Advanced Analytics for the Internet of Things

Rosaria Silipo
KNIME.com AG
Rosaria.silipo@knime.com
IoT

THE INTERNET OF THINGS
AN EXPLOSION OF CONNECTED POSSIBILITY

Industry

Energy

City

Wearables

Health

Household
The Challenges

• Handling very large amounts of data created over time

• Forcing sensor-equipped objects (house or city) to learn, and therefore to become smarter
IoT

Use Public Data Please....
Use Cases

Energy Smart Meters
- Energy Consumption Prediction
- Energy Profiling

Smart Cities
- Re-stocking Strategies
- Geo-localization and Traffic Prediction

Manufacture Industry
- Anomaly Detection
This Use Case
This Use Case

Capital Bikeshare in Washington DC
What Is Capital Bikeshare?

Capital Bikeshare puts over 2500 bicycles at your fingertips. You can choose any of the over 300 stations across Washington, D.C., Arlington and Alexandria, VA and Montgomery County, MD and return it to any station near your destination. Check out a bike for your trip to work, Metro, run errands, go shopping, or visit friends and family. Join Capital Bikeshare for a day, 3 days, a month, a year or try our new daily key option, and have access to our fleet of bikes 24 hours a day, 365 days a year. The first 30 minutes of each trip are free. Each additional 30 minutes incurs an additional fee.
Sensors!
The Business Challenge:
Even MORE of a Business Challenge

• Any Station without bikes for 1 hour:  
  $XXXX Per Violation

• Any Station with no free slots for 1 hour:  
  $XXXX Per Violation
Capital Bikeshare Response

Over 3 years

307 Stations
2963 Bikes
19.4% Casual Bikers
5.9m Bike Moves
Advanced Analytics for the Internet of Things

Pre-processing

Data Visualization

Predictive Analytics

- Read Bike
  - Read all the Raw Bike Transaction Records

- Enrich
  - Enrich with new data

- Expand
  - Expand from bikes to stations and paths

- Investigate
  - various Investigations

- Drill Down
  - Drill down and explore interesting cases

- Predict
The KNIME Platform: Open for Innovation

- **Powerful:** Legacy ↔ Future Tools
- **Collaborative:** Scientists ↔ Analysts
- **Integrative:** Legacy ↔ Future Data
- **Transparent:** Existing ↔ Future Expertise
- **Agile:** Internal ↔ External Wisdom
The KNIME Analytics Platform
From Access to Visualization and Deployment
Gartner Recognition

Source: Gartner (February 2015)
Data Pre-processing
Reading all Sensor Data
Enrich

Topology / Elevations

Weather

Holiday Schedules

<table>
<thead>
<tr>
<th>HOLIDAY</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Year's Day</td>
<td>Tuesday, January 1</td>
<td>Wednesday, January 1</td>
<td>Thursday, January 1</td>
</tr>
<tr>
<td>Martin Luther King Jr.</td>
<td>Monday, January 21</td>
<td>Monday, January 20</td>
<td>Monday, January 19</td>
</tr>
<tr>
<td>Memorial Day</td>
<td>Monday, May 27</td>
<td>Monday, May 26</td>
<td>Monday, May 25</td>
</tr>
<tr>
<td>Independence Day</td>
<td>Thursday, July 4</td>
<td>Friday, July 4</td>
<td>Friday, July 3</td>
</tr>
<tr>
<td>Labor Day</td>
<td>Monday, September 2</td>
<td>Monday, September 1</td>
<td>Monday, September 7</td>
</tr>
<tr>
<td>Thanksgiving</td>
<td>Thursday, November 28</td>
<td>Friday, November 29</td>
<td>Thursday, November 27</td>
</tr>
<tr>
<td>Christmas</td>
<td>Tuesday, December 24</td>
<td>Wednesday, December 24</td>
<td>Thursday, December 24</td>
</tr>
<tr>
<td>New Year's Eve</td>
<td>Tuesday, December 31</td>
<td>Wednesday, December 21</td>
<td>Thursday, December 31</td>
</tr>
</tbody>
</table>
Elevation from Google API

Connecting to **Google API** and other **REST** services available on the Web

Use REST GET Resources to retrieve elevation

![Diagram of KNIME workflow with Google API integration](image_url)

**Table: Elevation Data**

<table>
<thead>
<tr>
<th>Row ID</th>
<th>URL</th>
<th>lat</th>
<th>lon</th>
<th>elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row0#0</td>
<td><a href="http://maps.googleapis.com/maps/api/elevation.json?locations=38.856">http://maps.googleapis.com/maps/api/elevation.json?locations=38.856</a>, -77.051 &amp; sensor=false</td>
<td>38.856</td>
<td>-77.051</td>
<td>15.476</td>
</tr>
</tbody>
</table>
Expand/Transform
Data Visualization
Station and Bike Facts

Over 3 years

307 Stations
2963 Bikes
19.4% Casual Bikers

5.9m Bike Moves
Stations: deficits and surpluses

16th and Harvard

Bike sources

Bike Sinks

Union Station
Top 250 Routes

- Urban/Students
- Dupont Circle
- Tourists
- Suburban
- Union Station
Total Bikers Number per Hour

- Registered (blue) vs. Casual (red) Bikers
The Goal

- Restocking alert signal

- 1 hour warning! $\text{Lag(flag-1)}$
Input Features

- Weather related features
- Number of registered and casual people showing up
- Station infos (name and max. number of docks)
- Calendar infos (working day, holiday, date)
- Past infos
- Number of bikes added and removed at each hour
- Adjusted cumulative sum = number of bikes available at the station at a given hour
- Bike ratio = adjusted cumulative sum/total docks available
Restocking Alert System

78% accuracy
Feature Elimination Loop

Two options:

1. Use all the input features (no thinking required, just a powerful machine)
2. Select the most useful input features via the “Feature Elimination” loop

At each step, one input feature is removed - i.e. the input feature whose removal produces the smallest error increase.
Input Attribute Impact

Dialog - 0:43:2 - Backward Feature Elimination Filter

- Column Selection
  - Include target column
  - Select features manually
  - Select features automatically by error threshold
    Prediction error threshold: 0.5

- Flow Variables
  - Error
    - 0.163
    - 0.183
  - Nr. of features
    - 3

- Memory Policy
  - season
  - month
  - hour
  - holiday
  - weekday
  - workingday
  - weather
  - temp
  - atemp
  - humidity
  - windspeed
  - casual
  - registered
  - year
  - day of month
  - day of year
  - Terminal
  - station
  - day
  - Flag(-1)
  - bike ratio
    - bike ratio(-1)
    - bike ratio(-2)
    - bike ratio(-3)
    - bike ratio(-4)
    - bike ratio(-5)
    - bike ratio(-6)
    - bike ratio(-7)
    - bike ratio(-8)
The input feature subset with the smallest error (81% accuracy):

- Hour of the day
- Working day (Y/N)
- Current Bike Ratio
- Terminal (station code)

Past of bike ratio and weather infos do not seem to be relevant!
Is this because most bikers are registered members?
Lessons Learned

• Enrich / Expand
  – KNIME Transformations
  – REST calls to external sources

• Explore
  – Visualization with Open Street Map integration
  – Network Analysis (Graph Visualization)

• Prediction
  – (Lean) Restocking Alert System
  – Weather influence does not seem to be important!
Where can I find all this?


White paper, Workflows, and Data is available on the KNIME web site:
http://www.knime.com/white-papers
Resources

• **KNIME** ([www.knime.org](http://www.knime.org))
  • **BLOG** for news, tips and tricks([www.knime.org/blog](http://www.knime.org/blog))
  • **FORUM** for questions and answers ([tech.knime.org/forum](http://tech.knime.org/forum))
  • **EXAMPLE SERVER** for example workflows
  • **LEARNING HUB** ([www.knime.org/learning-hub](http://www.knime.org/learning-hub))

• **KNIME TV** channel on ![YouTube](http://www.youtube.com)

• **KNIME** on ![Twitter](http://twitter.com) @KNIME

• **KNIME** on ![Facebook](http://facebook.com) https://www.facebook.com/KNIMEanalytics
Thank You

Free Copy of KNIME Beginner’s Luck Book at KNIME Press

https://www.knime.org/knimepress

Promotion Code: OSCON_KNIME