GAD Project Outline

Here is the project outline and related resources for this project.

Table of Content

1. Background

We want to use GPS dataset to detect areas which may have jamming activity(Anomalies).

This GPS project will help the warfighter determine safe zones versus danger zones where there is a hotbed of jamming activity. The project aims to detect EMI (jamming/spoofing) on GPS/GNSS frequencies using the sensors available on the Pixel 2 phones, which were implemented using emerging Open Geospatial Consortium standards that are of interest to USSOCOM stakeholders.

2. Problems we facing

a. Explore the dataset. Make sure all schemas are understood.
b. Determine what kind of ML problem it is. If we have a labeled dataset and a unlabeled dataset, then it should be supervised ML problem, which should be binary or multiclass classification. Otherwise, we could only try clustering.
c. Determine model to use. For clustering, try K-mean the GMM; for classification, we should determine the model by dataset size. I suggest to start with a logistic regression model then digging deeper into tree-based models and neural network.

Other Problems listed in 5th Janurary:

a. Provide a machine learning model to predict anomalies (priority- due Feb 1st) - Hao & John
b. Perform Time Series Analysis on the data - John & Christina
c. Perform exploratory data analysis to determine static & trend indications of electromagnetic interference with GPS / GNSS satellite signals. CN0 and agc fields likely to be most relevant. - Christina

3. Dataset

3.1. Recourse

Here is the link to get access to database. This is a postgreSQL database so we should do ETL in pgadmin and download dataset for ML model training.

http://torgi-pgadmin.vmhost.devwerx.org/
Login: torgidocker@sofwerx.org
Password: dockertorgi

3.2 Database Description

See detail session downward

4. Build processing pipeline

Possible pipeline:

a. Import data
b. Data Explanation
c. Standardization, transformation
d. Feature Engineering
e. Model fitting
f. Evaluation

5. Model Comparison, optimization, and evaluation

Logistic-decision tree-RF-XGB-Ensemble

K-mean-GMM(or DBSCAN, EM or Mean Shift)

Evaluation metric for clustering:
https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation

6. Related Links

https://youtu.be/_rh4Nl0JSGk Time Series Clustering
https://www.gpsworld.com/detecting-false-signals-automatic-gain-control-12804/ interference and agc (Automatic Gain Control) levels
https://www.kaggle.com/albertmistu/detect-anomalies-using-gmm GMM for clustering
https://www.youtube.com/watch?v=ukzFI9rgwfU Overview of ML

3.2 Detail database description

Server: torgi-postgis

Sub-database: 4 databases, which are: gis, postgres, template_postgis, torgi. We found out that the torgi database have the majority of the dataset, gis also have 2 tables. Postgres and template_postgis are somehow meaningless.

Schema:

In gis: 2 schema, spatial_ref_sys have more informations

In torgi: Following tables seem to be more useful: gps_observation_points, motions, sat_data, and spartial_ref_sys. GPS observation should be the mostly important since that directly related to GPS data prediction.