Auto-tagging with machine learning

Lucy Kidd
About me

Lucy Kidd
Data Scientist at Bueno
Our product
Our product

Haystack tagging library is used to:

- Map each site to a common model
- Represent the complex relationships between point and equips
- Give meaning to a range of building systems
Root cause analytics: Analytics which identify as close as possible to the root cause of an issue, without physical inspection of the equipment.
Challenges of tagging

- Each building has a lot of points that need tagging
- Labor intensive
- Human error
- Quality assurance
- Points getting more complex as we create new tags
Solution: Unicorn

• Learns context for tags
• Suggests tags per point
• Deployments can easily accept or decline tagging suggestions
How Unicron Learns

Raw Data
- Tagging guide
- All data points

Clean Data
- Relevant points
- Meaningful strings

Feature pool
- Features of each point

Model
- Features mapped to tags
How Unicron Learns

Point data

<table>
<thead>
<tr>
<th>Name</th>
<th>ID</th>
<th>Temp</th>
<th>Sensor</th>
<th>untagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZT_SP</td>
<td>alc-b-55gi</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Zone-Temp</td>
<td>alc-b-56k3</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>sound_peak</td>
<td>alc-b-48ia</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Tagging guide

<table>
<thead>
<tr>
<th>Point</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone Temp</td>
<td>zone, air, temp, sensor</td>
</tr>
<tr>
<td>Zone Temp</td>
<td>zone, air, temp, sp</td>
</tr>
<tr>
<td>SP</td>
<td>zone, air, temp, sp</td>
</tr>
<tr>
<td>HHWV</td>
<td>heat, cmd</td>
</tr>
</tbody>
</table>
### How Unicron Learns

#### Raw Data

#### Clean Data

#### Feature pool

#### Model

### Relevant points

<table>
<thead>
<tr>
<th>Name</th>
<th>ID</th>
<th>temp</th>
<th>sensor</th>
<th>sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZT_SP</td>
<td>alcbl-55gi</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zone-Temp</td>
<td>alcbl-56k3</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>ZoneTempSp</td>
<td>alcbl-48ia</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

### Meaningful strings

- zone
- temp
- sp
- zt
How Unicron Learns

Raw Data

Clean Data

Feature pool

Model

Features for point 1
ID: a1cb-55gi
Name: ZT_SP
Entropy: 0
Tags: air,temp,zone,sp
Features:
- Equip: VAV
- Unit: °C
- Kind: number
- zt,sp

Features for point 2
ID: a1cb-56k3
Name: Zone-Temp
Entropy: 0
Tags: air,temp,sensor,zone
Features:
- Equip: zone
- Unit: °C
- Kind: number
- zone,temp
How Unicron Learns

Raw Data
- **Features:**
  - Equip: VAV
  - Unit: °C
  - Kind: number
  - zt, sp
- **Tags:** air, temp, zone, sp

Clean Data
- **Features:**
  - Equip: zone
  - Unit: °C
  - Kind: number
  - zone, temp
- **Tags:** air, temp, sensor, zone

Feature pool

Model
- **Features:**
  - Equip: ahu
  - Unit: %
  - Kind: number
  - hwvlv, cmd
- **Tags:** cmd, heat
Matching points to tag sets

Data collected previously

Raw Data

Features

Suggestions

New points

Features of new points

Tags for each point

Meaningful strings

Model
Matching points to tag sets

Raw Data

New point 1
ID: a1cb-7832
Name: ZT_SP
EquipRef: AHU_1
hisEnd: 2019-02-06
Unit: °C
Kind: Number

New point 2
ID: a1cb-88gy
Name: HWVLV_CMD
EquipRef: AHU_1
hisEnd: 2019-02-06
Unit: %
Kind: Number
Matching points to tag sets

Raw Data

New point 1
ID: a1cb-7832
Name: ZT_SP
Entropy: 0
Features:
- Equip: AHU
- Unit: °C
- Kind: number
- zt, sp

Features

New point 2
ID: a1cb-88gy
Name: HWVLV_CMD
Entropy: 0
Features:
- Equip: AHU
- Unit: %
- Kind: number
- hwv1v, cmd

Suggestions
Matching points to tag sets

New point 1
Name: ZT_SP
Tags: air, sp, temp, zone
Confidence: 1
Features:
- Equip: AHU
- Unit: °C
- Kind: number
- zt, sp

New point 2
Name: HWVLV_CMD
Tags: heat, cmd
Confidence: 1
Features:
- Equip: AHU
- Unit: %
- Kind: number
- hwvlv, cmd
Matching points to tag sets

Raw Data: Proportion of point name matched

Features: Normalised Levenstein distance between feature sets

Suggestions:

New point 2
Name: HWVLV_CMD
Tags: heat, cmd
Confidence: 1
Features:
- Equip: AHU
- Unit: %
- Kind: number
- hwvlv, cmd
Matching points to tag sets

Raw Data

Features

Suggestions

Proportion of point name matched

Normalised Levenstein distance between feature sets

New point 3

Name: rmtmpadj
Tags: air,sensor,temp,zone
Confidence: 0.69
Features:
- Equip: VAV
- Unit: °C
- Kind: number
- rmt
How deployments use it

- Deployments tell bot what site they want to tag
- Bot delivers a suggestion
- Deployments will accept (or decline) suggestions
- Points on the site will be tagged with all accepted tags

### rmtmpadj
- **Tag:** air, sensor, temp, zone
- **Confidence:** 0.69
- **Features:**
  - Equip: VAV
  - Unit: °C
  - rmt
- **Members:** 1
- **Accept:** FALSE

### HWVLV_CMD
- **Tag:** heat, cmd
- **Confidence:** 1
- **Features:**
  - Equip: AHU
  - Unit: %
  - hwvlv, cmd
- **Members:** 103
- **Accept:** TRUE
Architecture
• 60% less bugs in our analytics due to tagging issues
• Deployment time spent tagging has dropped by 50%
• Cheaper barrier to entry for clients
In Summary

- High quality tagging is essential for our analytics
- Manually tagging sites is unsustainable as we scale
- Automated tagging has allowed us to:
  - Speed up the tagging process
  - Improve the quality of our analytics
- Future direction:
  - Apply this to equips
Thank you