Machine Learning and Pipeline Replacement Prioritization

Clifford Chan
David Katzev

June 1, 2018
EBMUD System & Service Area

Customers
- 1,400,000 customers

Raw Water System
- 2 upcountry reservoirs
- 5 local reservoirs

Treatment System
- 6 water treatment plants

Distribution System
- 4,200 miles of pipeline
- 122 pressure zones
- 164 reservoirs
- 135 pumping plants
- 100 regulators/RCS
Addressing Real Water Loss

- Active leakage control
- Pressure management
- Speed and quality of repairs
- Infrastructure management
Pipeline Inventory

- Steel 26% (1300 miles)
- Asbestos Cement 30% (1100 miles)
- Cast Iron 35% (1300 miles)
- PVC 9%
Pipeline Break History

Leaks by Material Per Month (1991-2016)

- Cast Iron (12-month Avg)
- AC (12-month Avg)
- Steel (12-month Avg)
- Total Cast Iron Breaks
- Total AC Breaks
- Total Steel Breaks

Number of Leaks per Month

Jan-91 to Jan-16
Infrastructure Renewal Program -- Ramping Up Replacements

Industry Benchmarks for a well maintained system

= < 15-30 leaks/100 miles/year

How do we make sure we are replacing the right pipe?
Prior Pipeline Replacement Model

Cost Benefit Ratio = \frac{\text{Repair Costs}}{\text{Replacement Costs}}

> 1.0 , more beneficial to replace pipe
New Pipe Replacement Risk Model

**LOF**

RISK = (Likelihood of Failure) \times (Consequence of Failure)

The probability a pipeline will leak.

**COF**

The resulting magnitude of consequence if the pipe does leak.
### Pipe Replacement Risk Model

**Likelihood of Failure**

<table>
<thead>
<tr>
<th>Consequence of Failure</th>
<th>Very Low</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Pipe Leak History
- Pipe Age

- Creek crossing
- Diameter
- Consumption
- Access
- Slope
- Backbone/Critical
- Highway crossing
- Railroad crossing
How we choose projects using the Risk Model
New Decision-Making Tools

Investment in Infrastructure

Need for innovation

Smarter selection of pipes

Limited resources
Data Never Sleeps
Operational Data

• SCADA
  - Enterprise database: ~150 million readings per year
  - Data historian: ~7 billion readings in the database

• What’s Next
  - Pipeline data
  - Maintenance data
  - AMI data
  - Pressure and leak detection data
The secret of change is to focus all of your energy, not on fighting the old, but on building the new.
Machine Learning
How It Works?

1. Wrangle and import water main and geographic data
2. Layer geodata and run machine learning algorithms
3. Visualize vulnerabilities and apply LOF results
Step 1: Wrangle and Import Data

- Cleaning and normalizing the data
- Correcting wrong/outlier data points and filling in missing values
- Geocoding breaks with pipe segments

Example:
Received Pipe Data as ESRI Shapefile
23876 Pipe Segments
~15% Total Missing/Error

Received Break data as two Excel sheets
+ GIS file
  Clean
  Correct
  Geocode and assign to pipes
Step 2: Machine Learning Analysis

**Utility Data**
- Pipe Parameters
  - Length, Material, Diameter, Install Year, Pressure etc.
- Break History
  - Break info assigned to pipe segment

**Geo Variables**
- Environment
  - 40+ Soil Properties from USGS
- Location
  - Slope, Elevation
  - Proximity to Transportation Features (roads, rail etc.)
  - Proximity to Water (salt water, river etc.)
- Population and Buildings
  - Population, Zoning, Buildings etc.

**Variables derived from Utility Data**
- Age-Derived Variables
- Pipe and Leak Density

**Variables derived from Geo Data**
- Min/Max/Mean Distance
- Density of Soil Type Changes

<table>
<thead>
<tr>
<th>Utility variables</th>
<th>Derived Utility variables</th>
<th>Geo variables</th>
<th>Derived Geo variables</th>
</tr>
</thead>
</table>

1000+ variables
Step 2: Layer Geodata and Run Machine Learn

- Process in a repeatable and scalable manner
- Layer additional variables
- Look for correlations
- Calculate 5-year LOF probability scores
How to Measure LOF Accuracy

Prediction

Actual

- RANK 5
- RANK 4
- RANK 3
- RANK 2
- RANK 1

Main Break
# Measuring Accuracy

Used historical data, 1990 → 2011, to predict next 5 years LOF

<table>
<thead>
<tr>
<th>Fracta LOF Focused Ranking</th>
<th>Rank F5</th>
<th>Rank F4</th>
<th>Rank F3</th>
<th>Rank F2</th>
<th>Rank F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>41.6 miles (1%)</td>
<td>166.3 miles (4%)</td>
<td>415.8 miles (10%)</td>
<td>831.8 miles (20%)</td>
<td>2702.9 miles (65%)</td>
</tr>
<tr>
<td>Pipe Segments</td>
<td>979</td>
<td>2692</td>
<td>7151</td>
<td>17462</td>
<td>76245</td>
</tr>
<tr>
<td>5-yr forecasted breaks</td>
<td>397</td>
<td>901</td>
<td>1160</td>
<td>1251</td>
<td>1183</td>
</tr>
<tr>
<td>5-yr forecasted break rate</td>
<td>191.0</td>
<td>108.3</td>
<td>55.8</td>
<td>30.1</td>
<td>8.8</td>
</tr>
</tbody>
</table>
Comparing to Actual Events
Break Data (2012-2016)

<table>
<thead>
<tr>
<th>Fracta LOF Focused Ranking</th>
<th>Rank F5</th>
<th>Rank F4</th>
<th>Rank F3</th>
<th>Rank F2</th>
<th>Rank F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>41.6 miles (1%)</td>
<td>166.3 miles (4%)</td>
<td>415.8 miles (10%)</td>
<td>831.8 miles (20%)</td>
<td>2702.9 miles (65%)</td>
</tr>
<tr>
<td>Pipe Segments</td>
<td>979</td>
<td>2692</td>
<td>7151</td>
<td>17462</td>
<td>76245</td>
</tr>
<tr>
<td>5-yr forecasted breaks</td>
<td>397</td>
<td>901</td>
<td>1160</td>
<td>1251</td>
<td>1183</td>
</tr>
<tr>
<td>5-yr forecasted break rate (x/100mi/yr)</td>
<td>191.0</td>
<td>108.3</td>
<td>55.8</td>
<td>30.1</td>
<td>8.8</td>
</tr>
<tr>
<td>Actual Breaks 2012-2016</td>
<td>436</td>
<td>991</td>
<td>1177</td>
<td>1038</td>
<td>838</td>
</tr>
<tr>
<td>Actual Break Rate 2012-2016</td>
<td>209.5</td>
<td>119.2</td>
<td>56.6</td>
<td>25.0</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Broad Correlation of Projected & Actual break rates
Demo & Questions