Proactive and Predictive Policing through the Application of Visualization and Data Analytics

Dr. Abish Malik (Research Scientist, Purdue University)
Chief John Cox (Chief of Purdue University Police)
Sgt. Thomas Gerber (Ohio Dept. of Public Safety)
Panelists

• Abish Malik, Ph.D.
  • Research Scientist, Purdue University, IN
  • Decision making using data analytics and visualization

• John Cox
  • Chief of Police, Purdue University, IN
  • 28+ years career in Law Enforcement

• Thomas Gerber
  • Deputy Chief of Operations for the Ohio Homeland Security’s Strategic Analysis & Information Center
  • 10+ year veteran of the Ohio State Highway Patrol
Overview

• VACCINE center
• Visual analytics
• Big data challenges
• Visual analytic solutions for law enforcement
  • Visual Analytics Law Enforcement Toolkit (VALET)
  • Social Media Analytics and Reporting Tool (SMART)
Who we are: VACCINE

• To solve today’s and tomorrow’s problems by exploring, analyzing, and reasoning with massive, multisource, multi-scale, heterogeneous, streaming data

• Mission: Provide visual analytic and scalable solutions for all 2.3 million extended homeland security personnel
  • 185,000 DHS personnel, 350,000 law enforcement personnel, 750,000 homeland security practitioners, 2.3 million extended personnel
VACCINE: International Team of Experts, 75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

• Federal Operating Components:
  • US Coast Guard
  • US Transportation Security Agency
  • US Citizenship and Immigration Service
  • US Federal Emergency Management Agency
  • US Customs and Border Protection
  • US CERT

• Law Enforcement
  • Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

• State Fusion Centers
  • Ohio (SAIC), Indiana (IIFC)

• Companies
  • Motorola, Kimberly Clark, Duke Energy, Bank of America
What We Do: Visual Analytics

Visual Analytics\(^1\) is the science of **analytical reasoning** facilitated by **interactive visual interfaces**

Interactive visualization, 
data analysis, 
exploration, and 
decision making 
with human in the loop!

---

1. Illuminating the Path: The R&D Agenda for Visual Analytics
Big Data

30 ZETTABYTES
The amount of data that will be stored in the world by 2020.
One zettabyte is the equivalent of streaming 36 million years of HDTV video
SOURCE: IDC

10 TERABYTES
The amount of data Facebook generates daily.
SOURCE: IBM

7 TERABYTES
The amount of data Twitter generates daily.
One terabyte would hold about 250,000 songs, or nearly a year’s worth of continuous music
SOURCE: IBM

40 PERCENT
The annual growth rate of global data generated per year.
SOURCE: MCKINSEY

10 BILLION+
The number of mobile-connected computing devices in use by 2016, up from less than 7 billion in 2012.
SOURCE: CISCO
Big Data: Solution to Global Challenges?

• Big Data is often defined along three dimensions:
  • **Volume** – size
  • **Velocity** – rate of input, update, change
  • **Variety** – different types, sources, variables

• Need:
  • Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information
    (TechAmerica Foundation)
  • **Enable effective, efficient analysis, decision making, planning, and action**
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• Qualitative, fuzzy, and social data
  • Preferences, significance of one relationship over another; Trust

• Context
  • Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make
    it into the collected data

• Spurious vs. Significant
  • Big data means more statistically significant events and correlations,
    but they may not have any relevance
  • Increases noise to signal ratio

• Big problems
  • Complex, multifaceted, multiparameter big challenges with
    unquantified dependencies
What Visual Analytics Offers to Stakeholders

• Enable users to be more effective through innovative interactive visualization, analysis, and decision making tools
• Provide the right information, in the right format within the right time to solve the problem
• Turn data deluge into a pool of relevant, actionable knowledge
• Enable user to be more effective from planning to detection to response to recovery
• Enable effective communication of information
Visual Analytics Law Enforcement Toolkit (i)VALET
Background

• Request from Tippecanoe, IN Sheriff:
  • How to harness all the digital crime data in the county

• Interactive crime analytics and visualization tools to provide the right information in the right amount of time to enable effective decision making and action
Our Solution: Visual Analytics Law Enforcement Toolkit – (i)VALET

Impacts:
• In use to analyze crime patterns and to connect strings of activities (200+ downloads)
• Investigating correlation factors
• Analyzing time of day problems to improving accuracy of police record management system
• Novel statistical predictive model incorporated for planning
• Incorporating predictive alerts

VALET delivered:
• Spring 2011: WL, Lafayette Police
• Fall 2013: Ohio State Highway Patrol
• Spring 2014: NYPD
• Fall 2014: Evansville PD, New Albany PD
iVALET delivered:
• October 2011: Purdue, WL Police, Lafayette PD
VALET Demo
Mobile VALET Demo

iVALET
Officer on patrol
Example of Law Enforcement Tasks and Uses

• **Police Chief/Sheriff – strategic goals**
  - Resource allocation, return on investment
  - Yearly effectiveness (Are we safer?)
  - Emerging trends

• **Police Captain – short-term effectiveness**
  - Force allocation for the day/week
  - Emerging trends and anomalies

• **Officer – improved situational awareness**
  - What happened near me around the same time in the past?
  - What will make me better prepared when I respond to this call?

• **Crime analyst/investigator – connections**
  - Correlations, dependencies, trends, predictions
Suicides/Attempts/Threats – Fall 2014

Crime For Drawn Rectangles

Home
Away

Accuracy 92%

Well-Being Check: 29
Suicide/Attempts/Threats: 19
Suicide: 1

Day-of-the-Week

VACCINE

May 2015
Prediction Features

• Novel prediction techniques based on seasonal, day of week, and time of day patterns
• Determination of right spatial and temporal level for reliable predictions
• Correlation of multiple types of crime with lag/lead indicators included
• Location aware updating of relevant information
Geospatial natural scale template: Filter out low activity regions
Predicting for 3/11/2014: Allocating Resources by Police Beats
User Refinement Using Domain Knowledge

- > 8%
- > 7%
- > 6%
- > 5%
- > 4%
- > 3%
- > 2%
- > 1%
- > 0%
- Exclude
User Refinement Using Domain Knowledge

Reg. 1

Reg. 2

Reg. 3

Reg. 4

TCPD

Walmart

Tipp Mall

Purdue
Applying Natural Time Templates

9AM – 3PM

9AM – 5PM (Day)

3PM – 7PM

9PM – 3AM (Night)

> 89%
> 78%
> 67%
> 56%
> 45%
> 34%
> 23%
> 12%
> 0%
VALET: Using Visual Analytics to Stay Ahead of Crime

• Visual analytics beneficial for all members of the police force
  • Chief
  • Shift supervisor
  • Detective/analyst
  • Officer on the street
• Actionable, justifiable information for effective decision making
• Real-time monitoring and alerts
Overview

• VACCINE center
• Visual analytics
• Big data challenges
• Visual analytic solutions for law enforcement
  • Visual Analytics Law Enforcement Toolkit (VALET)
  • Social Media Analytics and Reporting Tool (SMART)
Social Media Analytics and Reporting Tool (SMART)
Social Media is Ubiquitous

• About 1 in 7 people use a social networking site at least once a month
  • Number of active social media users in 2013:

<table>
<thead>
<tr>
<th>Platform</th>
<th>Users 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>1,190 million</td>
</tr>
<tr>
<td>YouTube</td>
<td>1,000 million</td>
</tr>
<tr>
<td>Google+</td>
<td>540 million</td>
</tr>
<tr>
<td>Twitter</td>
<td>232 million</td>
</tr>
<tr>
<td>Flickr</td>
<td>150 million</td>
</tr>
<tr>
<td></td>
<td>87 million</td>
</tr>
</tbody>
</table>

• Massive amounts of rich data generated
  • 400 million tweets per day
  • 72 hours of video uploaded to YouTube every minute
  • 1.54 million photos uploaded to Flickr per day
Finding a Needle in a Stack of Needles

Discovery of critical information from random daily chatter

Law enforcement and investigative analysis
How To Harness Social Media

- Law enforcement monitors social media
  - Costly
  - Inefficient
  - Information overload
- Purchase automated tools
  + Improves efficiency
  + Solves most problems but confused by what’s popular on social media

Need to find the relevant tweets - SMART
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
• Used by U.S. Coast Guard
  • District 8 for events in 2014 and 2015: Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations
• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston marathon bombing
  • Superbowl 2014 in NJ

• During recurring events
  • Football games

• During unexpected events
  • Natural disasters
  • Civil movements, riots

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
  • Gang activity
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston marathon bombing
  • Superbowl 2014 in NJ

• During recurring events
  • Football games

• During unexpected events
  • Natural disasters
  • Civil movements, riots

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
  • Gang activity
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston marathon bombing
  - Superbowl 2014 in NJ
- During recurring events
  - Football games
- During unexpected events
  - Natural disasters
  - Civil movements, riots
- Detecting daily issues
  - Teen threatening violence – Louisville, KY
  - Gang activity
Example From Purdue Football 10/12/14

• Prevented violence – the poor tuba player
OSU vs. Michigan 2014 Football Game
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston marathon bombing
  - Superbowl 2014 in NJ
- During recurring events
  - Football games
- During unexpected events
  - Natural disasters
  - Civil movements, riots
- Detecting daily issues
  - Teen threatening violence – Louisville, KY
  - Gang activity
Where Can It Be Helpful?

Hurricane Sandy, October 2012
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston marathon bombing
  • Superbowl 2014 in NJ

• During recurring events
  • Football games

• During unexpected events
  • Natural disasters
  • Civil movements, riots

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
  • Gang activity
Louisville, KY Oct 16 – School Shooting Threat

• WeakndMLG 10-16 23:07:50 EDT EVERYONE SKIP SCHOOL TOMORROW RT "@TaZeNaToR1: IM LITERALLY GOING TO KILL SOMEONE"  http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125203,-85.81704

Follow Up Tweets

• WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets?  http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
• WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs  http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072
SMART: Harnessing Social Media to Enable Proactive Policing

• Powerful tool for daily use and during special events
• Situational awareness for:
  • Large-scaled planned events
  • Unexpected events
  • Daily monitoring of incidents
• Realtime monitoring and email alerts
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Balance human cognition and automated analysis and modeling
  • Often applied on-the-fly for specific components identified by the user
• Interactivity and easy interaction
• Understandability of why predicted value is what it is
• Empowering the user, not replacing them
Acknowledgements

• This work has been supported by the U.S. Department of Homeland Security’s VACCINE Center under award number 2009-ST-061-CI0006
Discussion and Questions

For further info.: www.VisualAnalytics-CCI.org

Dr. Abish Malik
amalik@purdue.edu

Chief John Cox
jkcox@purdue.edu

Sgt. Tom Gerber
ThJGerber@dps.ohio.gov
Proactive and Predictive Policing through the Application of Visualization and Data Analytics

Discussion
Backup Slides
Multivariate Correlative Predictive Analytics

• Two Key Features:
  • Automatic correlation computation against lead/lags
  • Temporal and spatial windowing
• Enables detection of periodic properties among different data variables
• Allows correlations at multiple spatial and temporal granularities
• Utilizes Pearson’s Correlation in current implementation
Multivariate Correlative Predictive Analytics

- Two Key Features:
  - Automatic correlation computation against lead/lags
  - Temporal and spatial windowing

- Enables detection of periodic properties among different data variables
- Allows correlations at multiple spatial and temporal granularities
- Utilizes Pearson's Correlation in current implementation

Graph showing global correlation over time with statistical significance indicated for days -26, -11, and -1.

Burglary crimes vs. drug violations
Big Data

These forces fuel an explosion of data – a new economic asset forming the basis of opportunity.

**In just two days**
we now generate as much data as was generated in total through 2003.

**80% of all data**
is unstructured and growing at 15-times the rate of structured data.

**Over 1 billion tweets**
are sent every 3 days.

**5 million trade events**
are clocked every day.

We are here.

2013

2015

- **Sensors and devices**
- **Percentage of uncertain data**
- **Social media**
- **VoIP**
- **Enterprise data**
Where Can It Be Helpful?
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order: 10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC: 10/29, 8:00 PM
Keene, NH Pumpkin festival
10/18/14 – 10/19/14 Overview

Major topics:

Security classifier became prominent.
Keene, NH

Time range: 10/18 12:00pm – 10/19 06:00am, Security classifier)
The picture was later used by ABC news(link).
Keene, NH
Geospatial natural scale template

Threshold for Max. Value (i.e., filter out low activity regions)

STL Forecasting

Time Steps without Incidents

Distribution of Incidents

Binary Data
Geospatial prediction

Predicting for geospace at natural scale templates

- Man-made boundaries
- Rectangular grids

STL Forecasting

Count/KDE Signal

Time #

Count/KDE Value

1  3  5  7  9  11  13  15  17  19  21  23  25  27  29
Case Study: Forecasting crime incidence levels for 3/11/2014 (Tue) using historical data

Overall historical distribution

Past 2 Years

March + Tuesday Filter
(Past 10 Years)
Geospatial natural scale template: Filter out low activity regions
Predicting for 3/11/2014: Allocating Resources by Police Beats
User Refinement Using Domain Knowledge

- > 8%
- > 7%
- > 6%
- > 5%
- > 4%
- > 3%
- > 2%
- > 1%
- > 0%
- Exclude
User Refinement Using Domain Knowledge
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)

<table>
<thead>
<tr>
<th>User ID</th>
<th>Creation Date</th>
<th>Tweets Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1315016449</td>
<td>13-04-15 17:01:26 EST</td>
<td>Our hearts go out to anyone affected by the Boston Marathon bombing today.</td>
</tr>
<tr>
<td>1315016449</td>
<td>13-04-15 17:01:26 EST</td>
<td>JFK Library is a separate incident. Electrical fire. No relation to #BostoMarathon explosion.</td>
</tr>
<tr>
<td>65893882</td>
<td>13-04-15 17:01:32 EST</td>
<td>My heart goes out to the people in the explosion today.</td>
</tr>
<tr>
<td>268911912</td>
<td>13-04-15 17:01:30 EST</td>
<td>too think that I was on boylston right where the explosion happened an hour before though...</td>
</tr>
<tr>
<td>515614871</td>
<td>13-04-15 17:00:30 EST</td>
<td>I'm near the location the explosion occurred. My prayers goes to everyone who was hurt.</td>
</tr>
<tr>
<td>338849700</td>
<td>13-04-15 17:01:03 EST</td>
<td>JFK Library is a separate incident. Electrical fire. No relation to #BostoMarathon explosion.?</td>
</tr>
<tr>
<td>170292353</td>
<td>13-04-15 17:01:08 EST</td>
<td>?@BUNewsService: JFK Library is a separate Incident.</td>
</tr>
<tr>
<td>642820457</td>
<td>13-04-15 17:01:12 EST</td>
<td>@JSKFG17</td>
</tr>
<tr>
<td>622671783</td>
<td>13-04-15 17:01:17 EST</td>
<td></td>
</tr>
<tr>
<td>356162725</td>
<td>13-04-15 17:01:41 EST</td>
<td>Plan was a drone strike.</td>
</tr>
<tr>
<td>87716672</td>
<td>13-04-15 17:01:41 EST</td>
<td>Glad to be alive and safe but this is an unspeakable terror. People freaking out everywhere and I'm glad I was no closer than I was.</td>
</tr>
<tr>
<td>127240079</td>
<td>13-04-15 17:02:16 EST</td>
<td>Marathon explosion cause probabilities: 10% gas main, 10% foreign criminals, 80% drone strike.</td>
</tr>
</tbody>
</table>
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)
Our Solution: Social Media Analysis and Reporting Toolkit (SMART)

Hurricane Sandy, October 2012
(Emergency management perspective)

Super Bowl, February 2014
(News media perspective)