R&D project defense Jithin Sasikumar

Date: Friday, 26.03.2021
Time: 09:00
Room: Online

Title: Explaining Deep Learning Models for Detecting Anomalies in Time Series Data

Abstract:
The problem with many state-of-the-art machine learning models is the lack of transparency and interpretability, which is a major drawback in many applications. Explainable AI aims to address how the decisions of AI systems are made. It facilitates in understanding the decisions (or) predictions of a model. Anomalies refers to outliers (or) abnormal points. The process in which the anomalies are detected in the datasets, that deviates from the normal ones (deviation) is called anomaly detection. Spacecraft return a huge amount of telemetry data, so there is a huge requirement for enhanced anomaly detection systems that are needed to reduce the operational risk. This deviation occurs due to any potential hazards or unexpected behavior. When there is a failure to detect and respond to these deviations, it could jeopardize the entire spacecraft.

Anomaly detection is a crucial mechanism to caution about the unexpected behavior. In order to develop an efficient anomaly detection system, the explanation that specifies how the model classified its results as anomalies must be well understood. Long Short Term Memory (LSTM) networks which comes under the category of Recurrent Neural Networks (RNN) are used. The sequential data can be handled by LSTMs. Multiple LSTM networks with different model architectures are trained to detect anomalies from the satellite telemetry data (type of time-series data). The datasets used in this research are collected from NASA public repositories.

The most prominent methods for explaining the deep learning models are Layer-wise Relevance Propagation (LRP), Shapley Additive Explanations (SHAP), Deep Learning Important Features (DeepLIFT), Local Interpretable Model-Agnostic Explanations (LIME), Integrated Gradients (IG) and so on. These existing approaches (or) methods have already been applied for image data and text data but not for time-series data. This research focuses on comparing the existing approaches to explain the decisions of models trained using time-series data and the best fit method that generates explanations for a deep neural network will be proposed. The proposed approach is used specifically for explaining LSTM networks for anomaly detection task in time-series data (satellite telemetry data).