# Satellite Telemetry Anomaly Detection Based on Gradient Boosting Regression with Feature Selection

Zhidong Li[1(✉)], Bo Sun[1], Weihua Jin[2], Lei Zhang[1], and Rongzheng Luo[1]

[1] Beijing Institute of Spacecraft System Engineering, Beijing 100094, China
lizhidongcas@163.com
[2] Research Center of Satellite Technology,
Harbin Institute of Technology, Harbin 150080, China

**Abstract.** A data-driven satellite telemetry data anomaly detection method is proposed. The gradient boosting regression algorithm combined with feature selection, including feature scoring and recursive lowest-score feature elimination, can automatically mine the correlative telemetry variables through iterations and establish a nonlinear regression model for their functional association, which can be used as a health baseline for anomaly detection of telemetry data. This method requires no expert to specify correlative telemetry variables based on domain knowledge beforehand. It has the advantage of self-adaption for satellite operating conditions, which can overcome the problem of functional association altering under different operating conditions caused by orbit or sunshine condition changes. The validity and effectiveness of the method is verified by the telemetry data of the power subsystem.

**Keywords:** Anomaly detection · Satellite · Gradient Boosting · Feature selection

## 1 Introduction

The on-orbit satellite will generate a large number of telemetry data, which can reflect its on-orbit healthy state. Anomaly detection of telemetry data is important for the reliability and security of satellites.

With the development of AI and machine learning, data-driven methods are mainly used in anomaly detection nowadays, which include (1) one-class classification such as SVDD [1], OC-SVM [2], (2) statistical method such as Gaussian Mixture Model [3], (3) prediction based method such as LSTM network prediction [4], (4) supervised multi-class classification. For one-class classification and mixed Gaussian model, although it has strong universality, it has poor interpretability and unable to locate the cause of the problem further when an anomaly occurs. Prediction based method cannot handle the unpredictable telemetry variables. Moreover, due to the influence of conceptual drift, the prediction error will be accumulated gradually, affecting the accuracy of anomaly


Q. Wu et al. (Eds.): WiSATS 2020, LNICST 358, pp. 210–219, 2021.
https://doi.org/10.1007/978-3-030-69072-4_18

detection. For supervised classification, it is necessary to rely on experts to provide labels.

In engineering practice, single–variable-threshold method is often adopted. The upper and lower threshold values are set for a single telemetry variable. It is considered abnormal when the telemetry value exceeds the threshold. This method is simple and easy to use. Still it does not take into account the association between the various telemetry, nor does it take into account the impact of different satellite operating conditions on telemetries. For example, when telemetry violates certain constraints of consistency or does not conform to relevant rules in a particular context, it should also be judged as abnormal. However, the telemetry variables may not exceed their thresholds at the time. Therefore, multiple telemetry variables should be considered synergistically to achieve anomaly detection through joint analysis. When the telemetry variables deviate from their normal behavior pattern, also known as health baseline deviation, anomalies are considered to occur. When the satellite is in orbit, the change of orbit position and sunshine condition will lead to the satellite in different operating conditions, and the functional association between telemetry variables are usually different under different operating conditions. Multiple telemetry variables collaborative anomaly detection methods usually need to identify the operating conditions first, and then call the condition associated model to deal with the anomaly detection. However, the identification of operating conditions itself will introduce additional errors, and it is also difficult to achieve crisp segmentation of operating conditions during their transition. Moreover, expert knowledge like the functional hypothesis is usually needed when establishing a telemetry functional association model.

To solve the above problems, this paper uses a gradient boosting regression algorithm to automatically learn the nonlinear complex functional association model by data-driven manner, which is self-adaptive to the change of telemetry association altering caused by the change of operating conditions. Also, a novel feature selection method, which combines feature importance scoring through gradient boosting regression model training process and recursively lowest-score feature elimination, can automatically mine telemetry variables with the association through multiple iterations without domain knowledge. The method is characterized by high interpretability, high accuracy, and intelligence.

## 2 Problem Description

Data-driven anomaly detection requires first preparing training datasets, usually the historical telemetry data of satellites in orbit. Then, an anomaly detection model is learned from the training dataset through a data-driven manner. Finally, the trained anomaly detection model is used for anomaly detection for new on-orbit telemetry data.

A telemetry dataset containing N telemetry variables is represented as dataset = $\left\{\vec{X}_1, \vec{X}_2, \vec{X}_3, ..., \vec{X}_N\right\}$. $\vec{X}_k$ indicates the Kth telemetry variable, which is a vector with the timestamp and can be represented as $\vec{X}_k=[x_{k1}, x_{k2}, x_{k3}, ..., x_{kP}]^T$, and $x_{kp}$ indicates the value of the Kth telemetry variable at time p.

We propose a novel anomaly detection method that can automatically mine the telemetry with correlation from the dataset and establish the association model. Taking

$\vec{X}_1$ as an example, we first mine the telemetry variables correlated to $\vec{X}_1$, assuming that the results are $\vec{X}_2$ and $\vec{X}_5$. Then take $\vec{X}_1$ as the dependent variable and take $\vec{X}_2$ and $\vec{X}_5$ as the independent variable. The regression model $\vec{X}_1 = f_1^*\left(\vec{X}_2, \vec{X}_5\right)$ stands for the association between correlated variables can be established through learning from the training dataset.

When we performing the anomaly detection, the regression model $f_1^*(\cdot)$ is used to receive newly on-orbit telemetry data, and the expected value of $\mathrm{x}_{1\mathrm{p}}$ at certain time p is obtained. The expected value is represented as $\hat{\mathrm{x}}_{1\mathrm{p}} = f_1^*(\mathrm{x}_{2\mathrm{p}}, \mathrm{x}_{5\mathrm{p}})$. The regression error between the true value and the expected value of telemetry is represented as $\varepsilon_1 = \mathrm{x}_{1\mathrm{p}} - \hat{\mathrm{x}}_{1\mathrm{p}}$.

Anomaly detection criteria can be set by statistical methods or expert knowledge based on regression errors. For example, telemetry data may be considered abnormal if the regression errors exceed the limit one or more times in a given time period.

Similar to $\vec{X}_1$, anomaly detection can be applied to other telemetry data in the same way.

## 3 Anomaly Detection Algorithms

The training method of $f^*(\cdot)$ based on the gradient boosting regression algorithm is first introduced and then the feature selection method.

### 3.1 Gradient Boosting Regression

Gradient boosting regression algorithms is used to construct an association model $f^*(\cdot)$ through ensemble learning from the correlated telemetry variables. The algorithms take the regression tree as base model to jointly construct $f^*(\cdot)$. Assuming $f^*(\cdot)$ is composed of k regression tree models [5] $\{f_1, f_1, f_3, \ldots, f_\mathrm{K}\}$. In these tree structures, leaves represent continuous regression values and branches represent conjunctions of features that lead to those values. Overall regression value $\hat{y}_i$ equals to the sum of the regression results of all base models.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \tag{1}$$

$F$ is the functional space, $x_i$ is the input of the based model.

The objective function to be optimized is given by

$$\mathrm{obj} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \tag{2}$$

The left part $\sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right)$ is the training loss, and the right part $\sum_{i=1}^{t} \Omega(f_i)$ is the regularization term. $\Omega(\cdot)$ is model complexity of regression tree.

XGboost library is used for training the regression model. We give a brief review of XGboost training process in the following sections. People who are interested in XGboost can refer to [6].

Model training process of XGboost can be seen as additive training. In the training process, the new tree model is constantly being added. We write the prediction value at step $t$ as $\hat{y}_i^{(t)}$. Then we have

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{3}$$

$f_t(x_i)$ is the new tree model at step $t$. And the objective function at step $t$ becomes

$$\text{obj}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \sum_{i=1}^{t} \Omega(f_i) \tag{4}$$

Mean-Squared-Error (MSE) is used as the loss function, the objective becomes

$$\begin{aligned} \text{obj}^{(t)} &= \sum_{i=1}^{n} \left(y_i - \left(\hat{y}_i^{(t-1)} + f_t(x_i)\right)\right)^2 + \sum_{i=1}^{t} \Omega(f_i) \\ &= \sum_{i=1}^{n} \left[2\left(\hat{y}_i^{(t-1)} - y_i\right) f_t(x_i) + f_t(x_i)^2\right] + \sum_{i=1}^{t} \Omega(f_i) \end{aligned} \tag{5}$$

Taylor expansion of the loss function is taken up to the second-order, we obtain

$$\text{obj}^{(t)} = \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + 1/2 h_i f_t^2(x_i)\right] + \sum_{i=1}^{t} \Omega(f_i) \tag{6}$$

The $g_i$ and $h_i$ are defined as

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right) \tag{7}$$

$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right) \tag{8}$$

After all the constants are removed, the specific objective at step $t$ becomes $\sum_{i=1}^{n} \left[g_i f_t(x_i) + 1/2 h_i f_t^2(x_i)\right] + \Omega(f_t)$. This is the optimization goal for the new tree. The computation of $g_i$ and $h_i$ can refer to [6].

It can be seen that the objective function decreases gradually with the increase of the number of base models in the training process. An intuitive understanding is that each base model is good at different aspect and try to *complement* each other. Therefore, the method is self-adaptive to the operating conditions.

### 3.2 Feature Selection

Although gradient boosting regression can construct $f^*(\cdot)$, we must know which telemetry variables are correlated with each other beforehand. For example, we must know $\vec{X}_2$ and $\vec{X}_5$ correlated to $\vec{X}_1$, then we can construct an association model $\vec{X}_1 = f_1^*\left(\vec{X}_2, \vec{X}_5\right)$. This correlation usually provided by some domain experts. If we have no domain knowledge, we should mine the correlated telemetry variables first.

We proposed a novel feature selection method, which can automatically mine the correlated telemetry variables without domain knowledge. It includes (1) feature scoring and (2) recursively lowest-score feature elimination.

For feature scoring, the telemetry variables in the dataset are taken as features, and the feature importance degree can be scored through the training process of gradient boosting regression. Construction of gradient boosting regression tree model work top-down by choosing a feature at each step and optimized its split. The score of each feature can be defined as the number of times a feature is used to split the data across all trees. For example, after constructing $\vec{X}_1 = f_1^*\left(\vec{X}_2, \vec{X}_5\right)$, the score of feature $\vec{X}_2$ and $\vec{X}_5$ can be obtained.

**Table 1.** Computational process of Recursively feature elimination algorithm

| Recursively feature elimination algorithm | |
|---|---|
| step 1 | Take $\vec{X}_1$ as dependent variable. Take $\vec{X}_2, \vec{X}_3, ..., \vec{X}_N$ as independent variables.<br>Initialize $\vec{X}_1\_\text{correlated}$ as $\vec{X}_2, \vec{X}_3, ..., \vec{X}_N$. |
| step 2 | Learn $\vec{X}_1 = f_1^*\left(\vec{X}_1\_\text{correlated}\right)$ from the dataset.<br>Inclemently record $\vec{X}_1\_\text{correlated}$ and regression error as key-value pairs. |
| step 3 | Compute all the feature scores in $\vec{X}_1\_\text{correlated}$. |
| step 4 | Update : Eliminate the feature from $\vec{X}_1\_\text{correlated}$ that has the lowest score. |
| step 5 | If no feature in $\vec{X}_1\_\text{correlated}$, go to step 6, else return to step 2. |
| step 6 | According to all the key-value pairs, compute the best $\vec{X}_1\_\text{correlated}$ based on a utility function self-defined, which takes into account the regression error and the number of features. |

Feature scores combined with recursively lowest-score feature elimination can be used to mine correlated features. For example, if we want to mine the variables correlated to $\vec{X}_1$ from the dataset $\left\{\vec{X}_1, \vec{X}_2, \vec{X}_3, \ldots, \vec{X}_N\right\}$.The computational process of the recursively feature elimination is as shown in Table 1:

By using the utility function, we can obtain the best $\vec{X}_1$_correlated that the number of features in $\vec{X}_1$_correlated is smaller, and the regression error is low. If the regression error maintains large enough during the whole recursively feature elimination algorithm, then the dependent telemetry variable has no correlated variable.

Other features in the dataset can be applied feature selection to mine its correlated variables in the same way as $\vec{X}_1$.

## 4 Experimental Results and Analysis

The satellite power subsystem is taken as an example to verify the anomaly detection method. The application scope of the method is not limited to the power subsystem. For satellite, power subsystem is responsible for power supply to whole satellite, and it has great significance.

The subsystem is mainly composed of solar array panels, lithium battery pack, power conditioning controller (PCU), main bus, and so on. In the sunshine area, the solar array panels usually supply power for the whole satellite and charge the battery pack. Under the condition of the high-power loads, the solar array and battery pack are combined to supply the whole satellite. In the shadow area, the solar array panels have no output power, and the whole satellite is powered by the battery pack. PCU controls the operating mode of the subsystem through the internal main error voltage MEA feedback. The charging and discharging behavior of the battery pack are coordinated controlled by MEA and the battery error voltage BEA.

**Table 2.** Description of telemetry data

| Telemetry name | Distribution | Data type |
|---|---|---|
| MainBusVoltage | Main bus | Voltage |
| MainBusCurrent | Main bus | Current |
| −Y_solarCurrent | −Y_solarArray | Current |
| +Y_solarCurrent | +Y_solarArray | Current |
| −Y_BatVoltage | −Y_battery | Voltage |
| +Y_BatVoltage | +Y_battery | Voltage |
| −Y_charging | −Y_battery | Current |
| −Y_discharging | −Y_battery | Current |
| MEA | PCU | Voltage |
| −Y_BEA | PCU | Voltage |
| +Y_BEA | PCU | Voltage |

A set of key telemetry variables of the power subsystem is taken as the dataset, as shown in Table 2.

Power subsystem telemetry data curves are shown in Fig. 1. MainBusVoltage, +Y_solarCurrent, +Y_BatVoltage, −Y_BEA, and +Y_BEA are not shown in the figure because there are too many variables.

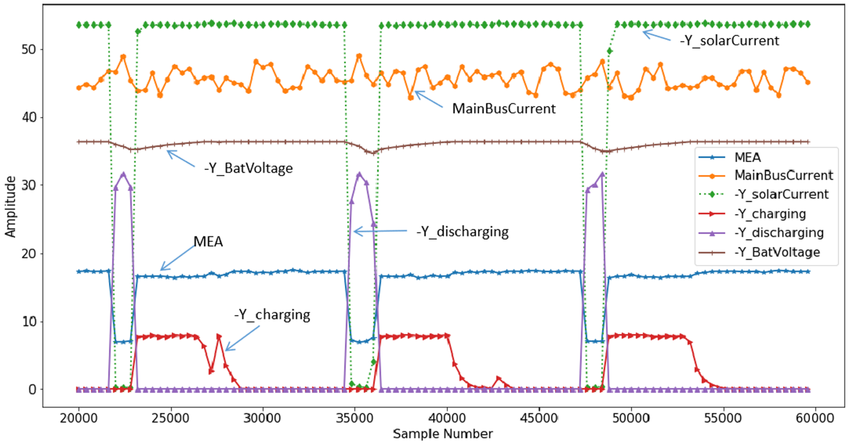


**Fig. 1.** Display of the satellite power subsystem telemetry data

From Fig. 1, it is shown that the telemetry curves show some correlation, and show different pattern under different operating conditions.

Take MEA as an example to verify the effectiveness of the method. Other telemetry variables can be handled in the same way.

MEA correlated telemetry variables are mined by the feature selection algorithm. The results are shown in Table 3. Due to a large number of process steps, the second and third feature elimination results were not shown.

From Table 3, it is shown that regression errors in the $9^{th}$ and $10^{th}$ elimination round is 0.3410 and 0.0350, which is much bigger than errors in the preceding round. Considering regression errors and the number of features, it is easy to know that the correlated telemetry variables to MEA are −Y_solarCurrent, MainBusCurrent and −Y_BEA. This feature selection result complies with expert's expectations. MEA and its correlated telemetry variables are shown in Fig. 2.

MEA is taken as dependent variables. −Y_solarCurrent、MainBusCurrent and −Y_BEA are taken as independent variables. The association model $f^*(\cdot)$ is constructed through gradient boosting regression algorithms learning from the training dataset. The true value and regression value of MEA are shown in Fig. 3.

It is shown that the regression value is very close to the real value, indicating that the regression model has high accuracy.

The regression error can be used as input of the anomaly criterion. The regression error is shown in Fig. 4.

**Table 3.** Results of feature selection

| Error feature /Score/ | 1st 0.0652 | 4th 0.0663 | 5th 0.0646 | 6th 0.0660 | 7th 0.0664 | 8th 0.0649 | 9th 0.3410 | 10th 0.035 |
|---|---|---|---|---|---|---|---|---|
| MainBusCurrent | 0.314 | 0.3370 | 0.337 | 0.338 | 0.336 | 0.363 | 0.430 | – |
| −Y_solarCurrent | 0.292 | 0.343 | 0.320 | 0.342 | 0.350 | 0.455 | 0.570 | 1 |
| −Y_discharging | 0.105 | – | – | – | – | – | – | – |
| −Y_BEA | 0.081 | 0.0959 | 0.090 | 0.117 | 0.199 | 0.182 | – | – |
| +Y_solarCurrent | 0.079 | 0.118 | 0.134 | 0.120 | 0.114 | – | – | – |
| +Y_BEA | 0.066 | 0.0559 | 0.073 | 0.183 | – | – | – | – |
| −Y_charging | 0.048 | 0.050 | 0.045 | – | – | – | – | – |
| −Y_BatVoltage | 0.007 | 0.00 | – | – | – | – | – | – |
| +Y_BatVoltage | 0.007 | – | – | – | – | – | – | – |
| MainBusVoltage | 0.00 | – | – | – | – | – | – | – |

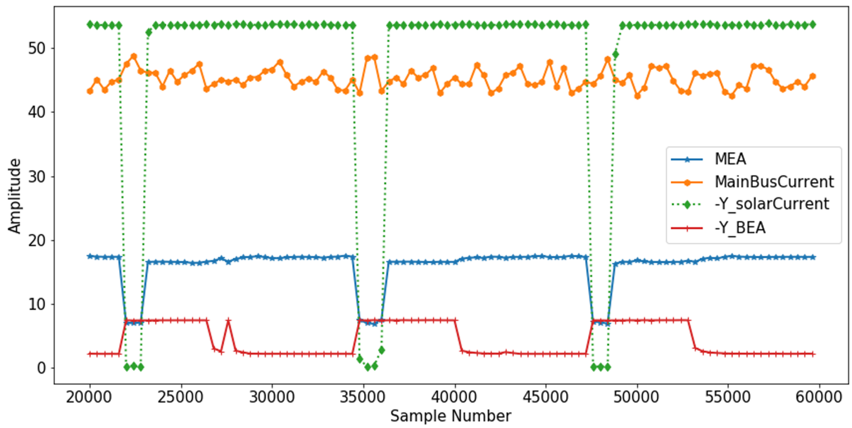


**Fig. 2.** Display of MEA and its correlated telemetry variables

From Fig. 4, It is shown that the regression error is small, and the threshold of anomaly criteria of regression error can be constructed by using statistical methods or based on expert experiences.

After artificial injecting some faults of MEA in the test dataset, anomaly detection result is shown in Fig. 5.

From Fig. 5, It is shown that the anomaly detection method proposed in this paper can detect anomalies effectively.

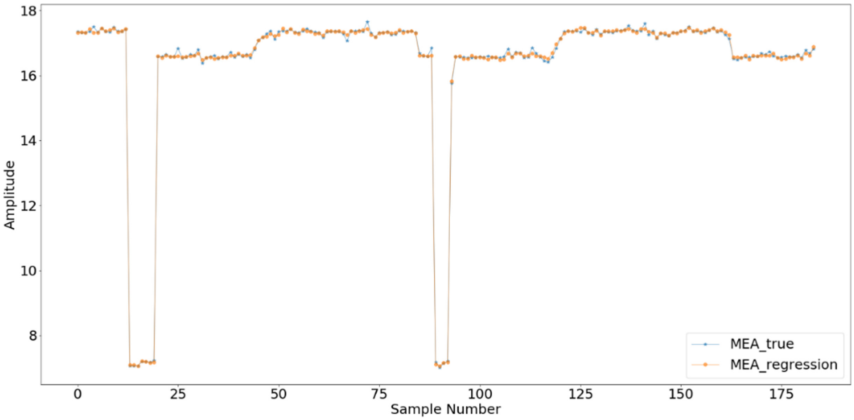


**Fig. 3.** True value and regression value of MEA

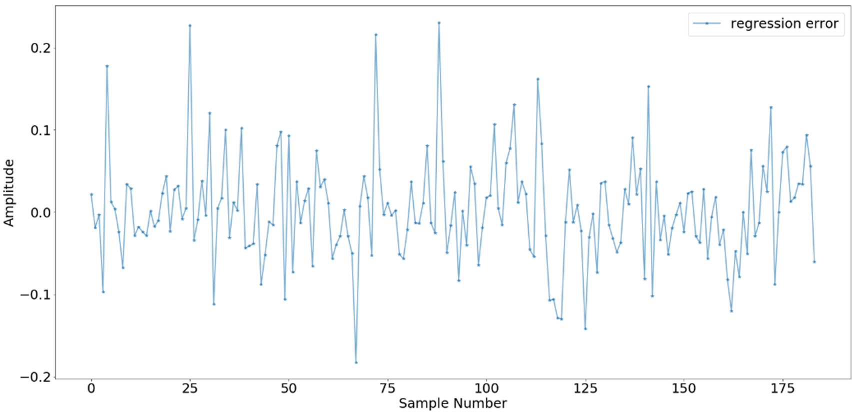


**Fig. 4.** Regression error of MEA

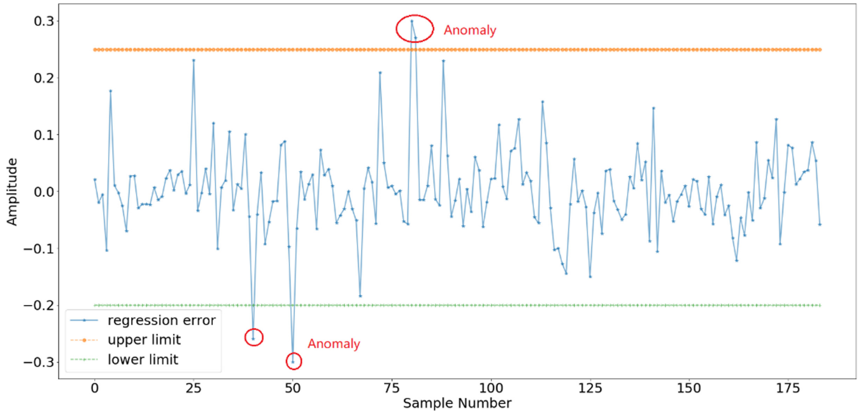


**Fig. 5.** Anomaly detection result

## 5 Conclusion

This paper proposes a new anomaly detection method. A gradient boosting regression algorithm is used to automatically learn the nonlinear complex functional association model by data-driven manner, which is self-adaptive to the change of telemetry association altering caused by the change of operating conditions. Also, a novel feature selection method, which combines feature importance scoring through gradient boosting regression model training process and recursively lowest-score feature elimination, can automatically mine telemetry variables with the association through multiple iterations without domain knowledge. The method is characterized by high interpretability, high accuracy, and intelligence. The validity and effectiveness of the method is verified by the telemetry data of the power subsystem.

## References

1. Zhao, Y., Wang, S., Xiao, F.: Pattern recognition-based chillers fault detection method using support vector data description (SVDD). Appl. Energy **112**, 1041–1048 (2013)
2. Zio, E.: A support vector machine integrated system for the classification of operation anomalies in nuclear components and systems. Reliab. Eng. Syst. Saf. **92**(5), 593–600 (2007)
3. Laxhammar, R., Falkman, G., Sviestins, E.: Anomaly detection in sea traffic-a comparison of the gaussian mixture model and the kernel density estimator. In: 2009 12th International Conference on Information Fusion, FUSION 2009. IEEE, pp. 756–763 (2009)
4. Malhotra, P., Vig, L., Shroff, G., et al.: Long short term memory networks for anomaly detection in time series. In: Proceedings. Presses universitaires de Louvain, p. 89 (2015)
5. Liaw, A., Wiener, M.: Classification and regression by randomForest. R News **2**(3), 18–22 (2002)
6. Chen, T., Guestrin, C., XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM (2016)