Deep Metric Learning for Abnormal Rotation Detection of Satellites from Irregularly Sampled Light Curve

Jun Yoshida[1], Ryosuke Togawa[2], and Taichiro Sano[3]

[1, 2, 3] *NEC Corporation*

j.yoshida@nec.com
togawa@nec.com
taichiro.sano@nec.com

ABSTRACT

In recent years, satellites have become an indispensable infrastructure in our lives. The number of satellites is increasing yearly and becoming increasingly active. To use satellites safely, it is crucial to manage them and detect the anomaly of satellites as much as possible. However, it currently takes skilled operators to detect an anomaly, and it is difficult for even skilled operators to detect the anomaly early without the telemetry data in cases such as an abnormal rotation. To address these challenges, we tested the feasibility of using deep metric learning for early anomaly detection from the irregularly sampled light curve. One of the characteristics of a light curve is unequally spaced because the optical sensor on the ground can only observe the subject at night and not when the weather is terrible. Given an irregularly sampled light curve, our model employs a long short-term memory (LSTM) unit of encoding the temporal dynamics and learns the embedding on the feature space using triplet loss. Then, an anomaly score is calculated based on pairwise distances between segments from the learned embedding in the feature space. With actual data from the satellite being operated, we showed the effectiveness of our model and the feasibility of early anomaly detection. Also, by exploring learned embedding in the feature space, we show that our model could capture the continuous state of the satellite.

1. INTRODUCTION

In recent years, the use of satellites has become increasingly active. It is becoming an indispensable part of our lives. As the use of satellites becomes increasingly active, operating costs are increasing. To use satellites safely and securely, it is necessary to manage their status of satellites. The telemetry data is typically used to identify the status of satellites. However, for some reason, it may not be available, in which case it is necessary to detect the anomaly of a satellite by other means. There was an event in which the HITOMI manufactured by NEC became out of control, and the investigation led to the belief that the cause of this event was abnormal rotation. Concerning this event, it was challenging to detect the anomaly of the abnormal rotation early when telemetry data was unavailable. In reality, such an event of abnormal rotation may occur for some reason. If it is impossible to utilize the telemetry data in real-time, a new way of anomaly detection that does not rely on telemetry data is needed. To achieve this, we focused on using a light curve as anomaly detection that does not rely on telemetry data. Some studies use a light curve for anomaly detection. However, these studies are based in a supervised manner, requiring a lot of labeled data for anomaly detection.

To address these issues, this paper examines the feasibility of the early abnormal rotation detection of satellites with Deep Metric Learning for multivariate time series retrieval. Metric Learning has been successfully used in various fields, including object identification, natural language processing, speech recognition, and medical diagnosis. Metric Learning aims to acquire the representation of data similarity on a distance basis and can perform similarity discriminations on unknown data. Furthermore, when combined with deep learning, metric learning has achieved remarkable results in recent years. In doing Deep Metric Learning, we employ triplet loss, which is found by Schroff, Kalenichenko, and Philbin (2015), the most popular loss function in deep metric learning. Given a raw multivariate time series segment, our model performs pseudo-labeling based on a distance of time series segment to create a dataset for triplet loss, encodes temporal dynamics using Long Short-Term Memory (LSTM) which is found by Sepp H. and Jürgen S. (1997), and uses triplet loss to feature segments of raw time series learning to embed on the feature space. In this way, the temporal dynamics in the raw time series segment are learned, and the feature space embedding can be obtained. The distance of the embedding on the feature space is then computed to measure the anomaly score of segments. Based on that measurement, the anomaly of a satellite is determined. Our model is the first unsupervised learning-based approach for early anomaly detection that can acquire temporal dynamics in the raw light curve and embedding on the feature space. To demonstrate the effectiveness of our model, we used a dataset of the irregularly sampled light curve of KIZUNA (NORAD ID: 35200). Communication with KIZUNA has been disrupted since February 8, 2019. After investigating the cause, it is believed that abnormal rotation is occurring. Telemetry data was unavailable when the cause was investigated, so the light curve was used to investigate. One of the characteristics of a light curve is unequally spaced because the optical sensor on the ground can only observe the subject at night and not when the weather is terrible. The irregularly sampled light curve of KIZUNA was observed on the ground when the abnormal rotation occurred.

2. Related work

2.1. Anomaly Detection of Satellites

There are existing technologies that utilize a variety of techniques ranging from statistical methods to machine learning and deep learning. In particular, machine learning and deep learning have made remarkable progress in recent years, and applied research using machine learning and deep learning has been increasing in the anomaly detection of a satellite. For example, some studies use telemetry and deep learning to conduct anomaly detection of satellites. Sara Abdelghafar, Ashraf Darwish, Aboul Ella Hassanien, Mohamed Yahia, Afaf Zaghrout (2019) show that Grey Wolf Optimization and an Extreme Learning Machine are used for the anomaly detection of satellites. However, it cannot be applied when telemetry data is unavailable for some reason (e.g., malfunction). Therefore, conducting early anomaly detection without the telemetry data is crucial.

Several studies use light curves to estimate the attitude of satellites, which differs slightly from our research objectives. However, some similarities exist in using a light curve: Gregory P Badura, Christopher R Valenta, and Brian Gunte r (2022) showed that the CNN model effectively estimates the attitude using a light curve. Gregory P. Badura, Christopher R. Valenta, Layne Churchill, Douglas A. Hope (2022) show the spin stability classification based on an irregularly sampled light curve. This is one of the closest to our research objectives. This is conducted using LSTM AutoEncoder for the classification of spin stability of satellites based on irregularly sampled light curves. These studies require large amounts of data with labels and unique features such as observation error, which limits the applicable environment. Nevertheless, these studies are slightly different from our objective of the early anomaly detection of a satellite using an irregularly sampled light curve. We explore a method that does not require a large amount of labeled data and unique features such as observation error for the early anomaly detection of a satellite.

3. Deep metric learning for anomaly detection

In this section, we explain our model based on deep metric learning in an unsupervised manner. We state the research objectives and then describe the architectural details of our model. Specifically, we explain how to acquire segment embedding using LSTM, and learn similarity based on metric learning using the features acquired by LSTM, the triplet loss function, conduct Pseudo-Labeling to segment data, and calculate an anomaly score.

3.1. Problem Statement

We introduce some main notions used in this paper. The description here is more general, assuming a multi-variate time series inputs, but includes the case of a univariate time series input. We denote a set of a multi-variate time series, k-th segment and n input series a time t as $S = \{\boldsymbol{s}^p\}_{p=1}^{N}$, $\boldsymbol{s}^k = \left(\boldsymbol{x}_{k_1}, \boldsymbol{x}_{k_2}, \boldsymbol{x}_{k_e}, \cdots, \boldsymbol{x}_{k_T}\right) \in R^{n\times T}$ and $\mathbf{x}_\mathrm{t} = (x_t^1, \mathrm{x}_\mathrm{t}^2, \cdots, x_t^n) \in R^n$, where T is the length of window size. Note that the interval in T do not have to be equally spaced. Given a multivariate time series segment $\boldsymbol{s}^q \in R^{n\times T}$, we aim to obtain the anomaly score of $\boldsymbol{s}^q$ as follows;

$$a^q = \sum_{m=1}^{k} \mathrm{TopK}\left(D(\boldsymbol{s}^p, \boldsymbol{s}^q)\right)_m,$$

where p denotes the index for p-th segment ($\forall p \in [1, N]$), N denotes the number of segments, $D(\cdot)$ represents a distance measure function, and $TopK(\cdot)$ represents a set of k items extracted in order of decreasing distance calculated by distance metric function.

3.2. Raw Segment Representation

Long Short-Term Memory (LSTM) is one of the most commonly used recurrent neural networks for processing time series data. It combines the previous hidden state and the current input to generate a new hidden state. Each LSTM layer has an architecture consisting of three gates: a Forget gate, an Input gate, and an Output gate. By using these gates, LSTM can capture both long-term and short-term dependencies. Specifically, the LSTM architecture is as follows.

$$\boldsymbol{f_t} = \sigma\left(\boldsymbol{W}_f[\boldsymbol{h}_{t-1}; \boldsymbol{x_t}] + \boldsymbol{b}_f\right), \tag{1}$$

$$\boldsymbol{i_t} = \sigma(\boldsymbol{W_i}[\boldsymbol{h}_{t-1}; \boldsymbol{x}_t] + \boldsymbol{b}_i), \tag{2}$$

$$\boldsymbol{o_t} = \sigma(\boldsymbol{W}_o[\boldsymbol{h}_{t-1}; \boldsymbol{x}_t] + \boldsymbol{b}_o), \tag{3}$$

$$\boldsymbol{c_t} = \boldsymbol{f_t} \cdot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \cdot \tanh(\boldsymbol{W}_c[\boldsymbol{h}_{t-1}; \boldsymbol{x}_t] + \boldsymbol{b}_c), \tag{4}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \cdot \tanh(\boldsymbol{c}_t), \tag{5}$$

where $[\boldsymbol{h}_{t-1}: \boldsymbol{x}_t] \in R^{m+n}$ is a concatenation of the previous hidden state and the input at time t. $\boldsymbol{W}_f, \boldsymbol{W}_i, \boldsymbol{W}_o, \boldsymbol{W}_c \in R^{m\times(m+n)}$ and $\boldsymbol{b}_f, \boldsymbol{b}_i, \boldsymbol{b}_o, \boldsymbol{b}_c \in R^m$ are learnable parameters. In this model, with segments as input, LSTM learns to map multivariate time series data to feature space. Then, the LSTM's final output $\boldsymbol{h}_t$, containing the information of all segments, is employed. These LSTM final outputs are fed into the linear layer as follows

$$\boldsymbol{s}_t = \boldsymbol{W}_s^T \boldsymbol{h}_t + \boldsymbol{b}, \tag{6}$$

This $\mathbf{s}_t$ is adopted as the representation of a segment because $\mathbf{s}_t$ is considered to encode temporal information.

3.3. Triplet Loss

Triplet Loss is a method for learning distances between data in the feature space using triples, which are tuples of three instances: a positive example, a negative example, and an anchor. By embedding the distance between anchors and positive examples closer and the distance between negative examples and anchors farther, the model trains the feature space, representing the similarity between data in terms of distance. The model is trained by minimizing the Triplet Loss on the training data and can represent the similarity of segments in the feature space concerning the distance. We use the triplet form, representing the relative similarity of segments. (e.g., in the case of class classification, $\mathbf{s}_a$ and $\mathbf{s}_p$ belong to the same class, and $\mathbf{s}_a$ and $\mathbf{s}_n$ belong to different classes.).The triplet loss is described as follows:

$$L_{triplet} = \max\left(0, \left\|\boldsymbol{s}_a - \boldsymbol{s}_p\right\|_1 - \|\boldsymbol{s}_a - \boldsymbol{s}_n\|_1 + m\right), \tag{7}$$

where $\|\cdot\|_1$ is used for the representation of l_1 norm, and m is the margin. Applying the above Triplet Loss to each triples allows similar segments to be embedded closer together and dissimilar ones farther apart. This allows the model to learn similarities that consider temporal information in the feature space.

3.4. Pseudo-Labeling

There is a labeling technique for data whose labels are unknown. Pseudo-labeling is a standard method of adding new labels to unlabeled data based on machine learning predictions. By using pseudo-labeling, it is possible to label many unlabeled data to increase the training data. Our model used pseudo-labeling based on Euclid distance, designed to learn the normal state from segments. Given two normal time series segments $\boldsymbol{s} = (\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots, \boldsymbol{x}_T), \boldsymbol{s}' = (\boldsymbol{x}'_1, \boldsymbol{x}'_2, \cdots, \boldsymbol{x}'_T)$,their Euclid distance can be calculated with:

$$d(\boldsymbol{s}, \boldsymbol{s}') = \sum_{t=0}^{T} \|\boldsymbol{x}_t - \boldsymbol{x}'_t\|, \boldsymbol{s}, \boldsymbol{s}' \in S, \boldsymbol{s} \neq \boldsymbol{s}', \tag{8}$$

where $S = \{\boldsymbol{s}_i\}_{i=1}^{N}$ is a set of segments. We can obtain the Euclidian distance vector between an i-th segment and all remaining segments from the above calculation. In order to take advantage of supervised learning, positive and negative segments are obtained from the top k similarities calculated based on the distance measure function entered with $\boldsymbol{s}^i$ as the query is defined as the similarity ranks in order of increasing similarity. Here, we denote k as the number of nearest neighbors we will extract as positive segments. Let S_{pos}^i denote the set of the extracted top k segments for an i-th segment, and the remaining segments S/S_{pos}^i are treated as negative segments. This process is performed across all segments to create a dataset with labels. Our model can be trained in a supervised manner using the created dataset. Note that our model leverages labels of this dataset in only a training phase.

3.5. Anomaly Score

There are various methods for calculating an anomaly score. We introduce a distance-based anomaly score calculation. This method is versatile and can be used in various cases. Given the segment representation of a light curve $\boldsymbol{s}$, the calculation of the anomaly score of an i-th segment is as follows:

$$a_i = \sum_{m=1}^{k} \underset{j}{\mathrm{TopK}} \left(\left\|\boldsymbol{s}_i - \boldsymbol{s}_j\right\|_1\right)_m, i \neq j, \forall j \in [1, n], \tag{9}$$

where n denotes the number of segments. Using the segment representation, we can calculate an anomaly score considering temporal information.

4. Experiment

4.1. Dataset

The data used in this experiment are the light curve of KIZUNA observed from optical sensors on the ground from January to February 2019. Figure 1 shows the irregularly sampled light curve of KIZUNA. One of the characteristics of a light curve is unequally spaced because the optical sensor on the ground can only observe the subject at night and not when the weather is terrible. In this experiment, since KIZUNA behaved abnormally on February 8, 2019, we define before and after February 8, 2019, as a standard and abnormal condition, respectively.

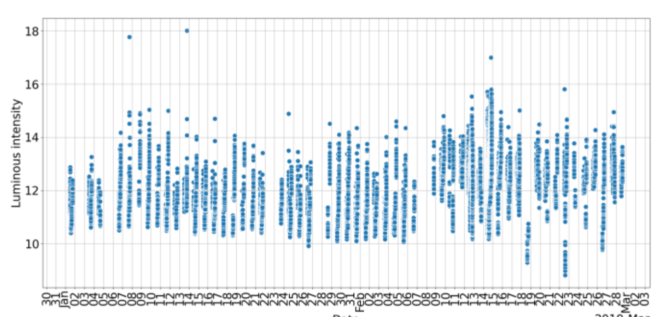

Figure 1. This figure shows the optical sensor observation data.

4.2. Parameter Setting

Our model has 5 hyper-parameters. We experimented with the hidden size of LSTM, the size of segment embedding, and window size of a segment set to 64. For the margin in triplet loss, we set it to 1. We set the number of nearest neighbors to 1000, which means the number of positive segments at each segment. Our model is trained on a server with NVIDIA GTX 2080 graphics cards.

4.3. Evaluation

In this experiment, the data before January 20 are trained as standard data. The anomaly scores after the above date are calculated; the period after the date includes both standard and suspected anomalous intervals. Figure 2 shows the plotting results. The anomaly scores calculated by the proposed method showed a clear upward trend in the anomaly scores when anomalies were suspected of having occurred. This result indicates the effectiveness of the proposed method in detecting anomalies. Furthermore, there is an increase in the trend of anomaly scores after January 24. This result suggests that the model could capture the anomaly before recognizing the abnormal behavior.

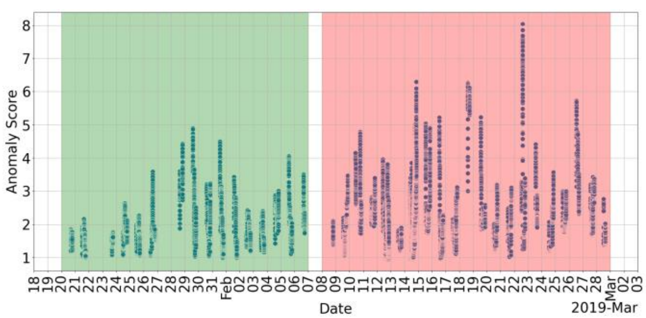

Figure 2. This figure shows the anomaly score. The green shade represents the interval of the normal state and the red shade represents the interval of the anomaly state of KIZUNA.

4.4. Explore The Segment Embedding

We explored the segment embedding learned by the model. Specifically, we used the t-SNE, which is found by Van der Maaten, Laurens, and Geoffrey Hinton (2008), to conduct dimensionality reduction of feature space. t-SNE is a widely used dimensionality reduction method characterized by its ability to project high-dimensional feature onto low-dimensional feature through a nonlinear transformation while retaining the characteristics of high-dimensional feature. Figure 3 illustrates how the segment embedding changed with the value of the anomaly score. This figure plots anomaly score values divided into intervals in increments of 1, with the colors becoming darker as the values increase. This result shows that relatively high and low anomaly scores form different clusters in the feature space, indicating that our model is likely to have captured the anomaly as a different state compared to the normal state. Also, the segment embedding is continuously changing, indicating that our model will likely capture continuous states of KIZUNA. This result indicates that our model can capture, such as between anomaly and routine, and detect the failure sign that would not be possible using telemetry data.

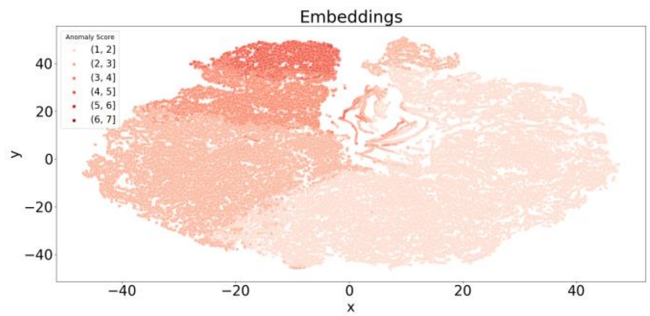

Figure 3. The t-SNE visualization of the learned embedding. The clear separation by the coloring indicates that the embedding represent continuous state changes of KIZUNA

5. CONCLUSION

We propose a new method for the early anomaly detection of a satellite from an irregularly sampled light curve. Given a light curve series segment, our model performs pseudo-labeling based on Euclid distance to create a training dataset and employ LSTM units to encode the temporal dynamics of light curve segments. Subsequently, the feature space embedding was learned by employing triplet loss. Finally, the distance of embedding on the feature space was employed to compute the anomaly score of a satellite. Empirical studies on the light curve dataset demonstrated the effectiveness and efficiency of our model. The future works include improvement in the accuracy of early anomaly detection.

REFERENCES

Dongjin Song, Ning Xia, Wei Cheng, Haifeng Chen, Dacheng Tao (2018). *Deep r-th Root of Rank Supervised Joint Binary Embedding for Multivariate Time Series Retrieval. KDD 2018.*

K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. (2014). *Learning phrase representations using RNN encoder-decoder for statistical machine translation.* arXiv:1406.1078 .

Sepp H. and Jürgen S. (1997). *Long short-term memory. Neural Computation* 9, 8 , 1735–1780.1

Gregory P Badura, Christopher R Valenta, and Brian Gunte r (2022). *Convolutional neural networks for inference* *of space object attitude status. The Journal of the Astronautical Sciences*, pages 1–34, 2022.

Gregory P. Badura, Christopher R. Valenta, Layne Churchill, Douglas A. Hope (2022). *Recurrent Neural Network Autoencoders for Spin Stability Classification of Irregularly Sampled Light Curves. (AMOS)* – www.amostech.com

Phan Dao, Kristen Haynes, Stephen Gregory, Jeffrey Hollon, Tamara Payne, and Kimberly Kinateder (2019). *Machine classification and sub-classification pipeline for geo light curves. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference, 2019.*

Sara Abdelghafar, Ashraf Darwish, Aboul Ella Hassanien, Mohamed Yahia, Afaf Zaghrout (2019). *Anomaly detection of satellite telemetry based on optimized extreme learning machine. Journal of Space Safety Engineering December 2019*, Pages 291-298

Schroff, Kalenichenko, and Philbin (2015). *"FaceNet: A Unified Embedding for Face Recognition and Clustering." CVPR 2015.*

Van der Maaten, Laurens, and Geoffrey Hinton (2008). *"Visualizing data using t-SNE." Journal of machine learning research* 9.11.