Fault Impact Analysis

Towards Service-Oriented Network Operation & Maintenance by ITU Predict an NE's average data rate change when a fault occurs

Emelike Caleb caleb.emelike@gmail.com

INTRODUCTION

The goal of fault management in telecom O&M (Operation & Maintenance) is to ensure stable & reliable networks and services. In the RAN (Radio Access Network), the most significant part of O&M activities is network fault management, including fault monitoring, analysis, diagnosis, and repair processes. Among these processes, fault analysis is an essential part of troubleshooting.

In the current practices with large-scale and complex network structure, O&M engineers have to face massive faults & alarms in daily work. The traditional way to analyze faults is by setting rules based on network experts' experience, such as duration of faults or predefined categories of faults, to determine which faults need to be handled with higher priority.

However, with the traditional method, the impact of each fault on network KPIs (Key Performance Indicator), such as coverage and data rate, are not explicitly assessed and taken into consideration. As a result, the reliability of network service cannot be quantified and it is impossible to optimally schedule O&M resources. For example, even if some faults are so critical that NEs (Network Elements) are out of service because of them, they might not be urgent and could still be handled with lower priority. More precisely, consider a heterogeneous network where several NEs provide multiple layers of coverage (e.g., the co-existence of 4G and 5G, multiple frequency bands). In this case, even if a NE is out of service due to a sudden fault, there might not be coverage hole since users may migrate to neighboring NEs to access the network and obtain the same level of service quality. Therefore, the fault can be handled with low priority if other faults with worse impacts exist. In another case as stated in what follows, some faults need to be prioritized although they might only cause service capability deterioration of a NE rather than a service outage. When this kind of fault occurs, users can still access the network through the NE, but with a dropped service quality, which may lead to complaints and subscriber churn.

DATASET

A dataset with RAN KPI data, collected from over 100 5G and 4G NEs, will be provided to train machine-learning models. For each NE in the network, it provides 7 key KPIs, the duration of a fault, and the “distance” of the NE to the fault in an hourly basis. Specifically, each row includes:

- NE ID, which is a unique identifier for each NE;
- Hourly timestamp of the data;
- 6 RAN KPIs that is related to data rate, including Access Success Rate, Resource

Utilizing Rate, TA (Time Advanced), BLER (Block Error Rate), CQI (Channel Quality Indicator), MCS (Modulation and Coding Scheme);

- Data rate of the hour;
- Fault duration (in seconds) during the hour.
- “Distance” (Relation) of the NE to the fault, where
  If the value is 0, it means that there is no fault during the hour;
  If the value is 1, it means that the fault occurred exactly at the NE;
  If the value is between 0 and 1, it means that the fault occurred at a neighboring NE, and that a larger value indicates a higher adjacency between the 2 NEs.

DATA PROCESSING

The initial step in the analysis is data preprocessing, which involves collating data from multiple files and refining it to extract meaningful insights. Firstly, the data is loaded from numerous files present in a designated path. The files are meticulously examined to discard any where faults occurred throughout, ensuring a more precise analysis. Moreover, the 'NE ID' is renamed based on the file name, amalgamating the data into a cohesive unit ready for the subsequent steps of analysis. This stage is pivotal in setting the stage for a more detailed and focused feature engineering process.

FEATURE ENGINEERING

The feature engineering phase is a multifaceted process that involves creating new features from the existing data to enhance the model's predictive power. This process is carried out through a series of functions, each performing specific operations:

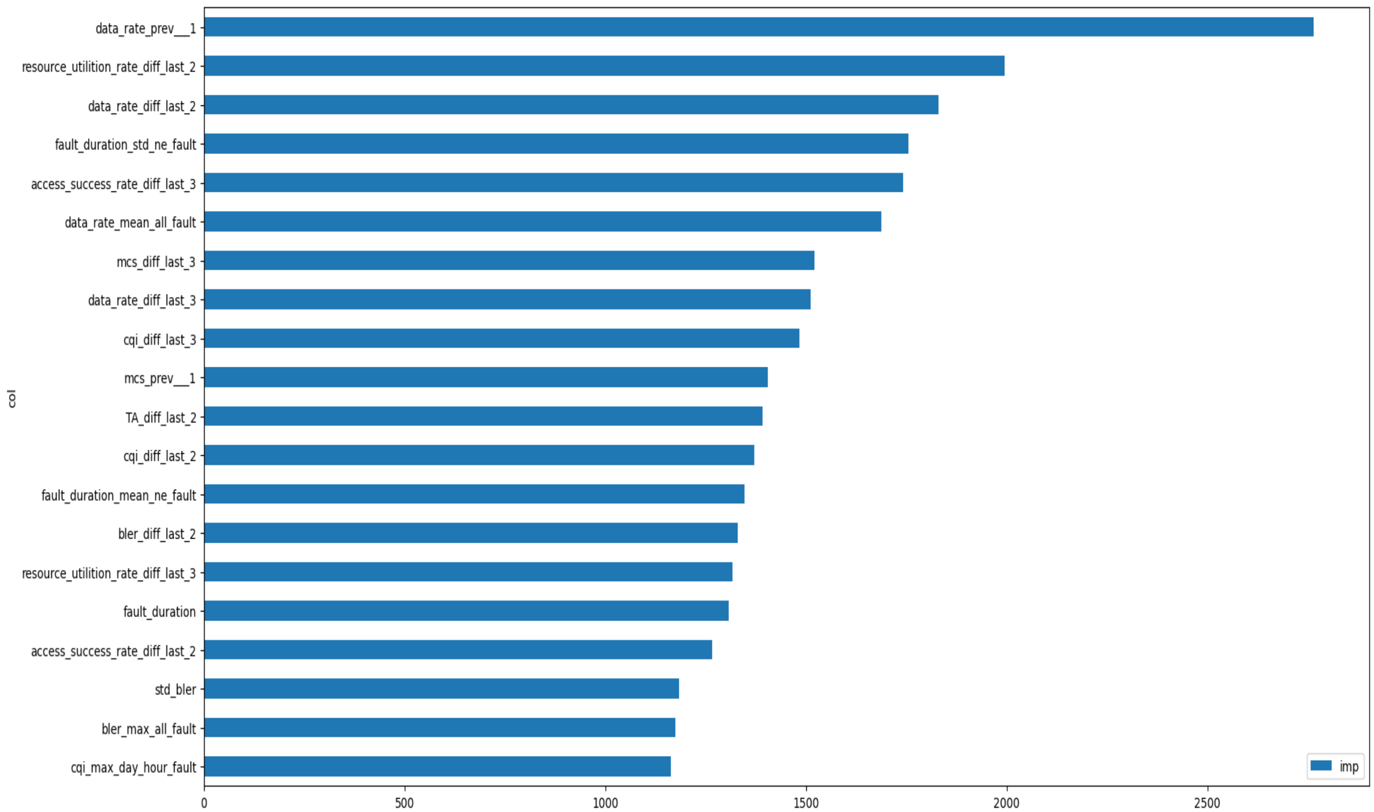
1. **Shift and Aggregate:** This function computes the differences in selected features based on specified shift values, followed by an aggregation process which calculates several statistical measures like sum, mean, standard deviation, maximum, and minimum values, for each hour.
2. **Aggregation Operations:** Different aggregation operations are carried out, including computing the mean, maximum, minimum, standard deviation, and skewness of selected columns when the fault duration is greater than zero.
3. **Extract Relevant Data:** This function is at the heart of the feature engineering process, orchestrating the extraction of relevant data and combining various computed features to form a comprehensive dataset. It involves intricate operations including grouping data based on various parameters, computing differences between consecutive rows, and aggregating data based on specific conditions.
4. **Focus on Fault:** A more focused analysis is carried out in this function where statistical measures like mean, standard deviation, maximum, and minimum values of various features are computed, concentrating specifically on instances where fault duration is zero. This function works in tandem with the extract relevant data function, bringing a sharper focus on fault analysis.
5. **Feature Computation:** A range of features are computed using different approaches, including computing differences between consecutive rows, calculating various statistical measures for groups of data, and merging various computed features to form a comprehensive dataset that serves as the foundation for the modeling process.

Together, these functions craft a rich feature set that is primed to feed into the modeling process, facilitating a more nuanced and insightful analysis.

MODELS

In the final phase of the analysis, a predictive model is developed using the LightGBM algorithm, a gradient boosting framework that uses tree-based learning algorithms. The model is fine-tuned using a StratifiedKFold strategy, which ensures that each fold is a good representative of the whole dataset, enhancing the model's predictive performance. The model excelled in identifying the fault impacts with a remarkable F1

Score of 0.7177355947433449, showcasing its efficacy in predicting fault impacts with high precision and reliability. Before are the feature importance of my Model.



CONCLUSION

This report delineates the comprehensive approach undertaken for fault impact analysis in telecommunications networks. Through meticulous data preprocessing and sophisticated feature engineering processes, a rich dataset was crafted, paving the way for insightful analysis. The deployment of the LightGBM model, fine-tuned with a StratifiedKFold strategy, culminated in a predictive tool with a noteworthy F1 Score, demonstrating its prowess in fault impact analysis. This endeavor lays a solid foundation for further explorations in enhancing the robustness and reliability of telecommunications networks.