LTE Spectrum Anomaly Detection using RNN models (LSTM)

Nihal Mehta
California State University, Sacramento
nihalnmehta@csus.edu

Sanket Modak
California State University, Sacramento
sanketmodak@csus.edu

Abstract

Managing spectrum in cellular networks is a tough task as the complexity in hardwares, configurations and new access technology for example LTE in IoT devices keeps growing. It is required that the wireless providers have a robust and flexible set of tools so that it is easier to check and detect faults that are observed in the physical spectrum, thus deploying it at scale. Anomaly detection for Long Term Evolution (LTE) spectrums of two carriers: AT&T and T-Mobile is implemented in this paper. We apply a deep Long Short-Term Memory (LSTM) model trained on 100 consecutive records of spectrums to predict 25 future spectrums.

Outliers are detected between the predicted and actual spectrum based on deviations of Root Mean Squared Error (RMSE). Overall, 121 anomalies for AT&T and 37 for T-Mobile are detected. This method assured that we could detect the errors in spectrum in definite time under the entire network. Anomalies of spectrum with noise can also be determined with the help of LSTM model.

Keywords: LTE, cellular networks, wireless providers, LSTM, anomaly detection, deep neural networks, spectrum, AT&T, T-Mobile.

1. Introduction

The ever-increasing complex modern cellular networks and continuous evolution has made the tasks of data analysts difficult. First, the signaling exchange between a mobile device and a cellular tower in LTE networks is so frequent that it prohibits use of naïve algorithms that do not scale well. Second, with the enormous complexity of LTE equipment, there is a demand for automated fault detection. For these reasons, cellular networks require a robust and flexible tool so that it can detect misbehavior efficiently in spectrum anomalies. One of the approaches to detect such faults relies on anomaly detection. Continuous evolution and increase of complexity of modern cellular networks creates numerous challenges for data analysts. First, the signaling exchange between a mobile device and a cellular tower in LTE networks is so frequent that it prohibits use of naïve algorithms that do not scale well. Second, with the enormous complexity of LTE equipment, there is a demand for automated fault detection. One of the approaches to detect such faults relies on anomaly detection. In this project, we apply deep LSTM model to two sample datasets with LTE spectrum records:

1. AT&T (880Mhz); 2. T-Mobile (729Mhz),

and detect anomalies based on the RMSE errors between the model's predictions and actual signal.

Cellular networks have been paying billions of dollars so that they can gain radio spectrum based on the capacity and coverage of the network. But spectrum management becomes complex and costly in terms of detection of spectrum interference faults as the ad hoc process generally involves the diagnosis to be manual followed by customer complaints and operational failure logs. Interference can originate from a range of complex sources anywhere at physical location, ranging from deliberate misuse of the spectrum and poorly configured transmitters to cable plant and connector RF leakage. Cellular IoT interfaces are rising and are tailored for IoT devices' network and energy needs. Such interfaces are introduced with the side effect of the security risks for LTE and surrounding spectrum bands.

Such difficulties lead to serious issues because the progress in spectrum usage and reconfigurable hardware has eventually made it simpler for the attackers to exploit spectrum without authorization. A manipulated system that operates on an attacker's behalf may do jamming or denial of service attacks on cellular bands. Cellular networks require powerful and effective tools to detect faults and spectrum misconduct, which we refer to as spectrum anomalies. This paper helps to examine the design of scalable systems that detects spectrum anomalies under wide-area LTE networks.

The report is organized in following manner: -

- Introduction
- Data
- Background: LSTM & Anomaly detection
- Methodology
- LSTM model
- Experimental Evaluation

- Related work
- Conclusion
- Future Work
- References

2. Data

In this project we have used two datasets representing spectrum of LTE networks of AT&T that operate at 880Mhz and T-Mobile working on 729Mhz. Each dataset contains over 25,000 records with 128 dimensions.

Signal strength in AT&T dataset ranges from -112.205 to -27.414 with the average of -39.17, while T-Mobile's signal ranges between -112.205 and -27.414 averaging at -38.86.

Samples of the spectrums are shown on Figure 1 (T-Mobile) and Figure 2 (AT&T).

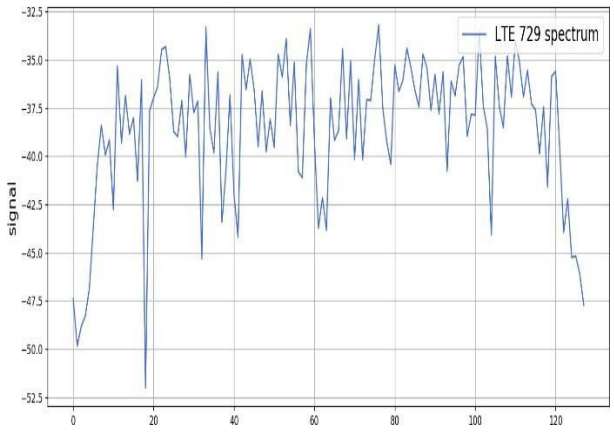

Figure 1. Sample spectrum collected from T-Mobile dataset

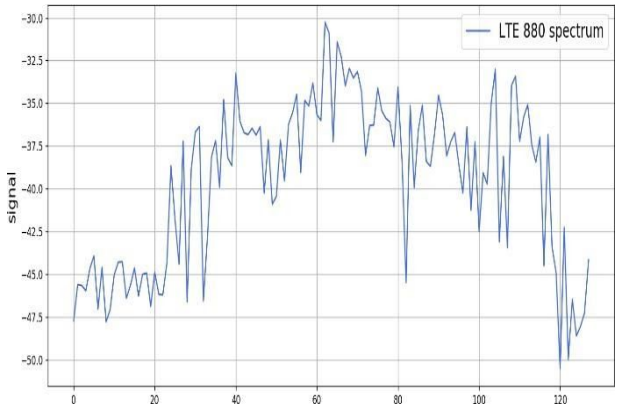

Figure 2. Sample spectrum collected from AT&T dataset

3. Background: LSTM & Anomaly detection

LSTM is an unique Recurrent Neural Network (RNN) type, well-known for acquiring detailed, complex sequential data-embedded patterns. In each RNN unit, an LSTM model maintains an inner state and consists of several stacked layers that create an architecture comparable to the neural feed-forward network. This allows for sequential data to be learned from complex relationships. Normally, at the end of the model, another completely connected layer is attached for classification or prediction.

A deep auto-encoder is a DNN model designed to learn about efficient data representation in an unsupervised way. This uses the Output Layer representations to reduce the input layer data into representations and then reconfigure the actual data. This technique forces the auto-encoder to delete the main useful features of the data.

The predictive models enable anomalies to be detected without knowledge of prior anomalies. Since each classifier is trained using standard spectrum data, the assumption is that this can not properly predict anomaly-containing information, resulting in large forecast errors that cause anomaly detection.

4. Methodology

Since the LTE spectrums are very dynamic, naïve approaches of detecting the outliers based on the Received Signal Strength (RSS) are not efficient. The key idea for our approach for automated anomaly detection is to compare the forecasting errors calculated with the LSTM model to see where the model's performance is the worst. Higher than regular errors would indicate the presence of outliers in the data used for training. We use a standard measure of forecasting errors – RMSE

$$RMSE = \sqrt{\overline{E(y - \hat{y})}},$$

where $\hat{y}$ stands for values predicted with the model, y stands for the actual observed values.

Our LSTM model was trained based on the input of 100 consecutive records of signal spectrum and the output of 25 future records of the spectrum, each record has 128 dimensions. To apply RMSE to such a high dimensional data, we flatten all 25 records of 128 values each into 3200 values and compare them with the flattened records from the datasets.

To detect outliers in the resulting errors we employed a simple heuristic – errors that more than 4 standard deviations above the mean are considered outliers. As a result, index of the RMSE error that is an outlier will correspond to the index of a spectrum record from the original dataset making it an outlier in the original dataset.

5. LSTM model

We used the following configuration for the deep LSTM model: 3 LSTM layers with 64 units and batch normalization and 1 dense layer with 3200 units for the output. The model accepts 100 consecutive 128-dimensional spectrum records and outputs 3200 values representing 25 future records of 128 dimensions each. Figure 3 shows the configuration of the deep LSTM model we have used.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 100, 64)	49408
batch_normalization_1 (Batch	(None, 100, 64)	256
lstm_2 (LSTM)	(None, 100, 64)	33024
batch_normalization_2 (Batch	(None, 100, 64)	256
lstm_3 (LSTM)	(None, 64)	33024
batch_normalization_3 (Batch	(None, 64)	256
dense_1 (Dense)	(None, 3200)	208000

Total params: 324,224
Trainable params: 323,840
Non-trainable params: 384

Figure 3. LSTM model parameters

6. Experimental Evaluation

We applied the deep LSTM model described in paragraph 4 to both datasets – AT&T and T-Mobile and recorded the RMSE errors. Since we were required to predict 25 future records of the spectrum, the overall number of overlapping records that we used was 24,900.

Figures 4 and 5 show the results of the outliers detection among the RMSE errors calculated for LSTM predictions on AT&T and T-Mobile datasets.

Following our approach, we have detected 121 outliers for AT&T dataset. Figure 6 demonstrates the 25 flattened spectrums of the AT&T dataset for which our deep LSTM model performed the worst in terms of RMSE which by our definition means the spectrums have an outlier.

Figure 7 demonstrates the forecasts of AT&T spectrum made with our model. Although the magnitudes of the forecast are slightly off, the model managed to capture the periodic patterns in the spectrum quite well. The average standardized RMSE error for the 25 flattened spectrums is 0.7

It is clear that the spectrum of the original dataset contains an outlier as detected with our approach. There is a "flat" line on the plot near the value #100 indicating something irregular happened.

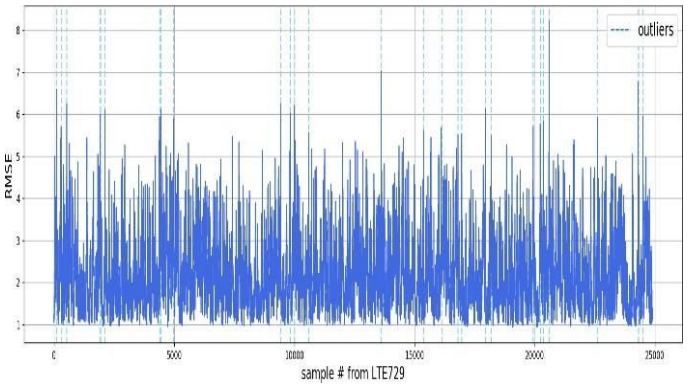

Figure 4. Outliers detected among RMSE errors in T-Mobile spectrum dataset

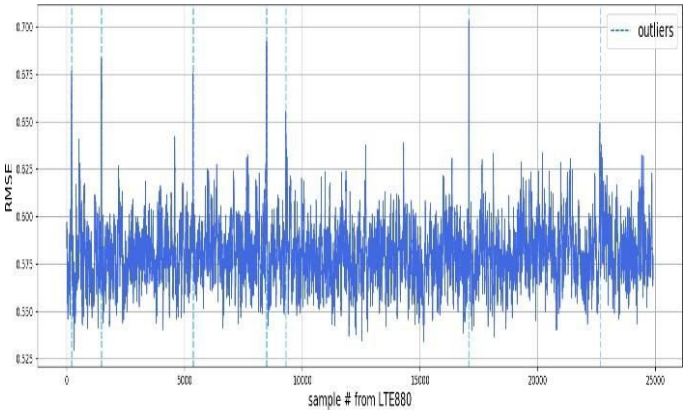

Figure 5. Outliers detected among RMSE errors in AT&T spectrum dataset

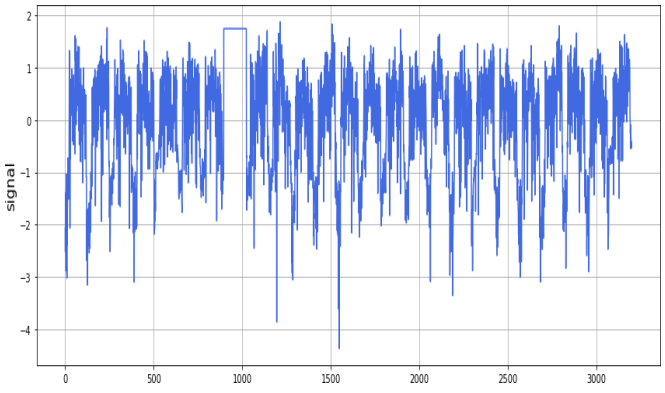

Figure 6. 25 consequent spectrums of AT&T dataset containing an outlier

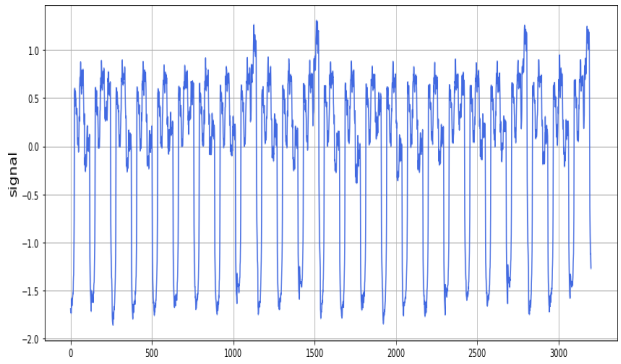

Figure 7. 25 consequent spectrums of AT&T dataset predicted by the model

Following our approach, we have detected 37 outliers for T-Mobile dataset. Figure 8 demonstrates the 25 flattened spectrums of the T-Mobile dataset for which our deep LSTM model performed the worst in terms of RMSE.

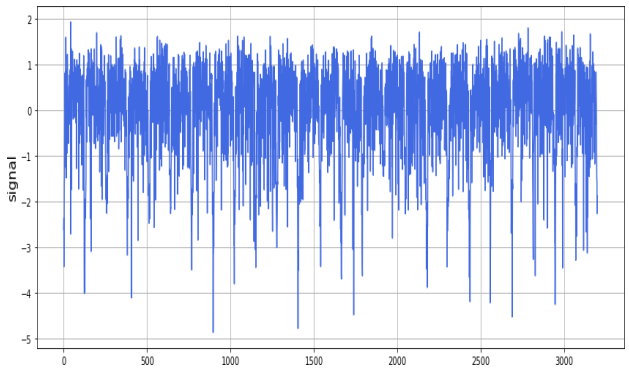

Figure 8. 25 consequent spectrums of T-Mobile dataset containing an outlier

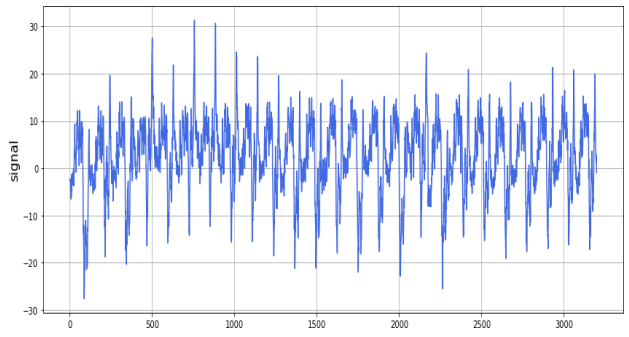

Figure 9. 25 consequent spectrums of T-Mobile dataset predicted by the model

Figure 9 demonstrates the forecasts of T-Mobile spectrum made with our model. Although the magnitudes of the forecast are slightly off, the model managed to capture the periodic patterns in the spectrum quite well. The average standardized RMSE error for the 25 flattened spectrums is 8.

Unlike in the case of AT&T spectrum where we can visually find a simple outlier, it is not clear why the largest error was obtained for this specific spectrum sample. It requires further analysis and investigation.

7. Related work

1. Crowdsourced wireless spectrum anomaly detection

As we know, Due to the large number of measurements manual and complexity of the spectrum use landscape, manual and fine-grained spectrum analysis is becoming tough day by day. For the detection of unexpected behaviors in the wireless spectrum from the collected data is the most important task of the automated monitoring. For enabling the interaction between users and automated systems, the most important function is control of detected anomalies. This system firstly analyzes the nature of these anomalies which are then converted from higher dimensional input data to a common feature space across sensors and to close the anomaly detection loop, developing the schemes for generalizing user feedback across sensors is helpful. Here, It aim to solve wireless spectrum anomaly detection problems by formulating it as a crowdsourced active learning problem and can focus on generalizing user feedback across different sensors and optimize anomaly detection based on user feedback from different users in a crowdsourced network.

2. Unsupervised anomaly detection

Different specific anomaly detectors are proposed in this study such as One-Class Support Vector Machine (OSVM) and Local Outlier Factor (LOF) which makes the common assumption that outliers are possible anomalies. These models normally perform poorly with increasing input data dimensionality. A Dynamic Spectrum Access (DSA) anomaly detector is presented, where distributed power measurements are performed. collective sensing

Anomaly detection used. For the specific case of DSA, the proposed detector is limited to allow user anomaly detection only. Similarly, it makes use of Hidden Markov Models (HMM) on probabilities of spectral amplitude that can detect interference on the DSA domain channel of interest again. The researchers used a model of Long Short- Term Memory (LSTM) to predict the next 4 IQ samples from the past 32 samples and detect an anomaly focused on the error in prediction. By the researcher, an adapted version of SAIFE is used for the initial unsupervised anomaly detection part of the proposed framework.

3. Semi-supervised anomaly detection

There are a few semi-supervised anomaly detection algorithms in literature, such as Lightweight on-line anomaly detector. The authors use an unregulated ensemble-based algorithm that is further enhanced by an active learning process. Uses a semi-supervised precision-at-the-top loss function to adjust LODA weights based on user input. The base anomaly detection algorithm is not limited to LODA and the authors also demonstrate that the active anomaly detection process can also be combined with tree-based algorithms. Recently,they proposed Semi-Supervised Detection of Outliers (SSDO), a semi-supervised anomaly scoring algorithm which relies on clustering and user labeling.

8. Conclusion

For network subsystem engineers and researchers, the constantly growing complexity of cellular networks poses new problems. Signal exchange requires more computing resources to execute, leading to greater network complexity.

As the range of LTE networks is very diverse in nature, It turns out to be expensive to evaluate naive methods that analyze basic variations in obtained signal intensity.

In order to address the disadvantage, we defined the outliers in this project based on the capacity of a prediction model that forecasts future spectrum based on historical observations. We have employed a deep LSTM model that takes as input 100 historical records and provides as output 25 potential records.

We have developed a simple heuristic for error outlier detection after measuring the model's RMSE errors-RMSE error is considered an outlier if it is 4 standard deviations above the mean error observed for the dataset. In total, in the AT&T dataset, we detected 121 outliers and in the T-Mobile results, 37 outliers.

We assumed that the deep LSTM model, which is actually a major demerit, was trained on the dataset without outliers. Without outliers in real life, it is not technically possible to collect data. This might eventually hamper our model's success.

9. Future work

In the follow-up work we are planning to direct our efforts to adding various spectral transformations, including Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT), and several others. We think that spectral transformations can help us to detect some invariant features of the normal signals that differentiate them from signals with outliers.

Another direction for future research is to apply different configurations of LSTM models – adjust number of hidden layers, number of units per layer, number of historical observations, and number of records that the model forecasts.

Yet another direction would employ clustering techniques like k-nearest neighbors or DBscan. The idea here would be to cluster the spectrum records in addition to fitting a deep LSTM model in order to compare its performance on similar spectrums and decrease errors caused by noise.

Finally, we are planning to apply our approach to other datasets representing cellular networks of other companies collected in other locations.

10. References

[1] Zhijing Li, Zhujun Xiao, Bolun Wang, Ben Y. Zhao and Haitao Zheng. Scaling Deep Learning Models for Spectrum Anomaly Detection. University of Chicago. +University of California, Santa Barbara

[2] spectrum_anomaly_detection.https://github.com/0x10cxR1/spectrum_anomaly_detection

[3] Sreeraj Rajendran , Wannes Meert , Vincent Lenders. Crowdsourced wireless spectrum anomaly detection by Sreeraj Rajendran. https://arxiv.org/pdf/1903.05408.pdf

[4] Sreeraj Rajendran , Wannes Meert , Vincent Lenders. Unsupervised Wireless Spectrum Anomaly Detection With Interpretable Features. https://lenders.ch/publications/journals/tccn19.pdf