A Study on Line-of-Sight Prediction in Urban Manhattan-Like Environments Using an Optimized Machine Learning Algorithm

Simone Del Prete(1), Nicola Di Cicco(2), Mohammad Hossein Zadeh(1), Franco Fuschini(1), Marina Barbiroli(1), Vittorio Degli-Esposti(1), and Enrico M. Vitucci(1)

(1) Department of Electrical, Electronic, and Information Engineering "G. Marconi", CNIT, University of Bologna, Italy; e-mail: simone.delprete4, mohammad.hossein3, franco.fuschini, marina.barbiroli, v.degliesposti, enricomaria.vitucci@unibo.it;

(2) Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Italy; e-mail: nicola.dicicco@polimi.it

Abstract

This paper addresses the prediction of Line-of-Sight (LoS) in Manhattan-like urban topologies. A classifier based on the Gradient Boosting Decision Tree (GBDT) model is used for this purpose, whereas training and testing datasets are generated using a standard Ray Tracing algorithm. Compared to a previous work on the same topic, the confidence of the prediction provided by the classifier is compared to an optimized decision threshold, improving the reliability of the prediction with respect to the ground truth, represented by Ray Tracing simulations. Finally, LoS curves as a function of the distance are provided for micro- and macrocellular cases and different configurations of the environmental features. As a reference, results are compared with the WINNER LoS model. In the microcellular case, it is shown that that the LoS probability curve tends to become more similar to the WINNER one as the height of the base station increases.

1 Introduction

Line of Sight (LoS) is an important factor in wireless communication, as the basic characteristics of the radio channel such as path loss, fading statistics, time and angular dispersion, etc. are affected by the LoS presence. knowledge about LOS occurrence is important from a modelling perspective, as some propagation models like the WINNER [1] and the ITU [2] models provide different path loss formulas depending on whether LoS condition is satisfied or not. From an application point of view, real-time assessment of LoS presence can helpfully assist beamforming (BF) strategies, as BF is obviously easier when the transmitter and the receiver are in clear visibility. In this framework, the development of satisfactory models to determine LoS condition – often in terms of some LoS probability – can be useful. Unfortunately, existing models are often not flexible, i. e. they can't be adapted to different scenarios. For instance, the WINNER model just expresses the LoS probability in an urban context as a decreasing exponential function with distance, without any parameters to take into account the actual geometry of the environment, like the building height, their densities, etc.

In recent past, the field of electromagnetic propagation has been affected by Machine Learning (ML) algorithms [3], leading to a number of successful applications so far. Since ML-based methods aim at learning an arbitrarily complex nonlinear function from raw data, they can be successfully applied to radio propagation modelling, where propagation markers like signal strength, fading effects, channel spreads are related to many geometrical and system features in a quite complex fashion. The use of machine learning therefore represents a useful tool for predicting LoS probability in a variety of environments with different building densities, street widths, and link distance.

The purpose of this work is to make use of ML techniques to extract a LOS probability model, using synthetic data obtained by Ray Tracing (RT) as a training and testing dataset. Gradient Boosting Decision Tree (GBDT) is the classification model considered in this paper.

There have been many investigations in the field of radio wave propagation using machine learning for path loss (PL) estimation or for user localization [3]. In [4] data collected from actual measurements are used to train a ML regression model to predict the PL of new network connections based on RSSI measurements. In [5] regression of probability for an indoor wireless link is shown to achieve values comparable to classical models. In this work, similarly to what done in [6], prediction of LOS condition for individual links is addressed, which can have interesting real-time applications. Nevertheless, LOS probability curves as a function of the distance are derived directly by using the confidence levels provided by the GBDT classifier instead going through a decision test based on a pre-set threshold, repeated on several Rx points. Moreover, with respect to [6] we consider a wider range of cases showing that the LoS probability curve tends to become similar to the WINNER one as the height of the base station increases.

2 Method and Results

Similarly to what done in [6], we have analyzed the performance of a trained GBDT classifier in predicting the LoS/NLoS condition in a reference urban scenario, the Manhattan-like topology. Training, validation and test datasets have been derived through RT simulations over several Tx/Rx positions distributed on the map, and by

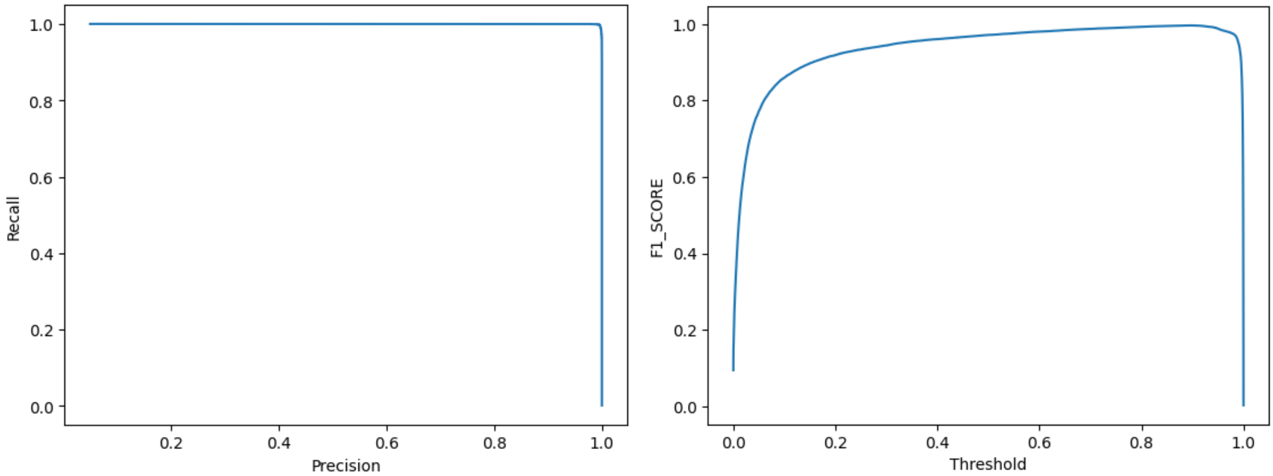

Figure 1. Example of Precision vs Recall curve (a) and F1-score curve as a function of the threshold (b). The threshold corresponding to the maximum F1 score is 0.897.

varying several environmental parameters such as length of the building blocks, street width, building height, and height of the transmitter. These parameters are then used as features for the ML algorithm, in addition to (x,y) coordinates of Tx/Rx and Tx-Rx link distance.

In the previous work [6], LoS probability curves as a function of the link distance were derived by running a Monte-Carlo simulation, i.e. averaging many thousands of Rx points distributed over circular crowns with a 50m radius, with increasing distances from the Tx. Each point was considered in LoS or NLoS based on the output of the trained classifier.

However, the ML classifier can provide both the label LoS/NLoS, and the confidence of the prediction.

In this work, the confidence level of the prediction is utilized and compared with an optimized threshold, to properly label each Tx-Rx link as LoS/NLoS. The procedure to determine the optimal threshold is the following:

- Compute the precision (p) and recall (r) parameters by employing the validation dataset. p is the ratio between the number of true positives (i.e. correctly classified LoS cases) and the number of true positives plus false positives; r is the ratio between the number of true positives and the number of true positives plus false negatives (i.e., LoS samples mistaken as NLoS).
- Compute the F1 score as follows:

$$F1 = \frac{2 * p * r}{p + r}$$

- Determine the optimal threshold as the one that maximizes the F1 score.

This procedure is summarized in Fig. 1 that shows an example of precision-recall curve (Fig. 1a), from which it is possible to determine the F1-score, as a function of the decision threshold (Fig. 1b). The maximum of the F1 curve correspond to the optimal threshold (equal to 0.897 in the case of Fig. 1b), that allows to obtain the best results on the decision. Once the threshold has been computed, the confidence values are compared with it and then the labels LoS/NLoS are determined for each Tx-Rx link. This method allows to have full control on the decision taken by the classifier and allows to optimize the decision based on the validation dataset.

In the following, some results in term of LoS probability vs link distance are presented, using some test datasets (not used previously for training and validation), with different configurations (microcellular and microcellular) and values of the features. After applying the threshold optimization method explained before, the results presented here have a better matching with the ground truth (i.e. RT simulations), compared to the results shown in [6].

Fig. 2 shows the LoS probability for a microcellular case with building blocks 20 m large, street width of 15 m, building height of 15 m, and Tx placed 2 meters above the building level (i.e. Tx height=17 m). This case corresponds to the precision-recall curve shown in Fig. 1a), which appears to be very close to the ideal case: in fact, the LoS curve predicted by GBDT is overlapped with the ground truth (RT simulation), so the error is virtually zero in this case. The curve predicted by the WINNER C2 model is reported as a reference: as it can be noticed from the figure and also discussed in [6], the WINNER macrocellular model appears to overestimate the LoS probability for short distances. However, the reason for this mismatch is probably that the WINNER Model is more suited for suburban environments, where the average building height is about 7-8 meters, and the height of the macrocell site is of the order of 20-30 m, so much higher than the average building level. If the Tx height is greater, an increase in the LoS probability is expected, especially for short Tx-Rx distamces.

In fact, Fig. 3 shows the LoS probability for another test case where the Tx height has been increased to 8 m above the building rooftop level, so closer to a “real” macrocellular case, while the other features are unchanged: as it can be noticed, the performance of the GBDT classifier w.r.t. the ground truth remains very good, but the LoS probability slightly increases for short distances and has a better match with the WINNER model, thus confirming the hypothesized trend.

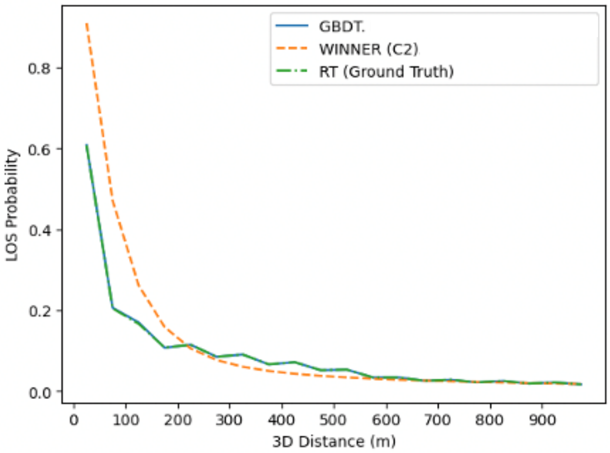

Figure 2. LoS probability as a function of the distance for a microcellular case with the following values of the environmental features: building length=20 m, street width=15 m, building height=15 m, Tx placed 2 m above the building rooftop level.

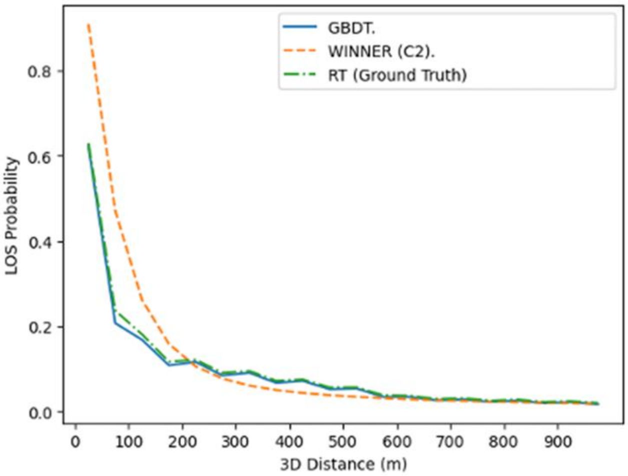

Figure 3. LoS probability as a function of the distance for a microcellular case with the following values of the environmental features: building length=20 m, street width=15 m, building height=15 m, Tx placed 2 m above the building rooftop level.

As a last case, in Fig. 4 we show the LoS probability vs distance for a microcellular case, with Tx placed at a height of 3 m above ground, and with the following values of the environmental features: building length=20 m, street width=20 m, building height=20 m. As it can be noticed, the GBDT and RT predictions have a very good match also in this case (despite slightly worse than in the macrocellular case), and, furthermore, are in good agreement with the values predicted by the WINNER micro (B1) model.

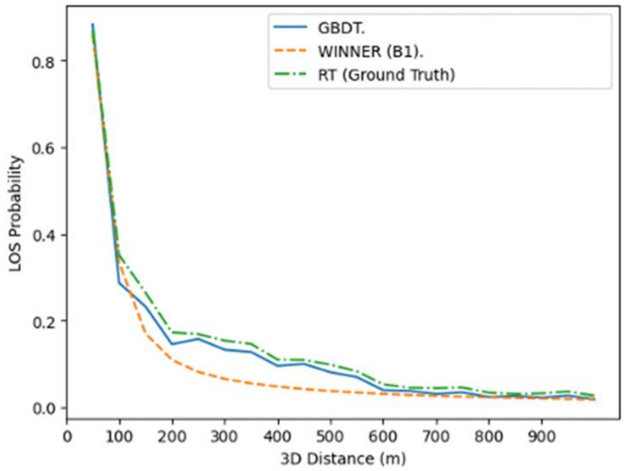

Figure 4. LoS probability as a function of the distance for a microcellular case with the following values of the environmental features: building length=20 m, street width=20 m, building height=20 m, Tx height = 3 m above ground level.

References

[1] P. Kyösti, J. Meinilä, L. Hentila, X. Zhao, T. Jämsä, C. Schneider, M. Narandzic, M. Milojevic, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. Jong, and T. Rautiainen, “IST-4-027756 WINNER II D1.1.2 v1.2 - WINNER II channel models,” Inf. Soc. Technol, vol. 11, 02 2008.

[2] ITU-R, “Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz,” 2021.

[3] A. Seretis and C.D.Sarris, “An overview of machine learning techniques for radio wave propagation modeling,” IEEE Trans. Antennas Propag., vol. 70, no. 6, pp. 3970–3985, Jun. 2022.

[4] C. A. Oroza, Z. Zhang, T. Watteyne, and S. D. Glaser, “A machine learning-based connectivity model for complex terrain large-scale low- power wireless deployments,” IEEE Trans. Cogn. Commun. Netw., vol. 3, no. 4, pp. 576–584, Dec. 2017.

[5] W. Yang, J. Zhang, and J. Zhang, “Machine learning based indoor line-of-sight probability prediction,” in Proc. of 2019 International Symposium on Antennas and Propagation (ISAP), 2019, pp. 1–3.

[6] N. Di Cicco, S. Del Prete, S. Kodra, M. Barbiroli, F. Fuschini, E. M. Vitucci, V. Degli Esposti, M. Tornatore, “Machine Learning-Based Line-Of-Sight Prediction in Urban Manhattan-Like Environments,” 17th European Conference on Antennas and Propagation (EuCAP 2023), Florence, Italy, 26-31 March, 2023.