

Gear Fault Detection and Classification Using Learning Machines

Paul C. Moster, Datatek Applications, Inc., Bridgewater, New Jersey

This article demonstrates that a machine can learn to distinguish different types of faults in gear motors after being shown only a few examples. Simulated cracks were introduced into the gear trains of a group of identical gear motors, and each was instrumented with an accelerometer. Pairs of motors had cracked gears installed in the same position in their trains, allowing one motor to be used as an example, while the second motor was used to test the success in learning. Demodulation techniques were applied to the accelerometer signals, then the modulation frequency spectra were used as input to a learning machine. First, we show the machine the spectra of a set of motors with damaged gears installed at several different positions in the motor's gear train, then demonstrate that it correctly classifies a different set of motors with damaged gears in the same positions. Next, we train the machine with a small set of motors with no gear damage, then show that it accurately distinguishes motors with no gear damage from those with gear damage in any position.

Could a machine learn to do the job of a vibration analyst? Probably not, but learning machines could become a valuable addition to the analyst's toolbox. Recent developments in the field of machine learning have produced simple machines that can recognize handwritten digits as well as a human. Could these machines learn to recognize the vibration signature of a faulty machine and to discern the type of fault from the signature? In this article, we conduct experiments on the test bench that explore these possibilities.

Gear Motor Test Fixture

For our test bench, we obtained eight Merkle Korff SF series gear motors. These gear motors consist of a shaded-pole induction motor connected to an eight-gear speed reducer. This is shown schematically in Figure 1. The gear sizes and gear mesh frequencies are shown in Table 1. The motor shaft S1 drives pinion P1 that meshes with gear G1. Gear G1 drives shaft S2, which drives pinion P2. This pattern repeats for shafts S3 and S4. Gear G4 drives the output shaft S5.

Kistler type 8704B50 accelerometers were attached to each motor. These accelerometers have integrated charge amplifiers with a nominal sensitivity of 100 mV/g and a nominal bandwidth of 50 kHz. A 10-32 tapped hole was made in the casting of the motor approximately 2.5 cm from the axis of the output shaft. The area around the hole was scraped smooth and the accelerometer stud screwed into the hole. Figure 2 shows the gear motor with the accelerometer attached.

Four of the motors were disassembled. Gear G1 was removed from two of the motors and G2 was removed from the other two. Simulated cracks were introduced into these gears by sawing 0.8 mm wide slots between two gear teeth on a horizontal milling machine. The slot depth was approximately 75% of the

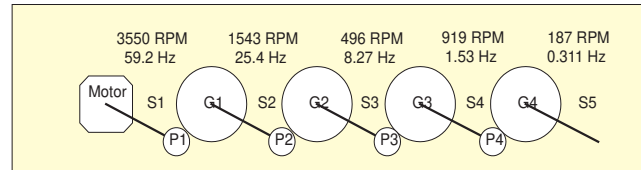


Figure 1. Schematic view of gear motor.

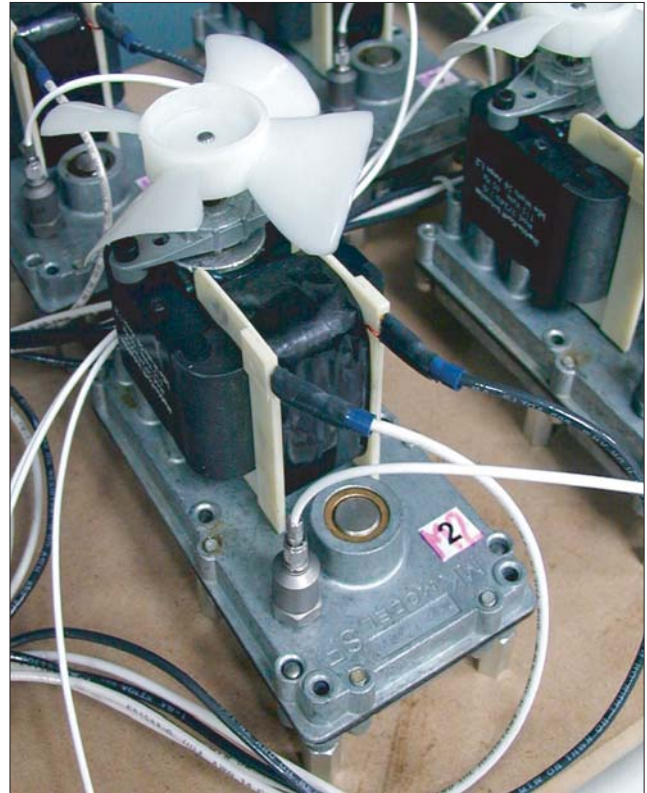


Figure 2. Gear motor with accelerometer attached.

radius of the gear. A modified gear is shown in Figure 3. The modified gears were installed and the motors reassembled.

Six motors were attached to a 16 in. x 36 in. piece of 3/4 in. medium density fiberboard, together with power switches for the motors and our data acquisition and analysis equipment. This formed our gear motor test fixture and is shown schematically in Figure 4. Two motors had simulated cracks in gear G1, two had simulated cracks in gear G2 and two had no modified gears. Table 2 shows the configuration of each motor on the board. Each motor's accelerometer was attached to a separate input channel.

Time Domain Analysis

We collected examples of the signals from the accelerometers on motors M1, M3 and M5. This set includes one good motor, one with a crack in gear G1 and one with a crack in gear G2. A sequence of 1024 samples was collected for each motor with a sampling period of 320 μ s. Figure 6, Figure 7 and Figure 8 show these examples. The signals from the accelerometers on motors M2, M4 and M6 are similar, and are not shown.

The accelerometer output suggests that misalignment of the gear teeth in the neighborhood of the simulated crack gener-

Table 1. Gear sizes and mesh frequencies.

Shaft	Pinion	Gear	Pinion Teeth	Gear Teeth	Shaft Speed (RPM)	Shaft Rotation Frequency (Hz)	Gear Mesh Frequency (Hz)
S1	P1	-	20	-	3555.0	59.167	-
S2	P2	G1	18	46	1543.5	25.725	1183.3
S3	P3	G2	10	56	496.1	8.269	463.0
S4	P4	G3	11	54	91.9	1.531	82.7
S5	-	G4	-	54	18.7	0.312	16.8

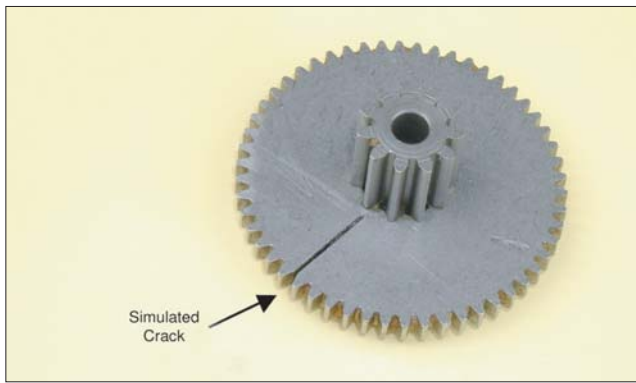


Figure 3. Gear with simulated crack.

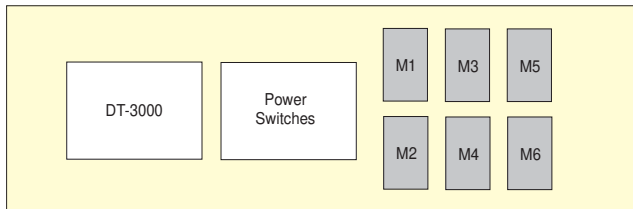


Figure 4. Schematic view of test fixture.



Figure 5. Test bench in laboratory.

ates a periodic impulse that excites a resonance in the motor case. The periods of these impulses correspond to the rotational periods of the damaged gears. The signal from the motor with no gear damage does not contain this periodic impulse. All motor signals contain background noise consisting of a variety of shaft rotation and gear mesh frequencies. This is consistent with what one hears when simply listening to the motors. All the motors are noisy when running, but the motors with damaged gears produce a distinct ticking at the rotating frequencies of the damaged gears.

While time domain data could be used as input to a learning machine, we can provide a less ambiguous input by moving to the frequency domain.

Demodulation Analysis

We performed a demodulation analysis over a modulation frequency range of approximately 5 to 75 Hz with a resolution of 0.37 Hz and a nominal carrier frequency of 3000 Hz with a bandwidth of 200 Hz. We used the wide-range technique described in the **Spectral Correlation vs. Demodulation** sidebar. Five spectra were averaged together to form the results. In total, 163,840 time samples were collected over approximately 13.5 sec. Figures 9-14 show the results for each of the motors.

The spectra for motors M1 and M2 show strong peaks at multiples of the rotation frequency of damaged gear G2, while

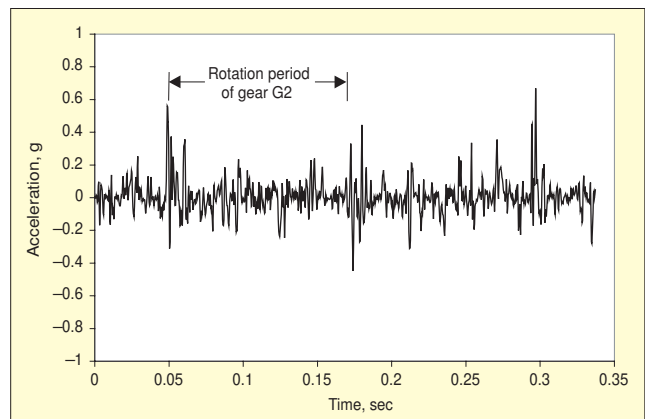


Figure 6. Accelerometer signal from motor M1 with defective gear G2.

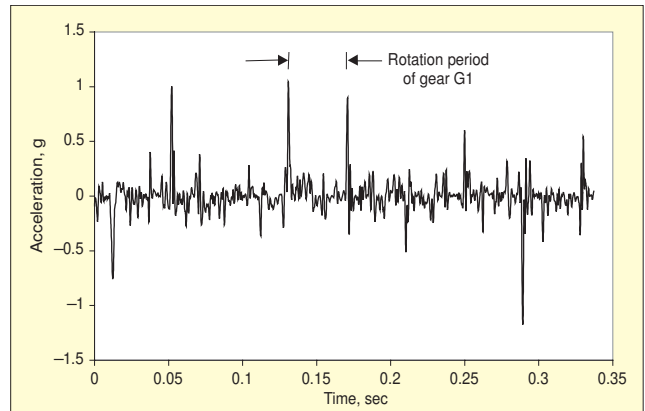


Figure 7. Accelerometer signal from motor M3 with defective gear G1.

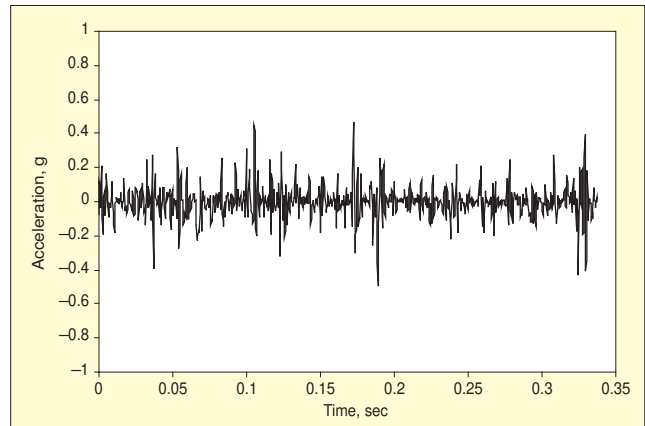


Figure 8. Accelerometer signal from good motor M5.

motors M3 and M4 show strong peaks at the rotation frequency of damaged gear G1, and smaller peaks at the rotation frequency of G2. All the spectra show a peak at the rotation frequency of shaft S1 and pinion P1, but it is most apparent for motors M2, M5 and M6. Although the sizes of the peaks vary between pairs of motors with identical faults, the locations are the same. When normalized, their spectra appear nearly identical. Overall, the modulation spectra are consistent with the respective time domain data in Figures 6-8.

Table 2. Gear motor configuration.

Motor	Configuration
M1	Simulated crack in gear G2
M2	Simulated crack in gear G2
M3	Simulated crack in gear G1
M4	Simulated crack in gear G1
M5	No gear modifications
M6	No gear modifications

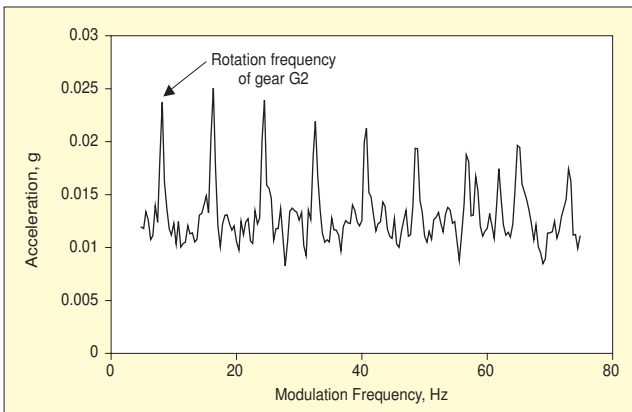


Figure 9. Modulation spectrum for motor M1 with defective gear G2.

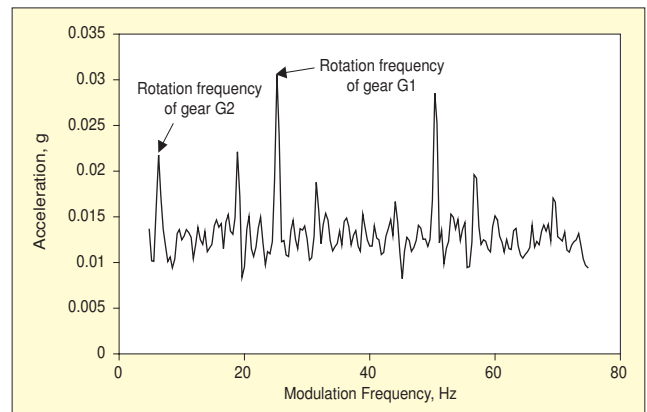


Figure 12. Modulation spectrum for motor M4 with defective gear G1.

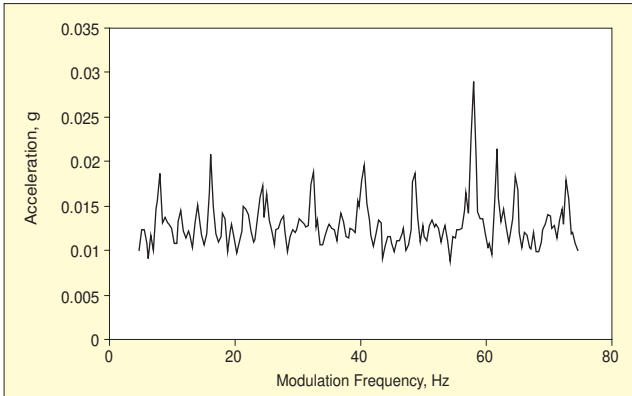


Figure 10. Modulation spectrum for motor M2 with defective gear G2.

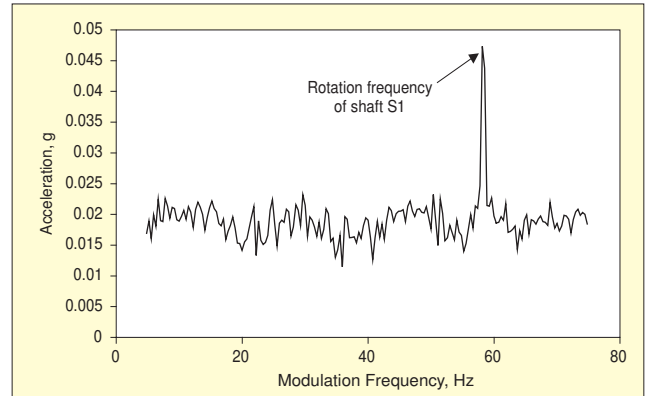


Figure 13. Modulation spectrum for good motor M5.

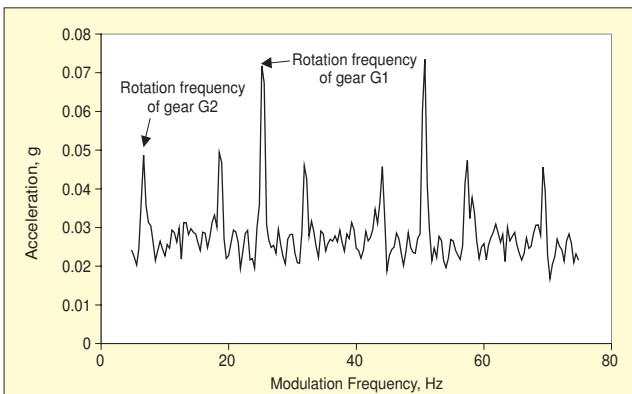


Figure 11. Modulation spectrum for motor M3 with defective gear G1.

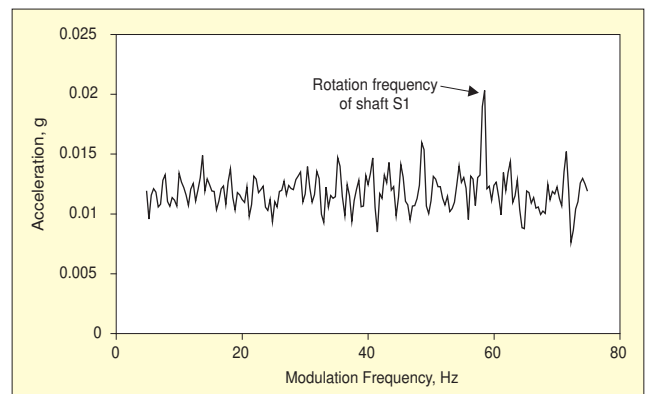


Figure 14. Modulation spectrum for good motor M6.

Fault Classification

Now we test the ability of a machine to learn to recognize the modulation spectrum associated with a particular gear fault. For these experiments, we used a Support Vector Machine, or SVM. (See the **About Support Vector Machines** sidebar for a description of how they work.) First, we collected 20 spectra from motors M1, M3 and M5 to serve as examples from which we trained the SVM. We gave the SVM multiple examples of each motors' spectrum since they vary somewhat over time, and we want it to see the range of variation. In addition, each spectrum was normalized such that the average value is one. We did this because we want the SVM to consider the shape of the modulation spectrum and to worry less about whether one motor is simply louder than another.

Having trained the SVM, we used it to classify at least one hundred spectra from each motor. The SVM renders its judgment about the condition of the gear train based only on its observations of the examples. Table 3 shows a tally of these judgments for each spectrum classified.

The SVM performed perfectly on all the motors. Note that the SVM was not trained on the spectra of M2, M4 and M6, yet

it still managed to classify them correctly. This shows that the SVM successfully generalized its training and was able to extend it to data it had never seen before.

Novelty Detection

Next we tested the ability of a machine to distinguish good motors from those with simulated gear cracks, using only examples of good motors for training. In the world of machine learning, this is referred to as the novelty detection problem. It may seem easier than fault classification, but in practice, it is often more difficult.

We collected 50 spectra from the good motors, M5 and M6, and used them as training examples for the SVM. Then we gave the SVM at least 50 spectra from each of the motors, and had it render a good/bad judgment. Table 4 shows the results.

The SVM performed well given the small number of training examples. However, it made some mistakes: defective motor M2 was classified good 4% of the time. This occurred on spectra where the peak at the rotation frequency of shaft S1 was much larger than the peaks at the rotation frequency of the defective gear. More significantly, the SVM indicated that good

Spectral Correlation vs. Demodulation

Demodulation or envelope analysis is a standard tool of the vibration analyst. From a statistical signal processing perspective, demodulation analysis is just a way of estimating the spectral correlation density function. It is simply a difference in vocabulary.

Figure 15 shows the signal processing involved in a typical demodulation analysis system. The input signal is first band-pass filtered at carrier frequency f_1 , then the filter output is demodulated. The demodulation process consists of frequency shifting the filter output to base-band, then taking the magnitude of the resulting analytic signal. Finally, the demodulator output is band-pass filtered to select a particular modulation frequency α_1 . Digital systems typically use a bank of band-pass filter channels implemented with an FFT to show the modulation frequency spectrum.

In the vocabulary of statistical signal processing, the second band-pass filter has the effect of auto-correlating the demodulator output over a period of $1/\Delta\alpha$ seconds, where $\Delta\alpha$ is the filter bandwidth. While this interpretation is of little interest to the vibration analyst, the statisticians provide an important insight about this system: for a reliable estimate, $\Delta\alpha$ must be much smaller than Δf , where Δf is the bandwidth of the first band-pass filter.¹

In the typical system, α_1 – the center frequency of the second band-pass filter – must be less than Δf . This is intuitive, since the first filter must allow a sideband of carrier f_1 at $f_1 + \alpha_1$ to pass through it if we are to detect modulation

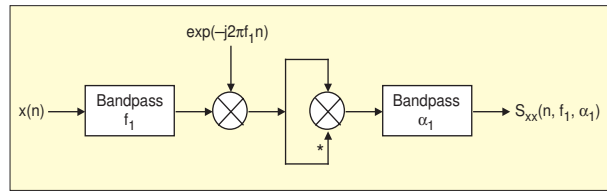


Figure 15. Typical demodulation analysis.

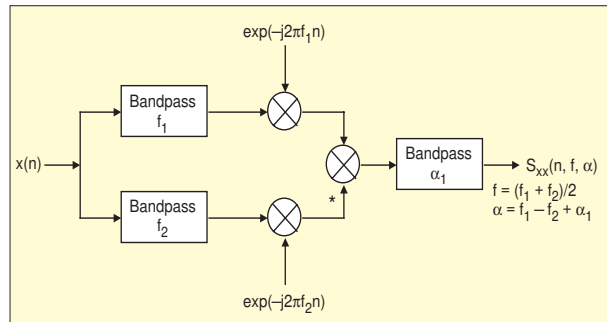


Figure 16. Wide-range demodulation analysis.

motor M6 was faulty 20% of the time. This suggests that the boundary the SVM has established between the spectra of the good motors and the faulty motors is placed too close to the side of the good motors. Further tuning of the training parameters is likely to improve these results. In addition, the current results could be improved by simply applying a threshold. For example, if a motor is deemed good by receiving five or more ‘goods’ in the last ten classifier results, the overall accuracy

Table 3. SVM classification results.

Motor	G2 Fault	G1 Fault	Good	Accuracy
M1	118	0	0	100%
M2	170	0	0	100%
M3	0	126	0	100%
M4	0	111	0	100%
M5	0	0	100	100%
M6	0	0	107	100%

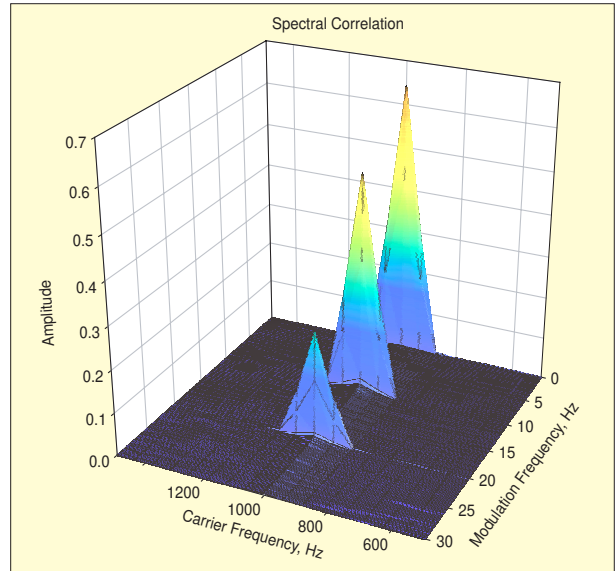


Figure 17. 1000 Hz sine wave with 10 Hz AM.

of f_1 by α_1 . Thus, when the first filter has a narrow bandwidth, you are limited to examining a small range of modulation frequencies starting at zero.

The system in Figure 16 overcomes this limitation and can be used to examine a wide range of modulation frequencies. The input sequence is supplied to two band-pass filters with center frequencies f_1 and f_2 respectively. The filter outputs are shifted to base-band, then the upper sequence is multiplied by the complex conjugate of the lower sequence. Finally, the output is fed to a second band-pass filter with center frequency α_1 . The output of the second band-pass filter represents the power at $f = (f_1 + f_2)/2$ and modulation frequency $\alpha = (f_1 - f_2) + \alpha_1$. When $f_1 = f_2$, the system in Figure 16 is identical to the system in Figure 15. However, by adjusting the separation in the center frequencies of the first band-pass filters, we can examine a wide range of modulation frequencies without sacrificing selectivity in carrier frequency.

Figure 17 shows the spectral correlation for a simulated 1000 Hz carrier with 10 Hz amplitude modulation computed using the wide-range system. Here we show the output as we vary both the carrier and modulation frequencies, producing a 3-D plot. Note that the carrier frequency resolution is much higher than the modulation frequency resolution, a consequence of the requirement that $\Delta\alpha \ll \Delta f$. In the gear motor experiments, we dispense with the 3-D plots and simply show amplitude vs. modulation frequency for a fixed carrier frequency.

would be perfect.

Conclusions

Our results indicate that a learning machine can accurately identify the type of fault in gear motors, given examples of each fault. Similarly, it can tell good motors from faulty motors, given only examples of good motors. It accomplished these tasks without human intervention.

Table 4. SVM novelty detection results.

Motor	Bad	Good	Accuracy
M1	78	0	100%
M2	66	3	96%
M3	244	0	100%
M4	63	0	100%
M5	0	50	100%
M6	19	76	80%

About Support Vector Machines

Here is a trivial, but illustrative example of how an SVM classifier works. Suppose our problem requires that we distinguish between two classes of gear motors, good and bad, and that we have examples of each. Further, suppose that our spectrum analyzer provides just two spectral lines at frequencies f_1 and f_2 . First, we collect data with our spectrum analyzer from the motors we are using as examples. This is our training set. We plot the amplitude of the line at f_1 vs. the amplitude of the line at f_2 , making circles for the good motors and squares for the bad motors (Figure 18). The SVM training computation finds the line that separates the circles from the squares and that minimizes the distance from any point to the line. In this example, three points are at this minimum distance, or margin, and support the separating line. These are the support vectors, since you can view each point as the tip of a vector coming from the origin. To classify the spectrum of an unknown motor, we simply determine whether the corresponding point on the plane resides above or below the separator.

It is easy to extend this technique to hundreds of spectral lines, but it becomes exceedingly difficult to graphically represent examples. Where our two-line spectrum is a point on a plane, a 200-line spectrum is a point in a 200-dimension hyperspace. Instead of a line, the separator is a hyperplane.

In our example, we were lucky that a straight line separated our spectra. Figure 19 shows an example where a straight line will not separate the good motors from the bad, since the bad motors have the good ones surrounded. Fortunately, SVMs provide an elegant method for handling this by allowing us to map our input space to a higher dimension feature space. For example, we can take our two-dimensional plane and wrap it around a three-dimensional globe.

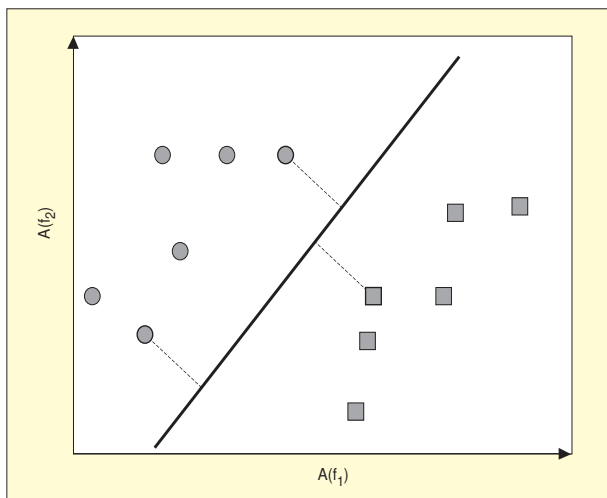


Figure 18. Simple SVM classifier.

While our tests used inexpensive gear motors that would not be candidates for vibration monitoring, we believe that this does not detract from the applicability of our results to larger machines. To the contrary, the gear motors used in this experiment are a complex mechanical system, albeit small, and they produce vibration signals with a complex spectrum. Gear mesh frequencies are modulated by shaft rotation frequencies as a result of mechanical tolerances, and the harmonic frequencies of the slower shafts overlap with the fundamental frequencies of the faster shafts. Resonances in the castings further color the vibration spectra. In larger systems, where the motors and gears are much larger relative to the size of the accelerometer, and where castings and weldments are more rigid, the vibration signals will be dominated by the component to which they are

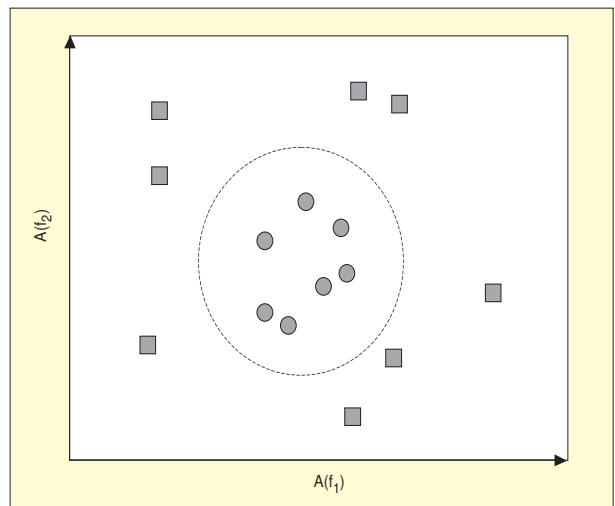


Figure 19. Training set that requires mapping to a higher dimension.

This puts the good motors on the North Pole, and the bad motors in the Northern Temperate Zone. A plane cutting through the Arctic Circle will separate the good from the bad.

While moving to a higher dimension feature space can separate the training data, the computations would be intractable for an already high dimension input space were it not for the kernel trick. For some very useful mappings, scalar (or dot) products in the feature space can be computed in the input space by using the appropriate kernel functions. For SVMs, the training and classification problem can be arranged so that the only computation we need to perform in the feature space is a scalar product.

In the real world, our training data will be noisy. If we try too hard to separate every good point from every bad one, we risk overfitting our data. Suppose that a nearby lightning strike drives one of the good points into a distant sector of hyperspace. It would be better to discount it, rather than to shift our separator about to include it. To this end, SVMs use a regularization constant to assign a cost for violations of the separator in the training set. The higher the cost of violations, the less noisy we assume our training data to be.

SVMs can also address the problem of separating good motors from bad when our training set only includes examples of good motors. Returning to our image of Figure 19 wrapped around a globe, a plane that separates the good motors from the center of the Earth will do a good job of separating them from points that appear anywhere else on the globe.

SVMs are just one example of kernel-based learning algorithms. The mathematics underlying these algorithms is subtle and delightful, and we have provided only the most superficial treatment here. Reference 2 provides an excellent overview. For more information, see www.kernelmachines.org.

closest.

Our experiments also show that the output of a wide-range demodulation analysis is well matched to the input of an SVM. Specifically, SVMs perform well with a high dimension input space, or in the context of vibration analysis, with a large number of spectral lines. Our tests worked well with a 200-line spectrum, but we introduced faults into only two of the eight gears. To identify faults in the remaining six gears, the SVM would need to observe a wider range of modulation frequencies without sacrificing resolution. This increases the number of spectral lines and the dimension of the input space.

Finally, our results suggest that learning machines won't replace the vibration analyst, but could serve as valuable adjuncts. In our experiments, a vibration analyst examined the

About the DT-3000

The DT-3000 Autonomous Vibration Monitoring System is the first product to combine data acquisition, feature extraction, novelty detection, classification and alarm reporting in a compact, industrial network appliance. Where conventional vibration monitoring equipment simply collects data for an engineer to analyze later, the DT-3000 can be trained to recognize and report the presence of signals that indicate incipient failures.

Beyond simple data acquisition, the DT-3000 uses advanced statistical signal processing techniques, like Spectral Correlation Density, with variable bandwidth and resolution to reveal features in vibration signals that are invisible to conventional spectrum analysis. But the DT-3000 is more than just a better spectrum analyzer. The processed signals are supplied to classifiers based on advance machine learning techniques, like Support Vector Machines, to provide a complete interpretation of the signal.

Driven by a 900 Mflop DSP, the standard DT-3000 provides 8 A/D channels with 16-bit resolution, 10 or 30 kHz bandwidth and built-in 4 mA current sources for ICP® sensors. Plug-in data acquisition cards can support different types of transducers. A 10/100 Base-T port provides network access to the DT-3000, while two serial ports provide local



Figure 20. Autonomous Vibration Monitoring System.


or modem access. Ten general-purpose opto-isolated I/O ports allow the DT-3000 to drive relays and actuators or sense contact closures. The DT-3000 comes with a graphical user interface for easy setup and configuration

Datatek Applications, Inc. specializes in the design and development of high-performance real-time hardware and software for multiple industries, including telecommunications, aerospace/defense and scientific research.

For more information on Datatek and the DT-3000, contact Steve Cokinos at scokinos@datatekcorp.com or visit the website at www.datatekcorp.com.

motors, selected the frequency ranges and supervised the training, using skills that no machine yet possesses. However, once trained by an expert, the machine did well at monitoring and fault classification. This suggests that learning machines could take over repetitive tasks, or those requiring 24×7 monitoring, or those where the sheer volume of data overwhelm a human analyst. This is the promise of machine learning and we look forward to its fulfillment.

References

1. Roberts, R. S., Brown, W. A., and Loomis, Jr., H. H., "Computationally Efficient Algorithms for Cyclic Spectral Analysis," *IEEE SP Magazine*, pp 38-49, April, 1991.
2. Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., and Scholkopf, B., "An Introduction to Kernel-Based Learning Algorithms," *IEEE Trans. Neural Networks*, Vol. 12, No. 2, 2001. 

The author can be contacted at: pcmoster@datatekcorp.com.