HOW GOOD IS YOUR PREDICTOR? EXPANDING CONFIDENCE INTERVALS TO DEFINE PROBABILITY DENSITIES ON ADAPTIVE PARAMETERS

Mark Dzwonczyk & Teresa H. Y. Meng

Department of Electrical Engineering, Stanford University, Stanford, California 94305
Primary Author Email: mdz@leland.stanford.edu

ABSTRACT

A method is proposed to measure the performance of linear predictors as they track non-stationary stochastic processes. Classical linear regression techniques are combined with a novel use of instantaneous error to define the likelihood that the coefficients of a linear predictor adequately capture a system's state. The resultant probability measure serves as a metric of predictor performance: a probability near unity indicates that the predictor is performing well, while a probability near zero indicates the state of the system is poorly captured by the coefficients. The approach is extended to trace coefficients, weighted by these probabilities, as they move about in a space of possible states. The probability measure provides an instantaneous confidence measure of the route that the system proceeds upon within that space: a likelihood roadmap of the state of the system through time. Specifically, the method is applied to the important problem of predicting the vibration signature of rotorcraft gearboxes as they mechanically fail. Actual data from US Navy drivetrain teststands are used to validate the method and underlying assumptions.

1. INTRODUCTION

Adaptive predictors attempt to learn the underlying process parameters as the system that they are tracking evolves. Although these predictors can be used to track stationary stochastic processes — in which they first "learn" the process by modifying their coefficients to minimize some error criteria and then maintain those coefficients — they are more valuable for non-stationary processes whose parameters, by nature, change with time. Unsupervised adaptation, whereby a feedback path is provided from the predictor error to the coefficient modifier, is an especially valuable learning method for non-stationary processes since it allows for autonomous tracking of the changing phenomenon.

Although the principal use of predictors is to suggest the most likely outcome of the stochastic process at some sample point in the future, the predictor coefficients themselves can provide valuable information about the state of the process being modeled. That is, as an estimator of the process in the future, the predictor contains information about the state of the process. For processes that naturally progress from one state to another, although at an unknown rate, the current state information can be quite valuable. The application of interest to the authors — health prognosis of rotorcraft drivetrain assemblies — is an important example of such a process. The cyclic nature of the machinery means phenomena in the assemblies is best observed by monitoring time-series vibration signatures [1]. As these systems fail, they evolve from a "healthy" state to one of several possible "failed" states. Interim states between these two end states can provide information about the likely type of failure that is causing degradation. When this information is coupled with prediction error, process history, and past trends, prognoses can be conjectured about the remaining usable lifetime of the gearbox or other assembly under observation.

An important determinant of predictor performance is obviously the accuracy of the coefficients, since it is the coefficients that weight previous inputs to form a sum that represents some future value. When considering predictor coefficients to be representative of the system state, their accuracy is even more consequential to performance. This paper presents a novel method for determining the accuracy of predictor coefficients in the context of representing system state. A probabilistic approach is used to determine a likelihood measure for the coefficients of linear predictors that track non-stationary processes. A classical linear regression error measure is used to define confidence intervals on the coefficients. This is coupled with a short-term error measure that addresses the time-varying transitions from one state to another. Specifically, for the application described above, the short term error tracks coefficient accuracy as the machinery moves from "healthy" to "faulty". Confidence intervals based upon regression error are created once the process becomes temporarily stationary. A probability density function is then created over coefficient space — representing process state space — to provide a map that suggests where and the how the machinery is progressing from healthy to faulty. This map will allow rotorcraft operators to make strategic decisions about flight risks related to the drivetrain.

2. LINEAR PREDICTORS & ERROR

Consider a conventional discrete-time m-stage linear predictor [2] in which some future output $y(n+1)$ is formed by the vector product of a set of previous inputs $\boldsymbol{x}$ and predictor coefficients $\boldsymbol{w}$.

$$y(n+1) = \boldsymbol{x} \bullet \boldsymbol{w}\,; \qquad \boldsymbol{x} = \begin{bmatrix} 1 \\ x(n-m) \\ x(n-m+1) \\ \vdots \\ x(n-1) \\ x(n) \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{m-1} \\ w_m \end{bmatrix}$$

where the bias weight is included in $\boldsymbol{w}$ as w_0. The predictor will have an instantaneous, error $e(n) = y(n+1) - x(n+1)$ which can be used to modify the predictor weights using Least Mean Squares (LMS) minimization or its variants [3] [4].

2.1 Long-Term Regression Error

If the predictor weights are modified using LMS, then these weights can be treated as regressors, that is, regression coefficients in classical linear estimation [5]. Moreover, confidence intervals can be defined for each of the $m+1$ individual predictor coefficients in $\boldsymbol{w}$. The width of each confidence interval, or wholly, the volume of hyperplane defined by $m+1$ confidence intervals, is a measure of how well the predictor has captured the process being modeled.

Specifically, the 100(1-α) percent confidence interval on coefficient w_j is

$$\hat{w}_j \pm t\left\{\frac{\alpha}{2}, M-m\right\}\sqrt{\sigma^2 C_{jj}}$$

where $t\{\frac{\alpha}{2}, M-m\}$ is the t-distribution of order $\frac{\alpha}{2}$ and degree $M-m$; M is the number of training epochs, and $(\sigma^2 C_{jj})$ is the j^{th} diagonal element of the covariance of $\boldsymbol{w}$.

2.2 Short-Term Instantaneous Error

Confidence intervals can provide a good estimate of the accuracy of the predictor coefficients for a stable, i.e., stationary, system. However, non-stationarities in the process will cause the confidence intervals to grow very large and render them meaningless. When this occurs, a different error measure must be used.

To complement the "long-term" regression-based error analyses, a "short-term" instantaneous error, namely the mean-square error (MSE) is used. It is computed from the instantaneous error with an exponential decay. Growth in MSE is indicative of poor predictor performance. Thus, MSE is used as a primary gauge on coefficient accuracy. Once MSE has settled to a nominal value and coefficients become stable, the long-term regression-based analyses can be used to form confidence intervals. Conversely, if (when) MSE again rises, MSE is used as a measure on the accuracy of the predictor coefficients and the confidence intervals are ignored (and reset) until the process stabilizes

3. DEFINING A LIKELIHOOD MAP OVER PARAMETER SPACE

Thus, we propose a system in which two error metrics are used to track predictor performance. MSE is the primary indicator and regressor confidence intervals are used once MSE has settled. Both are inversely proportional to the probability that the predictor coefficients have captured the state of the process being observed: as MSE decreases and as confidence intervals become tighter, the likelihood is greater that the predictor has adequately captured the system state evolution. More rigorous mathematical work is in progress that quantifies the link between the two "error" terms and a true probability density function. For this presentation, we generalize the concept and provide real-world example data and results.

As described above, if MSE is high (or changing) the confidence intervals provide little information about the underlying process. Thus, during those times, the MSE should be used exclusively as a some measure of likelihood that the predictor's coefficients have captured the underlying process state.

Consider the MSE of the predictor and only one predictor coefficient, w_j. In general, if MSE is low, then that coefficient is likely to be correctly capturing the system state. In the context of that single parameter, the system can be assumed to be near that parameter value. If the MSE is high — again, in general — then the actual system state may or may not be near that coefficient value.

One can view error in terms of a likelihood measure: as error increases, the system is likely to be moving from one state to another. If error is consistently high, a number of different states are likely. If error is low, the state is specified with a higher likelihood. Thus, we can consider MSE to define a cone over state space within which probabilities can be determined. If MSE is high, the true value of w_j may take on a range of values and one can assert a relatively low probability that the current value of w_j is an accurate choice. Conversely, for example, if MSE is precisely zero, then one may consider claiming that w_j is correct with high probability. This is illustrated in Figure 1, below, in one dimension (for one coefficient). For two coefficients, MSE over parameter space should be considered as a planar ellipse and in higher dimensions, it will be a hyper sphere (hyper ellipse).

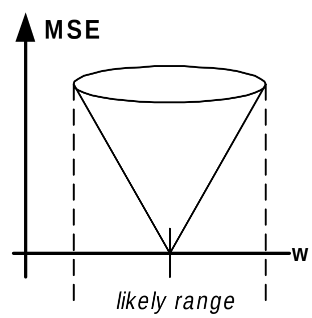

FIGURE 1. *Probability Measure Over Single Parameter*

Now consider a system that is moving between two stationary states, or from one well-defined state to another unknown state. This process is shown below (Figure 2) in the context of gearbox failures. If one adds another dimension of "likelihood" to this trajectory map, that is, the likelihood measure based upon MSE, then one might be able to depict probabilistically the trajectory from a healthy gearbox to a faulty gearbox. Of course, previous states will be well known (if they were captured by the predictor) and likelihoods of future states will be poorly defined. At any given time, however, the likelihood map could be depicted that characterizes the system evolution.

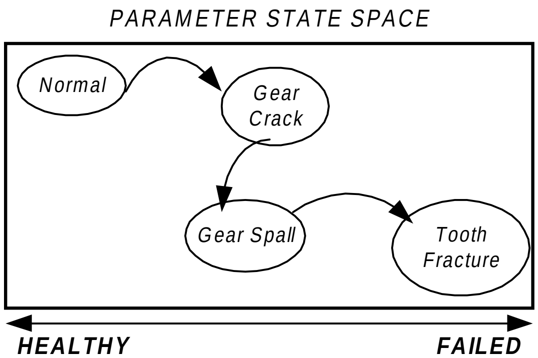

FIGURE 2. *Map of Trajectory Through Parameter Space.*

Confidence intervals are used once MSE is reduced to an acceptable level — that is, once the likelihood of being in particular state, as defined by the coefficient vector, is unacceptably large. Confidence intervals, based upon regression analyses, can then be used to give a more detailed measure of how well the coefficient vector captures the state of the underlying process.

4. A REAL-WORLD PROBLEM

This method has been applied to an important problems in the domain of aircraft health and usage monitoring: *rotorcraft transmission health prognosis*. With diminishing maintenance budgets and fleet sizes, the US Department of Defense is facing serious strategic and logistical issues involving mission-readiness for their helicopter deployments [6]. The realities of this problem are clear from the perspective of costs and safety. For example: the average costs of removing, repairing, and reinserting a gearbox in the US military fleet is $75K [7]; and nearly one-third of all serious (Type-A) military helicopter accidents can be traced to the drivetrain components [6]. A health prognosis system which assists operators in determining the expected remaining lifespan of the transmission components can both reduce costs by requiring only condition-based maintenance (versus time-based maintenance) and increase safety by detecting anomalies before they can evolve to become catastrophic.

An active research program at the NASA Ames Research Center [10] [11], in cooperation with its sister Lewis Research Center, the Army Aviation and Troop Command, and the Naval Air Systems Command, has been focused on the health and usage monitoring of rotorcraft drivetrain assemblies. As part of this program, vibration data from both transmission testrigs and actual vehicles have been collected and archived at Ames for analysis.

Vibration data from a controlled teststand for the intermediate gearbox of a TH-1L helicopter was used to evaluate the probabilistic methods presented in the previous section. This so-called *Hollins Dataset* [8] is well-suited for this evaluation because it contains multiple sets of vibration data under healthy conditions, plus five different types of gearbox faults (Figure 3).

Each of these test conditions was monitored by three accelerometers mounted at various locations on the gearbox and sampled at high rates during controlled experiments. These multiple looks at different types of faults have been concatenated together into several larger data sets. As such, they resemble a series of short-term stationary processes and are thus ideal for testing the efficacy of the system.

DATA LABEL	CONDITION OF GEARBOX
B, C	No Defect
D, E	Inner Race Fault
F	Rolling Element Fault
G, H	Outer Race Fault
I	Gear Spalls
J	Half Gear Tooth Removed

FIGURE 3. *Types of Faults in the Hollins Dataset*

For the sake of brevity, details about the source data sets have been omitted from this paper, but the reader is directed to previous works that have used the same data set for similar evaluations, for example, [9]. The original data was down-sampled to speed-up processing, but the discarded samples showed no effect on system performance except to lengthen the necessary filter size.

5. RESULTS WITH TH-1L DATA

Predictors were first trained with individual runs of healthy and faulty vibration data sets with good results. Figure 4 shows the predictor convergence for data set *B, No Defect*. The predictor was m=16 stage with accumulated LMS error minimization applied every 12 cycles. The linear predictor tracked the system well; similar results were achieved for the faulty data sets.

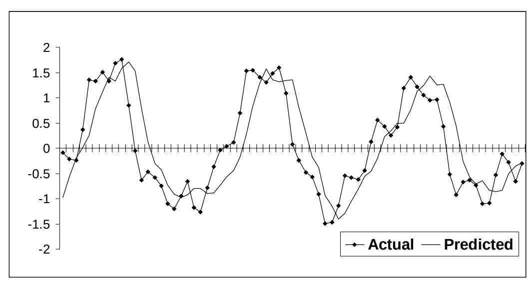

FIGURE 4. *No Defect Example of Predictor Tracking Actual*

To test the utility of the two error measures, non-stationarity in the data was modeled by imposing a sharp discontinuity: a faulty data set, specifically data set *J, Half Gear Tooth Removed*, was appended to the data set *B*. The predictor continued to adapt, but as expected, its MSE increased dramatically (constrained only by the lag in the mean term) before it could converge upon new coefficients that represented the new system state. The MSE immediately before and after the data discontinuity is shown in Figure 5.

Similarly, the confidence intervals expanded dramatically at the discontinuity. This is shown directly below in Figure 6. Once MSE had settled to nearly its previous level, the confidence intervals were reset (by hand) and recomputed.

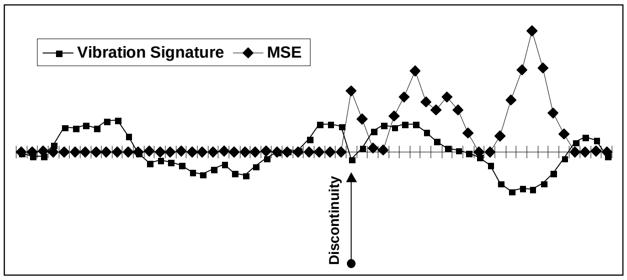

FIGURE 5. *Changes in MSE at the Due to "Fault Insertion" from No Defect (B) to Half Gear Tooth Removed (J)*

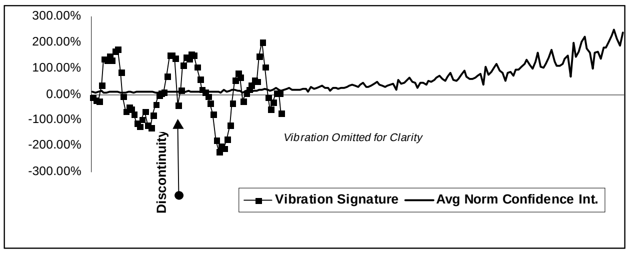

FIGURE 6. *Average Changes in (Normalized) Confidence Intervals at the Same Transition*

In Figure 6, the confidence intervals are represented by a normalized average of the 17 coefficients in ***w***. Normalization was first performed to represent each interval as a percentage of the individual coefficient. The 17 normalized coefficients were then averaged. This, of course, gives a qualitative view of the performance and is not truly indicative of how well the predictor captured the underlying process. Quantitative analyses of how individual regressors can be mapped into a probability measure are in development.

6. CONCLUSIONS & FUTURE WORK

This paper has illustrated a method of using both short-term MSE and longer-term regressor confidence intervals as probability measures on the performance on linear predictors. The results so far — with real data relating to an important problem — are appealing. Still, there is important work ahead.

We are presently formulating the mathematical constructs that map the error terms (MSE and regressor confidence intervals) into a probability metric. To facilitate this process, we are examining normalized error (and N-LMS error minimization), which should better lend itself to probability distributions. We will broaden that work to formulate a high-dimensionality probability function over parameter space that can be interpreted to provide a likelihood estimate on predictor performance.

We are automating the transition between MSE and regressor confidence intervals, for this particular application, in the hope of better understanding the performance tradeoffs and inter-dependencies of these two error terms.

Finally, we are applying these methods to other data sets in this same problem domain that are naturally less stationary (i.e., possess more non-stationarities). This work continues with the support and resources of NASA and the US Army at Ames Research Center with the goal of automating the prognosis of rotorcraft drivetrains.

7. REFERENCES

[1] Stevens, Hall, & Smith, "A Multidisciplinary Research Approach to Rotorcraft Health and Usage Monitoring," *American Helicopter Society 52nd Annual Forum*, 1996.

[2] Makhoul, "Linear Prediction: A Tutorial Review," *Proceedings of the IEEE*, April 1975, pp. 561-580.

[3] Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.

[4] Widrow & Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

[5] Montegomery & Peck, *Introduction to Linear Regression Analysis*, Wiley & Sons: NY, 1992.

[6] Dzwonczyk, *Assessment of Technologies Related to Integrated Mechanical Diagnostics for the H-53E Helicopter*, Signal Processing Associates Technical Report, SPA-97-047, 15 April 1997.

[7] Dobrin, "Expanding HUMS Horizons," *Helicopter World*, June 1995.

[8] Hollins, "The Effect of Vibration Sensor Location in Detecting Gear and Bearing Defects," *Proceedings of the 41st Meeting of the Mechanical Failures Prevention Group*, October 1986, pp. 49-58.

[9] Monsen, Dzwonczyk, & Manolakos, "Analog Neural Network Based Helicopter Gearbox Health Monitoring," *Journal of the Acoustical Society of America*, Dec. 95, pp. 3325-49.

[10] Dzwonczyk & Huff, "Helicopter Transmission Health Monitoring Using Real- Time Neural Computing Methods," Proceedings of the 1994 IEEE/AIAA 13th Digital Avionics Systems Conference, pp. 359-364.

[11] *NASA Ames Signal Analysis Laboratory Web Site*, http://siglab.arc.nasa.gov/

8. ACKNOWLEDGEMENTS

This work was wholly supported by a grant from the NASA Ames Research Center (NAG 2-1065). The authors gratefully acknowledge the continued support of Dr. Edward Huff of NASA Ames for his continued support of underlying objectives of the research.