MATLAB EXPO 2016
Get More From Your Data with Data Analytics

Dr. Roland Michaely
Applications Engineer
What do we have to work with?
Buildings Have Thermodynamic Properties

\[ \frac{\partial u}{\partial t} - \alpha \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = 0 \]
Temperatures change
Electricity demand varies
Humans have comfort bounds
\[ \frac{\partial u}{\partial t} - \alpha \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = 0 \]
BuildingIQ Develops Proactive Algorithms for HVAC Energy Optimization in Large-Scale Buildings

Office buildings, hospitals, and other large-scale commercial buildings account for about 30% of the energy consumed worldwide. The heating, ventilation, and air-conditioning (HVAC) systems in these buildings are often inefficient because they do not take into account changing weather patterns, variable energy costs, or the building's thermal properties.

BuildingIQ has developed Predictive Energy Optimization™ (PEO), a cloud-based software platform that reduces HVAC energy consumption by 10–25% during normal operation. PEO was developed in cooperation with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), HVAC pressure sensors, as well as weather and energy cost data. A single building often produces billions of data points, and the scientists and engineers needed tools for efficiently filtering, processing, and visualizing this data.

To run their optimization algorithms, the scientists and engineers had to create an accurate mathematical model of a building's thermal and power dynamics. The algorithms would use this calculated model to run constrained optimizations that maintained occupant comfort while minimizing energy costs.

BuildingIQ needed a way to rapidly develop mathematical models, test optimization algorithms...
Aspects of Data Analytics Applications

1. Diverse and/or Big Data

2. Advanced Algorithms

3. Deployment
Why MATLAB?

1. Analytics that increasingly require **both business and engineering data**

   DATA
   - Engineering, Scientific, and Field
   - Business and Transactional

2. Enable **Domain Experts to be Data Scientists**

3. Develop **embedded systems** with analytics powered functionality

4. Develop analytics to run on **both enterprise and embedded platforms**

   - **Embedded Systems**
   - **Business Systems**
Why MATLAB?

1. Analytics that increasingly require **both business and engineering data**

2. Enable **Domain Experts to be Data Scientists**

3. Develop **embedded systems** with analytics powered functionality

4. Develop analytics to run on **both enterprise and embedded platforms**

**DATA**
- Engineering, Scientific, and Field
- Business and Transactional

**Smarter Embedded Systems**

**Business Systems**
“No matter what industry our client is in, and no matter what data they ask us to analyze—text, audio, images, or video—MATLAB enables us to provide clear results faster.”

Dr. G Subrahmanya VRK Rao, Cognizant
# Business and Engineering Data

## Business and Transactional Data

**Repositories**
- Databases (SQL)
- NoSQL
- Hadoop

**File I/O**
- Text
- Spreadsheet
- XML

## Engineering, Scientific, and Field Data

**Web Sources**
- HTML
- Mapping
- Financial datafeeds
- RESTful
- JSON

**File I/O**
- Text
- Spreadsheet
- XML
- CDF/HDF
- Image
- Audio
- Video
- Geospatial

**Communication Protocols**
- CAN (Controller Area Network)
- DDS (Data Distribution Service)
- OPC (OLE for Process Control)
- XCP (eXplicit Control Protocol)

**Real-Time Sources**
- Sensors
- GPS
- Instrumentation
- Cameras
- Communication systems
- Machines (embedded systems)
Accessing Data

Load the data and create feature vectors:

```matlab
trainingData = []; % Define the training data

while hasdata(ds)
    % Read the next data file
    rawdata = read(ds);

    % Summary Statistics
    varMeans = varfun(@mean, rawdata(:,1:6));
    varStdSts = varfun(@std, rawdata(:,1:6));

    % Signal Processing
    varRMS = varfun(@rms, rawdata(:,1:6));
    varMeanFreq = varfun(@meanfreq, rawdata(:,1:6));

    % Activity
    activity = rawdata(1:end);

    % Add preprocessed datapoint to training table
    trainingData = [trainingData; varMeans varStdSts varRMS varMeanFreq activity];
end
```

`ds` is a `TabularTextDatastore` with the following properties:

- **Files**: A list of file paths where each file contains data.
- **FileEncoding**: 'UTF-8'
- **ReadVariableNames**: 'true'
- **VariableNames**: acc_x, acc_y, acc_z
- **Text Format Properties**:
  - **NumHeaderLines**: 0
  - **Delimiter**: ' ' (space)
  - **RowDelimiter**: '\n' (new line)
  - **TreatAsMissing**: 'NaN'
  - **MissingValue**: NaN
  - **TextscanFormats**: '%f', '%f', '%f', '%f', '%f', '%f', '%f'...
  - **ExponentCharacters**: 'eE0'
  - **CommentStyle**: '
'
  - **Whitespace**: '\t'
  - **MultipleDelimitersAsOne**: false

Properties that control the table returned by `preview`, `read`, `readS`:
- **SelectedVariableNames**: acc_x, acc_y, acc_z, acc_z...
- **SelectedFormats**: '%f', '%f', '%f', '%f', '%f'...
- **ReadSize**: 20000 rows
### Table

<table>
<thead>
<tr>
<th>first_name</th>
<th>emp_id</th>
<th>mgr_id</th>
<th>photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Seth'</td>
<td>4867</td>
<td>4324</td>
<td>[130x130x3 uin]</td>
</tr>
</tbody>
</table>

```matlab
>> imshow(s.photo{1})
```
Preprocessing Data

```
% Preprocessing data

% Remove noise
filterWindow = 5;
for i = 1:5
    b = (1/filterWindow)*ones(1,filterWindow);
    a = 1;
end

% Filter sensor data
smoothData = sensorData;
smoothData(:,3:end) = filter(b,a,sensorData(:,3:end));
smoothData(:,1:filterWindow) = [];

% Plot smoothed data
figure
for ii = 1:9
    subplot(3,3,ii)
    plot(smoothData.Time,smoothData(:,1:ii),
         'Linewidth',2)
    hold on
    plot(sensorData.Time,sensorData(:,1:ii),
         'Linewidth',3)
    title(smoothData.Properties.VariableNames(1:ii))
    xlabel('Time')
    xlim([0,125])
end
```
Signal Processing

- cheby2
- filter
- rms
- pwelch
- periodogram
- xcov
- findpeaks
- movmean
- movstd
- ...

\[
|H(j\omega)|^2 = \frac{e^{2} C_2^2(\omega / \omega_0)}{1 + e^{2} C_2^2(\omega / \omega_0)}
\]

\[
C_2(\omega_1, \omega_2) = \begin{cases} 
\cos[N \cos^{-1}(\omega_1 / \omega_2)] , & |\omega_1| \geq \omega_2 \\
\cosh[N \cosh^{-1}(\omega_1 / \omega_2)] , & |\omega_1| < \omega_2 
\end{cases}
\]

\[
\rho_x(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho_x(\omega) e^{i2\pi f \omega} d\omega
\]

\[
\phi_k = \frac{kF_s}{N}, \quad k = 0, 1, ..., N
\]

\[
\psi_{mn}(n) = \begin{cases} 
\sum_{k=0}^{N-1} x(n + m) - \frac{1}{N} \sum_{k=0}^{N-1} x_k, & m > 0 \\
\sum_{k=-m}^{m-1} x_k - \frac{1}{N} \sum_{k=-N}^{N-1} x_k, & m < 0 
\end{cases}
\]
**Image Processing**

Perform image processing, analysis, visualization, and algorithm development

- Image display and exploration
- Image enhancement
- Image analysis
- Morphological operations
- Image registration
- Geometric transformation
- ROI-based processing
# Feature Engineering – Extracting Information from Data

<table>
<thead>
<tr>
<th>Data type</th>
<th>Common Techniques for Deriving Features</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensor data</strong></td>
<td>• Peak analysis</td>
<td><img src="image" alt="Sensor data example" /></td>
</tr>
<tr>
<td></td>
<td>• Pulse and transition metrics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Spectral measurements (power, bandwidth, mean frequency, median frequency)</td>
<td></td>
</tr>
<tr>
<td><strong>Image and video data</strong></td>
<td>• Bag of visual words</td>
<td><img src="image" alt="Image and video data example" /></td>
</tr>
<tr>
<td></td>
<td>• HOG (Histogram of Oriented Gradients)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Minimum Eigenvalue algorithm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Local feature descriptors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Edge detection</td>
<td></td>
</tr>
<tr>
<td><strong>Transactional data</strong></td>
<td>• Decomposing timestamps into components (day, month, year, day of the week, etc.)</td>
<td><img src="image" alt="Transactional data example" /></td>
</tr>
<tr>
<td></td>
<td>• Calculation of aggregate values</td>
<td></td>
</tr>
</tbody>
</table>
Why MATLAB?

1. Analytics that increasingly require both business and engineering data

DATA
- Engineering, Scientific, and Field
- Business and Transactional

2. Enable Domain Experts to be Data Scientists

3. Develop embedded systems with analytics powered functionality

4. Develop analytics to run on both enterprise and embedded platforms

Smarter Embedded Systems

Data Scientist

Business Systems
“MATLAB has helped accelerate our R&D and deployment with its robust numerical algorithms, extensive visualization and analytics tools, reliable optimization routines, support for object-oriented programming, and ability to run in the cloud with our production Java applications.”

Borislav Savkovic, BuildingIQ
Enabling Domain Experts to be Data Scientists
Apps - Classification Learner app
Object Recognition using Deep Learning

<table>
<thead>
<tr>
<th>Training (using GPU)</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Millions of images from 1000 different categories</td>
<td>Real-time object recognition using a webcam connected to a laptop</td>
</tr>
</tbody>
</table>

MathWorks Demo: Deep Learning for Computer Vision
Why MATLAB?

1. Analytics that increasingly require **both business and engineering data**
   
   **DATA**
   - Engineering, Scientific, and Field
   - Business and Transactional

3. Develop **embedded systems** with analytics powered functionality

2. Enable **Domain Experts to be Data Scientists**

4. Develop analytics to run on **both enterprise and embedded platforms**

Smarter Embedded Systems

Data Scientist

Business Systems
Smarter Embedded Systems

**RESEARCH**

**REQUIREMENTS**

**DESIGN**

- Environment Models
- Physical Components
- Algorithms

**IMPLEMENTATION**

- C, C++
- VHDL, Verilog
- Structured Text
- MCU
- DSP
- FPGA
- ASIC
- PLC

**TEST AND VERIFICATION**

**INTEGRATION**

- Airbus
  - Battery management
- GM
  - Climate control
- Festo
  - Industrial robots
- Sonova
  - Hearing implants
- Weinmann
  - Transport ventilator
- manroland
  - Printing presses
- FLIR
  - Thermal imaging
- ABB
  - Smart Grid controller
- Daimler
  - Cruise controller
function [y1] = netCodeGen(x1)

%NETCODEGEN neural network simulation function.

% Generated by Neural Network Toolbox function genFunction, 07-Mar-2016 15:51:16

% [y1] = netCodeGen(x1) takes these arguments:
% x = 60xQ matrix, input #1
% and returns:
% y = 5xQ matrix, output #1
% where Q is the number of samples.

%#ok<*RPMT0>

% ===== NEURAL NETWORK CONSTANTS =====

% Input 1
C Code Generation

```
/*
 * Trial License - for use to evaluate programs for possible purchase
 * an end-user only.
 * File: netCodeGen.c
 */

/* MATLAB Coder version : 3.1
 * C/C++ source code generated on : 03-Mar-2016 17:43:43
 */

/* Include Files */

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>double</td>
<td>1x60</td>
</tr>
<tr>
<td>y1</td>
<td>double</td>
<td>5x60</td>
</tr>
<tr>
<td>x1_step1</td>
<td>struct</td>
<td>1x1</td>
</tr>
<tr>
<td>xoffset</td>
<td>double</td>
<td>60x1</td>
</tr>
</tbody>
</table>
```
Why MATLAB?

1. Analytics that increasingly require both business and engineering data

2. Enable Domain Experts to be Data Scientists

3. Develop embedded systems with analytics powered functionality

4. Develop analytics to run on both enterprise and embedded platforms

DATA
- Engineering, Scientific, and Field
- Business and Transactional

Smarter Embedded Systems

Business Systems
## Where Does the Processing Happen?

<table>
<thead>
<tr>
<th>Devices</th>
<th>Business Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visibility</td>
<td>Self, Maybe Neighbors</td>
</tr>
<tr>
<td>Power</td>
<td>Battery (Low)</td>
</tr>
<tr>
<td>Processing</td>
<td>Microprocessor (Low)</td>
</tr>
<tr>
<td>Storage</td>
<td>Memory (Minimum Buffer)</td>
</tr>
<tr>
<td>Updates</td>
<td>Difficult</td>
</tr>
</tbody>
</table>
Why not Transfer All the Data?

1. Data privacy concerns
2. Cost of network/storage
3. Power required to transmit data from device (for wireless)
4. Response time
Splitting Computation
Enterprise Integration – Forecasting Model

Select Zone

Forecast

Plattsburgh International Airport
Station ID: KPSG
Contributes to zones: NYISO D-North (100%), NYISO F-Capri (5%)
Weather forecast

Comparision

Sharing MATLAB Based Applications and Algorithms

MATLAB EXPO 2016
MATLAB Differentiators

1. Analytics that increasingly require both business and engineering data

DATA
- Engineering, Scientific, and Field
- Business and Transactional

2. Enable Domain Experts to be Data Scientists

3. Develop embedded systems with analytics powered functionality

4. Develop analytics to run on both enterprise and embedded platforms

Smarter Embedded Systems

Business Systems

Data Scientist