The Process Of Predicting with Big Data

Art and Science Applied

Caron B. Kogan
Strategic Planning Director
March 6th, 2013
We Deal with Diverse Big Data Problems

Provided over 450,000 hours of Full Motion Video to warfighters.
Today

• Predictive Analytics
  – *Art and Science to pull out actionable Insights*

• Process Stages
  – *Align with Business*
  – *Execute analytics*
  – *Deliver Results*

• Challenges with Big Data

• Example Use Case
What is Predictive Analytics?

Predictive analytics is the branch of data mining concerned with the prediction of future probabilities and trends.

To discover that which you do not know!
Predictive Analytics Process

= Science + Art

• Systematic organization of knowledge extracted from data
• Delivered to provoke new insights
• Seeking to recognize possible future events
Predicting Future Events

Reasoning with Human and Physical Factors

Social Sciences
Mathematics Statistics Computer Science
Natural Sciences

Actionable Insights
Stages of a Predictive Analytics Project

Deliver Results

Execute Analytics

Align with Business
The Path to **Actionable Insights**
Identify the business problem

Know the data you have

Get professional help

Be sure the decision maker is prepared to act

Communicate the results in business language

Start with a quick win

Line up a business champion

Test, revise, repeat.

Align with Business
Deliver Results

Execute Analytics

Align with Business
Execute Analytics

CRISP-DM  Cross Industry Standard Process for Data Mining
Signals from the Noise

Big Data → Interesting Signals → Patterns Over Time
Classic Data Mining

- Problem Definition
- Data Gathering & Preparation
- Model Building & Evaluation
- Knowledge Deployment
Data Mining with Big Data

1. Problem Definition
2. Data Gathering & Preparation
3. Data Reduction & Extraction
4. Model Building & Evaluation
5. Knowledge Deployment

The process involves defining the problem, gathering and preparing data, reducing data and extracting knowledge, building models and evaluating them, and finally deploying the knowledge.
Model Techniques – The Science

Predictive Analytics Methods

- Classification
- Link Analysis
- Clustering
- Decision Trees
- Pattern Matching
- Regression Analysis
- Neural Networks
- Feature Extraction
Factor in Time

- Feature Extraction
- Classification
- Link Analysis
- Clustering
- Neural Networks
- Regression Analysis
- Decision Trees
- Pattern Matching

Past - Present - Future
Identify Connections with Analytics

Operational Data

Analysis

Information Clusters

CSV
DNS
AV
IPS
LDAP
Netflow
Logos
DCHP
SNMP
AD

Correlated Results
Art to Provoke……

Big Data requires new visualization techniques to **REDUCE** the complexity of data down into something a human can comprehend

...Design to Deliver.
Visual Data Exploration

User Interaction

Transformation

Data

Mapping

Visualization

Model Visualization

Model Building

Data Mining

Models

Parameter Refinement

Knowledge

Feedback Loop

Solving Problems with Visual Analytics (Keim, Kohlhammer, Ellis, Mansmann)
Activity Based Intelligence (ABI)
ANALYTICS IN ACTION
ABI: Defined

“ABI is a discipline of intelligence where the analysis and subsequent collection is focused on the activity and transactions associated with an area of interest”

“Activity Based Surveillance, built on a strong foundation of Human Dimension analysis, will form the intellectual underpinning of how we perform intelligence in the future.”

Honorable James Clapper, Undersecretary of Defense for Intelligence - 12/9/2010
ABI: Five Fundamental Objectives*

- Collect, characterize and locate activities and transactions
- Identify and locate actors and entities conducting the activities and transactions
- Identify and locate networks of actors
- Understand the relationships between networks
- Develop patterns of life

* “Surveillance Employment Strategy for Irregular Warfare”, Office of the Under Secretary of Defense for Intelligence, August 2010
145 Terabytes Per Hour

How to digest?
From “Big Data” to Signals
Progressing From $10^{14}$ to $10^2$ Data Sets

Activity Based Intelligence (ABI) Process
- Decrease Volume
- Minimize First-Order Importance
- Underlying Geospatial Datum
- Produce “Actionable” Value

- Recognize Tracks
  - Networks / Patterns-of-Life
  - Change over time
  - Kinematic Model – How “Dots” Move

- Heat Maps
  - Identifying events
  - Areas of Interest

- Multi-Source Data Sphere
  - Tracks
  - Tracklets
  - Entities
  - Events
  - Activities

- Movement Detections
  - Pixels
  - $10^{12} - 10^{14}$ Pixels

- Physical Mapping
  - $10^2$ Signals
From Data to Action...

...Balancing Science of Mathematical Models with Art of Visualization

Usama Fayyad “From Data Mining to Knowledge Discovery in Database” (1996)
Predictive Analytics

Maximizing science with art.
Thank You

Caron B. Kogan
caron.b.kogan@lmco.com
301.240.5588