From Insight to Action: Analytics from Both Sides of the Brain

Vaz Balasingham
Director of Solutions Consulting

vbalasin@tibco.com
Insight to Action – from Both Sides of the Brain

- Both Sides of the Brain
  - Fast & Slow
- Insight to Action
  - Visual Analytics
  - Numerical Algorithms
  - Insight Execution
- Wrap-Up / Questions
  - Come see the demos
Fun Problem from Kahneman

“A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?”
Themes: Thinking Fast and Slow... from Both Sides of the Brain

System 1: Association Engine

System 2: Monitor & Control

What you see is all there is

Lazy Controller

Thinking, Fast and Slow
Daniel Kahneman
Winner of the Nobel Prize in Economics
Themes: Thinking Fast and Slow... from Both Sides of the Brain

System 1: Association Engine

System 2: Monitor & Control

Spotfire

MATLAB
Themes: Insight without Action has Little Value

1. Event
2. Data Ready for Analysis
3. Analysis Completed
4. Decision Made
5. Action Taken
The TIBCO Insight Platform

Insights

EVENTS

Actions
The TIBCO Insight Platform

- Model
- Data
- Wrangle
- Analyze
- Events
The TIBCO Insight Platform

DATA

WRANGLE

ANALYZE

MODEL

Insights
The TIBCO Insight Platform

- Monitor
- Predict
- Decide
- Act
- Model
The TIBCO Insight Platform
The TIBCO Insight Platform

**Insights**

**Actions**

**MODEL**
The TIBCO Insight Platform

TIBCO Spotfire  
Matlab  
TIBCO Streambase

Insights  MODEL  Actions

TIBCO Live Datamart
Smart Visual Analytics

Be first to insight, first to action

Analytics Apps
Build and broadcast smart analytics

Streaming Analytics
Continuous algorithmic awareness and automation

Visual analytics is like a bicycle for your business mind.
Spotfire: Visual Analytics
Enhanced Data Connectivity

Connect Seamlessly with your Data Sources

- Hassle-free connectivity to your data sources
- Database drivers included with Spotfire software download
Recommendations

- Visualization recommendations based on your selected data
- Build one visualization – or entire dashboard – in just a couple of clicks
- Data previews accelerate discovery & insights
- Powered by Spotfire Analytics Recommendation Engine™

Jump Start Your Analysis
Gain a better understanding of the data you’re trying to visualize.

Speed up visualization design.

Visual Overview of data table sources & operations
Key Information At Your Fingertips

- See all dimensions and measure available to visualize.
- Expanded view gives extra insights without going to other screens.
- Change data type, category, sort order all without dropping what you’re doing.
- Data filters built-in.
- Views based on type of data in column.
Add Commentary To Your Analysis

- Overlay annotations on any part of your analysis.
- Select size, color, and font to match the look and feel of the rest of your visualizations.

We opened the Fremont Distribution Center this past September, which reduced our longest drive time by 23 minutes.
Connect Your Data To The Conversation

- Have conversations about specific issues and insights found in your data.
- Save your views and discoveries on each message in the conversation.
WHAT IS IT?

- High-accuracy, global base maps designed for data exploration & location analytics
- Visualize, explore, analyze data in the context of location
- Modern, interactive and easy map navigation
- Mash-in new data sources quickly and provide accurate enterprise geo-coding
- Use multi-layered analysis to understand geographic correlations in data and expand understanding

QUICK DETAILS

- Automatic, offline geo-coding
- Combine multiple layers of data on a same map
- Import & save geographic files for future use
- Cloud based base map always available, always up-to-date
- Drill-down to a deeper level of detail

FOR WHOM

- Analysts
- Data Scientist
- Business Users

MATURITY CURVE STEP

Diagnose (foundation for Predict & Optimize)

TIBCO GeoAnalytics
California Drought Conditions 2013 - 2015

Top Worsening Conditions:
- Alpine
- Amador
- Calaveras
- El Dorado
- Madera
- Mariposa
- Merced
- Nevada
- Placer
- Plumas
- San Joaquin
- Sierra
- Stanislaus
- Sutter
- Tuolumne
- Yuba

Top Improving:
- San Francisco
- Sonoma
- Mendocino
- Lake
- Santa Cruz
- Trinity
- Napa
- Shasta
- San Mateo
- Marin

Select counties here or in the list to the left to see details.

Yearly Breakdown:

- 11 2013:
  - Abnormal: 7%
  - Moderate: 99%
  - Severe: 1%
  - Extreme: 0%

- 11 2014:
  - Abnormal: 97%
  - Moderate: 2%
  - Severe: 0%
  - Extreme: 0%

- 11 2015:
  - Abnormal: 98%
  - Moderate: 2%
  - Severe: 0%
  - Extreme: 0%

Graphical representation of drought conditions across California for the years 2013 to 2015.
Analytics Apps
Build and broadcast smart analytics
Insight to Action: the Model

- PMML
  - R PMML (CRAN)
  - R2PMML (JPMML)
  - KNIME PMML
  - SAS PMML
  - ... many

- Matlab Model Object
  - MDS
  - PMML

- H2O
  - POJO
  - Model object

- R Model Object
  - *.rds
  - *.RData
  - *.mdl (TERR)
• Declarative & Heuristic Rules
• SPC and Anomaly Detection
• Machine Learning
  • Supervised
  • Unsupervised
  • Gradient Boosting Machines
  • Random Forests
  • Deep Learning
• Optimization
  • Linear & Quadratic Programming
  • Genetic Algorithms
  • Process optimization
  • Capacity constraints
Machine Learning finds predictive models in data without being told where to look

- **Supervised** – Solve known problems: $y=f(X)$
  - Build a model that predicts a condition (failure, success, ..)
  - What factors are driving customer network issues?
  - Decision Trees, Random Forests, Gradient Boosting Machines, Deep Learning

- **Unsupervised** – Identify new patterns, Detect anomalies X only
  - Are there new patterns or failure modes emerging?
  - Clustering, Principle Components, Deep Learning Autoencoder

- **Optimization** – Support Decision-making
  - Find best solution even when there are constraints on the process
  - What is the optimum allocation of resources for equipment maintenance?
  - Genetic Algorithm, Linear/Quadratic Programming
Anomaly Detection

Sensor Data / Anomaly Detection

Existence of known anomaly?

What if we don’t have this?

Unsupervised – no $y$ variable

Can still model! 😊

- SPC and Western Electric Rules (ESP)
- Single class Support Vector Machine
- Principal Components: PC Score
- Deep Learning: Auto encoding
Model: Gradient Boosting Machine

GBM Results

Predictor Importance - Effect on Yield

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 62</td>
<td>8</td>
</tr>
<tr>
<td>Mean 24</td>
<td>7</td>
</tr>
<tr>
<td>Mean 349</td>
<td>6</td>
</tr>
<tr>
<td>Mean 106</td>
<td>5</td>
</tr>
<tr>
<td>Mean 67</td>
<td>4</td>
</tr>
<tr>
<td>Mean 344</td>
<td>3</td>
</tr>
<tr>
<td>Mean 346</td>
<td>2</td>
</tr>
<tr>
<td>Mean 980</td>
<td>1</td>
</tr>
<tr>
<td>Mean 200</td>
<td>10</td>
</tr>
<tr>
<td>Mean 68</td>
<td>9</td>
</tr>
<tr>
<td>Mean 354</td>
<td>8</td>
</tr>
<tr>
<td>Mean 31</td>
<td>7</td>
</tr>
</tbody>
</table>

Predictor Interactions Summary Table

<table>
<thead>
<tr>
<th>Predictor 1</th>
<th>Predictor 2</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 62</td>
<td>Mean 24</td>
<td>0.20</td>
</tr>
<tr>
<td>Mean 349</td>
<td>Mean 67</td>
<td>0.65</td>
</tr>
<tr>
<td>Mean 24</td>
<td>Mean 106</td>
<td>0.03</td>
</tr>
<tr>
<td>Mean 24</td>
<td>Mean 980</td>
<td>0.05</td>
</tr>
<tr>
<td>Mean 24</td>
<td>Mean 67</td>
<td>0.08</td>
</tr>
<tr>
<td>Mean 106</td>
<td>Mean 67</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean 62</td>
<td>Mean 980</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean 62</td>
<td>Mean 349</td>
<td>0.01</td>
</tr>
<tr>
<td>Mean 62</td>
<td>Mean 106</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Heat Map Setup

Bin Continuous variables into N groups for display:

Eliminate small cells with unstable target average. Only show cells with n rows greater or equal to:

Predictor Effect on Yield Detail

Predictor Interactions Detail

Yield Loss (%) -
- Max (27.87)
- Average (6.78)
- Min (0.00)
Model: Gradient Boosting Machine

• Boosting Process
  • Sample the data: fit a tree: recursive partitioning
  • Drop the observations down the tree
  • Re-sample the data; up-weighting the observations that weren’t fitted well in previous model
  • Save all the trees and average them
  • Excellent fit + prediction
Reference Diagram
Increasing Capacity and Resilience

TIBCO Spotfire Web Player

MatWorks. MPSExtension

MATLAB Production Server

MATLAB Analytics

Load Balancer

TIBCO Spotfire Server

Mobile App

Spotfire Web

Spotfire Desktop
Spotfire & Matlab Demo
How does Spotfire Consume Data?

**In-Memory**
Load data from source into memory

**In-Database**
Leave data in DB Dynamically load and discard data to visualize

**On-Demand**
Dynamically swap data in and out of memory

**DATA SOURCES**
- RDBMS
- Spreadsheets
- Flat Files
- XML
- Hadoop & Big Data stores
- Analytical DWs e.g. Exadata
- Cubes
- Event Data Streams
- Active Spaces

© Copyright 2000-2013 TIBCO Software Inc.
Streaming Analytics
Continuous algorithmic awareness and automation
• Data collected from multiple sources and periodically placed in a persistent store.

• Analytical processes are executed against the stored data.

• Introduces too much “decision latency”.

• Responses are delivered “after-the-fact”.

• Decisions are made on old and stale data.

• Maximum value is lost.
The New Era: Fast Data Processing

- Events are analyzed and processed in **real-time** as they arrive.
- Decisions are **timely**, **contextual**, and based on **fresh data**.
- **Decision latency is eliminated**, resulting in
  - Superior Customer Experience
  - Operational Excellence
  - Instant Awareness and Timely Decisions
Anomaly Detection: Data Flow

Transactions
* 1 row / transaction detail

Matlab
Prepare Data for Modeling (Matlab, TERR)

“Featurize” Data
* 1 row / group

Spotfire
Dashboards
Anomaly Detection Models (TERR, Matlab)

Spotfire/Matlab
Anomaly Detection Models (TERR, Matlab)
- Autoencoder
- Gradient Boosting Machine

Anomaly Detection Model
* Rdata object

Streambase
Real Time Scoring

© Copyright 2000-2016 TIBCO Software Inc.
Streaming Analytics: Equipment Monitoring

- Load Reference Data to be used in Rules/Alerts, Cleaning, or Anywhere else
- Publish and Expose Clean Data to Whole Organization
- Continuously Publish Summary Statistics for Analysis
- Publish Raw Events to another StreamBase Workflow or other Applications
- Continuously Build Features and Publish
- Output to BPM Systems, Web Services, Databases, and anywhere else!
StreamBase Adapters and Connectivity

**Financial**
- Activ Financial
- Alpha
- Autobahn
- Bloomberg
- BM&F BOVESPA
- Inetco
- Kitco
- MarketFactory
- Nomura
- Thomson Reuters
- ICAP
- FXcm
- GTX
- UBS

**Databases**
- IBM
- Microsoft
- MySQL
- Oracle
- Sybase
- HP
- Vertica
- Big Data
- Hadoop
- Vertica

**IoT: Integration**
- OSI-PI
- WITSML
- MQTT

**Messages**
- Kafka
- Flume
- Rabbit MQ

**API's**
- C++, Java, .NET, JavaScript, Python

**Messaging**
- TIBCO RV & EMS, JMS, MQ Series, RMDS, Solace, Tervela, Wombat

**General Purpose Adapters**
- FIX, STAMP, JDBC, ActiveSpaces, POP3, IRC, HTTP, SMTP, IM, XML, RSS

© Copyright 2000-2016 TIBCO Software Inc.
Streambase & Matlab Demo
Learn more about TIBCO

• **www.tibco.com**

• **Accelerators & Community**
  - [https://community.tibco.com/](https://community.tibco.com/)
  - [https://community.tibco.com/wiki/accelerators](https://community.tibco.com/wiki/accelerators)

• **Spotfire micro site**
  - [http://spotfire.tibco.com/](http://spotfire.tibco.com/)
  - [http://spotfire.tibco.com/demos](http://spotfire.tibco.com/demos)
INTERCONNECT EVERYTHING

People & Processes

APIs

Data & Systems

Reporting & Dashboards

Streaming Analytics

Data Visualizations

AUGMENT INTELLIGENCE