A Comparative Analysis of N-Nearest Neighbors (N3) and Binned Nearest Neighbors (BNN) Algorithms for Indoor Localization

Serpil Ustebay[1(✉)], M. Ali Aydin[1], Ahmet Sertbas[1], and Tulin Atmaca[2]

[1] Department of Computer Engineering, Istanbul University, Istanbul, Turkey
{serpil.ustebay,aydinali,asertbas}@istanbul.edu.tr
[2] Laboratoire Samovar, Telecom SudParis, CNRS, Université Paris-Saclay, Evry, France
tulin.atmaca@telecom-sudparis.eu

Abstract. In this study, performances of classification algorithms N-Nearest Neighbors (N3) and Binned Nearest Neighbor (BNN) are analyzed in terms of indoor localizations. Fingerprint method which is based on Received Signal Strength Indication (RSSI) is taken into consideration. RSSI is a measurement of the power present in a received radio signal from transmitter. In this method, the RSSI information is captured at the reference points and recorded for creating a signal map. The obtained signal map is knows as fingerprint signal map and in the second stage of algorithm is creating a positioning model to detect individual's position with the help of fingerprint signal map. In this work; N-Nearest Neighbors (N3) and Binned Nearest Neighbors (BNN) algorithms are used to create an indoor positioning model. For this purpose; two different signal maps are used to test the algorithms. UJIIndoorLoc includes multi-building and multi floor signal information while different from this RFKON includes a single-building single floor signal information. N-Nearest Neighbors (N3) and Binned Nearest Neighbors (BNN) algorithms are presented comparatively with respect to success of finding user position.

Keywords: N3 and BNN algorithms · Indoor localization · RSSI · Fingerprint · Accuracy

1 Introduction

With a common infrastructure such as the Internet, different applications have became available anywhere at any time. It is possible to obtain the instant status information of the person or objects with the positioning systems, Positioning systems are divided into two basic groups, indoor and outdoor. Outdoor satellite-based positioning systems achieve high success by processing more than one satellite information with triangulation. For this, some information is needed like the arrival angle of signals, arrival time of signals and so on. Also, the signals must be within line-of-sight.

P. Gaj et al. (Eds.): CN 2017, CCIS 718, pp. 81–90, 2017.
DOI: 10.1007/978-3-319-59767-6_7

Many countries try to implement their own GPS applications. GLONASS [1], Galileo [2], Beidou [3] and IRNSS (NAVIC) [4] can be given as examples for developed applications. Although recommended indoor GPS [5] systems are realized for indoors, it is a disadvantage that the devices used are costly. The search for a high performance, low cost solution continues. There is no standard for any positioning systems. Offered solutions locate people inside a building using radio waves, Bluetooth and infrared signals, magnetic fields, acoustic signals or other sensory information. Indoor positioning is divided into two categories, active and passive. In active positioning, it is the direct transmission of information in positioning by means of a device carried by the users. RFID, Bluetooth, Infrared, ultra-wideband, IEEE 802.11 WLAN are examples of these. In passive systems they do not transfer information directly with any device they carry on them. Differential Air, Computer vision, Device-free Passive systems can be given as examples of passive systems [6]. Fingerprint, is a localization method based on obtaining signal map of place by power of signal of IEEE 802.WLAN. This technique is applied just other sensor data like magnetic field, RFID, Bluetooth.

The fingerprint consists two step that are called online and offline stage as shown on Fig. 1. Offline phase consists the radio map construction process. Although the offline phase takes a lot of time and effort, it provides high success rate and low cost implementation. Reference points are selected inside the building and Received Signal Strength (RSS) are measured from Wireless Access Points (WAP) with the help of a mobile device. Also Detailed information about

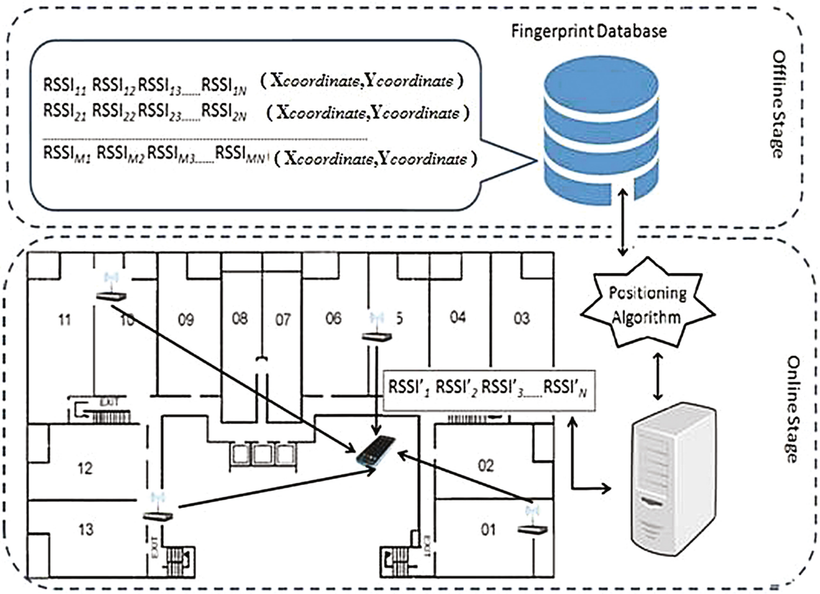

Fig. 1. Fingerprint algorithm stages

reference point as building id, floor id, space name...etc are recorded. RSS measurements can easily affected from noise (objects, people, wheatear.. etc.) to obtain high accuracy in the noisy wireless channel, different mobile devices are used to capture RSS. The signal map is used to generate a positioning model at second stage. The model can be grouped by mathematical formulas. Geometry based methods like triangulation. This method facilitates validation of data through cross verification from two or more sources. Probability based mathematical approaches can be used like Bayesian inference techniques. Machine learning algorithms especially classification methods [7–10], neural network [11–13], deep learning [14–16]...etc can be used to create high performed positioning model.

This study is aimed at finding position of a person in indoors by using fingerprint method. At positioning stage, N3 and BNN algorithms based on nearest neighbor approach are applied. Success and error distance of algorithms are tested on two different databases are presented.

2 Use of Wireless Signal Maps

Two different signal maps are used to test the algorithms. The first one is named RFKON and second one is named UJIIndoorLoc. The constructions from which both signal maps are obtained are shown in Fig. 2.

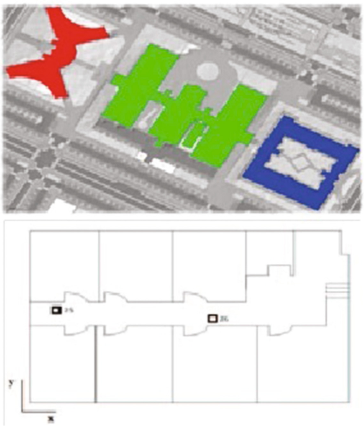

Fig. 2. Red, Green and Blue refers to the building of UJI Riu Sec Campus (top). KIOS Research Center floor plan (down) (Color figure online)

The Area was broken into 1.2×1.2 m grid squares and Received Signal Map Strength (RSS) from available Wireless Access Points (WAP) was measured at the center of each grid square to get RFKON [17]. Four different mobile devices are used to capture RSS and the number of reference points appearing in the

Table 1. Data samples in RFKON

ID	Time	Device ID	x	y	Floor	Battery	MAC1	MAC2	...	MACN
1	01.07.2015 09:10	1	1.2	1.2	1	100	−97	−65	...	100
2	02.07.2015 09:10	2	1.2	1.2	1	100	−95	−64	...	100
3	03.07.2015 09:10	3	1.2	1.2	1	100	100	−70	...	100
4	04.07.2015 09:10	4	1.2	1.2	1	100	100	100	...	100

Table 2. UJIIndoorLoc signal map attribute information.

Attribute	Information
Attribute 001 (WAP001)	Intensity value for WAP001. Negative integer values from −104 to 0 and +100. Positive value 100 used if WAP001 was not detected
...	...
Attribute 520 (WAP520)	Intensity value for WAP520. Negative integer values from −104 to 0 and +100. Positive Vvalue 100 used if WAP520 was not detected
Attribute 521 (Longitude)	Longitude. Negative real values from −7695.9387549299299000 to −7299.786516730871000
Attribute 522 (Latitude))	Latitude. Positive real values from 4864745.7450159714 to 4865017.3646842018
Attribute 523 (Floor)	Altitude in floors inside the building. Integer values from 0 to 4
Attribute 524 (BuildingID)	ID to identify the building. Measures were taken in three different buildings. Categorical integer values from 0 to 2
Attribute 525 (SpaceID)	Internal ID number to identify the Space (office, corridor, classroom) where the capture was taken. Categorical integer values
Attribute 526 (RelPosition)	Relative position with respect to the Space (1 - Inside, 2 - Outside in Front of the door). Categorical integer values
Attribute 527 (UserID)	User identifier (see below). Categorical integer values
Attribute 528 (PhoneID)	Android device identifier (see below). Categorical integer values
Attribute 529 (Timestamp)	UNIX Time when the capture was taken. Integer value

database is 54. Each recorded measurements attributes and sample values are given at Table 1. 100 dbm means that there is no WAP information at related reference point.

Second signal map is UJIIndoorLoc [18] which is the biggest multi-building and multi-floor database in the literature and first publicly available database that could be used to make comparisons among different methods shown. Totally 520 WAP data were recorded by more than 20 users using 25 different models of mobile devices. BuildingID is used to identify buildings which could be 3 different value. FloorID is used to identify floors of building. SpaceID is used to identify the particular space (offices, labs, etc.) where the capture was taken. All attribute information is listed at Table 2.

3 Algorithms

In this section, the structure of the N Nearest Neighbor (N3), Binned Neighbor Neighbors (BNN) algorithms [19] and the comparison parameters used to measure the success of the algorithms are presented.

3.1 N Nearest Neighbors (N3)

N3 algorithm is based on nearest neighbours which is take into account all neighbours except k-nearest neighbours. When a test sample is classified, all neighbours are sorted from most similar to the least similar with this a similarity rank vector is acquired. A weight matrix is W_{ig} calculated to classify the test sample. W_{ig} is defined as the ith object's g th class weight and calculated on the Formula 1.

$$W_{ig} = \frac{1}{\hat{n}_g} . \sum_{j=1, j \neq i}^{(n-1)} \frac{s_{ij}}{r_{ij}^{\alpha}} . \delta_j \qquad \delta_j = \begin{cases} 1 \ if \ C_j = g \wedge \frac{s_{ij}}{r_{ij}^{\alpha}} > \varepsilon \\ 0 \qquad otherwise \end{cases} \tag{1}$$

$\hat{n}_g$ is the number of neigbors that contribute to the class weight and calculated with the Formula 2. r_{ij} is a similarity rank and modulated by α which is a real-values parameter to be optimized in the range [0.1, 2.5]. c_i is the class of the jth object. ϵ (e.g., 10^{-7}) is a threshold parameter to define Dirac delta (δ_i).

$$\hat{n}_g = \sum_j \delta_j \tag{2}$$

S_{ij} is the similarity between ith and jth object and calculated from their distances with Formula 3. d_{ij} is a distance measure between ith object and jth object.

$$S_{ij} = \frac{1}{1 + d_{ij}} \tag{3}$$

Finally, the calculated W_{ig} matrix is normalized with Formula 4.

$$\acute{W}_{ig} = \frac{W_{ig}}{\sum_g W_{ig}} \tag{4}$$

All objects class definition is making with $\acute{W}_{ig}$ values. These values can be interpreted as fuzzy measures. Maximum $\acute{W}_{ig}$ value defines ith object class which is predefined at wireless signal map.

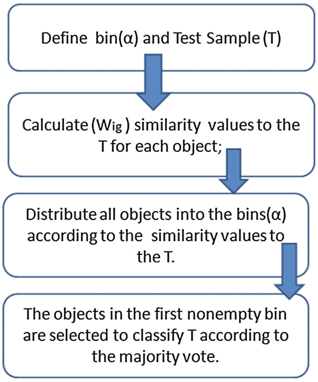

Fig. 3. The flow chart of the BNN algorithm

3.2 Binned Nearest Neighbors (BNN)

The BNN algorithm works similarly to the KNN algorithm. The test data are classified according to their nearest k neighbors. The BNN algorithm differs from the KNN algorithm in that it is the selection of neighbors used in classification. While k closest neigbors are selected in KNN algorithm, All neighbors are distributed into predefined similarity intervals (bins) and distributes into these intervals in BNN algorithm. Then, the first non-empty bin are chosen as the nearest neighbors in order to being used for prediction. The Prediction is taken as majority vote [19]. The flow diagram of the BNN algorithm is given in Fig. 3.

3.3 Comparison Parameters

Classification algorithm's performance are compared with different metrics. It affected by number of instances that are correctly classified, number of instances that are incorrectly classified, total number of instances, etc. The metrics which are used to compare performance of different classification algorithms are listed at Table 3.

Table 3. Algorithms comparison parameters

Name	Formula	Definition
Accuracy	$\frac{TP}{n}$	TP is the number of true predicted samples, n is the number of test instances
Error type 1	$\frac{\sum_i^n D_i}{n}$	D_i is ith class distance between predicted class and real class, n is the number of test samples
Error type 2	$\frac{\sum_i^n D_i}{FP}$	D_i is ith class distance between predicted class and real class. FP is the false predicted instance

4 Implementation and Results

In this study two different fingerprint map is used to test the BNN and N3 algorithms. The UJIndoorLoc dataset contains 21048 measurements (samples) taken from 520 KEN (attributes). However, 3329 measurements were used. The measurements used were re-labeled according to BuildingID, FloorID, and SpaceID, and 174 classes were distributed. RFKON data has 54 reference points recorded with coordinates values and 26 WAP measurements as listed at Table 4.

Table 4. Description of fingerprint signal map

Name	Number of sample	Number of variable	Number of class	Multi building	Multi floor
RFKON	18480	26	54	No	No
UJIIndoorLoc	33290	520	174	Yes	Yes

In machine learning algorithms; amount of data must be used in training process and amount of data must be used for to test accuracy of model. For this, 80% of data is separated for training, while 20% is for test is used both of fingerprint maps. For RFKON database, 14.784 records are used for training and 3696 records is used for test. This operation is repeated for 100 times so the average accuracy value is calculated. The optimal alpha value for N3 and BNN is determined 0.75. These values gave the highest accuracy according to the test results. Euclidean is used as the distance metric.

Table 5. Test results of BNN and N3 algorithms

	Algorithm	Error type 1	Error type 2	Accuracy
RFKON	BNN	0.3373	2.3949	85.9053
	N3	0.4031	2.3975	83.1872
UJIIndoorLoc	BNN	4.3420	20.6632	78.8905
	N3	2.546	14.2402	82.1514

Error distance shows the indicates how much error the position of a person is made. Error distance is a measurement may change related to reference point frequency. If frequent intervals of reference points are determined within the building, error distance will be reduced. Error distance of both signal maps is shown in Table 3. Obtained lower error types for RFKON could be a small grid squares as 1.2×1.2 m when measurements are taken at inside building.

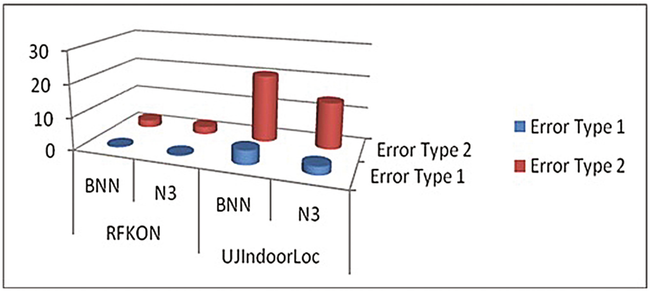

Fig. 4. Test results

5 Conclusion

In this study, two different indoor positioning models were created with fingerprint method. For this; N3 and BNN algorithms were used to define user/objects position and tested with the two different signal maps. RFKON is a data set based on signal strengths consisting of a single building and a single floor. Conversely, the UJIIndoorLoc data set consisting of multi-floor and multi-building signal map has been preferred. Effect of adding different features (building and floor information) to the N3 and BNN algorithms have been examined. When the test results are examined, it is seen that the performance of the algorithms changes according to the signal map. For UJIIndoorLoc, N3 algorithm has higher accuracy than BNN algorithm as shown Table 5. Besides, error distance is lower. N3 algorithm could not show the same performance on RFKON dataset as shown Fig. 4. In single floor data, BNN has high performance and low error distance. Lower accuracy can be obtained at multi floor signal maps than one floor signal map. Floor information affect RSSI captures.

The measurement number of fingerprinting signal data, capturing method, the number of the floor and building of location effects the accomplishment of positioning. The results of these algorithms do not sufficient for localization. For obtaining a more accurate localization system, new classification algorithms and hybrid approaches are required. For this purpose, different sensor data may be included to the system. Measured RSS values when floor information is considered can lead to errors and reduce accuracy.

Acknowledgements. This work is also a part of the PhD thesis titled "Design of an Efficient User Localization System for Next Generation Wireless Networks" at Istanbul University, Institute of Physical Sciences.

References

1. Bykhanov, E.: Timing and Positioning with GLONASS and GPS. GPS Solutions **3**(1), 26–31 (1999)
2. O'Keefe, K., Julien, O., Cannon, M.E., Lachapelle, G.: Availability, accuracy, reliability, and carrier-phase ambiguity resolution with Galileo and GPS. Acta Astronaut. **58**(8), 422–434 (2006)
3. Zeng, Q.H., Liu, J.Y., Hu, Q.Q., Yang, D.: Research on Beidou and GNSS multi-constellation integrated navigation. GNSS World China **1**, 013 (2011)
4. Zaminpardaz, S., Teunissen, P.J.G., Nadarajah, N.: IRNSS/NavIC and GPS: a single-and dual-system L5 analysis. J. Geodesy, 1–17 (2017)
5. Van Diggelen, F., Abraham, C.: Indoor GPS technology. In: CTIA Wireless-Agenda, Dallas, vol. 89 (2001)
6. Pirzada, N., Nayan, M.Y., Subhan, F., Hassan, M.F., Khan, M.A.: Comparative analysis of active and passive indoor localization systems. AASRI Procedia **5**, 92–97 (2013)
7. Talvitie, J., Renfors, M., Lohan, E.S.: Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals. IEEE Trans. Veh. Technol. **64**(4), 1340–1353 (2015)
8. Bozkurt, S., Elibol, G., Gunal, S., Yayan, U.: A comparative study on machine learning algorithms for indoor positioning. In: 2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA), pp. 1–8. IEEE, September 2015
9. Khudhair, A.A., Jabbar, S.Q., Sulttan, M.Q., Wang, D.: Wireless indoor localization systems and techniques: survey and comparative study. Indonesian J. Electr. Eng. Comput. Sci. **3**(2), 392–409 (2016)
10. Obeidat, H.A., Dama, Y.A., Abd-Alhameed, R.A., Hu, Y.F., Qahwaji, R.S., Noras, J.M., Jones, S.M.: A Comparison Between Vector Algorithm and CRSS Algorithms for Indoor Localization Using Received Signal Strength. University of Bradford Engineering and Informatics Publications (2016)
11. Lin, T.N., Lin, P.C.: Performance comparison of indoor positioning techniques based on location fingerprinting in wireless networks. In: 2005 International Conference on Wireless Networks, Communications and Mobile Computing, vol. 2, pp. 1569–1574. IEEE, June 2005
12. Fang, S.H., Lin, T.N.: Indoor location system based on discriminant-adaptive neural network in IEEE 802.11 environments. IEEE Trans. Neural Networks **19**(11), 1973–1978 (2008)
13. Dai, H., Ying, W.H., Xu, J.: Multi-layer neural network for received signal strength-based indoor localisation. IET Commun. **10**(6), 717–723 (2016)
14. Wang, X., Gao, L., Mao, S., Pandey, S.: CSI-based & fingerprinting for indoor localization: a deep learning approach. IEEE Trans. Veh. Technol. **66**, 763–776 (2016)
15. Faragher, R., Harle, R.: SmartSLAM-an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing. In: ION GNSS, vol. 13, pp. 1–14, September 2013
16. Gu, Y., Chen, Y., Liu, J., Jiang, X.: Semi-supervised deep extreme learning machine for Wi-Fi based localization. Neurocomputing **166**, 282–293 (2015)
17. Bozkurt, S., Yazıcı, A., Gunal, S., Yayan, U., Inan, F.: A novel multi-sensor and multi-topological database for indoor positioning on fingerprint techniques. In: 2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA), pp. 1–7. IEEE, September 2015

18. Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Avariento, J. P., Arnau, T.J., Benedito-Bordonau, M., Huerta, J.: Ujiindoorloc: a new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems. In: 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 261–270. IEEE, October 2014
19. Todeschini, R., Ballabio, D., Cassotti, M., Consonni, V.: N3 and BNN: Two new similarity based classification methods in comparison with other classifiers. J. Chem. Inf. Model. **55**(11), 2365–2374 (2015)