A priori verification and validation study of RFKON database

Sinem BOZKURT KESER¹, Ugur YAYAN², Ahmet YAZICI¹, Serkan GUNAL³

¹Computer Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey
²Inovasyon Muhendislik Ltd. Sti., Eskisehir, Turkey
³Computer Engineering, Anadolu University, Eskisehir, Turkey
Email: sbozkurt@ogu.edu.tr, ugur.yayan@inovasyonmuhendislik.com, ayazici@ogu.edu.tr, serkangunal@anadolu.edu.tr

ABSTRACT
In literature, there are studies that consider only one type of measurements such as Wi-Fi or Bluetooth RSS, but these values are not sufficient alone to overcome the problems in dynamically changed environments. In order to deal with this, we propose a novel fingerprint database that contains both Wi-Fi and Bluetooth RSS values in addition to magnetic field measurements obtained from mobile devices. On the other hand, this study presents a verification and validation of RFKON database to determine suitable machine learning algorithm and compare the performance of these algorithms. The aim of this study is to show performance of the classifiers in the RFKON database. For this purpose, different classifier algorithms which are deterministic algorithms such as k-nearest neighbor, Support Vector Machine, decision tree and probabilistic algorithms such as Naïve Bayes and Bayesian Networks are tested using this database. In addition to these tests, ensemble learning algorithms, namely AdaBoost and Bagging, are used to improve the performance of the selected classifiers. Also, feature selection algorithms are applied to enhance the performance of the selected classifiers. As a conclusion, selected algorithms test results are reevaluated using multi-criteria optimization technique in order to find admissible algorithm in terms of both accuracy and computation time.

KEYWORDS
RF Map — Indoor Positioning Systems — Classification — Feature Selection — Feature Extraction

1. Introduction
Positioning systems are becoming widespread and are used to estimate and track position of a user. These systems are divided into two types (i.e., outdoor and indoor) according to position of a user. Global Positioning System (GPS) that receives signals from satellites and applies triangulation methods to estimate the user location when a user is outside. GPS are unable to provide accurate location due to lack of line-of-sight propagation in indoor. There are number of technologies (infrared (IR), ultrasound, Wireless LAN (WLAN), Ultra-Wideband (UWB), Radio Frequency Identification (RFID), Bluetooth, Wireless Sensor Network (WSN), Inertial Measurement Units (IMUs), etc.) and techniques (triangulation, Pedestrian Dead Reckoning (PDR), fingerprinting, etc.) have been proposed for indoor localization.

WLAN (Wi-Fi) technology is more attractive than other technology to build indoor positioning system because of wide use and decreasing cost. In addition to these, it covers a large area compared with other types of indoor positioning systems; it does not need special hardware and is only based on ubiquitous Wi-Fi access points. In addition to these, positioning accuracy of this mechanism does not depend on time.

Fingerprint positioning technique is received signal strength (RSS) based positioning that associates location dependent characteristics with a location and uses these characteristics to estimate the position [1]. Fingerprinting based indoor positioning systems have two phases named as offline and online phases. Fingerprint database is established in offline phase and positioning is performed in online phase. Fingerprint database construction is started with dividing interested area into grids by floor plan. RSS values that are obtained from access points on a specific time period at the predetermined reference points are stored in a database that is called as fingerprint database. Each fingerprint in this database contains basically location information and RSS values obtained from surrounding access points at that location. In addition to these values, mobile device model, mobile device position, mobile device battery level, and measurements obtained from gyroscope and accelerometer are also stored. This additional information is used from position-
ing algorithms to improve the positioning accuracy of indoor positioning systems.

There are various studies in literature dealing with constructing databases for indoor positioning systems that keep fingerprint map information. Building number, interested region surface area, floor number, space number, training data record number, Wi-Fi access points' number and mobile device type used in measurements are stored in these databases [2-8].

In a study, RSS data were collected at KIOS Research Center which is a 560m² typical office environment that consists of offices, labs, a conference room and corridors. 9 APs and 5 different mobile devices (HP iPAQ hw6915 PDA with Windows Mobile, an Asus eeePC T101MT laptop running Windows 9, an HTC Flyer Android tablet and two other Android smartphones (HTC Desire, Samsung Nexus S)) were used for data collection process. Training data were constructed by collecting RSS measurements from all 9 APs, at 105 distinct reference locations by carrying all 5 devices at the same time. There were 20 fingerprints were collected per reference points, so total number of fingerprints in training data was 2100. Besides, test data were collected 2 weeks later by walking along a predefined route 10 times at the same time with all devices. There were 96 locations on this route most of which different from the reference points. One fingerprint was collected per each test location, so total number of fingerprints in test data was 960 [2, 3]. This database was used to solve device calibration problems in indoor positioning.

UJIIndoorLoc database [4] is the biggest and publicly available database in the literature. Data were collected from a surface of 108703m² containing 3 buildings with 4 or 5 floors depending on the building. There were 933 reference points and 520 different wireless access points (WAPs) including in the database. More than 20 users using 25 different mobile devices were collected fingerprints. There were 19938 fingerprints for training and 1111 fingerprints for testing were recorded. Test data were collected 4 months later after training data. This database could be used to make comparisons among different methods in indoor positioning. In [5], data collecting surface was a 2 building with 3 floors. There were 9358 sample points taken from 101 different WAPs in the database. Total number of reference points was 392. In this study, a new fingerprint-based method was proposed. This method uses previously stored map and similarity functions and majority rules were used to estimate the location.

In [6], the database has 2880 sample points taken from 1 building of 1 floor with 206 different WAPs. 2 different mobile devices were used to obtain measurements. The number of different reference points appearing in the database is 96. In this study, instead of RSS values obtained from surrounding access points, the rankings of the RSS values were used. This database was constructed in order to solve accuracy degradation caused by hardware and software differences between RSS measurements collected by user devices and RSS in the fingerprint map.

In [7], the measurements in the database were taken from 1 building with 9 different WAPs. This study deal with differing antenna attenuation among different devices and environments where not every beacon is visible everywhere in WLAN based fingerprint. In [8], the measurements were taken from Computer Science department building of Arab Academy for Science and Technology and Maritime Transport (AASTMT) at Alexandria in Egypt. There were 16 access points (4 at each floor) in this building. In addition to this, there were 39 reference points which were separated by 2 to 3 meters from each other. 25 fingerprints were collected at each point for training and testing data. So, total number of fingerprints for training and testing were 975 and 500, respectively. Since there are 16 access points, each fingerprint contains 16 RSS values. This study was done for comparing different machine learning algorithms in order to find the more accurate and more suitable one among selected algorithms.

As discusses so far, there are various studies for constructing fingerprint map in the literature. The common point of these studies is that they only consider one type of measurement. Due to the fact that obstacles in indoor environment cause reflections and multipath inference, radio signal spreading is difficult to predict. In addition to this, RSS values are also be affected by user’s presence close to the mobile devices’ antenna [5]. By combining different types of sensor measurements, these problems would be solved. For this purpose, we propose a novel multi-sensor and multi-topological fingerprint database for indoor positioning that classified as anchor based and mobile based databases. Anchor based database contains 4 type of measurements named as Wi-Fi, Bluetooth (BT), Bluetooth Low Energy (BLE) RSS values, and link quality indicator (LQI) measurements for supporting Wi-Fi based indoor positioning. Mobile based database contains Wi-Fi, BT RSS values, in addition to these magnetometer and accelerometer measurements are also included in the database. To increase precision and accuracy, fingerprints are collected in each grid of the room.

The main contribution of this work is to construct a database for fingerprint map to enhance accuracy, precision, and robustness of the location estimation system to dynamically changed environment. This database is publicly accessible for researchers to study different fields in indoor localization problem [9]. Besides, this database is validated using selected classifiers. In experiments, accuracy and computation time of each classifier are compared. Feature selection algorithms are applied to improve the performance of selected classifiers.

2. System Overview

RFKON is a system that is constructed to solve indoor positioning problem. It contains three units named as Gezkon, Haskon, and Konsens. Gezkon is a mobile application and it is responsible for collecting RSS values (WiFi, BT) and self-localization for mobile devices. Haskon is software and it is used to collect RSS values (WiFi, BT) and passive sensor values (encoder, IMU) for sensor fusion. It is aimed to use for positioning of mobile robots. Konsens is a server that collects RSS values from the access points (APs) in the region to the construction of RFKON.
RFKON database is used by Gezkon and Haskon for self-positioning applying machine learning algorithms. Konsens is also used to estimate position of access points in the region and also responsible for updating and calibrating of RFKON database. Sensor nodes emits RF signals (WiFi, BT) for positioning and gathers RF signals from the region. Sensor nodes are also responsible for communication. This communication is established by DDS layer. Konsens maintains sensor nodes. This indoor location scenario is given in Figure 1.

Konsens consists of two parts names as Konsens Server and Konsens Sensor Network (SensDug).

**Konsens Server**
The hardware of the Konsens Server is shown in Figure 2. Besides this hardware, Mikrotik RB951Ui is used for network communication for roaming. RSS data first comes to Mikrotik, then send to Konsens Server.

Konsens runs on Ubuntu 14.04 LTS Server operating system. It contains four nodes named as Konsens Control, SensDug DDS, Gezkon Web, and Haskon DDS. All these nodes use DDS layer as a middle layer for establishing data transferring process between Konsens and other units (Haskon, and Gezkon). These nodes are represented in Figure 3.

Konsens control node is deal with database construction phase. It collects RSS values via SensDug DDS, Haskon DDS node and Gezkon Web node and writes the collected values into MongoDB database. Because of environmental dynamics (wall, furniture, people) and device diversity, the RF database must be calibrated/updated for accurate positioning. For this purpose, this node is enhanced in future for constructing robust RF database.

SensDug DDS node reads and write RSS values (WiFi, BT) by WiFi and BT Publisher and Subscriber topics. It collects RSS values from the APs in the region and sends these values to Konsens control node using TCP/IP protocol. Data are written in JSON data format with MAC addresses and corresponding RSS values. The collected data are used for constructing anchor based database named as RFKON_AB. RFKON_AB database is separated as RFKON_AB_WiFi and RFKON_AB_BT according to sensor type. This database is used for sensor node location in Konsens control node side.

Haskon DDS node receives and sends RSS values to Konsens control node same as SensDug DDS node. It also collects magnetic field (MF) sensor information to construct RFKON_MB MF database in Konsens control node side. Gezkon Web node collects RSS values from the APs in the region and sends these data to Konsens control node for constructing RF database named as RFKON_MB_WiFi and RFKON_MB_BT according to sensor type. Data sending process is achieved using TCP/IP protocol.

**Konsens Sensor Network**
The hardware of Konsens sensor network contains Mikrotik RB951Ui (in network communication for roaming), and Raspberry Pi B+ that is illustrated in Figure 4. They are chosen
because of low price, ease of use, and ease of supply. It also involves USB WiFi and BT dongles for collecting WiFi and BT RSS values.

Figure 4. Hardware architecture of Konsens Sensor Network.

Data Distribution Server (DDS) is an open source software that is used for data transfer. DDS is a Machine-to-Machine Middleware layer that is constructed from Object Management Group (OMG). Middleware is the software layer that lies between the operating system and applications. It enables the various components of a system to more easily communicate and share data. DDS has a scalable and high performance architecture. It is real time, reliable and secured. DDS establishes a data transfer between Publishers and Subscribers.

Quality of Service (QoS) is aimed to supply system needs of working applications in the network and reduce time loss. A topic may have contained Writer/Reader that has various QoS by the help of DDS. Each Writer/Reader application uses the QoS of current Topic as well as has its own QoS. DDS is developed from some firms: Opensplice DDS, RTI DDS, and OpenDDS. RFKON is constructed from Opensplice DDS Community Editor. In our DDS architecture that is formed for RFKON, every Sensor Node has an DDS application for sending WiFi and BT data to Konsens. For this purpose, DDS Single Process architecture is applied. In this type of architecture, DDS application is maintained independent of operating system. In other words, DDS maintain and important services are attached to the application. Each DDS application contains all DDS infrastructure itself. Four topics are established for RFKON systems: WiFi_Publisher, BT_Publisher, WiFi_Subscriber, and BT_Subscriber for sending and receiving RSS values.

Durability service of Opensplice DDS is responsible for sending data to all compatible Subscriber in the domain in time. In addition to this, Historical Data for all nodes that are involved in system are hold after publishing data. Delay of data read/write process of RFKON is reduced by configuring QoS of DDS.

3. RFKON Database

RFKON database is publicly available database [9]. It was planned to be created for researchers used in the subject of indoor localization, classification comparison, self-localization, sensor nodes position calculation, improve the accuracy of WLAN based positioning, AP position calculation using geometric methods and remote monitoring. RFKON database is different from the ones defined in literature because it is constructed considering more sensor values instead of one type of sensor values. It is updated itself iteratively when a new device is identified in the study region [10].

3.1 Experimental Environment

Figure 5. Experimental environment of floor 1.

Figure 6. Experimental environment of floor 2.

Data were collected to construct our database from Eskisehir Osmangazi University Teknopark. It has 2 floors about 800m². This area was broken into 486 grid squares (each of size 1.2m × 1.2m) and the center of each grid square was noted.

As seen in Figure 5 and 6, red stars represent the reference points that are used for collecting sensor values from the access points and red squares represent sensor nodes in the test area. There are 7 sensor nodes: 5 of them are positioned in floor 1 and the remaining ones are in floor 2.

3.2 Sensor Types

RFKON database is constructed by collecting three different measurements such as WiFi, BT, and MF. Bluetooth RSS values are used for region based positioning. To estimate more sensitive positions WiFi and MF values are used.

RFKON database contains 2 different databases classified as anchor based that contains measurements (Wi-Fi, BT, BLE, and
LQI) from seven sensor nodes and mobile based that contains measurements (Wi-Fi, BT, and MF) collecting with five different mobile devices (LG g3, SONY Xperia z2, Samsung S4 mini, Samsung Note 10.1 Tablet and Nexus Tablet).

RFKON anchor-based BT database (RFKON_AB_BT) is a special database and it is used for researchers to make a study for sensor node localization in indoor environment. RFKON mobile-based Bluetooth database (RFKON_MB_BT) is used for estimating of a person or an object in indoor environment using machine learning algorithms and geometric based methods.

RFKON anchor-based WiFi database (RFKON_AB_WiFi) is used for WLAN based positioning to estimate location of sensor nodes and fixed access points. Besides machine learning algorithms, geometric approaches are also usable with this database. This database is also useful if any modem is identified or removed from the environment. RFKON mobile-based WiFi database (RFKON_MB_WiFi) is used for WLAN based indoor positioning. Probabilistic, deterministic methods and geometric-based approaches could be used with this database to estimate the position of user or any object in indoor environment.

RFKON mobile-based MF database (RFKON_MB_MF) could be used for machine learning algorithms and data fusion methods.

4. Database Verification and Validation

In this section, 7 different classifiers (Decision Tree (DT), Naïve Bayes (NB), Bayes Net (BN), Sequential Minimal Optimization (SMO), Nearest Neighborhood (NN), Adaboost (AB), Bagging (BAG) and 4 different feature selection methods (Correlation-based Feature Selection (CFS), Chi Square Selection (CHI), Filtered Attribute Selection (FILT), Gain Ratio Selection (GAIN)), and Principal Component Analysis (PCA) are used for indoor positioning. The selection methods are compared according to accuracy, computational time and selected attributes.

Figure 7 and 8 show the accuracy and computation time results of selected classifiers using all attributes in the RFKON_MB_WiFi and RFKON_MB_MF databases for floor 1 and floor 2.

As seen in the Figure 7, AB-DT classifier gives the best results for all databases in terms of accuracy. BN classifier gives the best computation time results for WiFi database and NB for magnetic field databases in Figure 8.

There are 26 Access Points in the RFKON_MB_WiFi database for floor 1. The number of access points is reduced to 23 when CFS is applied as a feature selection algorithm and 24 when other feature selection algorithms are used. There are 32 Access Points in the RFKON_MB_WiFi database for floor 2. This number is reduced to 24 when CFS is applied, 31 when CHI, FILT, and GAIN are used, and 28 when PCA is applied.
Figure 10. Computation Time Results of Correlation-based Feature Selection.

Figure 11. Accuracy Results of Chi-square Feature Selection.

Figure 12. Computation Time Results of Chi-square Feature Selection.

Figure 13. Accuracy Results of Filtered Att. Feature Selection.

Figure 14. Computation Time Results of Filtered Att. Feature Selection.

Figure 15. Accuracy Results of Gain Ratio Feature Selection.
5. Conclusion and Result Summary

In this section, all test results are reevaluated for selecting the most appropriate algorithm in terms of both accuracy and computation time. In indoor positioning, computation time is more important than accuracy because mobile device user can find his location instead of a margin error in accuracy. But, computation time is significant criteria because of battery of mobile device. When positioning phase takes long time, the battery of mobile device is attenuated. For this purpose, we introduce Equation (1) to incorporate both computation time and accuracy with the coefficients 0.7 and 0.3, respectively.

\[
\text{Result} = 0.7 \times \left( \frac{\text{Time}_{\text{selected}}}{\text{MaxTime}_{\text{selected}}} \right) + 0.3 \times \left( 1 - \frac{\text{Accuracy}_{\text{selected}}}{\text{MaxAccuracy}_{\text{selected}}} \right) \tag{1}
\]

The coefficients in Table 1 are selected empirically. According to this equation, minimum result value obtained from Equation (1) gives the most appropriate algorithm. These results are given in Table 2.

Table 1. Coefficients of Time and Accuracy

<table>
<thead>
<tr>
<th>Set</th>
<th>Time</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Set 2</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Set 3</td>
<td>0.3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Table 2. Most appropriate algorithms according to Equation (1)

<table>
<thead>
<tr>
<th>Set</th>
<th>WiFi Floor 1</th>
<th>WiFi Floor 2</th>
<th>RAW DATA Floor 1</th>
<th>RAW DATA Floor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>BN</td>
<td>BN</td>
<td>DT</td>
<td>DT</td>
</tr>
<tr>
<td>Set 2</td>
<td>BN</td>
<td>AB-DT</td>
<td>DT</td>
<td>DT</td>
</tr>
<tr>
<td>Set 3</td>
<td>BN</td>
<td>AB-DT</td>
<td>DT</td>
<td>DT</td>
</tr>
</tbody>
</table>

6. Future Works

In future works, we would intend to develop calibration and update methods/algorithms using this database. Also, sensor fu-
sion and classifier fusion methods will be implemented by using RFKON database. In addition to these, we are constructing new databases for developing RFKON database, check for updates [9].

Acknowledgments

This work is supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under grant number 1130024.

References


[9] RFKON database-
http://inovasyonmuhendislik.com/rfkondatabase/