A Review on Existing Feature Selection Metrics for Prognostics

M. Hwang[1], Y. Ban[2], and H. Oh[3*], S. Kim[4] and M. Mo[5]

[1,2,3]School of Mechanical Engineering, Gwangju Institute of Science and Technology, Gwangju, 61005, Korea
[1,4,5] Global Institute of Technology, KEPCO Plant Service and Engineering Co., LTD, Naju, 58217, Korea

[*]Corresponding author: hsoh@gist.ac.kr

1. Introduction

Feature selection is one of key steps for diagnostics and prognostics of mission-critical engineered systems. Relevant features among candidate features can be different between diagnostics and prognostics. A set of features for diagnostics should be selected to detect and identify failure modes, whereas another set of features for prognostics should be selected to capture degradation behaviors so that remaining life and/or future conditions can be predicted [1]. It is obvious that the selected feature subset for diagnostics may not always perform as intended for prognostics, vice versa.

Numerous studies presented feature selection metrics for diagnostics. However, a limited amount of studies was reported in feature selection metrics for prognostics. There is no agreement on desired characteristics for feature selection metrics. Multiple metrics for identical characteristic can be found. Thus, this study presents a critical review on the existing feature selection metrics.

The remaining section of this paper is organized as follows. In Section 2, existing feature selection metrics for prognostics are reviewed. In Section 3, counter examples of the existing feature selection metrics are presented. Section 4 concludes this paper with future works.

2. Review on Existing Metrics for Prognostics

Three desired characteristics of feature selection metrics were suggested by Coble et al. [2]: (1) monotonicity, (2) trendability, and (3) prognosability. The monotonicity is used to capture the underlying positive or negative trend for individual features since engineered systems undergo irreversible process during deterioration. Mathematically, the monotonicity (M) is defined as the absolute difference between feature fractions of positive and negative derivatives.

$$M(F) = \left| \frac{\# d/df_i > 0}{N-1} - \frac{\# d/df_i < 0}{N-1} \right| \quad (1)$$

where $F=\{f_i\}_{i=1:N}$ is the feature sequence with the feature value, f_i, at i^{th} observation and; N is the number of observations; $\#d/df_i>0$ and $\#d/df_i<0$ represents the number of positive differences and negative differences, respectively.

The trendability (T) is characterized by comparing the fraction of positive first derivative and second derivative in each feature. It indicates the degree to which parameters of a population of systems have the same underlying shape. The indication of trendabillity for individual features is given in Eq. (2). The trendability of a population of sample units is formulated as Eq. (3).

$$T_{i\text{-}Coble} = \frac{\# d/df_i > 0}{N-1} + \frac{\# d^2/df_i^2 > 0}{N-2} \quad (2)$$

$$T(F)_{Coble} = 1\text{-}std(T_{i\text{-}Coble}) \quad (3)$$

where $\#d^2/d^2f_i>0$ denotes the number of positive second derivative; and $std(T_{i\text{-}Coble})$ is standard deviation of $T_{i\text{-}Coble}$.

Prognosability (P) is a measure of deviation of the final failure values for each path divided by the mean range of the path. Lower value represents the wide spread in the final failure values.

$$P = \exp\left(-\frac{\sigma(f_E)}{\left(|f_E - f_0|\right)/M} \right) \quad (4)$$

where M is total number of feature sets under same degradation condition; f_E is final feature values at the end of life; f_0 is initial feature value; $\sigma(f_E)$ is standard deviations for vectors of f_E; and $(|f_E-f_0|)/M$ represents the mean of degradation path.

Javed et al. [1] presented another metric for the trendability as a correlation between features and time.

$$T(F)_{Javed} = \frac{N\left(\sum f_i t_i\right) - \left(\sum f_i\right)\left(\sum t_i\right)}{\sqrt{\left[N\sum f_i^2 - \left(\sum f_i\right)^2\right]\left[N\sum t_i^2 - \left(\sum t_i\right)^2\right]}} \quad (5)$$

where $\{t_i\}_{i=1:N}$ is the i^{th} observation value. Other metrics for trendability such as spearman coefficient and rank mutual information were described in [3].

Another characteristic, robustness, was discussed in [4]. The characteristic is used to select features with a smooth degradation trend and robustness to measurement noise.

$$R(F)=\frac{1}{N}\sum \exp\left(-\frac{f_i-\bar{f}}{f_i}\right) \tag{6}$$

where $\bar{f}$ is smoothed feature value.

From above observations, it was found that previous studies presented different characteristics and corresponding metrics for prognostics. There is no universally-accepted characteristics and metrics for prognostic feature selection.

3. Counter Examples for Existing Metrics for Prognostics

To predict the remaining useful life of engineered systems, prognostic models were proposed with the concept of degradation signals [5]. A degradation signal can be a linear combination of basis functions with a random noise term.

$$L(t)=\theta' + \beta t_i + \varepsilon(t_i) \tag{7}$$

where θ' represents the constant that represents the amplitude of the degradation signal; β is the constant that describe the weight of the basis functions; $\varepsilon(t_i)$ is the white Gaussian noise.

The degradation signal was generated as illustrated in Fig. 1. The fluctuation of the original degradation signal decreased as the window size of a moving averaging method is increased.

The metrics for monotonicity and trendability were calculated with the three degradation signals as shown in Table 1. Depending on the window sizes, different metric values were computed. For example, M(F) of the original degradation signal was 0.06, whereas M(F) of the moving-averaged degradation signal with the window size of 10% was 0.74, whose change is significant. The metrics can lead t o a large amount of deviation depending on how to smooth the original signal.

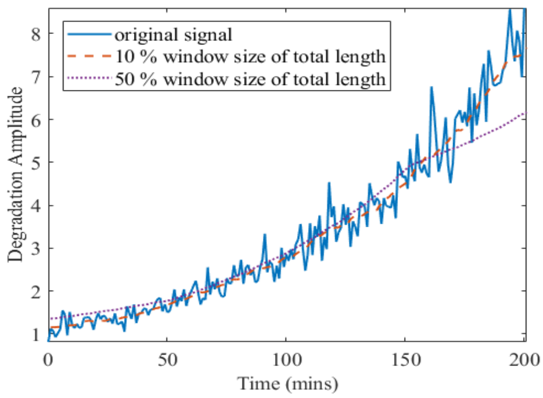

Fig.1 Original degradation signal and moving-averaged ones with different window sizes

Table 1 Effect of window sizes on existing metrics

Window Size	M(F)	$T_{i\text{-Coble}}$	$T(F)_{Javed}$
None	0.06	1.04	0.94
10%	0.74	1.36	0.96
50%	1.00	1.50	0.98

4. Conclusions and Future Work

This paper reviewed existing feature selection metrics for prognostics and presented the limitations of the existing metrics. The monotonicity and trendability metrics can change significantly depending how to smoothen the degradation signals since the metrics contains the first and second derivative terms. The metric provided different results depending on the window size. This can be problematic in feature selection for prognostics.

A feature selection metric for prognostics should not be affected by the random error term. To this end, the future work is to develop a reliable feature selection metric for prognostics, i.e., metric that can capture the degradation trend regardless of measurement noise.

Acknowledgment

This research is part of the projects that are financially supported by the KEPCO KPS (Project No.: C101800708).

References

[1] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, Enabling health monitoring approach based on vibration data for accurate prognostics, *IEEE Transactions on Industrial Electronics*, 62 (1) (2015) 647-656.

[2] J. Coble, J. W. Hines, Identifying optimal prognostic parameters from data: A genetic algorithms approach, *in Proc. Annu. Conf. Prognost. Health Manage. Soc.*, (2009) 1-11.

[3] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: a systematic review from data acquisition to RUL prediction, *Mechanical Systems and Signal Processing*, 104, (2018) 799–834.

[4] V. Atamuradov, K. Medjaher, F. Camci, P. Dersin, and N. Zerhouni, Railway Point Machine Prognostics Based on Feature Fusion and Health State Assessment, *IEEE Transactions on Instrumentation and Measurement*, (2018) 1-14.

[5] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, Residual-life distributions from component degradation signals: A Bayesian approach, *IIE Transactions*, 37, (6), (2005), 543-557.