Deep Learning Based Cell Outage Detection in Next Generation Networks

Madiha Jamil[1], Batool Hassan[1], Syed Shayaan Ahmed[1], Mukesh Kumar Maheshwari[1], Bharat Jyoti Ranjan Sahu[2]

[1] *Bahria University, Karachi Campus, Pakistan*
[2] *Siksha O Anusandhan University, Bhubaneswar,*
madihajamil970@gmail.com
batoolhassan960@gmail.com
shayaan_ahmed158@outlook.com
mukeshkumar.bukc@bahria.edu.pk
bharatjyotisahu@soa.ac.in

No Institute Given

Abstract. 5G and beyond wireless networks will support high data rate, seem-less connectivity and a massive number of users as compared to 4G network. It is also expected that the end to end latency in transferring data will also reduce significantly i.e 5G will support ultra low latency services. To provide the users with all these advantages, 5G utilises the Ultra Dense Networks (UDN) technique. UDN helps manage the explosive traffic data of users as multiple small cells are deployed in both indoor and outdoor areas, for seamless coverage. However, outage is difficult to detect in these small cells as these small cells have high density of users. To overcome this hindrance, Cell Outage Detection (COD) technique is utilised which aims to detect outage autonomously. This reduces maintenance cost and outages can be detected beforehand. In this paper, Long Short Term Memory (LSTM) is used for outage detection. The LSTM network is trained and tested on subscriber activities values which include SMS, Call and Internet activity. Our proposed LSTM model has classification accuracy of 85% and a FPR of 15.7303%.

Keywords: 5G, Cell Outage Detection, UDN, Call Detail Record, Deep Learning, LSTM

1 Introduction

5G network is expected to provide higher data rate, provide ultra-low latency and massive number of connect users. It is expected that Billion of devices will be connected to the network because of the 5G IoT ecosystem provided by the technology [1]. Ultra Dense Networks (UDN) is one of the emerging techniques of 5G that will help manage the explosive traffic of data. In UDN, multiple small cells i.e. femto, pico and microcells are combined to compensate large number

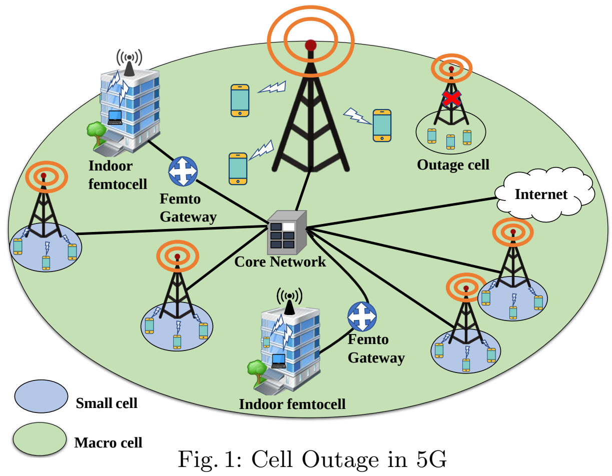

Fig. 1: Cell Outage in 5G

of users. Small cells are deployed in both indoor and outdoor areas to provide seamless coverage. These small cells encounter heavy traffic as connectivity has increased which enables thousands of users to be connected in a single cell [2].

Cell outage is encountered by users when they are unable to access the network. The probability of outage has increased in 5G network because of the immense traffic of users in a single cell and other reasons of outage can be: rain, snowfall, other environmental factors and software or hardware failure [3]. Cell Outage Detection (COD) is a method of detecting outage cells among healthy cells. The existing method of COD is through human monitoring [4]. It will be difficult to detect outage through this method because of small cells deployment and increase in users.

The new features of 5G come with a great deal of complexities and challenges. UDN allows large number of users to be accommodated in a single cell but also increases the probability of outage. 5G is a hybrid and integrated technology that combines existing technologies. The increasing number of base stations causes handover issues. Suppose number of base stations are available during handover, the network has to decide which base station the call has to be transferred. If calls of number of users is transferred to the same base station, the traffic data for the particular base station will increase and this results in outage [5]. Fig. 1 shows a case of outage in a small cell. The base station in outage cell is not able to receive network services from the core network.

The traditional way of COD is through visits to base stations, analysing users data and statistics and manual drive tests. These methods are costly and time consuming. The existing COD methods can impose difficulties for operators to detect outage in 5G. The densely situated small cells i.e. femto, pico and microcells makes it even harder for outage to be detected. Due to the complexity of these small cells, outage may not be detected for hours or even days [4]. This is the reason researchers are now working on deep learning algorithms that detect outage autonomously.

In previous researches, KPI parameters were used as input in [4] to train LSTM network for COD in multi-tiered networks. The authors used LSTM technique for COD and the algorithm had 77% classification accuracy. KPIs information was used as an input in RNN deep learning model for small cell outage prediction [6]. The model yielded 96% prediction accuracy. [7] and [8] utilised Autoencoder technique for outage detection. In [9], sleeping cells were detected in next generation networks and Deep Autoencoder and One Class SVM results were compared. [10] used Deep Convolutional Autoencoder to detect outage and obtained better accuracy than Deep Autoencoder. Hidden Markov Model was used in [11] for outage detection in UDN. The model predicted a cell as outage with 95% accuracy. K-Nearest Neighbour (KNN) classification algorithm was implemented in [12] for outage detection. Minimization of Drive Tests (MDT) measurements can not detect outage in small cells. [13] proposes a M-LOF algorithm that overcame this problem by using handover statistics for outage detection.

1.1 Motivations

Previous studies have mainly focused on working with KPI parameters for outage detection. The KPI parameters are not enough to accurately detect outage in small cells. Most of the studies also focused on detecting outage from readings of neighbouring cells. We were motivated to design an algorithm that uses real-time readings of users activities i.e. Call Detail Record (CDR) data of different cells.

Algorithm for Data Preprocessing

Inputs: CDR dataset: Data of 1000 Cell IDs of a single day
Output: Xtrain, Xnew
Method:
1. Import a file from CDR dataset and store randomly selected 1000 cell IDs data into a matrix.
2. Separate first 700 cell IDs for training of LSTM. Remaining 300 cell IDs will used for testing the trained model.
3. Remove the column containing country code and timestamp.
4. For each cell ID
   Sum each subscriber activity and store it into a matrix.
end
5. Store each matrix value as one example in Xtrain.
6. Repeat the same procedure for test data and store its value in Xnew.

1.2 Contributions

This research paper utilises the deep learning technique LSTM for outage detection. The neural network was trained and tested on a dataset that contained subscriber activities of different cells. The subscriber activities include SMS in, SMS out, Call in, Call out and Internet.

The rest of the paper is structured as follows: In Section II, we briefly discuss LSTM and its functionality. In Section III, the preprocessing of dataset and simulation results are discussed. Finally, the paper is concluded in Section IV.

2 Our Proposal

In this paper, we have used the deep learning technique Long Short Term Network (LSTM). The LSTM network has been previously utilised for outage detetction in [4]. LSTM is part of the Recurrent Neural Network (RNN). RNNs differ from other neural networks as they have the ability to remember their previous inputs. They use memory of previous input values to make decisions regarding the output value and current state. This makes RNN a useful tool in deep learning for time series data classification purpose [14]. We designed an algorithm that trains the neural network on subscriber activities of users i.e. CDR data of different cells [15]. The reason of using CDR data is because as [13] suggested that MDT measurements are not enough to detect outages in small cells. The UDN is densely populated with small cells and each cell has traffic of thousands of users. With an algorithm that is trained on CDR data it would be easier to detect anomaly in subscriber activities.

LSTM are especially used for time series data that requires memorisation of long term data. CDR data consists data of thousands of Cell IDs so it will be easier to train the dataset with LSTM since it can memorise the data pattern. The architecture of LSTM consists of gates i.e. the sigmoid layer and point wise multiplication operator, cell state, input gate, forget gate and output gate. The cell state represents memory of LSTM. The cell state changes when a memory is removed from the network through forget gate or when a new memory is added through input gate. The input gate decides which information is to be added to the memory. The forget gate discards the information that is not useful to the network. The output gate produces an output from the memory [16].

Hyperparameters	
No. of iteration/Epochs	1000
Learning Rate	0.001
Mini-batch size	5
Gradient Threshold	1
No. of Hidden Units	100

Table 1: Hyperparameters values utilized for ADAM

Cell ID	Time stamp (mili seconds)	Country Code	Subscriber Activities				
			SMS in	SMS out	Call in	Call out	Internet
10000	1385858400000	39	0.085994773	0.280452185	0.054231319	0.162693958	17.87857366
10000	1385859000000	39	0.085994773	0.054231319	0.208245313	0.434466179	18.11289499
10000	1385859600000	39	0.054231319	0.312215638	0.213019632	0.085994773	9.873021891
10000	1385860200000	39	0.004774318	0.171989546	0.004774318	0.085994773	13.39198912

*Every entry represent beginning of 10 minute interval in Unix epoch. For example, 1385858400000 represent Sunday, December 1, 2013 12:40:00 AM (GMT) [15].

Table 2: Sample of CDR dataset from 1st December,2013

Metric	Model testing	Random test*
Accuracy	85%	88.8000%
Error Rate	15%	11.2000%
Precision	78.9474%	87.0334%
Recall	89.0656%	90.5930%
FPR	15.7303%	12.9159%

*1000 Cell IDs of Day 2 were utilized for the test

Table 3: Performance Statistics of our COD model

LSTM network is trained in a supervised manner i.e. the inputs should be labelled and the output should be known. This is known an supervised learning. The dataset that we are utilising for training of LSTM network consists of CDR data which includes SMS, Calls and Internet activity. The LSTM network is trained in a manner where it remembers the pattern of user activities. The network can detect any anomaly that occurs in the values of user activities. For e.g. for a given cell if there is a spike in SMS activity than normal then the neural network will detect it. The network is trained in a similar manner for all the subscriber activities, hence any anomaly in the CDR data can be detected by the network. CDR data contains data of thousands of users, LSTM network is a useful deep learning technique to detect anomalies in the data. LSTM neural network as mentioned above is a powerful for time series classification. The CDR data contains data of Cell IDs at different time intervals. The LSTM deep learning network can be trained on this time series data and learn the pattern of the user activities. If outage is encountered by any cell, the value of CDR data will deviate from the normal values and this will be detected by the LSTM network.

3 Performance Analysis

The raw CDR dataset for training and testing of neural network was downloaded from [15]. Table I shows a sample of the CDR dataset. The dataset contained data of subscriber activities of different Cell IDs of 62 days. Each day consisted data of 10,000 Cell IDs. The dataset was in raw form so before utilising it for training of neural network, we preprocessed it. The preprocessing technique is similar to [17]. The preprocessing algorithm is summarized in Algorithm I [17]. In the preprocessing algorithm, the raw CDR data was imported and the data of 1000 Cell IDs of one day was extracted. The country code column was removed to ensure security. The data of each subscriber activity was summed and stored in matrices Xtrain and Xnew. The data was divided into training and testing with a 70% and 30% ratio i.e. 700 Cell IDs for training and 300 Cell IDs for testing.

As mentioned before, LSTM is based on supervised learning. It is necessary for the dataset to be labelled for training and testing. Supervised learning requires labeled input data and the output class should be known. Labels were created by manipulating the equation used in [17] since the data in CDR dataset was not labelled. The labels were created for Xtrain and Xnew matrices.

$$\|\mu - \sigma\| > \|\mu_x(i) - \sigma_x(i)\| \tag{1}$$

In equation (1), σ represents standard deviation and μ represents mean. Example *x(i)* is a single vector value of XTrain and Xnew. The corresponding output label *y(i)* is created for YTrain and Ynew respectively. Where i represents the ith row of the matrices. The label is marked outage if the norm of *x(i)* deviates more than the standard deviation ($\sigma \in R^5$) from mean ($\mu \in R^5$) otherwise healthy.

The training hyperparameters of the LSTM network are given in Table II. Since, the network is trained on the subscriber activities of users, the input size is set to 5. The neural network has a single LSTM layer with 100 hidden units. The mini-batch size is set to 5 with a learning rate of 0.001. ADAM optimizer was used for better training results. Fig. 2 demonstrates the training accuracy of the neural network and Fig. 3 demonstrates the training loss of the network. It can be seen that with every epoch the training accuracy increases and the training loss decreases. The model was tested and yielded an accuracy of 85% and a FPR of 15.7303%. To check the robustness and accuracy of our model we conducted random test on 1000 Cell IDs of other days. Table III: shows testing comparison of our model and random test conducted on 1000 Cell IDs of Day 2. It is noticed from Table III that that the random test yielded an accuracy of 88.8% with a FPR of 12.9159%. The error rate has also decreased from 15% to 11.2%.

Fig. 4 demonstrates the overall performance of the LSTM network. The precision and recall rate shows the accuracy of the testing model. The low error rate demonstrates how the network detects cells as outage or healthy with low chance of false alarm. The confusion matrix shown in Fig. 5 displays the performance of our testing model. Our model has an overall prediction accuracy of 85%. The tested data has 50% of healthy cells and 35% of outage cells.

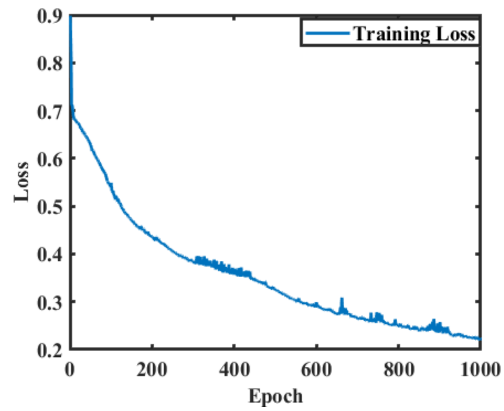

Fig. 2: Mini-Batch Loss

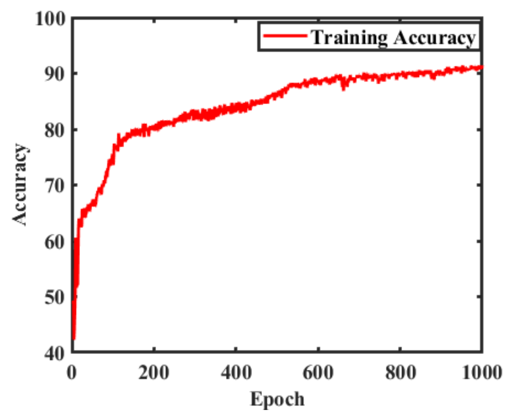

Fig. 3: Mini-Batch accuracy

4 Conclusion

In this research paper, we designed a deep learning algorithm using LSTM that classifies cells as healthy or outage. The model is trained and tested on subscriber activities values i.e. SMS in, SMS out, Call in, Call out and Internet activity. Previously, research studies have used KPI parameters for training of neural network which is a drawback as outage in small cells can not be detected through KPIs. The LSTM network is trained in such a manner where it can detect any anomaly in the values of subscriber activities. The model has a testing accuracy of 85% and FPR of 15.7303%. We also conducted random test on Cell IDs of day 2 which increased our testing accuracy from 85% to 88.8%. The model can be utilised in 5G network and next generation networks as it can predict outage in densely populated small cells. The classification accuracy can

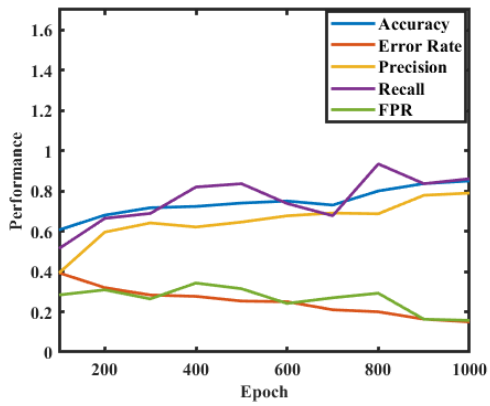

Fig. 4: Performance Metric

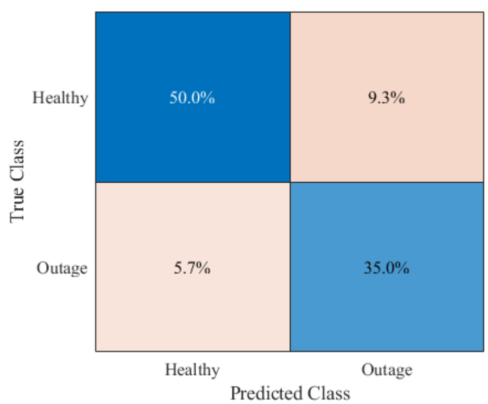

Fig. 5: Confusion Matrix

be further increased by working on larger datasets and considering subscriber activities of more users.

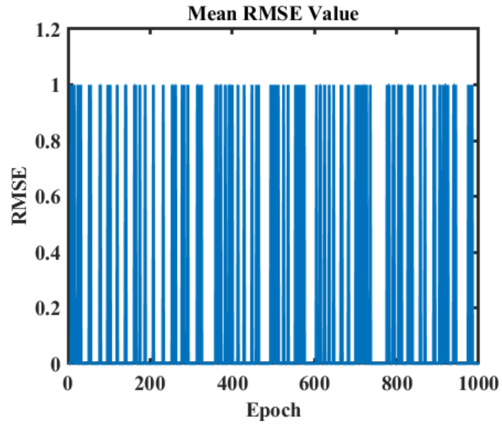

Fig. 6: Root mean Square Error

References

1. Agiwal M., et al.: "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Commun. Surveys Tuts.,* 18.3 (2016): 1617-1655.
2. 5G Ultra Dense Networks (5G-UDN). [Online]. Available: `https://icc2018.ieee-icc.org/`
3. The Top 5 Causes of Network Outages. [Online]. Available: `https://segron.com/`
4. Oğuz H. T. , Kalaycioğlu A., Akbulut A.:"Femtocell Outage Detection in Multi-Tiered Networks using LSTM," *2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Pitesti, Romania, 2019, pp. 1-5.
5. Handover Problems and Solutions for 5G Communication Technology. [Online]. Available: `https://www.ukessays.com/`
6. Ming Y.W., Lin Y.H., Tseng T.H., Hsu C.M.: "A Small Cell Outage Prediction Method Based on RNN Model." *Journal of Computers Vol. 30 No. 5*, 2019, pp. 268-278.
7. Lin, P.C.: "Large-Scale and High-Dimensional Cell Outage Detection in 5G Self-Organizing Networks." In 2019 *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 8-12. IEEE, 2019.
8. Asghar, M. Z., Mudassar A., Khaula Z., Pyry K., Timo H.: "Assessment of deep learning methodology for self-organizing 5g networks." *Applied Sciences* 9, no. 15 (2019): 2975.
9. Masood U., Asghar A., A. Imran A., Mian A. N.: "Deep Learning Based Detection of Sleeping Cells in Next Generation Cellular Networks," *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, 2018, pp. 206-212.
10. Ping, Y.H., Lin p. C.: "Cell Outage Detection using Deep Convolutional Autoencoder in Mobile Communication Networks." In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1557-1560., 2020.
11. Alias m., Saxena N., Roy A.: "Efficient Cell Outage Detection in 5G HetNets Using Hidden Markov Model," in *IEEE Communications Letters*, 20 (3), 62-565, March 2016.

12. XueW. W. , Peng M., Ma Y., Zhang H.: "Classification-based approach for cell outage detection in self-healing heterogeneous networks," 2014 *IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, Turkey, 2014. 2822-2826.
13. Zhang, T., Lei F., Peng Y., Shaoyong G., Wenjing L., Xuesong Q.: "A handover statistics based approach for cell outage detection in self-organized heterogeneous networks." In 2017 *IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pp. 628-631., 2017.
14. Understanding LSTM Networks. [Online]. Available: `https://colah.github.io/`
15. Telecommunications- SMS, Call, Internet- MI. [Online]. Available: `https://dataverse.harvard.edu/`
16. Basic understanding of LSTM. [Online]. Available: `https://blog.goodaudience.com/`
17. Hussain B., Du Q., Zhang S., Imran A., Imran M. A.: "Mobile Edge Computing-Based Data-Driven Deep Learning Framework for Anomaly Detection," in *IEEE Access*, vol. 7, pp. 137656-137667, 2019.