OMAINTEC 2018 CONFERENCE EXHIBITION

INTERNATIONAL OPERATIONS & MAINTENANCE CONFERENCE
IN THE ARAB COUNTRIES
UNDER THE THEME
"MANAGING MAINTENANCE WITHIN INDUSTRY 4.0"
CONICIDE WITH **THE 16TH ARAB MAINTENANCE EXHIBITION**

Machine Learning in Maintenance Optimization

Chi-Guhn Lee

Machine Learning

Machine Learning is to

- Quickly growing area of research.
- Enable algorithms to iteratively improve without explicit programming of models
- Utilize Statistical techniques

1. Supervised Learning
   - Majority of practical machine learning
   - Given data has the desired output value known
2. Unsupervised Learning
   - Given data has the desired output value **un**known
   - Powerful in that they operate beyond our preconceptions
3. Reinforcement Learning
   - Agent performs a particular goal by interacting with the environment
   - The agent is trained to make specific decisions

Supervised Learning

1. Linear Regression
   - Developed in the area of statistics
   - Output as a linear combination of input variables
2. Logistic Regression
   - One of the most commonly used ML methods
   - Developed in the area of statistics
   - Estimating probabilities using logistic function
   - Binary outputs for classification task
3. Neural Network
   - Inspired by biological neural network in human brain
   - Multi-layers of neurons: input, hidden, and output layer
   - Highly non-linear classification

Un-supervised Learning

1. K-means Clustering
   - Vectors are clustered based on distance
   - Vectors in a cluster have smaller distance to the centroid
   - The # of clusters should be set by the user
   - Computationally expensive
2. Affinity Propagation
   - Does not require user input
   - Clustering via message passing among observations
3. Hierarchical Clustering
   - Build clusters in a hierarchy
   - Observations belong to multiple clusters
   - Top-down vs. bottom-up

Reinforcement Learning

1. Markov decision processes
    - Mathematical framework: (T, S, A, Pr(s'|s,a), R(s,a))
    - Backward Induction
    - Successive approximation algorithms
    - Curse of dimensionality
2. Forward approximation
    - Sampling along the time
    - Convergence provided learning rate
3. Learning Algorithms
    - Q-Learning
    - SARSA
    - Deep Reinforcement Learning

Machine Learning Data Requirements

1. Flexible and Big Data
    - Pooling effect
    - ML requires BIG data
2. Design of pattern recognition system
    - Data collection
    - Formation of pattern classes
    - Feature selection
    - Specification of algorithm
    - Estimation of classification error
3. ML for maintenance applications
    - Feature selection
    - User inputs such as the # of clusters
    - Parameter estimation

Case Studies

Data

- 480 generating units
    - Hydro power plant in Niagara Falls, Canada
    - Over 5 years (2012~2017)
- The units failed for various causes from 114 components.
- 0.6 million entries of maintenance records, failures, etc.

Parameters

- # of clusters = 3
- # of iterations = 1,000

472-megawatt steam turbine generator (photo credit: businesswire.com)

Case Studies

Data Preparation

1. Removing redundancy
2. Removing units with inadequate records
3. Removing or recovering incomplete/inconsistent observations

	ForceOut	MainOut	MaxCapability	NumofCommon	PlanOut	WorkingHour	G199999	G142100	G141100	G142115	...
HGU0001	11	13	77.0	0.0	12	33429.6	0.0	6.0	2.0	0.0	...
HGU0002	13	14	77.0	0.0	10	36574.6	0.0	9.0	1.0	0.0	...
HGU0003	5	14	77.0	0.0	6	35721.3	0.0	3.0	1.0	0.0	...
HGU0004	11	10	77.0	0.0	7	36983.2	0.0	4.0	0.0	1.0	...
HGU0005	11	13	77.0	0.0	9	39080	0.0	6.0	1.0	0.0	...
HGU0006	6	13	77.0	0.0	8	35225.6	0.0	6.0	2.0	0.0	...
HGU0007	5	4	150.0	0.0	7	40213.8	1.0	7.0	16.0	1.0	...

Case Studies

Clustering

	Cluster 1	Cluster 2	Cluster 0
Average number of Forced outages	25.985	13.603	17.460
Average number of Maintenance outages	30.758	15.026	23.400
Average number of Planned outages	15.833	10.250	26.100
Average number of Common modes	0.015	1.263	0.280
Average maximum capability	46.533	58.185	306.586
Average working hours	37738.700	38917.044	35084.975

Case Studies

Cluster 0

- Largest average maximum capacity/unit
- Medium reliability
- Highest planned outages number is scheduled on these units
- Cluster 0 seems mostly important to the company.

Cluster 1

- Smallest average maximum capability/unit
- The least reliable units
- Relatively large number of planned outages
- Failures seem isolated rather than caused by other units

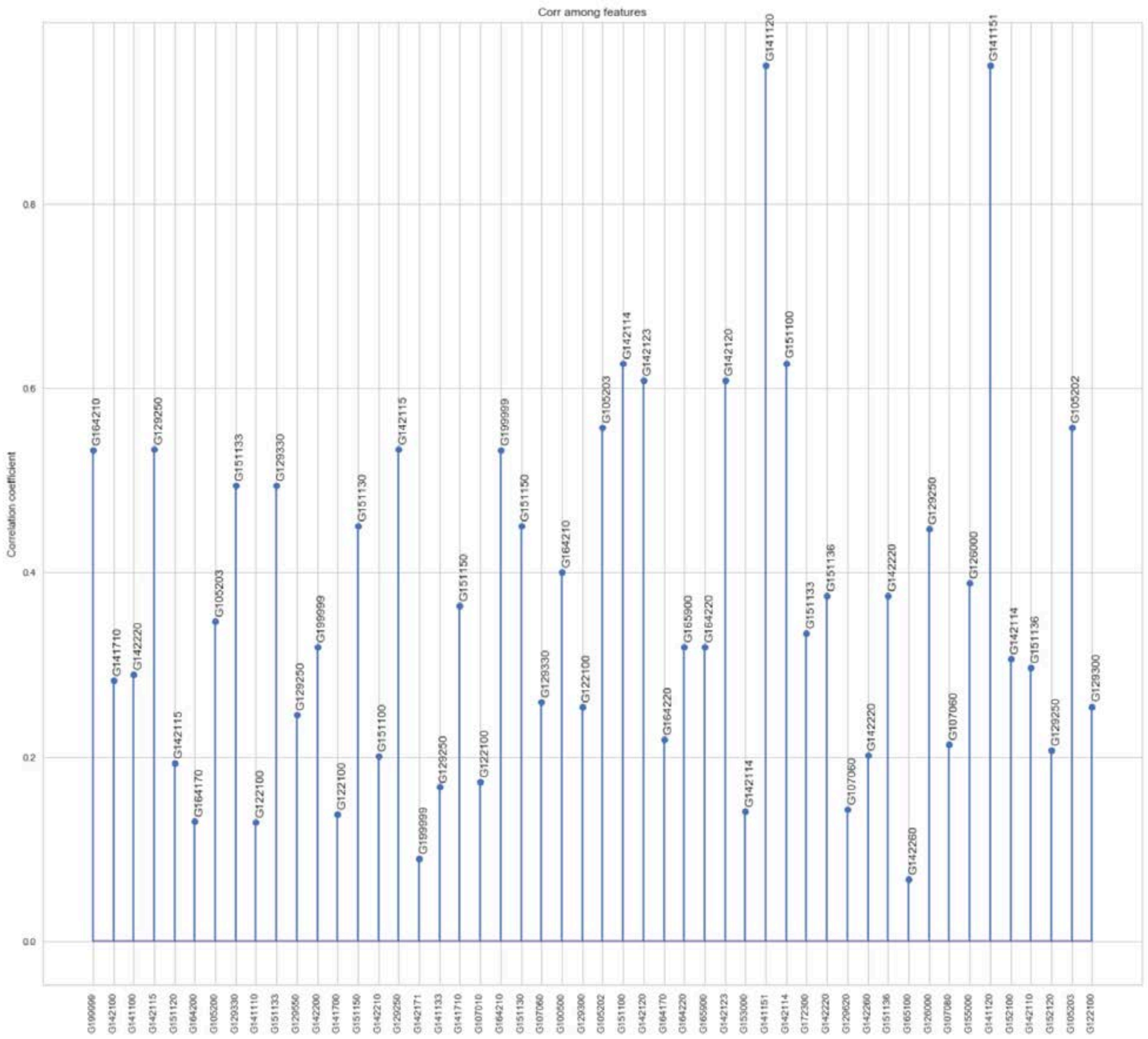
Cluster 2

- The most reliable unit
- Outages are likely caused by other units (leading to correlation analysis)
- Least maintained

Case Studies

Correlation Analysis

- Top 50 most commonly failing components



Case Studies

Correlation Analysis

- 3 pairs of components with the highest correlation

	Component 1	Component 2
Pair 1	41120 Rotor	41151 Shaft Seal: Packing, carbon Seal,…
Pair 2	42114 Slip Rings and Commutator	51100 Output Generator Voltage Equip.
Pair 3	42123 Generator Stator Winding	42120 Generator Stator

Case Studies

Occurrence Analysis

- Top 5 failing components in each cluster

	Cluster 1		Cluster 2		Cluster 0	
1	G199999	2,460	G199999	1,632	G141100	1,680
2	G142115	576	G142100	1,434	G142100	967
3	G151120	517	G141100	693	G142210	388
4	G142100	336	G142115	671	G105200	285
5	G142171	317	G141110	649	G164200	251

Challenges and Opportunities

Limitations:

1. Insufficient details about the generating units
   - Type, date to operate or which companies the units from.
   - Characteristics not provided but uniquely define each cluster
2. Unbalanced amount of records. After preprocessing,
   - There were 480 different units,
   - The average number of records is 1246.51
   - Many units have less than 100 records
   - The maximum number of records is as large as 17417.
3. Outliers need further investigation
   - In cluster 2, 0012, 0517, 0591, 0711 and 0841 have large forced outages
   - Some (591, 0711 and 0841) are acceptably large
   - Others (HGU 0012 and 0517) are unacceptably large, worth a further investigation.

Challenges and Opportunities

Recommendations:

1. Working with domain experts to get a deeper insight to the data
    - Validation of assumptions made during analysis
    - Improved quality of input data
2. Data not related to maintenance can be useful
3. Outliers can lead to new findings
4. Inter-unit analysis is required
    - Common mode can be a clue
5. Unsupervised results can be utilized to train supervised algorithms
    - Clustering results can be used in linear regression or NN analysis
    - Combined with statistical models such as PHM