Wi-Fi Based Indoor Positioning Systems

Burak Emre Ozer
Hacettepe University
Ankara, Turkey
burak.ozer@hacettepe.edu.tr

Furkan Caglar Gulmez
Hacettepe University
Ankara, Turkey
furkancaglargulmez@gmail.com

Huzeyfe Kocabas
Hacettepe University
Ankara, Turkey
kocabashuzeyfe@gmail.com

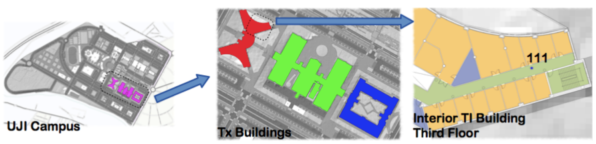


Figure 1: Source [12] Left: map of UJI campus and Tx buildings. Middle: red indicates ESTCE - Tx building. Right: example of a reference point.

Abstract

Wi-Fi Fingerprint-based positioning approach that detects the position of user or device is widely used in the indoor positioning systems instead of Global Positioning System (GPS). In this approach, Received Signal Strength (RSS) values that are known as Wi-Fi fingerprints used. Received Signal Strength values are the measurement of the power present in a received radio signals. We use UJIIndoorLoc dataset with 19937 training records and 1111 test records. This dataset of RSS values are collected by using previously placed wireless access points (WPAs) in Tx Buildings of University of Jaume I campus. We aim to predict location points with respect to floor IDs, building IDs, longitude and latitude values with supervised machine learning algorithms such as K-Nearest Neighbor Algorithm, Random Forest Algorithm, Support Vector Machine and Decision Tree Algorithm. Then we use the model with the highest accuracy in the rest of the progress. Classification techniques are used for building and floor classification and regression techniques are used for detection of location points.

Keywords— indoor positioning, received signal strength indication (RSSI), machine learning algorithms, classification, random forest

1. Introduction

Global Positioning System (GPS), which uses satellites,is the most popular outdoor positioning system, however its signals can be easily blocked by various structures and factors then it becomes useless for indoor environment because of signal loss. Unlike the GPS, Indoor Positioning Systems aims to detect the position of user or device by using Access Points signal also called Wi-Fi ngerprint. With the advancing technology and spread of wireless networks, Indoor Positioning Systems become even more important place in the fields of augmented reality, social networking, personal tracking, guiding blind people, tracking small children or elderly individuals and location-based advertising etc.

Wi-Fi-based fingerprint methods have some problems when positioning phase in indoor. These problems can be caused by the fact that the devices in which the radio signals are collected during the training stage and the devices in the test phase are different.

Another reason is that the number of access points in the environment varies greatly. Inevitably, these problems negatively affect positioning success. However, we will try to determine the position with regression algorithms using the real latitude and longitude values of the collected locations.

Then we will turn our problem into classification problem by using the building and floor features in the data set. In the test section, we will try to estimate which building is located or which floor of the building.

Different machine learning algorithms will be tried and we will decide the most suitable algorithm for indoor positioning.

We will use the UJIIndoorLoc database throughout the entire project. Classification and regression problems will be solved using the RSS values from 520 wireless access points (WAPs). In the classification part, since the data set contains 3 buildings, we will divide the data into three and try to estimate which floor is located.

The rest of the paper is organized as follows: Presentation of related studies, explanation of the data set used in the experiment, explanation of solution approaches, experimental results and conclusion.

2. Related Work

There are many designing alternative based on lots of technologies such as infrared (IR), ultrasound, radio-frequency identification (RFID) by using electromagnetic fields, wireless local area network (WLAN) by using wireless network to communicate between two device, Bluetooth by using radio wave, sensor networks, ultra-wideband (UWB), magnetic signals, vision analysis and audible sounds for Indoor Positioning Systems.

Some articles [13], [4], [5] contains an overview of different technology options for Indoor Positioning Systems design. Each system has advantages of using a particular positioning technology(or combining many of these technologies) and also has disadvantages of limitations of these technologies.[14]

This paper is related to the Indoor Positioning System based on RSSI (Received Signal Strength Indicator) fingerprint positioning using wireless local area network (WLAN) signals and effective machine learning methods for this system.

There are many different indoor positioning methods for wireless network such as time-based positioning[2], angle-based positioning[1], received signal strength indicator(RSSI) based modeling positioning[9] and RSSI-based fingerprint positioning[[6],[8],[7],[11]].

The references presented in the [[6],[8],[7]] are the researches that use k-nearest neighbor algorithm and Euclidean Distance to find nearest points. In the reference [11] , RSSI-similarity degree which was used as weight factors to estimate the coordinates of measured points is introduced as a different point of view.

3. Dataset

We decided to use UJIIndoorLoc dataset which is downloaded from University of California Irvine Machine Learning Repository, because it contains Multi-Building Multi-Floor indoor localization which is the most appropriate repository for our solution.

The subject of the dataset is WLAN fingerprint that is almost 110.000 m2 which includes three buildings from Universitat Jaume with 4 or more floors. This database consists of 20 different users with different heights and 25 different brand android mobile phone which has different android versions.

There are 19937 training and 1111 test instances. Each instance has attribute of RSSI (Received Signal Strength Intensity) that are taken from different WAPs (Wireless Access Points).

We are expected to predict each test coordinate values of instances which are latitude, longitude, floor and building ID. Floor and building ID features are used for to compare the classification algorithms. And remaining features which are latitude and longitude are used for to detect the position by regression algorithms.

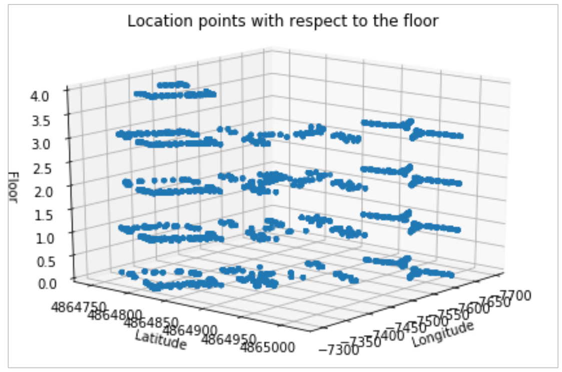


Figure 2: Floors inside the building

4. The Approach

This section will provide technical details about the project. We will roughly explain k-Nearest Neighbour, Support Vector Machines, Random Forests and Decision Tree algorithms then share the result we use. We will analyze the results.

4.1. k-Nearest Neighbour

The k-Nearest Neighbour algorithm aims to search the nearest k-neighbor from the training set by measuring the received signal strength indication (RSSI) node based on Euclidean distance.

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

In the following example, we see the latitude and longitude points obtained from a sample data set. The algorithm performs operations according to the nearest 3 neighbors (k=3).

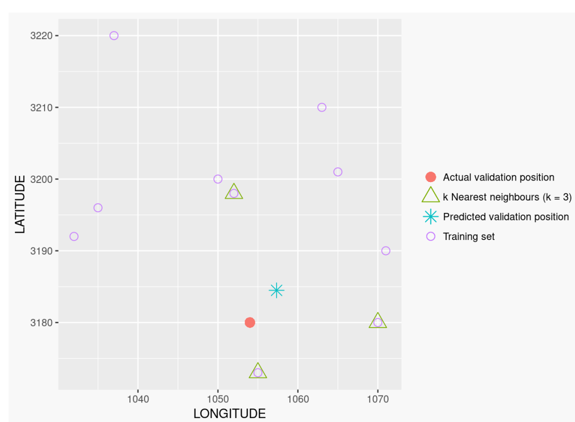
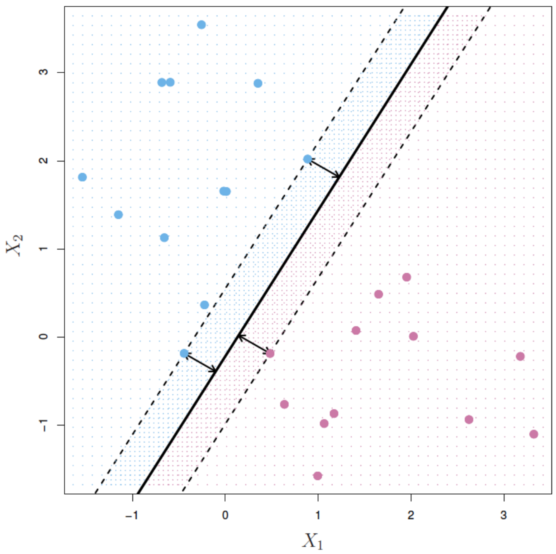


Figure 3: Source [10] This plot shows the predicted position.

The location corresponds to the average of the Euclidean training fingerprints on the latitude and longitude based on RSSI.

4.2. Support Vector Machines (SVMs)

Support vector machines try to pass a linearly separable hyperplane with the maximum margin through the data set to classify the data.



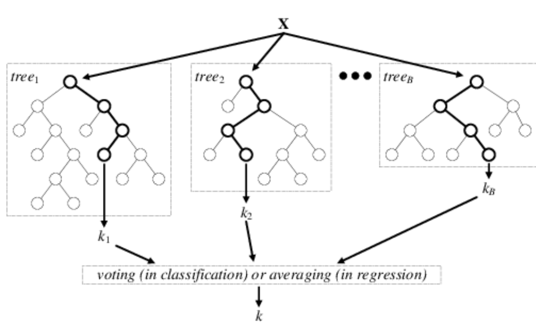
With some parameters we can set how much miss-classification will be. As with all supervised machine learning algorithms, there is a bias-variance trade-off. If the tuning parameter (denoted by C) is large the number of miss-classification will increase, and vice versa.

Also, if the data cannot be separated linearly, we can solve this problem by using the kernel functions.

4.3. Random Forests

Random forest is simple and successful algorithm that works without hyper-parameter also an ensemble method.

The algorithm is roughly a sample is selected from the training set and then a decision tree is created. Then random estimates are made from each node so that the most appropriate split is decided. Then the tree is saved as it is. Finally, the output is given based on average response from the all individually trained trees.



So random forest is simply a collection of decision trees. In addition, the more the number of trees, the more the over-fitting may occur.

4.4. Decision Trees

Decision Tree learning is one of the most widely used and practical methods for inductive inference over supervised data.There are various decision tree algorithms. We will use CART (Classification and Regression Trees) algorithm[15] which is implemented in sci-kit-learn library.

The most important thing for a decision tree is to choose root. Various thresholds were used when selecting root.But since the decision tree was over-fitting, this situation was called pruning.

4.4.1 CART (Classification and Regression Trees)

It is similar to other algorithms but also supports numerical variables. First, it creates the decision tree for classification. Then splits the instances according to the attribute with the highest information gain. The split continues until the classes of the leaf nodes are the same.

5. Experimental Results

In this section, we will share some graphs and analysis of the performance and success rates of the approaches we investigate.

The experiments were performed using the UJIIndoorLoc data set together with the sci-kit library.

Our goal is to determine position using regression and to compare classification algorithms and finally to create a model with high success rate. For this purpose, we used Support Vector Machines, Random Forests, Decision Trees and k-nearest Neighbour algorithms in the experiments.

5.1. Regression

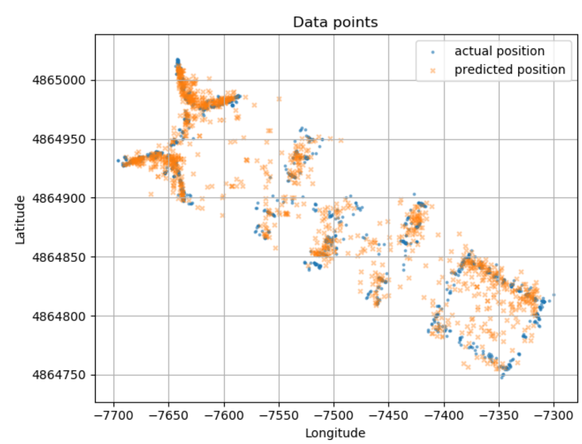


Figure 4: Location points representation

In this section, the results of the models related to positioning models will be shared and analyzed by using regression algorithms.

We have trained and tested our models for four different algorithms using the collected signals from the 520 wireless access points in the data set.

The positioning problems mentioned above are clearly seen in the table 1. As the number of access points varied and the devices used differed, the error values (meters) were much higher than expected.

Table 1: Regression accuracies for kNN, SVM, RF, Boosting Decision Tree

Model	min.(m)	mean(m)	max.(m)
kNN	0.12	15.59	136.42
SVM	0.91	39.35	113.26
Boosting Decision Tree	0.84	30.19	132.69
Random Forest	0.09	14.99	80.29

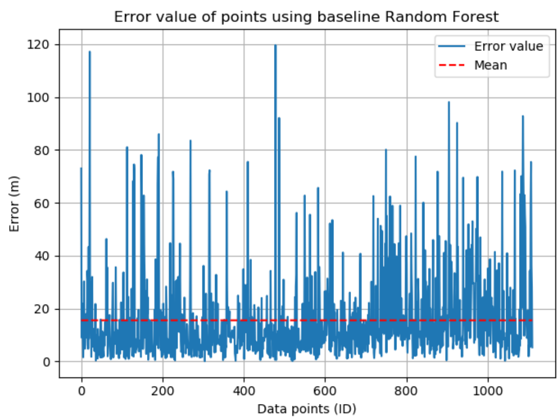


Figure 5: Random forest error values

For example, if we draw an error graph showing the error value of each point only for random forest, we can see it more clearly in the figure 5.

In order to avoid this problem, we repeated the regression process using only the data of the first building. First, we remove the data with latitude and longitude values of zero. Besides, only to select the TI building set and filtered the buildingID to 0.

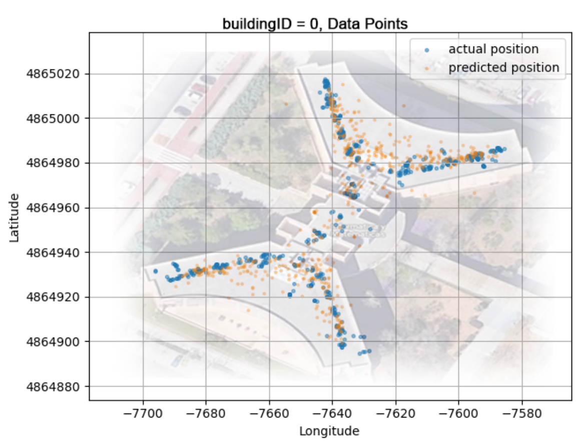


Figure 6: buildingID=0 points

Then we train random forest regressor model with parameters max_depth=20, n_estimators=100, random_state=42 and got the average error value: 8.30 meters as seen from the top location of each points on the map. The error value we found in the past experiments was almost halved.

5.2. Classification

In this section, the results of the models related to building and floor classification models will be shared and analyzed by using regression algorithms.

We have trained and tested our models for three different algorithms using the collected signals from the attribute 523 and attribute 523 wireless access points (WAPs) in the data set.

Attribute 523 (Floor): Altitude in floors inside the building. The values from 0 to 4.

Attribute 524 (buildingID): The ID used to specify the building. Measurements were taken for three different buildings. Identity values are 0 to 2.

First of all, we have classified the building number by using the whole data set.

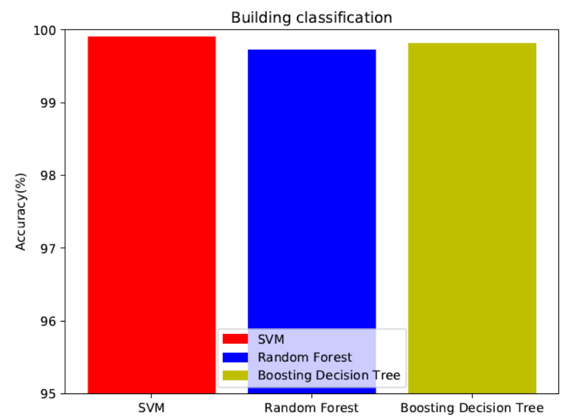


Figure 7: building classification using the whole dataset

The result of this experiment is shown in Fig. 7. It is clear that Support Vector Machines are more successful than the other classifiers.

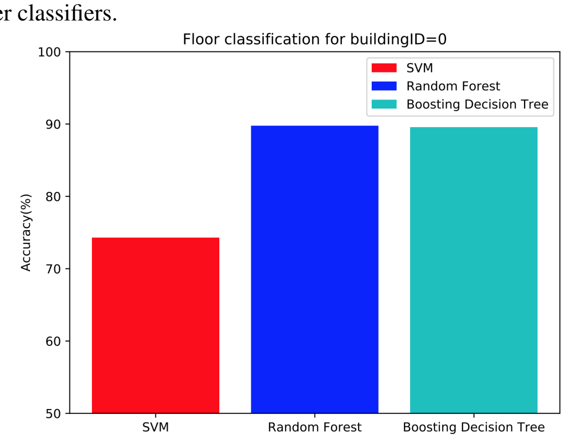


Figure 8: floor classification using the building0

We performed the floor estimation using the data of the building0 and classification algorithms. Although Random Forest and Boosting Decision Tree are close to each other, Random Forest looks better with a fine difference.

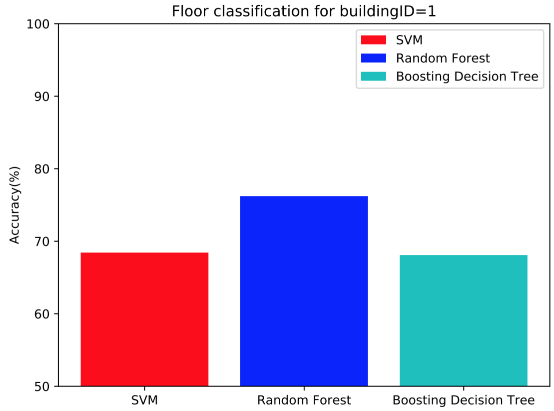


Figure 9: floor classification using the building1

In the second step, we made the classification for the building1. RF clearly looks better than other algorithms. Because the number of data collected in the middle building was high, this situation benefited from the random forest concept and we achieved higher results in this model.

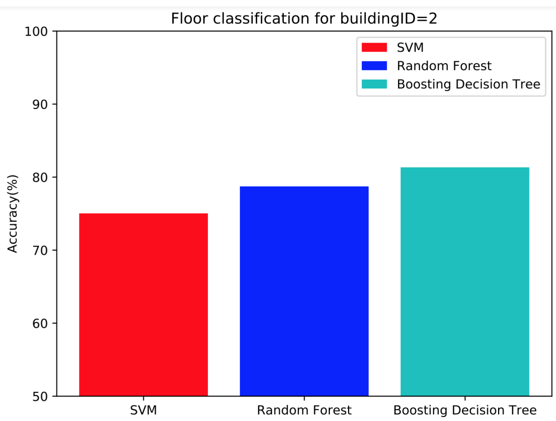


Figure 10: floor classification using the building2

In the last step, we made the classification for the building2. Boosting Decision Tree clearly looks better than other algorithms.

As seen above, the distribution of the data set according to the buildings has played a key role on the success of the models.

5.3. Tables

We can analogize the results of our predictive models by using three different methods which k-Nearest Neighbor, Random Forest, Decision Tree.

We can see that the Random Forest model outshout k-NN model and Decision Tree model. [3]

6. Conclusions

Since least amount of high signal data collected in the middle building, the data set covers huge area, the number of wireless access points received at most test points has varied greatly

Table 2: Regression accuracies for kNN, RF, Decision Tree

MODEL	ACCURACY	BETTER?
KNN	77.3	×
RF	86.0	√
DECISIONTREE	81.5	×

caused low accuracy in the position estimation process. In addition, the multi-floor and multi-building data set had a negative impact on the success of the regression models. Therefore, we re-educated our model with data obtained from the floor where the user is located and we have achieved a significant increase in the success rate on position prediction.

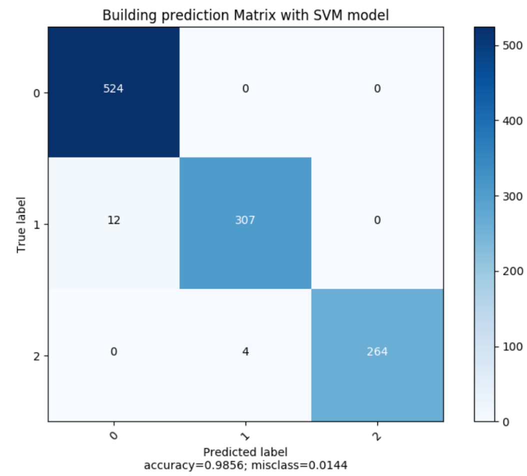


Figure 11: building prediction confusion matrix

For the classification part, we made a floor classification according to the building and building classification. The results were satisfactory. We shared the classification results we made with Support Vector Machine model in the above confusion matrix11.

As a conclusion, random forest algorithm for the regression is the best algorithm to predict the latitude and longitude which are coordinates. But for the classification of the floor and buildingID is varying. So that the best algorithm changes case to case. We selected random forest according to accuracy and efficiency of time of the computation and also it outshouts kNN and Decision Tree model.

6.1. Feature Works

In future works, adding threshold for the signals of the WAPs (wireless access points), and deleting undetected wireless access points from the data set will have a significant impact on the success of the position prediction model.

References

[1] F. Belloni, V. Ranki, A. Kainulainen, and A. Richter. Angle-based indoor positioning system for open indoor environment. *2009 6th Workshop on Positioning, Navigation and Communication*, pages 261–265, 2009.

[2] X. Cheng, A. Thaeler, G. Xue, and D. Chen. A time-based positioning schemes for outdoor wireless sensor network. *IEEE INFOCOM 4(4)*, pages 2685–2696, 2004.

[3] E. J. Gilbert. Wi-fi locationing.

[4] J. Hightower and G. Borriello. Location sensing techniques. *Technical Report UW CSE*, 2001.

[5] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *Technical Report UW CSE*, pages 57–66, 2001.

[6] X. Liang, X. Gou, and Y. Liu. Fingerprint-based location positioning using improved knn. *2012 3rd IEEE International Conference on Network Infrastructure and Digital Content*, pages 57–61, 2012.

[7] L. Ni, Y. Liu, Y. Lau, and A. Patil. Indoor location sensing using active rfid. wireless networks 10(6). *LANDMARC*, pages 701–710, 2004.

[8] T. Peerapong and C. Xiu. Indoor positioning based on wi-fi fingerprint technique using fuzzy k-nearest neighbor. *Proceedings of 2014 11th International Bhurban Conference on Applied Sciences and Technology*, pages 461–465, 2014.

[9] W. Ren, L. Xu, Z. Deng, and C. Wang. Positioning algorithm using maximum likelihood estimation of rssi difference in wireless sensor networks. *Journal of Data Acquisition —& Processing 21(7)*, pages 1247–1250, 2008.

[10] E. S. Sansano, R. M. Cols, scar Belmonte Fernndez, and J. Torres-Sospedra. Indoorlocplatform.

[11] F. Tian, Y. Dong, E. Sun, and C. Wang. Nodes localization algorithm for linear wireless sensor networks in underground coal mine based on rssi-similarity degree. *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, pages 1–4, 2011.

[12] Torres-Sospedra, J. Montoliu, R. Martnez-Us, A. Avariento, J. P. Ar-nau, T. J. Benedito-Bordonau, M. Huerta, and Joaqun. Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint- based indoor localization problems.

[13] M. Vossiek, L. Wiebking, P. Gulden, J. Wiehardt, C. Hoffmann, and P. Heide. Wireless local positioning. *IEEE Microwave Mag*, 4:77–86, 2003.

[14] A. L. Yanying Gu. A survey of indoor positioning systems for wireless personal networks.

[15] F. Zhu, M. Tang, L. Xie, and H. Zhu*. A classification algorithm of cart decision tree. 14:17–25, 2018.