2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT) | 979-8-3503-9229-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIoT63253.2024.00033

Navigating Uncertainty: Ambiguity Quantification in Fingerprinting-based Indoor Localization

†Junwei Ma, §Xiangyu Wang, ‡Jian Zhang, †Shiwen Mao, §Senthilkumar CG Periaswamy, and §Justin Patton
†Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
‡Department of Electrical and Computer Engineering, Kennesaw State University, Kennesaw, GA 30144, USA
§RFID Lab, Auburn University, Auburn, AL 36849, USA
Email: {jzm0175, xzw0042}@auburn.edu, {jianzhang, smao}@ieee.org, {szc0089, jbp0033}@auburn.edu

***Abstract*—In this paper, we present a conformal prediction (CP) based method to evaluate the performance of a fingerprinting localization system through uncertainty quantification. The proposed method emphasizes a standalone module that is compatible with any well-trained fingerprint classifier without incurring extra training costs. It provides rigorous statistical guarantees for revealing true labels in the fingerprinting multi-class classification problems with high efficiency. Uncertainty quantification of the predictions is accomplished by leveraging a small calibration dataset and a given error tolerance level. Three specific metrics are introduced to quantify the uncertainty of the CP-based method from the perspective of efficiency, adaptivity, and accuracy, respectively. The proposed method allows developers to track the model state with minimal effort and evaluate the reliability of their model and measurements, such as in a dynamic environment. The proposed technique, therefore, prevents the intrinsic label inaccuracy and the additional labor cost of ground truth collection. We evaluate the proposed method and metrics in two representative indoor environments using vanilla fingerprint-based localization models with extensive experiments. Our experimental results show that the proposed method can successfully quantify the uncertainty of predictions.**

***Index Terms*—Chanel State Information (CSI), Conformal Prediction (CP), Indoor localization, Uncertainty measurement.**

I. Introduction

Fingerprinting-based localization is an emerging localization technique that covers the gap where traditional GPS-based location services are not available in indoor environments. Due to the nature of the fingerprinting method, the localization problem is treated as a pattern recognition task. Thus, the method is friendly to any stable signals, which has led to its widespread adoption in localization systems [1]–[3]. For example, MaLoc [4] uses smartphone sensors to collect magnetic readings as indoor fingerprints, achieving centimeter-level localization with the implementation of a particle filter. NaviLight [5] creatively adopted light intensity values as indoor fingerprints for pattern matching. Moreover, benefiting from the basic idea of the fingerprinting method, fingerprinting-based indoor localization evolves with the progress of pattern recognition in machine learning and deep learning. For example, LTLoc [6] proposed a device-free passive fingerprinting localization scheme using a novel adaptive Deep Neural Network (DNN). CiFI [7] adopted a Convolutional Neural Network (CNN) as a classifier to identify WiFi fingerprints generated from channel state information (CSI). Furthermore, the deep reinforcement learning (DRL) method is applied in [8] to select the optimal APs for fingerprint collecting.

However, nothing is perfect. The inherent structure of the fingerprinting method could hamper the performance of fingerprinting-based indoor localization systems. During fingerprint collection, the indoor environment is divided into grids. The density and accuracy of fingerprints directly affects the accuracy of fingerprinting-based localization systems. In [9], a SLAM-enabled robot is introduced to conduct a site survey and gather dense fingerprints with minimum effort. A Generative Adversarial Network (GAN) is adopted in [10] to create synthetic data to cut-down the cost of fingerprint collection. Similarly, a conditional GAN is proposed in TransLoc [11] to update fingerprints with unlabeled WiFi signals so that the time-consuming signal collection task is avoided when the indoor environment changes.

Nonetheless, when we examine the fingerprinting-based localization system, statistical assurance of system performance does not receive much attention. Distance error has long been the only metric used to depict the performance/accuracy of localization systems. Most existing studies concerning model parameters or ablation experiments are carried out by determining how the modification affects distance error. In fact, distance error implies that ground truth coordinates are essential, which is a strong prerequisite. On one hand, it is difficult to gather positional coordinates that are dispersed throughout a large indoor environment for evaluating the overall performance of the localization model. On the other hand, an unreliable ground truth, or test data, may negatively influence the performance evaluation.

Therefore, a performance assessment metric that is able to depict the reliability of the generated location prediction has become an essential need when deploying the majority of fingerprinting-based systems in real-world scenarios. For instance, either with contaminated data or an outdated model, trustworthy fingerprinting systems are supposed to answer: “We don’t know” [12] rather than generate the most likely point estimation. Also, an optimal metric should take into account the limitations of groundtruth labeling. The proposed metric should be able to depict system performance using a small number of labeled fingerprints.

979-8-3503-9229-6/24/$31.00 ©2024 IEEE
DOI 10.1109/AIoT63253.2024.00033

To this end, conformal prediction (CP) has been explored as a promising tool for addressing the challenges and concerns stated above [13]. It is a model-agnostic and distribution-free framework [14] to estimate the uncertainty of a machine learning model [15]. Recently, it has drawn considerable interest in the Natural language processing (NLP) field [14], especially with the rapid development of large language models (LLMs). The authors in [16] proposed a novel CP method for LLM without accessing the logits output. More important, the CP algorithm has been deployed in various real-world applications that require risk control, such as autonomous driving and AI-aided medical diagnosis. For example, in [17], the CP method is leveraged for planning paths in unknown dynamic environments with probabilistic safety guarantees. The procedure of the CP method is to first train a black-box classifier with the training dataset. Then a small calibration dataset will be utilized to calculate the nonconformity score, which describes the similarity between the given sample and the ground truth of the prediction [18]. As a result, a threshold of the nonconformity score is determined with a given confidence level $(1-\alpha)$.

In this work, we proposed a CP-based method to measure the uncertainty of a fingerprinting-based localization system. We highlight our key contributions in this work as follows:

- This work raises concerns about the inadequate performance evaluation that exist in the majority of fingerprinting-based localization research. We utilize the CP technique to assess the reliability of the system by performing uncertainty quantification. This approach provides the possibility for confidence level calibration without laborious ground truth measurement or anchor settings.
- Based on the proposed CP approach, we introduce three metrics to estimate the system uncertainty comprehensively. The metrics could be a trace to distinguish ambiguous predictions.
- We apply the proposed method in two representative indoor environments using WiFi CSI and Bluetooth Low Energy (BLE) received signal strength indicator (RSSI) as fingerprints. Extensive experiments demonstrate that our method can accurately measure the reliability of the fingerprinting system.

The remainder of this paper is organized as follows. Section I introduces the related work and motivation of our proposed framework. We present the system overview in Section III and our experimental study in Section IV. Section IV-B concludes this paper.

II. Preliminaries and Motivation

With the fingerprinting method, the localization problem is converted into a multi-class classification problem. This is a perfect match to the great power of deep learning models in pattern recognition, and contributes to the popularity of the fingerprinting-based indoor localization system. However, fingerprint collecting is essentially a process of discretizing an indoor space. The absence of ambiguity among fingerprints

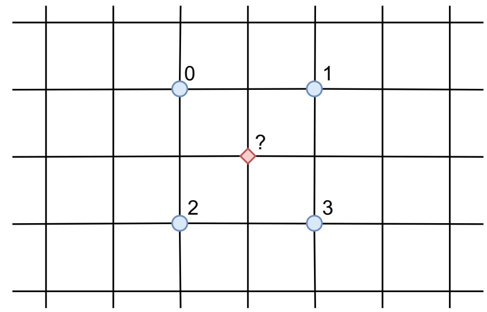

Fig. 1. The intrinsic ambiguity from ground truth locations.

makes pattern recognition a compromised approach to the localization problem.

For instance, see the example shown in Fig. 1. Fingerprints are collected at the grid points for training a deep learning model as a classifier. The model performance is evaluated by classification accuracy, by comparing the predicted label and ground truth. However, if the classification accuracy is evaluated for the red diamond location, a dilemma arises. The labels of the neighboring blue dots, i.e., label-0, 1, 2, 3, could potentially be considered true label candidates for the red diamond. Yet only one of them would be assigned as the ground truth label. If label-0 is assigned to the red diamond in the verification dataset and the model's classification result is label-3, we cannot conclude that the prediction is incorrect in the actual world, but the accuracy would be low. In summary, it will produce false negative predictions that impair the performance of the fingerprinting system.

This unavoidable error not only hampers the model training but also invalidates the classification accuracy, as the false negative rate increases significantly. Ideally, if the target could move to a random direction to break this unexpected symmetry, classification accuracy would still be an appropriate criterion for assessing the model performance. However, it is challenging to address this dilemma case with existing classification models.

Furthermore, the mismatch among models, fingerprints, and environments may also trigger a similar issue. Due to the device update or environmental change, the newly generated fingerprint could mislead the outdated classification model. The existing model will be unable to provide uncertain prediction. Thus, in this study, we used CP as an assessment tool, and propose metrics for assessing ambiguities in the fingerprinting-based localization system.

III. System Overview

A. Fingerprinting Model

Fig. 2 shows the typical architecture of a fingerprinting-based localization system. The system operates in two stages. In the offline stage, wireless signals, such as WiFi CSI and BLE RSSI, are collected and labeled with the corresponding positional coordinates. The labeled signal serves as fingerprint

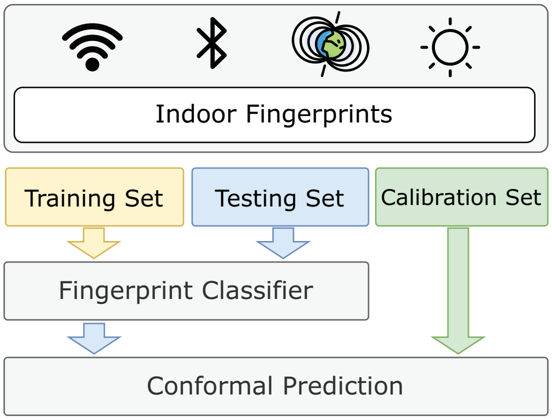

Fig. 2. Conformal Prediction (CP) is an independent component to the well-trained fingerprint classifier.

and makes up the training dataset. Then, the fingerprint classifier is trained with signal-coordinate pairs. In the online stage, users feed newly collected signals into the well-trained classifier to obtain the label prediction. The final location estimation could be the output of the classifier, or it could be derived with the classifier's output using post-processing algorithms. In this paper, we incorporate an additional conformal prediction block to quantify the uncertainty of prediction from these fingerprinting models. It is an independent block of the well-trained fingerprint classifier but as a plug-and-play component.

B. CP-based Fingerprinting

CP is a statistical method that offers uncertainty estimation for the prediction of any machine learning model [12], [18]. This work focuses on the classification problem and develops a prototype for quantifying the uncertainty of a fingerprinting localization system. It guarantees the production of a minimum-sized prediction set containing the correct label at a specific confidence level.

In the training stage, we train a fingerprinting classifier with the training dataset. The classifier utilizes a softmax layer as activation function to generate the prediction of a probability distribution for each input sample. This probability indicates a heuristic confidence in the prediction and does not align with the likelihood of correctness (e.g., the ground-truth prediction is associated with the highest score). This prediction process can be formulated as follows:

$$\mathcal{F}: x_i \rightarrow C(x_i) = \{(\hat{l}_1, p_1), (\hat{l}_2, p_2), \ldots, (\hat{l}_c, p_c)\}. \quad (1)$$

Specifically, for a given input $x_i \in \mathcal{X}$, the model $\mathcal{F}$ produces a set of predictions $C(x_i)$. Each predicted label $\hat{l}_i \in \mathcal{L}$ in the label space $\mathcal{L} = \{1, 2, \cdots, c\}$ is associated with a heuristic confidence $p_j \in (1, 0)$. Then, we propose a prototype system to convert this empiric set $C(x_i)$ to a rigorous confidence guaranteed set $C_\alpha(x_i)$ that consists of the true label l_i with a probability greater than or equal to $(1-\alpha)$.

$$\mathbb{P}\{l_i \in \mathcal{C}_\alpha(x_i)\} \geq (1-\alpha), \quad (2)$$

where $\alpha \in (0,1)$ is the user-specified error tolerance level, $\mathcal{C}_\alpha(x_i) \subseteq C(x_i)$, and it has the minimum length. To enable any well-trained model $\mathcal{F}$ to satisfy the above goal in (2), we propose a prototype system of CP-based fingerprinting. It utilizes a calibration dataset $\mathcal{D}_{cal}$: $\{(x_1, l^1_{true}), (x_2, l^2_{true}), \cdots, (x_n, l^n_{true})\}$, $x_i \in \mathcal{X}$, $l^i_{true} \in \mathcal{L}$, $i = 1, 2, \cdots, n$, to regularize $\mathcal{F}$. The calibration dataset contains n samples of measurement x_i and its corresponding ground truth label l^i_{true}. Additionally, the $\mathcal{D}_{cal}$ is usually a small dataset comprising representative samples of the target domain and can be independent of the training set D_t.

In the calibration stage, we first calculate a non-conformity score for each sample in $\mathcal{D}_{cal}$. We deploy a score function to obtain the non-conformity score S_i for sample $x_i \in \mathcal{D}_{cal}$, as:

$$S_i = 1 - (p_i)_{l^i_{true}}, \quad (3)$$

where $(p_i)_{l^i_{true}}$ is the heuristic confidence value produced by (1), and when its associated prediction is the ground truth label l^i_{true}. This non-conformity score indicates the degree to which the trained classifier's prediction will deviate from the entire calibration dataset. To be more specific, a larger non-conformity score indicates a smaller probability of the corresponding label being the correct one. Then, we obtain the non-conformity score set for $\mathcal{D}_{cal}$: $\{S_i\}^n_{i=1} = \{s_1, s_2, \cdots, s_n\}$. With a given user-specified error tolerance level α, the adjusted $(1+1/n)(1-\alpha)$ empirical quantile $\hat{q}$ of $\{S_i\}^n_{i=1}$ can be further established as:

$$\hat{q} : \mathbb{Q}\left\{\left\lceil \frac{1+n}{n}(1-\alpha)\right\rceil ; \{S_i\}^n_{i=1}\right\}, \quad (4)$$

where $\mathbb{Q}$ is the quantile function, and $\lceil \cdot \rceil$ is the ceil function. For any test data $x_{(n+1)} \sim \mathcal{D}_{cal}$, its conformal prediction set $C_\alpha(x_{n+1}) \in \mathcal{L}$ is determined when the calculated $\hat{q}$ achieves the guaranteed marginal confidence level $(1-\alpha)$, as:

$$\hat{C}_\alpha(x_{n+1}) = \{\hat{l}_j : S_j \leq \hat{q}\}^c_{j=1}, \quad (5)$$

where $\hat{l}_j$ is the predicted label by $\mathcal{F}$ in (1) and S_j is the associated non-conformal score. Thus, the determined conformal prediction set $\hat{\mathcal{C}}(x_{n+1})$ is guaranteed to satisfy (2). We summarize the above proposed conformal prediction method in this study in Algorithm 1.

C. Ambiguity Quantification

This section introduces ambiguity quantification of the predictions to evaluate the fingerprint system's statistical performance. The above proposed conformal fingerprinting system provides a solution for uncertainty evaluation of the predictions. It is a post hoc method compatible with a pre-trained fingerprinting system, including the deep learning model and observation from certain locations. As an unambiguous classification problem, fingerprinting is always compromised to predict the most likely location in the search space. The

Algorithm 1 Conformal Prediction based Fingerprinting

Input: Training dataset $\mathcal{D}_t$, calibration dataset $\mathcal{D}_{cal}$
Input: Test sample $x_{n+1} \sim \mathcal{D}_{cal}$, error tolerance level α
Output: The conformal prediction set $\hat{\mathcal{C}}_\alpha(x_{n+1})$

1: // n represents the total number of data samples within the calibration dataset
2: Train the fingerprinting model $\mathcal{F}$ with training dataset $\mathcal{D}_t$ that satisfies (1);
3: **for** $t = 1$ **to** n **do**
4: Use (3) to calculate the non-conformity score S_i;
5: **end for**
6: Compute the adjusted $(1-\alpha)$ empirical quantile $\hat{q}$ as in (4);
7: Predict $C(x_{n+1})$ by the trained model $\mathcal{F}$;
8: Based on $C(x_{n+1})$ and $\hat{q}$, determine the conformal prediction set $\hat{\mathcal{C}}_\alpha(x_{n+1})$ by (5);
9: **return** Conformal prediction set $\hat{\mathcal{C}}_\alpha(x_{n+1})$;

proposed conformal prediction fingerprinting generates the estimated conformal prediction set to satisfy the rigorous statistical guarantee. By evaluating the characteristics of this set, we can determine the uncertainty of the fingerprinting system. Therefore, we define the potential uncertainty of conformal prediction fingerprinting as:

$$\mathbb{U}_\alpha(x_{n+1}) : \mathbb{1}\left\{\mathbb{dim}\left[\hat{\mathcal{C}}_\alpha(x_{n+1})\right] \neq 1\right\}, \tag{6}$$

where $\mathbb{dim}\left[\hat{\mathcal{C}}_\alpha(x_{n+1})\right]$ indicates the dimension of $\hat{\mathcal{C}}_\alpha(x_{n+1})$ with the given error level α and test measurement x_{n+1}, and $\mathbb{1}\{\cdot\}$ represents the indicator function that maps the input to zero or one. Thus, $\mathbb{U}_\alpha(x_{n+1}) = 0$ only happens in the case where the CP set has a single label. Otherwise, we have $\mathbb{U} = 1$, which indicates an uncertain prediction has been made. There are two possible scenarios in uncertain prediction, the first of which is that the cardinality of the prediction set is larger than 1, meaning that the prediction set contains multiple labels that could be the ground truth label. The other is that the prediction set is empty, indicating that the system does not know which label should be assigned to the measurement x_{n+1}.

Based on $\mathbb{U}$, three metrics are introduced to perform quantitative analysis on a test dataset $\mathcal{D}_{test}$: $\{x_{n+1}, x_{n+2}, \cdots, x_{n+m}\}$, $x_i \in \mathcal{X}$, $i = 1, 2, \cdots, n$, which has the same distribution as the calibration set $\mathcal{D}_{cal}$. The first metric is defined as the Average Prediction Set Size (APSS), which is computed based on the test dataset $\mathcal{D}_{test}$, as:

$$\mathrm{APSS}_\alpha(\mathcal{D}_{test}) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{dim}\left[\hat{\mathcal{C}}_\alpha(x_{n+i})\right], \tag{7}$$

where $\hat{\mathcal{C}}_\alpha(x_{n+i})$ represents the estimated conformal prediction set of sample x_{n+i} in $\mathcal{D}_{test}$. By assessing the prediction of both certainty and uncertainty, the overall efficiency of the proposed conformal predictor is acquired. Generally, if APSS is closer to 1, the system will achieve better efficiency. For instance, a very large APSS (e.g., 80) indicates a lack of precision for the CP process due to an outmoded classifier or a contaminated dataset. Note that the smallest APSS does not necessarily indicate the highest efficiency of the CP.

The metric of Uncertainty Detection Rate (UDR) is introduced to describe the distribution of uncertainty, defined as:

$$\mathrm{UDR}_\alpha(\mathcal{D}_{test}) = \frac{1}{m}\sum_{i=1}^{m} \left|\mathbb{U}_\alpha(x_{n+i})\right|, \tag{8}$$

where UDR is computed as the ratio of uncertain cases detected, which implies the average uncertainty over the entire test dataset. UDR is always a number within the range of $[0, 1]$. By estimating the UDR of a dataset based on the fingerprinting model, the confidence level of the overall predictions can be acquired. Similar to APSS, a larger UDR indicates that a more uncertain prediction exists, and a smaller UDR shows more confidence in the system.

The Conformal Accuracy (CA) metric illustrates the accuracy of prediction when no uncertainty is detected. The CA is calculated as follows:

$$\mathrm{CA}_\alpha(\mathcal{D}_{test}) = \frac{\sum_{i=1}^{m} \mathbb{1}\left\{\hat{\mathcal{C}}_\alpha(x_{n+i}) = l_{true}^{n+i}\right\}}{\left|\mathbb{U}_\alpha(x_{n+i}) = 0\right|}, \tag{9}$$

where $|\cdot|$ is the function of cardinality. The CA metric only covers cases when the conformal prediction set size is equal to one, where no uncertainty is detected, and the CP system claims confident predictions.

IV. Experimental Study

A. Experiments Setup

To evaluate the proposed CP-based metrics, we set up the prototypes of fingerprinting-based localization systems in two typical indoor environments. First, a public dataset is adopted to investigate the proposed metrics. The dataset was collected on the third floor of the Engineering Office Wing (EOW) at the University of Victoria [19]. The indoor map is generated by a Turtlebot3 robot using the ROS2 SLAM toolbox. It covers a square corridor with a side of 20 meters. The dataset includes 102,998 BLE datapoints collected from 85 coordinates with 7 BLE routers. Each unique location in the dataset receives a class label. To simplify the experiment, we randomly selected 15% datapoints as the test dataset. The datapoints share the label with the closest training data. A 5-layer MLP model is trained as a classifier. For each layer, the output feature sizes are 256, 512, 128, and 85, respectively.

Second, we adopt a Radio Frequency Identification (RFID) scenario representing a small-scale setting, as shown in Fig. 3. In this scenario, we collect RSSI from 50 RFID channels by interrogating the 8 anchor RFID tags to create RFID fingerprints. Thus, the fingerprints are arrays of dimension $(8, 50)$. The testbed covers an area of $1.5m \times 1.5m$. Fig. 4 illustrates a corner of the testbed. The area is divided into grids with side lengths of 10 cm. Training fingerprints are gathered at the grid points, which are marked as green diamonds. Blue dots represent the test dataset, *dataset-hard*; all the blue dots are located in the center of the training fingerprints, and the

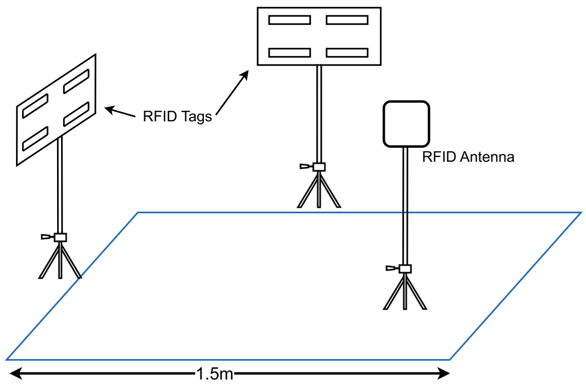

Fig. 3. The layout of the RFID testbed.

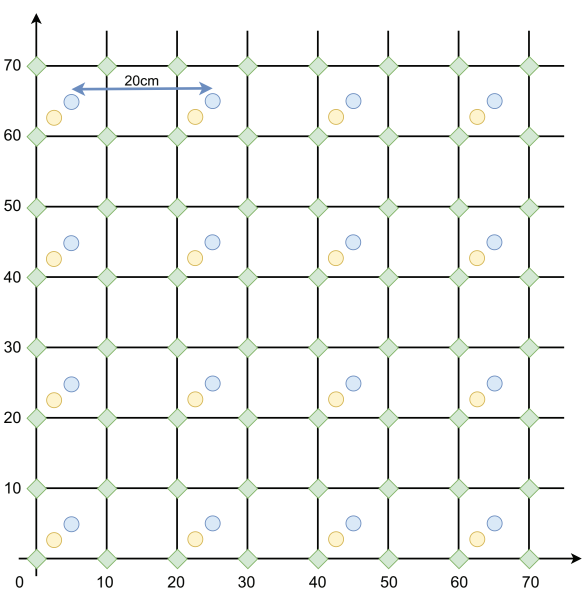

Fig. 4. Details of the RFID datasets.

TABLE I
RFID CLASSIFICATION MODEL

Layers	Input_features	Output_features	Kernel size
conv1D_0	8	3	1
conv1D_1	3	16	3
conv1D_2	16	32	3
linear_0	640	512	
linear_1	512	256	
linear_2	256	256	

gap between two adjacent blue dots is $20cm$. On the other hand, the test dataset, *dataset-easy*, is made up of the RSSI from the yellow dots that are near the grid points. In this experiment, three 1D convolution layers are leveraged first to adjust data size and extract features, then a three-layer MLP serves as a classifier for producing logits. The detailed network architecture is presented in Table I.

TABLE II
UNCERTAINTY ESTIMATION WITH THE PUBLIC DATASET

Error Tolerance Level α	APSS	UDR
0.05	1.07	0.12
0.10	1.02	0.06
0.15	0.94	0.07
0.20	0.84	0.15
0.50	0.59	0.40

TABLE III
UNCERTAINTY ESTIMATION WITH THE RFID DATASET

Error Tolerance Level α	Dataset Easy			Dataset Hard	
	APSS	UDR	CA	APSS	UDR
0.1	1.33	0.14	0.98	86.8	1.00
0.2	1.03	0.21	1.00	63.7	1.00
0.3	0.85	0.31	0.99	47.2	1.00
0.5	0.52	0.47	0.92	17.4	0.98
0.75	0.29	0.71	0.84	2.1	0.72

B. Experimental Results

To demonstrate the efficiency of our proposed CP-based method, experiments are firstly performed with the test data from the public dataset. The metrics of APSS and UDR are used with 5 different error levels ranging from 0.05 to 0.5. As Table II shows, for the error level $\alpha = 0.10$, the APSS is 1.02 and the UDR is 0.06. We consider this as the best scenario that achieves the highest efficiency for the proposed CP-based system. This is not only because the APSS is closest to 1, but also because that the lowest UDR indicates the least uncertainty detection. It is obvious that the prediction set size decreases and UDR increases when a larger α is prescribed. Therefore, the error rate $\alpha = 0.10$ is a suitable parameter that guarantees the $(1 - \alpha) = 0.9$ confidence level.

Table III illustrates the comparison of two different test datasets of the RFID scenario: dataset-hard and dataset-easy. These two datasets have different uncertainty levels depending on the sampling strategy. The dataset-hard covers the samples of measurement that represent hard cases that potentially increase the false negative rate. The dataset-easy contains the observations from the points that are relatively closer to certain labels compared with dataset-hard. It can be seen that the proposed method achieves the high efficiency for dataset-easy with error rates $\alpha = 0.10$ and $\alpha = 0.20$, referring to APSS of 1.33 and 1.03, and UDR of 0.14 and 0.21, respectively. However, the system efficiency for dataset-hard is extremely low when APSS is 86.8 and 63.7, and UDR is 1.00. This demonstrates the adaptability of the proposed method. Smaller APSS for easy cases, and larger APSS for hard inputs that show the underlying uncertainty. Table III also shows the Conformal Accuracy (CA) for dataset-easy of the RFID scenario. We observe that the it is initially high (>0.95) for small α values, and then it decreases when the error rate α is increased.

We further explore the uncertain predictions that contain multiple labels. Fig. 5 depicts three examples of corresponding

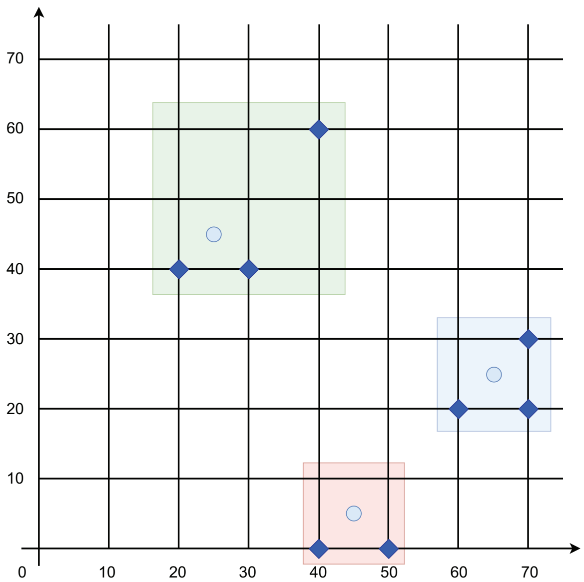

Fig. 5. Examples of the Uncertain Conformal Prediction Sets with the corresponding labels.

non-single prediction sets when uncertainty was reported in dataset-hard. The grid map covers the RFID testbed area of $0.7m \times 0.7m$, where the fingerprints are measured. Within each colored area, the circle marker represents the true label, and the diamond markers are the labels that form the prediction sets. Each example indicates a different level of uncertainty individually. In the bottom part of the grid map, the area in red presents the instance that a prediction set with the dimension of 2. There is a 5 cm displacement in the vertical direction for this prediction set. On the right side of the map, we observe the situation when three out of four nearest neighbor points are identified as the prediction set, which implies less uncertainty compared with the previous one. Nevertheless, more uncertain cases are also been discovered in the investigation. For instance, the green color area contains a predicted label at the top right corner, which is far away from the ground truth. Therefore, these measurements from ambiguous points express significant uncertainty for the final prediction results, which should be carefully evaluated when a reliable fingerprinting system is desired.

CONCLUSIONS

In this study, we presented a CP-based method to study the uncertainty of multi-class fingerprinting classification. The proposed method provides a statistical guarantee of accurate indoor localization predictions for a user-specified error tolerance level. With a few samples from the calibration dataset, uncertainty quantification is executed to assess the reliability of the system including the deep learning model and input fingerprints. Our experiments in two scenarios demonstrated that our method was effective for evaluating the intrinsic uncertainty based on the introduced metrics from the perspectives of efficiency and accuracy. Our future work shall further explore the potential of conformal prediction techniques, which is a promising tool for dealing with the challenging problem of dynamic environments with minimal effort.

ACKNOWLEDGMENTS

This work is supported in part by the NSF under Grants CCSS-2245607 and CCSS-2245608.

REFERENCES

[1] Q. D. Vo and P. De, “A survey of fingerprint-based outdoor localization,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 491–506, First-Quarter 2016.

[2] X. Wang, Z. Yu, and S. Mao, “DeepML: Deep LSTM for indoor localization with smartphone magnetic and light sensors,” in *Proc. IEEE ICC 2018*, Kansas City, MO, May 2018, pp. 1–6.

[3] M. Azizyan, I. Constandache, and R. Roy Choudhury, “SurroundSense: Mobile phone localization via ambience fingerprinting,” in *Proc. ACM MobiCom’09*, Beijing, China, Sept. 2009, pp. 261–272.

[4] H. Xie, T. Gu, X. Tao, H. Ye, and J. Lu, “A reliability-augmented particle filter for magnetic fingerprinting based indoor localization on smartphone,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 1877–1892, Aug. 2016.

[5] Z. Zhao, J. Wang, X. Zhao, C. Peng, Q. Guo, and B. Wu, “NaviLight: Indoor localization and navigation under arbitrary lights,” in *Proc. IEEE INFOCOM 2017*, Atlanta, GA, May 2017, pp. 1–9.

[6] Z. Li and X. Rao, “Toward long-term effective and robust device-free indoor localization via channel state information,” *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3599–3611, Mar. 2022.

[7] W. Wang, X. Wang, and S. Mao, “Deep convolutional neural networks for indoor localization with CSI images,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 316–327, Jan./Mar. 2020.

[8] S. A. Junoh and J.-Y. Pyun, “Enhancing indoor localization with semi-crowdsourced fingerprinting and GAN-based data augmentation,” *IEEE Internet of Things Journal*, vol. 11, no. 7, pp. 11 945–11 959, Apr. 2024.

[9] S. Dai, L. He, and X. Zhang, “Autonomous WiFi fingerprinting for indoor localization,” in *Proc. IEEE ICCPS 2020*, Sydney, Australia, Apr. 2020, pp. 141–150.

[10] M. Nabati, H. Navidan, R. Shahbazian, S. A. Ghorashi, and D. Windridge, “Using synthetic data to enhance the accuracy of fingerprint-based localization: A deep learning approach,” *IEEE Sensors Letters*, vol. 4, no. 4, pp. 1–4, Apr. 2020.

[11] Y. Tian, J. Wang, and Z. Zhao, “Wi-Fi fingerprint update for indoor localization via domain adaptation,” in *Proc. IEEE ICPADS 2021*, Beijing, China, Dec. 2021, pp. 835–842.

[12] M. Cauchois, S. Gupta, and J. C. Duchi, “Knowing what you know: Valid and validated confidence sets in multiclass and multilabel prediction,” *Journal of Machine Learning Research*, vol. 22, no. 1, pp. 81:1–42, Jan. 2021.

[13] V. Vovk, “Conditional validity of inductive conformal predictors,” in *Proc. Asian Conference on Machine Learning*, Singapore, Nov. 2012, pp. 475–490.

[14] M. M. Campos, A. Farinhas, C. Zerva, M. A. Figueiredo, and A. F. Martins, “Conformal prediction for natural language processing: A survey,” *arXiv preprint arXiv:2405.01976*, May 2024.

[15] A. N. Angelopoulos and S. Bates, “A gentle introduction to conformal prediction and distribution-free uncertainty quantification,” *arXiv preprint arXiv:2107.07511*, Dec. 2022.

[16] J. Su, J. Luo, H. Wang, and L. Cheng, “API is enough: Conformal prediction for large language models without logit-access,” *arXiv preprint arXiv:2403.01216v2*, Apr. 2024.

[17] L. Lindemann, M. Cleaveland, G. Shim, and G. J. Pappas, “Safe planning in dynamic environments using conformal prediction,” *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 5116–5123, Aug. 2023.

[18] G. Shafer and V. Vovk, “A tutorial on conformal prediction,” *Journal of Machine Learning Research*, vol. 9, no. 3, pp. 371–421, Mar. 2008.

[19] B. Yuen, Y. Bie, D. Cairns, G. Harper, J. Xu, C. Chang, X. Dong, and T. Lu, “Wi-Fi and Bluetooth RSSI SQI indoor localization.” IEEE Dataport, May 2024. [Online]. Available: https://ieee-dataport.org/documents/wi-fi-and-bluetooth-rssi-sqi-indoor-localization