Final Project Proposal
March 20, 2024

Anomaly Detection with Convolutional Autoencoders

Description

Anomaly detection is a significant area in the field of data science, with broad applications ranging from fraud detection to spacecraft monitoring.[1] The ability to identify unusual patterns or outliers that do not conform to expected behavior is crucial across many sectors.

The primary objective of this project is to explore 5 temporal datasets for anomaly detection in spacecraft telemetry by using Convolutional Autoencoders (CAE) technique within the unsupervised machine learning realm.

In our analysis, we plan to implement Convolutional Neural Network (CNN) and Autoencoder algorithms – including backpropagation, applying down sampling operations and feature selections within encoder, fine tuning the hyperparameters such as stride, number of convolutional layers, and filter sizes. We will be comparing the performances of configurations of the neural network using metrics such as precision, recall, and F1-score.

Questions to answer

The primary question we are trying to answer to is how can convolutional layers be integrated into autoencoder architectures to enhance anomaly detection in spacecraft telemetry data? We also want to find out the optimal configurations of our algorithm for detecting anomalies within our spacecraft telemetry data. In efforts to answer this question, we will be using Convolutional Neural Network (CNN) within the frameworks of Autoencoders; particularly Convolutional Autoencoders (CAE).

Motivations for Choosing This Project

Ankur Baliga

I'm choosing to focus on unsupervised machine learning and anomaly detection because of its huge potential in real applications. This area is fascinating because it involves working with data that hasn't been labeled in advance. This means the challenge is to figure out from the data itself what's normal and what's not, which is important for finding problems in areas like online security or healthcare.

[1] https://en.wikipedia.org/wiki/Anomaly_detection

I'm doing this project as part of my unsupervised ML course. It's a perfect match for me because I'm eager to apply what I'm learning in class to actual problems. We have been comparing different algorithms, I'm more interested in how to use these techniques work in practice and goes on in the backend. I believe that understanding and applying anomaly detection can lead to real improvements in technology and society, particularly my interest in unsupervised algorithms.

Sayantan Datta

Anomaly detection is key in identifying outliers in data, crucial for areas like fraud detection and error identification. My project aims to examine and compare various anomaly detection algorithms across different datasets. This work is relevant to my goals as a Computer Science Master's student with a keen interest in data science. I got interested in this area after working with the DBSCAN algorithm in a class assignment. By evaluating different algorithms, I aim to gain a deeper understanding of their strengths, weaknesses, and applicability to various types of data. This exploration will not only bolster my analytical skills but also contribute to the broader conversation on enhancing anomaly detection methods for real-world applications.

Tenzin Kunsang

The first time I heard about the concept of anomaly detection in the context of machine learning was during my undergraduate in an "Introduction to Artificial Intelligence" coursework. Although it was mentioned in passing, I was intrigued by the technique because most of what I have learnt in ML thus far has been about not letting a model be impacted strongly by outliers – the outliers are always an afterthought. I believe anomaly detection attends to the samples that are considered inessential in realizing patterns – maybe there is no clear pattern. Therefore, I believe understanding and adopting various anomaly detection algorithms is crucial in not just deeming one algorithm better than the other – but to categorize how these algorithms are better in relation to the types of datasets provided. In the age of Artificial General Intelligence, I think it necessitates a deep understanding of how to interpret and act upon sparse or unlabeled data with unrecognizable patterns. With this project, I am curious to see how our unsupervised models behave differently based on our various feature configurations.

Division of Project Work:

We will all individually be reviewing multiple literatures specifically on CAE and CNN applications in anomaly detection in the spacecraft domain. We will regularly be meeting to touch base on our progress to discuss how we can improve on our feature configurations and in making selections of instances or attributes. Sayantan will work on understanding the datasets on a deeper level including exploratory analysis, data preprocessing, and performance tracking. Ankur will work data visualizations, evaluating metrics, and algorithm behavior analysis. Tenzin will work on algorithm implementation, feature engineering, and building the models. Most importantly, we will be working as a team as the datasets and each process are very closely related and will need be configured continuously so that we can build a model that detects anomalous instances with high accuracy and low error.

References

- Polson, Shawn. "Unsupervised Machine Learning for Spacecraft Anomaly Detection in WebTCAD." Laboratory for Atmospheric and Space Physics, June 7, 2019.
- Dataset used: Satellite data
- Wikipedia contributors. "Anomaly detection." Wikipedia, The Free Encyclopedia. Last modified March 19, 2024. https://en.wikipedia.org/wiki/Anomaly_detection.