Analysis of Social Media Trends for Planning Network Resources

Shubham Tiwari

Overview

1. Objective
2. Issues
3. Possible solution
4. Methodology: Dataset
5. Methodology: Estimating BS Resource Utilization
6. Methodology: Estimating Traffic Distribution
7. Methodology: Train an MLP Classifier
8. Implementation: System Model
9. Implementation: Algorithm
10. Evaluation
11. Results

Objective

Objective

- Reduce Capital Expenditure (CAPEX) and Operating Expenses (OPEX) for a cellular network by intelligently planning network resources.

Issues

Issues

- Traditional deployment - configuration (ON/OFF) of network cells depends on the time of the day.
- For example: During office hours, cells are in an 'ON' configuration.
- But, network resources might get exhausted due to higher demand at an unexpected time!

Possible solution

Possible solution

- Predict user density in the region using social media data.
- Some social media platforms provide streaming APIs, for example, Twitter.
- Can use streamed Twitter data to estimate traffic distribution in the region?

Methodology: Dataset

Methodology: Dataset

- Twitter dataset consisting of 3267 geo-tagged tweets in the city of Tokyo from 12-04-2016 to 21-04-2016.

Date	# Tweets
2016-04-21	252
2016-04-20	299
2016-04-19	353
2016-04-18	346
2016-04-17	456
2016-04-16	513
2016-04-15	293
2016-04-14	370
2016-04-13	359

Methodology: Estimating BS Resource Utilization

Methodology: Estimating BS Resource Utilization

- Model the cellular network using an M/M/1/K-EPS (egalitarian processor sharing) queueing model to estimate KPIs[1].
- Given traffic distribution, estimate BS resource utilization for each cell.
- Define threshold probability above which the cell should be switched ON.

[1] H. Klessig et al. "A Performance Evaluation Framework for Interference-Coupled Cellular Data Networks". In: *IEEE Transactions on Wireless Communications* 15.2 (2016), pp. 938–950.

Methodology: Estimating BS Resource Utilization

- How to estimate the Poisson arrival process for the queue?
- It has been shown that there is a high correlation between the Tweet densities and spatial traffic distribution[2].
- Estimate spatial traffic distribution using Tweet co-ordinates and Voronoi estimate[3].

---

[2] H. Klessig et al. "Twitter as a Source for Spatial Traffic Information in Big Data-Enabled Self-Organizing Networks". In: *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. 2017, pp. 1–5.

[3] Christopher D Barr and Frederic Paik Schoenberg. "On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process". In: *Biometrika* 97.4 (2010), pp. 977–984.

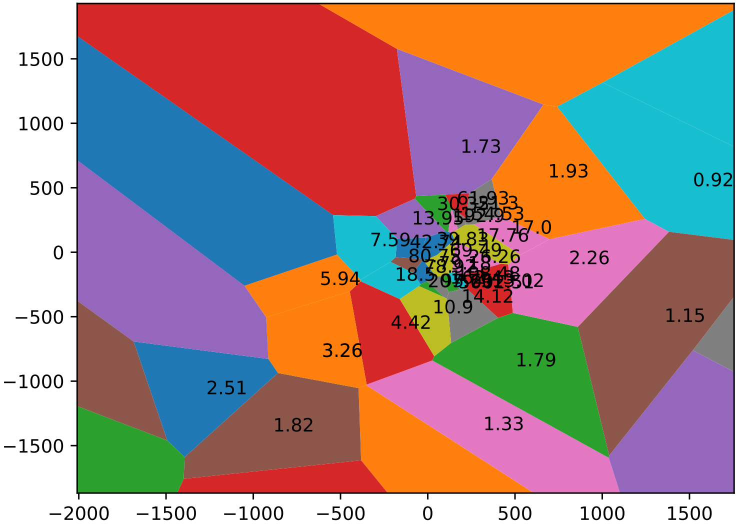
Methodology: Estimating Traffic Distribution

Methodology: Estimating Traffic Distribution

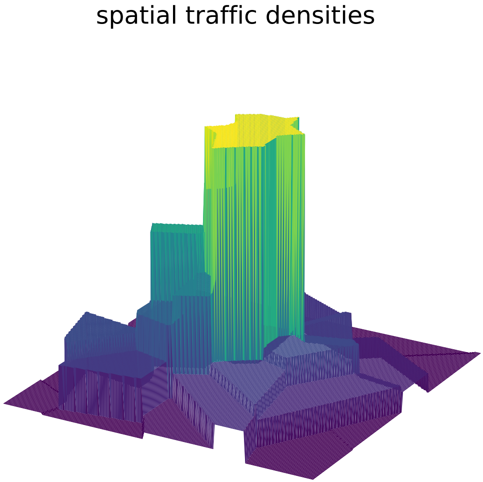
- Let $\mathcal{P} = \{p_1, .., p_n\}$ be co-ordinates of n geo-tagged tweets in the region.
- To obtain the user-density in the region, construct a Voronoi Tessellation of the set of points $\mathcal{P}$.
- User-density in each of the convex polygons obtained is the inverse of the area of the polygon.
- Scale the obtained densities to obtain traffic distribution.

Methodology: Estimating Traffic Distribution

- An example:



- An example:



Methodology: Train an MLP Classifier

Methodology: Train an MLP Classifier

- Computational complexity of BS resource utilization increases with an increase in the spatial resolution and size of the region.
- BS utilization threshold plays an important role.
- User-densities low => lower threshold values.
- Can train a classifier (for example, Multilayer perceptron) to predict an optimal ON/OFF configuration for the cells.

Implementation: System Model

Implementation: System Model

- We consider a region $\mathscr{L}$ composed of N cells, where region covered by cell i is denoted by $\mathscr{L}_i$. Each UE assumed to be connected to a unique BS, such that $\mathscr{L}_i \cap \mathscr{L}_j = \phi \forall i, j \in \{1, .., N\}$ and $\bigcup_{i=1}^{N} \mathscr{L}_i = \mathscr{L}$.
- Let η_i be the resource utilization for BS serving the area $\mathscr{L}_i$. Let $\eta = (\eta_1, ..., \eta_N)$. Also, let $\chi_i \subset \{1, .., N\}$ be the set of BS interfering with the frequency of BS serving $\mathscr{L}i_i$ The SINR at each point u in $\mathscr{L}_i$ is given by:

$$\gamma_i(u, \eta) = \frac{p_i(u)}{\sum_{j \in \chi_i} \eta_i . p_j(u) + N_o} \tag{1}$$

Implementation: System Model

- Other parameters can be estimated in the following way:
  Max. achievable rate for cell i:

$$c_i(u,\eta) = aBmin\{log_2(1 + b\gamma_i(u,\eta)), c_{max}\} \tag{2}$$

Normalized user density for the region $\mathscr{L}_i$:

$$d_i(u) = \frac{d(u)}{\int_{\mathscr{L}_i} d(u)du} \tag{3}$$

Average rate in region $\mathscr{L}_i$:

$$C_i(\eta) = \left[\int_{\mathscr{L}_i} \frac{d_i(u)}{c_i(u,\eta)}\right]^{-1} \tag{4}$$

Implementation: System Model

Assuming each flow is of size Ω Mbps on an average, the mean service rate μ_i offered by the BS serving the region $\mathscr{L}_i$ is given by:

$$\mu_i(\eta) = \frac{C_i(\eta)}{\Omega} \tag{5}$$

Load of BS i, ρ_i, serving the region $\mathscr{L}_i$ is given by:

$$\rho_i(\eta) = \frac{\hat{\lambda}_{\mathscr{L}_i}}{\mu_i(\eta)} \tag{6}$$

For systems with admission control, the BS resource utilization η_i for the BS serving the region $\mathscr{L}_i$ is given by:

$$U_i(\eta) = \rho_i(\eta)\frac{1-\rho_i^K(\eta)}{1-\rho_i^{K+1}(\eta)} \tag{7}$$

Implementation: System Model

- Finally, we obtain the following equation:

$$\eta = (U_1(\eta), .., U_N(\eta)) \tag{8}$$

- Eq. 8 can be solved by fixed-point iteration.

Implementation: Algorithm

Implementation: Algorithm

Algorithm 1: compute_resource_util_iter

function compute_res_util (η);

Input : resource utilization vector η

Output: updated resource utilization vector η_u

$\eta_u = (U_1(\eta), .., U_N(\eta))$

return η_u

function solve (η_o, tol);

Input : initial resource utilization vector η_o and tolerance tol

Output: final vector η_f after performing fixed-point iteration

$\eta_{new} = \eta_o$

do

$\eta_{prev} = \eta_{new}$

$\eta_{new} =$ compute_res_util(η_{prev})

while ($\left\|\eta_{new} - \eta_{prev}\right\|_{max} > tol$)

return η_{new}

Evaluation

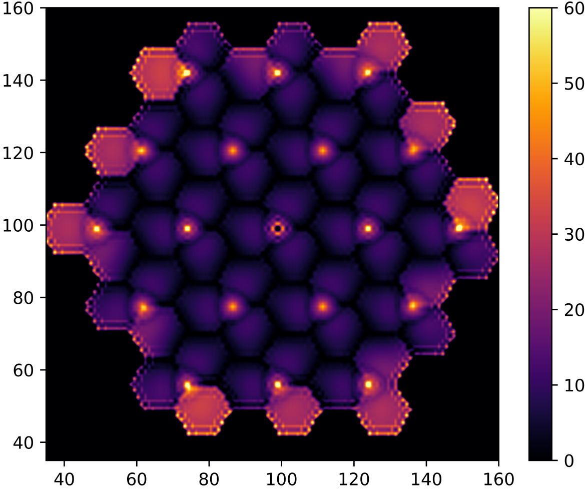
Evaluation

- Geographical region: region surrounding `Tokyo station` $(35.681308°N, 139.767127°E)$.
- Deployment scenario: urban macro-cell scenario as specified in ITU Report[4].
- Network topology: hexagonal layout with 19 cells and 3 sites per cell, with center at `Tokyo Station`.

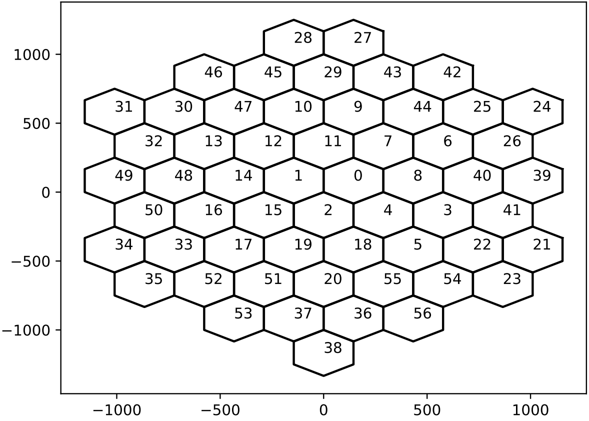
[4] M Series. "Guidelines for evaluation of radio interface technologies for IMT-Advanced". In: *Report ITU* 638 (2009), pp. 1–72.

Evaluation

- SINR map for the topology:



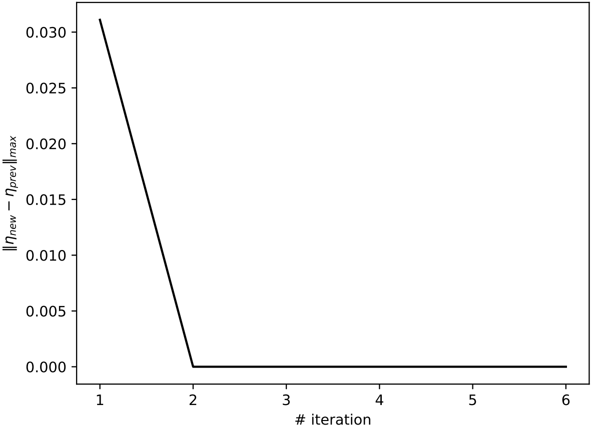
- Cell numbering convention: cells are numbered from 0 to 57 in the following way:



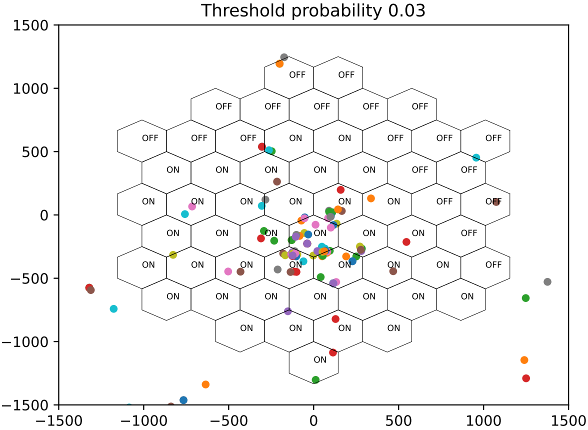
Results

Results

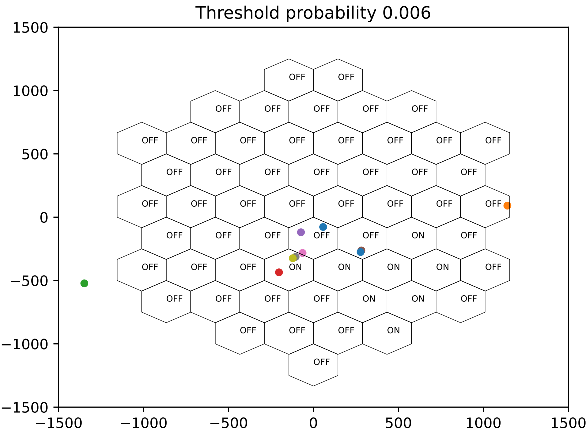
- Convergence of BS utilization vector:



- Example:

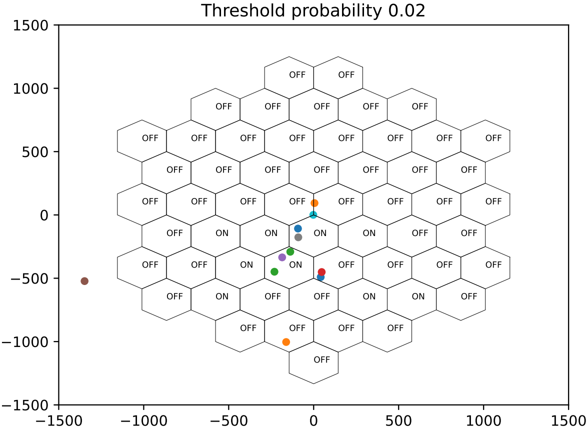

Threshold probability 0.03
OFF OFF
OFF OFF OFF OFF OFF
OFF OFF OFF ON ON OFF OFF OFF
ON ON ON ON ON OFF OFF
ON ON ON ON ON ON OFF OFF
ON ON ON ON ON ON OFF
ON ON ON ON ON ON ON OFF
ON ON ON ON ON ON ON
ON ON ON ON
ON
1500
1000
500
0
−500
−1000
−1500
−1500 −1000 −500 0 500 1000 1500

- Example:


Threshold probability 0.006
1500
1000
500
0
−500
−1000
−1500
−1500
−1000
−500
0
500
1000
1500
OFF OFF
OFF OFF OFF OFF OFF
OFF OFF OFF OFF OFF OFF OFF OFF
OFF OFF OFF OFF OFF OFF OFF
OFF OFF OFF OFF OFF OFF OFF OFF
OFF OFF OFF OFF OFF ON OFF
OFF OFF OFF ON ON ON ON OFF
OFF OFF OFF OFF ON ON OFF
OFF OFF OFF ON
OFF

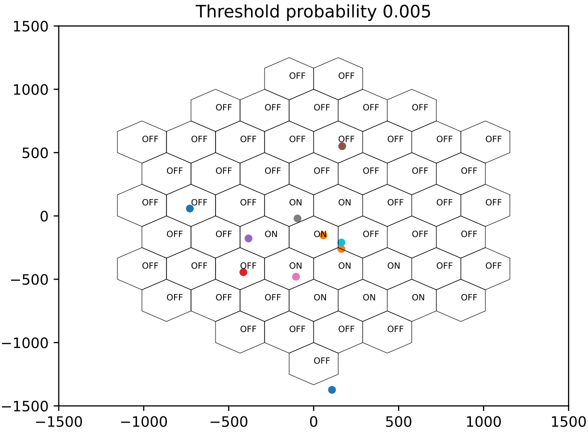
Results

◎ Example:



Results

- Example:



Conclusion

- Social Media trends can be useful in optimal allocation of cellular resources.
- Present work considers a simple ON/OFF configuration.
- Social Media trends maybe be useful in determining other configurations such as Antenna Tilt, Emission Power.
- Sentiment analysis of tweets to determine customer sentiments.

THE
END