1. Damian TERLECKI[1], 2. Teodora Dimitrova-GREKOW[2], 3. Jacek GREKOW [3]

Bialystok University of Technology (1,2,3)
ORCID: 2. 0000-0003-3138-4660; 3. 0000-0003-2094-0107

doi:10.15199/48.2023.08.07

Indoor Localisation Based on Wi-Fi Infrastructure

***Abstract**. Over the last decade, indoor positioning has played an increasing role in the navigation market. In this work we present an indoor attempt for localization based on radio-environment properties. Our research concerns implementation of machine learning algorithms for a Wi-Fi fingerprints-based positioning system. The algorithms we chose are kNN, NB and RF. The method was evaluated using several different mobile devices, with samples collected in different locations of a school building. The results we achieved are very promising.*

Streszczenie.** W ciągu ostatniej dekady pozycjonowanie w pomieszczeniach odgrywa coraz większą rolę na rynku nawigacji. W niniejszej pracy przedstawiamy podejście lokalizacji wewnętrznej na podstawie właściwości środowiska propagacyjnego. Przedstawiamy zastosowanie algorytmów uczenia maszynowego do systemu pozycjonowania opartego na odciskach palców Wi-Fi. Wybrane zostały algorytmy kNN, NB oraz RF. Skuteczność/Dokładność metody została oceniona przy użyciu kilku różnych urządzeń mobilnych, z próbkami pobranymi w różnych miejscach budynku szkolnego. Otrzymane wyniki są bardzo obiecujące. **(Lokalizacja wewnątrzbudynkowa w oparciu o strukturę Wi-Fi).

Keywords: dynamic in-door localization, Wi-Fi infrastructure, mobile application.
Słowa kluczowe: lokalizacja wewnątrzbudynkowa, Wi-Fi infrastruktura, aplikacja mobilna.

Introduction

Location, logistics, service and interaction with the client, virtual reality - these are some of the many areas in which various types of positioning systems are created. It turns out that people spend much more time indoors, where the use of GPS is not practical. The disadvantages of this system, when used inside buildings, include: a frequent lack of coverage, signal reflection, and above all, very poor accuracy of several dozen meters. Use of other location methods that would meet the expectations of customers and users becomes necessary.

In this work, a positioning system was tested that could be used to carry out navigation at a very satisfactory level. This system is based on a Wi-Fi infrastructure. We analyze selected methods of position detection inside buildings on the example of a university (Białystok University of Technology).

Positioning is determining the location of an object. The algorithm we introduce in this work is based on Wi-Fi fingerprint collection methods combined with machine learning. The positioning accuracy evaluation was done on three different mobile devices, with samples collected at 32 different locations on the last two floors of a school building, consisting of numerous classrooms and offices next to each other.

Methodology

Analysis of location techniques inside buildings [1-3] allows to distinguish three main types of systems: inertial [4], infrastructural [5, 6], hybrid [7, 8].

The most popular infrastructures used to implement positioning systems inside buildings are radio infrastructures - mainly Wi-Fi and Bluetooth [9, 10]. Active landmarks like RFID (Radio-Frequency IDentification) are also very common. Many methods used in radio location are based on various models of signal propagation: TDoA (Time Difference of Arrival), DoA (Direction of Arrival), etc. All of these methods assume that the position of the transmitters are well known and on this basis the location of the receiver is determined. For this purpose, triangulation and multilateration are used. The first technique is based on angular values and requires two transmitters. The second one needs at least three transmitters, but allows you to use time or distance values.

The fingerprinting method assumes that a place can be registered with a unique signature of that place. In the standard Wi-Fi fingerprinting method [11, 12], the location area is divided into a set P of reference points.

In practice, the creation of a map takes some time, therefore averaging over time and number of samples is used. The creation of a radio map defined in this way completes the phase of the Wi-Fi fingerprint collection process known as the ”offline” stage. In the online (live) phase, the user of the mobile device measures signal strength values without location information. On the basis of these data, the position is determined by classification (digitized location) or regression (coordinates: x, y, z). For this purpose, three different approaches are used: deterministic, probabilistic, and based on pattern searching (e.g. using neural networks) [13].

Implementation

We developed a fingerprint attempt to detect the user's position. The basis is the existing Wi-Fi infrastructure in the school building. For this reason, we collected fingerprint data. Then an appropriate set of attributes was selected, which allows for maximum accuracy of machine learning algorithms. The attributes of the objects were defined supposing possible transformations of both the test and the training sets. Finally, an available filtration point was considered.

The implemented system is intended for mobile devices. Using Android as an example, a developer can scan for access points that returns the following information:

- BSSID (Basic Service Set Identifier) - access point identifier,
- SSID (Service Set Identifier) - network identifier (name),
- RSSI (Received Signal Strength Indicator) - indicator of received signal strength [dBm],
- timestamp counted from the system startup to the last access point visibility [µs].

This information is grouped into a single result, and for a single scan, a result list with a length equal to the number of access points detected is returned.

From the user's point of view, a sufficient form of location is to determine in which part of the building it is located. The accuracy required in this approach is not expressed in meters, but is defined at the level of places, classrooms or offices. This significantly simplifies the designed positioning system. On the one hand, discretization is associated with loss of certain information, which can also reduce accuracy. On the other hand, it allows us to simplify the data collection stage a bit. This is an undoubted advantage in the case of involving the user, without special measuring equipment, in the stage of collecting data in a new location [14, 15]. Considering all the advantages, disadvantages and the practical purpose of the system, the choice is to take the advantage of including the users in the data collection stage, which allows for trouble-free updating of training data in the event of subsequent changes in the infrastructure. Therefore, the chosen method was classification of places based on the signal and not the estimation of x, y, z coordinates using regression.

In summary, the decision class is a defined location, e.g. "position opposite room 123." The attributes, on the other hand, are associated with access points BSSID. They are assigned RSSI values at a defined location. The radio map is built from these data.

There are several different types of attributes for a Wi-Fi fingerprint that you can choose from. The most popular, however, is the strength of the received signal, which value we consider on **linear** and **logarithmic** scales.

In the literature, the terms RSSI and RSS are used interchangeably. They usually refer to a logarithmic scale.

In a logarithmic scale, the received Wi-Fi signal strength is presented in terms of one milliwatt in the form of decibels [dBm]. Values on this scale are obtained by scanning for Wi-Fi access points using the Android API.

The standard range of RSS values received on a mobile device is from about -30 dBm to -95 dBm. If the signal is not detected, the value is -95 dBm or -100 dBm. Experiments [16] show that the signal value from the same access point may fluctuate by +/- 5 dBm. This can be affected by the direction of the device, the presence of other interfering signals, and multipath propagation.

The transition to the linear scale allows for a slightly changed interpretation of the distance between Wi-Fi fingerprints, as in the kNN (k-Nearest Neighbours) [17] method. The more two objects are dissimilar, the greater the difference between the signals. The difference also increases with power. For example, the distance when comparing -30 dBm and -40 dBm signals is much longer than -80 dBm and -90 dBm. In this way, dependencies towards access points with the highest signal values (points closest to the user) are prioritized.

Additional Transformations and Access Point Filtration

Signal fluctuations can affect the classification and the result in the received signals being classified elsewhere. The reason for this is the insufficient uniqueness of Wi-Fi fingerprints. To improve the classification, additional data transformations can be proposed:

1) mean - easy to analyze (combined with variance);

2) median - as a measure independent of extreme values, it may turn out to be effective in the case of momentary signal fading (represented as -100 dBm).

These functions are very useful by creating a Wi-Fi fingerprint corresponding to a given reference point and access point.

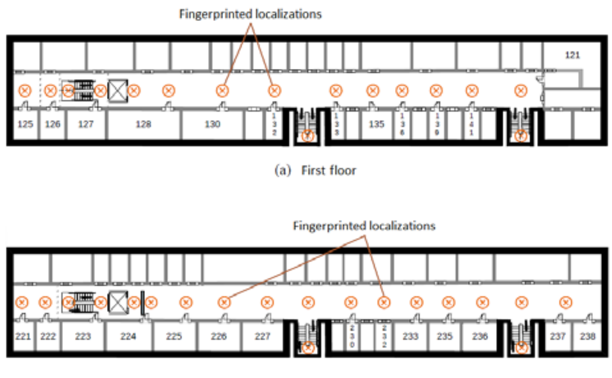
Use of online transformations implies that the user is in the same place for the last n scans. A change of position can, however, be associated with an independently working pedometer algorithm that could mark the start of scanning for a new location. The variability of the Wi-Fi infrastructure manifests itself in the disappearance and appearance of certain access points. However, this is not a problem, as the fingerprint approach takes such conditions into account. The problem arises when the access points move, which will affect the signal. The classification may then turn out to be wrong due to an outdated model, so it is necessary to select elements with a stable position.

In the case of the Białystok University of Technology, the eduroam network is basically a permanent element of the Wi-Fi infrastructure. The auxiliary elements include various types of open access points (hotspots). Ultimately, it is possible to add other devices such as network printers to temporarily improve the positioning (they don't always have to work). Other infrastructure elements, e.g. mobile hotspots, should be ignored in order not to compromise accuracy.

Filtering is done based on the SSID that identifies the network. Depending on the infrastructure, one network may consist of multiple access points. Often, a dual-band infrastructure is also built. This configuration entails that a single access point is assigned two identifiers (BSSIDs), one each for the 2.4 GHz and 5 GHz bands.

Collecting data to build Wi-Fi radio maps

To complete the Wi-Fi fingerprint classification, an appropriate data amount was needed. We gathered such data at the Faculty of Computer Science of the Białystok University of Technology, Poland. For the experiments, the last two floors of the building were selected. We chose 32 reference points, evenly spaced apart. The locations of these points are presented on the illustrative maps of the floors of the building (Figs. 1a and 1b), with the points on the mezzanines being repeated. The dimensions of the floors are approximately 100 m by 18 m. The built Wi-Fi radio maps consist of 13 reference points on the 1st floor, 16 reference points on the 2nd floor (slightly larger corridor space and an additional wall) and 3 reference points on the landing/entresol.



(a) First floor

(b) Second floor

Fig.1. Location of the used reference points

Due to the division of the data into the test and the training sets, the number of K samples for each reference point was chosen as 30 and 10, respectively. This is partly due to the time of a single scan. Depending on the device, it can usually last from 1 to 5 seconds. In addition, an extra pause of about 1 second was introduced between consecutive scans to maintain the signal diversity in case of faster results. In this way, creating a Wi-Fi fingerprint for a given reference point takes from 20 to 60 seconds for the test set or from 1 to 3 minutes for the training set.

The training set was created using a Pentagram Combo 4-Core device, while the test sets were collected using three different devices: 1) Pentagram Combo 4-Core, 2) Goclever Tab A103, 3) Samsung Galaxy Note 10. Data analysis shows that the signal received by different devices (and Wi-Fi modules) may differ despite the same conditions, which required separate models for different devices or the use of data set transformations. To examine the reliability of the study, the data were collected on different days during standard working conditions of the University.

Experimental results

The data collected in the form of training set and test set were used for the positioning accuracy tests. Both sets contained 32 decision classes (rooms). The training set had 960 records (30 per decision class) and the test set had 320 records (10 per decision class). The research was conducted with the machine learning software Weka [18]. The classification methods we discuss are: kNN, NB (Naive Bayes) [19], and RF (Random Forests) [20]. Location accuracy is the ratio of the number of correctly classified features to the total number of features, expressed as a percentage. The experiments were conducted in three parts:

Experiment 1: building models on training set and evaluating using 10-fold cross validation CV-10 (one device, same environment by training and testing).

Experiment 2: building models on training set and evaluating using test sets. The gathered test sets on days different than the training set allowed to estimate the influence of the environment (different devices, different environment by testing).

Experiment 3: building models on merged test sets and evaluating using training set (different devices, different environment by training).

Experiment 1

To check if the number of access points was sufficient, we conducted initial experiments on one device. For each attempt, two types of filtering against network identifiers were taken into account. This resulted in a different number of attributes (access points): The following SSID regular expressions were used "(eduroam)": L = 9 and "(eduroam | pb-guest |. * Hotspot. *)": L = 23.

The obtained accuracies of classification proved that the created radio map contains sufficiently unique fingerprints. The results are presented in Table 1. Increasing the number of access points L from 9 to 23 results in an increase in accuracy to within one percentage point. The most correctly classified objects can be recorded for RF: 95% and 96%, respectively. The kNN method for Euclidean and city distance also achieved good results 94% and 96%.

Table 1. Accuracy of classifiers evaluated using CV-10

Classifier	Parameters	Accuracy	
		L = 9	L = 23
kNN_e	K=1, Euclidean distance	94.17	94.90
kNN_c	K = 1, city distance	94.48	95.94
kNN_e’	K=1, Euclidean distance considering not normalized noise	90.52	94.06
NB	Gaussian distribution of class attributes	89.48	92.19
RF	I = 100 (trees);	95.42	96.46

Experiment 2

To stabilize the precision of our location, we focused on the neighbouring access points. Considering the distance to the nearest neighbour reference point and the distances to the two next ones, we defined **Accuracies Grade** or just **Grade**. Hence **Grade 2** considered the nearest neighbour point and **Grade 3** considered the next two points, respectively. **Grade 1** stands for the accuracy used previously. For example, if you are at location 125, Grade 2 would also include location 126, and Grade 3 would additionally include location 127 (Fig. 1a). Furthermore, the results of two transforming the data from linear to logarithmic scale and vice versa are shown, additionally using the median and averaging.

The logarithmic indicator of the received signal strength is the standard way by which radio maps are constructed. However, this method may not be sufficient for tests between different devices. The best results in the distribution graph are for the RF method 78% (Grade 3). Using the distance to the nearest neighbour reference points improve significantly location accuracy. Detailed results for this option are presented in Table 2.

Table 2. Location accuracy using logarithmic scale and RF

Device	Grade 1	Grade 2	Grade 3
Goclever Tab A103	30.00	56.56	72.19
Pentagram Combo 4-Core	41.56	73.44	92.19
Samsung Galaxy Note 10	24.30	48.91	70.09
Average	31.95	59.64	78.16

RF shows high results again (Grade 3). However, the best one is achieved by kNN (kNN_e’ - Euclidean distance combined with noise elimination +/- 5 dBm). The results for kNN_e’ (Table 3) are very stable across devices. In comparison with logarithmic scale using RF (Table 2) the obtained mean accuracy of Grade 1 and Grade 2 (36%, 64%) are higher, and the mean accuracy of Grade 3 differs only by two percentage points.

Table 3. Location accuracy using linear scale and kNN_e'

Device	Grade 1	Grade 2	Grade 3
Goclever Tab A103	36.56	60.31	73.44
Pentagram Combo 4-Core	35.63	68.75	79.06
Samsung Galaxy Note 10	34.89	64.17	75.39
Average	35.69	64.41	75.96

Experiment 3

One of the advantages of such an indoor navigation system is the possibility of adding the data collected by users to the system database. Our last test examined the impact of the data collected by different users with different devices on the accuracy of the localization. In order to check this situation, individual test sets obtained from three different devices were merged. Thanks to this, the number of registered access points for the new set was increased from 21 to 52.

The results of these experiments were very impressive and worth presenting. Thus, we gathered all of them into Table 4, which requires a short explanation for proper reading of the cumulated information. For each of the classifiers (and classifiers’ variations: ***kNN_e, kNN_c***, ***kNN_e’***, ***NB*** and ***RF***) we discussed previously, the results for both logarithmic ***Log*** and linear ***Lin Scale*** are shown. The number of registered access points is ***L***. Data for all three Accuracy Grades ***Gr 1***, ***Gr 2*** and ***Gr 3*** are exposed. Further there are two experimental results in one table. Left from the classifier column, which is common for both, are presented results from the access point verification. Right from the classifier column are shown results from the evaluation by an external set. The last column presents the normal N data and the averaged A ones using the previous training set as a test set. The first question was whether the aggregated fingerprints were still sufficiently unique. Algorithms were evaluated using CV-10 on both sets: L = 21 and L = 52. This allowed us to claim that the correct localization is still possible. More access points did not negatively affect accuracy.

The accuracy achieved with the use of RF for Grade 2 and Grade 3 was 84% and 97%, respectively, for the logarithmic, and 82% and 96%, respectively for the linear scale. The kNN_e' method ranked second.

The results presented in Table 4 show that averaging the data increased the accuracy of the location using Wi-Fi fingerprint classification. Once again, the RF method emerged as the best solution, resulting in an accuracy for Grade 3 of 98% for the logarithmic and 100% for the linear scale.

Table 4. Accuracy model trained on data from different devices

L	Access points verification			Classifier/ Scale	Final model			
	Gr 1	Gr 2	Gr 3		Gr 1	Gr 2	Gr 3	
21	63	88	96	kNN_e/Log	29	62	86	N
52	63	86	91	kNN_e/Lin	29	56	69	
21	63	86	96	kNN_e/Log	30	65	87	A
52	64	87	92	kNN_e/Lin	32	65	72	
21	67	89	96	kNN_c/Log	38	68	87	N
52	64	86	92	kNN_cLin	32	60	73	
21	72	92	98	kNN_c/Log	35	73	92	A
52	68	88	93	kNN_c/Lin	36	64	72	
21	56	85	94	kNN_e'/Log	36	72	90	N
52	79	89	94	kNN_e'/Lin	38	70	83	
21	62	83	94	kNN_e'/Log	41	76	93	A
52	67	89	94	kNN_e'/Lin	43	72	84	
21	43	71	84	NB/Log	30	64	84	N
52	45	71	77	NB/Lin	38	67	82	
21	53	81	93	NB/Log	32	65	86	A
52	60	81	84	NB/Lin	44	75	79	
21	83	95	97	RF/Log	55	84	97	N
52	82	95	98	RF/Lin	53	82	96	
21	87	98	99	RF/Log	59	97	98	A
52	88	98	100	RF/Lin	57	85	100	

Conclusion

For location detection based on Wi-Fi fingerprint classification, a fairly good accuracy of 96.46% of correctly classified objects was obtained. The use of various devices in different environmental conditions causes a drastic decrease in the location accuracy to 10-20%. It is then necessary to introduce some transformations of the values of attributes and sets, e.g. change of scale, averaging, or the median. It turns out that good results can be obtained after switching from the logarithmic to the linear scale and averaging the signals at reference points on the test sets.

Wi-Fi location from the user's point of view is less demanding. It only requires Wi-Fi to be turned on. The necessary condition for the proper operation of this system, however, is the prior creation of a radio map of all the navigated buildings. This research shows that data from different devices can be mixed together without any loss of accuracy.

Furthermore, the users themselves can be involved in the creation of the databases of benchmarks. For a Wi-Fi-based system, it is worth testing other methods that allow for better classification of places between different devices. An interesting aspect may be the combination of this solution with an inertial algorithm based on an accelerometer into one integrated system.

Acknowledgments This research was realized as part of study in Department of Information Systems and Computer Networks, Faculty of Computer Science of Bialystok University of Technology, and financed with funds from the Ministry of Science and Higher Education, funds for 2023.

Authors: *mgr inż. Damian Terlecki, E-mail: damian.terlecki@gmail.com; dr inż. Teodora Dimitrova-Grekow, E-mail: t.grekow@pb.edu.pl; dr hab. inż. Jacek Grekow, E-mail: j.grekow@pb.edu.pl, Politechnika Białostocka, Wydział Informatyki, Wiejska 45A, Białystok 15-351.*

REFERENCES

[1] Correa A., Barcelo M., Morell A., Vicario J. L., A Review of Pedestrian Indoor Positioning Systems for Mass Market Applications, *Sensors*, 17 (2017), No. 8

[2] Bernard R., Indoor Positioning Systems. Securities Industry A ssociation, https://www.securityindustry.org/2017/11/21/indoor-positioning-systems/ (accessed on 03.01.2023)

[3] He S., Chan S. H. G., Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons, *IEEE Communications Surveys & Tutorials*, 18 (2016), No. 1, 466–490

[4] Beauregard S., Haas H., Pedestrian Dead Reckoning: A Basis for Personal Positioning, *Proceedings of the 3rd Workshop on Positioning, Navigation and Communication*, (2006), 27–36

[5] Kunhoth J., Karkar A., Al-Maadeed, S. *et al.*, Indoor positioning and wayfinding systems: a surve, *Hum. Cent. Comput. Inf. Sci.*, 10 (2020), No. 18

[6] Popoola Or., Sinanović S., Ramirez-Iniguez R.et al., LED-based indoor positioning system using novel optical pixelation technique, *Healthcare technology letters*, 6 (2019), No. 3, 76-81

[7] Kanrar S., Dawar K., Pundir A., Pedestrian localisation in the typical indoor environments, *Multimed Tools Appl*, 79 (2020), 27833–27866

[8] N Styła M, Adamkiewicz P., Optimization of building control systems with use of alternative tracking methods using the RSSI index and artificial intelligence mechanisms, *Przegląd Elektrotechniczny*, 98 (2022), nr 2, 199-202

[9] Baichuan H., Jingbin L., WeiS., Fan Y., A Robust Indoor Positioning Method based on Bluetooth Low Energy with Separate Channel Information, *Sensors*, 19 (2019), 16

[10] Leplawy M., Indoors localization system with the use of WiFi and other network standards, *Przegląd Elektrotechniczny*, 91 (2015), No. 1, 173-174

[11] Pérez-Navarro A., Torres-Sospedra J., Montoliu R., et al., Challenges of Fingerprinting in Indoor Positioning and Navigation, In Intelligent Data-Centric Systems, *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*, Academic Press, (2019), 1-20

[12] Wysocki M., Nicpoń R., Trzaska M., Czapiewska A, Research of Accuracy of RSSI Fingerprint-Based Indoor Positioning BLE System, *Przegląd Elektrotechniczny*, 98 (2022), No. 9, 86-91

[13] Wang W., Marelli D., Fu M., Fingerprinting-Based Indoor Localization Using Interpolated Preprocessed CSI Phases and Bayesian Tracking, *Sensors*, 20 (2020)

[14] Xia Sh., Liu Y., Yuan G., et.al, Indoor Fingerprint Positioning Based on Wi-Fi: An Overview, *ISPRS Int. J. Geo-Inf.*, 6 (2017)

[15] Yoo J., Park J., Indoor Localization Based on Wi-Fi Received Signal, Strength Indicators: Feature Extraction, Mobile Fingerprinting, and Trajectory Learning, *Applied Sciences*, 9 (2019), No.18

[16] Firdaus F., Ahma. N., Sahibuddin S., Indoor positioning system based Wi-Fi fingerprinting for dynamic environment: Experimental preliminary result, *J. of Engineering and Applied Sciences*, 12 (2017), No.17, 4442–4447

[17] Aha D., Kibler D., Instance-based learning algorithms, *Machine Learning*, 6 (1991), 37-66

[18] Frank E., Hall M. A., Witten I. H., The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, (2016)

[19] John G. H., Langley P., Estimating Continuous Distributions in Bayesian Classifiers, *In: Eleventh Conference on Uncertainty in Artificial Intelligence*, (1995), 338-345

[20] Breiman L., Random Forests, *Machine Learning*, 45 (2001), No. 1, 5-32