Fuzzy logic in indoor position determination system

Michał Socha\textsuperscript{1*}, Wojciech Górka\textsuperscript{1†}, Iwona Kostorz\textsuperscript{1‡}

\textsuperscript{1}Instytut Technik Innowacyjnych EMAG, Leopolda 31, Katowice, Poland

Abstract. The article outlines how to use the convergence of collections to determine the position of a mobile device based on the WiFi radio signal strength with the use of fuzzy sets. The main aim is the development of the method for indoor position determination based on existing WiFi network infrastructure indoors. The approach is based on the WiFi radio infrastructure existing inside the buildings and requires operating mobile devices such as smartphones or tablets. An SQL database engine is also necessary as a widespread data interface. The SQL approach is not limited to the determination of the position but also to the creation of maps in which the system defining the position of the mobile device will operate. In addition, implementation issues are presented along with the distribution of the burden of performing calculations and the benefits of such an approach for determining the location. The authors describe how to decompose the task of determining the position in a client-server architecture.

Keywords WiFi; indoor location; fuzzy sets

Received 11 JAN 2016 Revised 11 JUL 2016 Accepted 18 JUL 2016

This work is published under CC-BY license.

1 INTRODUCTION

Nowadays a number of WiFi radio signals are present in public buildings (offices, shopping centres, etc.) as well as in private apartments. It enables the determination of position by the utilization of WiFi signals on mobile devices inside these buildings. The main problem is the accuracy of the determination [1, 2].

To determine the position of a mobile device the strength of the WiFi radio signal of nearest access points can be used. Unfortunately, this signal is very susceptible to interference originating from the people present inside. This causes significant changes of the signal strength [3]. In

\textsuperscript{*}E-mail: msocha@ibemag.pl
\textsuperscript{†}E-mail: wgorka@ibemag.pl
\textsuperscript{‡}E-mail: ikostorz@ibemag.pl
this case, the relative position of a mobile device to the access point does not change while the interpretation of the WiFi signal strength indicates the mobile device movement in relation to the access point. The floating position of the mobile device WiFi antenna to the access point affects the position determination process as well. Obviously, there are other sources of the interference but their influence on the position determination is unpredictable. As the result of observations, it has been established that the presence of people inside and the changes of the mobile device antenna positions have significant influence on lowering the accuracy of calculations. It seems to be impossible to find a mathematical model taking into account all interferences [4, 5, 6].

2 Position determination

Taking into account the results of the previous research (also the authors’ research) [1, 2, 3, 4, 7, 8, 9, 10] on the WiFi signal attenuation and requirements of OpenArt project, the authors decided to find a method for the position determination meeting the following conditions:

- the position must be determined with the ‘one room’ precision,
- the calculations must be adjusted to the speed of a person’s walk,
- the computational complexity must not influence the other applications running on the mobile device,
- the preparation time of the whole system to run indoors must be as short as possible.

The main elements of the proposed solution have been shown in Fig. (1).

The data measured in all rooms that have been scanned are stored in one database system. All data is aggregated in one view. The aggregation process uses statistical functions provided by the database engine: average $\text{avg}$ and standard deviation $\text{stddev}$. The processed data (the view) makes the map of rooms in the position determination system. The view may be downloaded on demand to the mobile device local database. Aggregated data for each room or the area in the system is the map set $M$. The set $M$ consists of triples

$$
M = \{ N_{ID}, \text{APNS}, \text{STD} \},
$$

where:
- $N_{ID}$—the network identifier,
- $\text{APNS}$—the average power of a WiFi network signal (the RSSI of the received power in dB without normalization between different devices),
- $\text{STD}$—the standard deviation of all measurements.

\[^{1}\text{All research and implementation have been carried out as a part of the project ‘Sztuka współczesna dla wszystkich OpenArt’ (IS-1/021/NCBR/2013)}\]
The mobile device that is situated in the position determination system area measures and stores the data as a set $P$ of values. The set $P$ is given by a collection of pairs

$$P = \{N_{ID}, \text{APNS}\},$$

where:

- $N_{ID}$—the network identifier,
- $\text{APNS}$—the average power of the WiFi network signal measured by the mobile device.

It was assumed that the comparison of the data set from the mobile device with the map would determine the position of the device. Thus, the similarity of sets must be calculated.

## 3 DATABASE ENVIRONMENT

The PostgreSQL database engine has been used for the research. The PostgreSQL functions `avg` and `stddev` were used \[11, 12\]. The average value is calculated as the arithmetic mean value. These functions are very easy to calculate and are available in each database. Further experiences could lead to more advanced functions but they seem to be useful as a starting point.

The database structure shown in Fig. (2) is a foundation of the map of rooms in the position determination system. The fragment of the map was presented in the List. [1] The map is built with the use of the SQL query.

The `HAVING` clause has been used for filtering the data that cannot be processed by the database statistical functions. The minimal value of the WiFi signal measurement is 3.
Fuzzy logic in indoor position determination system

Figure 2 Database diagram of localization system

Listing 1 SQL query for building the map of venues

```sql
select venue.name,
       network.bssid,
       avg(measure.strength),
       stddev(measure.strength)
from venue,
     point,
     measure,
     network
where venue.domain_id = 4 -- Testing area
  and venue.id = point.venue_id
  and point.id = measure.point_id
  and measure.network_id = network.id
group by venue.name,
        network.bssid
having count(*) > 2 -- Three or more measures
```
4 Similarity of the sets

The List. 1 representing respectively the map set and the mobile device set, are fuzzy sets according to definitions [5, 13, 14, 15]. The authors assume that to determine the position of the mobile device with a sufficient probability one has to compare the set representing the position of the mobile device with the sets of the map and to compute the level of their similarity.

According to Sachwa [16], the similarity of sets can be determined by various methods: the fuzzy inclusion method, using a metric on the set and using cross sections of the sets. In the case of the data collected during the work of the position determination system, the fuzzy inclusion method cannot be implemented. This method is based on a graphical analysis of the overlapping areas of the sets and their not overlapping parts. Additionally, the fuzzy inclusion method must operate on continuous data sets. All available sets in the system consist of discrete elements. For our purposes the remaining two methods of the similarity determination could be implemented (the method using the metric on the set or the cross section of the sets).

The metric on the fuzzy sets space $F(X)$ over the universe $X$ is the function [16]

$$m : F(X) \times F(X) \rightarrow \mathbb{R}^+ \cup \{0\}, \tag{3}$$

where:

- $m(A, B) \geq 0$,
- $m(A, B) = m(B, A)$,
- $A = B \implies m(A, B) = 0$,
- $m(A, C) \leq m(A, B) + m(B, C)$.

The determination of the similarity of fuzzy sets can be calculated as a complement of their normalized distance. The level of the similarity is given by the formula

$$E(A, B) = 1 - m(A, B). \tag{4}$$

In practice, the similarity of the sets is determined by the means of the metric on the set, and can be performed as follows. For each element belonging to the map set $M$, a one-to-one corresponding element from the set $P$ in the mobile device is found upon the network identifier. Each pair of the sets must meet the condition that a measured value is in the range of the average value and the standard deviation. The value of the similarity can be defined as a quotient of the number of elements of the map set $M$ and the number of elements of the set $P$ measured by the mobile device. This value must be in the range of $R = (\text{avg} - \text{stddev}, \text{avg} + \text{stddev})$. The result is the similarity of two sets. The higher is the value, the more similar the sets are. Such a method of the similarity determination is based on the experimental observations. The research has shown that the more networks in the range of $R$, the more similar the sets are.
Depicting metric similarity is shown in Fig. 3. The hatched bar represents the range $R = (\text{avg} - \text{stddev}, \text{avg} + \text{stddev})$ for the element of the map set $M$. The points marked with $\times$ represent the signal strength measured on the mobile device during WiFi scan (the set $P$).

The map set $M$ includes five elements. Three elements of the mobile device set $P$ are in the range of $R$ (mobile devices in the range of networks 1, 2 and 4) and two of them are out of the range of a given network (mobile devices out of the range of networks 3 and 5). One element of the $P$ set cannot be found in the map set $M$ at all. Thus the similarity of the sets is $\frac{3}{5}$.

In this method two sets are equal if and only if the $P$ set includes WiFi network identifiers being the elements of one of the map sets.

The position determination is performed with the use of the SQL query shown in the List. 2.

The proposed algorithm can be slightly modified by considering the ratio of the measured signal strength to the adequate average value in the map set.

For each obtained pair from $M \cap P$ the difference between the power of the measured signal and the mean power of the analyzed network signal must be calculated. The obtained value of the difference has to be compared with the width of the signal changes determined with the use of the standard deviation parameter. The sum of the values computed for each common element of the sets should be compared with the cardinality of the map set. The result of the comparison is the value of the similarity of two sets. The higher is the value, the more similar the sets are.
Listing 2 SQL query for the position determination
Graphical interpretation of the sets comparison laid the foundation of the method of the similarity determination (see Fig. 4). The points marked as $\times$ represent the element of the measured set $P$ (from the mobile device) for a given network. The hatched vertical bar represents the range $R = (\text{avg} - \text{stddev}, \text{avg} + \text{stddev})$ with the average signal strength (marked with vertical ticks). The dotted bar represents the difference of the average signal strength and the one measured on the mobile device. The values on the right side are the metrics based on the difference of the average signal strength and the measured one on the mobile device related to the standard deviation, for each network according to the Eq. (6).

The higher number of $P$ set elements belong to the map set and the closer they are to the signal average value, the closer is the mobile device to the measurement point.

The equality of sets occurs when the $P$ set contains WiFi network identifiers belonging to the one of the map sets and the values of the signal strength are equal to average values of the map set.

This modification of the algorithm of sets comparison is like the similarity determination based on the cross section analysis. However, due to the different method of the set generation process, it is not exactly the performance of the cross section method. In practical use the cross section of the sets determination process is based on the numerical values analysis for the same identifier networks.

According to Lachwa [16] the cross section level is determined by a network identifier. The number of levels depends on the map set. Each set of the map has a different number of cross section levels. For each cross section level the union and the product of the sets ($M$ and $P$) can be calculated. As a union of the sets the size of the range $R = (\text{avg} - \text{stddev}, \text{avg} + \text{stddev})$ has been taken. The product of the sets is the absolute value of the difference between the value measured
by the mobile device and the average value of the given network signal

\[ |s_{mac} - \text{avg}_{mac}|. \]  

(5)

The similarity of two sets is given by the formula

\[ E(M, P) = \frac{1}{n} \sum_{i=1}^{n} \left( 1 - \frac{|s_i - \text{avg}_i|}{\text{stddev}_i} \right), \]  

(6)

where:

- \( M \)—the map set (the network identifier, the average strength of the network signal, the standard deviation),
- \( P \)—the mobile device set generated during the WiFi network scanning (network identifier, the strength of the signal),
- \( n \)—the cardinality of the map set \( M \),
- \( s \)—the strength of the network signal in the \( P \) set,
- \( \text{avg} \)—the average strength of the network signal in the \( M \) set,
- \( \text{stddev} \)—the standard deviation of measurements.

The above considerations were put into the SQL in List. 3

5 Practical implementation of the solution

The first implementation works were conducted only with the use of the PostgreSQL database. The server and client parts were in a single database and were both checked. The opportunity to test the efficacy of the maps already in the process of their development in the SQL console was a big advantage. It was possible to keep an eye on the impact of individual measurements for determining the result location. On the server side, the data gathering and view generation were checked. The view represented the map of the rooms. On the client side, determining the position based on the scanned WiFi’s was checked. Good preliminary results made it possible to carry out the implementation and separate the client and server side components. The server side was reduced to the database and a simple data diagram.

The data diagram allows to configure freely the map of the rooms which are covered by the localization system. All measurements are stored in the database. Each measurement is assigned to a specifically named point where the measuring device was located. The measurement points are combined into the rooms and that is how the aggregation is carried out. The evaluation of interference of the WiFi signal is the first aim of the aggregation. The second one is to consider different mobile devices that the users will use. Bearing in mind the second objective, the data should be collected by means of various devices.

The process of building the map of the rooms begins by defining and naming the rooms and the measurement points. Then, in such a defined environment, measurements can be carried out.
Generally speaking, the more measurements are made and the more devices are used, the better is the quality of the map. It should be noted, however, that thanks to the database approach after several measurements (at least 3 measurement records for one venue), the map is useful and can ...
be used by the system. However, meaningful results for localization can be obtained when at least 7 measurements for one point are made by three different devices. But these are just observations made during the start-up and not the results of systematic research. Further measurements can be carried out at the same time without any negative impact and can be entered into the database. Each measurement sent to the database improves the quality of the map.

This feature can be used back and forth too. Incorrectly performed measurements can be removed and thus the quality of the map improves. It is also possible to delete measurements which are not significant from the point of view of the room. This will reduce the amount of data and speed up the performance of the system. The Fig. (5) shows the general architecture of the solution.

Mobile devices get the copy of the table with a map of the rooms, for which the localization is to be performed, from the server. The copy is loaded into a local database with an SQL interface. The mobile device determines its location on the basis of this copy of the map. A fragment of the local database diagram is shown in Fig. (6).

The data for maps is stored in `wl_condition` table. This is aggregated data downloaded from the central server that describes the rooms. An additional table, `wl_measure`, represents the current measurement. Due to the fact that there is an SQLite database on the device, queries cannot be too complex (excluding nested queries). Data from each measurement is therefore stored to this (`wl_measure`) table and on its basis the relevant query is made. It is therefore a table with temporary data. It is possible to eliminate the local database on the mobile device.
and use the central server directly. However, given the fact that a query that delivers results for the determination of the position is quite complex and difficult to parameterize, the idea was abandoned. This query, and then the results should be sent over the wireless network between the client and the server. It seems to be an unnecessary network load.

The query in the version for the mobile device is presented in Listing 4 (FOUND_WIFI_COUNT field means an amount of matching networks and is dynamically set in the program).

```
Listing 4 SQL query for the position determination in the version for the mobile device

1  select vc.venue_id,
2    ( FOUND_WIFI_COUNT / count(vc.bssid) ) / avg(
3           abs(vc.condition_avg - m.rssi) / vc.condition_stddev ) as measure
4  from   wl_condition vc
5  left outer join wl_measure m
6    on vc.bssid = m.bssid
7  group by vc.venue_id
8  order by measure desc
```

6 Verification of the method

The applied metric was tested in a simple implementation without any additional development work. Several mobile devices were used for collecting measurements in order to build maps. Data was gathered manually from a WiFi network, using standard applications, and entered into the database. Measurements were also collected from different parts of the building in which the tests were proceeded. This data was used as example collections to verify the measurement method for determining the position.

Then the test queries were executed and their results were checked. The results (metrics) from the database pointed to places where measurements were made. A sample response is shown in Tab. 1.

After the first basic tests, in-depth verification was carried out. The verification was performed in an office building. The building consists of a ground floor and two floors. The measurements
were carried out in the corridors where the main office WiFi infrastructure is located. There are also additional access points in some rooms (additional phones, additional access points). During the verification there were 21 points defined. They were placed in different zones of the corridor. The plan of the building was drawn too. The Android application was used to gather and verify data about the WiFi signal strength. The data was recorded in the database. The data was collected by three mobile devices: two tablets and one smart phone. There were 28 measurements in each point (including all devices). The measurement data became the basis to calculate the average value of the power of signals for individual rooms which can be the basis for determining the location.

Then a reverse operation was performed. The accuracy of individual measurements was verified (for each measurement separately at a single point) based on a digital map resulting from the average of the data collected from all measurements.

The sum of all the measurements was treated as a source of the digital map of the rooms. A single measurement was thus treated as an attempt to locate the user. The location was determined for individual measurements. Each result consists of a list of rooms and the probability of being located in the given room. At the same time we know the actual location of the user during the measurement. On this basis it can be determined how many measurements were accurate or close to the accuracy of the proper location, i.e. which was the position of the proper room.

The results are as follows. There were 447 measurements (different devices, different measurement points, many measures for each point). The accuracy was as follows:

- taking into account the first position—171 (38%),
- taking into account the second position—268 (60%),
- taking into account the third position—322 (72%),
- taking into account the fourth position—358 (80%).

The accuracy on the first position is not satisfying. Only when the third and fourth positions were taken into account, the result could be considered satisfactory.

After a careful analysis of individual results it turned out that the weak accuracy is partly due to the position of the measured points. Several points are close to each other ‘vertically’—the first floor above the ground floor. The building is ‘L-shaped’ and some points from one part of the building are similar to points from the other part (similar WiFi propagation). It should also be noted that the results were collected by various mobile devices with different WiFi receivers.
7 SUMMARY AND CONCLUSIONS

The metrics analysis provided a number of conclusions. The first is the room (point) where the measurement were made. It can also be deduced what the nearby rooms are and which of them are closer and which of them are further from the current point. It seems that with constant conditions, the metric could be an absolute measure of the distance, and not only the value to generate a list of rooms in the hierarchy. However, obtaining such a ratio requires that each room has been defined by a set of an equally large map. This condition is not possible to meet in real locations. During the test the ability to modify the metrics was analyzed in order to adjust the sensitivity of the position. There are different scenarios to implement. One of them is to prefer the centre of the room, which means that the set of measurement must be close to the average value of the set of maps. Other results should have the metric with low value and should be rejected if the metric is below the threshold. The threshold can be introduced directly in the query. It is also possible to consider the scenario in which the area where the mobile device is placed is important. In this case, the range used to compare measurements can be extended for example to the value of $4 \times \text{stddev}$.

During the tests, it was observed that it is possible to examine the position in relation to the area, but also in relation to the part of the room. This requires only minor modifications of the query.

The test results showed some weaknesses too. The results of the tests indicate that the location is not precise in some situations. On one hand it can be assumed that these are situations within the expected level of the precision error (precision up to 10 m). This way of location is just to ‘coarsely’ estimate in which part of the building the person is localized. On the other hand, this method would be insufficient for a more precise location. The rooms that are presented on the map by a small number of collections are promoted by the algorithm. This influences the requirements for WiFi infrastructure in the physical location. The adequate density of WiFi signals has to be provided. Simultaneously, development analytical tools (appropriate SQL queries) can be introduced, to indicate areas where the additional WiFi beacons are needed.

Summarizing, the system was implemented with the use of the database, the data diagrams and database queries, which practically implement the concept of statistics and fuzzy sets used to determine the position on the basis of the WiFi signal inside the buildings. Although it does not provide precise and reliable location within a few meters, it can be used to estimate the room (zone) in which the individual is placed. The advantage of this method is the simplicity of measurement and computational independence of the number of AP to include (contrary to a method using angle determination). In addition, the method does not require the analysis of the architecture of the building by the person who prepares the configuration for the location—it simply adds the actual data and the location is automatically calculated on that basis.

REFERENCES


