

Equipment's Prognostics Using Logical Analysis of Data

Alireza Ghasemi¹, Sasan Esmaili¹, and Soumaya Yacout²

¹ Dalhousie University, Halifax, Canada

{alireza.ghasemi, sasan.esmaeili}@dal.ca

² École Polytechnique de Montréal, Montréal, Canada

soumaya.yacout@polymtl.ca

Abstract. This paper demonstrates the implementation of Logical Analysis of Data (LAD) methodology in the field of prognostics in Condition Based Maintenance (CBM). In this paper the LAD classification methodology, based on Sensitive Discriminating and Equipartitioning methods for data binarization, Mixed Integer Linear Programming (MILP) and Hybrid Greedy methods for pattern generation, is used. Using the generated patterns, two methods of calculating the survival function are introduced. The methodology is applied on Prognostics and Health Management Challenge dataset, which is a condition monitoring dataset provided by NASA Ames Prognostics Data Repository. The results obtained by using LAD methodology, are compared with that obtained by using the Proportional Hazards Model (PHM).

Keywords: Condition Based Maintenance (CBM), Logical Analysis of Data (LAD), Prognostics, Condition Monitoring.

1 Introduction

Widely applied in maintenance, *Condition Based Maintenance (CBM)* [1] was introduced in order to use the equipment's health condition in optimizing or improving the maintenance activities. In recent years, many statistical methodologies have been introduced to model the relation between equipment's age and its health condition and its failure. Since the statistical assumptions are not always met, new methodologies which aim at presenting structural relations between the equipment's conditional survival probability and its age and/or health condition are required.

Logical Analysis of Data (LAD) [2], is a combinatorial optimization and Boolean logic based methodology for the analysis of datasets. The aim of LAD is to extract knowledge hidden in observations of a dataset in order to detect patterns that lead to different classes of observations. A pattern is defined as a conjunction of features or indicators' (*Attributes*' values which are found in one class and never in another class. A class is a distinct state of equipment. In CBM, an attribute can be the monitored equipment's age or a health condition of the equipment. Observations are classified into two classes: observations of failure during the coming period, referred to as the *Positive Class*, and observations of survival at least until the end of the coming period, referred to as the *Negative Class*. A *Positive (Negative) Pattern* is a

conjunction of literals that is reflected in one or more of the observations of the positive (negative) class while not reflected in any of the observations of the negative (positive) class. A *literal* is either a binary attribute or its *Negation*. The historical data base is divided into two sets; the training set and the testing set. LAD constructs the classification model based on the given historical dataset, which is called *Training Set*. Then, by using the classification model, it tests the quality or accuracy of this model by classifying the other part of the historical dataset, which is called *Testing Set*. The former process is *Training Phase*, and the latter one is the *Testing Phase*.

LAD has been widely applied to the analysis of datasets in different fields. Recently, it has been used in equipment's failure diagnostic [3-6], and has proved to be a promising technique that gives practical and interpretable results. Herein, our focus is on calculation of the probability of failure at a certain future moment, which is the prognostic objective of CBM, which has been comparatively untested when using LAD. We have extended the LAD methodology to predict the equipment's chance of survival at each observation's moment, when new observation of the equipment's attributes is available. Unlike the earlier applications of LAD in failure diagnostic, we consider both age and the equipment's condition as LAD's attributes.

The paper is structured as follows: Section 2 describes two data binarization methods and two pattern generation methods. Section 3 introduces two prognostic methods. Section 4 provides the experimental prognostic results obtained by using LAD methodology along with a comparison with those obtained when using PHM. Conclusions are presented in the section 5.

2 Data Binarization and Pattern Generation

LAD deals with Boolean attributes, whereas the monitored attributes may appear in numerical form (e.g. temperature), nominal form (e.g. color), or ordered form (e.g. temperature which can be high, medium or low). Therefore, all the attributes have to be converted to a binary format. Binarization transforms each non-binary attribute into several binary ones, by comparing attribute's values to certain thresholds, called *Cut-Points*. In the literature, there are several approaches to define cut-points [7-9], among which, we will use *Sensitive Discriminating* method and *Equipartitioning* method. In the sensitive discriminating method, a cut-point is defined as the average of two consecutive attribute's values, each belonging to a different class. The outcome cut-point represents a threshold that differentiates between positive and negative classes. In the equipartitioning method, the cut-points are defined in such a way that all the attribute's values are approximately equally divided into a pre-defined number of intervals.

A pattern discriminates between one or more of the observations of a class from all or many of the observations of the opposite class. The basic pattern generation algorithms are mainly based on generating all combinations of literals, and examining whether each of the combinations can be considered as a pattern. Recently, some heuristic methods have been introduced that require less computational effort while providing equivalent performance [10-12]. Among all the proposed pattern generation

algorithms, we will use the *Mixed Integer Linear Programming (MILP)* method and *Hybrid Greedy* method in order to solve the MILP.

Reference [12] introduced a Mixed Integer Linear Programming to generate optimal *Strong Pure* patterns. A pattern is strong if the set of observations covered by the pattern is not a subset of that covered by any other patterns. A pattern is pure if it does not cover any observation from the opposite class. An observation will be considered *Covered* only if all the attributes forming the pattern are reflected in the observation. Since the optimality of a strong pattern is measured with respect to its coverage, the MILP's objective is to generate a pattern with the maximum number of covered observations. By solving the MILP model, a generated pattern doesn't necessarily cover all the observations of its class. So, the model is reconstructed to generate more patterns, up to a point that all the observations are covered by at least one of the generated patterns. Although each generated pattern differs from previously generated ones, yet it might cover some or all of the previously covered observations. In order to avoid generating redundant patterns, all the observations that were previously covered by the generated patterns will be removed from the training set before reconstructing the model.

Reference [13] introduced two heuristic algorithms, called *Bottom-Up* and *Top-Down*, to obtain optimal *Prime* pure patterns. A pattern is prime if removal of any of its literals results in coverage of observations from the opposite class. If the restriction of obtaining pure patterns is relaxed, a pattern becomes a combination of literals covering at least a minimum number of observations of the pattern's class, and at most a maximum number of observations of the opposite class. These two parameters are defined by the user, and are called *Coverage* and *Fuzziness* parameters, respectively, where fuzziness means the maximum number of observation from the opposite class that is covered by the pattern. Obviously, a pure pattern is a better indication of a certain class than a pattern that has certain fuzziness. However, this relaxation is frequently used when the separation between classes is difficult. In order to solve the MILP, we use a hybrid greedy method, which is composed of two phases: the bottom-up phase and the top down phase. If any observation is left uncovered by the end of the first phase, the second phase is performed.

3 Failure Prognostic

In the following section, we introduce two methods to calculate the conditional survival probability of the equipment, based on the estimated survival functions using *Kaplan-Meier (KM) estimation* [14].

Assume a training set includes 5, 3 and 4 observations corresponding to three pieces of equipment of the same type. The observations of equipment one are indicated by 1-0, 1-1, 1-2, 1-3, 1-4, where the first number means equipment number 1, and the second number is the equipment age. This means that equipment one has failed sometime during the fifth period. Each observation has a health condition attribute corresponding to each observation that is not shown in this example. 1-0, 1-1, 1-2, and 1-3 are considered negative observations, and 1-4 is considered positive observation. Using MILP positive and

Table 1. List of the generated patterns and covered observations

+ Pattern	Covered Observations	- Pattern	Covered Observations
<i>P1</i>	1-3 , 1-4 , 3-3	<i>N1</i>	1-0 , 1-1 , 1-2 , 2-0 , 2-1 , 2-2 , 3-0 , 3-1 , 3-2
<i>P2</i>	2-2 , 3-3	<i>N2</i>	1-0 , 1-1 , 1-3 , 2-0 , 2-1 , 3-0 , 3-1

negative patterns are generated. For each generated positive or negative pattern, the list of its covered observations is presented in Table 1.

We associated to each pattern *P*, Pattern Conditional Survival Probabilities $SP_p(i)$ for $\forall i \in \{1,2,\dots,T\}$, which represent the pattern’s survival estimation of a piece of equipment for at least *i* periods, when the equipment’s observation is covered by the pattern. *T* is the maximum available survival period within training set. *T* is 5 in the example.

KM estimation of pattern conditional survival probability is defined as the proportion of the number of observations covered by pattern *P* whose corresponding pieces of equipment survived at least *i* periods after being covered by the pattern, to the total number of observations covered by pattern *P*.

$$SP_p(i) = \#(P \cap S; \tau > \tau_0 + i\Delta) / \#(P \cap S; \tau > \tau_0) \tag{1}$$

Where *S* is the set of observations in the training set, and $P \cap S$ represents the subset of observations in the training set *S* that are covered by the pattern *P*. Function $\#(N)$ counts the number of members of a set *N*. τ is the failure time of the corresponding equipment, and τ_0 is the current age of the corresponding equipment at observation moment at which it is covered by pattern *P*. Δ is the observation period length.

Table 2 shows KM estimation of the conditional survival probability of the generated patterns, based on their corresponding covered observations in the Table 1. For example, $SP_{P1}(1)$ is equal to 0.333 because *P1* covers observations 1-3, 1-4, 3-3 in the Table 1, but only observation 1-3 has corresponding equipment (i.e. equipment 1) that survives more than one period after being covered by *P1*. Both corresponding equipment of observation 1-4 and 3-3 have failed during next period, as soon as they were covered by *P1*.

Table 2. KM estimation of patterns conditional survival probability for the problem

<i>i</i> Δ	1	2	3	4
$SP_{P1}(i)$	0.333	0	0	0
$SP_{P2}(i)$	0	0	0	0
$SP_{N1}(i)$	0.889	0.667	0.333	0.111
$SP_{N2}(i)$	1	0.714	0.428	0.143

We also defined the Baseline Conditional Survival Probability to indicate time-based survival function, regardless of the equipment’s condition. KM estimation of the baseline conditional survival probability is calculated as the proportion of the number of pieces of equipment that survived at least *i* periods, to the number of all the pieces of equipment in training set.

$$SP_b(i) = \#(E; \tau > i\Delta) / \#(E) \tag{2}$$

where *E* is the set of all pieces of equipment in the training set. Table 3 shows the KM estimation of baseline conditional survival probabilities. $SP_b(3)$ equal to 0.667 means that two out of three pieces of equipment in training set have survived more than 3 periods.

Table 3. KM estimation of baseline conditional survival probability for the problem

$i\Delta$	1	2	3	4
$SP_b(i)$	1	1	0.667	0.333

By combining both, $SP_p(i)$ and $SP_b(i)$ for each observation at period i , we introduce two methods to calculate the conditional survival probability of the equipment from which a new observation is collected. The first method favors the Pattern Conditional Survival Probability (SP_p), while takes into account the Conditional Survival Probabilities that were calculated for the equipment previously (SP_{former}), less weightily. It also contains the Baseline Conditional Survival Probability (SP_b). Defining n as the number of patterns that cover an observation, the conditional survival probability of the equipment for period i is calculated as follows:

$$SP_{obs}(i) = \begin{cases} \left[\sum_{p=1}^n SP_p(i) + SP_b(i) \right] / n + 1 & , \text{if } t = 0 \\ \left[\sum_{p=1}^n SP_p(i) + SP_{former}(i+1) \right] / n + 1 & , \text{if } t > 0 \end{cases} \quad (3)$$

As an example, we assume four consecutive testing observation record for a piece of equipment covered by the patterns shown in Table 4. As introduced in eq. (3), the conditional survival probabilities of the equipment at different observation moments are shown in the Table 4. $SP_{obs}(2)$ for 1-0 is equal to 0.794 because the observation 1-0 is covered by patterns $N1$ and $N2$ for which $SP_{N1}(2)$ and $SP_{N2}(2)$ are equal to 0.667 and 0.714 respectively (see Table 2), and $SP_b(2)$ is equal to 1 (see Table 3). As a result $SP_{obs}(2)$ for 1-0 is equal to $(0.667+0.714+1) / 3 = 0.794$. $SP_{former}(1)$ for 1-1 is also equal to 0.794 because its corresponding equipment was formerly predicted to survive for at least 2 periods with the probability of 0.794 ($SP_{obs}(2)$ for 1-0 is 0.794).

Table 4. Conditional survival probabilities of hypothetical equipment at different observation moments—the 1st calculation method

Obs	Covering Patterns	$\Sigma SP_p(i)$				$SP_b(i)$				$SP_{former}(i)$				$SP_{obs}(i)$			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1-0	-N1, -N2	1.89	1.38	0.76	0.25	1	1	0.67	0.33	-	-	-	-	0.96	0.79	0.48	0.19
1-1	-N1, -N2	1.89	1.38	0.76	0.25	-	-	-	-	0.79	0.48	0.19	0	0.89	0.62	0.32	0.08
1-2	-N1	0.89	0.67	0.33	0.11	-	-	-	-	0.62	0.32	0.08	0	0.76	0.5	0.21	0.06
1-3	+P1, +P2	0.33	0	0	0	-	-	-	-	0.5	0.21	0.06	0	0.28	0.07	0.02	0

The second method also favours the latest observation to older ones, but, it attributes equal weight for Pattern and Baseline Conditional Survival Probabilities. The conditional survival probability of the equipment is calculated as follows:

$$SP_{obs}(i) = \begin{cases} \left[\left[\sum_{p=1}^n SP_p(i) \right] / n + SP_b(i) \right] / 2 & , \text{if } t = 0 \\ \left[\left[\sum_{p=1}^n SP_p(i) + SP_{former}(i+1) \right] / n + 1 + SP_b(i) \right] / 2 & , \text{if } t > 0 \end{cases} \quad (4)$$

The conditional survival probabilities of the same testing equipment at different observation moments using the latter method are shown in Table 5. $SP_b(1)$ for 1-3 is

equal to 0.5 because one out of two pieces of equipment that have survived more than 3 periods, has survived more than 4 periods.

Table 5. Conditional Survival probabilities of hypothetical equipment at different observation moments—the 2nd calculation method

Obs	Covering Patterns	$\Sigma SP_{A(t)}$				$SP_{A(t)}$				$SP_{Borne(t)}$				$SP_{end(t)}$			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1-0	-N1, -N2	1.89	1.38	0.76	0.25	1	1	0.67	0.33	-	-	-	-	0.97	0.85	0.53	0.23
1-1	-N1, -N2	1.89	1.38	0.76	0.25	1	0.67	0.33	0	0.85	0.53	0.23	0	0.96	0.65	0.33	0.04
1-2	-N1	0.89	0.67	0.33	0.11	0.67	0.33	0	0	0.65	0.33	0.04	0	0.72	0.42	0.09	0.03
1-3	+P1, +P2	0.33	0	0	0	0.5	0	0	0	0.42	0.09	0.03	0	0.38	0.02	0	0

4 Experiments

We applied the LAD methodology on Prognostics and Health Management Challenge dataset, a condition monitoring dataset provided by NASA Ames Prognostics Data Repository, which we randomly divided to one testing set and 10 training sets. The performances of the equipartitioning method and the hybrid greedy method are examined for 4 and 12 different settings, respectively. We also used the PHM [1] to calculate a survival function in order to compare the different LAD settings’ performances with that of the PHM.

For each observation in the testing set, we calculate the set of conditional probability of survival for the future predictable periods. However, this set is not meaningfully comparable to its matched pair set provided by other experiments. Therefore, we transform the information of the set into a single comparable value, the *Mean Residual Life (MRL)*, so that we can compare performance of different experiments. *MRL* represents the expected value of equipment residual life. The *MRL* is formulated as following [15]:

$$MRL = \sum_{i=1}^{\infty} i\Delta \times P(\tau > \tau_0 + i\Delta | \tau > \tau_0) \tag{5}$$

Where $P(\tau > \tau_0 + i\Delta | \tau > \tau_0)$ shows the probability of survival for at least i periods, knowing that the equipment has not failed until τ_0 . This conditional probability is the conditional survival probability $SP_{obs}(i)$, introduced in this work. So the *MRL* is formulated in terms of $SP_{obs}(i)$ as follows:

$$MRL = \sum_{i=1}^{\infty} i\Delta \times SP_{obs}(i) \tag{6}$$

The measurement compared between different experiments is the absolute value of differences between the *MRL* and the actual *Residual Life (RL)*. *Friedman matched-pair test* [16] is used to compare the performance of all Hybrid Greedy methods. The best Hybrid Greedy method (*coverage*>10% and *fuzziness*=0), then was compared with the MILP and PHM methods. Comparison shows that the PHM slightly outperforms both LAD methods, while the Hybrid Greedy and MILP methods are not statistically different (see Table 6). However, LAD methods profit from interpretable patterns, and no statistical assumptions.

Table 6. (MRL - RL): Hybrid Greedy vs. MILP vs. PHM

	Hybrid Greedy	MILP	PHM
Sensitive Discriminating	3.751	3.811	3.507 ⁺
Equipartitioning 20	3.748	3.867	3.51 ⁺
Equipartitioning 30	3.731	3.826	3.509 ⁺
Equipartitioning 40	3.723	3.801	3.51 ⁺
Equipartitioning 50	3.728	3.801	3.51 ⁺

⁺ Minimum mean value

Figure 1 shows the difference between the MRL and the actual RL of one of the 15 pieces of equipment, in testing set, using the Hybrid Greedy, the MILP, and the PHM methods. It shows that the LAD methods underestimate the MRL at the early observation moments. As time passes, the estimations get closer to the actual RL, and they correctly estimate the MRL around the mid-age observation moments. Later, when getting closer to the actual failure moment, they overestimate the MRL. The PHM method overestimates the MRL by about one period or more. Overestimation grows as the equipment approaches its failure.

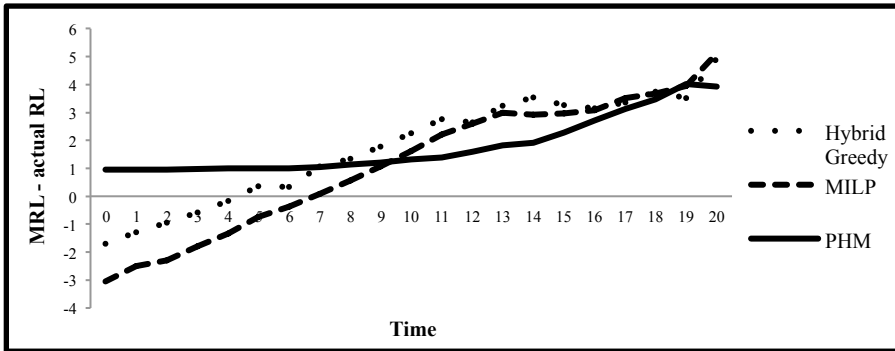


Fig. 1. Difference between MRL and actual RL using the best Hybrid Greedy, MILP, and PHM methods

5 Conclusions

In this research, we extended the LAD methodology to predict equipment’s conditional probability of survival. Prognostics using LAD methodology were compared with that using PHM. Comparison with respect to the accuracy of estimated MRL showed that: A conditional survival probability calculation method that equally favors the Baseline and Pattern conditional Survival Probability, statistically outperformed the one that favours the Pattern conditional Survival Probability. The Hybrid Greedy method with the parameters *coverage*>10% and *fuzziness*=0, statistically outperformed the Hybrid Greedy methods with other parameter settings. The PHM method statistically slightly outperformed both LAD methods, while the LAD methods were not significantly different.

Our results also showed that the PHM method had an optimistic outlook about the equipment’s survival. The LAD methods’ outlooks change gradually from pessimistic to optimistic, as the equipment health deteriorates over its lifetime.

However, the LAD has the added advantage of generating meaningful patterns that can be used as indicators of probable failure. It also has the advantage of being free of any statistical assumptions such as the correlation between the attributes, or independence of variables. It has no statistical parameters that need to be estimated, and it has also been proven [3, 18] that LAD is a robust method that reacts very well to incorrect or missing data, which is the case of most real databases. This last property has not been tested with other methods.

References

1. Jardine, A.K.S., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20(7), 1483–1510 (2006)
2. Crama, Y., Hammer, P.L., Ibaraki, T.: Cause-effect relationships and partially defined boolean functions. *Annals of Operations Research*, 16(1) 16(1), 299–326 (1988)
3. Bennane, A., Yacout, S.: LAD-CBM; new data processing tool for diagnosis and prognosis in condition-based maintenance. *Journal of Intelligent Manufacturing* 23(2), 265–275 (2012)
4. Mortada, M., Carroll, T., Yacout, S., Lakis, A.: Rogue components: Their effect and control using logical analysis of data. *Journal of Intelligent Manufacturing* 23(2), 289–302 (2012)
5. Mortada, M., Yacout, S., Lakis, A.: Diagnosis of rotor bearings using logical analysis of data. *Journal of Quality in Maintenance Engineering* 17(4), 371–397 (2011)
6. Yacout, S.: Fault detection and diagnosis for condition based maintenance using the logical analysis of data. In: 2010 40th International Conference on Presented at Computers and Industrial Engineering (CIE) (2010)
7. Boros, E., Hammer, P.L., Ibaraki, T., Kogan, A.: Logical analysis of numerical data. *Mathematical Programming* 79(1), 163–190 (1997)
8. Kotsiantis, S., Kanellopoulos, D.: Discretization techniques: a recent survey. *GESTS Int. Transact. Comput. Sci. Eng.* 32, 47–58 (2006)
9. Liu, H., Hussain, F., Tan, C.L., Dash, M.: Discretization: An enabling technique. *Data Mining and Knowledge Discovery* 6(4), 393–423 (2002)
10. Alexe, S., Hammer, P.L.: Accelerated algorithm for pattern detection in logical analysis of data. *Discrete Applied Mathematics* 154(7), 1050–1063 (2006)
11. Hammer, P.L., Kogan, A., Simeone, B., Szedmak, S.: Pareto-optimal patterns in logical analysis of data. *Discrete Applied Mathematics* 144(1-2), 79–102 (2004)
12. Ryoo, H.S.: MILP approach to pattern generation in logical analysis of data. *Discrete Applied Mathematics* 157(4), 749–761 (2009)
13. Boros, E., Hammer, P.L., Ibaraki, T., Kogan, A., Mayoraz, E., Muchnik, I.: An implementation of logical analysis of data. *IEEE Transactions on Knowledge and Data Engineering* 12(2), 292–306 (2000)
14. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53(282), 457–481 (1958)
15. Banjevic, D., Jardine, A.K.S.: Remaining useful life in condition based maintenance: Is it useful? In: *Proceedings of MIMAR 2007*, the 6th IMA International Conference, pp. 7–12 (2007)
16. Friedman, M.: A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics* 11(1), 86–92 (1940)
17. Boros, E., Crama, Y., Hammer, P., Ibaraki, T., Kogan, A., Makino, K.: Logical analysis of Data: Classification with justification. *Rutcor Research Report, RRR 5-2009* (2009)