Development and Application of Advanced Data-Driven Methods for Prognostics and Health Management of Industrial Components under Scarce Degradation Information

Francesco Cannarile

Politecnico di Milano, Department of Energy, Via la Masa 34/3, 20156, Milano, Italy
francesco.cannarile@polimi.it

ABSTRACT

My thesis project aims at developing confident and robust data-driven methods for PHM under scarce information and few data on the degradation process. Four techniques are expected to be developed in this thesis project: 1) A Fuzzy Expectation-Maximization (FEM) algorithm to estimate unknown parameters of a multi-state degradation model when the only information available is qualitative statements provided by an expert on transition times from a degradation state to another; 2) Development of an unsupervised clustering based-method for assessing the degradation state of industrial components when the only information available is *i*) measurements of a physical quantity indirectly related to the components degradation and ii) the beginning and the end of operation of the components, whereas no information is available on the component degradation state during its operation life; 3) A data-driven direct RUL prediction algorithm capable to exploit not only run-to-failure degradation trajectories data but also uncomplete, right-censored degradation trajectories; 4) A shrinking ensemble algorithm of prognostic methods based on coalition game theory. Different real case studies will be considered such as Piping Systems (PPs) in Nuclear Power Plants (NPPs), cutting tools in the packaging industry and turbofan engines in the aerospace industry to verify the proposed methods.

1. PROBLEM STATEMENT

In human health care, a medical analysis is made, based on the measurements of some parameters related to health conditions; the examination of the collected measurements aims at detecting anomalies, diagnosing illnesses and predicting their evolution. By analogy, Prognostics and Health Management (PHM) aims at to detect, diagnose and predict the future evolution of the state of industrial components (Zio, 2016). The implementation of PHM leads to the following benefits: 1) improving system safety, availability and reliability which increases the probability of mission success; 2) reducing processing and operation time, manpower, and costs. The knowledge of the state of industrial components and the prediction of its future evolution are at the basis of condition-based maintenance strategies: according to these strategies, maintenance actions are carried out when a measurable component condition shows the need for corrective repair or preventive replacement. From the point of view of production performance (Zio, 2016), by identifying the problems in the components at their early stages of development, it is possible to allow the equipment to run as long as it is healthy and to opportunely schedule the maintenance interventions for the most convenient and inexpensive times. The dynamic scheduling of condition-based maintenance represents a challenging task, which requires the prediction of the evolution of the monitored variables representing the component condition. Upon detection of failure precursors, prognostics becomes a fundamental task; this entails predicting the reliability or the probability of failure of the equipment at future times, and the Remaining Useful Life (RUL), i.e. the amount of time the equipment will continue to perform its function according to design specifications. This prediction/forecasting/extrapolation process needs to account for the current state assessment and the expected future operational conditions.

Yet, in many practical situations, developing the model of component degradation and performing experimental tests are very difficult or even impossible. These situations are typically encountered for the safety-critical and high-value components (nuclear, aerospace, military, oil and gas fields etc.), which are characterized by very high reliability, unique or new designed material composition, and for which performing run-to-fail test is too expensive or not feasible. This complicates the fault diagnostics and prognostics tasks, since few information and data are available and calls for the development of ad-hoc methods.

The target of this thesis work is to develop confident and robust data-driven methods for PHM under scarce information and few data on the degradation process.

2. EXPECTED CONTRIBUTIONS

The outcome of this research work is the development of data-driven methods to enhance fault diagnostics and prognostics when few information is available. In detail, we consider the following realistic situations where only the following information and data are available: c1) an expert can provide a qualitative judgment on the transition times among different degradation states (e.g., from healthy to low degraded state); c2) measurements of a physical quantity indirectly related to the component degradation; c3) few run-to-failure degradation trajectories and many right-censored degradation trajectories. Cases c1), c2) and c3) are discussed in Subsections 2.1, 2.2, and 2.3, respectively. Finally, in Subsection 2.4 we consider also the possibility of resorting to ensemble learning methods to enhance generalization of prognostic algorithms when few data are available.

2.1 Multi-State Degradation Modeling from Imprecise Maintenance Outcome

Multi-State (MS) reliability models are used in practice to describe the evolution of degradation in industrial components and systems. To estimate the MS model parameters, we propose a method based on the Fuzzy Expectation-Maximization (FEM) algorithm, which integrates the evidence of the field inspection outcomes with information taken from the maintenance operators about the transition times from one state to another. Possibility distributions are used to describe the imprecision in the expert statements. A procedure for estimating the Remaining Useful Life (RUL) based on the MS model and conditional on such imprecise evidence will be also developed.

2.2 Development of an unsupervised clustering based-method for assessing the degradation state of industrial component

We propose a method for on-line assessing the degradation state of components for which we have only available measurements of a physical quantity indirectly related to the equipment degradation and we know the beginning and end of operation life of the component, but we do not have any information on the component degradation state during its life. The method to identify the component degradation state is based on the following steps: S1) extraction of statistical and frequency-based features from the raw data; S2) trend extraction using Empirical Mode Decomposition (EMD); S3) selection among the set of extracted features of an optimal subset of Health Indicators (HIs) for the identification of the component degradation using monotonicity and trendability index (Coble & Hines, 2009); S4) unsupervised clustering of HI data in order to identify the component degradation state.

2.3 Evidential Fuzzy Similarity: a novel method to predict the Remaining Useful Life of industrial components based on Fuzzy Similarity and Evidence Theory

Data-driven direct methods assume the availability of a set of training samples data from run-to-failure degradation trajectories (‘faulty trajectories’) which are fed to the learning algorithm. Yet, in many industrial applications, components can be replaced before they fail, and corresponding right-censored degradation trajectories (‘healthy trajectories’) are not used in the training stage of these learning algorithms. To tackle this issue, we propose a novel method called Evidential Fuzzy Similarity (EvFS) which combines Fuzzy Similarity (FS) with Dempster-Shafer Evidence Theory. In this framework, when predicting the RUL of a test trajectory, both healthy and faulty trajectories in the training-set can be regarded as agents whose state of knowledge on the actual RUL of the test equipment, is encoded in a Basic Belief Assignment (BBA). The construction of these BBAs is based on the fuzzy similarity between the test and training trajectories. Then, Dempster’s rule of combination is applied to combine the BBAs of all training trajectories. The resulting combined BBA is, then, used to compute 1) its pignistic expectation which represents our RUL point prediction and 2) its Belief Cumulative Distribution Function (BCDF) which allows quantifying the uncertainty in the RUL prediction.

2.4 Coalition game based ensemble shrinking for enhancing RUL prediction

When few degradation data are available, prognostic algorithms may overfit. A possible solution to this problem is ensemble learning. In fact, the combination of different prognostic algorithms usually generalizes better than a single prognostic algorithm. Yet, ensemble methods must be used with caution, since they can negatively compromise the predictive performance of the ensemble members. In fact, most approaches to develop ensembles tend to generate an unnecessarily large number of predictive algorithms in order to minimize the training error. This can inevitably result in overfitting the training set, which in turn result in a reduction in the generalization capability of the ensemble. To cope this issue, ensemble shrinking methods have been introduced. Ensemble shrinking aims at selecting a subset of prognostic algorithms that optimizes a criterion suitable to obtain generalization. In this framework, we propose a method based on coalition game theory which is based on the following steps: 1) definition of a measure of diversity among different prognostic algorithms; 2) evaluation of the diversity contribution of each ensemble member using a power index (e.g., Shapley value, Banzhaf index, etc.); 3) selection of the shrinking ensemble as the minimal winning coalition in a weighted voting game constituted of the best ranked members.

3. RESEARCH PLAN

This PhD project is a four years PhD program in "Mathematical Models and Methods in Engineering" supported by the consulting company Aramis S.r.l., which is started on November 2014, and is expected to finish before March 2019. Methodologies discussed in Subsections 2.1 and 2.2 have been developed and verified on data concerning degradation of Piping Systems in Nuclear Power Plants (PSNPPs) and cutting tools in the packaging industry, respectively. The development of the methodologies discussed in Subsections 2.3 and 2.4 is currently under investigation.

3.1 Work Performed

The FEM algorithm for estimating the parameters of a multi-state degradation model from imprecise maintenance outcomes discussed in Subsection 2.1 has been successfully applied to a case study concerning the degradation of PSNPPs whose degradation process has been modelled as a four-state semi-Markov process with Weibull sojourn-time. We have assumed that the only information on the degradation state of PSNPPs is provided by experts who can give statements such as "The pipe transition from detectable flaw state to detectable leak state occurred between 1998 and 2000, March 1999 being the expected month for this transition". This kind of information has been properly encoded through possibility distributions and using FEM algorithm Weibull sojourn-time parameters have been estimated. The obtained results show that the combination of field data with expert knowledge can allow reducing the uncertainty in degradation estimation and RUL prediction.

The methodology discussed in Subsection 2.2 has been successfully verified on a real case study concerning the degradation state of cutting tools in the packaging industry. Two HIs have been selected and Fuzzy C-Mean (FCM) algorithm has been used to infer the unknown number of degradation states and label HI data with the corresponding degradation state.

3.2 Remaining Work

We are working on the development of the EvFS algorithm discussed in Subsection 3.3. By now the EvFS algorithm has been applied to real data from the packaging industry and compared with FS. Table 1 reports the comparison of the two methods with respect to some prognostic metrics (relative error R, precision P and α - λ metric (Saxena et al, 2010)).

		FS	EvFS
Relative Error *R*		0.48±0.02	0.40±0.03
Precision *P*		0.96±0.02	0.97±0.01
α=0.3	λ=0.25	0.80±0.04	0.80±0.01
	λ=0.50	0.80±0.04	0.90±0.03
	λ=0.75	0.50±0.05	0.70±0.05
Average α - λ		0.66±0.05	0.80±0.04

Table 1: Comparison of FS and EvFS

Table 1 shows that EvFS outperforms FS, EvFS provides an average relative error on the RUL prediction which is 16.66% lower than that of the FS, this is due to the capability of EvFS of exploiting all the available information (i.e., including also that provided by the healthy trajectories), however, to show the real strength of this method, it requires to be tested on other case studies. Finally, we are also investigating ensemble shrinking methods as discussed in Subsection 2.4. At early stage, we are working on the definition of "diversity" among different prognostic algorithms, then, we will focus on modeling the ensemble member selection process as a weighted coalition game and the proposed methodology will be applied to different case studies (e.g., the NASA turbofan dataset (Saxena et al., 2010)).

4. CONCLUSION

Performing PHM when scarce information on the degradation process of the component and few data is available is challenging and requires the development of ad-hoc methods. The contribution of this thesis is to solve this problem, developing different data-driven methods corresponding to different source of information.

REFERENCE

Coble, J., Wesley Hines, J., (2009). *Identifying optimal prognostic parameters from data: A genetic algorithms approach*. Annual Conference of the Prognostics and Health Management Society, PHM 2009.

Saxena, A., Celaya, J., Saha, B., Saha, S., Goebel, K., (2010). *Metrics for offline evaluation of prognostic performance*. International Journal of Prognostics and Health Management 1 (1).

Zio, E. (2016). *Some challenges and opportunities in reliability engineering*. IEEE Transactions on Reliability, 65 (4), pp. 1769-1782.