WHY TESTING AUTOMATION IS THE PERFECT DOMAIN FOR MACHINE LEARNING

Presented by Tamas Cser
AI – Man vs Machine
Machine Learning Opportunities
Current State

72%
AI in testing automation

**Human Intelligence**
Good for abstract feature identification, bad at scale.

**Machine Intelligence**
Great at scale, learning anomalies.
### Detect Anomalies in Large Dynamic Data

**Week 1:**
- Results 1..5: Result A, Result B, Result C, Result D, Result E, etc.

**Week 2:**
- Results 1..5: Result B, Result A, Result C, Result D, Result E, etc.

**Week 3:**
- Results 1..5: Result A, Result B, Result C, Result D, Result E, etc.

**Week 4:**
- Results 1..5: Result B, Result C, Result D, Result E, Result F, etc.

Fingerprint the data to determine numeric range for “normal”

Anomaly
SUPERVISED OR UNSUPERVISED?

- Supervised Learning
- Unsupervised Learning

dataaspirant.wordpress.com
ANOMALY DETECTION WITH UNSUPERVISED ML

MODEL SELECTION

• Unsupervised
  • Gaussian Mixture
  • Streaming K-Means
K-Means Clustering
Streaming K-Means – Adaptive Learning

- Partition objects into \( k \) nonempty subsets
- Repeat
  - Compute centroid (i.e., mean point) for each partition
  - Assign each object to the cluster of its nearest centroid
- Until no change
Algorithm (Streaming K-Means)

• Model Training (Normal dataset)
  • K: Number of clusters
  • Normalization of data
  • Engineering (categorical transformation/ dummy coding)
  • Labels/Entropy
• Trainer will yield centroid and threshold
• Validation
  • Anomalies: data points away from threshold from centroid
Algorithm (Streaming K-Means)

- $c_t$: previous centre of cluster
- $n_t$: number of points in a cluster
- $x_t$: cluster centre for current data
- $m_t$: number of points added in current batch
- Decay factor: $\omega$

\[
c_{t+1} = \frac{c_t n_t \omega + x_t}{n_t \omega + m_t} \quad n_{t+1} = n_t + m_t
\]
Happy Coding

Don’t Forget Machine Learning
THANK YOU

tamas@functionize.com