Real-time Diagnosis of Anomalies in Deep Space Network Operations using Machine Learning

Umaa Rebbapragada
Jet Propulsion Laboratory, California Institute of Technology

Research described in this presentation was carried out at the Jet Propulsion Laboratory under contract with the National Aeronautics and Space Administration. Copyright 2021 California Institute of Technology. All Rights Reserved. US Government Support Acknowledged. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.
Acknowledgements / Contributions

- Rishi Verma\(^1\)
- Kyongsik (KS) Yun
- John Mason
- James Montgomery
- Lauren Klein (USC), summer intern 2020

\(^1\)Mission-Critical, Real-Time Fault-Detection for NASA’s Deep Space Network using Apache Kafka
https://videos.confluent.io/watch/xRmEDYJAhIVu56xcmMy6Yr
“NASA’s Deep Space Network (DSN) was established in December 1963 to provide a communications infrastructure for deep space missions”
A Global Enterprise by Necessity

- GEO
- 30,000 Km
- LEO (600 Km)

Goldstone Complex
Madrid Complex
Canberra Complex
A Diverse Set of Missions

Not all supported spacecraft depicted
Deep Space Communications

Science
- Radar
- Radio Science
- Radio Astronomy

Spacecraft Uplink
- Pointing accuracy
- Antenna efficiency
- Scheduling

Spacecraft

Hot Body Noise

Cosmic Background Noise

DSN Antenna
Terminology: Track

- From a data science perspective, a track is associated with a collection of time series from the beginning to end of DSN communication with the spacecraft that measure:
  - Attributes and strength of the signal
  - Attributes and telemetry associated with the spacecraft and DSN equipment
  - Weather conditions at the site

Multivariate Time Series
A single track has many time series associated with it, collectively referred to as monitor data
There are hundreds of monitor data items
Challenges

- DSN has pressure to reduce costs while maintaining quality of support to DSN mission users
- DSN complexes were once staffed 24/7 and each operator monitored a single track
- The Follow the Sun initiative (launched 2017) staffs a single complex during their daylight hours only to monitor tracks at all three complexes simultaneously

Operators supporting up to 4 simultaneously tracks

DSN has seen data rates increase over time

More missions, especially cubesats at lunar distances and beyond
Matching ML Capabilities with DSN Challenges

● Supervised learning
  ○ Classification
  ○ Regression

● Unsupervised learning
  ○ Clustering
  ○ Density Estimation
  ○ Anomaly Detection
Matching ML Capabilities with DSN Challenges

- Supervised learning
  - Classification
  - Regression

- Unsupervised learning
  - Clustering
  - Density Estimation
  - Anomaly Detection

We have historical data stored in a database at a slightly coarser sampling rate.
Matching ML Capabilities with DSN Challenges

- **Supervised learning**
  - Classification
  - Regression

- **Unsupervised learning**
  - Clustering
  - Density Estimation
  - Anomaly Detection

DSN has a system of reporting problems. Discrepancy Reports can provide some labels for classification. The data also contains actual measurements that can be compared to predicted ones. Thus, regression is possible.

We have historical data stored in a database at a slightly coarser sampling rate.
Developing Use Cases

- Team of mostly data scientists and developers
- DSN operations engineer (who formerly worked as an operator) assisting
- Developed four use cases based on feasibility of ML solution
  - supported by initial prototyping
- Surveyed DSN link control operators and operations engineers

1. Compare tracks across all monitor data items for similarity (multivariate time series similarity)
2. Identify anomalies within monitor data items given previous history
3. Classify whether a track is losing lock with the spacecraft due to bad weather and radio frequency interference (RFI)
4. Detect timing abnormalities in acquisition/loss of spacecraft signal
Survey Results

Based off of responses from 21 DSN staff from Madrid, Canberra, and California

Summary

Please best match the below metrics with potential tools described above

Activity Comparisons
Anomaly Detection
Predicting DRs
Profiling Early/Lateness
Classification based on Discrepancy Reports

- Top-ranked use case
- Most problematic due to issues with labeled data
- DSN to pursue a new tool that will improve label quality for ML

RFI incident reported but no clear evidence seen in data

Incident reported in yellow band, but voltage spike happened later
Input a **target track** of ~10 monitor data items (~0.2Hz)

- Identify tracks of the same type (i.e., spacecraft, antenna)
- Normalize time-series monitor data items
- Compute mean distance, ratio, corr coeff of tracks

**Machine learning model to output similarity score (0~1)**

Output **top 10 similar tracks**

**Statistically compare target track with “normal” track (outlier detection)**

Output **within track anomalies compared with normal track**

Input **two target tracks** (current or previous tracks)

- Normalize time-series monitor data items
- Monitor data time-series windowing (every 10 time points)
- **Compute mean distance, ratio, corr coeff of tracks**
- Machine learning model to output similarity score (0~1)

**Statistical comparison of two tracks**

Output **statistical differences between two tracks**

Show **top 10 similar tracks** of each target track
Track Comparisons - preliminary results (206 tracks: 52 similar & 154 dissimilar tracks)

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>60%</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>73.3%</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>73.3%</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>66.7%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>73.3%</td>
</tr>
<tr>
<td>KNN</td>
<td>73.3%</td>
</tr>
<tr>
<td>Ensemble Bagged Trees</td>
<td>80.0%</td>
</tr>
<tr>
<td>Neural Net</td>
<td>88.8%</td>
</tr>
</tbody>
</table>

AUC=0.96

![ROC Curve](image)

Confusion Matrix

- **True Positive Rate**
  - 94.2% 7.6% 9.9%
  - 9.6% 73.1% 88.8%
  - 26.9% 11.2% 19.1%

- **False Positive Rate**
  - 70.4% 6.8% 91.2%
  - 5.8% 18.4% 80.9%
  - 88.8% 8.8% 19.1%

- **Output Class**
  - 145 14
  - 9 38
  - 1 1
Roadmap to Integration

The below roadmap details a development plan based on survey results on utility of tools, maturity of current tool development, and available resources.

**Release 1:**
**Track Comparisons**
- Given two tracks, compare quantitatively

**Release 2:**
**Anomaly Detection**
- Highlight anomalous regions within track, given recent history

**Release 3:**
**DR lookup**
- Highlight DRs in historical tracks

**Release 4:**
**DR prediction**
- Predict DRs in current track

**FY 2021**
**FY 2022**
Conclusion

- DSN is preparing for major shifts in its operational paradigm
- Machine learning and automation will certainly play a role
- Build trust and value for users
  - Survey needs
  - Integrate with simpler products
- Data quality must be addressed as we shift to supervised models