Spacecraft Telemetry Data Anomaly Detection Based On Multi-objective Optimization Interval Prediction

Xunjia Li
College of Systems Engineering
National University of Defense Technology
Changsha, China
lixunjia0526@qq.com

Tao Zhang
College of Systems Engineering
National University of Defense Technology
Changsha, China
zhangtao@nudt.edu.cn

Kaiwen Li
College of Systems Engineering
National University of Defense Technology
Changsha, China
1176414603@qq.com

Yajie Liu
College of Systems Engineering
National University of Defense Technology
Changsha, China
liuyajie@nudt.edu.cn

***Abstract*—Spacecraft telemetry data anomaly detection is crucial for the timely detection of potential malfunction in spacecraft systems. Because of the uncertainty of prediction, interval prediction models are more suitable for anomaly detection than point prediction and probability prediction. This paper first puts forward an anomaly detection framework based on the traditional LUBE model, and introduces a method to eliminate the error of the model itself in the framework of anomaly detection. Considering that the LUBE method judges the quality of the prediction interval, there are two indicators, interval width and interval coverage, which is essentially a multi-objective optimization problem. Therefore, this paper proposes a LUBE interval prediction model based on multi-objective optimization. Compared with the traditional model, the combination of the two indicators is obviously superior to the original method. Finally, the effectiveness is proved by anomaly detection experiments of public datasets and spacecraft telemetry data.**

Keywords-anomaly detection; interval prediction; LUBE model; multi-objective optimization

I. INTRODUCTION

Spacecraft as a large complex system, its telemetry data involves huge number of sensor measurements on various types of subsystems, but before the fault occurs, some symptoms in the telemetry data can help the ground monitoring to achieve timely adjustment and control of the spacecraft, thus avoiding major accident. Therefore, data-driven fault detection research is of great significance for the reliability and security of large and complex systems.

In the previous research of related scholars, the classification of anomaly detection methods was summarized [1,2], which are roughly classified into six varieties: fixed threshold based, statistical analysis based, nearest neighbors based, clustering based, partition-based and prediction-based methods. Each algorithm has its own application requirements and scenarios. The fixed threshold algorithm is simple and fast, but it is difficult to find the abnormal evolution process. Statistical analysis-based method lacks universality when modeling data sets with specific distribution. Nearest neighbors-based and clustering-based methods are subject to the assumption that normal points appear in some fixed areas, while abnormal points appear in unreasonable areas away from normal points [3], and the effects of the two methods are extremely sensitive to metric evaluation functions and outliers. Partition-based methods conduct supervised learning by marking normal and abnormal labels, but the labeling of data requires professional work, and this process is too cumbersome and expensive, at the same time, the integrity of the label can affect the effect of the entire algorithm [4,5]. In the background of massive data, the telemetry parameters will be measured and saved in real time, and will reach in form of streaming data, so the above methods are not satisfied in real time. The prediction-based algorithm is more appropriate than the above method in anomaly detection problem because it does not require prior knowledge, does not depend on tags, and can perform real-time detection. The prediction-based algorithm achieves anomaly detection by comparing actual data with model prediction output data. Therefore, this paper focuses on the prediction-based method.

The prediction-based method marks the sample by the difference between predicted value and observational value [6]. However, the current prediction method is mainly a point prediction method, which can only obtain deterministic point prediction, and cannot estimate the probability that the value may appear and construct a prediction interval. The accuracy and range of the approximation cannot be reflected, so the interval prediction model with uncertainty is more suitable for anomaly detection.

The existing interval predictions are mainly divided into probabilistic prediction methods and boundary estimation methods. Probabilistic prediction is one of the more studied methods. It is a process of predicting the probability distribution of physical quantity in the prediction period by using the relevant sample statistical information before the prediction period. In probabilistic prediction, uncertainty are represented by probability statistical indicators including probability density function (PDFs), quantiles, mean and

Project supported by National Natural Science Foundation of China (71571187) and Excellent Youth Foundation of Hunan Scientific Committee (2017JJ1001).

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

variance [7]. However, the upper and lower boundary estimation method directly predicts the upper and lower bounds of the interval, and does not involve probability distribution information of the target to be tested. It does not need any prior statistical assumptions and has strong versatility and intuitiveness, which is an typical benefit of this approach. The LUBE (Lower Upper Bound Estimation) method is a classical method raised in recent years [8]. It establishes a neural network with two outputs for predicting the upper and lower bound of the interval respectively. Because the LUBE method judges the quality of the prediction interval, there are two indicators of interval width and interval coverage, which is essentially a multi-objective optimization problem. Therefore, the idea of using multi-objective optimization method to optimize neural network structure is proposed. Pareto frontier of a series of optimal neural network parameters is obtained by multi-objective optimization algorithm. By selecting the points of Knee point region as the final neural network parameters, the most satisfactory interval prediction results for decision makers are obtained. The optimized prediction results make the combination of the interval width and the interval coverage better than the original method.

Based on this method, an anomaly detection framework is constructed, and the parameter indicators of anomaly detection are proposed as the decision basis. For the purpose of testing the proposed method, experimental data verification of the public datasets and spacecraft telemetry data proves the effectiveness and applicability.

II. Interval Prediction Anomaly Detection Model Based on Upper and Lower Bounds Estimation

A. Forecast interval evaluation index

Evaluating the performance of an interval mainly starts from two indexes: PICP and NMPIW. Among them, PICP is the abbreviation of the Prediction Interval Coverage Probability, and NMPIW is the abbreviation of the Normalized Mean Prediction Intervals Width [9].

PICP is calculated by counting the amount of real data that is within the prediction interval. The formula for defining PICP is as follows:

$$PICP = \frac{1}{N}\sum_{i=1}^{N} c_i \tag{1}$$

where N is the sample amount of the test set, c_i is 1 when the true value is between the boundaries of prediction interval, otherwise c_i is 0.

The larger PICP indicator, the better performance of the constructed interval. But an overly wide prediction interval will not make any sense. Therefore, the width of prediction interval should be considered at the same time. The MPIW indicator is defined as follows:

$$MPIW = \frac{1}{N}\sum_{j=1}^{N} (U_i - L_i) \tag{2}$$

The interval width is typically measured by the normalized MPIW (NMPIW). Its definition is as follows:

$$NMPIW = \frac{MPIW}{R} \tag{3}$$

where the value of R depends on the target to be measured. The smaller the NMPIW indicator, the better performance of the constructed interval.

Both PICP and NMPIW have a value range of [0,1]. Increasing the PICP generally represents increasing the NMPIW, so the two indicators are mutually exclusive. In some literatures, the integration index CWC (coverage width-based criterion) is raised for mutual exclusion characteristics of these two objectives. Definition is as follows:

$$CWC = NMPIW \cdot \left(1 + \gamma(PICP) e^{-\eta(PICP-\mu)}\right) \tag{4}$$

which $\gamma(PICP)$ is defined as:

$$\gamma = \begin{cases} 1, & \text{if } PICP < \mu \\ 0, & \text{if } PICP \geq \mu \end{cases} \tag{5}$$

where μ is installed to a confidence level of $1-\alpha$. If interval coverage is higher than $1-\alpha$, the performance of the interval is only evaluated by NMPIW. At the same time, if the PICP doesn't achieve $1-\alpha$, it will be punished. η is usually set to a larger penalty factor, such as 50. The performance of the constructed interval is getting better as the *CWC* gets smaller. [8].

B. LUBE interval prediction model

Based on the above-mentioned prediction interval evaluation index, the interval bounds are straightway predicted by neural network. The LUBE model proposed in 2010 is a classical algorithm and is the basic framework of many such methods. LUBE uses the exponential CWC to integrate the two into a loss function and then optimize network structure by simulation annealing algorithm [8].

The structural diagram of LUBE neural network is described in Figure 1.

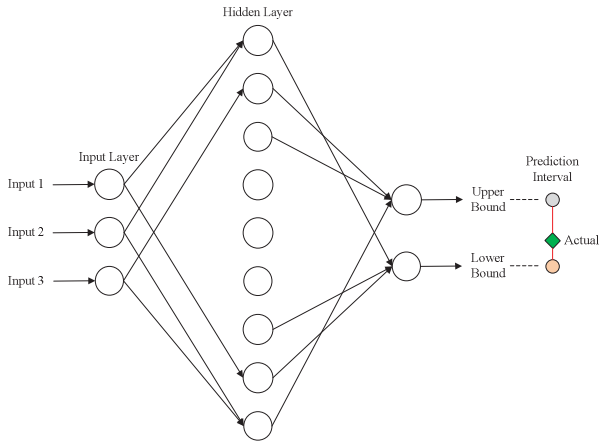

Figure 1. Interval prediction neural network structure

The algorithm steps are briefly described as follows:

- **Step1**: Preparation process: The raw dataset is separated into two portions, namely training dataset and test dataset, and the data is normalized.
- **Step2**: Parameter initialization: Initialize network parameter w_{opt} and initialize the temperature and CWC_{opt} of algorithm.
- **Step3**: Update the temperature until the temperature reaches the threshold.
- **Step4**: Generate a new neural network parameter w_{opt} by disturbance, use network parameters to get the prediction interval and compute index value CWC_{new}. If $CWC_{new} < CWC_{opt}$ is satisfied, then let $CWC_{opt} = CWC_{new}$, $w_{opt} = w_{new}$, otherwise generate a random number r, if $r \geq e^{-\frac{(CWC_{new} - CWC_{opt})}{\kappa T}}$, then still adopt the new solution.
- **Step5**: If temperature value attains the set threshold, stop training and proceed to step 6, otherwise enter step 3.
- **Step6**: Use the obtained neural network to construct interval on test dataset, then evaluate model performance by evaluation index.

C. *Anomaly detection based on LUBE interval prediction model*

As shown in Figure 2, based on the LUBE interval prediction model for anomaly detection, the framework includes offline training process and online testing process.

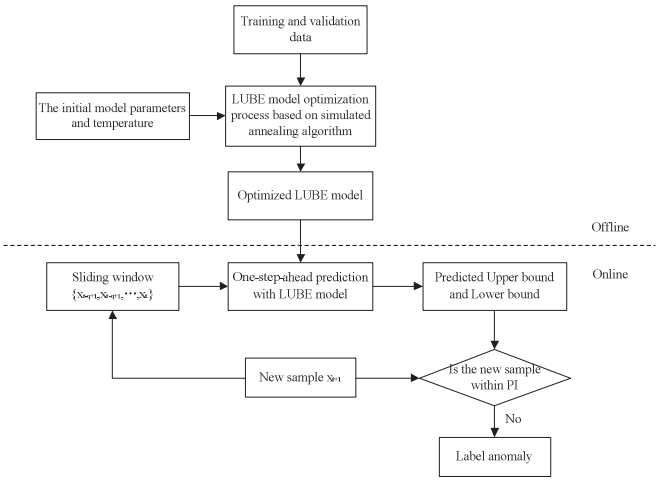

Figure 2. Anomaly detection framework based on LUBE model

- **Offline training**: Offline training uses the simulated annealing algorithm to iteratively optimize model parameters, and obtains the optimized LUBE neural network to meet the interval prediction.
- **Online testing**: In the online testing phase, Real-time data are gradually added by building a sliding window. The neural network model performs one-step prediction to obtain the predicted bounds of the new data. Then, actual value of the new data is compared with the upper and lower bound estimates to mark whether the sample is an abnormal value. By repeating these steps, online testing can be continuously performed.

For anomaly detection, the inaccuracy of the prediction interval will lead to some false alarms. The reason for this is mainly in the training process. There are two main situations: (1) Whether all the training data are located in the prediction interval of the model; (2) Whether the width of the prediction interval is too narrow to contain all training data. By analyzing the above two cases, we need to further discuss the applicability of the LUBE model for anomaly detection problems. Based on the above analysis, the improved implementation framework for the uncertainty arising from the data and the model is described in Figure 3:

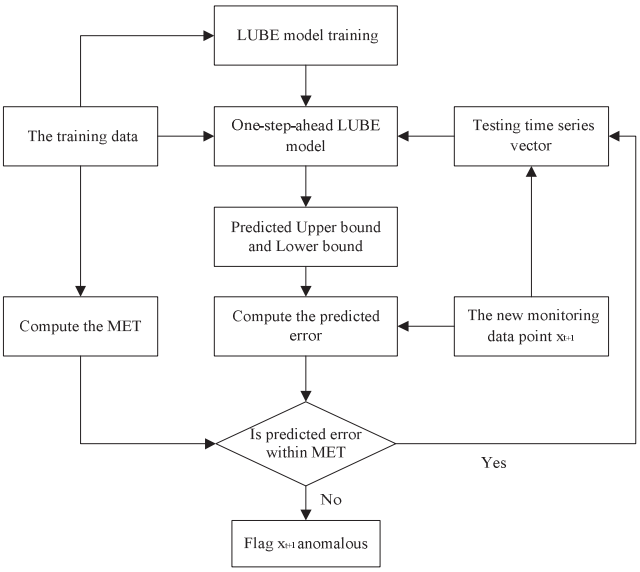

Figure 3. Improved framework

The proposed anomaly detection framework has three main components: First, the trained LUBE model is obtained through the training data. Then, the model error limit (MET) is calculated by the improved k-fold cross-validation. Finally, anomaly detection is carried out based on one-step LUBE model and MET.

Here we mainly introduce the calculation method of the model error limit (MET). For the uncertainty interval prediction in this paper, the measure of anomaly is reflected in whether the actual value is within the prediction interval and deviated from the prediction interval. Here, the variable is defined: prediction error (PE). The prediction error (PE) characterizes the actual value deviating from the prediction interval and is calculated as follows:

$$PE_t = \begin{cases} y_t - UB_t, & \text{if } y_t > UB_t \\ LB_t - y_t, & \text{if } y_t < LB_t \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where t is time stamp, y_t is model prediction value at time t, and UB_t and LB_t are the model prediction upper and lower bounds.

K-fold cross-validation is a classical method for performing error evaluation to eliminate the influence of model uncertainty [10]. The steps are as follows:

- Assuming that the total training dataset number is n, randomly select $1/k$ of training data as test input and target, and the remaining $(1-1/k)$ part are used as training input and target.
- Calculate the test sample PE.
- Repeat the above process k times, the total error reaches n.
- To simplify the calculation, assume that the model residuals obey the Gaussian distribution., MET is a range of symmetry about 0, and its calculated is as follows:

$$MET = \left[\underline{MET}, \overline{MET}\right] = \left[-t_{\alpha/2,n-1} \times s\sqrt{1+\frac{1}{n}}, t_{\alpha/2,n-1} \times s\sqrt{1+\frac{1}{n}}\right] \quad (7)$$

where S is the standard deviation of the error series, and $t_{\alpha/2,n-1}$ is parameter of a Students t-distribution.

The improved IPE is as follows:

$$IPE_t = \begin{cases} y_t - UB_t - \overline{MET}, & \text{if } y_t > UB_t + \overline{MET} \\ LB_t - y_t + \underline{MET}, & \text{if } y_t < LB_t + \underline{MET} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

D. Multi-objective optimization LUBE model

Firstly, the concept of Pareto front is introduced. For an optimization problem with multiple targets, there is often mutual exclusion between multiple targets. The improvement of one target usually leads to the deterioration of another target. Therefore, for a problem with multiple targets, the optimal solution is not singular, and there generally exist multiple Pareto optimal solutions. Solution x_1 is called the Pareto optimal solution, only if no solution exists in the decision space that can Pareto dominate x_1, it is also called the non-dominated solution. Therefore, the optimal solutions to the problem are generally a series of Pareto optimal solutions, which constitute the Pareto front in the target space.

So the traditional way of integrating multiple targets into one goal, such as CWC, is just a point on the Pareto front that corresponds to the problem. It is just the result of solving the problem in a certain direction. The result may not be the most satisfactory solution for the decision maker, it may be the point A on the Pareto frontier, and the best point should be point B, which is the Knee point in multi-objective optimization theory. The so-called Knee point on the Pareto front is defined as: At this point, slightly changing the value of any one target will cause the value of the other target to deteriorate drastically. Therefore, this point is the most ideal solution for decision makers, and the solution obtained by the traditional LUBE method does not necessarily obtain the most ideal solution. The obtained prediction interval is not necessarily the optimal prediction interval. This is a fundamental flaw in the traditional boundary estimation method. Define this method as KM-LUBE model.

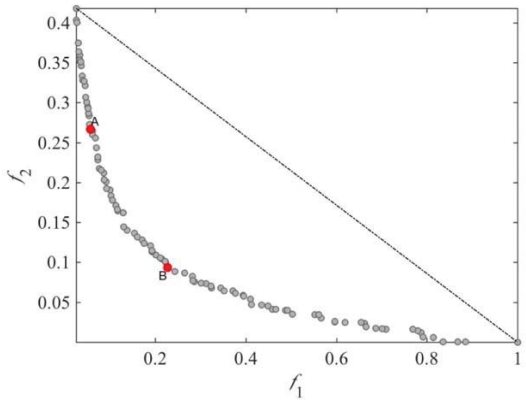

Figure 4. Schematic diagram of interval prediction Pareto front

The method uses a single layer neural network with one hidden layer and two outputs as basic model. The network structure is described in Figure 1. The input layer nodes number is N_{input}, the hidden layer nodes number is n_h, and the output layer nodes number is 2. The parameters of the network are $[\omega, b] = [\omega^{(1)}, b^{(1)}, \omega^{(2U)}, b^{(2U)}, \omega^{(2L)}, b^{(2L)}]$, where $\omega^{(1)}$ means the weight coefficient between the neurons of the input and hidden layer, $b^{(1)}$ means the bias coefficient of each neuron in the hidden layer, $\omega^{(2U)}$ and $\omega^{(2L)}$ are the weight connection coefficients between the hidden layer and the upper and lower bounds U_i and L_i of the output, $b^{(2U)}$ and $b^{(2L)}$ are the offset coefficients of U_i and L_i of the output.

For the model constructed in this paper, all the parameters that need to be optimized are the weight coefficient ω and the bias coefficient b between different layers of the neural network, so the dimensions of each parameter are

$$\dim(\omega^{(1)}) = N_{input} \cdot n_h \quad (9)$$

$$\dim(\omega^{(2U)}) = \dim(\omega^{(2L)}) = n_h \quad (10)$$

$$\dim(b^{(1)}) = n_h, \dim(b^{(2U)}) = \dim(b^{(2L)}) = 1 \quad (11)$$

The total number of parameters that need to be optimized is as follows:

$$N_{input} \cdot n_h + 2n_h + n_h + 2 = (N_{input} + 3) \cdot n_h + 2 \quad (12)$$

Therefore, the multi-objective optimization algorithm need to optimize $(N_{input} + 3) \cdot n_h + 2$ parameters, so that

corresponding model can obtain the optimal PICP and NMPIW. The specific steps of the Knee point-based multi-objective boundary estimation method are as follows:

- **Step1**: Preparation process: The raw dataset is separated into two portions, namely training dataset and test dataset, and the data is normalized.
- **Step2**: Model initialization: Firstly, construct a single-output point prediction neural network. Its structure is the same as the interval prediction except the last layer. Use the gradient descent method to quickly train to obtain the initial parameters of the model, and then carry out the parameter migration, that is, the parameters of the first layer of point prediction network framework are transferred to interval prediction network framework, and the parameters of second layer increase the random perturbation, thereby obtaining the initial population of interval prediction neural network model.
- **Step3**: Adopt the selected multi-objective optimization method to iterate and optimize the initial parameter population. Sample part of the data from the original training data in each generation. According to the calculated index, the population selection and evolution are performed according to the selected multi-objective algorithm, and the final Pareto front is obtained after a certain number of iterations. Each point on the Pareto front represents a neural network structure.
- **Step4**: Calculate the Knee point, which is the best selected neural network parameter.
- **Step5**: Construct interval prediction of the test set using the obtained neural network.

In the aspect of multi-objective algorithm selection, this paper adopts NSGA-II algorithm [11], which is one of the most widely studied and used multi-objective optimization algorithms currently. It has the advantages of low complexity, fast speed and excellent solution convergence, and proves to be the benchmark for evaluating the performance of other algorithms.

The basic flow of classic NSGA-II algorithm is briefly described in following steps:

- **Step1:** First, execute fast non-dominated sorting on the initial population, then carry out the calculation of the congestion distance.
- **Step2:** Secondly, select, cross and mutate to generate progeny population according to the result of step 1.
- **Step3:** Finally, composite the progeny population with the original population to form a new mixture population, execute fast non-dominated sorting on the initial population, then carry out the calculation of the congestion distance like Step1, select N individuals, which represent elite strategy, and then repeat steps 1-3 until the end of the iteration.

III. EXPERIMENTAL RESULTS AND ANALYSIS

For the purpose of confirming the good performance of the anomaly detection algorithm put forward in this paper, this paper uses a series of data sets to conduct experiments. First, the Ma_Data and ECG data are public data used to test and verify the universality of the detection algorithm in this paper. Afterwards, the true spacecraft telemetry data is used in experiments to verify the practicability and effectiveness of the proposed algorithm in practical scenes.

Since the positive and negative values of PE and MET have no effect on the abnormal judgment decision, in the experiments conducted below, we only adopt the positive value of PE and MET to reduce the amount of calculation. For the method proposed in this paper for uncertainty, the IPE indicator is calculated to perform anomaly detection.

There are four possible cases for the test results, as shown in Table 1.

TABLE I. 4 POSSIBLE PREDICTIONS RESULTS

Results	Real normal data	Real abnormal data
Detected normal data	*TP*	*FP*
Detected abnormal data	*FN*	*TN*

Based on the performance indicators of the machine learning classifier, this paper uses three indicators of detection rate (*DR*), false acceptance rate (*FAR*) and accuracy (*ACC*) to verify the effectiveness of the anomaly detection algorithm.

1) The detection rate (*DR*) is the ratio of the real abnormal data that is correctly detected as abnormal, *DR* is calculated by the following formula.

$$DR = \frac{TN}{FP + TN} \times 100\% \tag{13}$$

where molecule *TN* indicates the quantity of data marked as abnormal in the real abnormal data, and denominator *FP+TN* indicates the total quantity of the real abnormal data.

2) The false acceptance rate (*FAR*) is the ratio of real normal data that is wrongly detected as abnormal, *FAR* is calculated by the following formula.

$$FAR = \frac{FN}{FN + TP} \times 100\% \tag{14}$$

where molecule *FN* indicates the quantity of data marked as abnormal in real normal data, and denominator *FN+TP* indicates the total quantity of the real normal data.

3) Normally, the classification of normal and abnormal is unbalanced. In addition, *DR* and *FAR* are contradictory. In order to effectively estimate performance through an indicator, *ACC* is used and is calculated by the following formula.

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{15}$$

where molecule *TP+TN* indicates the quantity of correct detection data, denominator *TP+FP+FN+TN* indicates the

quantity of all data detected, that is, the accuracy (*ACC*) is the ratio of the data correctly detected in the entire detection dataset.

The larger the *DR* and *ACC* and the smaller the *FAR*, the better the performance of the anomaly detection.

A. Experiments on public data sets

First, the public data set is used to test and verify the universality of the detection algorithm in this paper, and an artificial data set and two ECG data are selected.

Ma_Data is a set of data generated by simulation as following formulas [10]:

$$Y_1 = \sin\left(\frac{40\pi}{N}t\right) + n(t) \tag{16}$$

$$Y_2 = \sin\left(\frac{40\pi}{N}t\right) + n(t) + e_1(t) \tag{17}$$

where $n(t)$ is Gaussian white noise with mean of zero and standard deviation of 0.1, and $e_1(t)$ is a period of simulated Gaussian white noise with mean of zero and variance of 0.5.

The ECG data contains two ECG data series: ECG1 and ECG2, which have periodic characteristics but exhibit different trends [12]. For ECG1, faults happen roughly in the sample index of 2340 to 2390 and 2410 to 2460. Accordingly, the faults of ECG2 happen in the sample index of 2830 to 2910.

The data of Ma_Data and ECG is shown in Figure 5, where the red part is the abnormal section.

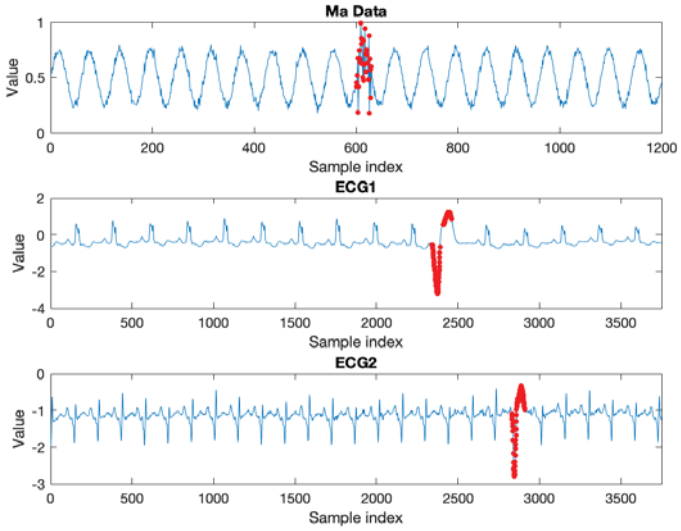

Figure 5. Public data sets containing abnormal data

The detection results of the LUBE and KM-LUBE algorithms on the ECG data set are shown in the figures below. It can be seen from the figures that the prediction interval of the KM-LUBE algorithm is better than the LUBE algorithm. Both algorithms can detect anomalies well.

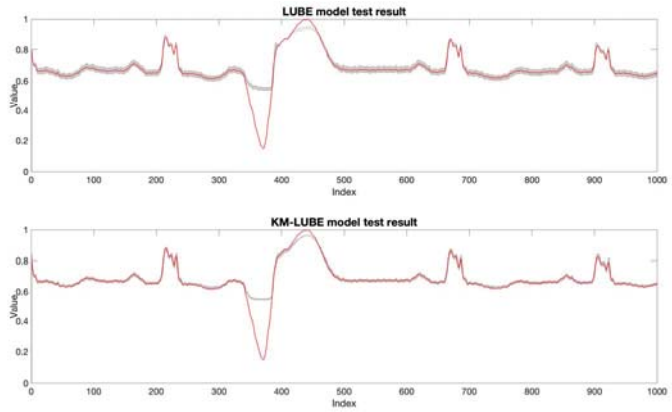

Figure 6. Performance comparison of detection algorithms of LUBE (top) and KM-LUBE (bottom) algorithms on ECG1

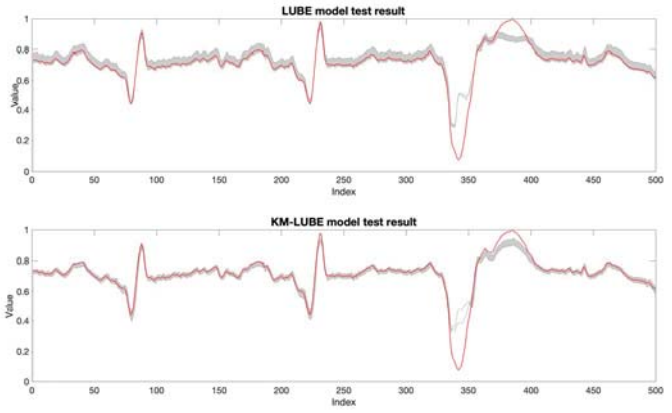

Figure 7. Performance comparison of detection algorithms of LUBE (top) and KM-LUBE (bottom) algorithms on ECG2

B. Experiments on spacecraft telemetry data

In the aerospace field, thousands of sensors in different subsystems of spacecraft conduct real-time measurement of a large number of telemetry data. In this paper, satellite battery temperature and battery voltage are used for experiments, and the anomaly detection effect is also good.

The telemetry data of the spacecraft battery is shown in Figure 8, where the red part is the abnormal section. In the test data, both the battery temperature and the battery voltage are faulty in the third cycle.

The anomaly detection results of the LUBE and KM-LUBE algorithms for spacecraft telemetry data are described in Figure 9 and Figure 10. It is clearly that both algorithms can detect the anomaly, but the interval coverage and interval width obtained by the KM-LUBE method are better than the LUBE method. Abnormal detection effect is also better.

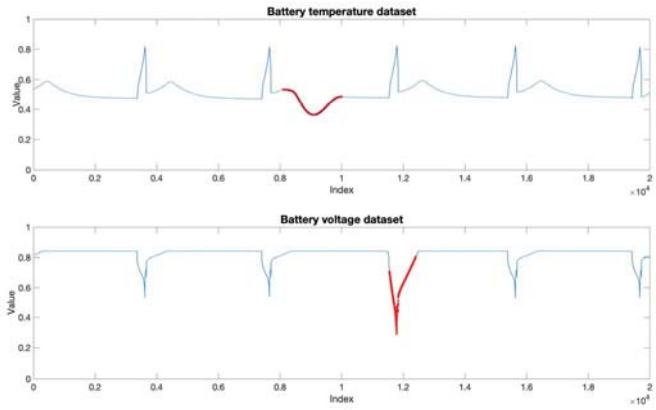

Figure 8. Spacecraft Battery telemetry dataset containing abnormal data

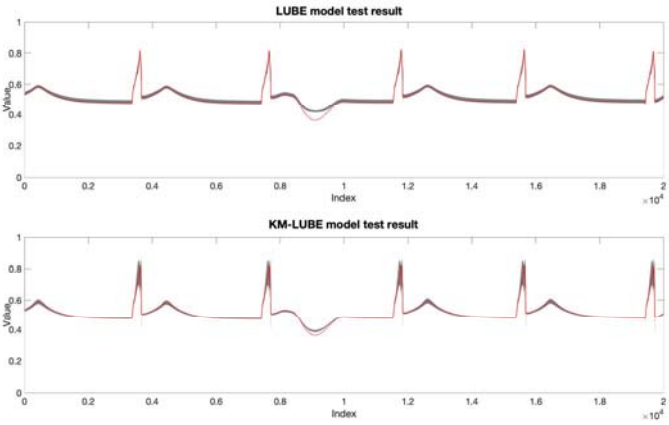

Figure 9. Performance comparison of detection algorithms of LUBE (top) and KM-LUBE (bottom) algorithms on Battery temperature data

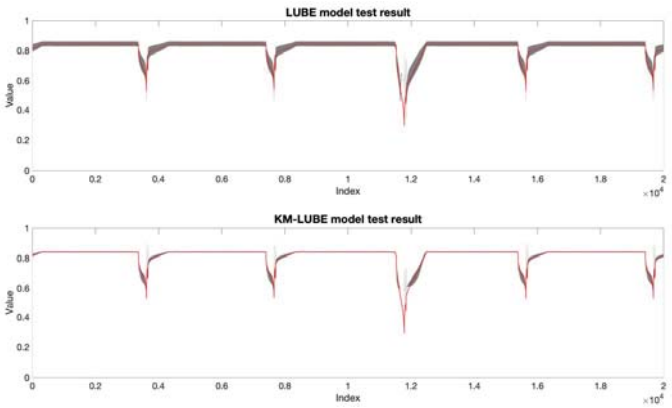

Figure 10. Performance comparison of detection algorithms of LUBE (top) KM-LUBE (bottom) algorithms on Battery voltage data

For all the above data sets, firstly, the effects of the two methods on the prediction interval are calculated. The PICP and NMPIW indicators of the two algorithms are calculated respectively, all results are described in detail in Table 2. It can be seen that the combination of the KM-LUBE algorithm on the two indicators is better than the traditional LUBE algorithm (the bold data in the table is the better algorithm under this index). For the evaluation of the anomaly detection effect, the previously proposed *DR*, *FAR* and *ACC* indicators are used to calculate the detection results to compare the performance of the two algorithms used on different data sets, and the indicators information are described in detail in Table 3. It is clearly seen that for anomaly detection problem, the KM-LUBE algorithm is superior to the traditional LUBE method in most indicators (the bold data in the table is the better algorithm under this index). It can be proved that the optimized KM-LUBE algorithm not only outperforms the original algorithm in predicting interval evaluation index, but also improves the performance of the algorithm in the anomaly detection problem.

TABLE II. COMPARISON OF ALGORITHM PREDICTION INTERVAL RESULTS

Algorithm / **Dataset**	**LUBE**		**KM-LUBE**	
	PICP	***NMPIW***	***PICP***	***NMPIW***
Ma_Data	0.811	**0.091**	**0.887**	0.118
ECG1	0.797	0.030	**0.858**	**0.018**
ECG2	0.786	0.050	**0.895**	**0.035**
Battery Temperature	0.954	0.024	**0.992**	**0.018**
Battery Voltage	0.993	0.040	0.993	**0.011**

TABLE III. COMPARISON OF ANOMALY DETECTION ALGORITHMS BASED ON IPE INDICATORS

Algorithm / **Dataset**	**LUBE**			**KM-LUBE**		
	DR	***FAR***	***ACC***	***DR***	***FAR***	***ACC***
Ma_Data	67.74%	5.69%	93.61%	**73.68%**	**4.51%**	**94.78%**
ECG 1	88.23%	8.63%	91.00%	**92.15%**	**5.28%**	**94.93%**
ECG 2	59.26%	6.85%	95.38%	**75.31%**	**3.24%**	**96.13%**
Battery Temperature	**58.42%**	5.08%	91.45%	55.42%	**0.86%**	**95.00%**
Battery Voltage	34.80%	**0.03%**	98.33%	**67.00%**	0.35%	**98.84%**

IV. Conclusion

This paper proposed a LUBE interval prediction anomaly detection model based on multi-objective optimization. The Knee point idea was introduced into the interval prediction. Compared with the traditional LUBE interval prediction model, the combination of the optimized interval width and interval coverage was obviously better than the original method. The method of eliminating the error of the model itself was introduced in the framework of anomaly detection, and the model uncertainty was combined with the interval prediction detection framework based on the improved k-fold cross-validation. By calculating the proposed IPE, more information can be provided than the traditional 0, 1 binary test results. Finally, the good performance of the proposed algorithm on the experiments by public data set and spacecraft telemetry data was proved.

In addition, some work should be carried out in the future to increase the accuracy and reliability of proposed algorithm in anomaly detection. First of all, some data sets targeted in this paper are mostly periodic and obvious data, so in the future work, the algorithm is further applied to aperiodic data to assess the performance. Secondly, the training of the model is carried out offline. For online real-time detection, the algorithm frame need to be implemented in time to improve its time adaptability of the model to consider the drift of the training pattern with time, thus enhancing the self-learning ability of the model.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under grants 71571187 and the Excellent Youth Foundation of Hunan Scientific Committee under grant 2017JJ1001.

References

[1] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Computing Surveys. vol. 41, No. 3, pp. 1–72, 2009.

[2] A. Patcha, J.M. Park, An overview of anomaly detection techniques: Existing solutions and latest technological trends, Computer Networks. vol. 51, No. 12, pp. 3448–3470, 2007.

[3] J. Ding, Y. Liu, L. Zhang, J. Wang, Y. Liu, An anomaly detection approach for multiple monitoring data series based on latent correlation probabilistic model, Applied Intelligence. vol. 44, No. 2, pp. 340-361, 2016.

[4] Chitrakar R, Chuanhe H. Anomaly Based Intrusion Detection Using Hybrid Learning Approach of Combining k-Medoids Clustering and Naïve Bayes Classification. 2012.

[5] M. Sheikhan, Z. Jadidi, Flow-based anomaly detection in high-speed links using modified GSA-optimized neural network, Neural Computing and Applications. vol. 24, pp. 599–611, 2014.

[6] Ergen T , Mirza A H , Kozat S S . Unsupervised and Semi-supervised Anomaly Detection with LSTM Neural Networks. 2017.

[7] Y. Zhang, J. Wang, and X. Wang, Review on probabilistic forecasting of wind power generation, Renewable & Sustainable Energy Reviews, vol. 32, pp. 255–270, 2014.

[8] Khosravi, A., Nahavandi, S., Creighton, D., Atiya, A.F., Lower upper bound estimation method for construction of neural network-based prediction intervals. IEEE Transactions on Neural Networks. vol. 22, No. 3, pp. 337–346, 2011.

[9] Khosravi A, Nahavandi S, Creighton D, et al. Comprehensive review of neural network-based prediction intervals and new advances. IEEE Transactions on Neural Networks, vol. 22, No. 9, pp. 1341-1356, 2011.

[10] Pang J , Liu D , Peng Y , and Peng X. Anomaly detection based on uncertainty fusion for univariate monitoring series. Measurement, vol. 95, pp. 280-292, 2017.

[11] Deb K, Agrawal S, Pratap A, and Meyarivan T. A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. Parallel Problem Solving From Nature. vol. 6, No. 6, pp. 849-858, 2000.

[12] E. Keogh, J. Lin, and A.Fu, https://www.cs.ucr.edu/~eamonn/discords/, 2005.