International Journal of
Intelligent Systems and Applications in Engineering

ISSN:2147-6799 www.ijisae.org Original Research Paper

# A Hybrid Approach for Indoor Positioning

**Sinem BOZKURT KESER*[1], Uğur YAYAN[2]**

*Accepted 3rd September 2016*

***Abstract:*** Positioning systems have wide range of applications with the developing technology. Global Positioning System (GPS) is an efficient solution for outdoor applications but it gives poor accuracy in indoor environment. And, various methods are proposed in the literature such as geometric-based, fingerprint-based, etc. In this study, a hybrid approach that uses both clustering and classification is developed for fingerprint-based method. Information gain based feature selection method is used for selection of the most appropriate features from the WiFi fingerprint dataset in the initial step of this approach. Then, Expectation Maximization (EM) algorithm is applied for clustering purpose. Then, decision tree algorithm is used as a classification task for each cluster. Experimental results indicate that applied algorithms lead to a substantial improvement on localization accuracy. Since, cluster specific decision tree models reduce the size of the tree significantly; computational time of position phase is also reduced.

***Keywords:*** *Fingerprinting, indoor positioning, access point selection, clustering, classification, feature selection, expectation maximization, decision tree, received signal strength, WLAN.*

## 1. Introduction

Positioning systems are becoming very important with the developing technology. These systems are used to determine the position of the mobile device and they are used by location-based services for navigating, or tracking etc. They are classified as outdoor and indoor positioning systems. Global Positioning System (GPS) is used for outdoor in our daily life. But, it cannot be used for indoor environment due to lack of GPS signals including multipath and signal blockage [1]. Therefore, different kinds of indoor positioning systems (IPSs) have been developed. WiFi-based indoor positioning systems have become an attractive solution in indoor area since WiFi access points (APs) can be found widely in indoor buildings such as airports, shopping malls, or office buildings, etc. WiFi-based indoor positioning is an inexpensive solution because they have not need any additional installation cost [2].

Fingerprinting technique is the most accurate technique that is based on Received Signal Strength (RSS) in WiFi-based indoor positioning systems [3]. In this technique, the RSSs obtained from WiFi APs are used to determine the position of a mobile device. It contains two phases named as offline and online phases. In offline phase, RSS values are collected at known reference points (RPs) in the experimental area and then are used to construct radio map. The dimension of the radio map is grown while including all the detectable APs in the region. But all these APs are not contributed positively to the accuracy of position. Therefore, the redundant APs will be removed from the radio map, since they increase the computational cost and also cause deterioration in the accuracy of position. For this purpose, the redundant APs are eliminated using feature selection methods.During the online phase, the radio map is used to estimate the location of a mobile device using a new RSS measurement using machine learning (ML) algorithms such as Decision Tree (DT), K-nearest neighbour (KNN), or Support Vector Machine (SVM) etc. [4].

In this study, a hybrid approach that combines clustering and classification algorithms is applied. Information gain (InfoGain) based feature selection is firstly used to remove redundant APs in the radio map. Then, EM clustering algorithm is utilized to divide radio map into distinct groups or clusters. Finally, after assigning each test data to correct cluster, DT classifier algorithm is used as a classification task. In the experiments, DT and applied hybrid approach is compared in terms of positioning accuracy using WEKA open source machine learning toolbox and RFKON database [5].

The paper is organized as the follows. Related works in literature for fingerprinting based indoor positioning are given in Section.2. Section.3 focuses on applied hybrid approach introducing the algorithms using in this approach. Experimental area is given in Section.4. Test results for proposed algorithm are given in Section.5. As a result of this paper, conclusion part summarizes all written things in the paper with a short paragraph including future works in Section.6.

## 2. Related Works

A large number of studies that adopt fingerprinting as the position estimation method are proposed in the literature. Fingerprint-based positioning algorithms can be categorized into two groups: deterministic algorithms [2, 6-8] and probabilistic algorithms [9-13].

Deterministic algorithms are used to find the minimum signal distance between the newly measured RSS vector and pre-measured fingerprints which are vectors of RSSs from detectable APs in the region (radio map). Each fingerprint in the radio map

[1] *Department of Computer Engineering, Eskisehir Osmangazi University, Eskisehir, Turkiye*
[2] *R&D Department, Inovasyon Muhendislik Ltd. Sti., Eskisehir, Turkiye*
* *Corresponding Author:Email: sbozkurt@ogu.edu.tr*
*Note: This paper has been presented at the 3rd International Conference on Advanced Technology & Sciences (ICAT'16) held in Konya (Turkey), September 01-03, 2016.*

is associated with a known RP. In [2], KNN is used to estimate the position of a mobile WiFi-enabled device. A feature scaling based KNN (FS-kNN) algorithm is proposed to improve localization accuracy[6]. In FS-kNN, different weights are assigned to signal differences at different RSS levels when estimating the similarity between two RSS vector. Artificial neural network (ANN) classifier is utilized to classify link quality patterns for each location in [7]. In [8], DT algorithm is used as the classification algorithm.

In probabilistic algorithms, the probability of the mobile device position to be each reference point is calculated and then maximum probability is returned as a mobile device position. In [9], particle filter, a Bayesian based method, is employed. Observed signal strengths are obtained using Bayesian inference in [10] and the estimated position is determined as the highest probability in the resulting distribution. An extended Kalman filter based approach is presented in [11], where the intra cell position of a cellular device is estimated using RSS readings from base stations. This estimate, movement pattern data and velocity vectors are combined in order to predict the next cell crossing. In [12], a Bayesian filter based approach is proposed. In this study, a posterior probability distribution over the target's location is obtained by inverting Bayesian belief network. In [13], subset of the strongest APs is considered instead of all APs and target location is predicted using Bayesian estimate.

Various algorithms are proposed in the literature to reduce the computational cost by eliminating redundant APs in the radio map. In [14], according to strength of the signal a weight is assigned to each AP and then the APs with minimum weights are dropped from each fingerprint. Various AP significance measures such as average RSS, entropy, variance, maximum RSS are examined in [15]. Fast Orthogonal Search (FOS) and modified FOS (mFOS) algorithms are implemented in [16] in order to reduce the dimensionality of the radio map in an IPS.

## 3. Proposed Hybrid Approach for Wi-Fi Based Indoor Positioning

In this study, a hybrid approach that combines clustering and classification algorithms is applied. InfoGain- based feature selection is firstly used to remove redundant APs in the radio map. Then, EM clustering algorithm is utilized to divide radio map into distinct groups or clusters. Finally, after assigning each test data to correct cluster, DT classier algorithm is used as a classification task. These algorithms are described in the following subsections.

### 3.1. InfoGain-based Feature Selection

InfoGain is the most commonly used feature selection method in the machine learning field that is based on the entropy [16]. Information gain of each feature is calculated using (Equation.1).

$$\mathrm{IG}\ (\mathrm{f}) = -\sum_{c,\bar{c}} P(c)\log(c) + \sum_{f,\bar{f}} P(c\mid f)\log P(c\mid f) \qquad (1)$$

where f is the feature and c is the class.

### 3.2. Expectation Maximization (EM) Clustering Algorithm

Clustering algorithms assign similar data to same cluster without the prior knowledge about the data's characteristics. Since the data's labels' are not known, these algorithms are also called as unsupervised learning algorithms [17]. EM algorithm is a clustering algorithm that assigns data to particular clusters by computing one or more probability distributions. It then maximizes the overall probability of the data belonging to a certain cluster [18]. EM algorithm consists of two steps: determination of expectation and maximization of expectation iteratively. In this study, EM algorithm in WEKA is used for clustering purpose.

### 3.3. Decision Tree Classifier Algorithm

Decision Tree predicts an output by tracking the decisions in the tree from the root node down to a leaf node according to the outcome of the tests along the path [19]. In this study, C4.5 that is a benchmark tree (J48 in WEKA) is applied in the classification step.

## 4. Experiment

Data were collected to construct our database for Eskisehir Osmangazi University Teknopark. It has two floors of area of $800m^2$. This area was broken into grid squares (each of size $2.4m \times 2.4m$). We collect the data from the first floor of Teknopark and the center of each grid square was noted as seen in Figure.1.

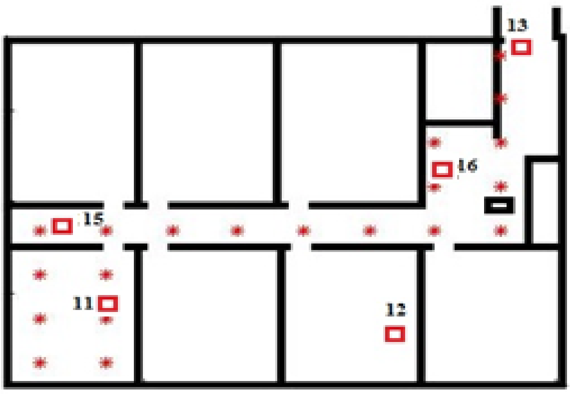


**Figure 1.** Experimental environment of floor 1

As seen in Figure.1, red stars represent the reference points that are used for collecting sensor values from the APs and red squares represent sensor nodes in the test area. The database and the experimental area are briefly described in [5]. There are 5 sensor nodes and 20 reference points in the original RFKON database. And, WiFi RSS values of 80 are obtained from per reference point (RP). There are 1600 instances in the train data and 1600 instances in the test data. In experiments, we use mobile-based WiFi RFKON database that is obtained using Sony Xperia mobile device.

## 5. Experimental Results

In experiments, we first apply "InfoGainAttributeEval" function in WEKA to determine the number of most important APs using both train and test data with Decision Tree classifier. These results are given in Figure.2.

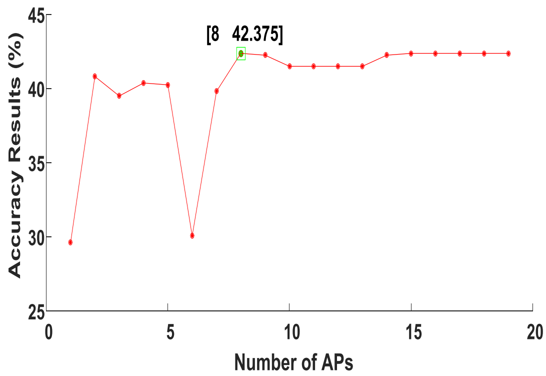


**Figure 2.**Number of APs determination using both train and test database with InfoGainAttributeEval

There are 19 APs in the database before applying Information gain based feature selection algorithm. As seen in Figure.2, we obtain best accuracy results after selecting 8 APs using InfoGain-based feature selection algorithm. This is an important improvement of reducing computational time.

In the clustering phase, EM algorithm is applied. In experiment, different number of clusters are tried to select the best number of clusters. 5 clusters give best accuracy results among the attempted cluster numbers. The RPs are numbered as seen in Figure.3 to illustrate the cluster assignments clearly.

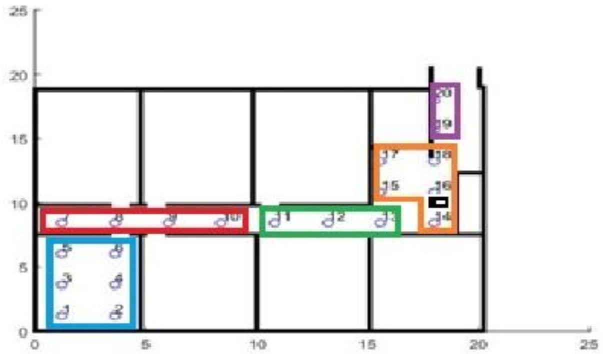

**Figure 3.**The representation of each reference point

Each RP that illustrated with numbers in Figure.3 is assigned to a cluster using EM algorithm as seen in Table.1.

**Table 1.**EM Clustering Assignments

| *Cluster Name* | *RP Number* |
|---|---|
| Cluster0 | 7, 8, 9, 10 |
| Cluster1 | 19, 20 |
| Cluster2 | 11, 12, 13 |
| Cluster3 | 14, 15, 16, 17, 18 |
| Cluster4 | 1, 2, 3, 4, 5, 6 |

Finally, DT classifier is applied for each cluster. In this step, cluster specific DT models are constructed. This step is reduced the total computational time of classification process while increasing the accuracy results. The comparison of the applied hybrid algorithm with DT algorithm is given in Table.2.

**Table 2.** Comparison of Accuracy Results

| *Algorithm* | *Number of APs* | *Accuracy Results (%)* |
|---|---|---|
| DT | 19 | 42.25 |
| DT | 8 | 42.375 |
| Hybrid Approach (EM-DT) | 8 | **66.42** |

The accuracy result of DT classifier using all APs in the database is 42.25. However, we obtain nearly same result (42.375) when we select 8 APs after applying InfoGain-based feature selection algorithm. As a result of Table.2, the hybrid approach increases the accuracy (66.42) about %25 using 8 APs. To demonstrate the applied hybrid algorithm improvement on the decision tree size, Table.3 is constructed.

**Table 3.** Decision Tree Size

| *Train Database* | *Number of Leaves* | *Size of the Tree* |
|---|---|---|
| 19 APs | 34 | 67 |
| 8 APs | 35 | 69 |
| Cluster0 | 3 | 5 |
| Cluster1 | 8 | 15 |
| Cluster2 | 6 | 11 |
| Cluster3 | 2 | 3 |
| Cluster4 | 5 | 9 |

As seen in Table.3, when decreasing the number of APs to 8 after applying InfoGain-based feature selection algorithm, number of leaves and size of the tree are increased. So, feature selection does not make an improvement on the tree size solely. Since, same instances are grouped together when applying EM clustering algorithm, the size of the decision tree is dramatically reduced. This causes lower computational time in the classification step in addition to improvement on the accuracy results as seen in Table.2.

## 6. Conclusions

In this study, a hybrid approach that uses both clustering and classification is applied in WiFi fingerprint-based method. Redundant APs are eliminated from the WiFi fingerprint dataset using InfoGain-based feature selection method in the initial step of this hybrid approach. In the clustering step, Expectation Maximization (EM) algorithm is applied. In the last step named as classification step, decision tree models are constructed for each cluster. Experimental results indicate that applied algorithms lead to a substantial improvement on localization accuracy. In addition to this, by the help of the clustering phase of the applied approach and constructing cluster specific DT models reduce the size of the tree significantly.

In a future work, different clustering and classification algorithms will be evaluated in a hybrid approach to get better results in terms of accuracy and computational time.

## Acknowledgements

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

## References

[1] G. M. Djuknic, and R. E. Richton, “Geolocation and Assisted GPS,” IEEE Computer, vol. 2, pp. 123–125, Feb. 2001.

[2] P. Bahl and V. N. Padmanabhan, “RADAR: An InBuilding RF-based User Location and Tracking System,” IEEE INFOCOM, 2000, pp. 775–784.

[3] A. Abusara and M. Hassan, “Enhanced fingerprinting in wlan-based indoor positioning using hybrid search techniques,” in International Conference on Communications, Signal Processing, and their Applications (ICCSPA), 2015, pp. 1–6, Feb. 2015.

[4] H. Liu, H. Darabi, P. Banerjee, and J. Liu, “Survey of wireless indoor positioning techniques and systems,” Systems, Man, and Cybernetics, Part C: IEEE Transactions on Applications and Reviews, vol. 37, pp. 1067–1080, Nov. 2007.

[5] S. Bozkurt Keser, U. Yayan, A. Yazici, S. Gunal, "A priori verification and validation study of RFKON database", International Journal of Computer Science: Theory and Application, vol. 5, 20-27, 2016.

[6] D. Li, B. Zhang, Z. Yao and C. Li, "A feature scaling based k-nearest neighbor algorithm for indoor positioning system," 2014 IEEE Global Communications Conference, Austin, TX, 2014, pp. 436-441.

[7] Y. Ha, E. Ae-cheoun, and B. Yung-cheol, "Efficient sensor localization for indoor environments using classification of link quality patterns", International Journal of Distributed Sensor Networks, 2013.

[8] S. Eisa, J. Peixoto, F. Meneses, and A. Moreira, "Removing useless APs and fingerprints from WiFi indoor positioning radio maps", International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp.1-7, Oct. 2013.

[9] V. Seshadri, V. G. Zaruba, and M. Huber, "A Bayesian sampling approach to in-door localization of wireless devices using received signal strength indication", Third IEEE International Conference on Pervasive Computing and Communications (PERCOM 2005), pp.75-84, March 2005.

[10] X. Chai, and Q. Yang, "Reducing the calibration effort for location estimation using unlabeled samples", Third IEEE International Conference on Pervasive Computing and Communications (PERCOM 2005), pp. 95-104, March 2005.

[11] T. Liu, P. Bahl, and I. Chlamtac, “A hierarchical position-prediction algorithm for efficient management of resources in cellular networks”, Global Telecommunication Conference (GLOBECOM ’97), IEEE, vol. 2, pp. 982-986, Nov. 1997.

[12] M. Isard, and A. Blake, “Contour tracking by stochastic propagation of conditional density”, Computer Vision (ECCV ’96), pp. 343-356, 1996.

[13] M. A. Youssef, A. Agrawala, A. U. Shankar, and S. H. Noh, “A probabilistic clustering-based indoor location determination system”, Tech. Report, University of Maryland at College Park, CS-TR 4350, March 2002.

[14] J. Ledlie, “Method and apparatus for on-device positioning using compressed fingerprint archives,” June 2011.

[15] E. Laitinen, E. Lohan, J. Talvitie, and S. Shrestha, “Access point significance measures in WLAN-based location,” in 2012 9th Workshop on Positioning Navigation and Communication (WPNC), pp. 24–29, Mar. 2012.

[16] A. Abusara, M. S. Hassan, and M. H. Ismail, "RSS fingerprints dimensionality reduction in WLAN-based indoor positioning", IEEE 2016 Wireless Telecommunications Symposium (WTS), April 2016.

[17] A. G.Karegowda, A. S. Manjunath, and M. A. Jayaram, “ Comparative study of attribute selection using gain ratio and correlation based feature selection”, International Journal of Information Technology and Knowledge Management, vol. 2,pp. 271-277, Dec. 2010.

[18] I.H. Witten, E Frank, MA Hall. Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques. Elsevier, 2011.

[19] L. Rokach, O. Maimon, Decision Trees. In The Data Mining and Knowledge Discovery Handbook. Springer, pp. 165–192, 2005.