FUZZY APPROACH TO DEVICE LOCALIZATION BASED ON WIRELESS NETWORK SIGNAL STRENGTH

Michał Socha, Wojciech Górka, Marcin Michalak
Research Network Łukasiewicz, Institute of Innovative Technologies EMAG, Katowice, Poland

Abstract. The paper presents an original approach to device location detection in a building. The new method is based on a map of individual interiors, drawn up based on the measurements of the strength of wireless network signals for each building venue. The device is initially assigned to all venues whose descriptions sufficiently correspond with the current measurements taken by the device. A fuzzy assignment level for each of the potentially considered venues depends on the difference between the averaged network strengths for the venue and the signal strengths currently measured with the device for localization purposes. Ultimately, the device is assigned to the venue with the highest level of assignment.

Keywords: wireless networks, fuzzy sets, device location

ROZMYTE PODEJŚCIE DO LOKALIZACJI URZĄDZEŃ NA PODSTAWIE SIŁY SYGNAŁU SIECI BEZPRZEWODOWYCH

Streszczenie. W pracy przedstawiono oryginalną metodę lokalizowania urządzeń w budynku. Nowa metoda bazuje na mapie poszczególnych pomieszczeń, stworzonej w oparciu o pomiary sygnałów sieci bezprzewodowych zmierzonych w tych pomieszczeniach. Wstępnie urządzenie przypisywane jest do tych pomieszczeń, których opis w odpowiednim stopniu pokrywa się z pomiarami dokonanymi przez urządzenie. Stopień rozmytej przynależności do każdego z wstępnie wytypowanych pomieszczeń zależy z kolei od różnicy pomiędzy uśrednionymi wartościami sygnałów sieci bezprzewodowych i aktualnie zmierzonymi do celów lokalizacji. Ostatecznie urządzeniu przypisywane jest to pomieszczenie, dla którego stopień przynależności jest największy.

Słowa kluczowe: sieci bezprzewodowe, zbiory rozmyte, lokalizowanie urządzeń

1. Introduction

The use of location tracking is becoming increasingly popular in many services used on a daily basis. Traffic navigation or location context services that can be used outdoors are mainly based on GPS [5]. There are many possible uses of indoor location tracking. A good example can be guidance apps at museums [8, 9]. A museum app can contextually present descriptions of nearby exhibits. It can also be an indoor map supported with information about the user's current position. Location tracking can be also useful for disabled people and health care buildings [2, 14]: monitoring people entering or leaving restricted areas and venues, equipment location, searching for people, guiding services. Unfortunately, GPS localization does not work properly indoors. There is a strong need to provide indoor location tracking in a different manner. There have been many solutions for indoor localization developed so far. Just to mention systems based on infrared sensors [15, 18], ultrasound [16], or magnetic fields [4, 6]. In recent years we have seen the growing popularity of Bluetooth-based location systems named iBeacons [17]. However, these systems require the installation of additional devices.

On the other hand, WiFi-based location tracking is also becoming popular: as there are many access points in offices using WiFi, such location tracking should be costless. Since RADAR [1], many WiFi-based indoor localization solutions were proposed and developed [2, 7, 9, 11–14]. In general, WiFi location tracking uses a map as a reference. There are many types of maps, which can be divided into discrete and continuous maps [10]. Continuous maps allow determining the position as a point on the map. Such precision in WiFi location tracking is possible to obtain under certain conditions: homogeneous devices, defined obstacles, no interference from people. It is useful when we have to know the precise position, e.g. to control a robot [11]. A discrete map usually defines certain zones which are treated as positions on the map. Such location tracking is less precise (because we cannot determine the position, only the zone) but higher location precision cannot be reliably achieved using the most popular WiFi devices (different radio parameters, different measurement techniques).

The continuous approach usually does not require the venue map preparation stage. The system should have information about the location of access points, the structure of obstacles, the strength of the signal in the current position, and based on this information it determines the location in the venue [11]. However, the calculation is vulnerable to errors due to many factors, such as weather conditions, interference from human bodies, different equipment [7].

Discrete positioning usually requires building a pre-learned set of fingerprints to infer the position of a device. There are many solutions to this issue: hidden Markov models [12], Bayesian filtering [21], clustering techniques [19], and genetic algorithms [3].

WiFi-based localization is currently easy to implement in many buildings using existing infrastructure. Discrete positioning seems to be a good choice considering the variety of WiFi devices, the precision of signal strength measurement, and the most common needs. The main contribution of this paper is to present a new discrete, WiFi signal-based positioning algorithm with the support of an SQL database and specified queries. The new method is based on a pre-learned set of fingerprints of defined locations. It uses fuzzy logic to determine the location and the SQL database and queries directly on the collected data. Additionally, it is possible to regularly update the data – location fingerprints – and the operation does not require restarting the system. In opposition to the mentioned solutions, the location tracking is based on the rooms specified by the system operator, not on the zones calculated by the algorithm. As a result, it is possible to correct a location tracking procedure or a venue map definition manually. From the functional side, we receive specific information about whether the user is in a given room, if the user is missing, or their location cannot be determined.

The paper presents only the logical aspects of the SQL-based fuzzy location procedure: the process of building the map and the algorithm of fuzzy location. The other aspects of the device location such as software and hardware requirements, database structure, and SQL queries are not provided.

The paper is organized as follows: it starts with a short description of existing approaches to device location tracking issues with the discussion of the necessity of a new solution; afterward, the definition of the building map is presented, which is based on the wireless network's signal strength measured in the venues in question; the next part presents the fuzzy location tracking procedure in details and with a simple example incorporating artificial data; finally, the description of experiments and their results are presented. The paper ends with conclusions and possibilities of future works.

2. Related works

As was already mentioned in the previous section, there are many known implementations of WiFi-based indoor location tracking procedures. One of the first and most frequently mentioned is a system called RADAR [1]. In the context of WiFi location tracking, a k-NN classifier was used to decide if a person is in a discrete location (close to the previously measured point). The authors are also considering the direction of the localized person in relation to the transmitter. Eventually, the direction is considered unnecessary.

Another work on WiFi localization is related to SLAM (Simultaneous Localization and Mapping) systems [11]. The authors aimed to build a solution that does not need to use fingerprint maps (building a fingerprint map may be an expensive task). Their solution has to precisely obtain the position which is used to navigate the robot. The position is evaluated based on the measured signal strength of several WiFi sensors. Such algorithms also need to evaluate the signal loss due to propagation and obstacles.

The fingerprint solution is a hierarchical topology based on WiFi Indoor Localization [10]. The test environment is built and then split into smaller sub-zones with a reduced number of WiFi Access Points (AP) and reference positions to be identified. The hierarchical partition of the map is created using a KMeans clustering algorithm and the Calinski–Harabasz Index. The authors try to use different classifiers: K-NN, SVM, FURIA. According to the authors, the results are superior to those of the RADAR system. The hierarchical approach makes the solution scalable from small to rather huge environments (several floors, many venues).

The authors of the Fuzzy Logic Based System [7] focus on measurement issues. The signal level can be influenced by weather conditions, nearby devices, obstacles, people in rooms, different radio signal frequencies. These all lead to drops and peeks in the signal strength of access points. The authors propose several pre-processing and post-processing techniques to improve the quality of measurements. They also propose the fuzzy calculations approach to position detection.

3. Building map

The natural solution to the issue of determining the current position is to measure distances from the reference points of known coordinates. Based on these data and after some necessary calculations the location can be specified. Generalizing the problem, it can be stated that the reference points impose (constitute) a local coordinate system in relation to which furthe analyses will be carried out. In the defined area, it is possible to measure the distance to reference points and in this way determine the coordinates of a point in the local coordinate system.

A generalized localization case consists of two stages. The first stage is to define the local coordinate system and determine the reference points appropriate to the local system. The second stage is to compute the relation between the measuring point and the reference points. Finally, after making the calculations, the measuring point receives the coordinates in the local system coordination system.

The first stage involves installing the infrastructure, e.g. setting reference points. This was called defining a map of a building covered by the location system.

The process of making the map involves selecting rooms in the building. Not all rooms in the building were considered to be attached to the building map. The selection of rooms was conducted according to the following criteria:
- proximity of the rooms on the same floor, horizontal distance;
- proximity to the rooms on different floors, vertical distance;
- different types of walls separating the rooms;

- distant rooms with multiple separating walls;
- rooms between which there is visibility through the windows;
- large rooms.

A large room was divided into 4 subzones. This selection of rooms aimed at using the results collected during the experiment for later analysis.

The data collection procedure was as follows:
1) A person enters the room and stops near its centre, e.g. at the reference point. (The room divided into 4 subzones is an exception: the procedure was performed separately for each subzone).
2) All WiFi signals detected by the phone are measured (this step required a dedicated application for Android OS).
3) Measurements (WIFIs, strengths, room, device id, etc.) are stored in the database.

Additionally, the order of measurements in the location and the order of locations are saved.

The sample data are presented in Table 1, where:
- id – database technical row identifier,
- device_id – identifier of the device used to provide measurements,
- BSSID – network identifier,
- point_id – reference point identifier which represents a room,
- strength –signal strength in dB measured at the reference point,
- m_in_point – order number, grouping rows in one measure probe and numbering according to the reference point,
- m_in_total – order number, the same as m_in_point but numbering according to all gathered samples.

For better generalization of measurements, five different devices operating under Android OS were used. Data acquisition was performed at different times and at various intervals to avoid measurement result contamination with particular environmental features in which the measurements were made. Additionally, the measurements were taken manually so the position of the measuring device relative to the designated reference point in the room changed slightly with each measurement. Over 30 measurements were made at each point. It was the consequence of the strength variability of the measured WiFi signal resulting from many factors independent of the measurement process and of the WiFi technology itself.

After 30 individual measurements in all reference points have been taken, each of them was described by the following dataset:
- BSSID,
- average signal strength,
- standard signal strength deviation.

This dataset defines the coordinates of the reference point in the map space. It should be emphasized that due to the attenuation of the WiFi signal, the sets describing reference points contain a different number of BSSIDs. Of course, the average signal strength and standard deviation are also different. Table 2 presents sample descriptions of two reference points.

Finally, after collecting all the measurements for each reference point, the coordinates identifying the point in the building space were determined using the WiFi signal strength. For each reference point, the average and standard deviation of the WiFi signal strength of each WiFi network available at that point were determined. The networks were identified by BSSID. As a result, each reference point is identified by the coordinates resulting from all the WiFi networks available at the point. Additionally, for each average signal strength, the standard deviation is calculated that will be used in the localization procedure.

Table 1. Sample data describing WiFi signal strength in different locations measured with different devices

id	device_id	BSSID	point_id	strength	m_in_point	m_in_total
1800	1	06:d6:aa:35:1b:5f	12	-42	6	70
1801	1	00:17:c5:e3:75:5e	12	-66	6	70
1802	1	00:17:c5:e3:75:5d	12	-66	6	70
1803	1	00:17:c5:e3:76:e9	12	-75	6	70
1804	1	00:17:c5:e3:76:ea	12	-75	6	70
1805	1	00:17:c5:e4:42:9b	12	-82	6	70
1806	1	00:17:c5:e4:42:9c	12	-83	6	70
3249	1	06:d6:aa:35:1b:5f	12	-26	7	105
3251	1	00:17:c5:e3:75:5e	12	-65	7	105
3255	1	00:17:c5:e3:75:5d	12	-66	7	105
3258	1	00:17:c5:e3:76:e9	12	-66	7	105
3261	1	00:17:c5:e3:76:ea	12	-66	7	105
3287	1	06:d6:aa:35:1b:5f	12	-30	8	106
3289	1	00:17:c5:e3:75:5e	12	-67	8	106
3291	1	00:17:c5:e3:75:5d	12	-67	8	106
3293	1	00:17:c5:e3:76:e9	12	-76	8	106
3295	1	00:17:c5:e3:76:ea	12	-76	8	106
3298	1	68:a0:f6:17:32:fa	12	-82	8	106
3303	1	00:17:c5:e4:42:9b	12	-83	8	106
3307	1	00:17:c5:e4:42:9c	12	-86	8	106

Table 2. Sample descriptions of two reference points

point_id	BSSID	strength_avg	strength_stddev
13	06:d6:aa:35:1b:5f	-46.48	9.25
13	0c:37:dc:b7:96:f6	-80.54	4.33
13	e8:de:27:b7:6a:fe	-83.79	2.42
13	4c:5e:0c:7e:ca:ed	-61.09	7.83
13	56:67:11:e4:40:f4	-89.80	4.21
13	92:5c:14:56:13:31	-90.00	3.46
13	92:5c:14:ac:20:26	-90.20	1.30
13	90:5c:44:ac:20:26	-90.89	1.05
13	54:67:51:e4:40:f4	-88.00	4.56
13	9c:b2:b2:5f:b1:67	-84.00	7.44
13	68:a0:f6:17:32:fa	-70.50	6.15
14	00:1d:7e:bc:45:af	-87.33	6.35
14	00:17:c5:e3:71:2f	-80.83	4.46
14	00:17:c5:e3:71:fe	-68.41	7.27
14	00:17:c5:e3:76:ea	-92.17	2.71
14	00:17:c5:e3:77:b0	-93.00	0.00
14	00:17:c5:e3:81:01	-82.96	3.78
14	00:17:c5:e3:f3:c1	-88.40	3.51
14	00:17:c5:e3:f7:2a	-77.91	4.07
14	00:17:c5:e3:71:2e	-80.87	4.20

4. Fuzzy location procedure with an example

In the research, the fuzzy representation of venue assignment was assumed. The notion of fuzziness – proposed over 50 years ago in [20] – is one of the most common approaches of uncertainty representation. Such ability of uncertainty processing is a significant advantage when variable signals – like the WiFi network strength – are being analyzed.

The positioning of the device is similar to the map development procedure described earlier. Unlike map creation, positioning is an unattended process and is performed once for each location determination attempt. At the time of the measurement, the position of the device (the person holding the telephone) is unknown and the measurement is performed once or several times at short intervals. Repeating the signal strength measurement in several second intervals is intended to reduce the impact of WiFi signal variability. As a result, the networks whose signals are very week (or cannot be detected by the devices) are eliminated. Data concerning networks and signal strength collected by the device that wants to know its location are compared with data describing reference points. The similarity of an unknown-location point to the reference points (for which the location is known) is determined. As a result of the algorithm, we get a list of similarities. After sorting, at the top of the list is the reference point nearest to which the measuring device (smartphone) is located. This reference point is considered the device position. The presented fuzzy location procedure will now be described in more detail.

All visible WiFi networks are read at the unknown location point. The networks are identified thanks to BSSID. For each network, the signal power that reaches the measuring device from the access point is read. An example dataset is provided below in Table 3.

Table 3. Detected WiFi networks and their signal strength

BSSID	signal strength
4e:8e:ff:8c:5a:8f	-31
4c:9e:ff:8c:5a:8e	-33
4c:9e:ff:8c:5a:8f	-43
4e:8f:ff:8c:5a:80	-43
4c:9e:ff:8e:1c:9a	-54
4e:1f:ff:8e:1f:10	-59
4e:20:ff:8e:1f:21	-63
4c:9e:ff:8e:1f:20	-64
4e:9a:ff:8e:1c:9b	-67
4c:9e:ff:8e:1c:9b	-67
4e:9b:ff:8e:1c:9c	-67

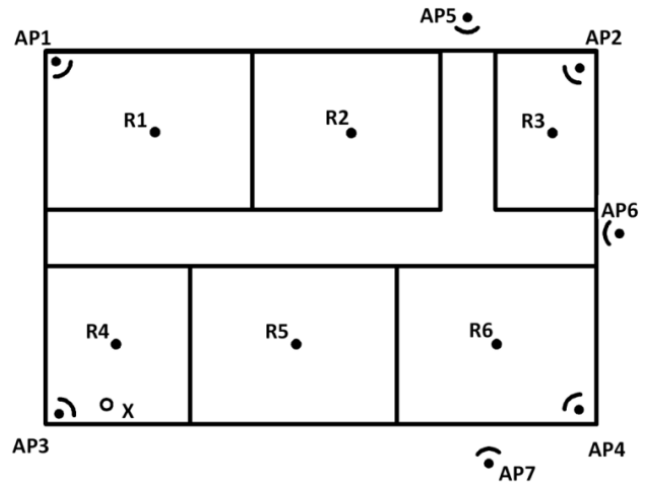*Fig. 1. A sample building floor plan*

Table 4. Detected WiFi networks and their signal strength

Description of point R1		
Net ID	Average signal strength (s)	Standard deviation (std)
AP1	-42	3
AP2	-68	7
AP3	-55	6
AP5	-91	4

Description of point R2		
Net ID	Average signal strength (s)	Standard deviation (std)
AP1	-62	5
AP2	-59	6
AP3	-69	5
AP4	-72	7
AP5	-82	9
AP6	-94	6

Description of point R3		
Net ID	Average signal strength (s)	Standard deviation (std)
AP1	-70	8
AP2	-38	6
AP4	-65	7
AP5	-74	9
AP6	-71	7

Description of point R4		
Net ID	Average signal strength (s)	Standard deviation (std)
AP1	-57	9
AP2	-98	7
AP3	-41	6
AP4	-79	8
AP6	-95	5
AP7	-87	9

Description of point R5		
Net ID	Average signal strength (s)	Standard deviation (std)
AP1	-61	7
AP2	-89	8
AP3	-54	5
AP4	-68	8
AP6	-89	7
AP7	-78	5

Description of point R6		
Net ID	Average signal strength (s)	Standard deviation (std)
AP1	-85	9
AP2	-78	7
AP3	-79	6
AP4	-39	8
AP5	-87	5
AP6	-74	9
AP7	-77	7

The process of measuring and determining the position of the device can be presented in several steps described below. Prior development of the map is necessary for determining the position, as described above. In order to illustrate the position determination process, we assume that there is a map prepared for the building fragment as shown in Fig. 1.

Points R1 to R6 are the map reference points, example descriptions of these points are presented in Table 4. Point X is the measuring point for which the location will be determined, an example measurement made here is shown in Table 5. Points AP1 to AP7 represent the location of the wireless network access points. AP5 to AP7 are located outside the analyzed area, but the signal of these access points is visible in the area covered by the map. In the example tables, instead of the full identifier of the BSSID network, short identifiers were included to increase the transparency of the example. Similarly, most numbers are represented as integers.

The fuzzy location procedure consists of the following steps:

Step 1: A person who wants to determine their position in the building, goes to any place covered by the map. Using the mobile device, the person measures WiFi networks visible there. Together with the BSSID network, the strengths of individual signals are read. An example measurement taken in room 4 at point X can provide the following results as presented in Table 5.

Step 2: The collected results are compared with the building map. The quantity condition of the network is checked first. For each reference point, we compare a set of network identifiers assigned to that point with a set of network identifiers visible at point X. Next we determine a common part of these two sets.

Table 5. Results of measurement made at location X

Net ID	signal strength (s)
AP1	-68
AP3	-38
AP4	-78
AP7	-75

Using the sample data, it can be seen that the measurement point X and the reference point R3 have three common network identifiers: AP1, AP4. In the case of X and R4, common identifiers are: AP1, AP3, AP4, AP6.

Step 3: For each reference point, a network factor is determined that reflects the ratio of found networks (in reference to the number of networks assigned to the reference point). If the factor takes on a value of more than 0.5, the reference point is taken into account in the next steps. Otherwise, the reference point is rejected and is not included in further calculations. This approach reduces the computational complexity of position determination but with the possible cost of decreased location accuracy (or the time of location increase).

The networks factor for the pair measurement point and the reference point is determined by the formula:

$$f_{X,R_j} = \frac{n_{X,R_j}}{n_{R_j}} \tag{1}$$

where f_{X,R_j} is the network factor, n_{X,R_j} is the number of common network identifiers for sets describing the X measurement point and reference point, R_j, and n_{R_j} means the number of network identifiers describing the reference point R_j.

For sample data, the factor summary is presented in Table 6.

Table 6. Network factors for a reference point, based on measurement from location X

j	n_{X,R_j}	n_{R_j}	f_{X,R_j}
1	2	4	0.50
2	3	6	0.50
3	2	5	0.40
4	4	6	0.67
5	4	6	0.67
6	4	7	0.57

Based on network factors, reference points R1, R2 and R3 will be rejected from further processing.

Then, for the remaining reference points, the strengths of the network signals from the measurement point set are analyzed, along with the mean signal strengths and the standard deviation of the network in the set for each reference point. For each pair of sets the distance factor is determined according to the following formula:

$$d_{X,R_j} = \sum_{i=1}^{n_{X,R_j}} rm_i(X, R_j) \cdot c_i(X, R_j) \qquad (2)$$

For each considered network – i – the multiplication of two functions is performed. The first can be called the reference point assignment and is defined as the fuzzy assignment of the measurement point X to the reference point R_j:

$$rm_i(X, R_j) =$$
$$\begin{cases} 1, & |s(R_j, i) - s(X, i)| < 1 \\ 0.5, & |s(R_j, i) - s(X, i)| \in [std(R_j, i), 2 \cdot std(R_j, i)) \\ 0.25, & |s(R_j, i) - s(X, i)| \in [2 \cdot std(R_j, i), 3 \cdot std(R_j, i)) \\ 0, & |s(R_j, i) - s(X, i)| \geq 3 \cdot std(R_j, i) \end{cases} \qquad (3)$$

where $s(R_j, i)$ is the average signal strength for the network from the description of the reference point, R_j is the measured signal strength for the network at the measuring point R_j, and $std(R_j, i)$ is the standard deviation of the signal strength for the network i from the description of the reference point R_j.

In the early stages of location procedure development, many different fuzzy assignment functions were taken into consideration: triangular, piece-wise linear, trapezoidal, trigonometrical (cosine and arcus tangent-based), and many more. However, from the point of view of location accuracy and implementation complexity (SQL queries) the presented solution gave the best results.

The second function c represents the signal strength correction factor and is defined as follows:

$$c_i(X, R_j) = \begin{cases} 0.2, & s(R_j, i) \in (-100; -90] \\ 0.5, & s(R_j, i) \in (-90; -80] \\ 0.8, & s(R_j, i) \in (-80; -55] \\ 1, & s(R_j, i) \in (-55; -35] \\ 0.8, & s(R_j, i) \in (-35; 0] \end{cases} \qquad (4)$$

The results of using distance only based on reference point assignment functions were unsatisfactory. This was mostly due to the fact that the radio signal strength results from many other factors such as transmitter power, the output of the transmission and reception antennas, attenuation on the receiver and transmitter, and signal reflections (amplifying or weakening the signal in the receiver). This implied introducing an additional correction factor whose final form, presented above, was obtained through trial and error.

Step 4: After performing the above calculations, the determined distance factor is corrected with the previously calculated network factor f_{X,R_j}. The correction takes place for each reference point.

This is expressed by the formula:

$$f_{X,R_j} = \frac{n_{X,R_j}}{n_{R_j}} \qquad (5)$$

As a result, after performing the above calculations, we obtain a measure matching measurement point X to individual reference points. As the obtained all room assignment levels have positive, non-limited values, they must be normalized. The range $[0; \max_j m_{X,R_j}]$ is scaled linearly to $[0;1]$.

The results showing the measure of matching the X measuring point to the map reference points are presented in Table 7.

Table 7. Raw levels of assignment and normalized levels of assignment

j	m_{X,R_j}	normalized m_{X,R_j}
4	1.63	1.00
5	1.07	0.65
6	0.60	0.37

After sorting the table according to the value of the measure of matching from the highest to the lowest value, in the first place we obtain the location of the measurement point. In the example, this is the reference point, i.e. based on the performed calculations it can be concluded that the measurement was conducted in room 4. The obtained results can be filtered by rejecting the values of the matching measure which did not exceed the limit value. Such filtering, however, requires additional research that will allow to deliberately set a results rejection threshold. Without the use of such filtering, the order in the list shows the distance of the measuring point from the reference point. The measuring point is located closest to the reference point in the first position and farthest from the reference point in the last position.

5. Experiments

5.1. Data acquisition

The building, in which the procedure was developed and applied, is a three-storey one with several dozen rooms on each floor. In the experiments, 20 of them were taken into consideration. One of them was the conference hall which was divided into four separate areas, so the final number of locations was 23. The criterion for selecting these rooms was the frequency of changes of the monitored people's locations.

As it occurred, inside these rooms over 100 different wireless network IDs were detected. In each room, 30 measurements of wireless network signal levels were conducted. These measurements were taken using several different types of mobile phones. The highest number of detected signals in a room was almost 35. This implies that the raw data were rather sparse. It also confirms that in the presented problem a non-resistant method could not be applied for missing values.

Finally, the collected dataset consisted of 690 objects and over 100 features. The measured signal strength was expressed in dB and, as it is assumed in the wireless network area, the signal strength is a negative number, and as the level of the signal decreases, the value of the strength decreases too.

A sample part of the collected data is presented in Table 8.

Table 8. Sample data acquisition results

mes_id	dev_id	n001	…	n012	n013	n014	n015	…	n100	ven_id
595803	2				-71	-71				16
595804	1			-33	-63	-63	-73			16
595812	2									17
595813	2									17
595814	1			-63	-62	-63	-61			17
595815	1									17
595831	2				-80	-80				14
595832	1				-67	-66	-76			14
595833	1			-80						14

The mes_id column represents the specific measurement identifier, the dev_id column is connected with the specific measurement device, columns from n001 up to n100 deal with wireless networks, and finally – ven_id – is the venue identifier in the building. The collected data, due to the nature of the

environment (dozens of networks and hundreds of measurements detecting only several networks), are characterized by a very high number of missing values. The total amount of non-missing ones was approximately 6%. For that reason, the application of the most popular multiclass classifiers was impossible.

5.2. Simple experiments

All experiments were carried out in the leave-one-out/stratified cross-fold validation method. In each of the 30 iterations, exactly one measurement from each venue was moved to the test set. This ensured that the fuzzy maps of venues were built on 29 measurements (each of the maps). On the other hand, the proportion of venue representation in the training and test set was equal. The proposed scheme also ensured that the prediction accuracy, measured as the fraction of correctly and incorrectly classified positions, is equal to the class-weighted accuracy. Finally, it is possible to present the prediction results in one confusion matrix as each measurement is only taken once as the testing object.

The results of the localization procedure are presented in Table 9.

Table 9. Confusion matrix for the classified objects

original \ predicted	1	2	3	4	5	6	7	8	9	10	12	13	14	17	18	19	20	21	22	24	25	26	27
1	21	3	6																				
2	2	20	1	5					2														
3	1		28						1														
4	3	13	1	13																			
5					30																		
6		1				28													1				
7	3			1	1		13	2						6	1				3				
8						1		28											1				
9									28	2													
10									3	26									1				
12											25						1		4				
13									2			27							1				
14			2										23				4		1				
17					9	1	1							8	9	2							
18					1										27				1				1
19		1							1						5	19	3		1				
20													1			1	28						
21															10	1		15	4				
22			1						2								4		22	1			
24											1				1				4	24			
25					1														3		26		
26	1								2							1	3	4	12			7	
27															1								29

The total accuracy reaches a level of almost 75% (74.64%). The least class prediction accuracy was at the level of 23.33% but it should rather be considered the outlier because the median is 83.33%, which is quite more than the average. It is also worth noticing that none of the objects was unclassified. In other words: each wireless network signal level measurement was classified to one of the possible venues.

Let us remember that the first four places can be interpreted as one location (the conference hall divided into four regions). That simplification increases the prediction accuracy up to 78.63% and the median up to 86.66%. However, from the perspective of utility, we would not consider the class aggregation in further experiments.

5.3. Haste makes waste

Typically in the case of classification or regression tasks, a situation where an object is not classified to any known class (it is not possible to predict the real value of the dependent variable) is usually considered the misclassification (the error of predicting a known value). This results from the fact that the environment of independent variables remains unchanged. This is very easy to explain: in the case of character recognition, the image of acquired pixels does not change in time; while in the case of real values prediction, the constant current value of the predicted variable should be constant due to constant independent variables values. However, in the case in question, when a person enters the room, we can agree that as long as that person stays in the same room the prediction may be done at a certain cost in terms of prediction duration time. In other words: in such a case – haste makes waste. We agree for the longer time of prediction provided that the prediction accuracy also increases. This issue will be explained later in this section.

The maximum number of networks in the venue maps is 33. This leads to the remark that we may define the condition of the minimal percentage of venue map networks to be recognized by the measuring device (network factor) as the necessary condition for taking the map into consideration. For example: if the map is built on the strength of 10 network signals and the assumed minimal percentage of the measured signal is 25%, at least three signal levels must fulfill the map criteria to begin to consider the location pointed by that map.

Let us consider all possible signal threshold levels, varying from 0 to 1. The smallest sensible value of the threshold is 0.03 as there is no map built on more than 30 network signals. Then, similar experiments were carried out for each considered threshold. The results are presented in Figure 2.
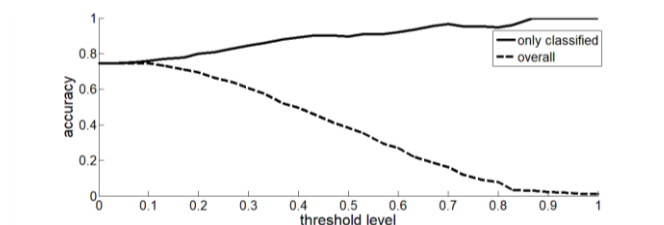
Fig. 2. Data classification accuracy

The solid line represents how the classification accuracy of the map covered the sample changes due to the threshold increase. Starting from the value of 0.7464 (which is the same as in the previous experiments) it increases to 1.0 (for the threshold of 0.8700). The dotted line also represents the classification accuracy but in this case, unclassified objects are considered wrongly classified. We can observe the intuitive effect: as the threshold increases, the accuracy decreases. The decrease in the accuracy is caused by the decrease in the data coverage. In the next figure (Figure 3) we can compare both tendencies.
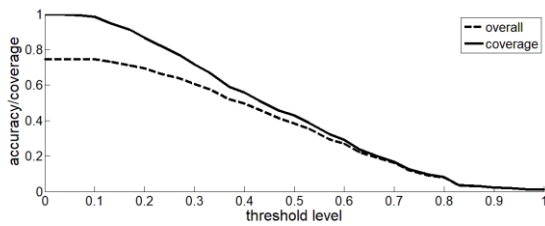
Fig. 3. Comparison of accuracy and coverage decrease

The intuitive reflection about the increasing range of unclassified objects says that the model degrades with the threshold level.

It is a well-known procedure of parametrized classifier quality description to present the chart called ROC – Receiver Operator Curve. In terms of the classification issue, it shows how the true positive and false positive rates change due to the classifier parameters modification. In our case, it is worth taking into consideration two other criteria. We are interested in high accuracy and low non-coverage (difference between a set of objects and a set of objects classified by the model/map - the measure complement to the coverage). Let us check how this dependence looks like in the case of our predictor. ROC is presented in Figure 4.
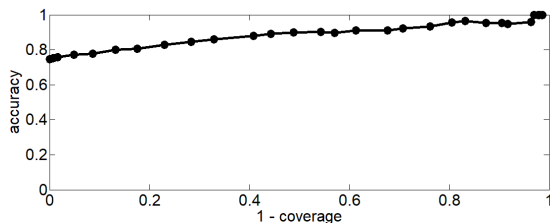
Fig. 4. Comparison of accuracy and coverage decrease

The presented figures confirm that in this specific case, with changing environmental conditions, the "lazy" decision is preferred due to its correctness and it is worth considering the threshold increase in the real application. The following section will be devoted to proper threshold level selection.

5.4. Network factor level selection

As it was mentioned in the previous section, the main goal of the issue is to predict the proper location even at the cost of time spent to obtain this prediction. So it is important to define the appropriate compromise between the coverage of the model and the accuracy of the model predictions (limited only to the final statements – unclassified objects do not degrade the prediction accuracy).

One of the most intuitive approaches is to compare the coverage decrease and the accuracy increase due to the threshold increase. It is easy to find the threshold level in which the mentioned two measurements intersect. The comparison of data coverage and covered data prediction accuracy is presented in Fig. 5.
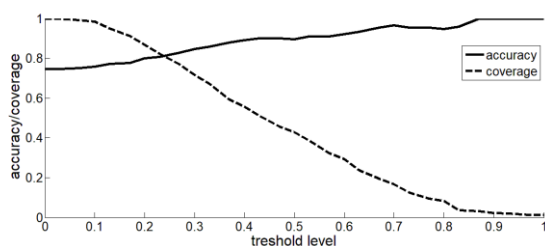

Fig. 5. Coverage and covered data accuracy

The analysis of that figure should imply the proper threshold level on the value equal to 0.23. A higher level will cause the coverage to be lower than the accuracy. With the assumed threshold, the accuracy would be equal to 0.807.

Another approach is to build the average of mentioned measures. The average result is presented in Fig. 6.
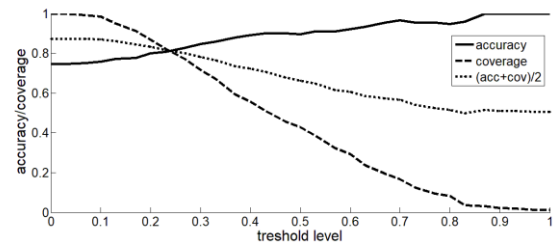

Fig. 6. Coverage and covered data accuracy together with their average

It occurs that such averaged measure decreases at the threshold level equal to just 0.1. That means that the optimal threshold value is a step-before one (0.066) and this situation corresponds with the initial prediction accuracy: 0.7518.

The mentioned two approaches are based on the raw quality measurement values. However, the most important quality criterion should take into consideration the real application results. The increase of the threshold value simultaneously increases the accuracy and decreases the data coverage, which was presented in the previous section. The crucial question is "How long can the threshold be increased to assure all venues be covered by the prediction results?''. In other words, what is the maximum threshold level that assures at least one correct classification into each of the possible venues?

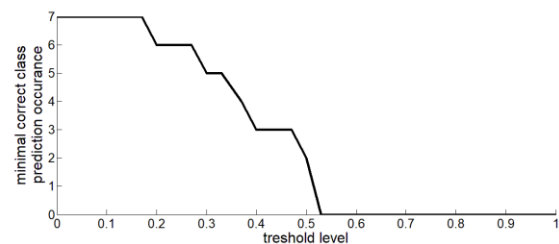This leads us to the results presented in Fig.7.

Fig. 7. The minimum number of objects classified correctly to the class

It occurs that the maximum threshold level that assures at least one correct object classification (for each possible object class) is 0.50. Up to this level, it is only a matter of time to obtain the correct location for the considered device. Over this level, there is a possibility that we are in a venue that we will never be classified into. The classification accuracy at this threshold level is 0.8953.

6. Conclusions and further works

The paper presents a novel approach to the location issue based on the wireless network signal measurement. A single measurement observed with the mobile phone is compared with the previously built description of all venues in the building. The final venue assignment is made based on the fuzzy assignment to each of the considered rooms.

The developed strategy of building the description (map) and the fuzzy location procedure is now applied in the people location tracking system developed at the Institute of Innovative Technologies EMAG.

The presented solution is adaptable but it requires an advanced implementation phase: the process of creating the map of the building based on averaged measurements. Additionally, the analysis of the proper network factor should be performed for each building separately.

—— IAPGOŚ 2/2021 ——

Our later works will focus on further fuzzy location improvement – especially on using other wireless signal measures to improve the location accuracy (e.g. beacons).

The data acquired during experiments are also a valuable source of information on how the location procedure duration depends on the network factor: it was shown that increasing this factor improves the classification accuracy but only if the condition of the minimum number of detected networks is fulfilled. A more detailed analysis of these durations may help to define a longer time of procedure application with the advantage of better location accuracy.

References

[1] Bahl P., Padmanabhan V. N.: RADAR: An in-building RF-based user location and tracking system. Proceedings IEEE INFOCOM 2000, 775–784.
[2] Benavente-Peces C. et al.: Global System for Localization and Guidance of Dependant People: Indoor and Outdoor Technologies Integration. Lecture Notes in Computer Science 5597, 2009, 82–89.
[3] Chintalapudi K., PadmanabhaIyer A., Padmanabhan V.: Indoor localization without the pain. Proceedings of the sixteenth annual international conference on Mobile computing and networking, 2010, 173–184.
[4] Chung J. et al.: Indoor Location Sensing Using Geo-Magnetism. Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, 2011, 141–154.
[5] Enge P., Misra P.: Special issue on Global Positioning System. Proceedings of the IEEE 87(1), 1999, 3–15.
[6] Galván-Tejada C. E., Carrasco-Jimenez J. C., Brena R.: Location Identification Using a Magnetic-Field-Based FFT Signature, Lecture Notes in Computer Science 8276, 2013, 9–16.
[7] Garcia-Valverde, T., Garcia-Sola, A., Hagras, H., Dooley, J. A., Callaghan, V., Botia, J. A.: A fuzzy logic-based system for indoor localization using WiFi in ambient intelligent environments. IEEE Transactions on Fuzzy Systems 21(4), 2013, 702–718.
[8] Górka W., Piasecki A. Socha M.: Mobile application supporting universal access to culture, taking into account the needs of disabled people. Proceedings of the 11th Scientific Conference Internet in the Information Society, 2016, 191–201.
[9] Hammadi O.A., Hebsi A. A., Zemerly M. J., Ng J. W. P.: Indoor localization and guidance using portable smartphones. Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2012, 337–341.
[10] Hernández N., Alonso J. M., Ocaña M.: Hierarchical Approach to Enhancing Topology-based WiFi Indoor Localization in Large Environments. Journal of Multiple-Valued Logic and Soft Computing 26(3-5), 2016, 221–241.
[11] Herranz, F., Llamazares, Á., Molinos, E., Ocaña, M., Sotelo, M. A.: WiFi SLAM algorithms: An experimental comparison. Robotica, 34(4), 2016, 837–858.
[12] Krumm J., Horvitz E.: Locadio: Inferring motion and location from Wi-Fi signal strengths. Prpceedings of The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004, 4–13.
[13] Liu H. et al: Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 37:6, 2007, 1067–1080.
[14] López E., Barea R., Bergasa L. M., Escudero M. S.: A human-robot cooperative learning system for easy installation of assistant robots in new working environments. Journal of Intelligent and Robotic System 40(3), 2004, 233–265.
[15] Polito S. et al.: Performance evaluation of active RFID location systems based on RF power measures. IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, 2007, 1–5.
[16] Priyantha N. B., Chakraborty A., Balakrishnan H.: The cricket location support system. Proceedings of the Annual International Conference on Mobile Computing and Networking, 2000, 32–43.
[17] Steinbuss S., Holtkamp B., Opriel S.: HANDELkompetent – Situation Aware Learning in Retail. Procedia Manufacturing 9, 2017, 245–253.
[18] Want R., Hopper A., Falcao V., Gibbons J.: The Active Badge Location System. ACM Transactions on Information Systems 10(1), 1992, 91–102.
[19] Youssef M., Agrawala A.: The Horus WLAN location determination system. Proceedings of the 3rd international conference on Mobile systems, applications, and services, 2005, 205–218.
[20] Zadeh L.: Fuzzy sets. Information and Control 8(3), 1965, 338–353.
[21] Zaruba G., Huber M., Kamangar F., Chlamtac I.: Indoor location tracking using RSSI readings from a single Wi-Fi access point. Wireless Networks 13(2), 2007, 221–235.

M.Sc. Michał Socha
e-mail: Michal.Socha@emag.lukasiewicz.gov.pl

Michał Socha was born in Poland in 1972. He received his M.Sc. Eng. in computer science from the Silesian University of Technology in 1997 and has begun Ph.D. degree study in 2020 at the same university. His scientific interests cover machine learning, geographic information systems, software architecture design. He is an author and co-author of over 20 scientific papers.

https://orcid.org/0000-0002-7439-4001

M.Sc. Wojciech Górka
e-mail: Wojciech.Gorka@emag.lukasiewicz.gov.pl

Wojciech Górka was born in Poland in 1979. He received his M.Sc. Eng. in computer science from the Silesian University of Technology in 2003. Since 2006 he has been working at the Łukasiewicz EMAG research institute (former ITI EMAG). His scientific interests are natural language processing, technology supporting the blind and deaf, and mobile devices and their ability to locate in indoor areas.

https://orcid.org/0000-0003-2124-7734

Ph.D. Marcin Michalak
e-mail: Marcin.Michalak@emag.lukasiewicz.gov.pl

Marcin Michalak was born in Poland in 1981. He received his M.Sc. Eng. in computer science from the Silesian University of Technology in 2005 and a Ph.D. degree in 2009 from the same university. His scientific interests cover machine learning, data mining, rough sets, and biclustering. He is an author and co-author of over 90 scientific papers.

https://orcid.org/0000-0001-9979-8208