Mean-offset Classifier based on Wi-Fi Indoor Positioning System

Pasungili Rajesh Ramakrishnan and Herman Myburgh

[1] University of Pretoria, Gauteng, RSA
[2] University of Pretoria, Gauteng, RSA

Abstract. A mean-offset classification technique was identified. It was found that the mean-offset classifier provides stability under dynamic indoor conditions and provides consistent results when training and test data combinations are swept from 10 – 95%. In this paper the mean-offset classifier is compared to the K-Nearest Neighbors (KNN) and Naïve Bayesian (NB) classifiers, with a view of developing an adaptable and computationally efficient indoor localization model using machine learning principles. It was seen that the mean-offset classifier improved results considerably and achieved an accuracy of 0.85 m and 1.15 m under line-of-sight (LOS) and non-line-of-sight (NLOS) conditions in residential areas.

Keywords: KNN, LOS, Mean-offset, Machine learning, Naive Bayesian NLOS.

1 Introduction

1.1 Overview

There has been a rapid growth in localization techniques and the application thereof in indoor environments [1]. The automation of locating people and objects in indoor environments, such as shopping malls, hospitals, warehouses and indoor sports centers, provides industries with valuable statistics that can be used to enhance their businesses. In this context, a plethora of indoor localization schemes have been proposed based on the type of signals used such as optical waves, Wireless Local Area Network (WLAN) radio signals and sound waves [2]. WLAN based location has become popular as wireless technologies are readily available, cost effective, scalable and most importantly received signal strength indicator (RSSI) can be extracted from most Wi-Fi receivers easily. There are two main groups in WLAN localization propagation model-based techniques and fingerprinting models. The former characterizes the indoor channel by building a site-general or site-specific path loss model based on received signal strength (RSS) and frequency fading statistics [3]. The latter obtains RSSI values and stores it in a database, after which a similarity metric is used as a differentiating factor to predict the user's location. RSSI has two major drawbacks, however. Multipath complicates and degenerates RSSI values as multiple line-of-sight (LOS) and non-line-of-sight (NLOS) signals with different phases, amplitudes and delays distort the shape of the signal, which leads to spatial ambiguity [3]. The second shortcoming is RSSI instability i.e. RSSI differs on different devices when recorded at the same place and time. To overcome the drawbacks faced by RSSI values, machine learning algorithms such as classification techniques and artificial neural networks (ANN) are implemented.

1.2 Related work

There has been a growth in implementing machine learning algorithms to RSSI based indoor localization. Table 1 summarizes the comparisons of localization using machine learning techniques. In the Kernel-based learning method a spatial filtering step is introduced to locate the estimated point to a subset of the environment and a kernelized distance for estimating the Euclidean distance between the observed RSS and the stored fingerprints is proposed, achieving an accuracy of 2.43m [2]. Support vector machines (SVM) classifiers that implement the linear and gaussian kernel achieve an accuracy of 2 m and 3.12 m respectively [4]. In [P], a hybrid approach combining PCA with a grid search-based Kernel SVM is proposed. The PCA algorithms decorrelates and denoises the data received in the offline phase before applying the grid search algorithm during the online phase to achieve an accuracy of 1.37 m [5]. The widely used KNN classifier achieves an accuracy of 3.08 m [5]. Decision trees are non-parametric supervised learning methods, which achieves an accuracy of 2.87 m [4]. A random forest classifier selects the tree with the highest votes after multiple decision trees are generated and achieves an accuracy of 3.1 m [4].

Table 1. Comparisons of indoor localization models using classification techniques

Proposed Method	Accuracy (m)
Kernel based [2]	2.43
KNN [5]	3.08
PCA-SVM [5]	1.37
SVM: Linear Kernel [4]	2
SVM: Gaussian [4]	3.12
Decision tree [4]	2.87
Random forest [4]	3.1

The main contributions of this paper are summarized as follows:

- A mean-offset classifier which computes the percentage of error between the users RSSI against the database of trained RSSI centroids is implemented. The classifier is tested under dynamic indoor conditions where human movement is present and across multiple days to see if temporal fluctuations affect localization.
- K-NN and Naïve Bayesian classifiers are implemented and compared to the mean-offset classifier under static indoor localization scenarios. Training data are swept from 10 – 90% to compare which classifier performs accurately when minimum training data are available.

2 Methodology

Machine learning algorithms improve the accuracy of localization systems. All machine learning algorithms are implemented in two phases, training and test phases [4]. In the first phase, a collection of RSSI values are stored in the database. The data are then pre-processed by scaling the features and splitting the data into training and test sets. The classifier then uses the database of RSSI values to learn and build a model by which location can be predicted. In the test phase, the classifier that has the most accurate model is used to predict location of the new set of RSSI.

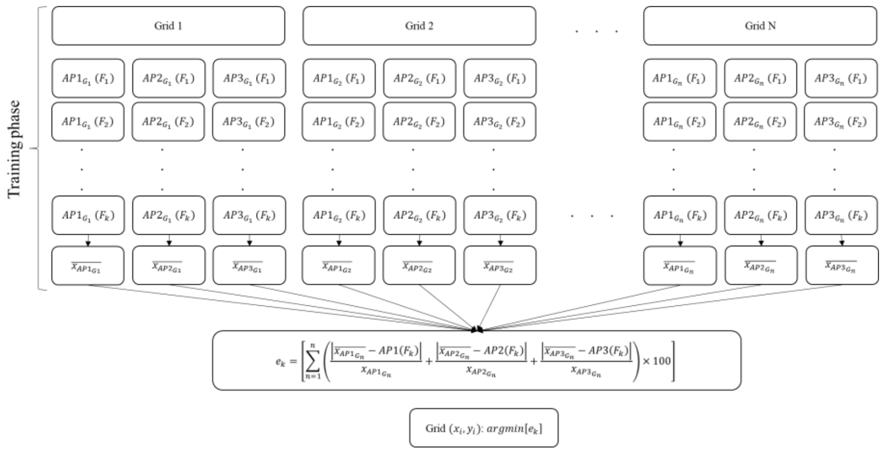

Fig. 1. Proposed mean-offset algorithm

2.1 K-Nearest Neighbours (KNN)

K-nearest neighbour is a classification technique that calculates the distance between features. The stored RSSI features correlate to the distance between the AP and mobile device. The algorithm calculates the P −norm of N- dimensions RSSI vector x_i, where x_i is the value on the RSSI database given as D^N [4]:

$$x_i \in D^N \tag{1}$$

Select the number K of neighbour and take the nearest K neighbours of the new data point according to the Euclidean distance. The distance between measured $\bar{x}$ and the RSSI value from the database $\bar{x}_i$ is represented as [4]:

$$d(\bar{x} - \bar{x}_i) = \left(\sum_{j=1}^{|\bar{x}|} \left|\overline{x_j} - \overline{x_{ij}}\right|^p\right)^{p^{-1}} \tag{2}$$

Where d is the measured distance, $\overline{x_{ij}}$ is the average and j represents the selected APs. The value for $p = 2$ as the Euclidean norm-distance is used. The K-NN classifier chooses the minimum distance of the K-neighbour points where W_k is the list of points corresponding to the K fingerprints on the database. Each E contains the RSSI feature from the vector x_i which must satisfy the following conditions [4]:

$$d(\bar{x} - \overline{x_i}) \leq d(\bar{x}, \overline{x_j}) \quad (3)$$

$$W_K = \{E_1, \dots, E_K\} \quad (4)$$

$$\overline{x_{1:K}} = \{\overline{x_1}, \dots, \overline{x_K}\} \quad (5)$$

The estimated location, $\hat{E}$, by averaging the coordinated of the KNN classifier is given as

$$\hat{E} = \frac{1}{K}\sum_{i=1}^{K} E_i \quad (6)$$

2.2 Naïve Bayesian Classifier (NB)

The Naïve Bayes classifier uses the Bayes theorem to makes classification decisions with an assumption of conditional independence and uses conditional probabilities to make classification decisions. In the indoor localization system, the location of a user needs to be determined given the RSSI feature from APs. To achieve this the probability of RSSI given in each region and the probability of the regions needs to be calculated. The probability of the RSSI values is given by [6]:

$$P_{Reg} = \arg max[f(RSSI_i)|k)] \quad (7)$$

$$f(RSSI_i)|k) = \frac{f(k|RSSI) * f(RSSI_i))}{(\sum_1^N f(k|RSSI_i) * f(RSSI_i))} \quad (8)$$

Where $f(RSSI_i)|k)$ and $f(k|RSSI)$ are prior likelihood distributions. $(\sum_1^N f(k|RSSI_i) * f(RSSI_i))$ and $f(RSSI_i)$ are constant in all cases as there is no information about the user's position, hence making it a maximum likelihood estimate [6]:

$$P_{ML} = \arg max[f(k|RSSI_i)] \quad (9)$$

Hence, maximising the likelihood probability will help make a decision. This problem is simplified to [6]:

$$g(K) = \arg max\left[p(k)\prod_1^n p(RSSI_i|\,k)\right] \quad (10)$$

As the NB classifier assumes these probabilities are conditionally independent.

2.3 Mean-offset classifier

The mean-offset classifier computes the percentage of error between the users RSSI against the database of trained RSSI centroids, after which the grid with the lowest percentage of error is chosen as the location, Fig. 1. Assuming the area of interest has A APs and N grids, each grid with the physical location $G_n(x_i, y_i)$ has a corresponding fingerprinting vector $f_i = \{AP1_{G_1}, AP2_{G_1}, \ldots, AP(A)_{G_1})$ where $\overline{x_{AP(A)_{G_n}}}$ is the database of centroids computed across each grid as follows:

$$\overline{x_{AP1_{G_1}}} = \sum_{k=1}^{n} \frac{AP1_{G1}(F_k)}{n} \tag{11}$$

During the training phase, the stored RSSI features in a database are clustered based on the proximity measure, which quantifies the similarity between the RSSI vectors. In this scenario RSSI collected from the same room in an indoor environment are clustered together, Fig 2. Where the prediction of the location, (x_i, y_i), with the least error, $argmin(e_k)$, between a test sample and each grids centroid is given as:

$$e_k = \left[\sum_{n=1}^{n} \left(\frac{\left|\overline{x_{AP1_{G_n}}} - AP1(F_k)\right|}{x_{AP1_{G_n}}} + \frac{\left|\overline{x_{AP2_{G_n}}} - AP2(F_k)\right|}{x_{AP2_{G_n}}} + \frac{\left|\overline{x_{AP3_{G_n}}} - AP3(F_k)\right|}{x_{AP3_{G_n}}} \right) \times 100 \right] \tag{12}$$

$$(x_i, y_i) = argmin(e_k) \tag{13}$$

Where $AP1(F_k), AP2(F_k), AP3(F_k)$ are the test samples, Fig. 1.

3 Experiment and Results

3.1 Experimental setup

The indoor localization system was carried out on the second floor of a residential space. The dimension of the residential environment is 11 m by 5 m. It has three bedrooms (outlined in red, green and blue), a passage (outlined in purple) and one bathroom (outlined in orange), which are separated by walls and cupboards Fig. 2. There are 119 testing grids that are clustered as outlined in Fig. 2. In training, five RSSI measurements per AP, a total of 1785 RSSI measurements, were collected across all the grids. In the online phase, the user moved to five random locations and RSSI measurements from three APs were stored to a server via an API call and exported to Python, which applied the classifiers to locate the user's position.

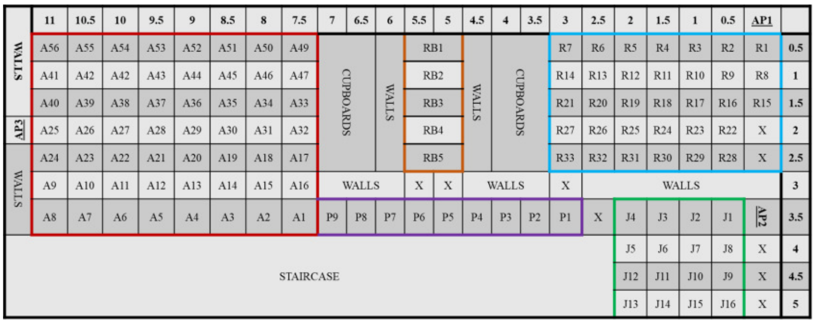

Fig. 2. Floor map of the residential area. The blue, green, purple, orange and red outlines indicate the clustered areas.

3.2 Classifier comparison

To evaluate the classification techniques, the RSSI measurements collected during the training phase where applied to K-NN, NB and Centroid-Offset classifiers. The classifiers selected where provided training data that ranged from 10% - 90% as an efficient classifier needs to be able to adapt to scenarios where only minimum training data is available. The mean offset classifier outperforms the KNN and NB classifier when the split between the training-test data is between 20 % and 75% Fig. (3). As more training data are provided, the NB classifier marginally outperforms the mean-offset classifier. However, the NB classifier also struggles when < 60% training data are provided. Overall, the mean-offset classifier consistently achieves accuracy of within 1.3 m.

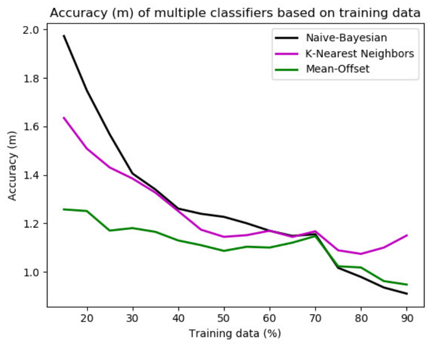

Fig. 3. Accuracy of Naïve Bayesian, KNN and mean offset classifiers under different training scenarios

3.3 Mean-offset classifier

The mean-offset classifier should account for dynamic changes presented by an indoor environment, such as temporal fluctuations in RSSI when recorded from the same point over multiple days and the movement of people and furniture. The Mean-Offset classifier was applied under two cases: tests were conducted on different times over multiple days and with movement of people.

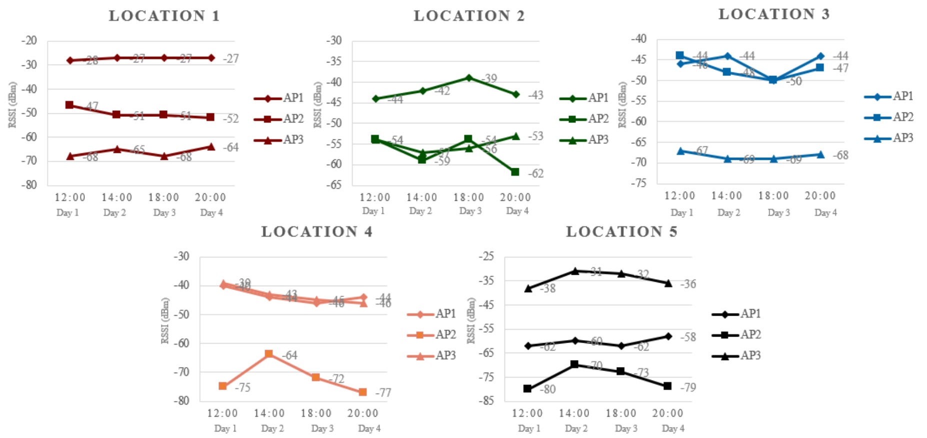

Fig. 4. Measured RSSI values from five different locations over four days indicating the fluctuation experienced in RSSI values

The test was conducted on five different locations over a period of four days. The fluctuations in RSSI across the three different APs was recorded and is represented in Fig. 4. The RSSI values had an average fluctuation of – 3dBm across the four days. The average accuracy over the locations at 12:00, 14:00, 18:00 and 20:00 was 1.08 m, 1 m, 0.83 m, 0.94 m and 0.58 m respectively Fig. 5. The accuracy over five locations across four days was 0.89 m.

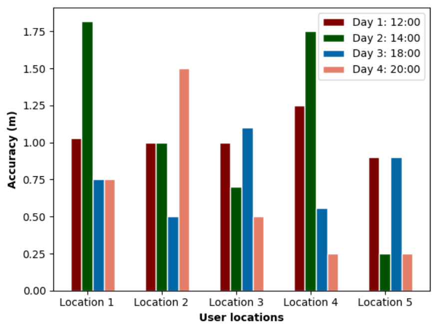

Fig. 5. Average localization errors across five different locations

Mean-offset classifier was selected based on its performance as it shows adaptability under uncertain conditions and in instances where the environment is changing rapidly and the RSSI values are fluctuating. The test data gathered in an indoor residential area are applied to the mean-offset classifier and the results are tabulated in Table 2. Under LOS conditions an accuracy of 0.85 and 1.15 m was achieved in residential area.

Table 2. Comparison in accuracy (m) under LOS and NLOS scenarios in a residential area are using mean-offset classifier

LOS conditions (Residential)		**NLOS conditions (Residential)**
Test	**Accuracy (m)**	**Accuracy (m)**
1	1.1	0.97
2	1	1.4
3	0.83	1.34
4	0.95	0.75
5	0.58	1.3
Avg	**0.86**	**1.15**

4 Conclusion and Future work

In conclusion, the mean-offset classifier was proposed for fingerprint indoor localization using Wi-Fi. The mean-offset classifier computes the percentage of error between the users of RSSI against the database of trained RSSI centroids, after which the grid with the lowest percentage of error is chosen as the location implemented. Experimental results have demonstrated that the mean-offset classifier achieves an average localization of 0.86 m with over 70% of errors under 1 m, which outperforms other kernel-based, KNN, PCA-SVM, linear-SVM, gaussian-SVM, decision tree and random forest classifiers. Furthermore, main challenges such as RSSI instability and temporal ambiguity have been mitigated by the mean-offset classifier. The mean-offset classifier needs to be implemented for dynamic localization under several indoor locations such as malls, office spaces and underground parking. A hybrid between Artificial Neural Networks (ANN) and mean-offset is a concept that will be further investigated.

References

1. Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI : Indoor Localization via Channel Response," *ACM Comput. Surv.*, vol. 46, no. 2, p. 25, 2013.
2. A. Kushki, K. N. Plataniotis and A. N. Venetsanopoulos, "Kernel-Based Positioning in Wireless Local Area Networks," in *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 689-705, June 2007.
3. X. Li, X. Cai, Y. Hei, and R. Yuan, "NLOS identification and mitigation based on channel state information for indoor WiFi localisation," in *IET Communications*, vol. 11, no. 4, pp. 531-537, 3 9 2017.
4. A. H. Salamah, M. Tamazin, M. A. Sharkas and M. Khedr, "An enhanced WiFi indoor localization system based on machine learning," *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alcala de Henares, 2016, pp. 1-8.
5. L. Zhang, Y. Li, Y. Gu and W. Yang, "An efficient machine learning approach for indoor localization," in *China Communications*, vol. 14, no. 11, pp. 141-150, Nov. 2017.
6. K. Zia, H. Iram, M. Aziz-ul-Haq and A. Zia, "Comparative study of classification techniques for indoor localization of mobile devices," 2018 *28th International Telecommunication Networks and Applications Conference* (ITNAC), Sydney, NSW, 2018, pp. 1-5.