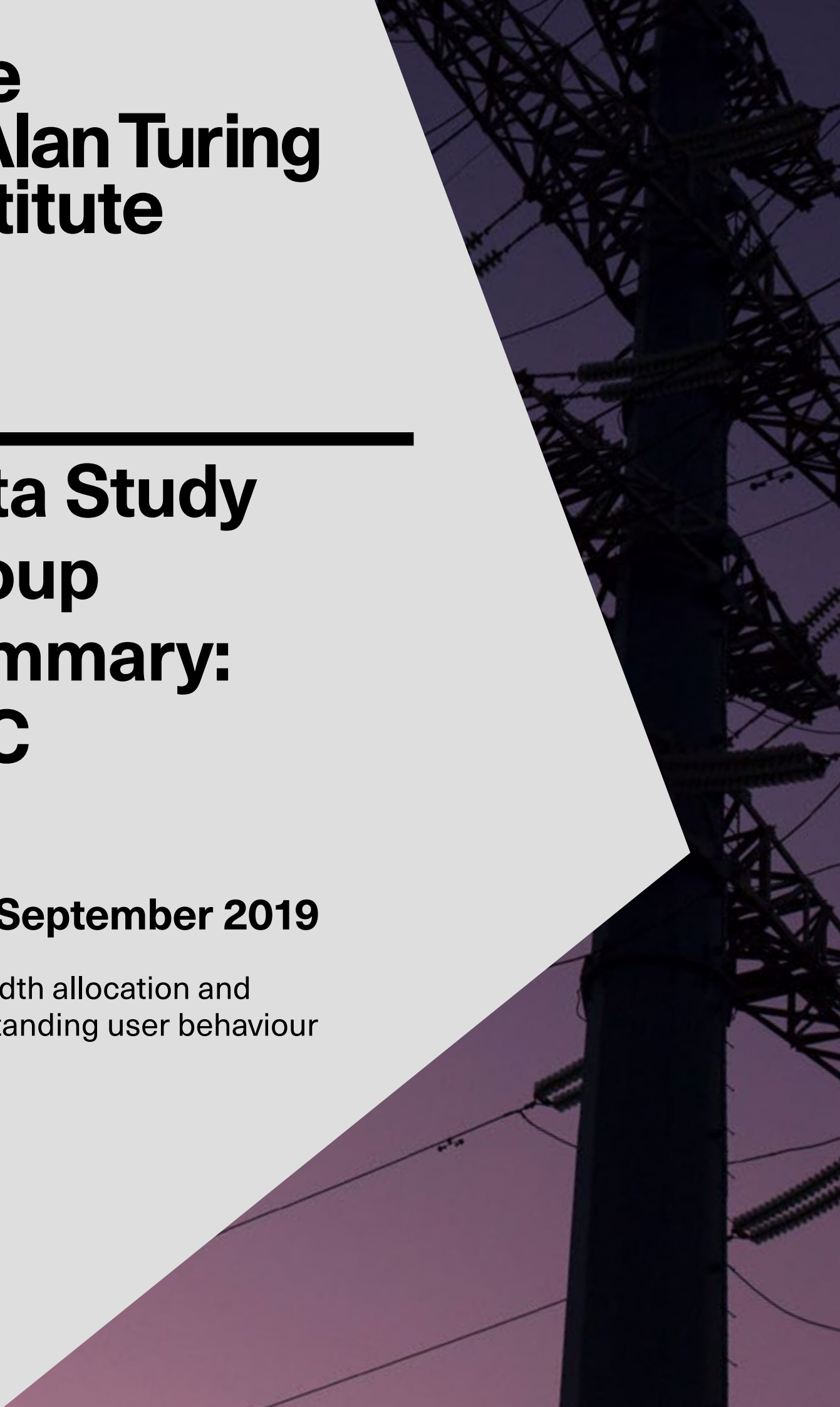
The Alan Turing Institute

Data Study Group Summary: STC

2 – 6 September 2019

Bandwidth allocation and understanding user behaviour



Challenge overview

The main challenge was to understand the user behaviour of the stc users, the majority of which use a fraction of their assigned bandwidth. Thus, there is a significant potential for cost savings both for stc and their users.

To solve this problem, we need to complete the following tasks:

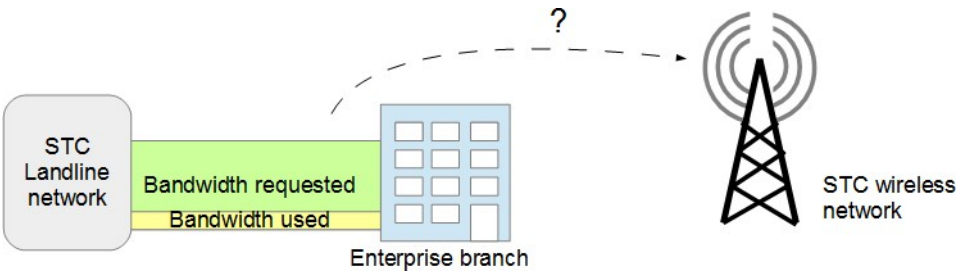


Figure 1: Problem of Assigning stc customers to the Wireless Network

1. Identify customers, who underutilise their connection and identify cell phone towers which have spare capacity to accommodate the additional traffic
2. Map customers to cells with sufficient capacity

Description of data

We were provided with the following four sampled datasets that cover around half a year of stc's customer's behaviour. The data was generally aggregated at a monthly level and covered customers in one specific city and the surrounding area.

1. This dataset provided an anonymised user ID, the category of application or service the customer used during the month, whether they used the application/service during peak hours (8am - 5pm), and the amount of traffic.
2. This dataset provided a description of stc customers in more detail. It contains the total volume of data the customers used each month and other details.
3. This dataset provided information about the cell towers.
4. This dataset contained data about the stc customer usage of their data circuit connection.

Approach

The first step was to explore whether stc's wireless network had enough spare capacity to accommodate a significant number of additional users. To do so, we employed a *K*-means classification algorithm. Then, we analysed stc customers with a fixed line connection to estimate their behaviour as well as the spare capacity of the cell sites. Once we had the spare capacity per cell site in bytes per hour, we matched customers to their closest cell if the cell had spare capacity. In order to guarantee that we did not overcrowd a cell and degrade the overall network experience for all mobile users we took the following precautions:

1. Use the maximum peak usage of a customer for the whole study period to find a realistic upper bound of an customer's maximum traffic use.
2. Use the minimum free capacity of a cell site for the study period.
3. Only use up to 95% of all available spare capacity of a cell site to ensure that there is a buffer left for unexpected surges in traffic usage.

Main conclusions

Given the problem of assigning customers which underutilize their fixed line connection to the wireless network, we developed a prototype of a greedy matching algorithm that is able to assign 80% of all customers to the wireless network. Moreover, we used K-means to identify three distinct classes of customers that could be used as input for future analysis. Last, we identified several data quality issues as well as two possible routes for future work: An improved version of the greedy matching algorithm, as well as dynamic bandwidth allocation based on the app usage of a customer.

Recommendations and future work

A potentially useful extension for allocating bandwidth is to use reinforcement learning to learn an optimal allocation pattern. In reinforcement learning, an agent is trained in a specific environment to maximize a reward function via feedback obtained from the environment. Allocating resources via reinforcement learning is an active topic of research. In our case, the agent would be trained to maximize the current throughput, while minimizing the maximum required throughput bandwidth. Two promising approaches for our problem are Simple Q-Learning and Reactive/Proactive Q-Learning.

turing.ac.uk
@turinginst