Multi-task Learning Model for Location Estimation Using RSSI of Wireless LAN in NLoS Environment

Zehui Wei †, Mingjie Liao †
†AsiaInfo Technologies (China) Inc., Beijing, China

***Abstract*—Two major localization approaches based on RSSI are model-based. In this report, we build machine learning models to explore the possibility of data-oriented localization technique using RSSI. More attentions are paid to the multi-task learning model, which forecasts the latitude and longitude simultaneously. Numerical experiment shows that the multi-task learning model is promising in the Wi-Fi RSSI based location estimation.**

I. Introduction

Global positioning system (GPS) is widely used for the outdoor environment. However, the global navigation satellite system (GNSS) requires a direct line-of-sight (LoS) and the connection of at least four satellites simultaneously [6], which is hardly satisfied for indoor positioning. Accurate indoor positioning is useful for many applications, such as objective localization in airports, people tracking in shopping malls, navigation in public transport, robot localization and so on [4], that can improve the quality of life.

In Wi-Fi received signal strength indicator (RSSI) signal based indoor positioning, the positioning system effectively utilizes the indoor Wi-Fi access points (APs) signal strength for localization. Traditional Wi-Fi RSSI based localization approaches such as trilateration [3], [5], [6] and fingerprint [1], [2] algorithms use the user distance information to estimate the user's positions. Wi-Fi fingerprinting is usually conducted in two phases: an offline phase (survey) followed by an online phase (query). The trilateration or triangulation methods decomposes the distance information into coordinate.

Artificial intelligent (AI) or machine learning (ML) methods, which can (potentially) improve performance in different scenarios, are implemented in location estimation. Random forest is a bagging approach taking advantages by its out-of-bag (OOB) data strategy [2]. [5] compared decision tree, random forest, and gradient boosting classifier. After making a fingerprint of the floor based on Wi-Fi signals, mentioned algorithms were used to identify device location at thirty different positions on the floor. CNN-LSTM takes RSSI heat maps as the input and predicts the user positions was studied in [4].

In this work, we follows the process of mining the distance information from the Wi-Fi RSSI signal by applying a multi-task learning (MTL) model. The MTL with deep multitask architectures refers to learning and optimizing multiple tasks simultaneously through the benefit of common information and specific information among tasks [7]. Three tree models are implemented in this work as the baseline, i.e. XGBoost, random forest and extra tree regression. The proposed MTL model outperforms tree models on the mean error of the ITU RSSI based location estimation dataset.

II. Problem description

Received signal strength indicator (RSSI) indicates how strong the signal being transmitted from a specific transmitters received at certain position. In this work, the positions (represented as latitude and longitude) of four access points (APs) are given. In the training dataset, positions and RSSI referring to certain AP/APs are offered, and a verification data with know RSSI are involved to predict points' locations. Moreover, the line-of-sight (LoS) or non-line-line-of-sight (NLoS) to APs is listed.

A. Data

The positions of both training and verification data as well as APs are shown in Fig. 1. There are 13 target points in the training dataset. Most points have RSSI records from three APs except two points that have only two RSSI records. Four points have records indicating NLoS environment. There are the same number of points the verification dataset of 13 different target nodes, with slightly different positions.

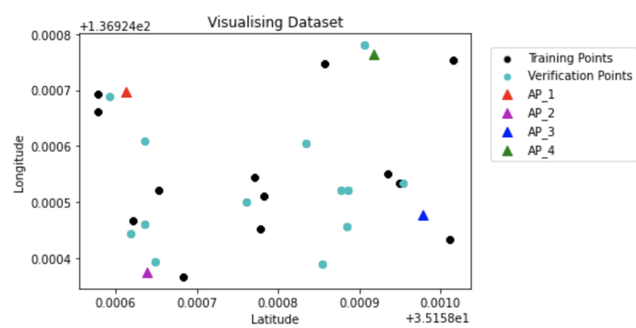

Fig. 1. Illustration of points' locations.

III. Data processing

This section introduces how we clean and augment the data.

A. Cleaning

To begin with, the information of four APs were incorporated into the training and verification dataset. The columns 'No' offers little information, and the 'Channel' contains constant value 20 throughout , hence, we remove them from the input. We notice that the time series was not continuous and records over many time periods were missing, and there is little dynamical fluctuation, hence, we assume that the RSSI values were constant during small time periods and do not consider the temporal dimension explicitly. Those different values over small time periods were regarded as outliers and replaced with mode of the periods. Finally, we dropped duplicated records and simplified our cleaned datasets. The cleaned data are shown in Table I.

TABLE I
CLEANED DATA OF A SINGLE AP

Feature	Latitude	Longitude	SSID	RSSI(dBm)	Obstacle
Type	Float	Float	Category	Integer	Category

Theoretically, three or more APs can accurately estimate the exact position of an unknown point from APs. on this condition, we remove those data with one record at a certain time point. For those points having records from two APs, we assume that they had three records and the third record was the same as one of the two records.

B. Augmentation

In practice, the order of the access points has no influence to the estimation. In this report, we permute the order of RSSI records from different access points to increase the data available. More precisely, assume one point refers to three APs, (AP1, AP2, AP3), the data of each AP has features listed in Table I. Though the order of the APs stored in the dataset does not matter, as structured data, the data of (AP1, AP2, AP3) and (AP2, AP1, AP3) is different to the model. Hence, we augment the data by permuting the order of APs without repetition of all target points.

IV. MODELS

The relation between the RSSI and the distance is well-studied. Popular models are the Lognormal model

$$P_{RX}(\text{dBm}) = A - 10\eta \log_{10}\left(\frac{d}{d_0}\right) + \mathcal{N}(0,\sigma), \quad (1)$$

the Frii's formula

$$P_{RX}(\text{dB}) = 20 \log_{10}\left(\frac{4\pi d}{\lambda}\right), \quad (2)$$

the free space path loss (FSPL)

$$P_{RX}(\text{dBm}) = 20 \log_{10}(d) + 20 \log_{10}(f) - 22.55. \quad (3)$$

The estimated distance can be further expressed as,

$$\hat{d}_i = d_i + b_i + n_i, \quad (4)$$

where $\hat{d}_i$ is the measured distance, d_i is the true distance, $n_i \sim \mathcal{N}(0,\sigma)$ and b_i is a bias introduced by the LoS/NLoS environment,

$$b_i = \begin{cases} 0, & \text{if point is LoS,} \\ 1, & \text{if point is NLoS.} \end{cases}$$

By introducing the bias, the LoS/NLoS information is considered in the location estimation. These facts inspire us to invoke the distance based method.

In this report, we chose three popular tree models: random forest, extra tree regression and XGBoost along with a multi-task learning model.

In the multi-task learning model, with the augmented input, we aim to predict the latitude and longitude as two tasks in one model. More precisely, as shown in Fig 2, after the input layer, a shared layer is used to extract the distance information from the RSSI, then the two towers are introduced to figure out the positional information from the distances.

The tree models requires two models for predicting latitude and longitude, respectively. The multi-task learning model forecasts latitude and longitude in one single model, which leads to better performance.

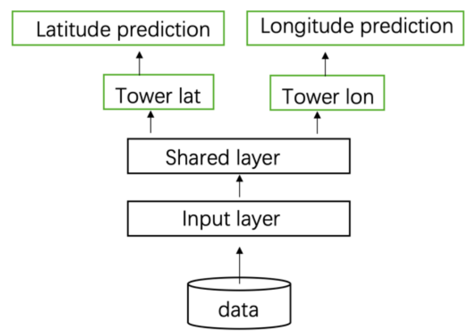

Fig. 2. Structure of multi-task learning model. The shared layer extracts important features.

V. EXPERIMENTS

To train the multi-task learning model, we run the process for 40 times, 1000 epochs each with 10^{-4} as the learning rate. We chose Adam as the optimizer. The performance is compared with tree models.

A. Results

Over the 40 repeats of training process, the mean of the mean error is 9.55552213 with standard deviation 0.9092437, while the mean of the maximum error is 34.51337252 with standard deviation 8.18946170.

The comparison of the models are shown in Fig. 3, where the result of the multi-task learning model is produced by the best model among the 40 runs. The extra tree regression outperforms the other tree models, while the multi-task learning model further improved the mean error of the extra tree regression by 11.67%. However, the maximum error of the multi-task learning models works worst. Fig. 4 illustrates the distribution of the errors in meter. It tells that the error is skewed to the smaller value, which shows the potential of the multi-task learning methods in the RSSI based location estimation.

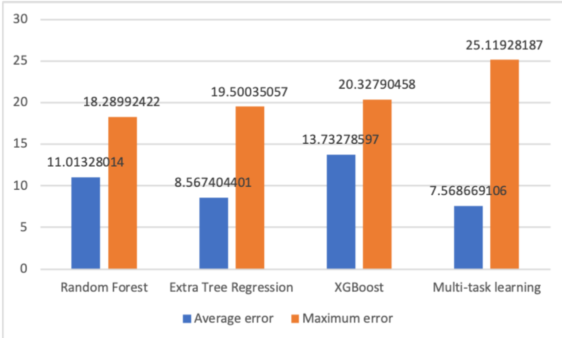

Fig. 3. Comparison of results.

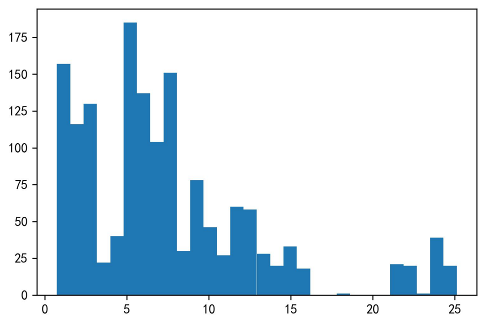

Fig. 4. Histogram of error by multi-task learning model.

VI. Conclusion

In this work, we implement the machine learning methods to the RSSI based location estimation, which has a significant value in the indoor location based service (ILBS). We proposed the multi-task learning model, which outperforms the basic tree models in the mean error. In this work, we predict both latitude and longitude in one model, to best mining the connections among features. However, due to the limit number of data, especially the LoS/NLoS points, we do not explicitly use the LoS/NLoS information. In the future, we will try to apply the multi-task leaning to explicitly extract the environmental information, which may probably improve the performance.

References

[1] S. He and B. S.-H. Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," IEEE Commun. Surv. Tutor., VOL. 18, NO. 1, 2016.

[2] S. Lee, J. Kim and N. Moon, "Random forest and WiFi fingerprint-based indoor location recognition system using smart watch',' Hum. Cent. Comput. Inf. Sci. VOL. 9, NO. 6, 2019.

[3] A. Poulose, O. S. Eyobu and D. S. Han, "A Combined PDR and Wi-Fi Trilateration Algorithm for Indoor Localization," International Conference on AI in information and communication, 2019.

[4] A. Poulose and D. S. Han, "Hybrid Deep Learning Model Based Indoor Positioning Using Wi-Fi RSSI Heat Maps for Autonomous Applications," electronics, 2020.

[5] M. A. Siddiqi, "Machine Learning Based Indoor Localization using Wi-Fi and Smartphone," Journal of Independent Studies and Research Computing, VOL. 18, 2021.

[6] V. Ilci, V. E. G*ü*lal, R. M. Alkan, and H. Cizmeci, "Trilateration Technique for WiFi-Based Indoor Localization," International Conference on Wireless and Mobile Communications, 2015.

[7] S. Ruder, "An overview of multi-task learning in deep neural networks," arXiv:1706.05098, 2017.