Clinical Prognosis Model

Rishikesh Naware, Amaan Shaikh and Manas Singh

February 12, 2022

Clinical Prognosis Model

Rishikesh Naware, Amaan Shaikh and Manas Singh

Department of Information Technology
Thakur College of Engineering and Technology

Abstract

I. INTRODUCTION

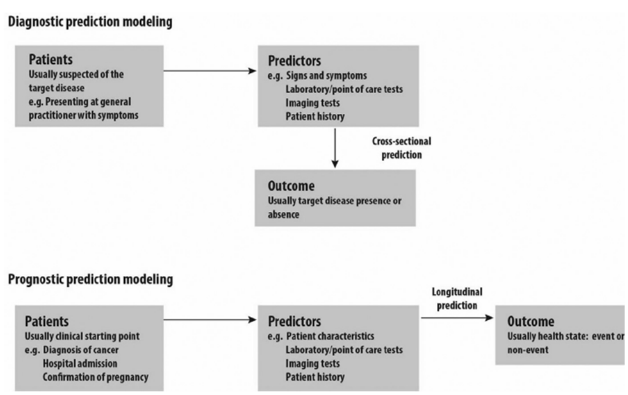
Clinical prediction models play an increasingly crucial function in current medical care, by means of informing healthcare professionals, sufferers and their family about outcome dangers, with the purpose to facilitate (shared) clinical choice making and enhance fitness outcomes. Diagnostic prediction fashion's purpose to calculate a character's hazard that an ailment is already present, at the same time as prognostic prediction fashions intention to calculate the hazard of specific heath states happening inside the destiny. This article serves as a primer for diagnostic and prognostic scientific prediction fashions, by discussing the simple terminology, some of the inherent demanding situations, and the need for validation of predictive overall performance and the assessment of impact of these models in scientific care. Prognostics refer to the estimation of the remaining useful life (RUL) of degrading systems and components based on the current health condition.

II. DEFINITIONS AND DIFFERENCES

A prognostic model is a formal combination of multiple predictors from which risks of a specific endpoint can be calculated for individual patients. For an individual with a given state of health, a prognostic model converts the combination of predictor values to an estimate of the risk of experiencing a specific endpoint within a specific period. The key difference among diagnostic and prognostic prediction models is inside the temporal courting be- tween the instant of prediction and the final results of interest. Prognostic fashions are vital at distinct stages in pathways leading to enhancements in health. The use of prognostic models ties in with the strong motion in the direction of stratified medicine, where choices concerning treatment alternatives are knowledgeable through an individual's profile of prognostic elements. In a diagnostic prediction version, the final results of interest are the modern fitness condition of the patient in the meanwhile of prediction. The natural observe design for diagnostic prediction fashions is a cross-sectional observe, in which statistics are accumulated from a set of patients suspected of having the goal situation, and measures each the outcome (goal situation popularity: presence as opposed to absence) and predictors on the equal second in time or with little or no time in among. An exception is the express use of observe-up to have a look at whether ailment happens in subjects. In whom the reference general cannot be completed, e.g., in imaging research in which it is impossible to perform a biopsy in a subject in which no lesion is detected for the duration of imaging (is predicated on the idea that detection of goal circumstance at follow-up accurately displays the target condition at moment of prediction). For prognostic prediction models, the focus is on predicting a future health final result that happens after the moment of prediction, additionally the use of predictors to be had in the meanwhile of prediction. Likewise, diagnostic models, which are commonly developed using logistic regression modeling or some variation thereof, may also suffer from incomplete outcome measurement.

III. PREDICTION MODEL DEVELOPMENT, PERFORMANCE AND IMPACT

The improvement of a diagnostic or prognostic version calls for numerous not unusual evaluation steps and decisions to be taken by way of the modeler. Some doctors believe that no prognostic version derived from one population can be generalized to patients drawn from some other, 27 within the same way that a few deny that scientific trials or overviews can tell man or woman choices approximately treatment. In brief, after the objectives of prediction are determined (e.g., outcome, target population and meant moment of use) and the dataset is prepared, constructing the prediction version requires decisions concerning the modeling framework (e.g., logistic regression), the candidate predictors to observe, a way to code the predictors and decide the practical form of the connection between the predictors and final results (e.g., a nonlinear impact of a continuous predictor, which include affected person age, using a spline feature), managing missing statistics in predictor and out- come, and in all likelihood choice amongst candidate predictors (e.g., the use of backward elimination in a regression model).



III. APPLICATION ON C-MAPSS DATA SET

A. Overview of the System and Data Set

In this section, the proposed JPM is illustrated and evaluated through the benchmarking C-MAPSS data set [25]. C-MAPSS is a tool developed by NASA for simulating a realistic large commercial turbofan engine that is monitored by multiple sensors. The C-MAPSS data set generated in [25] has been widely used as a benchmark system with multiple degradation signals in the prognostics and health management (PHM) field. C-MAPSS simulates an engine model of the 90 000-lb thrust with altitudes ranging from sea level to 40 000 ft, Mach numbers from 0 to 0.90, and sea-level temperatures from −60 ◦F to 103 ◦F. Users can adjust the conditions of aircraft altitude, Mach number, and throttle resolver angle to simulate different environmental conditions [25]. There are 14 inputs to simulate various degradation scenarios. The outputs include various sensor response surfaces and operability margins. A total of 21 variables out of 58 different outputs available from the model are used for analysis, as shown in Table I. To consider unit-to-unit variability, an unknown variance for initial wear level and a random noise were introduced.

A failure threshold for a hidden HI that is not accessible to users is predefined, beyond which the unit is considered failed. A total of four data sets with the corresponding failure modes and operational conditions were generated. In this article, we only consider two of the four data sets, FD001 and FD003, which are commonly used in performance evaluation and comparison. The FD001 data set has a single-failure mode (HPC degradation) and a single-operating condition, while the FD003 data set has one operating condition but two fault modes (HPC and fan degradation). For FD002 and FD004, there are six operating conditions mixed together, all of which affect the sensor values and the degradation process. It is inappropriate to estimate RUL based only on the 21-sensor data. Therefore, these two data sets are not considered here. For each data set considered in this article, there are 100 training units and 100 testing units. In the training data set, the fault grows in magnitude until system failure. In the test data set, the time series ends sometimes prior to system failure. A file of the actual remaining lifetime of the 100 testing units is also included for each data set. Sensor readings from the 21 outputs are collected at each observation epoch for each unit. The prognostic model is developed based on the available degradation patterns of the 100 training units and the testing data set is used for performance evaluation.

B. Variable Selection and Data Preprocessing

Among these 21 outputs, 14 outputs are highly related to the degradation process with an increasing or decreasing trend, while the other outputs are almost unchanged. Therefore, only these 14 degradation signals are included for further selection. The correlation analysis shows that there exist high correlations among these outputs (up to 0.96 for certain pairs). Signals with low correlation exhibit different signal patterns and involve different characteristics of the same unit. Therefore, the outputs are selected in such a way that the data show an obvious degradation trend and the pair-wise correlations of the selected outputs are as low as possible. To select the outputs based on the correlation, the hierarchical clustering algorithm is used, where (1 − correlation coefficient) is used as the distance or dissimilarity measure. The clustering dendrogram is shown in Fig. 5, where five clusters are obtained with a correlation threshold of 0.75. For each cluster, we randomly select an output, and the final outputs selected for prediction are Nc, T24, BPR, htBleed, and T30. The typical degradation forms of the selected sensor signals are shown in Fig. 6. All of these signals show an exponential functional form, which has been widely used to model cumulative damage processes [24], [28], [29]. Therefore, we use the exponential function to describe the degradation process of the turbofan engine. Following Liu et al. [1], we first perform log-transformation to the data and then apply linear models to the log-transformed data. Specifically, the quadratic polynomial function is assumed for the log-transformed data of each selected variable.

REFERENCES

[1] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," IEEE Trans. Autom. Sci. Eng., vol. 10, no. 3, pp. 652–664, Jul. 2013.

[2] M. S. Kan, A. C. C. Tan, and J. Mathew, "A review on prognostic techniques for non-stationary and non-linear rotating systems," Mech. Syst. Signal Process., vols. 62–63, pp. 1–20, Oct. 2015.

[3] M. Pecht, "A prognostics and health management for information and electronics-rich systems," in Engineering Asset Management and Infrastructure Sustainability. London, U.K.: Springer, 2012, pp. 317–323.

[4] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: A review," Int. J. Adv. Manuf. Technol., vol. 50, nos. 1–4, pp. 297–313, Jan. 2010.

[5] Z.-S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," Appl. Stochastic Models Bus. Ind., vol. 31, no. 1, pp. 16–32, Oct. 2015.

[6] Y. Wen, J. Wu, and Y. Yuan, “Multiple-phase modeling of degradation signal for condition monitoring and remaining useful life prediction,” IEEE Trans. Rel., vol. 66, no. 3, pp. 924–938, Sep. 2017.
[7] Y. Wen, J. Wu, D. Das, and T.-L. Tseng, “Degradation modeling and RUL prediction using Wiener process subject to multiple change points and unit heterogeneity,” Rel. Eng. Syst. Saf., vol. 176, pp. 113–124, Aug. 2018.
[8] Y. Wen, J. Wu, Q. Zhou, and T.-L. Tseng, “Multiple-change-point modeling and exact Bayesian inference of degradation signal for prognostic improvement,” IEEE Trans. Autom. Sci. Eng., vol. 16, no. 2, pp. 613–628, Apr. 2019.
[9] P. Lim, C. K. Goh, K. C. Tan, and P. Dutta, “Multimodal degradation prognostics based on switching Kalman filter ensemble,” IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 1, pp. 136–148, Jan. 2015.
[10] K. Liu and S. Huang, “Integration of data fusion methodology and degradation modeling process to improve prognostics,” IEEE Trans. Autom. Sci. Eng., vol. 13, no. 1, pp. 344–354, Jan. 2016.