Using Visual Analytics to Generate Salient Features for Neural Network based Outage Prediction

Ayush Kumar
Stony Brook University

Russell Fischer
AT&T Labs Inc.

Jim Strydio
AT&T Labs Inc.

Prafulla Verma
AT&T Labs Inc.

Klaus Mueller
Stony Brook University

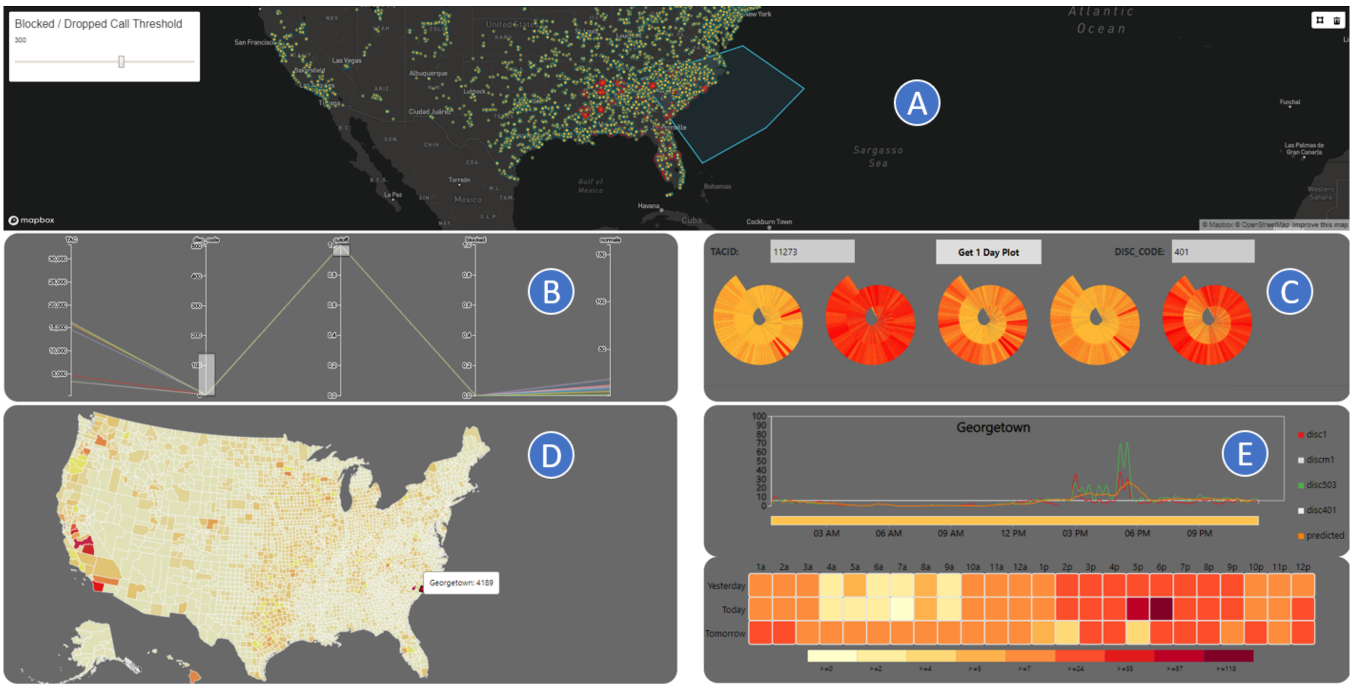

Figure 1: Overview of visual analytics interface with all major visualizations components. (A) All the located cellular towers are mapped on the US Map with an additional marker widget to select region of interest as shown in blue color polygon. Data associated to the region are transferred to Parallel Coordinate plot in (B). Parallel coordinate is used to analyse the selected cellular towers to find out operationally relevant performance indicators which are visualized in (C) using Spiral Chart. Relevant performance indicators are then used as salient features to feed neural network for outage prediction. Output of neural network model is aggregated over county level which is visualized as Choropleth map as shown in (D), and is linked to (E) for detailed exploration.

ABSTRACT

Providing high quality network performance with growing complexity in modern cellular network is challenging. Such network generates millions of voice data per minute, which is equally challenging to analyse for anomalous behavior due to predominantly normal performances. Therefore, its important to find out the relevant performance indicator of cellular networks to detect the small proportion of anomalous behavior from the pool of large generated data. In this paper, we present a visual analytic tool that enable us to find out the relevant performance indicator of cellular network data. We discovered that using this visual analytic tool to find out these indicators in timely manner is very significant from operation perspective. Indicators found during analysis are fed into the proposed neural network model as salient features to predict the unplanned network outages. We used the outages record from the past to train a Long Short Term Memory based neural network predictor. The predicted outages from various cellular towers are aggregated

XYZ, , USA

ACM ISBN 978-1-4503-11111111…$15.00
https://doi.org/10.1145/11111111

over county level and are visualized in the form of choropleth map. For detailed analysis of county's aggregated network performance in past, and future, we have used line chart and calendar chart respectively.

CCS CONCEPTS

• **Human-centered computing** → **Visualization**; • **Computing methodologies** → **Machine learning algorithms**.

KEYWORDS

Visual Analytics, Visualization, Neural Network, Time Series Prediction, Cellular Data

ACM Reference Format:
Ayush Kumar, Russell Fischer, Jim Strydio, Prafulla Verma, and Klaus Mueller. 2019. Using Visual Analytics to Generate Salient Features for Neural Network based Outage Prediction. In *2019 Symposium, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/11111111

1 INTRODUCTION

There has been high demand of uninterrupted services due to increasing mobile usages and fierce competition. In order to have high throughput and lesser operational expenditure to improve the efficiency of cellular network maintenance it's very much needed to have a tool that could help to solve this problem in timely manner. The main challenge with operation team is the difficulty of analysing large volumes of data and extracting features to find the outages in timely manner before it actually happens. Although more research is being conducted into ways of dealing with these problems for predicting the outages, only few researchers focus on visual analytic approaches in identifying salient features needed for prediction. There is already great demand of scientist to deploy predictive models from huge amount of data that can help domain experts to have better insight into it and have actionable plan. However, using machine learning models to deploy such predictive capability demands fine tuned features to perform better. Also not all features which are part of massive amount of data are of interest to the experts and are fed into the model to unecessary increase the complexity of the system. Researchers in past have shown that visual analysis can help to find the relevant features in time manner and reduces the complexity of the predictive model as well. In this paper, we have introduced a visual analytic tool using multiple visualizations to represent the massive amount of data in simpler way. Our analytic tool significantly helps the experts to identify the relevant performance indicator and feed those in the predictive model to find out the outages before it actually happens.

The rest of the paper is organized as follows: Section 2 discusses the exploratory analysis part, where we use multiple visualization to visualize both the spatial and temporal component of the dataset for selection of relevant features. Then, Section 3 discusses the proposed predictive neural network model to forecast the outages as shown in Fig. 1(D) & (E) on the basis of analysed salient features with the help of visualization shown in Fig. 1(A), (B) & (C) respectively. Section 4 discuss about the we have got in combination of both visual analytic too and neural network model and is visualized in the form of Line Chart in Fig. 1(E). Finally, Section 5 discusses the future work and gives concluding remarks.

2 DATA & VISUAL ANALYTICS

In this section, we discuss about the dataset generated from cellular networks and the characteristics of normal and anomalous call behaviour in the network. We look into the call record for each cellular phone call that includes the information about the call, the sequence of cell towers, calls that are dropped or blocked due to network or other unplanned reasons. We call this information as Call Detail Records (CDRs) provided by a large U.S. service provider (privately). This dataset comprises of over 10 million CDRs generated from a one a week in July 2019. The CDR data we have used has been completely anonymized for privacy reasons. Data is collected over an interval of ten minutes and aggregated for each cellular towers. Data visualized in Fig. 1(A) shows the number of call dropped during the time interval of ten minutes i.e. 4:50 PM to 5:00 PM. we have used a thresholding of 300 to distinguish between the cellular towers which are performing normal or abnormal. We have color coded the cellular towers which are normal functioning in yellow. The towers having number of call dropped or blocked more than 300 are assigned red color. There is region selection marker to the right of Fig. 1(A), that is used to select the region of interest and is marked in the blue color polygon. This region is of interest to the operation team to explore it further in case of abnormal network performance. CDR data of selected region is then filtered to the Parallel Coordinate Plots in Fig. 1(B) which is analysed for all cellular towers and their corresponding disconnect codes to filter out the relevant performance indicator which is further visualized in Fig. 1(C) in spiral plot visualization.

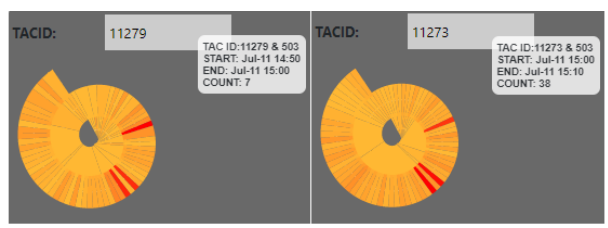

Figure 2: Visualization of relevant features from cellular tower ID 11273 **and** 11279 **for disconnect code** 503**. It shows an early trend of outages at** $3PM$**, where actual outage happened between** $5PM$ **to** $6PM$ **as visualized in this figure.**

3 LSTM BASED PREDICTION

Long Short Term Memory (LSTM) is a well known time series prediction model, which has been successfully applied in natural language processing, video frame prediction, weather forecasting, stock market and has been widely used in many fields and achieved great success, such as in music generation, image caption, speech recognition and machine translation. Since the LSTM has a promising outcomes when it comes to sequence observation and time series forecasting, we found LSTM to be the perfect match for our problem of predicting outages. LSTM improves the hidden-layer cell on the basis of RNN. The improvement of cell can make up for the gradient disappearance problem of RNN. LSTM adds some memory units, including forget gate, input gate and output gate. The memory units can further control the data and decide which should be retained and which should be deleted. We trained our model for single LSTM layer to begin with, for both the disconnect codes 1 and 503 which are considered to be the most relevant performance indicator for outage prediction. Neural network model used to predict the outages for both the codes are shown in Fig. 3 and Fig. 4 respectively. Selecting disconnect codes 1 and 503 as features was an outcome of visual analysis done in Fig. 1(A), (B) & (C), which is shown in the form of spiral visualization in Fig. 2.

```
Model: "sequential_5"
_________________________________________________________
Layer (type)          Output Shape         Param #
=========================================================
lstm_8 (LSTM)         (None, 128)          66560
_________________________________________________________
dense_5 (Dense)       (None, 1)            129
=========================================================
Total params: 66,689
Trainable params: 66,689
Non-trainable params: 0
_________________________________________________________

Train Score: 2.65 RMSE
Test Score: 3.27 RMSE
```

Figure 3: Neural Network Model with single LSTM layer used for predicting outage considering feature from Cellular ID 11273 and disconnect code 1.

```
Model: "sequential_6"
_________________________________________________________
Layer (type)         Output Shape        Param #
=========================================================
lstm_9 (LSTM)        (None, 128)         66560
_________________________________________________________
dense_6 (Dense)      (None, 1)           129
=========================================================
Total params: 66,689
Trainable params: 66,689
Non-trainable params: 0
_________________________________________________________

Train Score: 2.57 RMSE
Test Score: 3.01 RMSE
```

Figure 4: Neural Network Model with single LSTM layer used for predicting outage considering feature from Cellular ID 11273 and disconnect code 503.

4 RESULTS

In this section, we investigate the performance of our neural network model for predicting outages ahead of time which is shown in Fig. 5 for Cell Tower Id 11732 with disconnect code 1 and Fig. 6 for disconnect code 503 respectively. We can see that earlier trends at $3:00$ *PM* changes the normal predicted trend of the network in both the visualizations in Fig. 5 and 6 respectively. This helps operation team to explore the reason for outages in timely manner and fix it before it actually occurs. The predicted and the normal result is also visualized in the Fig. 1(E) line chart in detail, which is aggregated over the county's performance shown in Fig. 1(D).

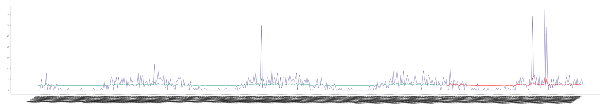

Figure 5: Predicted time series of network performance on disconnect code 1 using proposed neural network model over 4 days period. Predicted time series on test data is highlighted in red line chart where blue is the original time series and green is the predicted time series on training data respectively.

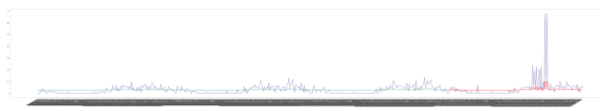

Figure 6: Predicted time series of network performance on disconnect code 503 using proposed neural network model over 4 days period. Predicted time series on test data is highlighted in red line chart where blue is the original time series and green is the predicted time series on training data respectively.

5 CONCLUSION

We presented a novel visual analytic approach to figure out relevant cellular features form vast amount of cellular network data for predicting outages in advance. Our visual analytic tools help users to dig down deeper into the data and find out anomalous behaviour in timely manner. We demonstrated the advantage of our tool over traditional approach that involved hit and trial running the code for every cellular tower and their corresponding disconnect codes. By extracting just a few relevant features from CDR dataset can be used to predict the outages well ahead in time using the neural network model mentioned in this paper.

Future work on this project will mainly comprise of two complementary areas. Currently, the proposed LSTM model does not consider the spatial-temporal characteristics of the CDR data and the relationship between many other attributes. Secondly, the root cause of the network outage is still not diagnosed from the proposed solution, which needs further visual analysis of the data to find out correlation among various attributes of CDR.