Predictive Maintenance
Sensor Rich but Uncertain Information Quality Environment
Case Study in Railroad

Hongfei Li (Research Staff Member, IBM Research)
Joint work with Dhaivat Parikh, Qing He, Zhiguo Li, Buyue Qian, Dongping Fang, and Arun Hampapur
Currently, PBTG (a third-party application - dynamic simulation of various car types) exceptions are created but not acted upon

- Look at sub-critical track geometry measurements/ defects combined with PBTG output to create new value added exception tags that predict (and prevent) mainline derailments
- How should yellow tags be prioritized given the scarcity of resources?

Increase network velocity by ---

- Reduce (or at minimum, do not increase) the number of derailments (attributed to mechanical (car and locomotive) faults as primary cause)
- Reduce intermediate maintenance calls due to false positives/alarms (by 5% from current level)
- Can we automatically “learn rules” from historical data to better predict which rail cars are likely to have problems and maximize the hit rate of our “setouts”? 
Monitoring Train Operations to reduce service interruptions

The project is exploring the application of data mining, machine learning and predictive modeling to predict impending failure of critical rail car components. The prediction will drive proactive inspection and repairs, reducing "operational equipment failure". The project expects to deliver significant gains in network velocity and safety.

A large class 1 US Rail Operator with 20,000 miles of track has 1000’s of wayside sensors monitoring the condition of wheels & axles of train traffic, providing alarms when critical parameters (wheel temperature, wear, trucking errors, wheel impact) reach emergency levels.

1. Stop the train
2. Slow orders
3. Keep an eye on this one.
4. Pull this car at next station
5. Pull this car at end of trip
Moving from Reactive to Proactive Action

We are putting a safer train on the track

Wayside detectors

Alarms

Network Operations Center – Alarm Actions
1. Stop the train
2. Slow orders
3. Keep an eye on this one.
4. Pull this car at next station
5. Pull this car at end of trip

Predictive Model Based on Historical Data Mining
Data mining, machine learning, predictive modeling and failure analytics

List of “suspect equipment”

Railyard train formation

Long Term – time series analysis
Multi-system analysis
Detailed Approach to Predicted Rail Car Maintenance

1. Failure Pattern Analysis
   - SME’s Leverage Patterns in day to day decisions

2. Failure Causal Analytics
   - Failure Causal Map
   - Historical Covariate Information
   - Historical Failure Information

3. Learning Failure Prediction Models
   - Analytical models

4. Predicted Failure / Optimized PM Program
   - Maintenance SME’s execute proactive maintenance

Integrated Data Model for Composite Detector Predictive Maintenance

- Wayside Detector Data
- Traffic Data & Network Data
- Track Inspection Data
- Weather Data
- Set-Out & Failure Data
- Tear Down / Repair Data
- Live Sensor Data
Addressing Analytical Challenges

- Big Data – the data needs to be processed is around 2 TB across ~100s of tables
  - High Performance Appliance – Netezza with hooks for SQL & SPSS and R

- Noisy data - missing date, incorrect values, missing values
  - Custom Algorithms to impute missing value

- Extremely imbalanced sample - historically, there are only 898 equipments that have alarms, but around 900,000 unique equipments in the data
  - Statistical techniques for sample replication

- Requirement of low false positives - only 20 alarms is allowed in every 150,000 equipments

- Hard to predict - The patterns before alarm and after alarm are quite similar & Equipments go in-and-out of service breaking the pattern
  - Original Machine Learning algorithm development and efficient implementation in C and Pearl
  - Merging data from different detectors and environmental variables to achieve information density

- Simple rules are preferred, but simple rules may not work in such complex data
  - Original method developed for simplification and logicalization of ML algorithm output that trade-off some accuracy for human interpretability
Analysis Structure – Composite Detector Analysis

Wayside Detector Observations

Scoring Algorithm

1. Predict Level 1 alarms 7-days in advance
   - Hot Box / Warm Bearing Detector & Wheel Impact Load Detector

2. Detect patterns in wheel movement error, wheel dimension and wheel impact load to detect truck issues earlier
   - Machine Vision and Optical Geometry Detector & Wheel Impact Load Detector

3. Detect patterns in wheel dimensions, movement errors and wheel impact load that predict the wheel replacement earlier
   - Machine Vision, Optical Geometry Detector, Wheel Impact Load Detector

Create new rules

Wheels that wear asymmetrically could have a shorter lifespan and can lead to truck performance issues

Wheels and Trucks are replaced when they create high impact or wear-out

Detect asymmetrically wearing wheels in advance

- Machine Vision and Optical Geometry Detector

Existing Decision Model

Proposed Decision Model
## Value Proposition – Composite Detector Analysis

### Improving Reaction Time

**Results:**
The proposed scoring engine can capture ~39% of the level 1 warm bearing alarms in advance, but with ~140 additional cars alarmed per week (or 7% of L1 WB alarms with ~0 additional car)

**Bearing**

### Field Validation

**Results:**
With 99% confidence interval, ~30 wheel-sets are identified with asymmetric wear patterns. Need field validation for these results.

**Wheel-set**

### New Rule Development

**Results:**
The proposed ~1500 rules can capture ~95% of truck related Bad Orders and ~84% of wheel replacement earlier than FRA condemnable limits.

**Trucks and Wheel-set**
WB L1 Alarm Prediction

Background
- The company issues L1 alarm based on warm bearing temperature observations to prevent melting bearings or other types of defects.
- Once the L1 alarm is issued and validated, the train will be stopped immediately which causes the impact on network velocity.
- Efficient prediction of L1 alarms in advance will give sufficient operation time for the company to inspect and fix the problem and thus improve the network velocity.

Goal
- Develop composite detector rules to predict L1 alarms 7-day ahead using HBD, WILD and ABD in order to reduce the cost of immediate corrective maintenance action.

7-day ahead prediction
Challenges

- Existing alarm of catastrophic failure is usually generated when the component failure is imminent, leaving little time for planning and putting the maintenance organization in reactive mode.

- Around 2 TB amount of big data collected from about 1,000 detectors of multiple varieties in 7 months poses a big challenge when merging the information across detectors for one particular bearing.

- High standard industry requirement of prediction only allows extremely low false positive rate.

- The rules developed to trigger alarms should allow for the rationale for its actions under various situations to be understood and interpreted by humans. Black box type of rules cannot be easily operated.

Solutions

- Predict alarms associated with catastrophic failures instead of predicting failures directly.

- Use hashing and parallelization techniques.

- Model and tune a set of parameters in our customized SVM.

- Develop human interpretable rules.
Overview of Machine Learning for Alarm Prediction

The SVM (machine learning) “learns the intricate patterns in the data” which cause equipment to alarm.

- **Feature Extraction.** In this step, we aggregate the readings in the historical reading window and extract features using quantiles for each numeric value variable.

- **Dimension Reduction.** The goal here is to linearly project the sample features to a lower dimensional space while maintaining a comparable learning performance, which in turns reduces both the time and memory complexities required by the learning model.

- **Prediction.** This step is to predict if a bearing will issue a L1 alarm in the next few days based on its “location” (in feature space) to the “support vectors” (the key samples that lie in the border area between positives and negatives).

- **Confidence Level Estimation.** In addition to predict whether an alarm will be issued or not, we also estimate the corresponding confidence based on the relative relations to the support vectors.
Customized SVM

The Primal Formulation

\[
\begin{align*}
\min_{w, \xi} & \quad \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \\
\text{s.t.} & \quad y_i (w^T x_i + b) \geq \rho - \xi_i; \\
& \quad \xi_i \geq 0; \\
& \quad \rho \geq 0,
\end{align*}
\]

- \(x\) is a feature description
- \(y\) is the class label, \textit{i.e.}, alarm and non-alarm.
- \(w\) defines a hyperplane
- \(b\) is the offset, both \(b\) and \(w\) together draw a decision boundary in the feature space.
- \(C\) is a parameter that controls the trading-off between the “margin” and the training error (in terms of misclassified labels)
- \(\rho\) is used to reduce the number of samples that will be labeled as positive
- \(\xi\), a non-negative “soft margin” slack variable. This allows the classification result on the training set to be slightly distant to (on the wrong side of) the given labels.
## Results

### Table I

<table>
<thead>
<tr>
<th></th>
<th>7-7</th>
<th>14-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>HBD</td>
<td>45.411%</td>
<td>0.175%</td>
</tr>
<tr>
<td>WILD</td>
<td>40.373%</td>
<td>0.211%</td>
</tr>
<tr>
<td>HBD+ABD</td>
<td>26.906%</td>
<td>0.084%</td>
</tr>
<tr>
<td>HBD+WILD+ABD</td>
<td>39.135%</td>
<td>0.120%</td>
</tr>
<tr>
<td>HBD+WILD</td>
<td>60.387%</td>
<td>0.199%</td>
</tr>
</tbody>
</table>

**Table I**

TRUE POSITIVE RATES AND FALSE POSITIVE RATES USING DIFFERENT COMBINATION OF DATA SETS FOR 7-7 AND 14-3 SCENARIOS, RESPECTIVELY.

### Table II

<table>
<thead>
<tr>
<th></th>
<th>7-7</th>
<th>14-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>Highest True Positive Rate</td>
<td>97.585%</td>
<td>5.657%</td>
</tr>
<tr>
<td>Highest True Positive Rate with about 0.014% False Positive</td>
<td>38.542%</td>
<td>0.014%</td>
</tr>
<tr>
<td>Highest True Positive Rate with about 0 False Positive</td>
<td>7.459%</td>
<td>0.000%</td>
</tr>
</tbody>
</table>

**Table II**: TRUE POSITIVE RATES AND FALSE POSITIVE RATES USING HBD AND WILD COMBINATION FOR 7-7 AND 14-3 SCENARIOS UNDER DIFFERENT RATE REQUIREMENTS, RESPECTIVELY.
Results

(a) Decision map using coarse search

(b) Probability map using coarse search

(c) Decision map using fine search

(d) Probability map using fine search
# Business Value Illustration – Level 1 Warm Bearing Alarm Prediction

## # of total L1 alarms issued per year
1,200

## # of total cars per year
54,750,000

## Cost of False Alarm
$1,500

### Prediction Horizon

<table>
<thead>
<tr>
<th>Cost of L1 Alarm</th>
<th>True L1 Alarms Captured</th>
<th>False Alarms Issued</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$50,000</td>
<td>1,196</td>
<td>2,135,250</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>$15,000</td>
<td>1,196</td>
<td>2,135,250</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>$1,500</td>
<td>1,196</td>
<td>2,135,250</td>
<td>&lt; 0</td>
</tr>
</tbody>
</table>

### TP = 99.7% FP = 3.9%

<table>
<thead>
<tr>
<th>Cost of L1 Alarm</th>
<th>True L1 Alarms Captured</th>
<th>False Alarms Issued</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$50,000</td>
<td>544</td>
<td>6,570</td>
<td>17,325,000</td>
</tr>
<tr>
<td>$15,000</td>
<td>544</td>
<td>6,570</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>$1,500</td>
<td>544</td>
<td>6,570</td>
<td>&lt; 0</td>
</tr>
</tbody>
</table>

### TP = 45.3% FP = 0.012%

<table>
<thead>
<tr>
<th>Cost of L1 Alarm</th>
<th>True L1 Alarms Captured</th>
<th>False Alarms Issued</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$50,000</td>
<td>107</td>
<td>-</td>
<td>5,340,000</td>
</tr>
<tr>
<td>$15,000</td>
<td>107</td>
<td>-</td>
<td>1,602,000</td>
</tr>
<tr>
<td>$1,500</td>
<td>107</td>
<td>-</td>
<td>160,200</td>
</tr>
</tbody>
</table>

### TP = 8.9% FP = 0%

<table>
<thead>
<tr>
<th>Cost of L1 Alarm</th>
<th>True L1 Alarms Captured</th>
<th>False Alarms Issued</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$50,000</td>
<td>1,196</td>
<td>2,135,250</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>$15,000</td>
<td>1,196</td>
<td>2,135,250</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>$1,500</td>
<td>1,196</td>
<td>2,135,250</td>
<td>&lt; 0</td>
</tr>
</tbody>
</table>

TP = 14-Days of History, 3-Days in advance L1 alarm Prediction

TP = 45.3% FP = 0.012%
Track Geometry Exceptions

Figure 1 Track geometry parameters [19]

Table 1 Geo-defect summary (in alphabetical order)

<table>
<thead>
<tr>
<th>Defect Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALIGN</td>
<td>ALIGN is the average of the left and right a certain chord alignment.</td>
</tr>
<tr>
<td>CANT</td>
<td>Rail cant (angle) measure the amount of vertical deviation between two flat rails from their designed value. (1 degree = 1/8” for all Rail Weights, approximation)</td>
</tr>
<tr>
<td>DIP</td>
<td>DIP is the largest change in elevation of the centerline of the track within a certain distance moving window. Dip may represent either a depression or a hump in the track and approximates the profile of the centerline of the track.</td>
</tr>
<tr>
<td>GAGE_C</td>
<td>Gage Change is the difference in two gage readings up to a certain distance.</td>
</tr>
<tr>
<td>GAGE_TGHT</td>
<td>GAGE_TGHT measures how much tighter from standard gage (56-1/2”).</td>
</tr>
</tbody>
</table>
Summary of Analytics Focus – Track Geometry Exception analysis

- **Red Tags**: Fix Immediately
- **Yellow Tags**: Inspect within 30 days, Fix if required
- **PBTG Exceptions**: No Action

**Existing Decision Model**
- Given the existing data, no action is recommended

**Proposed Decision Model**
- Rank and Prioritize Yellow tags based on Track Deterioration Rate for that defect type and Derailment Risk related to defect type and amplitude
- Predict Derailment Risk for a track segment

- Decision Analysis / Resource Optimization to classify the tracks and defects to be fixed immediately
Value Proposition – Track Geometry Exception Analysis

Yellow Tag Deterioration Rate

Yellow Tag Risk of Derailment

Track Derailment Prob. LS=179 TN=0

Average Derailment Prob. for Track 179 / 0 19.9%

Repair Prioritization

Next Steps:

- Field validate the track risk and yellow ticket prioritization
- Confirm business value with realistic costs
Summary

- IBM Research collaborated with a US Class I Railway Company on two initiatives
  - Composite Detector Analysis
    - Alarm Prediction
    - Bad Truck/Wheel Prediction
    - Asymmetric Wheel Detection
  - Track Geometry Exception Analysis
    - Track Deterioration Analysis
    - Track Derailment Analysis
    - Yellow Ticket Prioritization