The Alan Turing Institute

GSMA

telenor group

Data Study Group Final Report: Telenor

2-6 September 2019

Green Radio: Dynamic power saving configuration for mobile networks

The GSMA launched the Global AI Challenge in partnership with The Alan Turing Institute to explore areas where AI can make a significant impact on operators' businesses and also deliver societal and economic benefits on a global basis. By bridging the gap between mobile operators and the very best academic talent, deeper analysis and more diverse insights have been achieved when interrogating the complex and very valuable datasets that operators have.

As one of four mobile operators who became challenge owners, Telenor, along with the GSMA and the Turing embarked on this Data Study Group (DSG) addressing dynamic power saving for mobile networks. Artificial Intelligence will have a profound impact on mobile operators' businesses and the wider mobile ecosystem. Today, many operators are experimenting and deploying machine learning and other AI techniques at scale, but getting it right and scaling production is not an easy task, especially considering the limited resources in this nascent field which highlights the importance of initiatives such as this.

The GSMA represents the interests of mobile operators worldwide, uniting more than 750 operators and nearly 400 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and internet companies, as well as organisations in adjacent industry sectors. The GSMA also produces the industry leading MWC events held annually in Barcelona, Los Angeles and Shanghai, as well as the Mobile 360 Series of regional conferences.

https://doi.org/10.5281/zenodo.3786852

Contents

1 Executive summary

1.1 Challenge overview

Traffic demand of a cell site or sector exhibits strong spatial and temporal variation. For instance, commercial areas only experience high traffic demand during day-time while close to zero demand in the evenings and weekends. Power consumption in mobile networks varies in accordance with network traffic load [9]. Mobile network parameters are typically statically configured, hence the cell sites are dimensioned for handling peak traffic loads at all times. This leads to wastage of network resources during off-peak hours resulting in additional power consumption [3].

Goal of this challenge is to minimise the power consumption in radio-cells or sectors in order to reduce overall operational expenditure (OPEX) incurred by mobile operators. Specifically, our goal is to develop a framework that understands and takes advantage of the traffic load variations by automating next-day power saving schemes for each individual cell tower in a country. Expected outcome of the project is an algorithm which determines, based on current load and expected demand profile in the area, when to turn off radio-cells or sectors to save power.

1.2 Data overview

This project makes use of rich datasets from multiple sources. The main dataset consists of traffic demand from mobile users aggregated at the sector or radio-cell level. Traffic demand for each sector is provided at hourly intervals and span a period of over two years.

Traffic demand values are provided in the form of number of Physical Resource Blocks (PRB) – the smallest unit of bandwidth allocation in time and frequency [2]. A given number of PRBs, forming part of *coverage layer*, are always turned-on providing basic coverage for all sectors. During periods of high traffic demand, additional PRBs, part of *capacity layer*, are made available to each sector [1].

In addition to cellular traffic dataset, other datasets providing external information which were supposed to influence mobile traffic demand are also supplied. These include sector location data, weather data for each day as well as holiday information.

1.3 Main objectives

Mobile demand varies both spatially and temporally. For example, the temporal usage patterns are very different in residential vs business areas. During off-peak hours there will be periods with overcapacity, resulting in unnecessarily high power consumption. To reduce power consumption, the current best-practice has been to turn off cells during night time for all areas, but this approach is not optimal [19]. Hence, the goal of this project is to design an intelligent strategy which determines when to turn-on/off the additional capacity layer of radio-cells to save power. This problem is divided into the following main objectives:

1. Understand spatial and temporal trends in mobile data usage across a country, as this will allow to better plan future capacity upgrades.
2. Forecast the next-day traffic demand to design an informed optimal policy for turning off capacity layer of the radio cells.
3. Recommend an individual power-saving scheme depending on mobile usage in the area covered by the site.

1.4 Approach

The rich availability of mobile network data with Telecom operators serves as a strong motivation to explore data-driven solutions for solving the problem at hand [21]. Hence, the solution approach adopted in this project includes (a) understanding the underlying characteristics of mobile network traffic data using advanced data analytics techniques, (b) applying state-of-the-art machine learning methods to forecast next day traffic demands and (c) enabling the automated and informed decision decision-making leading to a reducted network OPEX.

The approach adopted in this project is structured into the following phases,

1. **Exploratory data analysis**: To understand traffic demand data better, spatial and temporal analysis is performed providing

descriptive statistics on the data. Details on data analysis are presented in Section 3.

2. **Traffic forecasting**: This is the data modelling phase in which multiple models to predict hourly traffic load of the subsequent day are developed. Various traffic forecasting models considered are described in Section 4.

3. **Power-saving strategy**: Finally, an activation policy is designed that recommends next day power-saving scheme based on traffic forecasts made in previous step. This is described in Section 5.

4. **Sector clustering**: To overcome the issue of scalability, sectors with similar temporal traffic demand profiles are clustered so that a common power-saving strategy can be designed for the group of sectors. Clustering mechanism is presented in Section 6.

1.5 Main conclusions

Traffic demand in sectors exhibits high temporal correlation or seasonality, thus serving as an enabler for traffic forecasting models to be used to predict next-day traffic demands. Among various forecasting models implemented, we found that multivariate Gaussian distribution based model achieved highest prediction accuracy, however it suffered with huge processing costs.

Based on traffic forecasts, an activation policy is proposed that determines the next day on-off period for various sectors based on their predicted load. As seen in the results, the parameters of proposed policy cost function allow us to adapt for different traffic models and thus performs well for both overestimating as well as underestimating models.

To overcome scalability issues, sector clustering is presented as a viable solution. The clusters obtained clearly segregate various sectors with distinguished traffic characteristics such as residential and industrial sectors.

1.6 Limitations

1. Traffic forecasting models presented in this report have not exploited spatial correlation to make predictions. One known model that could exploit spatial as well as temporal information jointly is a Gaussian Process model [8]. This model has been proven useful in areas such as crime, epidemiology, modelling demand, etc. where data is spatially correlated, and hence can potentially be beneficial for the current problem.

2. It is well understood that traffic demand of a sector evolves with time. While such changes are gradual, in certain cases they can be abrupt. It cannot be commented that whether the traffic forecasting models implemented in this project are robust against abrupt traffic demand changes and hence this aspect needs further investigation. More advanced forecasting algorithms may be needed to overcome this problem, in case it exists.

1.7 Recommendations and Future Work

1. There is a great potential to pursue this project further especially in performing detailed traffic characterisation and more accurate forecasting with the availability of more detailed datasets. This may include finer data temporally – for better prediction of peak mobile loads, and spatially, more data from rural areas – to overcome skewness issues, channel quality data etc. Information regarding the number of users requesting a service could also prove helpful in studying the impact of power-saving policy.

2. Changes in traffic demand profiles of a sector are very common in cellular networks due to the configuration changes or new service launches. Given that prediction models are trained on historical data, it is difficult to adapt these models to such scenarios. A promising future research direction is to exploit the use of *Reinforcement Learning* (RL) which has the potential to overcome this problem by updating the previously learned network parameters in a self-organising manner.

3. The number of power-saving windows is currently limited to only

one per day i.e., the capacity layer may only be activated for one continuous period per day. However, it is straight-forward to test for more than one switches in future work. Moreover, switching-off the capacity layer may incur certain operational costs. The activation-cost function can incorporate this cost by means of an additional parameter.

2 Data Overview

2.1 Dataset Description

2.1.1 Mobile Network Data Introduction

The utilisation of available bandwidth at each sector or a radio-cell is a measure of the traffic demand in that sector. Bandwidth utilisation is used to assess overall load on mobile networks and is measured as the fraction of occupied Physical Resource Blocks (PRB), where PRB is the smallest element of bandwidth that can be allocated to a mobile user by a sector [2]. A mobile cell site generally has 3 sectors, each covering 120 degrees. Each sector has a *coverage layer* and a *capacity layer*. The coverage layer has to be always turned-on to supply connectivity and basic services. The capacity layers are added to handle heavier loads e.g. during peak hours. Generally, the baseline coverage layer supplies 50 units of PRBs and operates at 800 MHz band. Capacity layer can supply up to 300 units of PRBs of capacity and uses higher bands like 1800 MHz or 2100 MHz [1].

2.1.2 Traffic Demand Dataset

The dataset for this challenge consists of 1310 sectors with an average of 3 sectors per cell site (hence 437 different cell sites). For each sector, we are provided hourly average PRB demand. The dataset spans over a period of 27 months from $01/04/2017$ to $01/06/2019$ and covers two big cities of *Denmark*.

Table 1 presents a snippet of traffic demand dataset. For each sector, the dataset consists of number of PRBs utilised (traffic demand) in the field prb_used_pdsch_avg. The field prb_avail_capacity consists of the maximum number of PRBs available in the basic coverage layer. If prb_used_pdsch_avg is less than prb_avail_capacity (i.e., 50), PRBs from basic coverage layer are used. However, if prb_used_pdsch_avg exceeds prb_avail_capacity, additional capacity layer is turned on to supply more PRBs.

S.No.	sector	period_start_ time	prb_used_ pdsch_avg	prb_avail_ capacity_layer
0	J4041xx2	2017-11-12 01:00:00	0.02	50
1	J0811xx3	2017-11-25 12:00:00	15.82	50
2	J2781xx1	2017-11-24 18:00:00	51.90	50
3	J0602xx2	2017-11-24 17:00:00	46.98	50

Table 1: A Snippet of Dataset for Representation Purposes

2.1.3 Sector Location Data

Mobile network traffic data is supplemented with location data consisting of latitude and longitude coordinates for all sectors. Table 2 below shows the format of location dataset.

sector	lat	lon
J3227xx1	10.17	56.12
J2254xx1	10.14	56.12
J3528xx1	10.09	56.18
J2284xx1	10.16	56.22

Table 2: A Snippet of Sector Location Dataset

2.1.4 Meteorological and Holiday Data

Meteorological data consist of weather records with one entry for each day. Dataset consists of daily minimum and maximum temperature, wind speed, precipitation and solar irradiation information. Although location dataset shows that sectors are mostly spread over the South Denmark, weather data is only provided for one city but is assumed to be homogeneous across all sectors located in various regions present in the dataset. Table 3 provides an overview of meteorological dataset.

date	temp_high	wind_10_min_mean_hightest	wind_gust_highest	temp_low	temp_mean	wind_mean	precipitation	sun
2017-05-01	12.7	11.8	15.5	1.9	7.5	6.2	0.0	13.2
2017-05-02	16.5	9.4	13.2	2.2	9.0	4.9	0.0	14.4
2017-03-03	16.4	7.3	8.7	2.1	10.1	3.9	0.0	11.4

Table 3: A Snippet of Weather Dataset

2.1.5 Holiday Data

Information containing school holiday dates as well as other public holidays is also supplied as part of the dataset.

2.2 Data Quality Issues

2.2.1 Missing Data

A noticeable number of sectors had missing PRB demand data for many hourly intervals. This could pose a problem while running traffic forecasting models. Interpolation was used to fill in PRB values for missing intervals.

Also, it was observed that a bunch of sectors did not have PRB demand data for the exact same period. This could be an error while data collection or aggregation, or could be due to power outage which may have led to multiple sectors reading being lost during the same period. Sectors with such huge anomalies were discarded from further analysis.

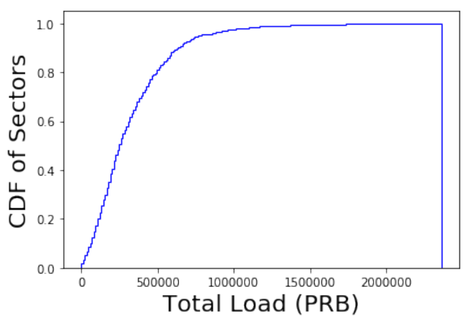
2.2.2 Abrupt Changes in Traffic Demand

Abrupt changes in the overall traffic demand were noticed for various sectors. Such changes were not consistent in terms of time period when they exhibited. As per Telenor, this could possibly be due to reconfiguration of the towers. For instance, addition of more towers could have abruptly decreased the demand of sectors next to them and vice versa.

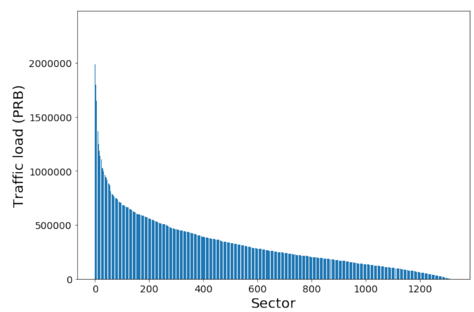
3 Exploratory Data Analysis

3.1 Network Data Behaviour

The variation of traffic demand across various sectors is represented in Figure 1. In Figure 1a, we plot CDF of base station sectors against total PRB load seen over two years of hourly data. Figure 1b shows sorted PRB load in sectors over two years. Both the CDF plot as well as PRB load plot show that there are very few sectors with significantly high or significantly low traffic demand. This shows that data does not contain significant outliers.



(a) Cumulative Distribution Function of number of sectors against PRB load.



(b) Total traffic demand at each sector (sorted) summed over 2 years

Figure 1: Network Traffic Distributions

3.2 Spatial Analysis

To study geographical context, PRB load across 437 cell locations is analysed using QGIS, a geographic information system application for viewing, editing, and analysing geospatial data [15]. Spatial variation of network traffic load is analysed along the land-use map of Denmark (provided by OpenStreetMap (OSM)). The provided dataset has most of the high density sectors belonging to two main cities in Denmark along with some low density sub-urban areas.

To understand spatial load patterns better, two regions with very different demographies are selected for detailed investigation, one representing the urban area while other from a rural locality. The *urban* region is one of the biggest cities in Denmark with 88 towers. The region consists of residential-cum-industrial areas with smaller green spaces and represents the urban traffic load profile. While the *rural* region is a small island characterised by green space with few scattered residential areas and a mere 4 towers representing the rural traffic load profile.

Average PRB load of various sectors were plotted as heatmaps over the geographical map using QGIS. It was observed that sectors located in the *urban* region present more distinctive and visible trends in data compared to the four sectors of *rural* region. These plots are omitted from the report for due to sensitivity concerns.

3.3 Temporal Analysis

To study temporal load variation, daily average loads of all sectors are plotted with the 75% confidence interval for the two regions over a years' course. We see in Figure 2 that the urban region witnessed highest amount of traffic load during winters, while the spring season, February till March, witnessed lowest traffic load.

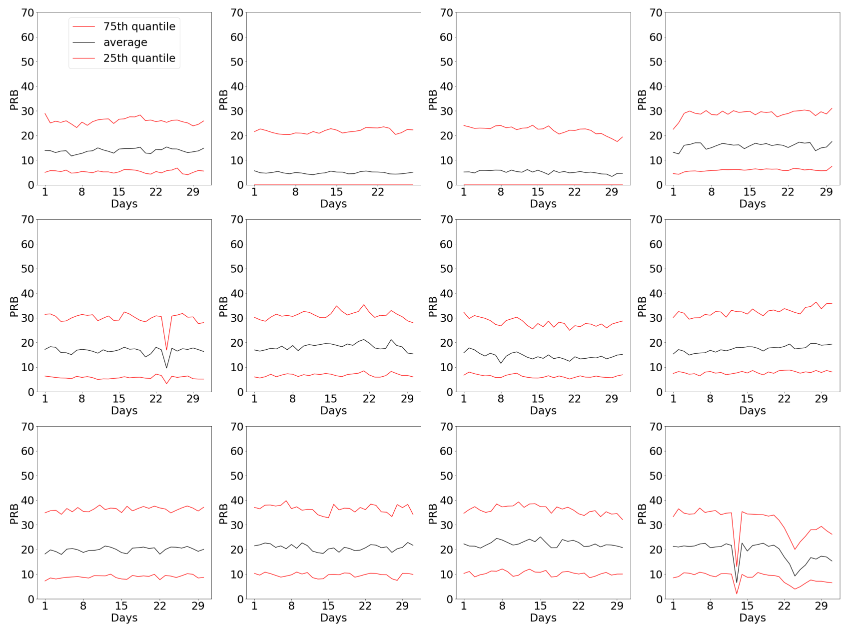


Figure 2: Summary of daily average loads of sectors in the *urban* region for 12 months (from left to right, from upper to lower).

In figure 3 representing the rural region, the load profile sees a strong weekly trend at the residential sectors with load being higher during the weekdays. The significant load increase in July results from the region specific annual festival.

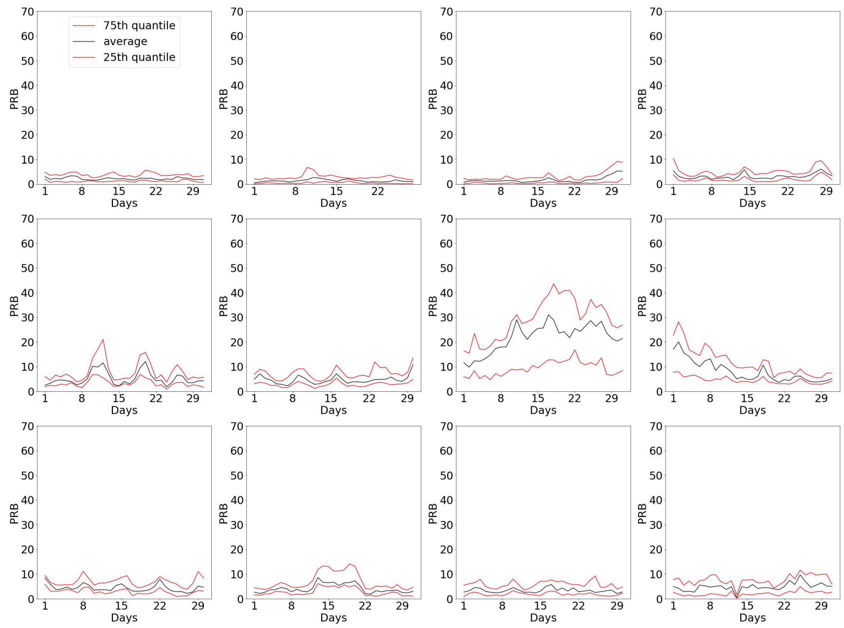


Figure 3: Summary of daily average loads in the *rural* for 12 months (from left to right, from upper to lower).

4 Forecasting Mobile Network Traffic

After understanding network traffic behaviour, next step is to forecast the traffic (PRB) load for next day. Multiple machine learning models are experimented with, in order to perform traffic prediction including:

1. Simple Linear Regression
2. Decision Tree
3. Light Gradient Boosting Machine (LightGBM)
4. Multilayer Perceptron (MLP)
5. Multilvariate Gaussian Distribution (MGD)

While estimating PRB load for the next day, these models are designed to consider the following:

1. Check the variation for different hours across each day,
2. Check the variation for different days of a week, and
3. Check the variation for holidays and weather.

4.1 Feature Engineering for Traffic Forecasting

For linear regression and decision tree, the following feature set is created:

- hour of the day,
- month of the year,
- holiday, and
- various weather features

For more advanced models i.e. Gaussian Process Regression, LightGBM, MLP and MGD, the following features are considered in addition to the initial feature set:

- twenty four hour lagged traffic demand,
- previous day mean traffic demand, and

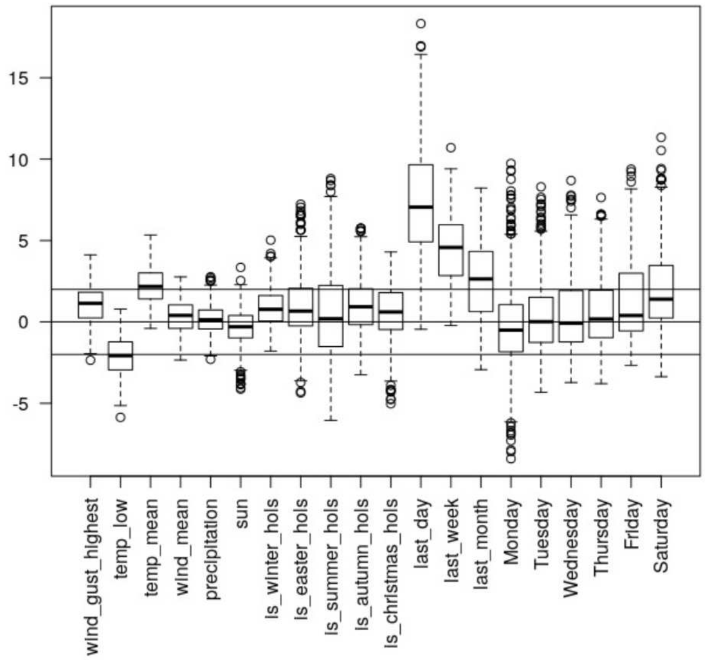


Figure 4: *t-statistics* value per regressor.

- previous day maximum traffic demand.

Categorical features such as hour of the day, month of the year, are represented in one-hot encoding format, except for tree models (decision tree and light gradient boosting machine) where these are label encoded. For MLP, input features are also normalised to zero mean and unit variance.

4.2 Simple Linear Regression

To evaluate the benefit of using non-linear models, a linear model is built to serve as a benchmark [12]. This linear model considers each hour separately to allow for a different marginal effect of each regressor. Various available regressors are considered including daily, weekly and monthly lags of the dependent variable, weather variables and dummies for public holidays.

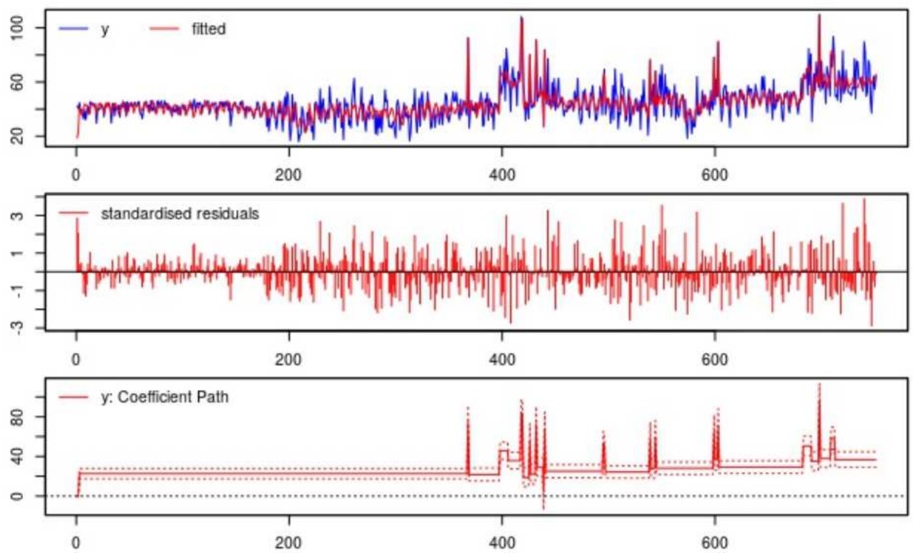


Figure 5: Step-indicator Saturation to model structural breaks.

Using this model is helpful in understanding what are the main features that determine PRB demand and by what factor. In total, there are 336 regressors. Figure 4 shows the range of *t-statistics* for each regressor [10]. We find that variables related to weather do matter significantly; in particular, `temp_low` and `temp_mean`. Surprisingly, weekdays don't seem to matter much over the whole range of models. As expected, the temporal regressors are very often relevant, with yesterday's realisation of hourly demand being most significant.

Ideally, linear regression model should detect and account for structural breaks in the data which can include location shifts and outliers. This could be done within an indicator saturation framework pioneered in [5] and implemented in R package `gets` [16]. We applied *Step-indicator Saturation* to model structural breaks [7]. Figure 5 shows that such an algorithm is needed, as the data is contaminated with many structural breaks.

Pros

- Fast to compute: model provides estimates in seconds except when robustification in regard to structural breaks are implemented, which are slow.
- Great explainability: easy to see magnitude of marginal effects and statistical significance of regressors.

Cons

- Does not on capture non-linearities and limited success in predicting bandwidth demand.

4.3 Decision Tree

The main intuition behind decision trees is to recursively partition the data space into sub-regions until a stopping condition is met. This partition can be expressed as a splitting rule, which can be combined into a tree [11]. The two common heuristics for decision trees to select the best split are Gini index and Information gain.

Pros

1. Decision trees are usually preferred due to its understandability and low computational cost, both when learning from the data as well as making prediction.
2. Decision trees can be visualised and interpreted, which allows the most deciding factors to be perceived immediately.
3. Decision trees are capable of capturing complex and non-linear relationships between features and outputs.

Cons

1. They do not guarantee robustness: a small change in the training data can lead to significant change in the tree structure, thus the final prediction.
2. Greedy approach in learning a decision tree is long-considered to be sub-optimal for simple concepts, which result in local optimality being made at each node.

3. Over-complex trees result in overfitting. To tackle such problems, pruning is necessary, which requires additional hyper-parameters.

4.4 Light Gradient Boosting Machine

The idea behind boosting is to combine multiple weak classifiers to form one final and stronger classifier by reducing statistical bias and variance [4]. It derives its methodology from gradient descent algorithm using a convex cost function.

Pros

1. LightGBM is an implementation of gradient boosting decision trees, which is capable of handling sparse data and suitable for parallel and distributed computing.
2. It exploits out-of-core computation which makes it suitable to analyse big datasets with more than say 10 million examples.
3. It also guarantees good predictive performance with minimal hyper-parameter tuning

Cons

1. LightGBM requires a large amount of data to extend its full potential. For applications where not data much data is available, traditional models such as Gaussian Process or Support Vector Machine are preferred.

4.5 Multilayer Perceptron

MLP is a class of feedforward artificial neural networks consisting of at least three layers of nodes: an input layer, one or more hidden layers and an output layer. The data travels through these layers in a forward direction from input layer to output layer of neurons, and are trained using back-propagation algorithm during which the error is calculated using least squares method [13].

Pros

1. The neurons include a non-linear activation function that can help model complex dependencies among various variables.
2. Network contains one or more layers of hidden neurons which enable the network to learn complex tasks.

Cons

1. MLPs may suffer from over-fitting, high computational cost as well as local minima can be a problem with optimisation.
2. Unlike some other neural networks, there is also a lack of biological plausibility.

4.6 Multivariate Gaussian Distribution

This methods uses a mixture of multiple Gaussian distributions with different means, (μ_i), and variances, (σ_i) [6]. In the scenarios where we observe significant outliers or periodic variation that are sort of repeated, pattern based models might not be a very good approach, especially when prediction accuracy is of utmost importance. But if those repeated patterns follow a certain distribution (say normal distribution), a mixture of different distributions can be used to capture outliers and reduce the noise effect. For MGD, time-series prediction algorithm in R is utilised (specifically Prophet package [14]) together with the concept of mixture model with some fixed effects.

Pros

1. Outliers can be detected with a high probability and hence noise effect can be reduced.
2. Long term seasonality (monthly or even yearly) is captured well using MGD, hence the traffic demands for longer periods of time can be easily estimated.

Cons

1. There might be risk of overfitting or underfitting data in the case that considerable amount of historical training data is not available.

As an enhancement, the MDG model is applied together with some random effects, so that we can obtain fixed as well as random effects. This helps in boosting the model's performance for unobserved features.

4.7 Performance of Traffic Forecasting Algorithms

Various models are trained on the first 24 month data while hyper-parameter tuning is performed on the following 2 months of data. Finally last one month is used as the validation set. Mean Absolute Error (MAE) is used as the metric to compare different traffic forecasting models. Table 4 compares the performance of various models in predicting traffic demand for the next 24 hour for a randomly chosen sector, `J0536xx1`. Notice that the MAE for MGD is least among all, hence it surpasses all other models in terms of accuracy of prediction. The table also also gives an idea about the execution cost of each algorithm where MGD is lagging far behind others questioning its scalability.

Model	Training time	MAE on test data
Linear Regression	–	–
Decision Tree	0.5s	7.68
LightGBM	2s	6.98
MLP	6s	7.1
MDG	190s	6.3

Table 4: Performance comparison of different forecasting models

Given highest prediction accuracy of MGD approach, traffic (PRB load) forecast for last week of March 2019 using MGD is presented in figure 6.

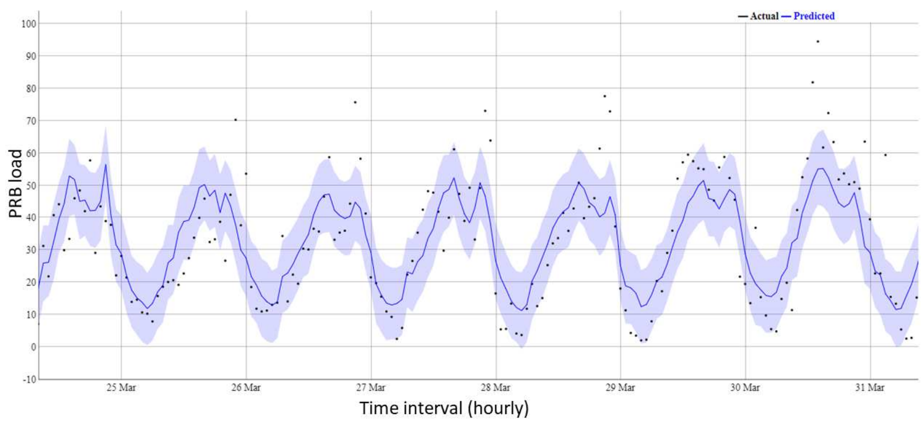


Figure 6: Predicted PRB load using Multivariate Gaussian Distribution

5 Optimising Energy Consumption

5.1 Radio-cell Activation Policy

The ultimate output of our models is an 'activation policy', which describes for each hour, whether the capacity layer for a sector should be enabled or not. Therefore, this policy should be calculated such that the cost (C) of the sector is minimised. The cost function is defined as a function of the following:

1. Activation Hours (A): this equals the cost of power consumption for the duration for which the extra capacity layer is turned on, and
2. Unmet Demand (U): this is the organisational cost of failing to meet demand when the excess capacity is switched off while demand exceeding the coverage layer of 50 PRBs.

The cost of this unmet demand is scaled by a factor that must be determined by the organisation and is based upon domain knowledge. For example, the decision to turn on capacity layer when demand is only 1 unit over the coverage layer is more difficult than when demand is 50 units over the coverage. In the former case, the cost of a slightly slower network may be acceptable. Considering this, a scaling factor λ, is introduced in our cost function which represents the cost of one unmet PRB demand as a proportion of the cost of running the capacity layer for one hour. For example, a λ factor of 0.25 defines a demand of 54 PRBs (i.e., 4 above the coverage layer) as the point at which it is neutral to turn on extra coverage. Hence, the final cost function of our activation policy is defined as:

$$C = \sum_i A_i + \lambda \sum_i U_i \tag{1}$$

Given this optimisation function, we can create an activation policy which minimises the cost for various levels of λ. This policy function takes hourly forecasts of PRB demand as input, which we derive from our models' predictions. To evaluate our models, we can compare our policy's cost against the original data as the gold standard i.e. a policy calculated directly on the real data. The difference between the costs of these policies is the true loss of the model.

5.2 Activation Policy Under Traffic Forecasts

The policy described above has no bearing on the cost function where a model incorrectly predicts a value if both prediction and actual value are within coverage layer bandwidth i.e. 50 PRBs (as either value is represented as 0 unmet demand and therefore a non-activation). Figure 7 provides better clarity on this where yellow curve represents actual traffic demand and green curve represents predicted demand. When both curves are either below or above capacity layer threshold of 50 PRBs, optimal decision making is straight forward. However, when one of them is below the threshold and other is above, optimal decision making becomes tricky. Hence a carefully selected value of λ plays an important role in activation policy.

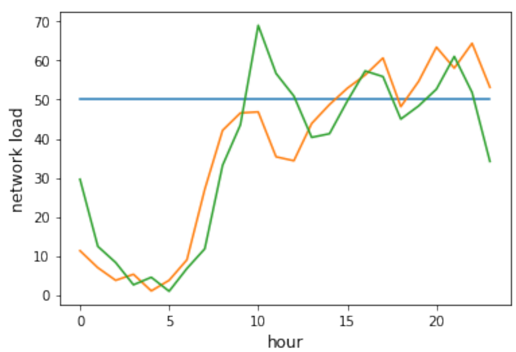
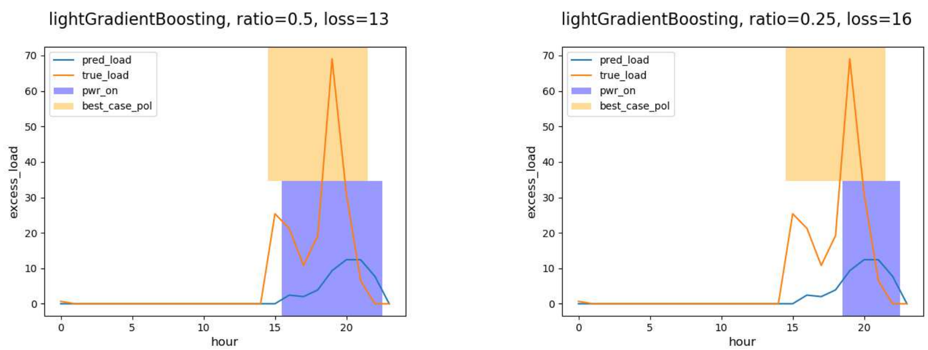


Figure 7: Policy implications for predicted demand vs actual traffic demand. Yellow - actual demand. Green - Predicted demand

The same is true for high values - if the λ is such that it causes an activation for a predicted value of x PRBs, it makes no difference to the cost when the true value is greater than x (or indeed, if the true value is less than x but still greater than the 'tipping point' as determined by the choice of lambda). This leads to models which over-predict to do better when the cost of not meeting demand when λ is high and under-predicting models to do better when λ is low. Figure 8 shows the affect of using different λ values against optimal policy.



(a) $\lambda = 0.5$ (b) $\lambda = 0.25$

Figure 8: Optimal activation policy (yellow) against proposed activation policy (blue) for different λ values. Y-axis represents traffic demand above 50 PRBs.

5.3 Forecasting Models against Activation Policy

Each of traffic forecast models are evaluated against the activation cost described in equation (1) for a range of λ values. In addition to the ML models described earlier, two naive models are also included for comparison. These models simply adopt the best policy under the assumption that the next day traffic demand will follow the same demand distribution as either its preceding day or the preceding same day of the previous week (i.e. 7 days prior).

Model \ λ	**0.05**	**0.1**	**0.25**	**0.5**	**1.0**
Decision Tree	1.28	3.24	6.20	13.50	28.27
Multivariate Gaussian	1.36	2.62	4.16	8.34	17.14
Light Gradient Boosting	1.16	1.16	6.79	12.44	26.89
Linear Model	2.22	3.35	9.12	18.29	37.23
Multilayer Perceptron	1.94	5.31	16.5	31.08	73.95
Last Week	2.85	6.42	8.10	16.45	33.35
Yesterday	1.81	3.20	6.50	12.80	26.68

Table 5: Cost C for different models at different λ values

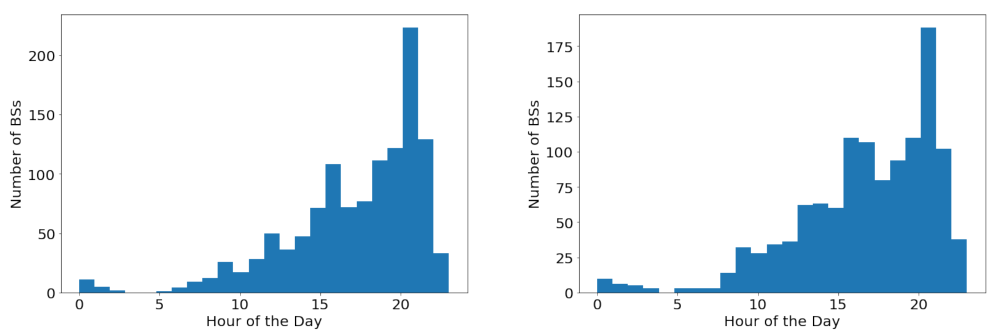
Table 5 shows the losses for each model as we adjust the λ value. For low values of λ, both the predicted and true models will not respect unmet demand and therefore design 'always off' policies. As this is an easy decision as we expect loss to naturally decrease for low λ values. The inverse is true as λ approaches 1 when model avoids unmet demand at all costs. The most interesting cases would arise for λ values that are in between these two extremes.

6 Clustering of Sectors

6.1 Clustering Motivation

Based on the discussions in Section 4, it is established that traffic forecasting algorithms with higher accuracy demand longer execution times. This may easily lead to scalability issues for predicting traffic demand at a nationwide scale. Also, it is generally observed that different sectors may have different behaviours in terms of traffic demand, they exhibit similarity at an aggregate level [17]. Hence, it is suggested clustering of sectors based on their demand profiles and forecasting the traffic demand for representative sectors of each cluster.

Figure 9 shows the histogram of sectors witnessing peaks at each hour of the day, separately for weekdays and weekends. Notice that most of the base stations experience peaks between $6 - 11$ pm. Although both weekdays and weekends show similar trend, but actual sectors under each bar (range) could be different.



(a) Weekdays

(b) Weekends

Figure 9: Histogram of sectors with peaks at each hour

A simplistic approach had initially been adopted to categorise sectors – divide time into multiple slots (e.g., morning, noon, evening, etc.) and categorise sectors by allocating them to each time slot based on which slots their peaks occurred in. However as seen in figure 9, this approach is not very effective because many sectors experience prolonged (or multiple) peak hours (e.g., hospitals or holiday resorts) leading to a biased or unbalanced clustering. Hence, a more advanced technique grounded in machine learning is adopted for clustering of sectors as explained below.

6.2 Feature Selection for Clustering

As per the exploratory data analysis, there are noticeable similarities among sectors in terms of hourly, weekly, and yearly seasonality. To capture such similar trends, for each time series, following feature sets were extracted:

1. average traffic demand by hour of the day (24 features)
2. average traffic demand by day of the week (7 features)
3. average traffic demand by month(12 features)

This gives a total of 43 features. Since only the trend and movement of the series were considered, all time series were normalised to the same scale by applying zero mean and unit variance normalisation.

6.3 Clustering Algorithm

Given the above feature set, a multi-step k-means clustering approach is used to categorise various sectors [20]. This approach was found to be very effective in identifying distinctive load patterns among sectors. Performing clustering on all features at the same time (43 features) could be problematic for distance-based methods due to the curse of dimensionality because of which distance loses meaning in high dimensions. Therefore, clustering was performed on each of the features sets separately. The results from each feature set were then used as input to another clustering step yielding the final grouping. The process is illustrated in figure 10.

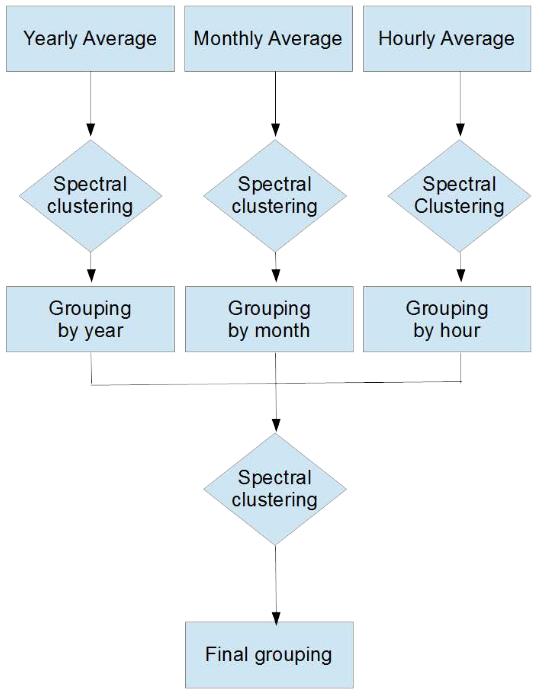


Figure 10: Multi-step Clustering Algorithm

6.4 Cluster Analysis and Visualisation

k-means clustering produced 4 different clusters. Upon investigating their traffic patterns, it is inferred that one of them belonged to the residential land-use, another one to the business area, third one belonged to rural or green space while the remaining one had a mixed traffic pattern. This was also verified by plotting various clusters on the map using QGIS (this image is not published due to privacy concerns). The clusters are therefore very distinctive in their traffic demand profiles ascertaining the quality of clusters obtained.

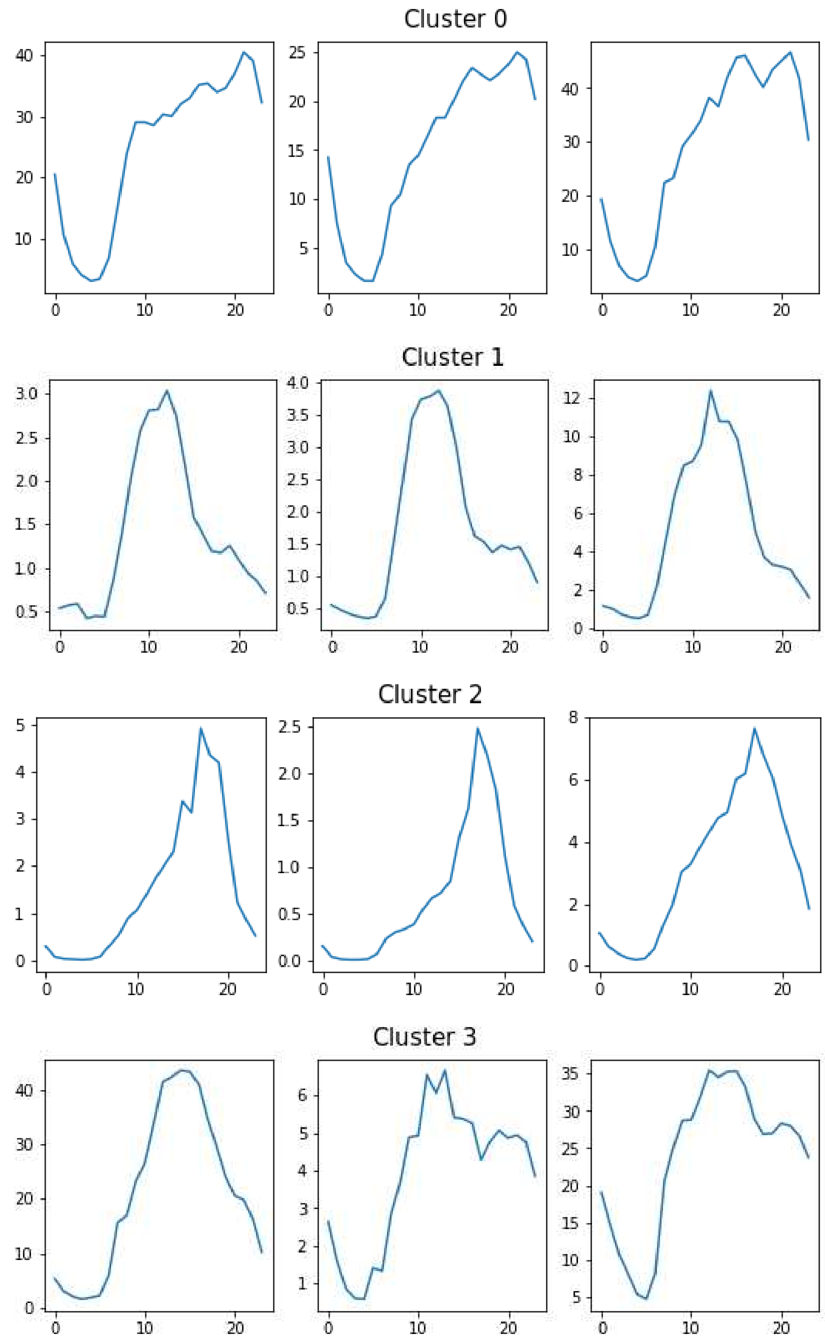


Figure 11: Hourly traffic demand for 3 randomly selected sectors from 4 clusters.

Figure 11 shows hourly traffic demand pattern of three randomly picked sectors from each cluster for one day. As can be seen, the representatives demonstrate similar patterns, and noticeably varying trends can be seen cluster-wise. Further analysis is provided below.

Figure 12 shows hourly traffic averaged over two years for all 4 clusters. Cluster 0 consists of 1105 sectors. Based on its traffic demand pattern, this cluster is tagged as *residential*. It was also noticed that the traffic demand of this cluster follows similar pattern on weekdays as well as weekends (Fig. 13(left)). Cluster 1 consists of 24 sectors. Based on its traffic demand pattern, this cluster is tagged as *business or commercial* as its sectors experience maximum traffic between 8am and 5pm. Moreover, weekend traffic demand of this cluster is significantly lesser than on weekdays reasserting that it belongs to business areas (Fig 13(right)). Cluster 2 consists of only 3. Although this cluster does not have any clear pattern, looking at the spatial map using QGIS, it is seen to belong to a forest area. Cluster 3 consists of 97. Traffic demand pattern in this cluster seems to be a combination of both residential and business areas.

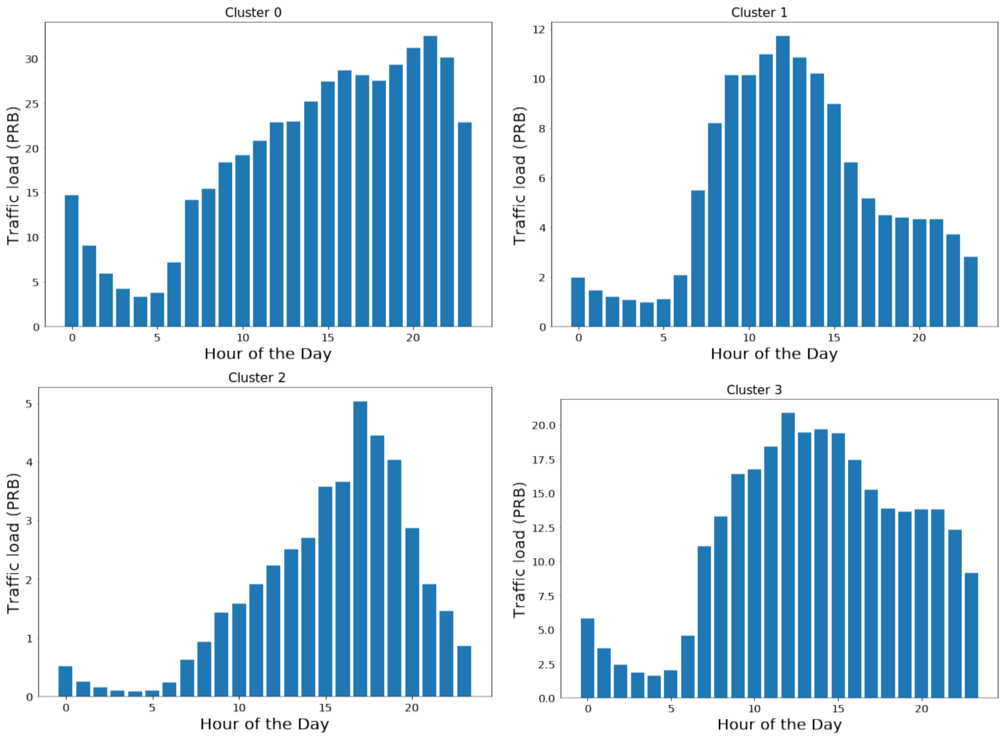


Figure 12: Hourly traffic of 4 clusters averaged over 2 years.

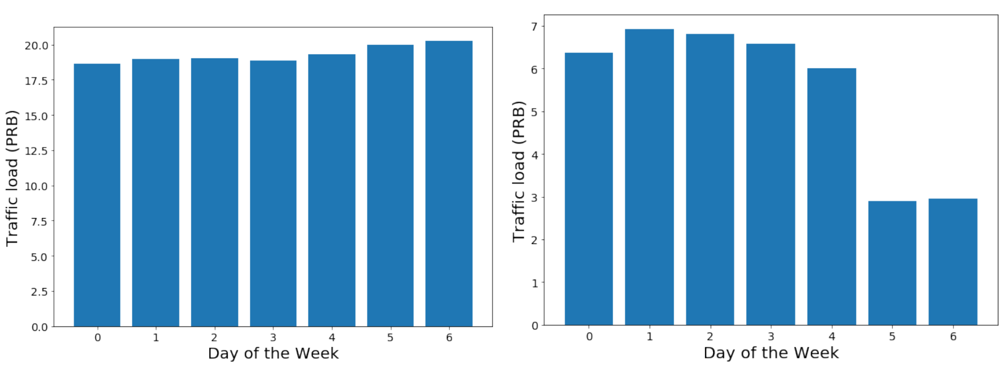


Figure 13: Daily traffic of cluster 0 (left) and cluster 1 (right) averaged over 2 years.

7 Future Work and Research Avenues

7.1 Operating on More Detailed Dataset

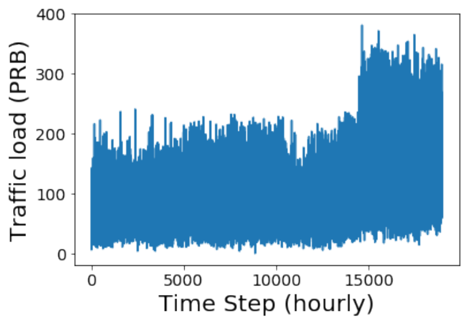
The dataset used in this challenge provides hourly average of PRB demand, however actual PRB usage per millisecond can be 6-7 times higher than the average usage. Not having more fine-grained information can be seen as a limitation since our prediction models may not be able to capture actual peak traffic demand at each second (or millisecond) level. Access to more fine grained data in time can resolve this issue.

Moreover, the dataset only consists of PRBs demand. This means that only the number of PRBs requested above coverage layer are known. If this information is supplemented with the number of users are requesting those extra PRBs, that can determine whether more users are getting bad quality of experience (which might be seen as more costly) or is it just a single user with high PRB demand.

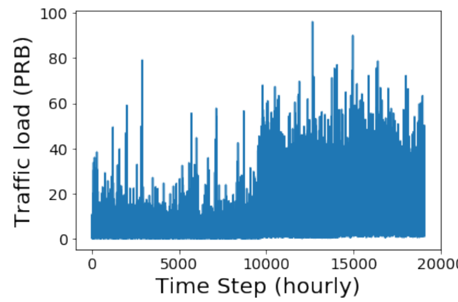
7.2 Adapting to Sector Reconfigurations and Traffic Changes

Figure 14 shows traffic demand for the most loaded sector (left) and a randomly selected sector (right) over two years. It is seen that in the former case, last 3 months have considerably higher load than the initial 24 months. Similarly, in the latter case, traffic pattern changes significantly after approximately 1 year.

Although, such changes in traffic patterns are expected in cellular networks due to configuration changes or new service launches, adapting network to such changes is the most challenging task. A promising future research direction is to exploit the use of *reinforcement learning* (RL) techniques to solve this problem [18]. RL can be used to update the previously learned network parameters in a self-organising manner.



(a) Sector with most traffic



(b) Randomly selected sector

Figure 14: Average PRB traffic load in two different sectors in the *urban* region over a period of two years.

7.3 Tackling Network Data Skewness

Network data is generally highly skewed based on whether a sector is installed in a highly urban area or a rural village. However, as seen in Section 3, the dataset provided is cleaned in a manner that it discards many sectors from rural areas, thus the skewness is missing. Although this type of cleaning has been helpful in providing non-sparse data, this could be an issue when the traffic forecast models built on the current dataset are applied to the real world data having the rural scenarios as well.

7.4 An Updated Activation Policy

There may be additional cost associated with switching-off the capacity layer as it requires reconfiguration and also causes users to be moved to other frequency bands. The technical details of this cost were unknown at the time of challenge, this cost is currently ignored. It can be included into equation (1) to design a more comprehensive activation policy.

The number of switches of capacity layer are limited to one per day; that is, the capacity layer may only be activated for one continuous period per day. However, it is also straight-forward to test for more switches in future work, and also implement the cost of switching (say γ) once that is supplied by the domain experts. Hence, the cost function can be updated as,

$$C = \sum_i A_i + \lambda \sum_i U_i + \gamma \sum_i C_i \tag{2}$$

8 Team members

Elizabeth Forde is a fourth-year Atmospheric Science PhD student at the University of Manchester. In this project, she conducted QGIS spatial data analysis, data preparation and assisted with cluster validation.

Farhad Hatami is a post-doctoral researhcer at Lancaster University. His work has mainly focused on the application of ML on Neurodegenerative diseases. Specifically, he is working with developing a methodology to handle prediction of disease progression (the method is implemented for high-dimensional longitudinal data). He has got his PhD in Mathematics at the Universitat Autonoma de Barcelona, Spain.

Juan Ungredda is a first-year PhD student at the University of Warwick. Main responsability was working on modelling using Gaussian processes to account for spatio-temporal relations in the data.

Matthias Qian is a Departmental Lecturer in Oxford, with a research portfolio of applying AI to financial, labour and property markets. His primary contribution to this project was the baseline model which evaluate the significance of the individual regressors.

Quy Vu is a data scientist at the Smart Infrastructure division of Mott MacDonald. His day to day work involves building and deploying machine learning pipelines into production. In this project, Quy was responsible for exploratory data analysis, time series clustering, feature engineering and model training (Decision Tree, LightGBM, and MLP).

Rajkarn Singh is a third-year PhD student at the University of Edinburgh working in mobile network data analytics and optimisation. In this project, Rajkarn was involved in exploratory data analysis and sector clustering. He also presented implications and future work of the project.

Ross Gales is completing his PhD in Oxford, where he studies how AI can aid intelligence-gathering by response organisations during natural disasters. His primary contribution to this project was in the design of optimisation function and evaluation of the set of models.

Yifu Ding is a first-year PhD student at University of Oxford, majoring in Engineering Science. She was responsible for data cleaning, spatial data analysis and processing the input data for the QGIS heatmap.

References

[1] *5G Spectrum - Public Policy Positioning*. URL: https://www.gsma.com/futurenetworks/wiki/energy-efficiency-2/.

[2] Sassan Ahmadi. "Chapter 9 - The IEEE 802.16m Physical Layer (Part I)". In: *Mobile WiMAX, Academic Press* (2011).

[3] G. Auer et al. "How much energy is needed to run a wireless network?" In: *IEEE Wireless Communications* 18.5 (Oct. 2011).

[4] *Boosting Algorithms: AdaBoost, Gradient Boosting and XGBoost*. URL: https://hackernoon.com/boosting-algorithms-adaboost-gradient-boosting-and-xgboost-f74991cad38c.

[5] M. P. Clements and D. F. Hendry. *Forecasting Economic Time Series*. Cambridge University Press, 1998.

[6] Chuong B. Do. *The Multivariate Gaussian Distribution*. Oct. 2010. URL: http://cs229.stanford.edu/section/gaussians.pdf.

[7] Jurgen A. Doornik, David F. Hendry, and Felix Pretis. *Step-Indicator Saturation*. Working paper no. 658, Economics Department, University of Oxford, 2013.

[8] *Gaussian Processes for Machine Learning*. The MIT Press, 2006. URL: www.GaussianProcess.org/gpml.

[9] Z. Hasan, H. Boostanimehr, and V. K. Bhargava. "Green Cellular Networks: A Survey, Some Research Issues and Challenges". In: *IEEE Communications Surveys & Tutorials* 13.4 (Fourth Quarter 2011).

[10] *Interpreting Regression Output*. Princeton University Library, 2007. URL: https://dss.princeton.edu/online_help/analysis/interpreting_regression.htm.

[11] Tom M. Mitchell. *Machine learning*. McGraw Hill, 1997. URL: http://www.cs.cmu.edu/~tom/mlbook.html.

[12] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. WILEY.

[13] *Multilayer Perceptron*. URL: http://deeplearning.net/tutorial/mlp.html.

[14] *Prophet - Forecasting at Scale*. URL: https://cran.r-project.org/web/packages/prophet/prophet.pdf.

[15] *QGIS - A Free and Open Source Geographic Information System*. QGIS User Guide, Release 3.4, Jan. 2020. URL: https://qgis.org/en/docs/index.html.

[16] James Reade and Genaro Sucarrat. *General-to-Specific (GETS) Modelling And Indicator Saturation With The R Package Gets*. Department of Economics, University of Oxford. URL: https://cran.r-project.org/web/packages/gets/gets.pdf.

[17] R. Singh et al. “Urban Vibes and Rural Charms: Analysis of Geographic Diversity in Mobile Service Usage at National Scale”. In: *The World Wide Web Conference*. 2019.

[18] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2018.

[19] J. Wu et al. “Energy-Efficient Base-Stations Sleep-Mode Techniques in Green Cellular Networks: A Survey”. In: *IEEE Communications Surveys & Tutorials* 17.2 (2015).

[20] Junjie Wu. “Advances in k-means clustering: A Data Mining Thinking”. In: *Springer Publishing Company* (2012).

[21] L. Yang X. Cheng L. Fang and S. Cui. “Mobile Big Data: The Fuel for Data-Driven Wireless”. In: *IEEE Internet of Things Journal* 4.5 (2017).

turing.ac.uk
@turinginst