection and prediction technique. It is based on a novel big data platform. Conventional hotspot detection methods use interpolation and tend to include non-hotspot regions. That can present problems. The breakthrough CVDI hotspots analytical tool is able to detect hotspots in a spatio-temporal context with a significant reduction in false positives -' a major advantage.
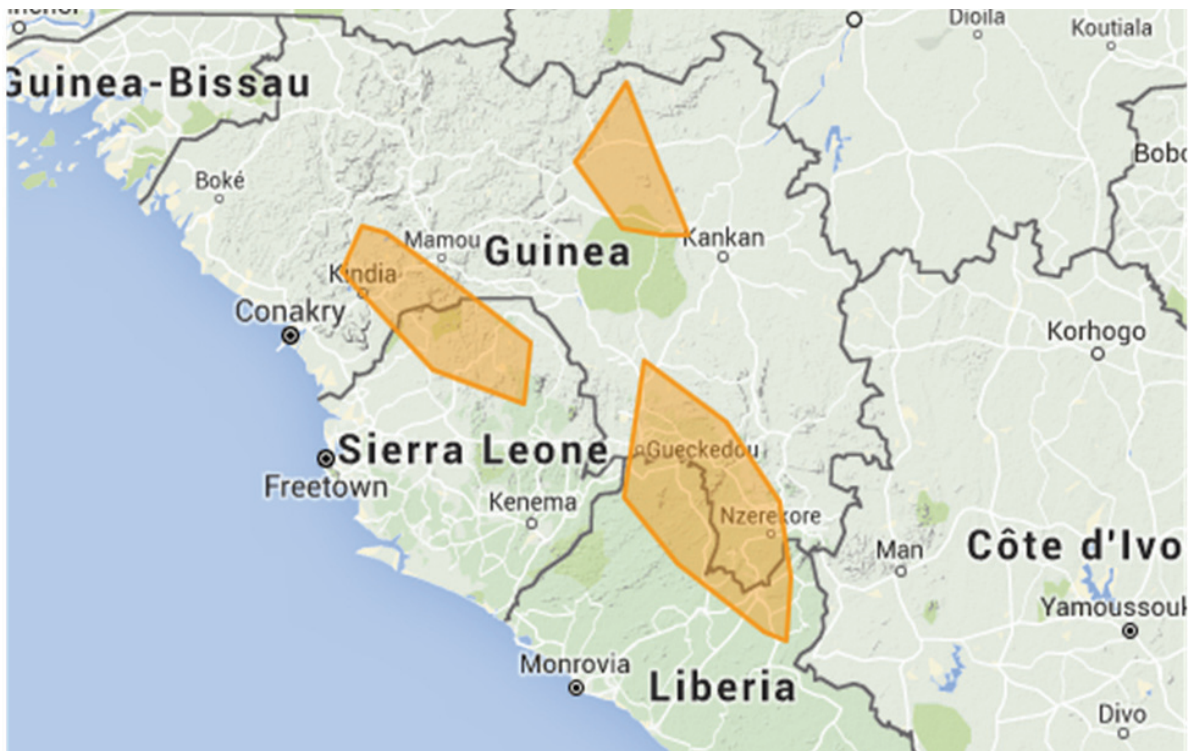

Spatiotemporal hotspots of Ebola outbreak detected around May 25, 2014 in west Africa

Visual and Decision Informatics (CVDI)

CVDI researchers extended the algorithmic approach by developing a distributed version of polygon propagation based on MapReduce. This MapReduce-based framework uses a polygon propagation based approach to detect compact hotspots tailored to the region(s) of interest. Polygon propagation is computationally expensive. MapReduce is a programming model for parallel processing of large data sets on a cluster of commodity machines. During empirical evaluations, the MapReduce-based algorithm is capable of reducing execution time by as much as 90% compared to serial implementations.

This breakthrough uses an ensemble-based hotspots prediction module that leverages multiple prediction models (temporal, seasonal, spatial, and their combinations) for forecasting hotspots. The modeling is tailored to a local time series to predict subsequent spatio-temporal hotspots. This ensemble-based prediction approach also improved prediction accuracy by more than 10% over similar techniques.
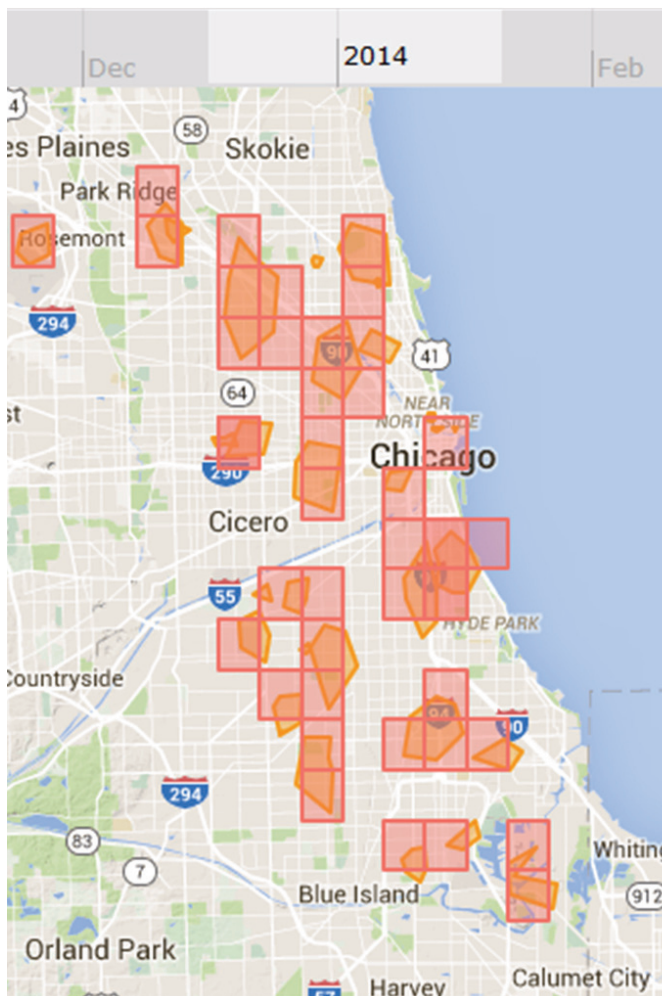
Most prediction models for hotspots use techniques such as kernel densities or time series forecasting. However, using a single model to make short-term forecasts can be prone to problems. These are due to sampling variations, model uncertainty, and structure changes over time. This breakthrough tool overcomes these problems by using an ensemble-based approach that leverages multiple models to predict outcomes with different conditions that vary the outcome and parameters.

This hotspots analytics framework was tested on the 2014 West Africa Ebola Outbreak, on Louisiana Historical Contagious Diseases, and on Chicago Crime datasets. Results have been promising. Furthermore, social media, the proliferation of sensors, and other crowdsourcing mechanisms have provided unprecedented opportunities to observe and predict hotspots around the globe. These new modalities of communication and messaging have resulted in an explosion of data. The scalability and fault tolerance nature of the MapReduce-based analytics provides the ability to perform the type of large-scale machine learning that will become increasingly important in the future.

Spatiotemporal hotspots of Chicago Crime in the first Quarter of 2014, the underlying orange polygons are the detected hotspots and the red squares on top are predicted hotspots area based on previous historical data.

Economic impact: Spatio-temporal hotspots detection and prediction can have a wide range of applications in areas such as homeland security, diseases surveillance, crime prevention, and environmental monitoring. The potential economic impacts of the framework are substantial because in many domains it can be an efficient proactive decision support tool. More than $3 billion of federal funds are spent annually on assisting state and local law enforcement in preventing crime. A severe contagious disease outbreak, for example, a pandemic flu, could result in a 5% reduction in the U.S. GDP, totaling $675 billion in lost worker productivity and decreased consumer spending. This breakthrough helps law enforcement allocate limited resources to the right place at right time to prevent and minimize crime. It can also assist public health officials and environmental protection agencies to analyze real-time surveillance sensor data to detect emerging abnormalities and to prepare for what is likely to happen next.

For more information, contact Jian Chen at the University of Louisiana at Lafayette, jchen@louisiana.edu, Bio http://www.nsfcvdi.org/peopleprofiles/jian-chen/, 337.482.0694, or Satya Katragadda, satyasivakumar@gmail.com, or Shaaban Abaddy, shaabanmahran@hotmail.com.

Forecasting Influenza Occurrence to Improve ED Operations

Given how the virus strain changes every year and related environmental factors seasonal influenza forecasting is a challenging problem. Researchers at the Center for Visual and Decision Informatics (CVDI) developed a novel big data based real-time seasonal influenza forecasting technique projects titled "Visual Analytic Approaches for Mining Large-Scale Dynamic Graphs."

This novel influenza forecasting model that uses a two-stage vectorized time series model that captures the influence of local environmental weather conditions (based on frequent associations between the flu severity and weather conditions). It can be used to forecast patients visiting emergency departments for influenza type illness. To forecast future flu occurrences, the impacts of environmental conditions and spatiotemporal flu spread characteristics are integrated into the vectorized time series model.

The United States Centers for Disease Control (CDC) monitors weekly flu projections and provides data that is anywhere from a week to two weeks old. There are also several real-time flu surveillance systems for flu monitoring based on search keys from Google, and social media trends from Twitter. These models only provide real-time monitoring capabilities rather than forecasting.

There is considerable evidence in literature about the influence of environmental factors (temperature, humidity, precipitation) on influenza virus survivability, and patterns of spread in space and time. However, the influence of environmental factors had not previously been captured adequately for real-time forecasting of influenza. The resulting model from CVDI researchers outperforms much better accuracy performance compared to existing time series based influenza forecasting.

The seasonal influenza forecasting model that was developed from this research is currently being adopted by the Schumacher Group. The influenza prediction model will be used to forecast emergency department visits for influenza type illness. It is expected to be deployed across the 130 emergency department facilities in the United States that the Schumacher Group manages. The predictive analytics capabilities of this model will help Schumacher Group and their partnering hospitals to better manage emergency department resources, to better staff emergency department facilities, and to more effectively allocate resources.

> **Economic impact**: The ability to accurately predict influenza volume contributes to the organization's ability to prepare for staffing and other operational impacts. Proactive resource management impacts hospitals' abilities to make most efficient use of its resources, including providers. Given the high percentage of expenses in emergency departments (EDs) that are associated with direct provider cost, modest reductions to provider cost can be expected to result in significant and positive impact on ED operating costs.
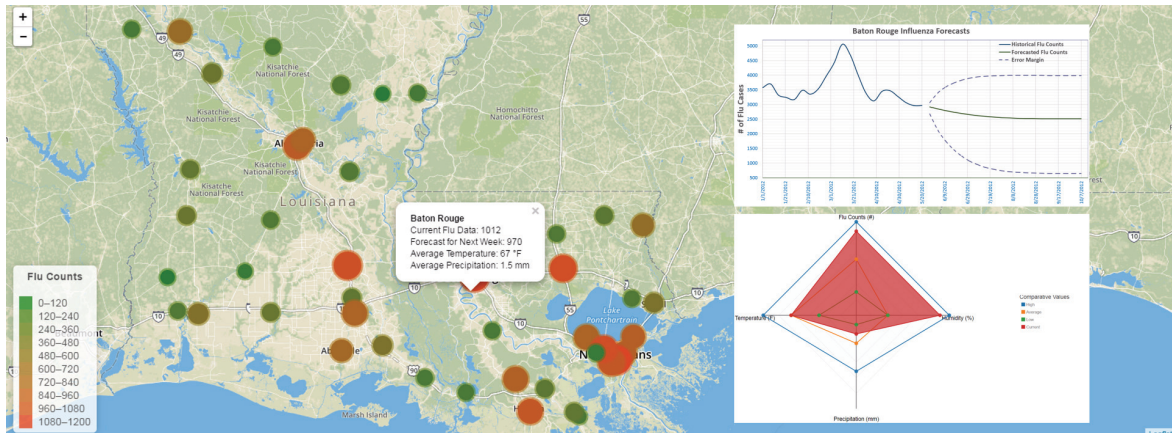
The dashboard shows flu map for the state of Louisiana, flu forecast for the city of Baton Rouge (top right) along with environmental conditions including temperature, humidity and precipitation at that time (bottom right)

For more information, contact Raju Gottumukkala at the University of Louisiana at Lafayette, raju@louisiana.edu, http://www.nsfcvdi.org/peopleprofiles/raju-gottumukkala/, 337.482.0632.