Changing Failure Rates, Changing Costs: Choosing the Right Maintenance Policy

Chris Drummond
Institute for Information Technology
National Research Council Canada
Ottawa, Ontario, Canada, K1A 0R6
Chris.Drummond@nrc-cnrc.gc.ca

Abstract

Over the life time of any piece of complex equipment, the likelihood of a failure and the cost of its repair will change. The best machine learning classifier, for predicting failures, is dependent on these values. This paper presents a way of visualizing expected cost which gives a clear picture as to when a particular classifier is the right one to use. Of equal importance, it also shows when a classifier should not be used and a more traditional maintenance policy is the better choice. It distinguishes the conditions when it is best to wait until a part breaks before taking action, or when it is best replace it routinely, at regular intervals. This paper demonstrates how overall, this visualization method gives maintenance personnel the means to adapt to changing failure rates and changing costs.

Introduction

In many industries, the equipment being maintained has a very long life span. Aircraft engines, this author's main focus (Drummond 2004; Létourneau *et al.* 2005), are no exception with expected lifetimes of over twenty years. In twenty years, many things will change. How often a part fails will change. The maintenance organization and original equipment manufacturer will address problems extending a part's life. Other aspects, such as the conditions to which the aircraft is exposed, will also affect failure rates. How much the part costs will change. Parts will fluctuate in price, due to improved manufacturing or other changes outside the maintenance organization's control. The cost of repairs will change. The maintenance organization will work to streamline its procedures to reduce repair time and costs. The consequence of a part failing will change, fuel prices and airport costs are seldom static tending to increase, sometimes rapidly, with time. So, any maintenance policy decided on early on may be totally inappropriate towards the end of an engine's life. The maintenance organization must adapt to these changes. This paper discusses a way of visualizing the trade-offs between different maintenance policies as changes occur.

One maintenance policy, at least when there is no safety issue, is to wait until a part breaks. This may still be a sensible alternative when a fault is rare. In this circumstance, it is very hard to find a predictive algorithm that is sufficiently accurate to do better than this policy (Drummond & Holte 2005). When faults are more common, or the consequence of failure is more costly, a common alternative is to periodically replace components based on some assessment of their lifetime. The policy that is of primary interest to people at this symposium is one based on accurate prognostics. The main focus of this author is in developing machine learning algorithms for predicting faults. Others will develop alternatives such physics of failure models. The performance of all these different policies must be compared. Each policy will do better for some failure rates, and costs, than for others. As conditions change, the best policy is also likely to change. This paper shows a way of visualizing the performance of these polices allowing the maintenance personnel to choose the best one for the prevailing conditions.

That the choice of policy is dependent on the prevailing condition also impacts how we evaluate our own prognostic algorithms. It is not enough to say our model predicts 95% of the equipment's faults, with a 5% false alarm rate. These seemingly impressive numbers may be insufficient if failures are sufficiently rare. We must demonstrate that using a predictive algorithm substantially improves on the performance of the maintenance policies already in place. We must further demonstrate that this is true for a reasonable range of costs and failure rates to have any reasonable confidence in our algorithms' being useful in practice. The method used to visualize the performance of algorithms, presented here, will also function as a valuable evaluation tool.

Visualizing Performance

This section introduces a way of visualizing the expected cost of different maintenance policies. The approach is a specialization of a general method for visualizing classifier performance called cost curves (Drummond & Holte 2006).

To introduce this visualization method, let us begin by imagining that failures are very rare. This is the lower left hand corner of figure 1. Let us suppose we wait for a failure to occur before doing anything. As failures become less rare, the arrow at the bottom of the figure, the overall cost will increase, the arrow at the left of the figure. The cost will be directly proportional to the probability of failure, the expected cost is just the cost of a failure times its probability. If the failure rate is constant but the cost of failure increases we would have the same effect. So the probability of failure and the cost of failure are closely related. If we normalize the product of probability and cost so that it ranges from zero to one we end up with the x-axis of figure 2. As we have normalized the x-axis, the y-axis, the expected cost also ranges from 0 to 1. One is the maximum cost that could occur.

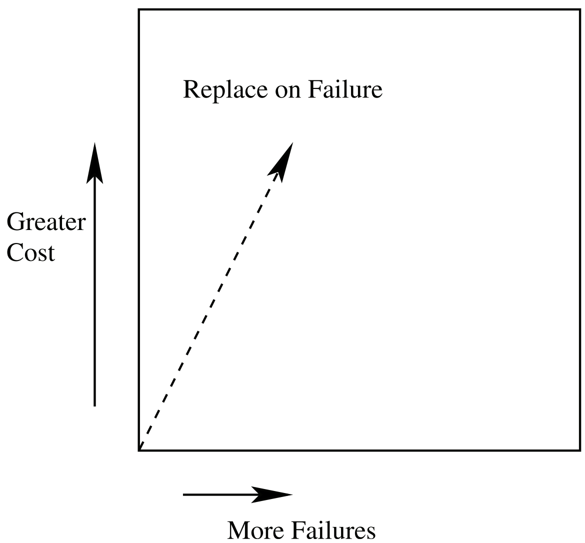

Figure 1: Increasing Costs

If the failures become too common, or the cost of a failure becomes too high, rather that wait until a part fails we would be better replacing it at every available opportunity. Of course, in practice, maintaining the equipment in this circumstance is probably futile. It does however establish a clear region, indicated by the cross-hatched triangle, where useful maintenance policies must operate.

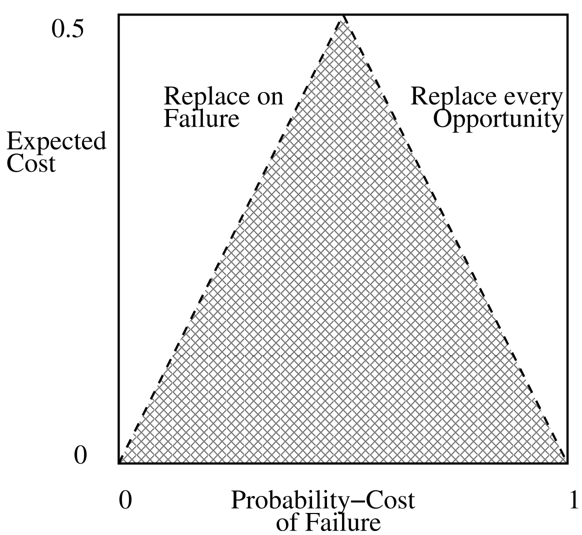

Figure 2: The Operating Region

Alternative Maintenance Policies

To investigate visualizing alternative maintenance policies using this approach, let us imagine monitoring the exhaust gas temperature of a gas turbine engine. As is common practice when using such engines, should the temperature exceed a threshold the engine must be repaired. Sampling from a lognormal distribution, commonly used to model "cycles-to-failure", 1000 samples of different failure times are generated. Figure 3 shows three such distributions. Faults generated by the solid black curve occur very quickly after the engine has been put in service. For the dashed black curve, the time to failure and the variance are increased. Progressively longer and wider distributions model more infrequent faults. The exhaust gas temperature is assumed to rise linearly but is overlaid with Gaussian noise. Many different values for the lognormal distribution will model the different fault probabilities, from very common to very rare.

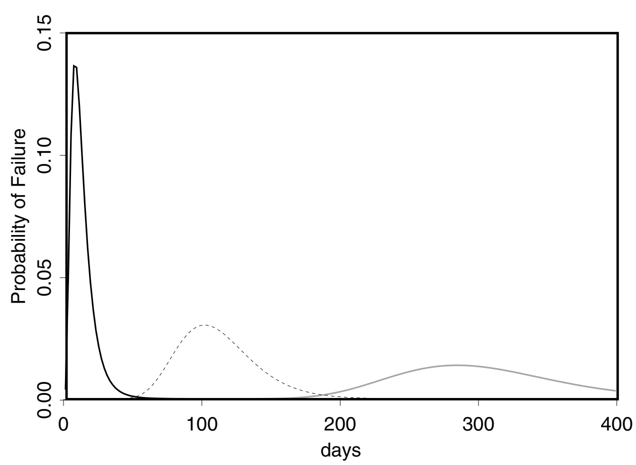

Figure 3: Distribution of Faults

For the purposes of this paper, fault prediction will be couched as a binary classification problem. The aim is to predict the fault within a 20 day period, 5 days before the fault. Let us first consider a routine maintenance policy that recommends replacing the part 45 days after being put in service. Thus it labels 20 days, and on, as positive examples. The performance of the policy is the gray solid curve in figure 4. It has the best performance, the lowest expected cost, in the center of the figure. This will occur when the number of positives and the number of negative is the same, i.e. the average time to failure is, indeed, 45 days. This policy could also be the best at other failure rates, when the cost of failure varies with respect to the cost of repair. When the cost of failure is much higher than the cost of repair, even when the time to failure is longer it is still cost effecient to replace the part at this time. Although routine maintenance is an effective policy, should the failure rate change significantly, it quickly becomes of limited utility. Ultimately, it is better to wait until the part breaks, or replace the part as the opportunity arises, than to rely on this policy.

Of course, we could change the routine maintenance period. The two gray dashed-dotted lines in figure 5 are for

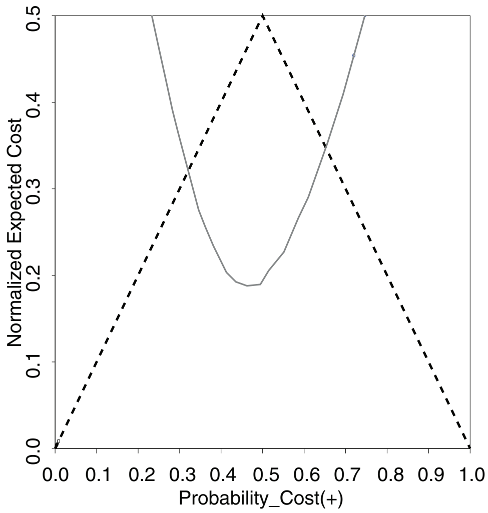

Figure 4: Routine Maintenance

two policies with longer and shorter time periods. These are optimum for different failure rates. What is noticeable, however, that they offer less an improvement over say the replace on failure policy.

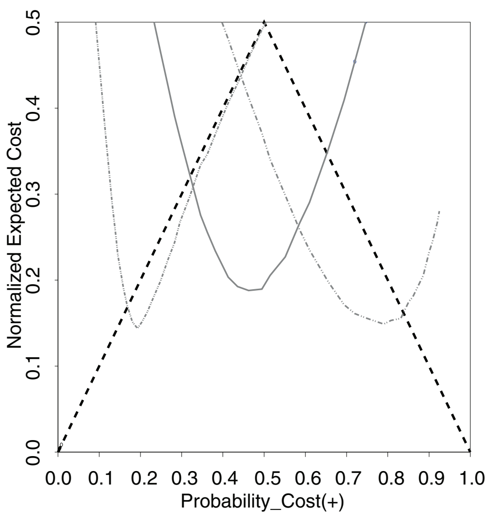

Figure 5: Other Maintenace Periods

Now let us look at the effect of a prognostic algorithm, figure 6. A training set, drawn from the same lognormal distribution, was used to learn at what exhaust gas temperature to recommend replacing the part. The algorithm simply calculated the average temperature 25 days before the problem, across the instances in the training set. When applied to the test data it labels everything as positive after this threshold has been reached. It performs generally better than routine maintenance, although notably its performance is not much better when failures are common, the right hand side of the figure. Indeed if failures are very common, the algorithm is worse than simply replacing the component at every opportunity. In the more likely case, when failure is rare, this prognostic algorithm outperforms any routine mainteance. But it should be noted, that as the rarity increases its advantage over the replace on failure policy decreases until the point that they are indistinguishable.

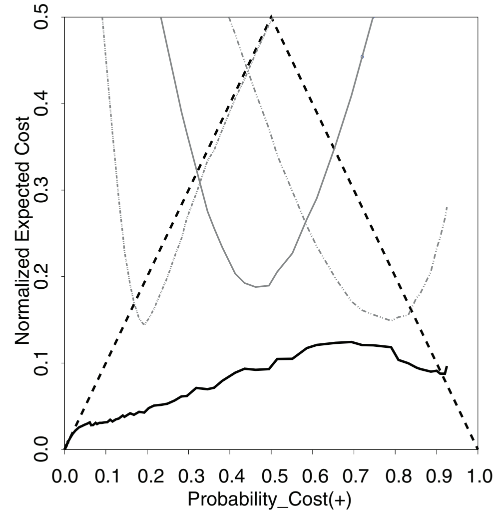

Figure 6: Using a Predictive Algorithm

Conclusions

This paper has shown a way of visualizing the expected cost of different maintenance policies. This will allow the maintenance personnel to make an informed choice when the frequency of failure or failure costs change. This sort of visualization is not only important in normal operation, it is also critical for evaluating prognostic algorithms. Sometimes existing polices are the best and we need to determine when our algorithms are going to be effective rather than spending time where little advantage can be gained.

References

Drummond, C., and Holte, R. C. 2005. Severe class imbalance: Why better algorithms aren't the answer. In *Proceedings of the 16th European Conference on Machine Learning*, 539–546.

Drummond, C., and Holte, R. C. 2006. Cost curves: An improved method for visualizing classifier performance. *Machine Learning* 65(1):95–130.

Drummond, C. 2004. Iterative semi-supervised learning: Helping the user to find the right records. In *Proceedings of the Seventeenth International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, 1239–1248.

Létourneau, S.; Yang, C.; Drummond, C.; Scarlett, E.; Valdés, J.; and Zaluski, M. 2005. A domain independent data mining methodology for prognostics. In *Proc. of Essential Technologies for Successful Prognostics, 59th Meeting of the Machinery Failure Prevention Technology Society*.