Conditional Probability-based Ensemble Learning for Indoor Landmark Localization

Zhongliang Zhao[a,*], José Luis Carrera V.[a], Torsten Braun[a], Zhiyang Pan[b]

[a]*Institute of Computer Science, University of Bern, Switzerland*
[b]*R&D, China Power Construction Engineering Consulting, Beijing, China*

Abstract

Indoor location awareness enables many location-based services, such as smart homes or smart offices. The huge amount of sensor data collected by nowadays' smartphones provides a solid basis for applying advanced machine learning algorithms to derive the correlation between indoor locations and sensor measurements. The combination of multiple sensor measurements, such as the Received Signal Strength of surrounding Wi-Fi access points and magnetic fields, is assumed to be unique in many locations, which can be derived to accurately predict smartphones' indoor locations. In this work, we propose a novel ensemble learning method to provide room level indoor localization in smart buildings. The proposal is based on a conditional probability model, which combines prediction results of multiple individual machine learning predictors using conditional probability concepts to predict class labels. We have implemented the system on Android smartphones and conducted extensive experiments in real-world office-like environments. The experiment results show that the proposed ensemble predictor outperforms individual and ensemble voting-based machine learning algorithms. It achieves the best indoor landmark localization accuracy of nearly 97% in office-like environments. This work provides a coarse-grained indoor room recognition, which can be envisioned as a basis for accurate indoor positioning.

*Corresponding author
Email address: `zhao@inf.unibe.ch` (Zhongliang Zhao)

1. Introduction

High localization accuracy within buildings would be very useful. In particular, large complex buildings like shopping malls, airports and hospitals would be well served by this feature. It would make orientation within these highly complicated structures much easier and would diminish the need for big floor maps scattered all around these buildings. However, walls, roofs, windows and doors of the buildings significantly reduce the GPS signals carried by radio waves, which leads to a severe accuracy loss of GPS inside buildings.

In contrast to outdoor environments, building interiors are normally covered by a large number of Wi-Fi access points that constantly emit signals. By scanning the area around a device, we can measure the received signal strength of each of the nearby access points, and we can assume that the list of all these values combined is unique at every distinct point in the building. Furthermore, we can assume that these values are constant over time when the Wi-Fi access points are fixed in place and are constantly emitting signals with the same signal strength. These assumptions are subject to change when indoor environments are modified. Different solutions already exist for indoor localization of mobile devices such as Pedestrian Dead Reckoning (PDR) and Wi-Fi fingerprinting based methods [1] [2]. In PDR the future location of a smartphone user is predicted based on the estimated current location, and the movement information is derived from inertial sensor measurements. In Wi-Fi fingerprinting, the Received Signal Strength (RSS) values of several Wi-Fi access points in range are collected and stored together with the coordinates of the location. A new set of RSS values is then compared with the stored fingerprints and the location of the closest match is returned.

With smartphones, it is easy to collect a lot of sensor data that provide a solid basis to apply advanced Machine Learning (ML) algorithms to find the correlations between indoor locations and smartphone sensor measurements. Therefore, we propose to use supervised machine learning methods to process the large amount of collected data. By training a classifier (supervised learning algorithm such as K-Nearest-Neighbor) on the collected labeled data, rules can be extracted. Feeding in the actual live data (RSS values, magnetic field values, illuminance level, etc.) of a moving user, the trained classifier can then predict the user's location in a coarse-grained level. We propose to apply existing machine learning methods (including both individual and ensemble predictors) and to develop novel conditional probability model-based ensemble predictors to solve this task due to the large amount of features that are available in indoor environments, such as Wi-Fi RSS values, magnetic field values, illuminance level, etc. In our previous work [13], we have validated the performance of different ML algorithms to recognize indoor landmarks. The results show that the Voting ensemble predictor outperforms individual machine learning algorithms and it achieves the best indoor landmark localization accuracy of 90% in office-like environments.

Whereas our previous work applied existing ML algorithms to recognize indoor locations [13], this work presents a novel ensemble machine learning algorithm to achieve high and stable room prediction performance. Compared to [13], the contribution of this article can be summarized as follows:

- We propose a novel ensemble learning predictor by combining conceptually different individual machine learning algorithms to predict class labels (i.e., rooms, landmarks). This combination is made by applying concepts of conditional probability and evidences about the prediction performance of individual predictors.
- We conduct experiments in a large indoor office environment that covers an entire floor, which has the double size of the experiment areas in [13].
- Experiment results show that our new algorithm outperforms other predictors, including our previous work [13].

The rest of the paper is organized as follows. Section 2 presents some related work in indoor localization and landmark detection. Section 3 describes our ensemble machine learning method and the used machine learning models.

Section 4 presents implementation and experiment details. Section 5 discusses the performance results of our approach. Section 6 concludes the paper.

2. Related Work

Various machine learning-based approaches have been proposed that use fingerprinting to estimate user indoor locations. Machine learning approaches can be classified into generative or discriminative methods. Generative methods build the machine learning model using a joint probability, while discriminative methods build the machine learning model with a conditional probability [1] [2]. K-Nearest-Neighbor (KNN) is the most basic and popular discriminative technique. Based on a similarity measure such as a distance function, the KNN algorithm determines the k closest matches in the signal space to the target. Then, the location of the target can be estimated by the average of the coordinates of the k neighbors [3]. Generative localization methods apply statistical approaches, e.g., Hidden Markov Model [4], Bayesian Inference [5], Gaussian Processes [6], on the Wi-Fi fingerprint database. Thus, the accuracy can obviously be improved by adding more measurements. In [6] for instance, Gaussian Processes are used to estimate the signal propagation model through an indoor environment. There is a limited number of works that have focused on reducing off-line efforts in learning-based approaches for indoor localization [7] [8] [9]. These approaches reduce the off-line effort by reducing either the number of samples collected at each survey point or the number of survey points or both of them (i.e., reducing the number of collected samples and number of survey points). Then, a generative model is applied to reinforce the sample collection data. In [7] for instance, a linear interpolation method is used. In [8], a Bayesian model is applied. In [9], authors propose a propagation method to generate data from collected samples. In [2], authors combine characteristics of generative and discriminative models in a hybrid model. Although this hybrid model reduces offline efforts, it still relies on a number of samples collected from fixed survey points (i.e., labeled samples). To maintain high accuracy, the number of survey points shall be increased in larger environments. Collecting samples from numerous survey points will become a demanding process, which makes the system unsuitable to large environments. In [10], authors validated the performance of different individual machine learning approaches for indoor positioning systems. However, they rather compare the results without any deep analysis of the performance difference. Moreover, they did not discuss how ensemble learning approaches could be used to enhance the system performance.

Recently, there are some efforts spent to improve the indoor localization accuracy by fusing multiple types of sensor information. In [14], authors proposed a system to improve the accuracy of Wi-Fi based localization by fusing information from alternate sources like LTE signals and magnetometers, collected through software defined radios and smart-devices. In [15], authors introduced a new indoor localization technique using off-the-shelf 802.11n multiple antennas Access Points (APs), which achieved better results by using fusion between multiple APs, instead of using only one AP. Authors of [17] designed a positioning system using narrow-band signals, particularly ZigBee signals, based on an enhanced fingerprinting algorithm by fusing received signal strength (RSS) and time information. However, few studies clearly show the impacts and importance of different sensor information. In this work, we compare the indoor localization performance of using the combination of different sensor data and show the efficiency of combining multiple sensor inputs.

3. Machine Learning Algorithm for Indoor Landmark Localization

An indoor landmark is defined as a small area within a room. The aim of the indoor landmark localization system presented in this work is to improve the accuracy of indoor landmark recognition using machine learning approaches. This section describes the details of our proposed conditional probability-based ensemble learning predictor, so called Conditional Performance Ensemble Learning Method (COND) approach.

3.1. Conditional Performance Ensemble Learning Method (COND)

It has been proven that ensemble learning methods can achieve better performance than from individual learning algorithms alone. Thus, we focus on developing an ensemble learning method that is based on the performance of its constituent learning methods. The proposed ensemble learning method is based on the concept of conditional probability. In probability theory, conditional probability is a measure of the probability of an event considering some evidences. For the real-time landmark localization problem, we define an event as an object that is located near a landmark. Moreover, we define an evidence as the outcome provided by some machine learning landmark prediction model (i.e., machine learning algorithm). Thus, the probability of being located at a landmark considering some evidences can be written as follows:

$$P(c_i \mid l_1, l_2, ..., l_n) = \frac{P(l_1, l_2, ..., l_n \mid c_i) \cdot P(c)}{P(l_1, l_2, ..., l_n)}, \tag{1}$$

where c_i is the landmark identifier (i.e., the class) and l_i is the *i-th* evidence provided by the *i-th* machine learning prediction model.

Equation 1 can be solved assuming conditional independence of events l_i given the event c_i. Conditional independence means that if some piece of information is known, the probability of other events become independent. For the real-time landmark localization problem, our assumption is that the probability of obtaining the outcome l_i becomes independent if the value of c_i is known. Thus, Equation 1 can be written as follows:

$$P(c_i \mid l_1, l_2, ..., l_n) = \frac{P(l_1 \mid c_i) \cdot P(l_2 \mid c_i) \cdot ... \cdot P(l_n \mid c_i) \cdot P(c_i)}{P(l_1, l_2, ..., l_n)} \tag{2}$$

Considering that individual predictors are independent from each other,

$$P(c_i \mid l_1, l_2, ..., l_n) = \frac{P(l_1 \mid c_i) \cdot P(l_2 \mid c_i) \cdot ... \cdot P(l_n \mid c_i) \cdot P(c_i)}{P(l_1) \cdot P(l_2) \cdot ... \cdot P(l_n)} \tag{3}$$

Therefore, the prediction of the landmark z can be calculated as follows:

$$z = \arg\max_{c} \frac{P(c) \cdot \prod_{i=1}^{n} P(l_i \mid c)}{\prod_{i=1}^{n} \sum_{j=1}^{m} P(l_i \mid c_j) \cdot P(c_j)} \tag{4}$$

where z represents the predicted class, n is the number of evidences given by n machine learning individual predictors, m is the number of landmarks (i.e, classes), and $P(c_j)$ is the initial emission probability. The initial emission probability is the likelihood of being located inside landmark c_i when the localization process starts.

3.2. Conditional Emission Probabilities

We define as emission probability the likelihood of observing a particular outcome l_n given that event c_i has happened. Therefore, the set B containing the emission probabilities $P(l_i \mid c_i)$ can be written as follows:

$$B = \{P(l_n \mid c_i)\}, \forall l_n \wedge \forall c_i \in Z, \tag{5}$$

where Z is the set of landmarks (i.e., classes) to be predicted. Therefore, $P(l_n \mid c_i)$ represents the sensitivity of the individual learning method n when the target is located inside the landmark i. Thus, $P(l_n \mid c_i)$ can be written as follows:

$$P(l_n \mid c_i) = \frac{P_{ln}}{P_{ln} + N_{ln}}, \tag{6}$$

where P_{ln} is the number of outcomes equal to ln and N_{ln} is the number of outcomes different to ln when the actual class is c_i. Both, P_{ln} and N_{ln} can be computed from the confusion matrix of each machine learning predictor.

3.3. COND Architecture

Figure 1 presents an overview of the COND learning method from a system view. The key idea of COND is to combine conceptually different individual machine learning models to predict indoor landmarks. This combination is made by applying conditional probability concepts and information about the individual prediction performance of each machine learning model. Thus, the outcome of each individual machine learning predictor can be regarded as an evidence l_i in COND. The prediction performance given the knowledge of the ground-truth class label defines the probability of the occurrence of an evidence given the ground-truth label class $P(l_i \mid c)$. From the system view, COND implements Conditional Performance Tables (CPT) to store $P(l_i \mid c)$ of each individual machine learning model. As explained in Section 3.1, $P(l_i \mid c)$ and l_i are used in the prediction process of COND. Thus, by considering the performance of each individual learning model, COND balance out their strengths and weaknesses to improve the prediction performance.

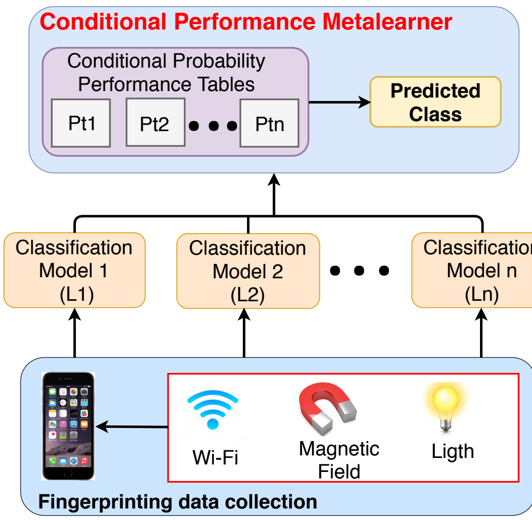


Figure 1: Conditional Performance Learning Method (COND) system architecture.

3.4. Features

In a machine learning-based classification task, the attributes of the classes are denoted as features. Each feature describes an aspect of the classes. In our case features are our measurements, e.g., an RSS value. To deliver good machine learning prediction accuracy it is very important to select the right attributes/features and to also modify certain features or even create new features out of existing features.

3.4.1. Wi-Fi RSS

These values provide the core data as they contribute the most to the performance of the ML methods. The smartphone scans the surrounding Wi-Fi access points and registers the RSS values of each access point. Wi-Fi RSS values depend on the distance between smartphone and access points.

3.4.2. Magnetic Field (MF)

The device's sensors measure the magnetic field in the device's coordinate system. As the user walks around, the orientation of the device may change all the time. Therefore, we have to collect all possible values from every orientation in every point in the training phase. This would result in a huge amount of data and the training performance would be inaccurate.

3.4.3. Light

Light sensors might also be helpful to identify rooms. For instance, a room facing a window will clearly be brighter than one surrounded by walls only. As shown in Section 4.3 this does improve the prediction accuracy. However, these assumptions are not stable, as the illuminance level might change over time. Therefore, it is better to work with light differences instead of absolute values.

4. Implementation, Experiments, and Results

This section explains the system implementation on Android smartphones and describes the details of the experiments conducted in an office environment.

4.1. Implementation

Figure 2 shows the data flow and the different components of our system. Sensor and Wi-Fi RSS values are measured by the smartphone and are received by the app. We then perform the data training process offline in a PC to pass the collected data to the Model Building component, which applies different machine learning algorithms to build the models. The built models are then optimized and transferred to the app on the smartphone for online experiments.

Considering that the landmark detection accuracy can be influenced by environmental parameters, we conduct some experiments to determine how parameters such as AP position or number of APs influence the accuracy of the Wi-Fi-based fingerprinting landmark detection approach. Additionally, we perform experiments to show how accuracy is improved by considering additional features such as magnetic field (MF) values and light illuminance level readings. As shown in Figure 3, we define 14 wall separated areas in our experiments.

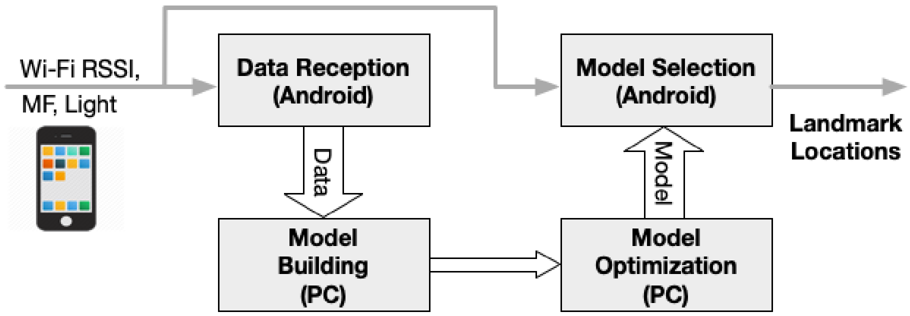


Figure 2: The architecture of the implemented Android app.

Hereafter, we refer to these areas as rooms. In our experiments, we do not need to know the exact locations of the Wi-Fi APs, while only the fingerprints of Wi-Fi RSSI, MF readings, and illuminance level readings are recorded during the data collection procedure.

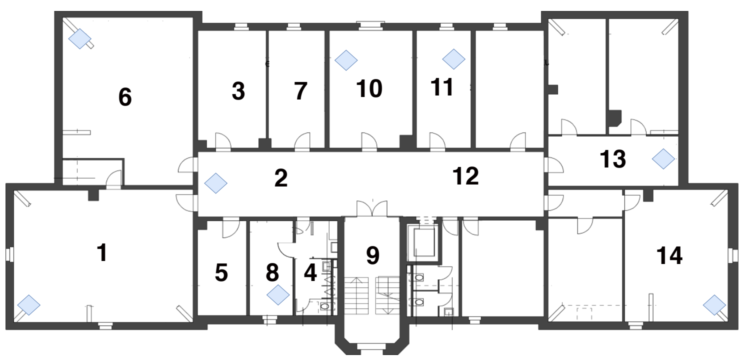


Figure 3: Experiment scenario, zone definition and ANs distribution (Diamond blue points: Wi-Fi APs as Anchor Nodes).

Parameters of learning-based algorithms are optimized from training data. Additionally, certain algorithms also have parameters that are not optimized during the training process. These parameters are called hyperparameters, which have significant impact on the performance of the learning-based algorithm. Therefore, we use a nested cross validation technique to adjust them. The nested cross validation technique defines inner and outer cross validation. Inner cross validation is intended to select the model with optimized hyperparameters, whereas outer cross validation is used to obtain an estimation of the

Table 1: Classifier's Hyperparamenters.

Classifier	Hyperparameter	Value/Option
KNN	G.B. Percent ratio	30%
SVM	Kernel	Polynomial single order
	Penalty parameter	1
	Kernel coefficient	0
MLP	H. Layers	Single
	Neurons	10
Bagging	Base estimator	CART
	N. Estimators	10
	Bootstrap	True

generalization error. Ten-fold cross validation was applied on both inner and outer cross validation. The classifiers were optimized over a set of hyperparameters. We optimized the global blend percentage ratio hyperparameter for KNN, the kernel type function for SVM, and the number of hidden layers and neurons per layer for MLP. Based on the parameter optimization process, we established the optimal hyperparameter values for the classifiers as follows: global blend percent ratio of 30% for KNN, single order polynomial kernel, $c = 1$, and $\gamma = 0.0$ for SVM, and single hidden layer with 10 neurons for MLP. Table 1 shows the final settings of all the relevant hyperparameters derived from the nested cross validation technique.

4.2. Experiments

To test the room landmark detection performance, we performed experiments on the third floor of the Computer Science building at the University of Bern, as shown in Fig. 3. During the experiments, we collected 17569 data points in total, which were equally distributed along the whole area in each room. Collecting the training dataset takes around 75 minutes. With the collected data, we build models with different data: the first one builds the fingerprint using only collected Wi-Fi RSS data, the second one using Wi-Fi RSS together with MF readings, and the third one using Wi-Fi RSS, MF readings, and illuminance level readings. Since the number of detected APs can differ on each landmark, we set the RSSI value of the AP that was not detected to 0 .

Thus, the length of the fingerprint remains constant, and the 0 RSSI value is used as part of the fingerprint.

To build the landmark fingerprint database, a person walks randomly around each room holding the phone in his/her hand. The data collection rate is only constrained by computational capabilities of the smartphones' Wi-Fi interface. Thus, in our experiments every data measurement was collected at a rate of 3 entries/second. Since our approach does not need any survey point, the time needed to build the landmark fingerprint database is proportional to the number of collected instances multiplied by the instance collection rate. Therefore, our approach needs much less time to build the landmark fingerprint database than traditional fingerprint-based solutions. To better support open dataset in the research of indoor localization, we have made our dataset openly accessible [18].

4.3. Results

This section presents and discusses results of the landmark detection model when different classifiers and features are used. To extend our previous work, we performed experiments in a bigger area compared to [13]. However, to test the performance of our approach with different environment configurations, we have divided our area of interest into two scenarios. We defined as landmark to each wall separated subarea inside the are of interest. Thus, each room corresponds to a landmark. Hereafter, we refer to room as landmark. We defined the left part of Figure 3 as scenario 1, which covers Zones 1-8. The right part of Figure 3 is defined as scenario 2, which includes Zones 9-14. For performance comparisons, we include five individual predictors: Classification and Regression Tree (CART) [16], Support Vector Machine (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Naive Bayes (NB), and two ensemble meta-classifiers: Soft Voting (SV) and Bagging. When comparing machine learning prediction performance, it is impossible to define a single metric that provides a fair comparison in all possible applications. In this work, we focus on the metrics of prediction accuracy, which refers to the rate of correct landmark detection.

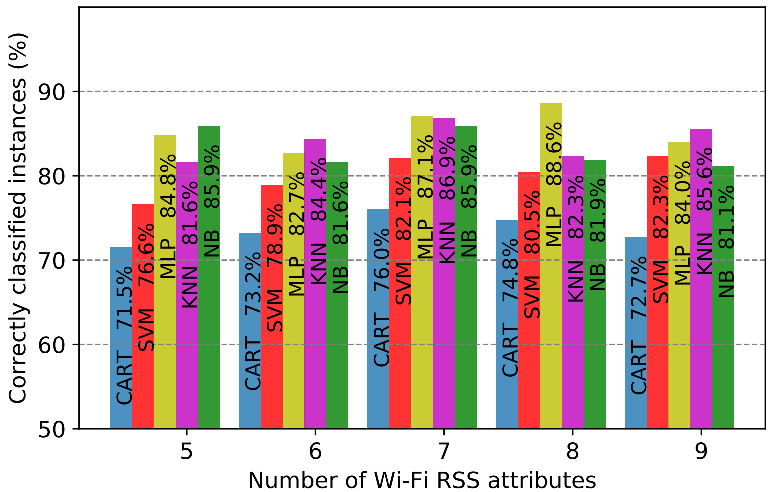


Figure 4: Individual predictors landmark prediction performance with different numbers of Wi-Fi RSS values in scenario 1.

4.3.1. Indoor Landmark Localization Accuracy

First, we only use Wi-Fi RSS values as inputs to machine learning algorithms in scenario 1. Figure 4 shows the classification accuracy of different individual predictors when different numbers of Wi-Fi RSS values are available and used. As we can see, starting from 5 RSS values, more RSS inputs increase the prediction accuracy for most of the predictors. Nevertheless, after 7 Wi-Fi RSS values are used, the improvement of adding more RSS values is almost negligible in almost all individual tested classifiers, and some of the predictors even got reduced accuracy when additional RSS values are considered. We think that the signal interference caused by additional Wi-Fi APs may be the reason for the worse performance when more than 7 Wi-Fi RSS values are utilized. Therefore, we take 7 Wi-Fi RSS values as the default configuration for the following experiments.

Next, we compare the classification accuracy when using only Wi-Fi RSS, Wi-Fi RSS plus MF, and Wi-Fi RSS plus MF and illuminance levels. Figure 5 shows the performance evaluation of the selected individual classifiers obtained with different feature combinations in scenario 1. The best performance is reached by the NB predictor, which classifies 90.3% of instances correctly when the fingerprint is composed by Wi-Fi RSS, MF readings, and illuminance levels.

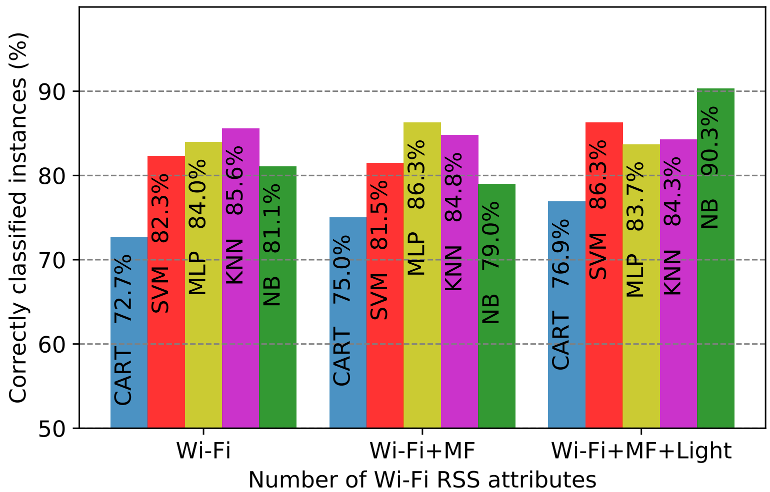


Figure 5: Individual predictors landmark prediction performance when using different features in scenario 1.

Hyperparameters have significant impact on the performance of the learning-based algorithm. Figure 6 shows the performance prediction of the selected individual and ensemble meta-classifiers in scenario 1. Figure 7 presents the performance prediction of the selected individual and ensemble meta-classifiers in scenario 2. The individual classifiers were set up with the hyperparameters optimized. The ensemble predictors use the outcomes of these individual classifiers as inputs. The classifiers are all fed with Wi-Fi RSS plus MF and illuminance levels. All the individual classifiers have improved performance, and NB even reaches an accuracy of 90.3%. The Bagging classifier uses a base classifier with random subsets of the original testing dataset. Then, it aggregates the individual predictions to determine a final prediction. We set up CART as base classifier for Bagging. Soft Voting uses the average predicted probabilities of CART, MLP, NB, KNN, and SVM to predict the room. Soft Voting (SV) can reach an accuracy of 87.7% in scenario 1 and 96.1% in scenario 2. Although all the tested traditional classifiers (i.e., CART, MLP, NB, KNN, SVM, Voting and Bagging) show high prediction accuracy, our proposed COND method outperforms them in both scenarios. COND achieves a prediction accuracy of 96.6% in scenario 1 and 96.8% in scenario 2.

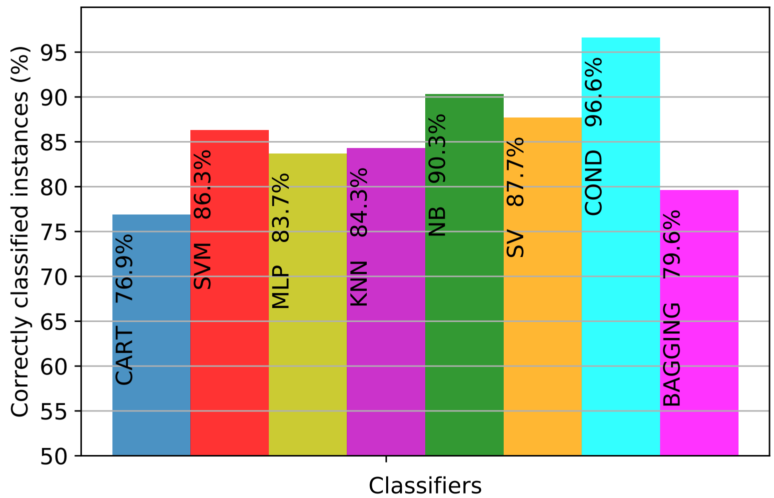


Figure 6: Landmark prediction performance of individual predictors with optimized hyperparameters and ensemble meta-classifiers in scenario 1.

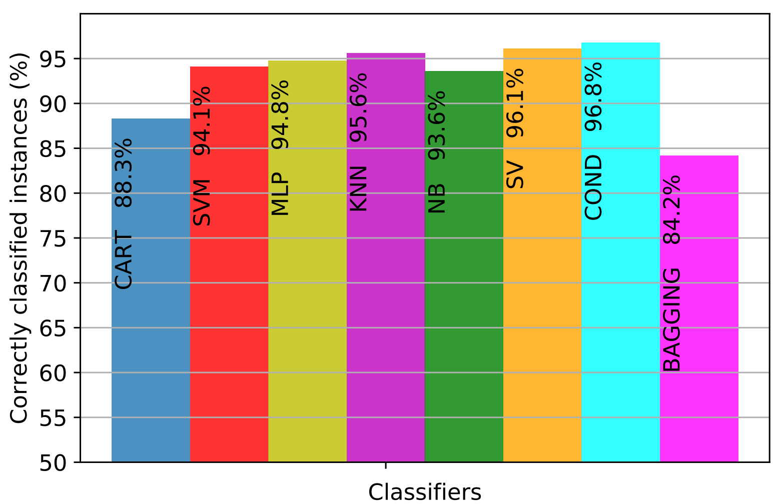


Figure 7: Landmark prediction performance of individual predictors with optimized hyperparameters and ensemble meta-classifiers in scenario 2.

4.3.2. Result Analysis

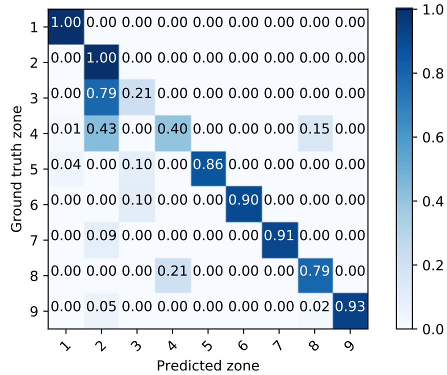
In indoor environments, Wi-Fi RSS and MF measurements vary according to locations. However, these values will remain similar on nearby positions. For example, on locations close to landmark borders, high similarities will be observed on the RSS values. These similarities could lead to misclassification problems. From Figure 5 we can see that KNN and MLP have better accuracy when both Wi-Fi RSS and MF readings are used. This is because KNN is an instance-based predictor, which uses entropy as a distance measure to determine how similar two instances are. Thus, this method is more sensitive to small variations upon the instance as unity. Since CART is a decision tree machine learning algorithm, it builds the classification model by parsing the entropy of information on attribute level. It means that CART measures entropy in the attribute domain to decide which attribute should be included. Thus, the classification model is prone to misclassification in this room prediction problem. When the illuminance level is included as input feature to predictors, Naive Bayes outperforms others. This is because the feature of illuminance level is completely independent from other radio signal measurements, which fits with Naive Bayes' strong assumptions about the independence of each input variable.

To further explain how the individual and ensemble predictors perform in scenarios 1 and 2, we show the confusion matrix of room recognition using MLP, Naive Bayes (NB), KNN, SV, and COND in Figures 8 and 9. We can observe that zone 3 is identified with accuracies of 21% by MLP, 100% by NB, 18% by KNN, 39% by Voting and 98% by COND. As a consequence, NB seems to be better in predicting zone 3 as compared to other individual and ensemble predictors. However, considering zone 4, NB achieves only 48% of accuracy, whereas KNN achieves 66%. We can see that the ensemble predictors adopt behaviors of different individual predictors. For instance, they adopt the good behavior of MLP and Naive Bayes, which leads to a much better prediction accuracy for zone 2. Unfortunately, ensemble predictors still have problems in some zones. For instance, SV achieves only 39% of instances correctly classified in zone 3. It is because most of the individual predictors that contribute to SV achieve low performance in that zone. It can be observed that in scenario 2 the SV ensemble predictor improves the accuracy compared to its constituent individual predictors. However, in scenario 1 the SV predictor does not achieve better performance than all its constituent individual predictors. It is because the prediction process of SV can be strongly influenced by individual predictors with low prediction performance. The performance of the Bagging classifier is affected by the performance of its base classifier. Unlike Bagging, the COND classifier relies on the performance of several individual predictors. As result, COND overcomes Bagging by 17.6% and 13.0% in scenario 1 and scenario 2 respectively. Since COND is based on the conditional probabilities of its constituent individual predictors, COND prediction is strongly influenced by individual predictors that show better performance in each scenario. Therefore, despite the different physical set up of scenario 1 compared to scenario 2, COND shows higher performance than the others tested predictors. Thus, we prove the generality of COND with different set of inputs.

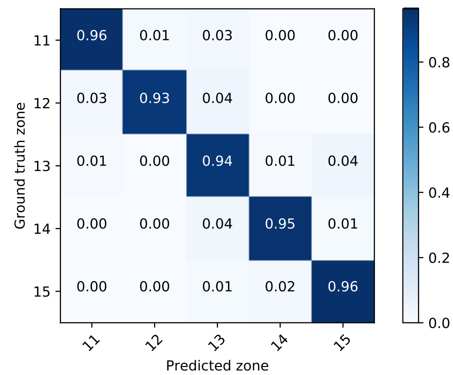
Our proposed COND method is able to outperform SV and all the individual tested predictors in both tested scenarios. Although COND and SV have the same constituent individual predictors, COND is able to perform better than SV. As it can be seen in Figure 9, COND overcomes SV in all zones. For instance, in zone 4, SV achieves 63% and COND 96% of correctly classified instances. This is because COND is able to balance out strengths and weakness of its constituent algorithms. This balance is made based on the observed prediction performance of each constituent classifier. Thus, we prove that COND is able to predict zones more reliably than the other tested methods. Our prediction model allows the production of better zone prediction performance compared to individual and ensemble voting models.

5. Conclusions

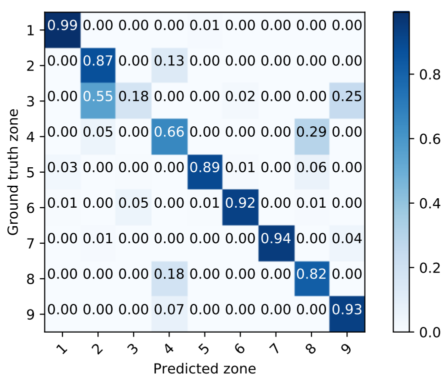
This work analyzes the performance of five traditional individual predictors and one ensemble predictor. Additionally, we propose a novel ensemble learn-



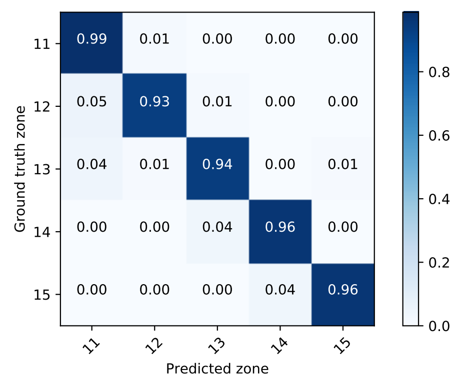
(a) MLP performance scenario 1.



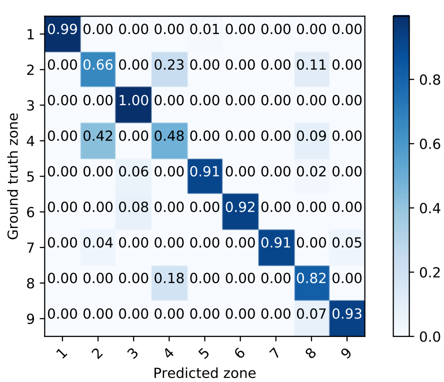
(b) MLP performacne scenario 2.



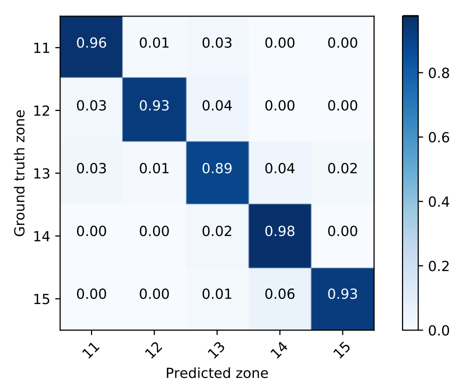
(c) KNN performance scenario 1.



(d) KNN performance scenario 2.

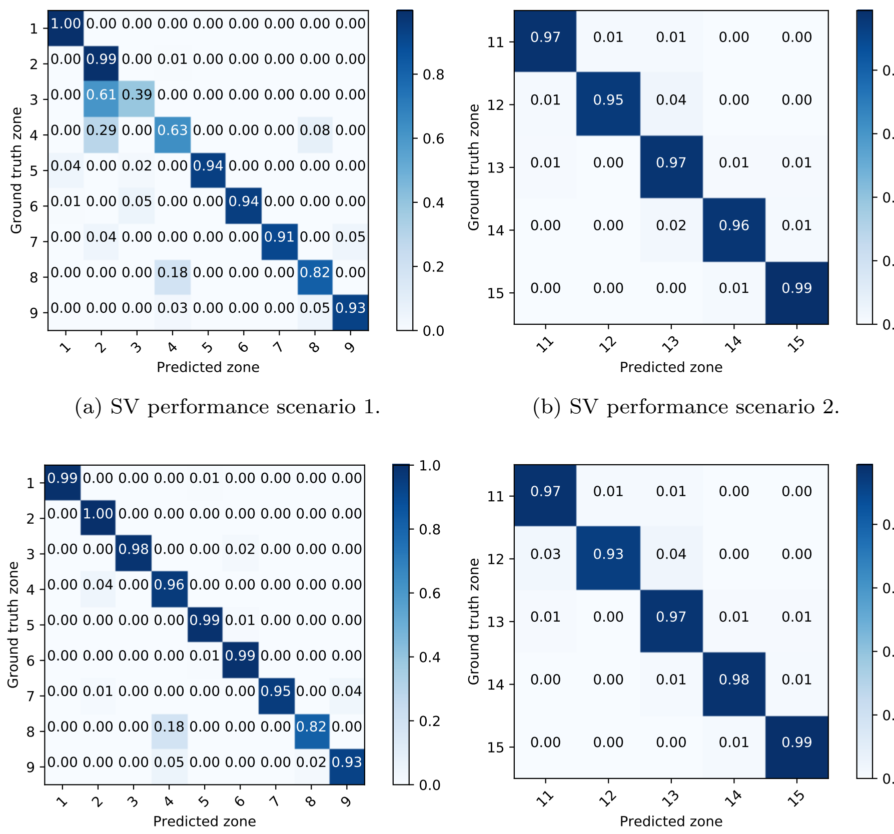


(e) NB performance scenario 1.



(f) NB performance scenario 2.

Figure 8: Individual predictors normalized confusion matrix in scenario 1 and 2.



(a) SV performance scenario 1.

(b) SV performance scenario 2.

(c) COND performance scenario 1.

(d) COND performance scenario 2.

Figure 9: Ensemble meta-predictors normalized confusion matrix in scenario 1 and 2.

ing algorithm, which is based on the concept of conditional probabilities. We tested prediction performance of these seven predictors in distinguishing zones on a floor. We have validated the performance of the system using different smartphone sensor measurements, such as Wi-Fi RSS, MF readings, and illuminance levels. Evaluation results show that the our proposed ensemble predictor COND achieves the best indoor landmark localization accuracy of 96.8%. To test the generality of our approach, we have divided our area of interest into two scenarios. This allows us to test our approach with two different environmental configurations. We show that despite the different physical set up, COND shows higher performance than the others tested predictors. Thus, we prove the generality of COND with different sets of inputs. To support open dataset for indoor localization research, we make parts of our experiment data openly accessible such that others can make fair performance comparison.

6. Acknowledgements

This work was supported by the Swiss National Science Foundation project #154458. The authors would like to acknowledge Joel Niklaus for sharing the dataset.

References

[1] S. He and S. Chan *Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons.* IEEE COMMUNICATIONS SURVEYS & TUTORIALS, 2016.

[2] R. W. Ouyang and A. K. S. Wong and C. T. Lea and M. Chiang *Indoor Location Estimation with Reduced Calibration Exploiting Unlabeled Data via Hybrid Generative/Discriminative Learning.* IEEE Transactions on Mobile Computing, 2012.

[3] B. Lakmali and W. Wijesinghe and K. de SIva and K. Liyanagama and S. Dias *Design, Implementation & Testing of Positioning Techniques in Mobile*

Networks. The 3rd International Conference on Information and Automation for Sustainability.

[4] Z. Ghahramani *An Introduction to Hidden Markov Models and Bayesian Networks.* International Journal of Pattern Recognition and Artificial Intelligence, 2001.

[5] O. Bousquet and U. von Luxburg and G. Ratsch *Bayesian Inference: An Introduction to Principles and Practice in Machine Learning.* Fresenius Environmental Bulletin 20(5), 2004.

[6] B. Ferris and D. Hahnel and D. Fox *Gaussian Processes for Signal Strength-based Location Estimation.* Procedures of Robotics Science and Systems, 2006.

[7] X. Chai and Q. Yang, *Reducing the calibration Effort for Probabilistic Indoor Location Estimation.* IEEE Tran. Mobile Computing, vol 6, no 6, pp. 649-662, 2016.

[8] D. Madigan and E. Einahrawy and R. Martin and W. Ju and P. krishnan and A. Krishnakumar, *Bayesian Indoor Positioning Systems.* IEEE. INFOCOM, vol 2, pp. 1217-1227, 2005.

[9] S. Liu and H. Luo and S. Zou, *A Low-Cost and Accurate Indoor Localization Algorithm Using Label Propagation based Semi Supervised Learning.* Fifth International Conference Mobile Ad-Hoc and Sensor Networks, pp. 108-111, 2009.

[10] D. Mascharka and E. Manley, *LIPS: Learning Based Indoor Positioning System using mobile phone-based sensors* 2016 13th IEEE Annual Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, 2016, pp. 968-971.

[11] J. Carrera and Z. Zhao and T. Braun and Z. Li and A. Neto, *A real-time robust indoor tracking system in smartphones.* Elsevier Journal of Computer Communications. DOI: http://dx.doi.org/10.1016/j.comcom.2017.09.004

[12] L. Buitinck and G. Louppe and M. Blondel and F. Pedregosa and A. Mueller and O. Grisel and V. Niculae and P. Prettenhofer and A. Gramfort and J. Grobler and R. Layton and J. VanderPlas and A. Joly and B. Holt and G. Varoquaux, *A*PI design for machine learning software: experiences from the scikit-learn project. PKDD Workshop: Languages for Data Mining and Machine Learning, pp. 108–122, 2013

[13] Z. Zhao and J. Carrera and J. Niklaus and T. Braun, *Machine Learning-based Real-Time Indoor Landmark Localization.* 2018 16th International Conference on Wired/Wireless Internet Communications (IFIP WWIC).

[14] M. Di Felice, C. Bocanegra and K. R. Chowdhury, *WI-LO: Wireless indoor localization through multi-source radio fingerprinting.* 2018 10th International Conference on Communication Systems Networks (COMSNETS), Bengaluru, 2018, pp. 305-311.

[15] A. M. Shata, S. S. A. El-Hamid, Y. A. Heiba and O. A. Nasr, *Multi-site fusion for WLAN based indoor localization via maximum discrimination fingerprinting.* 2018 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, 2018, pp. 214-218.

[16] Steinberg, D. and Colla, P., 2009. *CART: classification and regression trees.* The top ten algorithms in data mining, 9, p.179.

[17] Z. Li, T. Braun, X. Zhao, Z. Zhao, F. Hu and H. Liang, *A Narrow-Band Indoor Positioning System by Fusing Time and Received Signal Strength via Ensemble Learning.* in IEEE Access, vol. 6, pp. 9936-9950, 2018.

[18] Indoor Localization Dataset, `http://www.cds.unibe.ch/research/software/unibe_ips/`, 2019 March.