2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 18-21 September 2017, Sapporo, Japan

IndoorLoc Platform: A Public Repository for Comparing and Evaluating Indoor Positioning Systems

R. Montoliu, E. Sansano, J. Torres-Sospedra, O. Belmonte
Institute of New Imaging Technologies
Jaume I University
12071, Castellón, Spain
Email: [montoliu,esansano,jtorres,belfern]@uji.es

***Abstract*—This paper presents the *IndoorLoc Platform*, a public repository for comparing and evaluating indoor positioning algorithms and sharing datasets. The proposed web platform can be used to download datasets, learn how some well-known algorithms work, study the implementation of those algorithms, test the methods, and even upload indoor positioning estimations of the user's methods to check the accuracy when comparing against the results provided by other methods already included in a ranking, among other functionalities. This paper also presents a comparative study of the accuracy of two well-known fingerprinting-based indoor localization algorithms using the datasets included in the platform. This comparative study can be performed using the tools included in the platform.**

I. Introduction

Geolocation systems have been present during decades thanks to Global Navigation Satellite System (GNSS), of which the best known is the GPS. The most common are the navigation systems that guide us by car, foot or bicycle, but there are others such as evacuation services and social network services, among others. However, these services are useless in most of the situations, regarding that people spend most of their time in indoor environments such as offices, undergrounds, shopping malls, airports, etc., where these positioning systems do not work. This is one of the reasons why the development of new indoor positioning and navigation systems has attracted the attention of many researchers in the last years.

Many different technologies and approaches have been developed in the last years, being the ones based on RSSI (Received Signal Strength Indicator) fingerprinting [1], [2], [3] among the most popular. They are based on measuring the intensity of the received radio signals of the emitting devices (beacons) that are available at a particular place, and in comparing it with a previously built RSSI data set (also known as radio map). In this scenario, a fingerprint is an RSSI feature vector composed of received signal values from different emitting devices or beacons, associated to a precise position. This technique is becoming increasingly important for indoor localization, since Wi-Fi is generally available in indoor environments where GPS signals cannot penetrate, and the wireless access points (WAPs) can be used as beacons. Other types of indoor localization beacons (Bluetooth, RFID, etc.) can also be used in conjunction with Wi-Fi access points or as a standalone positioning system.

Although there are many papers in the literature trying to solve the indoor localization problem, each approach presents its estimated results using its own experimental setup and measures. Therefore, it is very difficult to compare different methods since the particularities of each experiment are hardly reproducible. In the Pattern Recognition and Machine Learning research fields, the common practice is to test the results of each proposal using several well-known datasets [4]. In this way, researchers are able to fairly compare different methodologies in the literature. For instance, the UCI Machine Learning Repository is a well-known example [5] in this sense. However, in the fingerprint-based indoor localization research field, there is a limited number of such kind of databases [6], [7], [8], [9], [10], [11].

In this paper, the *IndoorLoc Platform*[1], a public repository for comparing and evaluating indoor positioning algorithms, is presented to help solve this lack. The platform is a centralized website where researchers can: 1) access to a public repository of datasets for RSSI fingerprinting, 2) upload indoor positioning estimations on experimental setups included in the platform, 3) analyze positioning methods and 4) interact with the platform in a user friendly environment to tests the algorithms and datasets included.

In order to show a real example of the platform usage, this paper also presents a comparative study of the performance of two fingerprinting-based indoor localization methods included in the platform when using the datasets also included in the platform. In total, two different methods have been used with four different datasets. The methods differ in the methodology used to solve the indoor localization problem. They are a deterministic-based and a probabilistic-based method. The four datasets differ in the type of scenario where data has been captured, as for instance: the number of samples, the size of the scenario, the density of the samples, etc. All the experiments

[1] http://indoorlocplatform.uji.es

978-1-5090-6299-7/17/$31.00 © 2017 IEEE

presented are easily reproducible using the tools included in the platform.

As a summary, the main contributions of this paper are as follows:

1) We present the *Indoorloc platform*, a public repository for comparing, evaluating, testing and downloading indoor positioning algorithms and systems.
2) In order to provide an example of the usage of the platform, We also present a comparative study of the accuracy of two methods that apply different strategies to solve the indoor positioning problem using the datasets included in the platform.

The rest of the paper is organized as follows. Section II comments related work. Section III describes the main sections in which the platform is divided. Sections IV and V explain the datasets and methods, respectively, included in the platform. Section VI presents a set of experiments performed using the algorithms and datasets included in the proposed platform. Section VII describes a real case of use of the usage of the platform during a fingerprinting-based indoor positioning course. Finally, the most important conclusions arisen from this work are presented in Section VIII.

II. Related Work

As it has been commented in Section I, most of the indoor positioning methods found in the literature present the experiments using their own experimental setup. A second related problem is that those datasets are not made available to the research community, making it impossible to reproduce the presented results. Both issues make a fair comparison of localization methods developed by different groups not feasible in a rigorous manner, since scenarios may change in an uncontrolled way.

A better way to compare positioning algorithms is to use the same experimental setup, and for that purpose, the use of a repository of prerecorded data in a large variety of buildings and contexts can be very useful. Some good examples of data repositories in the machine learning community are the UCI Machine Learning Repository [5] and *Kaggle* [12], both created for evaluating machine learning algorithms with common databases.

Another alternative is competitions where several research groups should prepare their methods to obtain the best results using a common experimental setup, or even the same prerecorded data. Some examples of competition are: Microsoft-IPSN [13], [14], [15], [16], EvAAL [17] and EVARILOS [18]. The first off-site indoor location competition was the third track of the EvAAL-ETRI Indoor Location competition [19], called *Wi-Fi fingerprinting in large environments*, which was held during the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN'15). In this event, the competitors had access to the *UJIIndoorLoc* [7] dataset, that has been included in the proposed platform. A similar competition was held in the Seventh International Conference on Indoor Positioning and Indoor Navigation (IPIN'16) [20], where the dataset used was more challenging, since it was included the data provided by all sensors embedded in typical smartphones, acquired by different people moving in different types of buildings.

One of the main problem of such competitions is that when they finish, researchers cannot continue improving their methods. In addition, the different datasets are located in different web pages. The proposed web platform is focused on providing a common place for the researcher to access to fingerprint-related datasets. Another of the main objectives is to provide a continuous competition without deadlines. Therefore, researchers have not time restrictions to test their methods and submit their results to the platform.

The most similar work to the proposed one is [18], where the authors presented a web platform for evaluation of RF-based indoor localization algorithms with two core services: one focused on the storage of raw data and the other focused on automated calculation of metrics for performance assessment. They also include an SDK for convenient access to the platform from MATLAB and python. The two first characteristics are included in the proposed web platform. The SDK is not needed in our case, since users can directly interact with the web platform to upload their results. The main differences of the proposed web platform with respect to with [18] are: 1) it is more focused on fingerprinting methods, 2) it also includes a dashboard section where researchers can make experiments using the methods and datasets included in the platform in a user-friendly environment, and 3) there is a ranking section where the researcher can check the accuracy of their method against the methods of other researchers. In addition, the proposed web platform has been designed in order to easily upload new methods and datasets.

The proposed platform has been designed with a state-of-art visual style and with a user-centered interface making the access to all the elements of the platform very intuitive. For instance, the home web page (see Figure 1) directly presents the main sections of the platform. Another example is that users can download a dataset or upload a result with just a few mouse clicks. The platform is also responsive, and will automatically adapt to the device screen used to access it.

In addition, our platform has a high formative component, because even a user without programming knowledge can interact with the algorithms and datasets included. Although, obviously, it will be the users with a high programming skill who will be able to get a better advantage of the platform because, probably, they will be able to improve the results that can be obtained with the algorithms included in the platform.

III. Overview of the platform

The platform has four main sections: *Datasets*, *Ranking*, *Methods* and *Dashboard*. Figure 1 shows the home page of the platform. The *Datasets* section is a repository of several datasets. Users can download and use them in their own experiments. In the *Ranking* section, users can upload the results of their own algorithms to obtain an estimation of the accuracy of their methods when using the datasets included in the platform. In addition, the results can be included in the ranking, where the best results of each dataset are showed sorted by accuracy. The *Methods* section presents a set of well-known algorithms so users can study their implementation. Finally, in the *Dashboard* section, users can test the algorithms included in the platform using some of the datasets also included in a user-friendly environment. They are briefly described in the next subsections.

A. Datasets

Datasets section presents the basic information about all the datasets included in the platform. In addition, the links to download all the files related to each dataset are also included.

Each dataset has four files at maximum. The first one is a pdf file with information and features about the dataset. This description includes the name of the donors, the contact information, general information about the dataset, a description of the files included, the attribute information, the format of the result file and the citation request. The second one is the *Training* set, that should be used to train the localization models. The third file is the *Validation* set, that can be used to assess the performance of the previously created localization model. The last file is the *Test* set that can be also used to assess the performance of the previously created localization model. The difference between the validation and the test files is that the latter does not include the localization of the samples. To obtain an estimation of the accuracy of the test set, users can run their methods to obtain an estimation of the localization of the samples, and then upload their results to the platform. The training, validation and test files have a comma-separated values (CSV) file format. The three first files are accessible to everyone. Only registered users are allowed to access the test file. Not all the datasets included in the platform have a validation set. In this case, users can use techniques as Cross-validation [21] to assess the performance of the localization model generated.

At the time of writing this article, there are four different databases included in the platform. They are related to the Wi-Fi fingerprinting indoor localization problem, and they are described in Section IV.

Registered users can upload their own datasets following the instructions provided by the platform. Previously to be definitively added to the platform, each new dataset is rigorously examined by the administrators of the platform to ensure that it has the quality level required.

B. Ranking

One of the main objectives of the proposed web platform is to provide a tool to the indoor localization community to compare their methods using well-known datasets. This section has been devoted to this purpose. For each dataset, a list of the best methods, according to a figure of merit, are showed.

Registered users can upload the results of their methods. They must follow the instructions included in the description of the dataset (see Section III-A). Then, the platform calculates the figure of merit for this dataset using the estimated locations provided by the user and the ground truth internally (and privately) stored in the platform. Users have the choice of including or not the result in the ranking.

Each entry in the ranking has some information, provided by the user, about the experiment performed to obtain such us result, e.g. the parameters used.

Figure 2 shows the ranking page for the *IPIN2016 Tutorial* dataset. At the time that this paper was written, the ranking is composed by two experiments performed by the same user. According to the notes written by the user, the result of the leader was obtained using the probabilistic method and the one in the second position using a knn algorithm.

C. Methods

This section shows some basic information about the methods included in the platform. This information consists of the explanation of the method though R[2] source code using comprehensible examples. In addition, links to the Dashboard section, where users can test these methods, are also included.

At the time of writing this article, two methods have been included in the platform: deterministic-based [22] and probabilistic-based [23]. They are described in Section V.

To add new methods to the platform, users must contact with platform administrators. Similarly to the dataset case, each new method is rigorously examined by the administrators of the platform to ensure that it has the quality level required.

D. Dashboard

In the Dashboard section, users can test the methods included in the platform, using the datasets also included in the platform, in a friendly user interface. Figure 3 shows an example of a dashboard for the *UjiIndoorloc* dataset. In particular, the user selected the building 0, the floor 1, the validation set (to estimate the locations), the deterministic method. After clicking on the *Estimate error* button, the platform internally estimates the location of the validation samples and calculates some statistics, as the mean and the median of the estimation error. It also shows one figure with the error histogram and density, and another figure with the empirical cumulative density function (not showed in the Figure 3).

Registered users are allowed to use their own dataset using the methods included in the platform. This dataset must be formatted using a set of rules specified in the web platform.

E. Implementation details

The platform has been implemented using open source tools. In particular, the web service has been set up using the Django[3] framework to build the web application and Shiny[4] to build the interactive dashboards. The platform also makes use of the RMarkdown[5] technology and the **ipft** R package [24]. This

[2]https://www.r-project.org/
[3]https://www.djangoproject.com/
[4]https://shiny.rstudio.com/
[5]https://rmarkdown.rstudio.com/

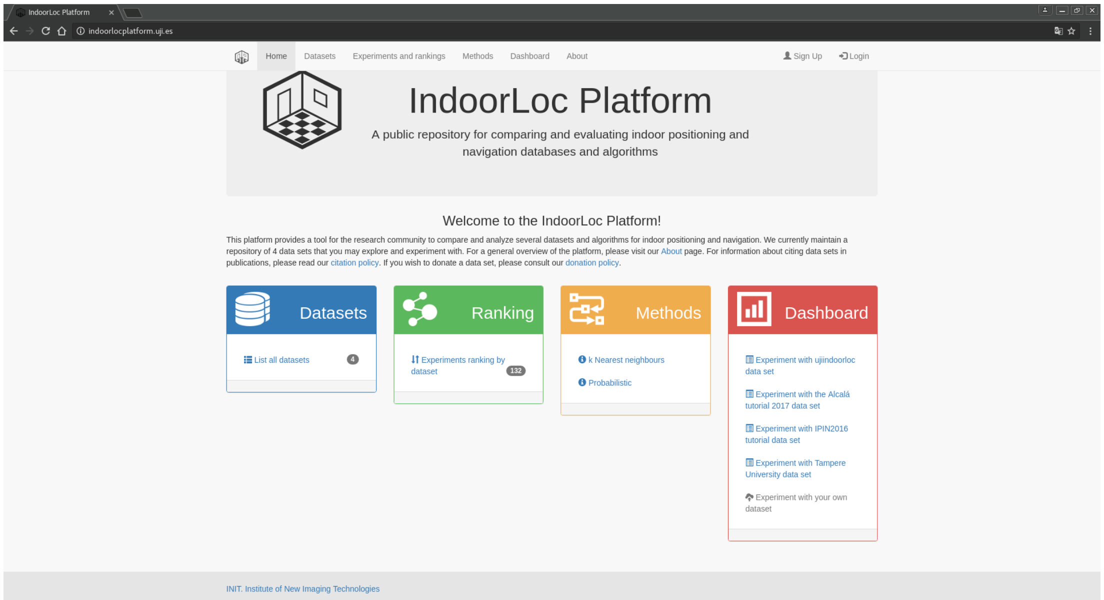

Fig. 1. Homepage of the Indoorloc platform

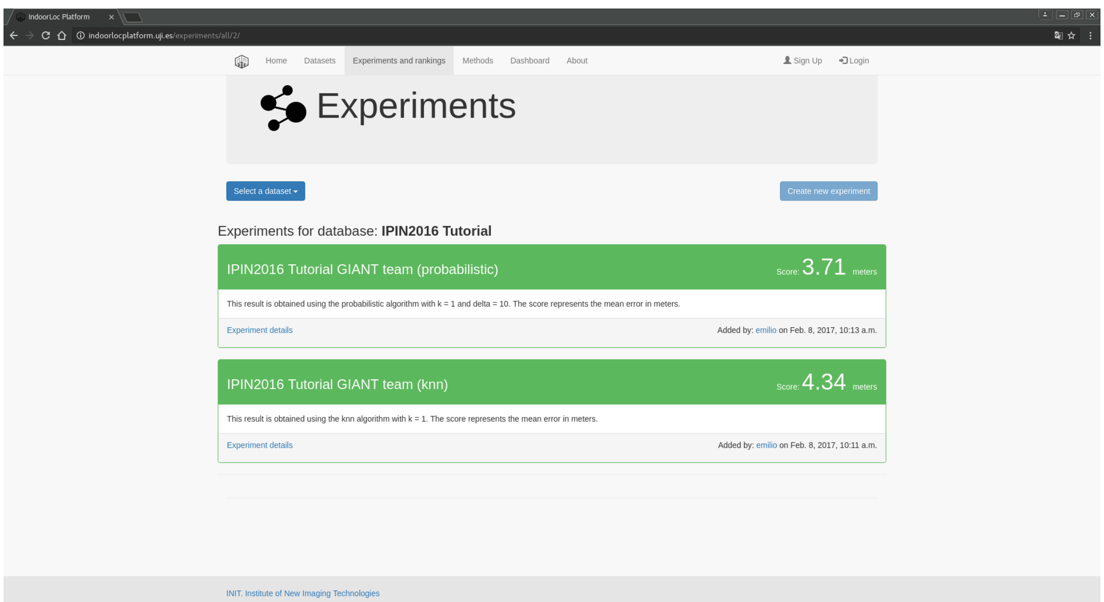

Fig. 2. Ranking webpage of the IPIN2016 dataset. Two experiments have been included in the ranking, the first one (according to the notes written by the contributor) using a probabilistic-based algorithm and the second one using a knn-based method.

system runs on a Linux machine running Apache [6] and Shiny servers.

IV. Datasets Included in the Platform

In the four datasets, Wi-Fi fingerprints are characterized by the detected Wireless Access Points (WAPs) and the corresponding Received Signal Strength Intensity (RSSI). The intensity values are represented as negative integer values near to -100dBm (extremely poor signal) to 0dBm. The positive value 100 is used to denote when a WAP was not detected. Table I shows a summary of the main characteristics of each dataset.

A. *UJIIndoorLoc*

The UJIIndoorLoc [7] database covers three buildings of Universitat Jaume I[7] (Spain), with 4 or more floors and an area of almost $110.000m^2$. It can be used for classification, e.g. actual building and floor identification, or regression, e.g.

[6] https://httpd.apache.org/

[7] http://www.uji.es

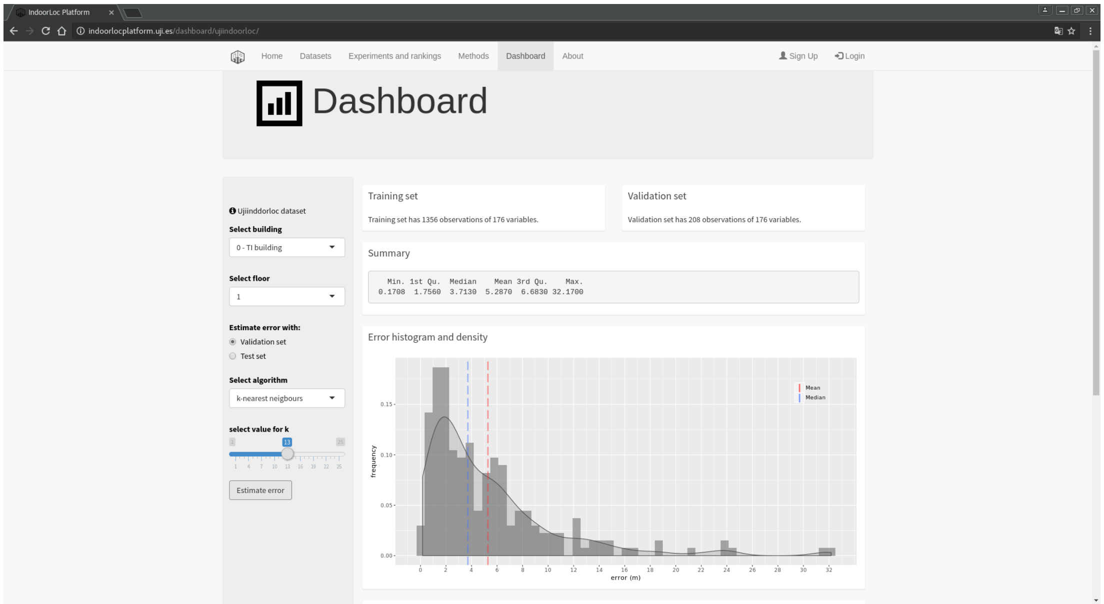

Fig. 3. Example of the *Dashboard* section of the platform

TABLE I
MAIN CHARACTERISTIC OF THE DATASETS INCLUDED IN THE PLATFORM. # STANDS FOR "NUMBER OF"

Dataset	# Buildings	# Floors	# WAPS	# training samples	# validation samples	# test samples
UjiIndoorLoc	3	4-5	520	19937	1111	4900
IPIN2016 Tutorial	1	1	168	927	0	702
Tampere	2	3-4	309-354	1478-583	0	489-175
Alcalá2017 Tutorial	1	1	152	670	0	405

actual longitude and latitude estimation. It was created in 2013 and 2014 by means of more than 20 different users and 25 Android devices. The database consists of 19937 training/reference records and 1111 validation records. There is also a test file where the ground truth is not accessible.

The 529 attributes contain the Wi-Fi fingerprint, the coordinates (latitude, longitude, floor) and Building ID, and other useful information such as the particular space (offices, labs, etc.) and the relative position (inside/outside the space) where the capture was taken, information about who (user), how (android device and version) and when (timestamp) Wi-Fi capture was taken, among other information. During the database creation, 520 different WAPs were detected. Thus, the Wi-Fi fingerprint is composed of 520 intensity values.

This dataset was used in the off-site track of the EvAAL-ETRI Indoor Localization Competition which was part of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN'15) [19]. The best results obtained in the competition have been included in the platform in the corresponding ranking.

Since the particular implementation of the localization methods included in the platform assumes that all the samples are in the same building and floor, the complete dataset has been divided into 11 different datasets.

B. IPIN2016 Tutorial

As an alternative of the *UJIIndoorLoc* dataset, the *IPIN2016 Tutorial* dataset is focused on the study of a small scenario. In particular, it covers a corridor of the School of Engineering of the University of Alcalá[8] (Spain). It is the place where a tutorial on Wi-Fi fingerprinting was held during the IPIN2016 conference. The database consists of 927 training/reference records and 702 test ones. The 177 attributes contain the Wi-Fi fingerprint (168 WAPs), the coordinates where it was taken, and other useful information.

C. Tampere University

This database [25] covers two building of the Tampere University of technology[9] (Finland), with 4 and 3 floors, respectively. In the first building, there are 1478 training/reference records and 489 test ones. The 312 attributes contain the Wi-Fi fingerprint (309 WAPs) and the coordinates (longitude, latitude, and height). In the second building, there are 583 training/reference records and 175 test ones. The 357 attributes contain the Wi-Fi fingerprint (354 WAPs) and the coordinates (longitude, latitude, and height). An important difference of this dataset, with respect the *UjiIndoorLoc*, is that in the

[8] https://www.uah.es/es/
[9] http://www.tut.fi/

former there is just one sample in each training location, while in the latter the number of samples is between 10 and 30.

Data from the two buildings can be considered as two separate datasets, with no implicit relationship between respective WAP labels and real access points MAC addresses, meaning that two columns with the same WAP name in either dataset may be assigned to different access points.

Similarly to the *UJIIndoorLoc* this dataset has been divided into 7 different datasets.

D. *ALCALA2017 Tutorial*

This dataset was created during the 2017 Fingerprinting-based Indoor Positioning tutorial held in the School of Engineering of the University of Alcalá. Data was acquired in the same corridor than the *IPIN2016 Tutorial* dataset. The main differences between both datasets are: 1) a thinner grid was used to capture training data; 2) some users made mistakes labeling the training fingerprints. These errors have not been eliminated since it is a situation that can occur in a real scenario. Users should take into account this situation in their methods.

The database consists of 670 training/reference records and 405 test ones. The 154 attributes contain the Wi-Fi fingerprint (152 WAPs) and the coordinates where it was taken.

V. Methods included in the platform

Two different approaches are considered here for the fingerprinting-based location process: a deterministic, or non-parametric method, and a statistical, or parametric method. In the former, no statistical behavior is assumed, and the location problem is solved according to a set of observations whose positions are known, while the latter makes explicit use of distributions and statistical parameters of the data stored in the radio map to optimize the probabilities in the assignment of the estimated position.

A. *Deterministic-based approach*

The deterministic approach [22], [26], [27] relies on the well known k-Nearest Neighbors algorithm (*knn*) [28] to, given an RSSI vector, select the k more similar training examples from the radio map. The similarity between the RSSI value vectors can be determined, for example, as the $Euclidean$ distance between them, but other distance functions can be used instead [27]. Once the k neighbors are selected, the method estimates the location of the user by calculating the weighted average of the neighbor's positions.

B. *Probabilistic-based approach*

Given the limitations of sensors accuracy and the complex character of signal propagation, the RSSI vector stored for a particular position cannot have completely reliable and accurate information about the emitters signal strength. This uncertainty has been usually modeled by a normal distribution [29], therefore many readings of the signals at the same position are needed to obtain a representative set of statistical parameters to model each RSSI present at that position. The more measurements for a particular location, the more reliable will be their inferred statistical parameters.

In the Probabilistic-based approach [29], [23], [30], the initial collection of RSSI observations associated to a particular point is transformed into a pair of vectors containing the means and the standard deviations of the RSSI for each beacon, and then the complete training data is stored as a set of statistical parameters. Then, given a test fingerprint, for each beacon, it is possible to estimate a probability value that expresses the similarity between the observation measurement at this beacon and the training data for a particular location. An evaluation of the total similarity for every location can be computed as a function of these individual probabilities, like its sum or its product.

The algorithm selects the k training samples with higher probability and, similarly to the deterministic method, it estimates the location of the user by calculating the weighted average of the selected samples positions.

VI. Experiments

The two methods explained in Section V have been tested with the four datasets described in the Section IV, using the tools included in the *Dashboard* section of the platform. Therefore, they are easily reproducible. All possible combination of the parameters have been tested. Only the combination of tuning parameters obtaining the best result is showed. In all cases, the test dataset has been used to assess the performance of the algorithms. The figure of merit used to provide and estimation of the performance of the methods is the mean localization error between the estimated position and the real one (internally known by the platform) of all test samples.

Table II shows the results obtained using the *UjiIndoorLoc* dataset. Note that there is no dataset for the building 0, floor 3 and for the building 2, floor 0, since there are not samples for these floors in the test set.

Table III shows the results on the *Tampere* dataset. In this case, only the results obtained with the deterministic approach are showed, since the probabilistic-based method can only be applied when there are enough samples at each position to calculate the estimation of the statistical parameters needed for the correct operation of this method.

Finally, Tables IV and V show the results on the *IPIN2016 Tutorial* and *ALCALA2017 Tutorial* datasets. In both cases, all the samples are in the same building and floor, therefore it is not necessary to divide the data into subsets.

In the case of the *UJIIndoorloc* dataset, the deterministic method provides better results than the probabilistic one in almost all the cases. The differences in the results obtained across buildings and floors depend on the quality of the radio map capture at each scenario and also on the structural characteristics of each scenario. According to the mean accuracy, the deterministic-based approach is preferable.

There is also a high variability across buildings and floors in the results obtained for the *Tampere* dataset due to the same reasons than in the *UJIIndoorloc* dataset.

TABLE II
MEAN POSITIONING ERROR (IN METERS) OF BOTH METHODS ON THE *UJIIndoorLoc* DATASET. # STANDS FOR NUMBER OF.

Building	Floor	# samples	Deterministic	Probabilistic
0	0	17	4.26	7.83
0	1	17	5.65	6.77
0	2	60	6.06	5.79
1	0	105	9.62	11.26
1	1	147	7.65	20.42
1	2	132	5.40	8.99
1	3	140	8.16	11.0
2	1	20	6.64	10.09
2	2	19	7.96	9.07
2	3	19	3.88	4.57
2	4	22	12.50	21.31
Mean		704	7.18	10.64

TABLE III
MEAN POSITIONING ERROR (IN METERS) OF BOTH METHODS ON THE *Tampere* DATASET. # STANDS FOR NUMBER OF.

Building	Floor	# samples	Deterministic
1	1	156	9.83
1	2	110	14.21
1	3	118	8.01
1	4	105	13.03
2	1	61	15.87
2	2	77	8.38
2	3	37	6.74
Mean		664	10.86

In the case of the *IPIN2016 Tutorial* and *ALCALA2017 Tutorial*, the results are very similar and in both cases the probabilistic approach is preferable. Note that, in the ALCALA2017, the difference is quite significant since the probabilistic-based approach can deal with the unintentional mistakes introduced by some of the dataset creators. Results obtained for these datasets are better than the ones obtained for the *UJIIndoorloc* and the *Tampere* datasets since the scenario of the former ones are more prone to provide lower positions errors, in terms of area covered and distance among samples, than the latter ones.

Taking into account all the datasets, the mean localization error of the deterministic-based method is 6.79 meters without including the results of the *Tampere* dataset, and 8.21 including them. The mean error of the probabilistic-based method is 9.47. Therefore, according to the results, in general, it seems that the deterministic-based method is preferable. However, taking into account the differences of the four scenarios, the deterministic-based approach gets better results in big scenarios with low density of data (*UjiIndoorloc* and *Tampere* datasets), while the probabilistic based one is preferable in small ones with high density of data (*IPIN2016 Tutorial* and *ALCALA2017 Tutorial* datasets).

TABLE IV
MEAN POSITIONING ERROR (IN METERS) OF BOTH METHODS ON THE *IPIN2016 Tutorial* DATASET. # STANDS FOR NUMBER OF.

# samples	Deterministic	Probabilistic
702	4.21	3.55

TABLE V
MEAN POSITIONING ERROR (IN METERS) OF BOTH METHODS ON THE *ALCALA2017 Tutorial* DATASET. # STANDS FOR NUMBER OF.

# samples	Deterministic	Probabilistic
405	5.03	2.53

Note that the results obtained with methods included in the platform can be effectively improved using more sophisticated algorithms, and also using modern machine learning techniques. For instance, the ranking of the *UjiIndoorloc* dataset include some better results than the presented in the Table II, since they are the best results obtained in the *Wi-Fi fingerprinting in large environments* (IPIN'15) competition.

VII. THE PLATFORM IN USE

The performance of the platform was tested during the 2017 Fingerprinting-based Indoor Positioning tutorial held in the School of Engineering of the University of Alcalá. As an activity of the course, an indoor localization competition took place using the *ALCALA2017 Tutorial* dataset.

The 15 attendees were invited to use the platform to upload the results of their proposals. Some of them used the *Dashboard* section of the platform to test different parameter configurations of the localization methods included in the platform, and others manually programmed their own method from the source code provided by the course instructors. After a very competitive and exciting competition, the winner team got an error of only 2.14 meters using a probabilistic method. This result is even better than the best one that can be directly obtained using the Dashboard included in the platform.

In general, tutorial attendees were able to easily use the platform, mainly the *Dataset* download section, the *Dashboard* section and, obviously, the *Ranking* sections. Almost no queries to the course introduction were produced, showing the effective user-centered design applied to the platform.

VIII. CONCLUSIONS

In this paper, the *IndoorLoc Platform* has been presented. It is a public repository for comparing and evaluating indoor positioning algorithms. The proposed web platform can be used to download datasets, learn how some well-known algorithms work, study the source code of those algorithms, test the methods, and even upload results of the user's methods to check the accuracy when comparing against the result provided by other methods already included in a ranking, among other functionalities.

To present a real example of the usage of the platform, a comparative study of the accuracy of two well-known fingerprinting-based indoor localization algorithms, using the datasets included in the platform, has been also presented. According to the results obtained, the deterministic-based approach gets better results in big scenarios while the probabilistic based one is preferable in small scenarios. This experiments are easily reproducible using the tools included in the platform.

This web platform is an ongoing project, and future versions will implement new algorithms and include more datasets, with the aim to provide an interesting tool for researchers and become a reference web platform for indoor positioning research. For this purpose, researchers are invited to include more methods and datasets in the platform.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the Spanish Ministry of Economy and Competitiveness in the projects: 'Proyectos I+D Excelencia' TIN2015-70202-P and 'Redes de Excelencia' TEC2015-71426-REDT.

REFERENCES

[1] S. He and S. G. Chan, "Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 466 – 490, 2016.

[2] C. Wu, Z. Yang, Y. Liu, and W. Xi, "Will: Wireless indoor localization without site survey," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 839–848, 2013.

[3] D. Han, S. H. Jung, M. Lee, and G.-W. Yoon, "Building a practical wi-fi-based indoor navigation system," *IEEE Pervasive Computing*, vol. 13, pp. 72–79, 2014.

[4] S. García, D. Molina, M. Lozano, and M. F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: A case study on the cec'2005 special session on real parameter optimization," *Journal of Heuristics*, vol. 15, no. 6, pp. 617–644, 2009.

[5] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[6] K. Nahrstedt and L. Vu, "Crawdad dataset uiuc/uim (v. 2012-01-24)," 2012. [Online]. Available: http://crawdad.org/uiuc/uim/20120124

[7] J. Torres-Sospedra, R. Montoliu, A. M. Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems," in *Proceedings of the 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN'14)*, 2014, pp. 261–270.

[8] J. Talvitie, E. Lohan, and M. Renfors, "The effect of coverage gaps and measurement inaccuracies in fingerprinting based indoor localization." in *Proceedings of International Conference on Localization and GNSS 2014 (ICL-GNSS'14)*, 2014.

[9] J. Torres-Sospedra, D. Rambla, R. Montoliu, O. Belmonte, and J. Huerta, "Ujiindoorloc-mag: A new database for magnetic field-based localization problems," in *Proceedings of the Sixth Conference on Indoor Positioning and Indoor Navigation (IPIN'15)*, 2015.

[10] P. Barsocchi, A. Crivello, D. Rosa, and F. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting," in *Proceedings of the seventh Conference on Indoor Positioning and Indoor Navigation (IPIN'16)*, 2016.

[11] N. Moayeri, O. Ergin, F. Lemic, V. Handziski, and A. Wolisz, "Perfloc: An extensive data repository for development and a web-based capability for performance evaluation of smartphone indoor localization apps," in *Proceedings of the 27th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'16)*, 2016.

[12] A. Goldbloom, B. Hamner, J. Moser, and M. Cukierski, "kaggle: Your home for data science," 2017. [Online]. Available: https://www.kaggle.com/

[13] D. Lymberopoulos, R. Choudhury, X. Yang, and S. Sen, "Microsoft indoor localization competition (ipsn'14)," 2014. [Online]. Available: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2014

[14] D. Lymberopoulos, J. Liu, X. Yang, A. Naguib, A. Rowe, N. Trigoni, and N. Moayeri, "Microsoft indoor localization competition (ipsn'15)," 2015. [Online]. Available: ttps://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2015

[15] D. Lymberopoulos, J. Liu, Y. Zhang, P. Dutta, X. Yang, and A. Rowe, "Microsoft indoor localization competition—ipsn 2016," 2016. [Online]. Available: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2016

[16] D. Lymberopoulos, J. Liu, M. Bocca, V. Sequeira, N. Trigoni, and X. Yang, "Microsoft indoor localization competition - ipsn 2017," 2017. [Online]. Available: https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2017

[17] F. Potortì, P. Barsocchi, M. Girolami, J. Torres-Sospedra, and R. Montoliu, "Evaluating indoor localization solutions in large environments through competitive benchmarking: The evaal-etri competition," in *Proceedings of the Sixth Conference on Indoor Positioning and Indoor Navigation (IPIN'15)*, 2015.

[18] F. Lemic, V. Handziski, N. Wirstrom, T. Van Haute, E. De Poorter, T. Voigt, and A. Wolisz, "Web-based platform for evaluation of rf-based indoor localization algorithms," in *In Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW'15)*, 2015.

[19] J. Torres-Sospedra, A. Moreira, S. Knauth, R. Berkvens, R. Montoliu-Cols, scar Belmonte-Fernndez, S. Trilles, M. J. Nicolau, F. Meneses, A. Costa, A. Koukofikis, M. Weyn, and H. Peremans, "Realistic evaluation of indoor positioning systems based on wi-fi fingerprinting: The 2015 evaal-etri competition," *Journal of ambient intelligence and smart environments*, vol. 9, p. 263279, 2017.

[20] J. Torres-Sospedra, A. R. Jimnez, S. Knauth, A. Moreira, Y. Beer, T. Fetzer, V.-C. Ta, R. Montoliu, F. Seco, G. M. Mendoza-Silva, O. Belmonte, A. Koukofikis, M. J. Nicolau, A. Costa, F. Meneses, F. Ebner, F. Deinzer, D. Vaufreydaz, T.-K. Dao, and E. Castelli, "The smartphone-based offline indoor location competition at ipin 2016: Analysis and future work," *Sensors*, vol. 17, no. 3, 2017.

[21] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.

[22] P. Bahl and V. N. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'00)*, 2000, pp. 775–784.

[23] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys'05)*, 2005, pp. 205–218.

[24] E. Sansano and R. Montoliu, "Indoor positioning and fingerprinting. the r package ipft," 2017. [Online]. Available: https://cran.r-project.org/web/packages/ipft/index.html

[25] A. Cramariuc and E. Lohan, "Open-access wifi measurement data and python-based data analysis," 2016. [Online]. Available: http://www.cs.tut.fi/tlt/pos/meas.htm

[26] N. Marques, F. Meneses, and A. Moreira, "Combining similarity functions and majority rules for multi-building, multi-floor, wifi positioning," in *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN'12)*, 2012, pp. 1–9.

[27] J. Torres-Sospedra, R. Montoliu, S. Trilles, O. Belmonte, and J. Huerta, "Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems," *Expert Systems with Applications*, vol. 42, no. 23, pp. 9263–9278, 2015.

[28] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Sep. 1967.

[29] A. Haeberlen, E. Flannery, A. Ladd, A. Rudys, D. Wallach, and L. Kavraki, "Practical robust localization over large-scale 802.11 wireless networks," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking (ModiCom'04)*, 2004, pp. 70–84.

[30] D. Madigan, E. Einahrawy, R. P. Martin, W.-H. Ju, P. Krishnan, and A. S. Krishnakumar, "Bayesian indoor positioning systems," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05)*, 2005, pp. 1217–1227.