DNV·GL

ENERGY

Give It All You Got

Using big data machine learning ensembles for condition monitoring

Elizabeth Traiger, Ph.D., M.Sc.

29 September 2016

Outline

- Motivating problem condition monitoring of gearbox issue detection in SCADA
- Developments around data
- Support vector machine failure model
- Gradient boosting machine failure model
- Summary of findings

The Problem

- Goal
  - Single model maximizing the use of high frequency (f ≥ 1 Hz) available SCADA signals real-time accurate condition monitoring (CM) to detect and classify gear box faults

- Challenges
  - Multiple mechanical causes of gearbox deterioration
  - Increasing number of component configurations
  - Increasing number of sensor configurations
  - More high frequency sensor signals available

Ungraded

The Problem

Current Serial CM Analysis Methods

Time Domain

- Trend analysis
- Time synchronous averaging
- …

Frequency domain

- Fourier transform and spectral analysis
- Order Analysis
- …

- Well established methodology and guidelines
- Capitalize on physical mechanism behind particular fault
- Multiple customized singular methods concurrently used
- Minimal use of hardware resources
- No dominate One-Size-Fits-All approach

Ungraded

Developments Around Data

- Big data is now just “Dysfunctional Data”

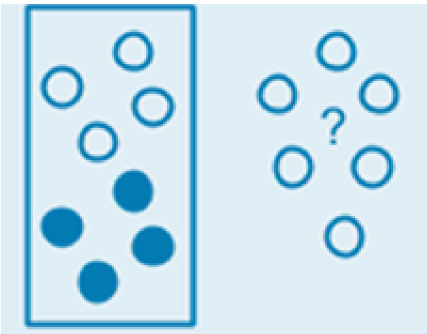
- Hardware constraints no longer a limiting factor
  - Sophisticated ontologies for relational databases in cloud computing
  - Distributed computing environments for alternative storage and processing

- Statistical tools now available
  - Development of fast parallel processing algorithms
  - Ensemble models now computationally feasible within standard software

- Collected wealth of data required for machine learning algorithms
  - Sufficient duration high quality data sets exist for pattern recognition
  - Synthesis of methods
    - Classification as well as detection possible

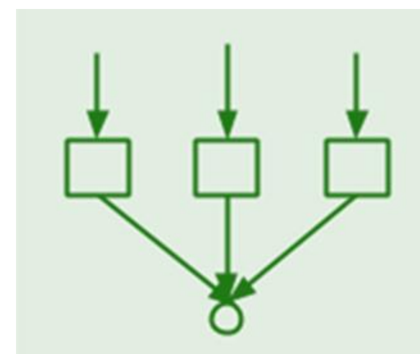
Ungraded

Machine Learning Ensemble Approaches



- Supervised Machine Learning
  - Excellent at anomaly detection in high dimensional spaces
  - Established results within conditional monitoring literature: artificial neural networks, self organizing maps, k-means clustering,…

- Ensemble Methods
  - Combine multiple weak models to create strong predictor
    - Support Vector Machines
    - Gradient Boosting Machines

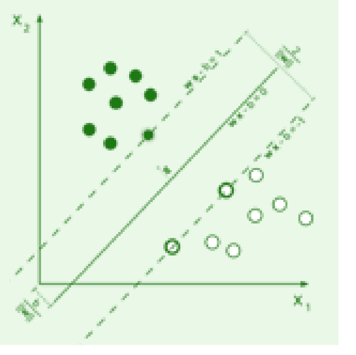


Ungraded

Support vector machine

- Support Vector Machine (SVM) is a non-probabilistic binary classifier which maps features into a separable space
  - Utilize a hyperplane with maximum margin to separate different classes of data
  - Can include non-linear features via kernel (RBF) trick
  - Hyperparameter tuning using grid search with k-fold cross validation

- Ensemble SVM method
  - Subsample the data run algorithm in parallel
  - Weight samples to compensate for unequal class sizes
  - Allow models vote on a prediction



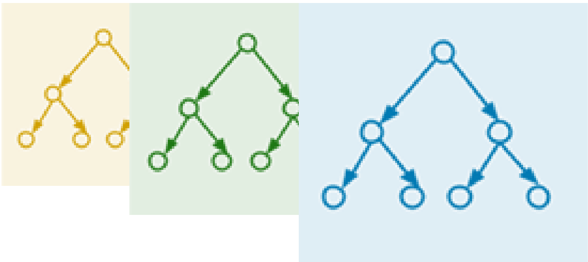
- One Versus All multi-class model
  - Consecutively build many binary classifiers learned to identify only one distinct class

Ungraded

Gradient boosting machine

- Gradient boosting machine (GBM) adaptively combines weak classifier models to form a strong classifier model minimizing pseudo-residuals
  – Decision trees are the base-learner models
  – Boosting sequentially adds new weak learners based on a loss function optimizing instances misclassified by previous learner
  – Intrinsic feature selection

- Rare inccident compensation
  – Additional weight vector for to the false positives error and false negatives added due to small incidents in singular classes
  – Weights are multiplied by the classification error at each iteration of the learning process

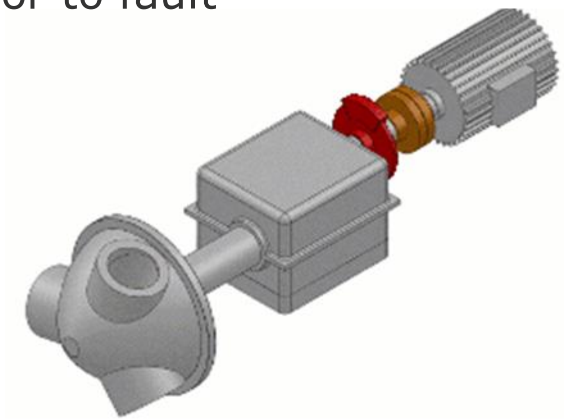


Ungraded

The Data

speed	acceleration	power	electrical	temperature
• generator RPM • wind • …	• rotor bearing • generator bearing • gear box • …	• turbine production • torque • …	• voltage phase • current phase • …	• nacelle • bearing • gear box • slip ring • generator • ambient • …

- Three different gearbox fault types to classify = Four classes
- Minimum of Two-months clear operational data prior to fault for training set
- Four year total series
- Multiple instances of each fault
  - Includes simulated fault data of each type



Ungraded

SVM & GBM Model Building

Pre-process Data*

Training / Test / Validation Partition

Feature Creation*

Iterative Model Creation

Predition Validations

Results Comparison

Classification Rate			
	Fault Type		
	Bearing Crack	Broken Gear tooth	Gear pitting
SVM	79%	76%	72%
GBM	92%	88%	84%

	Average Training Time (minutes)
SVM	14.22
GBM	38.19

	Average Prediction Time (seconds)
SVM	7.45
GBM	5.63

Ungraded

Closing Remarks

- Key Findings
  - Ensemble machine learning algorithms are a viable tool for condition monitoring mechanical fault detection applications utilizing many sensors with high frequency data
  - This study indicates ensemble Gradient Boosting Machines can outperform Support Vector Machines in multi-class gearbox fault classification

- Caveat Emptor
  - Missing values need to be handled with consideration
  - Domain expertise required
  - Model will not identify novel fault pattern
  - Dynamic
    - Models improve with additional fault library training, but grow
    - Retrain with the addition of new sensor signals

Ungraded

References

1. Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992.
2. Hastie, T., Tibshirani, R., and Friedman, J. H., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer, 2011.
3. D. Coronado and K. Fischer, "Condition Monitoring of Wind Turbines: State of the Art, User Experience and Recommendations Project Report", *Fraunhofer IWES*, January 2015. https://www.vgb.org/vgbmultimedia/383_Final+report-p-9786.pdf
4. GL Renewables Certification, "Guideline for the Certification of Condition Monitoring Systems for Wind Turbines", *Hamburg, Germany,* 2013.
5. M. Wilkinson, B. Darnell, T. van Delft and K. Harman, "Comparison of methods for wind turbine condition monitoring with SCADA data," *IET Renewable Power Generation*, vol. 8 np. 4, p.390-397, 2014.
6. Project Optique, http://optique-project.eu/.
7. Wang, Ke-Sheng, Vishal S. Sharma, and Zhen-You Zhang. "SCADA data based condition monitoring of wind turbines." *Advances in Manufacturing* 2.1 (2014): 61-69.

DNV·GL

Thank you for your attention

Elizabeth Traiger, Ph.D, M.Sc

elizabeth.traiger@dnvgl.com

www.dnvgl.com

SAFER, SMARTER, GREENER