LOCATION AWARE ANALYTICS IN THE CONTEXT OF MOBILE NETWORK PERFORMANCE OPTIMIZATION

Lucie Urbanová

Master Degree Programme (II), FEEC BUT

E-mail: xurban58@stud.feec.vutbr.cz

Supervised by: Philipp Svoboda, Martin Slanina

E-mail: psvoboda@nt.tuwien.ac.at, slaninam@feec.vutbr.cz

Abstract: The goal of this paper is to develop an estimation tool capable of predicting location aware network parameters, based on their previous field measurements and to perform a set of additional measurements to verify the accuracy of the tool. Additionally, the paper evaluates a number of regression methods in terms of their prediction accuracy, complexity and the amount of input data needed to create a prediction map of valid results.

Keywords: Estimation tool, GPR, IDW, Random Forest, Regression

1 INTRODUCTION

The radio resource management and optimization are one of the key opportunities to reduce the costs of cellular network infrastructure. One of such possibilities is an initial bandwidth allocation for the user connecting to the network. Without the knowledge of the network performance in user's location, the allocated bandwidth may be either too large (resulting in wasted resources and additional costs) or too small (resulting in limitation of the user's connectivity). The utilization of performance maps is one of such solutions. The performance maps carry information about the network characteristics depending on the location. The goal of this project is to utilize several regression techniques for creating network performance maps, namely Gaussian process regression [1], random forest [2], exponential smoothing [3] and inverse distance weighting [4]. The methods are utilized using various parameters and are benchmarked in terms of prediction accuracy, computational complexity and required amount of input data per area for a satisfactory prediction result.

2 REGRESSION METHODS

Regression analysis is a statistical process of estimating a value of a variable with relation to other variables. The goal of this paper is to find the estimate of, e.g. signal strength of LTE network (RSRP) in a previously unmeasured location based on the RSRP values measured in nearby locations. To achieve such goal, several methods have been utilized and compared in this paper.

Gaussian process regression (GPR) is a statistical tool built on the assumption, that all variables behave according to multivariate Gaussian distributions. The key aspect of GRP is defining the covariance function, which determines the behaviour of the regression on all possible points and using the maximum likelihood estimation fits the regression data.

Random forest (RF) technique is a supervised learning algorithm, which creates multiple decision trees and then merges them to get more accurate results. Random forest consists of several decision trees, which map data to outputs (results) based on the provided training data and from the results of those trees it 'decides' on the final result.

Exponential smoothing (ES) is a technique of smoothing the data based on exponential weighting of the previous observations related to the distance between the data points. There are plenty of types of exponential smoothing depending on how many parameters it utilizes. In this paper, I use double exponential smoothing, which applies the recursive filter twice to filter out inconvenient trends in the data caused by, e.g., varying speed of the measuring device and I utilize maximum likelihood estimation to minimize the error. Exponential smoothing is a technique used for smoothing the time-series, yet the principle is applicable to spatial smoothing as well.

Inverse distance weighting (IDW) is a simple regression technique that calculates the value in an unknown location based on the weighted sum of reference values depending on the inverse distance between the considered point and the reference locations.

3 METHOD ANALYSIS

For the purpose of comparing the individual methods, the regression was first based on the time domain measurements rather than location-based measurements to reduce the problem to a single dimension. The evaluation of the methods in time dimension provides the information in the whole range of the axis. The data were measured using Keysight NEMO [5] devices on the route through the centre of Vienna and were provided by the TU Vienna. In order to get enough data points, the same test was taken repeated while measuring the required parameters along the same path through the centre of Vienna, resulting in 9 repetitions of the test.

1D COMPARISON

Each of the regression methods was implemented in Matlab and later compared to the original dataset. Figure 1 left shows GPR of RSRP over all available samples of one measurement and Figure 1 right shows GPR of RSRP for all available measurements on normalized time. The normalization was made for each measurement to synchronize the start and the end of the timeline due to variations in the duration of the measurements while the path remained the same.

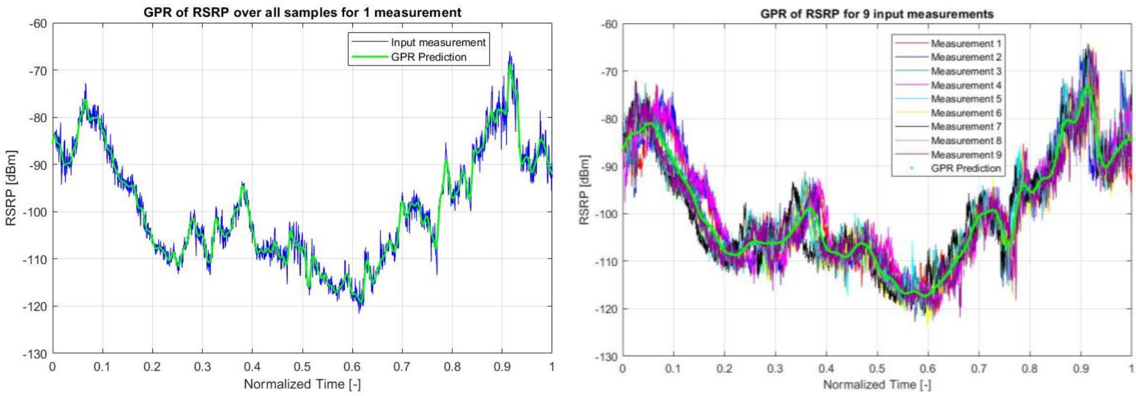

Figure 1: GPR of RSRP; Left: one measurement data input; Right: sum of 9 measurements as data input

By varying the input arguments of each regression, the parameters minimizing the deviation from the reference data have been found by comparing the output error of each setting based on the number of input data samples (100%, 10%, 1% and 0.5%) to estimate their dependency on the input data density. The chosen regression parameters have been then used to compare the regression methods against each other. The comparison of mean squared error (MSE) (see Equation 1) and mean absolute error (MAE) (see Equation 2) of various regression methods are shown in Figure 2. As a result, the best performing regression method is exponential smoothing and IDW.

$$MSE = \frac{1}{n} \cdot \sum_{1}^{n} (\bar{\bar{x}} - \bar{s})^2 \qquad (1)$$

$$MAE = \frac{1}{n} \cdot \sum_{1}^{n} |\bar{x} - \bar{s}| \quad (2)$$

where n is the number of samples, $\bar{x}$ is the vector of measured values and $\bar{s}$ is the vector of predicted values.

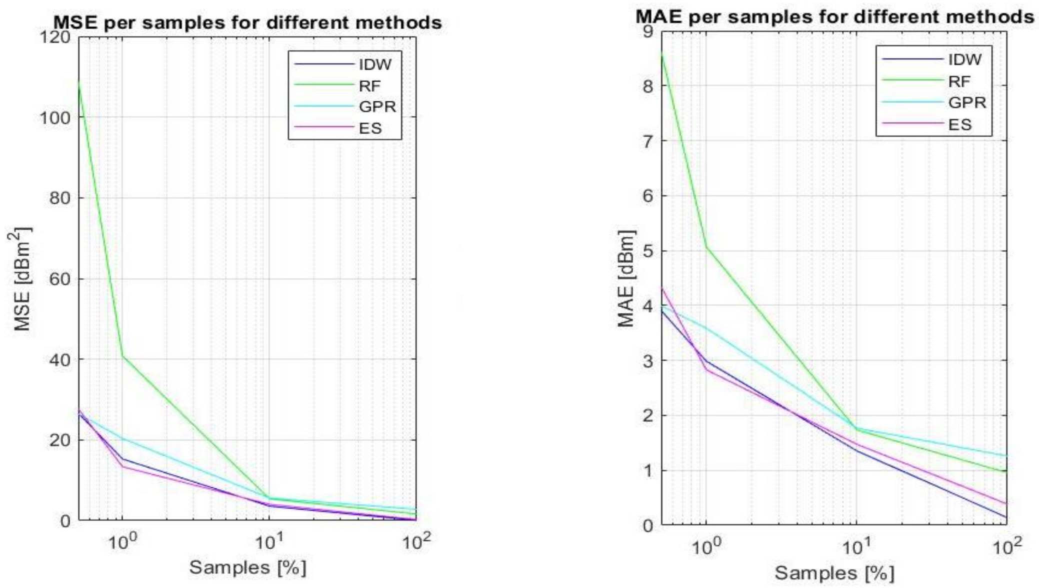

Figure 2: Comparison of error over sample density (MSE and MAE)

2D COMPARISON

It is apparent that some of the parameters of each regression had to be altered when applying the algorithms to the location-based grid. The 2D IDW regression of a single measurement, as the most accurate method for 1D with full-samples input (see Figure 2) is depicted in Figure 3, where the red line represents the trail of the reference measurement and the contour lines capture RSRP levels. The parameters of the two-dimensional IDW have been set to the power factor of 2, therefore the inverse of distances between the considered point and reference points are squared. This choice ensures the smoothness of regression in areas with sharply changing reference signal strength (due to e.g. shadowing). The radius parameter has been set to infinity, so the method calculates results over all grid, in comparison to 1D case, where minimizing the radius parameter gave slightly smaller deviations. This parameter setting takes all the reference points into account when calculating each point of the regression, while increasing the computation complexity. On the other hand, the combination of these parameters ensures validity of the regression around the area with nearby reference points, while returning mean of the reference point values in infinite distance.

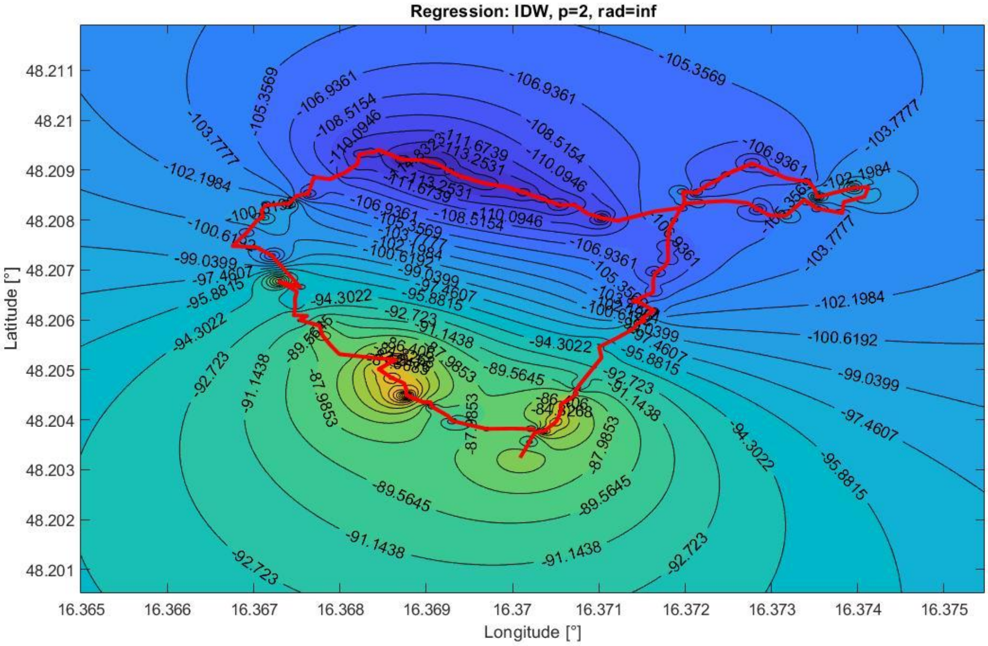

Figure 3: Location-based IDW regression

4 SUMMARY

This paper focuses on the comparison of different regression methods used for the estimation of network parameters (e.g. RSRP) based on previously taken field measurements. Four different regression methods are presented and their accuracy in time domain is evaluated. IDW was determined as the most accurate. Therefore, the space domain IDW regression is presented, displaying the estimations in the locations for which the measurements are not available. In the next part of the project, other regression methods will be evaluated (e.g. Kriging) and verified by applying on additional field measurements.

REFERENCES

[1] C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. c 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

[2] DONGES, Niklas. The Random Forest Algorithm. Towards Data Science. 2018.

[3] BROWNLEE, Jason. A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python. Time Series [online]. August 20, 2018 [ref. 2019-03-10]. Available at: https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/

[4] George Y. Lu, David W. Wong, An adaptive inverse-distance weighting spatial interpolation technique, Computers & Geosciences, Volume 34, Issue 9, 2008, Pages 1044-1055, ISSN 0098-3004, Available at: https://doi.org/10.1016/j.cageo.2007.07.010.

[5] Nemo Walker Air In-Building Measurement Solution: KEYSIGHT Technologies [online]. [ref. 2019-03-10]. Available at: https://www.keysight.com/en/pd-2767482-pn-NTD00000A/nemo-walker-air?cc=AT&lc=ger