Improving rail network velocity: A machine learning approach to predictive maintenance

Hongfei Li a,*, Dhaivat Parikh b, Qing He c, Buyue Qian a, Zhiguo Li a, Dongping Fang a, Arun Hampapur a

aIBM T.J. Watson Research Center, Yorktown Heights, NY, United States
bIBM Global Business Services, Dallas, TX, United States
cThe State University of New York at Buffalo, Buffalo, NY, United States

ARTICLE INFO

Article history:
Received 8 May 2013
Received in revised form 31 March 2014
Accepted 21 April 2014
Available online 16 May 2014

Keywords:
Big data
Condition based maintenance
Multiple wayside detectors
Information fusion
Predictive modeling
Rail network velocity

ABSTRACT

Rail network velocity is defined as system-wide average speed of line-haul movement between terminals. To accommodate increased service demand and load on rail networks, increase in network velocity, without compromising safety, is required. Among many determinants of overall network velocity, a key driver is service interruption, including lowered operating speed due to track/train condition and delays caused by derailments. Railroads have put significant infrastructure and inspection programs in place to avoid service interruptions. One of the key measures is an extensive network of wayside mechanical condition detectors (temperature, strain, vision, infrared, weight, impact, etc.) that monitor the rolling-stock as it passes by. The detectors are designed to alert for conditions that either violate regulations set by governmental rail safety agencies or deteriorating rolling-stock conditions as determined by the railroad.

Using huge volumes of historical detector data, in combination with failure data, maintenance action data, inspection schedule data, train type data and weather data, we are
Outline

1. Motivation
   ▶ Rail network velocity
   ▶ Train condition monitoring

2. A machine learning approach to predictive maintenance
   ▶ Challenges
   ▶ Application 1: predict alarms due to hot bearings
   ▶ Application 2: predict wear-out failures of wheels
   ▶ Summary

3. Related Work
1. **Motivation**
   - Rail network velocity
   - Train condition monitoring

2. **A machine learning approach to predictive maintenance**
   - Challenges
   - Application 1: predict alarms due to hot bearings
   - Application 2: predict wear-out failures of wheels
   - Summary

3. **Related Work**
Motivation

Rail network velocity - Definition

- The system-wide average speed of line-haul\(^1\) movement between terminals
- Is calculated by dividing total train-miles by total hours operated

Rail network velocity - Challenges

- Increased demand for railway services on relatively fixed rail networks requires increase in network velocity, without compromising safety
- One of the key problems for network velocity is service interruption
  - Lowered operating speed due to track/train condition
  - Delays caused by derailments\(^2\)

---

\(^1\)line-haul = Linienverkehr

\(^2\)derailment = Entgleisung
Train condition monitoring - Via wayside detectors

- Hot Box Detector (HBD): measures temperature of bearings and wheels
- Wheel Impact Load Detector (WILD): measures impact of wheels on the track
- Many more detector types, monitoring assemblies, axles, sound, ...

Images from: http://en.wikipedia.org/wiki/Train_inspection_system

\(^3\) wheel bearing = Radlager
Train condition monitoring - Reactive Maintenance

▶ When a detector’s measurement is violating a regulation, an alarm is issued
▶ Mostly a decrease in speed or even an immediate train stop is required
▶ The train then travels slowly to the nearest siding\(^4\) until maintenance men arrive to repair it
▶ This practice avoids derailments and further damages of train or tracks

So what is the problem?

▶ The schedule is disrupted, cargo remains undelivered
▶ Maintenance cannot be planned well, is therefore less efficient

\(^4\) siding = Abstellgleis
Train condition monitoring - Proactive/Predictive Maintenance

- With machine learning techniques maintenance can become proactive/predictive
- Historical detector data as well as maintenance data is stored by the railway companies and can be used for learning
- Alarms/Failures would then be predicted a certain time in advance
- Railroad could therefore plan their maintenance better and reduce service interruptions
- The results are reduced costs and higher network velocity
- Additionally aggregation of readings from multiple detectors would prevent false alarms due to measurement errors of a single detector
1. **Motivation**
   - Rail network velocity
   - Train condition monitoring

2. **A machine learning approach to predictive maintenance**
   - Challenges
   - Application 1: predict alarms due to hot bearings
   - Application 2: predict wear-out failures of wheels
   - Summary

3. **Related Work**
A machine learning approach to predictive maintenance

The rail network considered

- One of the leading railroads in the US
- Manages about 32,200 kilometers of tracks
- Has about 1,000 detectors installed along the network
- This corresponds to 1 detector per 32 kilometers
A machine learning approach to predictive maintenance

Challenges

- Combining spatio-temporal incompatible detector readings
  - Detectors are not co-located
  - Detectors vary in frequency, e.g. about 800 HBDs but only 12 WILDs
- Handling big data
  - E.g. data size for one-year HBD readings is about 3 TB in total
- Producing understandable rules
  - Accurate predictions require a certain complexity
  - But generated rules should be comprehensible to operators
- Extremely low false alarm rate
  - Railroad has limited maintenance budget
A machine learning approach to predictive maintenance

Application 1: Predict HBD alarms

Application 1:
Predict alarms due to hot bearings
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Alarms of Hot Bearing Detector

- An HBD alarm is issued if a bearing reaches a certain temperature threshold
- From high temperatures materials are softened, or sparks could inflame cargo
- Immediate train stop is required
- More than 1,000 HBD alarms at the considered railway each year

Features

- Features are aggregated statistics from historical readings of HBD and WILD for individual bearings (55 features in total)
- E.g. maximum, 95 percentile, mean, variation and trending
- Including features of further detectors did not improve the performance
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Parameters

- The first parameter is the time window for historical detector readings
  - Since WILD readings are so infrequent this parameter only effects HBD readings
  - For WILD the window has a fixed value of 4 weeks
- Second parameter is the prediction time window

Configurations

- 7-7:
  - Use HBD readings for a bearing of past 7 days (WILD readings of past 4 weeks)
  - Provide alarm prediction for that bearing at day 7 in the future
- 14-3:
  - Use HBD readings for a bearing of past 14 days (WILD readings of past 4 weeks)
  - Provide alarm prediction for that bearing at day 3 in the future
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Feature Extraction Example

7-7 Configuration
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Feature Extraction

- Since readings of every detector are stored in separate files information for single bearings must be linked
- This is based on string-matching of serial numbers which is very time-consuming for terabytes of data
- Therefore hashing and parallelization is used, described in detail in [3]
- For efficient model development the number of features is reduced from 55 to 12 by Principal Component Analysis
- Dataset: more than 800,000 non-alarm bearings and 600 alarm bearings with corresponding HBD/WILD readings for the given time window
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Model training

- The authors claim to have explored several machine learning techniques (including Decision Tree, k Nearest Neighbors, Artificial Neural Networks and Support Vector Machines)
- However they give only results for Decision Tree and Support Vector Machine (SVM)
- Performance is measured by false positive rate (FPR) and true positive rate (TPR)
- The railway considered demands a FPR of less than 0.014%
- For evaluation fivefold cross validation was used
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Results

Table: Results of SVM under constraint of FPR <= 0.014%

<table>
<thead>
<tr>
<th>Parameter-Configuration</th>
<th>7 - 7</th>
<th>14 - 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>0.014%</td>
<td>0.012%</td>
</tr>
<tr>
<td>TPR</td>
<td>38.542%</td>
<td>45.368%</td>
</tr>
</tbody>
</table>

Table: Results of Decision Tree for lowest possible FPR

<table>
<thead>
<tr>
<th>Parameter-Configuration</th>
<th>7 - 7</th>
<th>14 - 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>0.976%</td>
<td>1.073%</td>
</tr>
<tr>
<td>TPR</td>
<td>61.256%</td>
<td>68.463%</td>
</tr>
</tbody>
</table>
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Extracting Simple Rules from SVM model

- Feature space divided into cells, center points classified by SVM
- Example rule: if feature 1 > 100 and feature 2 > 6,300 then issue alarm
- Performance proportionally & rule-complexity inversely proportional with cell size
A machine learning approach to predictive maintenance
Application 1: Predict HBD alarms

Further decreasing of FPR by confidence estimates

- SVM assigns a confidence value to each prediction
- FPR can be further reduced by issuing alarms only at a high confidence level
Application 2: Predict wear-out failures of wheels
A machine learning approach to predictive maintenance
Application 2: Predict failures of wheels

Wear-out failures of wheels

- Wheels are vulnerable to wear-out failures\(^5\)
- Wheel failures cause about one half of all train derailments

Features

- 20 attributes from four different detector types, including WILD
- Problem: measurements are highly influenced by external factors (weather, train load)
- Therefore readings of 4 weeks are aggregated per wheel (mean, median, various quantiles, trending and variation)

\(^5\) wear-out failure = Verschleißausfall
A machine learning approach to predictive maintenance
Application 2: Predict failures of wheels

<table>
<thead>
<tr>
<th>Detector type</th>
<th>Attribute</th>
</tr>
</thead>
</table>
| MV            | Wheel flange height  
                Wheel flange thickness  
                Wheel rim thickness  
                Wheel diameter  
                Wheel tread hollow  
                Brake shoe upper thickness  
                Brake shoe lower thickness |
| OGD           | Truck hunting peak-to-peak (PTP) measurement  
                Truck hunting amplitude  
                Truck inter-axle misalignment (IAM)  
                Truck rotation measurement  
                Truck tracking error  
                Truck shift measurement |
| WILD          | Wheel average downward load reading  
                Wheel peak downward load reading  
                Wheel average lateral load reading  
                Wheel peak lateral load reading  
                Difference between peak and average downward load reading  
                Truck hunting index |
| TPD           | Ratio of later and vertical load |
A machine learning approach to predictive maintenance
Application 2: Predict failures of wheels

Parameters

- The first parameter is the time window for historical detector readings
- The second parameter is the prediction time window
  - Should be long enough for the staff to plan maintenance
  - But not too long to avoid removing components too early

Configuration

- Both parameters were set to 12 weeks based on discussion with experts
A machine learning approach to predictive maintenance
Application 2: Predict failures of wheels

Feature Extraction

- There is no specified detector for exactly identifying wear-out failures, so how to find positive instances?
- After an arbitrary detector issued an alarm and the train received maintenance, a corresponding repair record is stored
- Wear-out failures can be identified by information from these records
- Data: about 500 GB of raw multi-detector data and repair records from January 2010 to March 2012
A machine learning approach to predictive maintenance
Application 2: Predict failures of wheels

Model Training & Results

- Maintenance department emphasized importance of human interpretability of failure prediction rules
- Therefore a decision tree was trained on 80% of the data and evaluated on the remaining 20%
- The paper claims the „accuracy-requirements“ were met by the decision tree

Table: Results of Decision Tree for lowest possible FPR

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>0.23%</td>
</tr>
<tr>
<td>TPR</td>
<td>97%</td>
</tr>
</tbody>
</table>
A machine learning approach to predictive maintenance
Application 2: Predict failures of wheels

Extracting Rules
- Decision tree naturally produces human interpretable rules
- Every leaf of the decision tree corresponds to a rule
A machine learning approach to predictive maintenance

Summary
A machine learning approach to predictive maintenance

Summary

- Two different applications for predictive maintenance were presented: alarm prediction and failure prediction
- Using the proposed techniques can effectively lower service interruptions and improve network velocity
- Furthermore, the authors claim that this can save the railway between 200,000 and 5,000,000 USD per year, depending on alarm locations, traffic and TPR-FPR trade-off chosen in the implementation
- The techniques described are more generally applicable to many other industries that use sensor network for equipment monitoring
1. **Motivation**
   - Rail network velocity
   - Train condition monitoring

2. **A machine learning approach to predictive maintenance**
   - Challenges
   - Application 1: predict alarms due to hot bearings
   - Application 2: predict wear-out failures of wheels
   - Summary

3. **Related Work**
There is an extended version of the presented paper, which describes the alarm prediction in more detail [3]

Work on how to place detectors in the network to solve the problem of detectors being not co-located can be found in [4]

There are several other papers focusing on wheel failure prediction:
- Logistic Regression using two detector types (including WILD) is applied in [1], Results for predicting 30 days in advance: TPR=90%, FPR=15%
- Stacking of Naive Bayes Classifiers & Decision Trees using WILD is applied in [5], Results for predicting 20 days in advance: TPR=97%, FPR=8%
Related Work

Situation in Germany

- Deutsche Bahn network with 33,000 kilometers of tracks comparable to considered US rail network (32,000 kilometers)
- Wikipedia: 420 HBDs in Deutsche Bahn network in 2007 (1 HBD per 78.6 kilometers)
- 800 HBDs in the considered US railway network (1 HBD per 40.6 kilometers)
- Since bearings can go from a nearly undetectable problem to complete failure in less than 32 kilometers the number of HBDs in Deutsche Bahn Network seems to be rather small, unless it was remarkable increased in the meantime
- I did not found similar machine learning approaches for german railroad
H Hajibabaia, MR Saat, Y Ouyang, CPL Barkan, Z Yang, K Bowling, K Somani, D Lauro, and X Li.
Wayside defect detector data mining to predict potential wild train stops.

Hongfei Li, Dhaivat Parikh, Qing He, Buyue Qian, Zhiguo Li, Dongping Fang, and Arun Hampapur.
Improving rail network velocity: A machine learning approach to predictive maintenance.

Hongfei Li, Buyue Qian, Dhaivat Parikh, and Arun Hampapur.
Alarm prediction in large-scale sensor networks—a case study in railroad.
Yanfeng Ouyang, Xiaopeng Li, Christopher PL Barkan, Athaphon Kawprasert, and Yung-Cheng Lai.

Chunsheng Yang and Sylvain Létourneau.
Learning to predict train wheel failures. 2005.
End of my presentation

Thank you for listening!
### Appendix

**Table 1**  
Decision tree results for the settings of 7-7 and 14-3 under two scenarios.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>7-7</th>
<th>14-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPR (%)</td>
<td>FPR (%)</td>
</tr>
<tr>
<td>I: Highest true positive rate</td>
<td>91.546</td>
<td>6.849</td>
</tr>
<tr>
<td>II: Lowest false positive rate</td>
<td>61.256</td>
<td>0.976</td>
</tr>
</tbody>
</table>

**Table 2**  
Customized SVM results for the settings of 7-7 and 14-3 under three scenarios.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>7-7</th>
<th>14-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPR (%)</td>
<td>FPR (%)</td>
</tr>
<tr>
<td>I: Highest true positive rate</td>
<td>97.585</td>
<td>5.657</td>
</tr>
<tr>
<td>II: Lowest false positive rate</td>
<td>7.459</td>
<td>0.000</td>
</tr>
<tr>
<td>III: Highest true positive rate under constraint of 0.014% false positive</td>
<td>38.542</td>
<td>0.014</td>
</tr>
</tbody>
</table>
The parameter $p$ implicitly controls the number of support vectors. When $p > 1$ the model tends to have more support vector since there will be more active constraints. This would produce a more complex prediction model, such that FPR can be lowered (TPR is decreased as well).