A Similarity-based Prognostics Approach for Remaining Useful Life Prediction

O. F. Eker[1], F. Camci[2], and I. K. Jennions[1]

[1] *IVHM Centre, Cranfield University, UK*

o.eker@cranfield.ac.uk
i.jennions@cranfield.ac.uk
[2] *Industrial Engineering, Antalya International University, Turkey*
fatih.camci@antalya.edu.tr

ABSTRACT

Physics-based and data-driven models are the two major prognostic approaches in the literature with their own advantages and disadvantages. This paper presents a similarity-based data-driven prognostic methodology and efficiency analysis study on remaining useful life estimation results. A similarity-based prognostic model is modified to employ the most similar training samples for RUL estimations on each time instance. The presented model is tested on; Virkler's fatigue crack growth dataset, a drilling process degradation dataset, and a sliding chair degradation of a turnout system dataset. Prediction performances are compared utilizing an evaluation metric. Efficiency analysis of optimization results show that the modified similarity-based model performs better than the original definition.

1. INTRODUCTION

Prognostics is an essential part of condition-based maintenance, described as forecasting the remaining useful life of a system. There are two major prognostics approaches in the literature 1. Physics-based 2. Data-driven models. They both have their own advantages and disadvantages. Data-driven models employ routinely collected condition monitoring data and/or historical event data instead of building a mathematical model based on system physics or human expertise. They attempt to track the degradation of an asset using forecasting or projection techniques (e.g. regression, exponential smoothing, and neural networks), also rely on the past patterns of deterioration to forecast the future degradation. Since data-driven prognostics have no elaborate information related to asset or system, it has been considered as a black-box operation (Zhang et al., 2009). A detailed literature review on data-driven prognostics was conducted by Si et al.,

(2011) . Artificial Neural Networks (ANN) (Gebraeel and Lawley, 2008), Hidden Markov Models (HMM) and derivations (Camci and Chinnam, 2010), regression models (Guclu et al., 2010), Bayesian & Gaussian Processes (Saha et al., 2010) are employed in order to estimate the remaining useful of a component or system. Similarity-based prognostic approaches can also be categorized in data-driven prognostics. Details of the similarity-based prognostic models are discussed in section 2.4.

Physics-Based Models typically involve describing the physics of the equipment and failure mechanism. Mathematical models are usually employed which is directly tied to health degradation. In order to provide knowledge rich prognostics output; physics-based models attempt to combine defect growth formulas, system specific mechanistic knowledge and condition monitoring data. They assume that an accurate mathematical model for component degradation can be constructed from the first principles. Several examples of degradation modelling and physics-based prognostics, specific to the component or system, are found in the literature (Kacprzynski et al., 2002; Byington et al., 2004; Qiu et al., 2002).

This paper presents a data-driven prognostic methodology. Contribution of the paper is to modify a similarity-based prognostic approach which performs better prognostic results compared to its original definition. Comparison and the efficiency of the remaining useful life estimation results are discussed in the paper. The rest of the paper is organized as follows. In section 2, the details of the used datasets and the methodology are given. The prognostic and optimization results are discussed in section 3. Following that are conclusion and future works.

2. METHODOLOGY

This section provides the datasets used in prognostic modelling, the similarity-based prognostic approach methodology, and the modified version of it.

2.1. Virkler's Fatigue Crack Growth Dataset

In the structural health management (SHM) field, fatigue cracks are defined as one of the primary structural damage mechanisms caused by cyclic loadings. Cracks at the structure surface grow gradually. Therefore prediction of fatigue life or fatigue crack growth in structures is necessary.

The Virkler fatigue crack growth dataset (Virkler et al., 1979) contains 68 run-to-failure specimens. Each specimen used for the experiments is a center cracked aluminum sheet of 2024-T3. Specimens had a notch of 9mm initial crack and the experiments were stopped once the crack lengths reached around 50mm. Each specimen has 164 crack length observation points. Degradation for all specimens is shown in Figure 1.
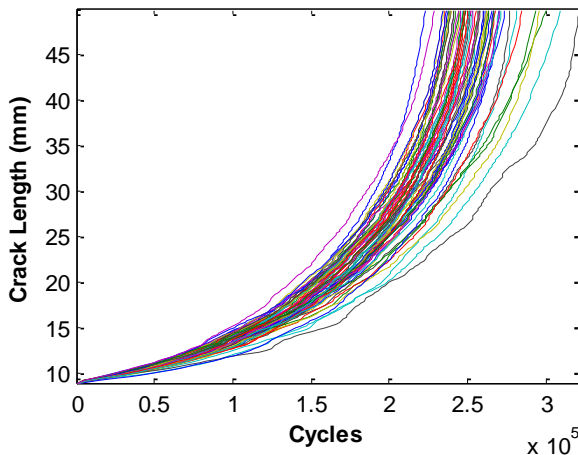

Figure 1. Crack length propagation samples under the same loading conditions

2.2. Drill-bit Dataset

Drilling processes are considered to be one of the most commonly used machining processes in industry (Lianyu Fu and Ling, 2002). For instance, up to 50% of all machining operations in the U.S. involve drilling (Furness et al., 1999). Drill bit breakage, excessive wear during the drilling process may cause fatal defects in the product. Drilled surface quality may affect the quality of the product. 60% of rejected parts are often granted to poor surface quality (Ertunc et al., 2001). Therefore, it is important to predict the failure of drill bit for obtaining good products.

The failure prediction for drill bits has been reported in (Camci and Chinnam, 2010; Baruah and Chinnam, 2005). Hidden Markov (HMM) based methods have been used for failure prediction in their methods. The dataset was collected by Chinnam et al., (2003).

Figure 2, shows the data acquisition system for drilling process. The dataset was collected from a HAAS VF-1 CNC machine. They used thin drill-bits to accelerate the aging process. The drill-bit dataset have twelve run-to-failure samples. The failure for each case is the breakage of the thin drill bit during the penetrating into work piece. Thrust-force and torque signals are collected during the actual drilling process. Concatenated thrust and torque signals, collected during the life of a drill bit, are displayed in Figure 3. In this figure, the degradation of a drill bit from brand new state to the failure state can be observed. This dataset will be used for comparison of the modified data-driven prognostic approach.
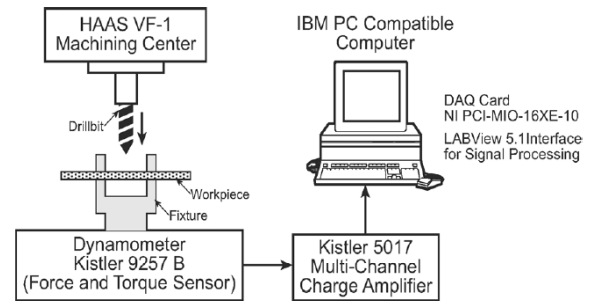

Figure 2. Experimental setup for data collection during drilling process (Camci, 2005).
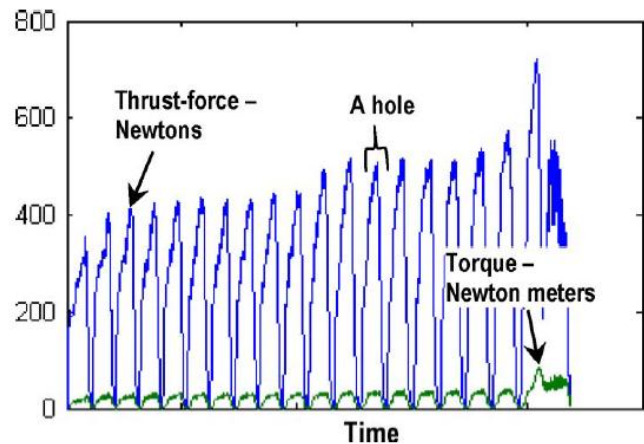

Figure 3. Trust-force and torque data from a drill bit

2.3. Turnout Dataset

Turnout systems are remote controlled electro-mechanical systems enabling trains to change their tracks as displayed in Figure 4. They are considered to be one of the most important components of the railway structure. The standard railway turnout system is a complex device with many potential failure modes. The dataset consist of five different sensors showing the degradation profile of ten different run-to-failure turnout mechanisms (Eker et al., 2011). We utilized the force sensor data among the other sensory information provided since they claimed the force sensor is capable of representing degradation process better than the rest of the sensors (Camci et al., 2014). They employed an exponential degradation model to organize the samples

collected from different discrete health states since there was no prior information about railway turnout degradation. They selected 'dry slide chair' as a failure mode for the turnout system. The dataset will be used for prognostic modelling and comparison. The dataset was collected under the project number '108M275(1001)' supported by TUBITAK (The Scientific and Technological Research Council of Turkey) in Turkey. The dataset is open to public and can be downloaded from their research group website (Camci et al., 2010).

Figure 4. Electro-mechanical turnout system

2.4. Similarity-Based Prognostics (SBP)

Zio and Di Maio, (2010) developed a novel similarity-based prognostics methodology for estimating the remaining useful life components of nuclear systems. Estimations of RUL requires evaluating the similarity between the test sample (i.e. 'q') and the training samples (i.e. 'r = 1: R') as shown in Eq. (2). This is done by calculating the point wise Euclidean distances in between 'n − long' sequences of observations. Distance score calculation in between training sample and the test sample at the ith time point shown in Eq. (1). Final RUL estimation of a test sample at a time instance (i.e. 'I') is achieved by taking the similarity weighted sum of training samples' remaining useful life values recorded on the same time instances as shown in Eq. (3).

$$d_i^r = \sqrt{\sum_{j=1}^{n} \left\| z_{I-n+j}^q - z_{i-n+j}^r \right\|^2} \qquad (1)$$

$$s_i^r = e^{-\frac{(d_i^r)^2}{\lambda}} \qquad (2)$$

$$RUL_I^q = \frac{\sum_{r=1}^{R} s_i^r rul_i^r}{\sum_{r=1}^{R} s_i^r} \qquad (3)$$

'λ' is the arbitrary parameter can be set to shape the desired interpretation of similarity whereas 'n' defines the number of latest observations involved in similarity calculations. The smaller is the λ, indicates the stronger the definition of similarity.

2.5. Modified SBP

This subsection discusses in detail the modifications made on the similarity-based prognostic model. The modifications have been made in the RUL estimation (i.e. Eq. (3)), in which the most similar K percent number of the training samples are utilized rather than using whole training set. The most similar K percent of training samples varies for every test sample and even it might vary for every time instance in a test sample. The best number K is required to be optimized by checking an error function, evaluating the prognostics efficiency. We calculated root mean squared error (RMSE) of RUL estimation results for performance evaluation. A genetic algorithm is employed to find the best number 'K' in terms of minimizing the RMSE values out of RUL estimations, shown in Eq. (4). Each K value provides its minimum RMSE value with the optimized 'n' and 'λ' parameters.

$$\min_K RMSE = f(RUL_{K_{\lambda,n}}) \qquad (4)$$

Comparison of different 'K' percentage values is discussed in the next section.

3. RESULTS

The optimization of K percentage values for their best 'n' and 'λ' parameters is shown in Table 1. By looking at the table, the lower RMSE from the RUL estimations for the Virkler dataset can be obtained when the most similar 18 numbers of training samples (38%) are utilized whereas this can be achieved in 44% and 100% for Drill-bit and Turnout datasets respectively. RMSE values of different percentage levels are shown in Figure 5. In the figure, K = 100% means all training samples are utilized in RUL calculation where it represents the original definition of the approach in the literature. Improvement in the estimation errors is anticipated as the percentages of training samples involved more in the RUL predictions. However as shown in Virkler's and drill-bit dataset plots in Figure 5 errors start to build up when K is around 40%. Drill-bit and Virkler

3

datasets show similar profile where the minimum RMSE values are obtained when 40% of the training samples are utilized in similarity weighted sum calculation of RUL values. However, the lowest error for the turnout dataset obtained at 25% and 100% levels.

Table 1. Optimization results for different datasets

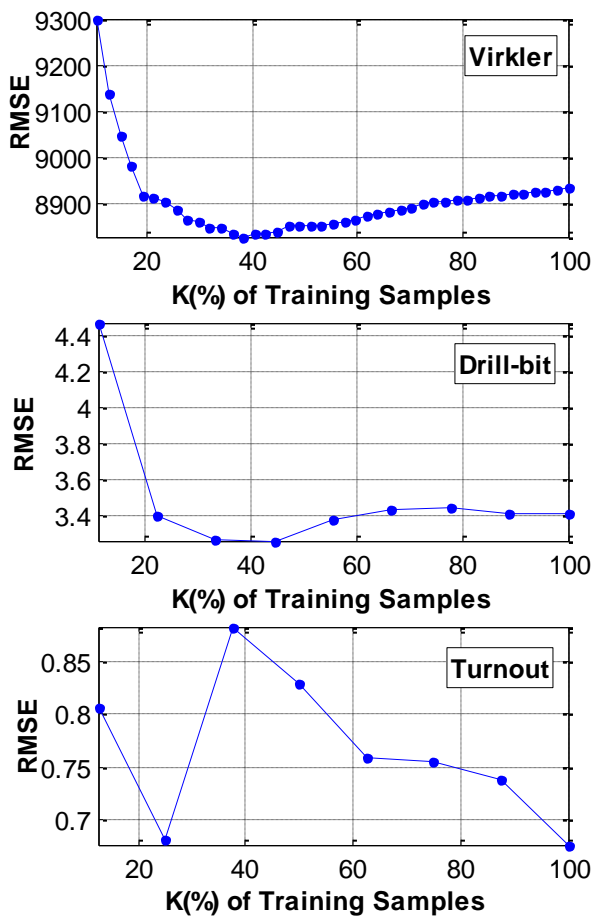| Dataset	n	λ	# of training samples	Best # of training samples	K (%)
Virkler	1.8e4	13.03	47	18	38
Drill-bit	6	0.013	9	4	44
Turnout	5	0.39	8	8	100

Figure 5. Optimization of the K percentage for different datasets

4. CONCLUSION & FUTURE WORK

This paper presents a modification on a pure data-driven similarity-based prognostic approach. The original model modified so that the most similar training samples to the test sample are involved in RUL estimation. Genetic algorithm is applied to optimize the parameters involved in similarity and RUL estimations. Results show that the modifications lessen the root mean squared error of the RUL estimations in two out of three datasets. Future studies will be on integration of a physics-based model with the modified similarity-based approach to achieve improved prediction of remaining useful life. And also the modified model prognostic performance will be compared with other prognostic approaches.

REFERENCES

Baruah, P. and Chinnam, R. B. (2005), "HMMs for diagnostics and prognostics in machining processes", *International Journal of Production Research,* vol. 43, no. 6, pp. 1275-1293.

Byington, C. S., Watson, M., Edwards, D. and Stoelting, P. (2004), "A model-based approach to prognostics and health management for flight control actuators", *IEEE Aerospace Conference Proceedings,* Vol. 6, pp. 3551.

Camci, F., Eker, O. F., Baskan, S. and Konur, S. (2014), "Comparison of sensors and methodologies for effective prognostics on railway turnout systems", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*.

Camci, F. (2005), *Process monitoring, diagnostics and prognostics using support vector machines and hidden Markov models* (PhD thesis), Graduate School of Wanye State University, Detroit, Lambert Academic Publishing.

Camci, F. and Chinnam, R. B. (2010), "Health-state estimation and prognostics in machining processes", *IEEE Transactions on Automation Science and Engineering,* vol. 7, no. 3, pp. 581-597.

Camci et al., (2010), *http://www.aiu.edu.tr/staff/fatih.camci/datasets.html*

Chinnam et al. (2003), Project Name: *'Diagnostic and Prognostic Algorithms for Condition-Based-Maintenance'.* Principal investigator: Ratna B. Chinnam, Funded by Advanced Manufacturing Technology Development (AMTD) group of Ford Motor Company for $65,000, 2003-2004.

Eker, O. F., Camci, F., Guclu, A., Yilboga, H., Sevkli, M. and Baskan, S. (2011), "A simple state-based prognostic model for railway turnout systems", *IEEE Transactions on Industrial Electronics,* vol. 58, no. 5, pp. 1718-1726.

Ertunc, H. M., Loparo, K. A., Ozdemir, E. and Ocak, H. (2001), "Real time monitoring of tool wear using multiple modeling method", *Electric Machines and*

Drives Conference, 2001. IEMDC 2001. IEEE International, pp. 687.

Furness, R. J., Tsu-Chin Tsao, Rankin, J. S.,II, Muth, M. J. and Manes, K. W. (1999), "Torque control for a form tool drilling operation", *Control Systems Technology, IEEE Transactions on,* vol. 7, no. 1, pp. 22-30.

Gebraeel, N. Z. and Lawley, M. A. (2008), "A neural network degradation model for computing and updating residual life distributions", *IEEE Transactions on Automation Science and Engineering,* vol. 5, no. 1, pp. 154-163.

Guclu, A., Yılboga, H., Eker, O. F., Camci, F. and Jennions, I. (2010), "Prognostics with Autoregressive Moving Average for Railway Turnouts", *Annual Conference of Prognostics and Health Management Society, Portland, Oregon, USA,* 10-14 October 2010, .

Kacprzynski, G. J., Roemer, M. J., Modgil, G., Palladino, A. and Maynard, K. (2002), "Enhancement of physics-of-failure prognostic models with system level features", *Aerospace Conference Proceedings, 2002. IEEE,* Vol. 6, pp. 6-2919.

Lianyu Fu and Ling, S. -. (2002), "Neural network based on-line detection of drill breakage in micro drilling process", *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on,* Vol. 4, pp. 2054.

Qiu, J., Seth, B. B., Liang, S. Y. and Zhang, C. (2002), "Damage mechanics approach for bearing lifetime prognostics", *Mechanical Systems and Signal Processing,* vol. 16, no. 5, pp. 817-829.

Saha, S., Saha, B., Saxena, A. and Goebel, K. (2010), "Distributed prognostic health management with gaussian process regression", *Aerospace Conference, 2010 IEEE,* pp. 1.

Si, X. -., Wang, W., Hu, C. -. and Zhou, D. -. (2011), "Remaining useful life estimation - A review on the statistical data driven approaches", *European Journal of Operational Research,* vol. 213, no. 1, pp. 1-14.

Virkler, D. A., Hillberry, B. M. and Goel, P. K. (1979), "The Statistical Nature of Fatigue Crack Propagation", vol. 101, no. 2, pp. 148-153.

Zhang, H., Kang, R. and Pecht, M. (2009), "A hybrid prognostics and health management approach for condition-based maintenance", *IEEM 2009 - IEEE International Conference on Industrial Engineering and Engineering Management,* pp. 1165.

Zio, E. and Di Maio, F. (2010), "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system", *Reliability Engineering & System Safety,* vol. 95, no. 1, pp. 49-57.

BIOGRAPHIES

Omer Faruk Eker is a PhD candidate in School of Applied Sciences and works as a researcher at IVHM Centre, Cranfield University, UK. He received his B.Sc. degree in Mathematics from Marmara University and M.Sc. in Computer Engineering from Fatih University, Istanbul, Turkey. He has got involved in a project funded by TUBITAK, and Turkish State Railways from 2009-2011. His research interests include failure diagnostics and prognostics, condition based maintenance, pattern recognition and data mining.

Dr. Fatih Camci works as a faculty member Industrial Engineering Department at Antalya International University in Turkey. He has worked on many research projects related to Prognostics Health Management (PHM) in USA, Turkey, and UK. After completion of his PhD at Wayne State University in 2005, he worked as senior researcher at Impact Technologies for two years. He has involved in many projects funded by US Navy and US Air Force Research Lab. He then worked as Asst. Prof. in Turkey. He has led a research project, funded by TUBITAK (The Scientific and Technological Research Council of Turkey) and Turkish State Railways, on development of prognostics and maintenance planning systems on railway switches. He has worked as faculty member at Cranfield University in the UK for three years and led several research projects at IVHM Centre.

Prof. Ian K. Jennions Ian's career spans over 30 years, working mostly for a variety of gas turbine companies. He has a Mechanical Engineering degree and a PhD in CFD both from Imperial College, London. He has worked for Rolls-Royce (twice), General Electric and Alstom in a number of technical roles, gaining experience in aerodynamics, heat transfer, fluid systems, mechanical design, combustion, services and IVHM. He moved to Cranfield in July 2008 as Professor and Director of the newly formed IVHM Centre. The Centre is funded by a number of industrial companies, including Boeing, BAe Systems, Rolls-Royce, Thales, Meggitt, MOD and Alstom Transport. He has led the development and growth of the Centre, in research and education, over the last three years. The Centre offers a short course in IVHM and the world's first IVHM MSc, begun in 2011. Ian is on the editorial Board for the International Journal of Condition Monitoring, a Director of the PHM Society, contributing member of the SAE IVHM Steering Group and HM-1 IVHM committee, a Fellow of IMechE, RAeS and ASME. He is the editor of the recent SAE book: IVHM – Perspectives on an Emerging Field.