Operational Data Based Anomaly Detection for Locomotive Diagnostics

Feng Xue, Weizhong Yan
GE Global Research
1 Research Circle,
Niskayuna, NY 12309

Nicholas Roddy
GE Transportation Systems
East Lake Road 12-2
Erie, PA 16531

Anil Varma
GE Global Research
1 Research Circle,
Niskayuna, NY 12309

Abstract. *Locomotives are complex electromechanical systems. Continuously monitoring the health state of locomotives is critical in modern cost-effective maintenance strategy. A typical locomotive is equipped with the capability to monitor their state and generate fault messages and a snapshot of sensed parametric readings in response to anomalous conditions. In our previous studies, we have developed and deployed a case-based reasoning system for locomotive diagnostics where fault codes were used as the inputs to the system. In order to increase the lead-time from detection to failure and allow for more proactive actions, one important effort in locomotive diagnostics is to perform anomaly detection on parametric operational data. In this paper, we present an anomaly detection strategy that is based on a combination of nonparametric statistical testing and machine learning methodology. We demonstrate the effectiveness of the anomaly detection strategy using real-world operational data from locomotives.*

Keywords**:** Anomaly detection, Non-parametric test, Mann-Whitney-Wilcoxon test, Generalized Regression Neural Network, Machine Learning

1. Introduction

Over the last decade, the General Electric (GE) Company has increasingly been moving into service-oriented businesses as a way to increase revenue beyond its base product sales. GE builds and sells a wide range of products including medical equipment, power turbines, aircraft engines, locomotives, and other industrial equipment. Control systems on these machines were built to optimize operation and protect the equipment from failure. Information that is produced from the control systems provides valuable information for how to fix the equipment, thus, is useful for field technicians in order to detect, diagnose, and fix equipment problems.

With the proliferation of more inexpensive telecommunications and computing capability, many businesses have been moving toward a business model to *remotely* capture machine data and proactively diagnose failures. Continuously monitoring the health state of machines or equipment is a critical part of modern cost-effective maintenance strategy. This approach has cost advantages in that it allows for more timely response to costly impending failures and allows for retention and constant improvement of diagnostic knowledge into corporate databases.

GE Transportation Systems has been performing remote diagnostics on locomotives since 1999. A major component of the off-board diagnostic system is a case-based reasoning (CBR) software application (Varma and Roddy, 1999; Roddy et al. 2005), which assists the end user in constructing a diagnostic strategy to send to technical staff in the field for resolving locomotive failures. GE Transportation Systems currently monitors thousands of locomotives in North America. As controller-logged faults and other anomalous conditions develop onboard the locomotive, they are collected and downloaded to a central monitoring center for analysis. An engineer examines this data and the output of multiple AI based diagnostic tools to determine if the equipment is experiencing a problem. If a problem is diagnosed, a troubleshooting strategy is created and sent to a field engineer in order to perform maintenance on the locomotive.

In the CBR system, fault logs generated from the control and monitoring unit of a locomotive are used as inputs for fault detection and diagnosis. Although a snapshot of sensed parametric readings in response to anomalous condition is also captured along with the fault messages, these parametric data have so far not been fully explored for the purpose of anomaly detection and diagnosis.

In this paper, we will show our initial efforts on the use of operational (parametric) data for fault detection and diagnosis of locomotives. In particular, we focus on the monitoring of the health status (normal or abnormal condition) of locomotives using operational data, which can provide complementary detection information to the CBR system. Specifically, we develop an anomaly detection strategy that combines non-parametric statistical testing and machine learning based decision fusion for estimating health status (normal or abnormal condition) of locomotives based on a series of operational data.

2. The proposed anomaly detection strategy

2.1 Overall structure of the proposed strategy

Figure 1 shows the overall structure of the proposed anomaly detection scheme. It essentially consists of two fundamental components, namely nonparametric statistical test and decision fusion. Nonparametric statistical test performs hypothesis test on the two samples of the same parametric measurement, one of which represents normal locomotive condition and another one represents current locomotive condition. The test is to statistically determine if the parametric measurement has deviated from normal condition. The test is performed on each of the parametric measurements.

Fusion, on the other hand, is to arrive at a final detection decision by intelligently integrating the individual results of the nonparametric statistical test performed on each of the parametric measurements. The following two subsections describe in details the two components of the proposed anomaly detection scheme.

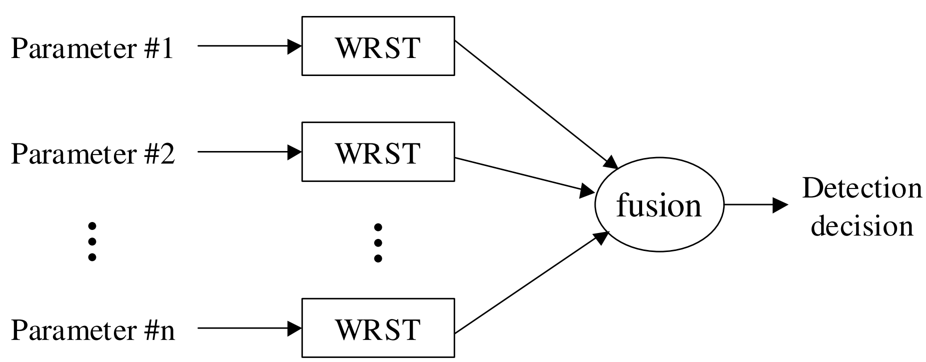

Figure 1: Overall structure of the proposed anomaly detection scheme

2.2 The Wilcoxon rank sum test

The Wilcoxon rank sum test (WRST) (Wilcoxon, 1947; Man and Whitney, 1945; Gibbons, 1971) is one of the most powerful nonparametric tests for comparing two populations. It is usually formulated in a way to test the null hypothesis that two samples come from identical distribution against the alternative hypothesis that the two distribution functions differ only with respect to location (median), if at all. It can also be used, in a more general way, to test if two samples come from identical distribution against the alternative hypothesis that one of the two distributions is statistically larger (smaller) than the other one. Eklund and Goebel (2005) have used this test based on a simulated statistic distribution for abnormal condition detection in an aircraft engine.

While the WRST is generally considered as a nonparametric alternative to the conventional two-sample *t*-test, it has one advantage that it does not require the assumption that the differences between the two samples are normally distributed. In addition, since it is based on the statistics of rank, rather than the original measurements, WRST is more resistant to the outliers, that is, WRST will be less affected by the presence of outliers unless the number of outliers becomes large relative to the sample size.

Let's consider two mutually independent random samples of sizes m and n, denoted by $x_1,x_2,\ldots,x_m$ and $y_1,y_2,\ldots,y_n$, drawn from continuous population F_X and F_Y, respectively. Assume that the two random samples are drawn from the two populations with the same distribution shape but possibly different locations. Our primary interest is to test whether these two distributions are the same, that is, whether the locations of these two distributions are the same. Let m_x denote the median of the *X* population and m_Y denote the median of the *Y* population. The hypothesis set can written as on of the following:

Two-sided alternative	$H_0: m_x = m_Y$	$H_a: m_x \neq m_Y$
One-sided alternative	$H_0: m_x = m_Y$	$H_a: m_x > m_Y$
	$H_0: m_x = m_Y$	$H_a: m_x < m_Y$

In order to serve this purpose, two statistics are used to perform the test: one is the Mann-Whitney *U* statistic; and the other is the Wilcoxon *W* statistic. These two statistics are actually the same since a linear relationship exists between the two statistics. This is also why we call this test as Mann-Whitney-Wilcoxon test.

The Mann-Whitney *U* statistic is defined as the number of times a *Y* precedes an *X* in the combined ordered arrangement of two independent random samples. If the *mn* indicator random variables are defined as

$$D_{ij} = \begin{cases} 1 & \text{if } Y_j < X_i \\ 0 & \text{if } Y_j > X_i \end{cases} \quad \forall \text{i}=1,2,\cdots,m;\, j=1,2,\cdots,n \tag{1}$$

a symbolic representation of the Mann-Whitney *U* statistic is

$$U = \sum_{i=1}^{m}\sum_{j=1}^{n} D_{ij} \tag{2}$$

The Wilcoxon *W* statistic is defined as the sum of the ranks of $X_i (i = 1, 2, \cdots, m)$ in the combined ordered sample of $X's$ and $Y's$. If we define an indicator vector $Z = [Z_1, Z_2, \cdots, Z_N]$ where $Z_i = 1$ if the i^{th} random variable in the combined ordered sample is an *X* and $Z_i = 0$ if the i^{th} random variable in the combined ordered sample is a *Y*. $N = m + n$ denotes the size of combined sample. In this way, the Wilcoxon test statistic can be written as

$$W_N = \sum_{i=1}^{N} iZ_i \tag{3}$$

A linear relationship exists between the Mann-Whitney *U* test statistic and Wilcoxon W_N test statistic:

$$U = W_n - \frac{m(m+1)}{2} \tag{4}$$

Under H_0: $F_X(u) = F_Y(u)$, the combined sample is essentially a random sample drawn from the same distribution. The Wilcoxon test statistic has an exact distribution or can be approximated by a normal distribution.

2.3 Decision fusion

After obtaining WRST results for each individual parameter, we arrive at our final detection decision by intelligently combining the individual WRST results. By taking account the complementary information from the individual results, the decision fusion can yield a more accurate and robust decision on the condition of the underlying system. There are a variety of fusion methods, ranging from simple averaging to more sophisticated Dempster-Shafer. In this paper, we use the generalized regression neural network (GRNN) as the fusion algorithm.

The generalized regression neural network (GRNN) is a universal approximator and typically used for function approximation. It has been shown that, given a sufficient number of neurons, GRNN can approximate a continuous function to an arbitrary accuracy. GRNN has two layers of artificial neurons. While the first layer consists of radial basis neurons (whose transfer function is a Gaussian with a spread factor), the second layer consists of neurons with a linear transfer function. In this paper, the inputs to the GRNN are the WRST results of individual parameters and the single output is the prediction of the locomotive engine condition (normal or abnormal). GRNN has one design parameter, spread factor, which is set to be 1.0 in this paper.

3. A case study

Figure 2 shows typical outputs from on-board controllers (each row is a fault log) and information contained in it. Each record of the fault log is composed of two distinct pieces of information. The first are fault codes, that are triggered when a predetermined set of conditions are met on the locomotive; while the remaining is an instantaneous reading of several significant operating parameters which are also recorded. Only a small portion of the parameters colleted from a locomotive is shown in Figure 2.

F	158309.95	158309.95	7080	01	01/22/2005 05:57	01/22/2005 05:57	0.0	C	1	186	0	0	0	0	P	0	0	0	I	B
*	158309.98	158309.98	7A07	00	01/22/2005 05:59	01/22/2005 05:59	0.0	C	1	436	4	0	180	186	R	0	0	0	I	B
*	158309.97	158309.98	7A07	00	01/22/2005 05:58	01/22/2005 05:59	0.0	C	1	439	4	0	179	186	R	0	0	0	1	
*	158309.97	158309.97	7A07	00	01/22/2005 05:58	01/22/2005 05:58	0.0	C	1	439	4	0	179	187	R	0	0	0	1	
*	158309.93	158309.97	7A07	00	01/22/2005 05:56	01/22/2005 05:58	0.0	C	1	439	4	0	177	187	R	0	0	0	1	B
*	158309.93	158309.93	7A07	00	01/22/2005 05:56	01/22/2005 05:56	0.0	C	1	439	4	0	178	187	R	0	0	0	1	
*	158309.93	158309.93	7A07	00	01/22/2005 05:56	01/22/2005 05:56	0.0	C	1	436	4	0	182	188	R	0	0	0	1	
*	158309.92	158309.93	7A07	00	01/22/2005 05:55	01/22/2005 05:56	0.0	C	1	335	3	0	182	188	R	0	0	0	1	
*	158309.32	158309.92	7A07	00	01/22/2005 05:19	01/22/2005 05:55	21.2	R	8	983	40	0	178	211	R	0	0	0	2	B
F	158309.32	158309.32	7080	01	01/22/2005 05:19	01/22/2005 05:19	21.9	R	8	248	43	0	168	201	P	0	0	0	A	B
F	158309.32	158309.32	773E	00	01/22/2005 05:19	01/22/2005 05:19	0.0	R	8	186	0	0	0	211	P	0	0	0	A	B
F	158309.30	158309.92	76D5	00	01/22/2005 05:18	01/22/2005 05:55	0.0	R	8	124	0	0	0	211	P	0	0	0	A	
F	158309.30	158309.30	773E	00	01/22/2005 05:18	01/22/2005 05:18	0.0	R	8	0	0	0	0	211	P	0	0	0	2	
*	158309.30	158309.30	773E	00	01/22/2005 05:18	01/22/2005 05:18	22.4	R	8	1062	219	282	178	211	R	0	0	0	2	

Figure 2: An example of typical fault log and the capture of an instantaneous snapshot operational data

There are many operational parameters one can collect from a complex electromechanical system such as a locomotive. In this paper, we focus on a particular turbo subsystem of the locomotive and conduct analysis on only those operational data that are relevant to this particular subsystem. To validate the proposed approach, we collect both normal operational and failure data from the GE Transportation Systems' remote monitoring and diagnostics service center. The results of the study based on a number of failure cases are presented in the next.

4. Results

We collected 5 groups of datasets, one from healthy engines and the other 4 are from engines under different faulty conditions within a subsystem. We used 2 of the 4 faulty datasets combining with a small portion of the healthy data for training GRNN fuser and used the other 2 faulty datasets combining with the rest of healthy data as the testing set to validate the performance of the system. The results reported here are for the testing dataset.

Figure 3 shows the 9 sensed parametric measurements. It represents a total of 579 snapshots (data points) over a period of x days. The first 510 data points are for healthy engine condition while the remaining 69 data points are taken from engines that have abnormal condition. For proprietary information protection purpose, the names of the 9 sensed parametric measurements are replaced by some general variable names and the raw sensor readings are normalized to the range of [0 1].

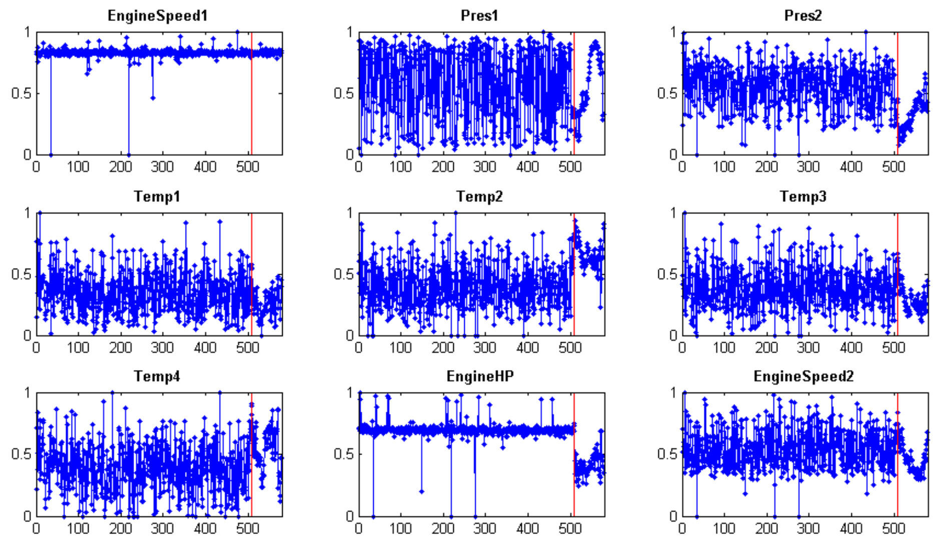

Figure 3: The 9 sensed parametric measurements

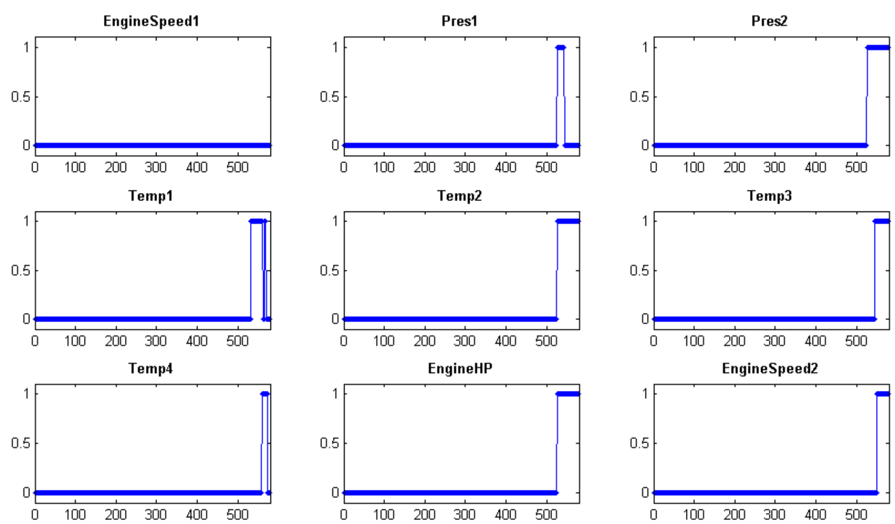

Figure 4: The individual WRST results for 9 sensed parameters

Figure 4 shows the results of the WRST for the corresponding 9 parameters, respectively. For each of the WRST, we perform the hypothesis test on the two samples, namely, "base" sample and "target" sample. The "base" sample is a collection of 100 data points that are randomly drawn from a large pool of data points representing normal operational condition. The "base" sample is fixed over entire tests. The "target" sample, on the other hand, is dynamic (i.e., changes over time). It is formed by 20 data points within a sliding window with a fixed window width of 20 points. The window slides one point at a time. The output of the WRST represents the test result for the last point of the window. A test result of zero indicates that the "target" sample is statistically no difference from the "base" sample, thus a normal condition is detected. On the other hand, if the test gives one as the result, the "target" sample has different distribution comparing to the "base" sample, that is, an abnormal engine condition occurs.

As can be seen from Figure 4, at any given time instance (or equivalently data point here), the results of the WRST for individual parameters may be conflicting (some say normal, while other say abnormal), which justifies the need for decision fusion to compromise. Figure 5 shows the outputs of GRNN fusion of the 9 WRST results. For the 579 snapshots considered, fusion shows a perfect performance, that is, it captures 100% abnormal cases while giving no false alarm for normal condition.

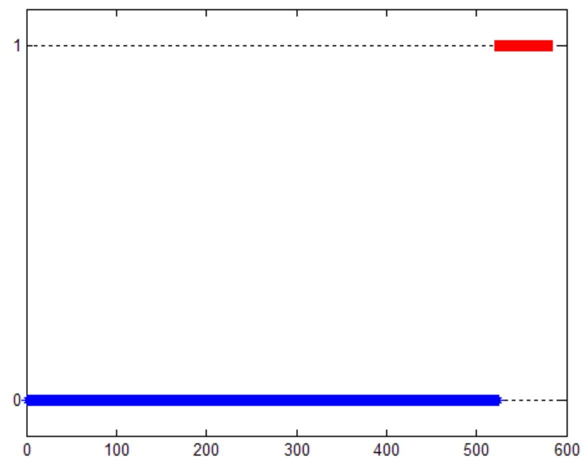

Figure 5: The final detection decision after fusion of the 9 WRST results

5. Conclusions

In this paper, we report our initial study on the use of operational parametric data for the anomaly detection for subsystem of a complex electromechanical system such as a locomotive. A proper analysis of operational data can often provide early fault detection and allow more proactive actions. In this paper, we have demonstrated the use of operational data for fault detection. Specifically, we report a combine strategy of non-parametric test and machine learning. We present the results of applying this strategy to real-world failure cases from locomotives monitored by GE Transportation Systems.

Reference:

[1] Varma, and N. Roddy,. ICARUS: design and deployment of a case-based reasoning system for locomotive diagnostics, Engineering Applications of Artificial Intelligence, vol. 12, no.6, pp 681-690, 1999.

[2] N. Roddy, A. Varma, S. Linthicum, Managing Case Quality in a CBR Diagnostic System, *International Conference on Artificial Intelligence*, 2005.

[3] Gibbons, J.D., *Nonparametric Statistical Inference*, McGraw-Hill, 1971.

[4] H. Mann, and D. Whitney. “On a test of whether one of two random variables is stochastically larger than the other”, *The Annals of Mathematical Statistics*, 18(1), pp. 50-60, 1947.

[5] F. Wilcoxon. “Individual comparisons by ranking methods.” *Biometrics*,1, pp. 80-83, 1945.

[6] N. H. Eklund, K. F. Goebel, “Using neural networks and the rank permutation transformation to detect abnormal conditions in aircraft engines”, *IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications*, 2005.