Gaussian Process based Remaining useful life Prediction for Electric Energy Metering Equipment

Ning Li¹, Junfeng Duan²*, Jun Ma², Wei Qiu², Wei Zhang¹, Zhaosheng Teng²

¹Marketing Service Center (fund intensive center, metering center), State Grid Xinjiang, Urumqi, Xinjiang Uygur Autonomous Region, 830027, China
²College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, 410082, China

*Corresponding author’s e-mail: duanjunfeng@hnu.edu.cn

Abstract. Electric energy metering equipment (EEME) will fail in advance not as designed running in extreme environments. A multi-kernel Gaussian process regression model using measurement error data to perceive remaining useful life (RUL) for EEME is proposed. Firstly, the gauss kernel and periodic kernel are used to match the health index trend of EEME under a variety of typical environmental stresses. Furthermore, the Bayesian method and Monte Carlo Markov chain method are used to solve the model, and the Weibull distribution is used to fit the posterior trajectory to get the probability density estimation of the RUL.

1. Introduction

Electric energy metering equipment (EEME) in the field will be affected by many complex factors such as load and operating environment, resulting in random degradation of performance and health status, and ultimately resulting in equipment failure [1]. Condition-based maintenance (CBM) technology makes the prediction of remaining useful life (RUL) a hot issue in the engineering field [2,3]. Therefore, the use of monitoring information to make fault diagnosis expectations in time and predict RUL based on the current state of health is of great significance to the reliable service and economic operation of EEME [4].

Generally, RUL prediction methods can be divided into two categories: accelerated life methods and degradation prediction methods. The former needs to impose extreme stress on the equipment in the laboratory environment to force the equipment to accelerate aging until it fails, obtain the pseudo-life data of the equipment under extreme stress, and then reverse the RUL under normal stress as a reliability reference. This method can obtain the life of the equipment in a short time, but has the disadvantages of high test cost and cannot reflect the actual working conditions [5]. At the same time, it is difficult to obtain life data for high-reliability equipment. The latter method pre-prescribes the acceptable performance degradation threshold of the equipment, through the monitoring of equipment performance degradation state data and using intelligent algorithms to predict the data trend, so as to obtain the RUL. The method has high applicability, can be used for laboratory degradation tests and field operation tests, and has the advantages of low test cost, flexibility, and high efficiency. In [6,7], some advanced artificial intelligence algorithms such as the least squares support vector machines are used to do RUL prediction, but the result of these methods are point estimation, and the interpretivity is poor. In [8,9], the Gaussian process (GP) is used to predict the RUL of the lithium battery, and the results obtained are estimated interval, which is a more reasonable prediction result.
In order to characterize the possible long-term dependence between the historical degradation data of the equipment and the characteristic that is easily affected by the environment, this paper selects the environmental stress that has the impact on the EEME as the input vector of the GP, and then conducts multi-kernel matching to track and learn the measurement error data trend. The first threshold time point of the future trajectory in a fixed time is predicted as the end time of life, thereby obtaining an approximate analytical solution of the RUL distribution. The method proposed in this paper is verified by numerical simulation and the field operation data of smart meters in Xinjiang’s typical environmental test station.

2. Algorithm introduction

2.1. Gaussian process regression

The Gaussian process is a stochastic process, which is suitable for dealing with the prediction problems of small samples, strong randomness and multi-dimensional complex factors [10]. Set $X = \{ t, E \}$ to be the vector of stress variables that affect the error of the smart meter, which represents the impact of time $t$ and environmental stress $E$ on the operating characteristics of electrical energy metering equipment.

From the perspective of function space, the GP model assumes that $f(X)$ is a Gaussian process, and the function $f(X)$ is completely specified by its mean function $m(X)$ and covariance function $k(X, X')$, which is defined as follows:

$$
\begin{align*}
& m(X) = E[f(X)] \\
& k(X, X') = \text{Cov}[f(X), f(X')] \\
& f(X) \sim \text{GP}(m(X), k(X, X'))
\end{align*}
$$

(1)

In the formula, $m(X)$ is the mean value function of Gaussian process, generally 0; $k(X, X')$ is the covariance function of Gaussian process. $X \in \mathbb{R}^d$ is a $d$-dimensional random variable.

Accordingly, it can be described as predicting the output $y_\ast$ under the new input $X_\ast$ according to the given dataset $D$, That is, the predictive function $y_\ast$ can be obtained by induction:

$$
\begin{align*}
& y = m(X), \\
& y_\ast = K(X_\ast, X)[K(X, X) + \delta^2 I]^{-1} y \\
& \text{cov}(y_\ast) = k(X_\ast, X) - K(X_\ast, X)K(X, X)[K(X, X) + \delta^2 I]^{-1} K(X_\ast, X)
\end{align*}
$$

(2)

Where $K(X, X_\ast)$ is the covariance of test data and training data, and $K(X_\ast, X_\ast)$ is the covariance of test data itself.

The posterior distribution of predicted $y_\ast$ can be obtained from above

$$
y_\ast|X, y, X_\ast \sim N(\overline{y}_\ast, \text{cov}(y_\ast))
$$

(3)

and,

$$
\overline{y}_\ast = K(X_\ast, X)[K(X, X) + \delta^2 I]^{-1} y
$$

(4)

$$
\text{cov}(y_\ast) = k(X_\ast, X) - K(X_\ast, X)K(X, X)[K(X, X) + \delta^2 I]^{-1} K(X_\ast, X)
$$

(5)

Where $\overline{y}_\ast$ is the mean value of test data $X_\ast$ corresponding to predicted value $y_\ast$, $\text{cov}(y_\ast)$ is its variance. GP can give the probability distribution of the predicted data and get the interval prediction with certain confidence. Under the confidence degree of $1-\alpha$, the probability results with the confidence interval as follows can be predicted:
\[ l, h = \left[ \hat{\mu} - z_{\alpha/2} \delta_{\gamma}, \hat{\mu} + z_{\alpha/2} \delta_{\gamma} \right] \]

(6)

Where \( l \) and \( h \) represent the lower and upper confidence bounds of the predicted value respectively, \( z_{\alpha/2} \) represents the \( \alpha \) loci of the standard Gaussian distribution.

2.2. Gaussian process kernel function selections and parameters solving

Affected by the environment and its own degradation characteristics, the degradation of the measurement error of the EEME presents a nonlinear characteristic, and it is difficult for a single kernel to describe its true degradation law [11]. Compared with the traditional and single gauss kernel, the combined kernel form is more diverse and can specifically describe the characteristics of different stresses. The gauss kernel function controls the sample distance correlation by \( \ell \) se, which can describe the long-term steady upward trend of the measurement error degradation data with time stress \( t \); the periodic kernel function describes the periodic change of the measurement error due to temperature and humidity stress. In order to fully characterize the influence of multi-dimensional stress on the measurement error, this paper combines the above kernel function based on prior knowledge, the expression of the combined kernel function of the Gaussian process is

\[
\begin{align*}
&f(X) = \text{GP}(m(X), \eta_{se}^2 \times k_{se} + \eta_{pe}^2 \times k_{pe} + k_{ma} + k_{on}) \\
&= \text{GP}(m(X), \eta_{se}^2 \times k_{se} + \eta_{pe}^2 \times k_{pe} + k_{ma} + k_{on})
\end{align*}
\]

(7)

In the formula, \( \eta_{se} \) and \( \eta_{pe} \) are hyperparameters that control the amplitude change.

Bayesian inference can be used to estimate the hyperparameters by calculating the posterior distribution of the parameters. Under the Bayesian framework, the prior distribution of hyperparameters needs to be determined first [12]. The kernel function needs to satisfy Mercer’s theorem, for the amplitude hyperparameter \( \eta \) of the kernel function and the noise term \( \delta_n \), the semi-Cauchy distribution is used to ensure its positive shape, which is initialized to \( \eta \sim \text{HalfCauchy}(2), \delta_n \sim \text{HalfCauchy}(1) \); characteristic length scale the hyperparameter \( \ell \) corresponds to the smoothness of different kernel functions, it uses a richly informative gamma function and is initialized to \( \ell \sim \text{Gamma}(4,3) \); for the period \( p \), the mean range can be observed from the data characteristics, so the normal distribution is used, it is initialized to \( p \sim \text{Normal}(12,0.05) \). Bayesian estimates are as following:

\[
\Pr(y) = \prod_{i=1}^{m} \Pr(y \mid \theta) \Pr(\theta)
\]

(8)

\[
\Pr(y \mid \theta) = \int \Pr(y \mid \theta) \Pr(\theta | y) \ d\theta
\]

(9)

Where \( \theta = \{ p, \eta_{pe}, \ell_{pe}, \ell_{ma,1}, \eta_{pe}, \ell_{se}, \ell_{ma,2}, \delta \} \) is the vector containing all the parameters of the Gaussian process combination kernel.

2.3. Principles of RUL prediction for EEME

In the process of instrument degradation, once the health index exceeds the given threshold, the moment is considered as the failure time, that is, the end of life time. The RUL can be defined as:

\[
T = \inf \left\{ t : y_t > \omega_{\text{upper}} \left| y_t < \omega_{\text{lower}} \right. \right\} - t_m
\]

(10)

where \( \omega_{\text{upper}} \) is the upper limit of the smart meter failure threshold; \( \omega_{\text{lower}} \) is the lower limit of the smart meter failure threshold, and \( t_m \) is the current time of prediction.

The prediction of system RUL is to calculate the distribution density of system RUL after obtaining the prediction result of the probability density of the state variable at \( t_{m+n} \). The principle of residual life distribution calculation of the system is shown in Fig 1:
The performance degradation and stress of the monitoring equipment at \( t_1 < t_2 < \ldots < t_m \) are \( Y_{t_1:m} \) and \( S_{t_1:m} \) respectively. It is assumed that the time-varying stress experienced by the product after \( t_m \) can be expressed as a discrete sequence \( \{S_{k+1}, S_{k+2}, \ldots \} \). When the stress changes with time, the Monte Carlo simulation (MC) method is used to approximate the remaining life distribution under the time-varying stress profile. The basic idea of MC method to solve residual life distribution is as follows: The degradation model updated to \( t_m \) and the stress profile \( \{S_{m+1}, S_{m+2}, \ldots S_{m+n}\} \) after \( t_m \) are used to generate multiple degraded orbit samples to predict the trajectory that crosses the threshold at time \( t_{m+n} \) to obtain the remaining life distribution of the evaluated product under time-varying stress.

The samples of the predicted results of system health indicator variables from \( t_m \) to \( t_{m+n} \) form a cluster trajectory in the state space, and the intersection of the cluster trajectory and the failure threshold form an intersection region. The trajectory weight is accumulated at the intersection of the posterior sample trajectory and the threshold, and the result reflects the probability of system failure at the corresponding time. Through the above method, a bunch of trajectory samples from the time \( t_m \) to \( t_{m+n} \) of the RUL distribution of the system can be obtained, and the corresponding weights are normalized, and then according to the total probability formula, the probability density of the remaining life distribution of the system can be obtained as:

\[
P(t_n) = \sum_{i=1}^{N} \Pr(y_{\text{upper}} \leq Y_i \leq y_{\text{upper}})w_i
\]

(11)

Where \( w_i \) is the normalized weight of the posterior trajectory.

3. Case study

To verify the performance of the proposed model in the RUL prediction for EEME, the measurement error data samples of smart meters produced by different manufacturers and operating on-site in Xinjiang’s high dry heat typical environmental test station were analyzed. The smart meters used in the test are all Class I meters, and the measurement error threshold is set to 60% of the accuracy level to provide early warning for maintenance.

The measurement error data of the smart meters in operation from four different manufacturers in Xinjiang are fused in high dry heat environment, and the potential model parameters are trained by the multi-kernel Gaussian process. Based on this, the future trend prediction is made according to the time specified by the measurement regulations. As shown in Fig 2, based on the current existing smart meter measurement error observation data for trend identification, the measurement error changes in a trend
with an annual cycle. The dark thin solid line is the predicted mean value, which represents the general changes of smart meter entities in the typical environment of high dry heat in Xinjiang. The light shaded area is an area composed of several posterior trajectories, which may fail in the positive direction or in the negative direction. This is related to the attributes of smart meters from different manufacturers.

![Measurement error prediction diagram of smart meters in Xinjiang](image1)

**Fig 2 Measurement error prediction diagram of smart meters in Xinjiang**

The key of RUL prediction is to solve the probability density function of RUL distribution. Weibull distribution is a model commonly used in reliability engineering to model material loss, mechanical wear, and performance degradation [13]. The Weibull curve fitting is performed on the time data when the posterior trajectory crosses the failure threshold to obtain the RUL probability density function of the smart meter in Xinjiang as shown in Fig 3. It can be seen that the Weibull probability density peak value of the remaining service life of the smart meter under the influence of the typical environmental stress of high dry heat in Xinjiang is 42.5. That is to say, based on the current operating conditions, the probability of general failure of smart meters in the 42 to 43 months after operation is the greatest, and attention should be paid to starting troubleshooting and verification. However, in the probability density histogram, the RUL peaks in the 30th and 55th months. This is due to the fact that the measurement error data of the smart meter is affected by the high dry heat environmental stress and periodically reaches the failure threshold. Considering that the time difference between peak occurrences is 12 months, that is, the failure of smart meters will occur concentratedly in a fixed month in a cycle.

![The posterior distribution of the RUL of smart meters](image2)

**Fig 3: The posterior distribution of the RUL of smart meters**
4. Conclusion
In a typical multi-stress environment, the acceleration index of the electric energy metering equipment will be out of tolerance. This paper proposes a multi-kernel Gaussian process regression based on measurement error data to predict the remaining useful life of smart meters. This method abstracts a type of actual fault prediction problem as a state estimation problem under the condition that the model contains unknown slowly varying parameters, and the estimated value of the state variable at a certain time in the future solved as a guide for the maintenance strategy of electric energy metering equipment.

References
[1] Lan, J., Yuan, M., Yuan, H. J., & Lv, P. (2017, July). Reliability assessment under real world environmental stress. In 2017 Second International Conference on Reliability Systems Engineering (ICRSE) (pp. 1-5). IEEE.
[2] Liu, X., Li, J., Al-Khalifa, K. N., Hamouda, A. S., Coit, D. W., & Elsayed, E. A. (2013). Condition-based maintenance for continuously monitored degrading systems with multiple failure modes. IIE transactions, 45(4), 422-435.
[3] Guo Jian, Xu Zongchang, Zhang Wenjun. Summary of state-based equipment failure prediction technology[J]. Journal of Artillery Launching and Control, 2019, 40(02): 103-108.
[4] Ning, L., Haiyang, L., Yongchao, W., & Qiaoya, L. (2021). Research on the Cause Analysis and Improvement Measures of Measurement Error Based on Model Forecast. In Journal of Physics: Conference Series (Vol. 1750, No. 1, p. 012005). IOP Publishing.
[5] An, D., Choi, J. H., & Kim, N. H. (2018). Prediction of remaining useful life under different conditions using accelerated life testing data. Journal of Mechanical Science and Technology, 32(6), 2497-2507.
[6] Lin, J., Su, L., Yan, Y., Sheng, G., Xie, D., & Jiang, X. (2018). Prediction method for power transformer running state based on LSTM_DBN network. Energies, 11(7), 1880.
[7] Dong, S., & Luo, T. (2013). Bearing degradation process prediction based on the PCA and optimized LS-SVM model. Measurement, 46(9), 3143-3152.
[8] Liu, J., & Chen, Z. (2019). Remaining useful life prediction of lithium-ion batteries based on health indicator and Gaussian process regression model. Ieee Access, 7, 39474-39484.
[9] Pang Jian Yue, Ma Yunyi, Liu Dabu & Peng Yu. (2014). Lithium-ion battery remaining useful life indirect prediction method. Chinese scientific papers (01), 28-36.
[10] Melkumyan, A., & Ramos, F. (2011, June). Multi-kernel Gaussian processes. In Twenty-second international joint conference on artificial intelligence.
[11] Rasmussen, C. E. (2003). Gaussian processes in machine learning. In Summer school on machine learning (pp. 63-71). Springer, Berlin, Heidelberg.
[12] Chen, Z., & Wang, B. (2018). How priors of initial hyperparameters affect Gaussian process regression models. Neurocomputing, 275, 1702-1710.
[13] Hallinan Jr, A. J. (1993). A review of the Weibull distribution. Journal of Quality Technology, 25(2), 85-93.