OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE

OPEN SOURCE INDICATORS (OSI)

IARPA

LEADING INTELLIGENCE INTEGRATION

Jason Matheny, Ph.D.
October 2014

Goal

- Develop and test methods for continuous, automated analysis of publicly available data in order to anticipate and/or detect significant societal events:

  – Disease outbreaks, political instability, political elections

- "Beat the news" by fusing early indicators of events from diverse data.

Approach

- 3-year forecasting tournament

- Research teams train machine learning models to detect patterns in publicly available data that have historically preceded societal events

- Teams are evaluated on the accuracy and timeliness of forecasts they deliver about real-world events in Latin America, the Middle East & North Africa

Example: Disease Outbreak

Data source	Indicator
Blogs, microblogs, web search queries, social networking sites	↑ keyword frequency: flu, sick, fever
Mobile devices	↑ non-work geolocation, reduced mobility
Webcams	↓ commuter traffic, public gatherings
Consumption	↑ retail sales of thermometers
Web traffic	↑ visits health-related websites
Imagery of health facility parking lots	↑ crowding
Reservation systems	↑ cancellation rates

Results

- Flu incidence
  - 26-day lead-time, 70% accuracy (4,320 events)

- Rare diseases
  - 6-day lead-time, 55% recall, 60% precision, 75% accuracy (70 events)

- Civil unrest
  - 8-day lead-time, 75% recall, 90% precision, 75% accuracy (>10,000 events)

- Elections
  - 14-day lead-time, 85% accuracy (20 events)

- Geolocation
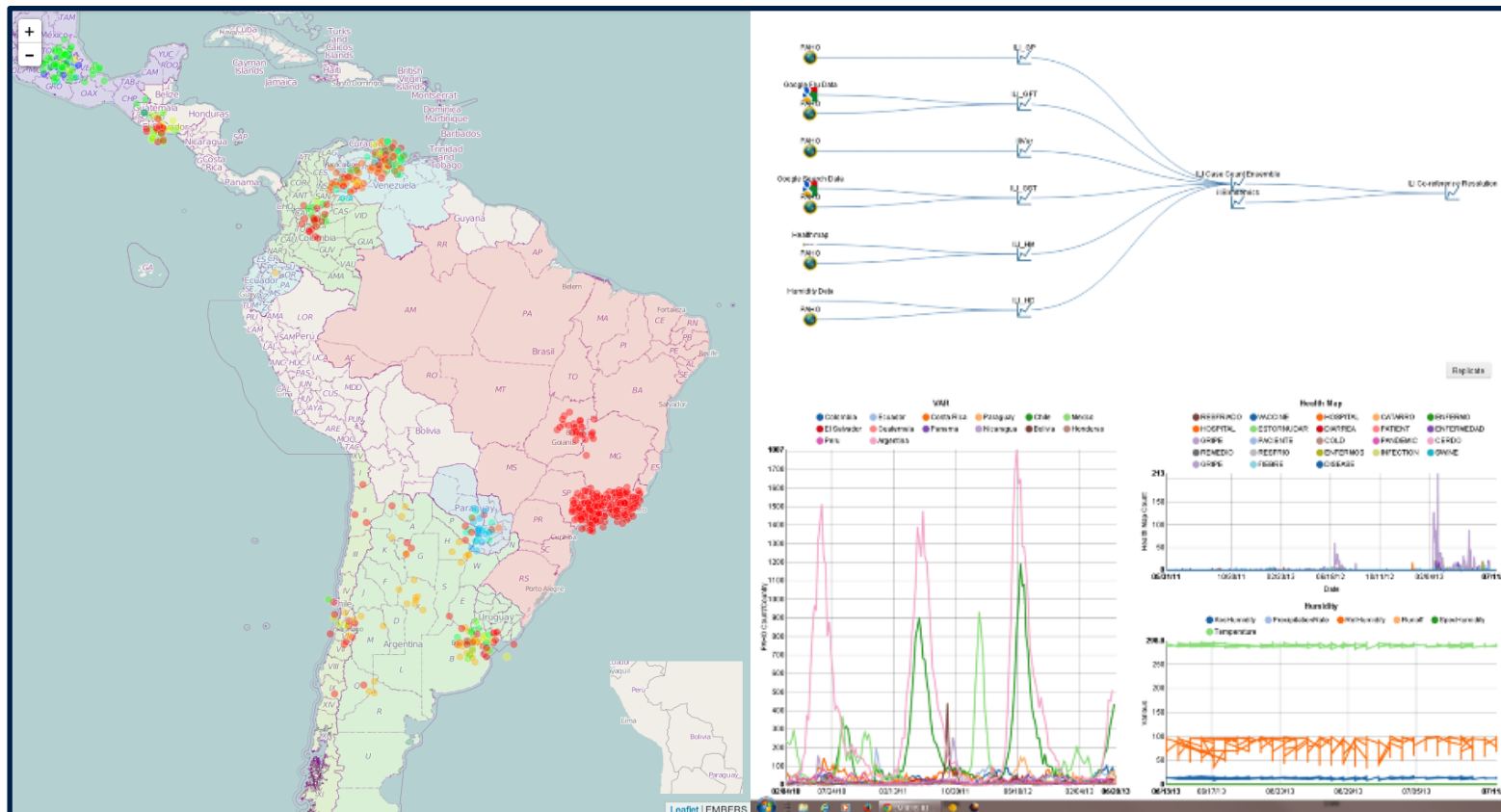 - ~90% of social media volume to the city-level

Examples of Successful Forecasts

- Riots after impeachment of Paraguay president (2012)

- The "Brazilian Spring" (June 2013)

- Hantavirus outbreaks in Argentina and Chile (2013)

- Venezuelan student uprising (Feb 2014)

- Recent elections in Panama, Colombia (2014)

Dashboard

Press coverage

USA TODAY
A GANNETT COMPANY

Data analysis allows researchers to predict disease outbreaks

BloombergBusinessweek
What Google Searches About the Future Tell Us About the Present

NOVAnext
The Inevitability of Predicting the Future

Slate
Symptoms of Violence
How technology can help spot civil unrest and even war before they start.

Science   The World's Leading Journal of Original Scientific Research.
BIG DATA
The Parable of Google Flu: Traps in Big Data Analysis

nature   International weekly journal of science
Spies to use Twitter as crystal ball

The Washington Post   PostTV   Politics   Opinions
The intelligence community gets social

Research Team

T&E

Points of Contact

IARPA
Jason Matheny, jason.matheny@iarpa.gov
Pete Haglich, peter.haglich@iarpa.gov
Philippe Loustaunau, philippe.loustaunau@iarpa.gov
Ben Cameron, benjamin.cameron@iarpa.gov

Virginia Tech
Naren Ramakrishnan, naren@vt.edu
Chris Walker, cwalk89@vt.edu