IMPROVED INDOOR POSITIONING TECHNIQUE BASED ON A GEOGRAPHIC WEIGHTED REGRESSION

A. Gholami [1], P. Pahlavani [2, *], S. Azimi [3], S.Shakibi [4]

[1] MSc. Graduate, Dept. of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran- amingholami328@gmail.com
[2] School of Surveying and Geospatial Engineering, College of Engineering., University of Tehran, Tehran, Iran-pahlavani@ut.ac.ir
[3] MSc. Graduate, Dept. of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran- sanazazimi@ut.ac.ir
[4] Dept. of RS and GIS, Faculty of Geography, University of Tehran, Tehran, Iran- sarashakibi95@gmail.com

KEY WORDS: Indoor Positioning, GWR, RSSI, Fingerprinting, Wi-Fi, Genetic Algorithm

ABSTRACT:

As technology and science develops and the coming of new equipment's, standards and different waves spread. Each of these standards and technologies have involved in indoor positioning by various scholars. Various methods have been developed based on different systems, all of which are based on specific methods and concepts. The research tries to do indoor positioning using local Wi-Fi fingerprints and signals. To reduce the error to collect local fingerprints, RSS values are recorded in 4 directions and two times. Geographic weighted regression method has been used to train the network. In this research, a genetic algorithm is used to select the appropriate parameters. Ultimately, the accuracy of the model has reached 1.76 cm. The results show that the increase in the number of access points does not affect the accuracy of position determination, but the choice of the effective access point will be effective in reducing the error.

1. INTRODUCTION

In GIS, the user's location is one of the priorities at any second, and the positioning is considered as an integrated one. In the outer space, due to the availability of GPS, this problem has been somewhat resolved, but the issue is to determine the position in the indoor environment. On the other hand, following the development of high buildings with complex architectures or crowded spaces such as shopping centres, the need for infrastructure to know the location of the user is more than ever. For this reason, many researches have been conducted over the past years and many technologies have been used to solve this need, the most common of which is efficiency of wireless networks. An internal positioning system with wireless network refers to the structure of a wireless network that generates local location information for each final user. A coordinate system or reference point within the pre - defined space is typically used to display the physical location of that unknown point. Recently, a path change has been made due to the focus in determining the outdoor positioning in determining the indoor position. Using the signal strength received by the user is the most obvious method. The received signal strength index can be used from the signal received from fixed reference points and reference tags that are placed in the specified locations for the user location.

One of the methods of determining the position based on the received signal strength is the wireless network signal, Wi-Fi, which is nowadays widely deployed on a wireless LAN. Wi-Fi is based on the IEEE802.11 standard and is primarily a local area network method that is designed for the inside of the building. The conventional Wi-Fi standard positioning algorithm can be divided into three categories: probability algorithm, triangulation algorithm and scene analysis.

In the scene analysis algorithm, the fingerprint location usually used based on the received signal strength indication (RSSI). The Scene analysis algorithm has the advantage of accuracy and does not require the location of the Wi-Fi receiver points, so it plays an important role in determining the indoor position. The fingerprinting method requires an offline phase where a set of calibrated points is measured, the measurement is the same, and while these RSSI and points of points are stored in a database. This information is known as radio map. Each database consists of a table in which the location of any ground point (x,y,z) and the RSSI vector are stored in those fingerprint points. The database is then used in an online and practical mode to estimate the position of the mobile device by the dependence of all available RSSI and the measured RSSI at that point.

Plenty of research have been done for indoor positioning. A number of aspects of Wi-Fi location fingerprinting based indoor positioning to improve the positioning accuracy was examined by Jekabsons and Zuravlyov (2010) with Locally Weighted Regression and was used in augmenting the radio map. The positioning error was a decent 7.14 m in term of using exclusively the APs from nearby buildings (Jekabsons and Zuravlyov, 2010).

Hwang et al. (2011) proposed the ZigBee wireless sensor network using ANN for accurate estimation of object's position when the signal strengths are unstable. They concluded that the accuracy of the polar coordinate system of the position of object is better than the rectangular form (Hwang et al., 2011).

Kehua et al. (2011) developed a hybrid Genetic Algorithm-Back Propagation Neural Network RFID. They concluded that the proposed method was used for indoor positioning within 10m with a proper precision (Kehua et al., 2011).

A positioning algorithm based on ANN that make use of time of arrival measurements and the angle of arrival information was proposed by Chen (2012) to estimate the location of a mobile station in non-line-of-sight environments. Their algorithm boost the precision of the mobile station location for the different levels of non-line-of-sight errors (Chen, 2012).

Guo et al. (2014) proposed two radial basis function neural networks based on RFID indoor positioning methods. These

https://doi.org/10.5194/isprs-archives-XLII-4-W18-441-2019

methods considered the received signal strength indication and the difference of the received signal strength as the inputs of the neural network. They improved the Gaussian filter to process received signal strength values and they utilized fuzzy clustering to determine the centers of the radial basis function. These methods enhance the reliability and precision of the RFID indoor positioning system (Guo et al., 2014).

Jan et al. (2015) developed a Wi-Fi indoor positioning based on the Kriging fingerprinting method. In this system, the error was estimated 3 dBm with a 72.2% probability than the surveyed RSS database without applying Kriging. Furthermore, the positioning error of their system is reduced by 17.9% in average compared to the case of without Kriging (Jan et al., 2015).

Han Zou et al. (2015) proposed a novel methodology based on standardizing RSS values in order to handle device heterogeneity and environmental changes. In addition, a robust indoor positioning system was developed in term of integrating the merits of both Signal Tendency Index (STI) and Weighted Extreme Learning Machine (WELM). The STI-WELM scheme improved the precision of indoor positioning by 39.89% compared to Received Signal Strength-Extreme Learning Machine (RSS-ELM), 33.54% compared with signal strength difference-Extreme Learning Machine (SDD-ELM) and 11.45% over Signal Tendency Index-Extreme Learning Machine (STI-ELM), respectively (Zou et al., 2015).

Pahlavani et al. (2017) introduced a multi-layer feed-forward (MLFF) artificial neural networks (ANN) for improving the positioning accuracy in indoor environment. In the offline phase of location fingerprinting, RSSs were collected at four directions in two time intervals (Morning and Evening). The Levenberg-Marquardt back-propagation algorithm was taken into consideration for the training phase of the proposed model. In the online phase of location fingerprinting, the location of user was estimated. The average positioning error for the proposed model containing 30% check and validation data was computed approximately 2.20 m (Pahlavani et al., 2017).

Zou et al. (2017) proposed a WiFi-based non-intrusive indoor positioning system for automatic online radio map construction and adaptation for calibration-free indoor localization. Moreover, Gaussian Process Regression (GPR) with Polynomial Surface Fitting Mean (PSFM-GPR) which was dedicated to predicting RSS on virtual reference points was proposed to capture the non-uniform RSS distribution over indoor environments. Regarding mentioned reliable regression technique, an average RSS estimation error and average localization accuracy were estimated 4.8 dBm and 1.718 m, respectively (Zou et al., 2017).

Subedi et al. (2019) proposed an extended two-step fingerprinting localization using the weighted centroid (WC) of nearby beacons, RSS, rank of locally available/seen beacons as a feature that defines reference points in fingerprinting to improve the accuracy of indoor positioning. Affinity propagation clustering was used in this method to decline the online computational cost of the system. Considering various fingerprinting features as well as RSS clustering in this method resulted in 1.05 m average localization estimation error in the corridor and also 1.38 m error in computer lab which was fully furnished (Subedi et al., 2019).

This study aims to increase the accuracy of positioning estimation by changing the input elements in the offline phase using the new learning method in the online phase. In order to learn the network in the online phase, the main objective of this research is to use the geographic weighted algorithm (GWR) to consider spatial autocorrelation and spatial non-stationary in the power vector regression of signals obtained by wireless networks and improving the location of this method. One of the secondary objectives of this study is determination of effective parameters by genetic algorithm among parameters for network training that result to the lowest error and computation time. One of the purposes of this study is to investigate the relationship between increasing the distance between fingerprint points and network error.

2. PROPSED METHOD

Since in different times an environment is changed to the system due to the variation of population density in that environment, the noise entered into the system can be changed. In the offline phase of this study, in addition to Access Point updates as inputs, the measurement time is also recorded at two in the morning or afternoon as input. As a result, a time parameter is added as an input parameter, on the other hand due to the variation of the received signal due to variation in the user's orientation. In the offline phase of this research, for every point at 4 angles, RSSIs were measured and stored in the database for measuring the point as an input parameter (as a result, a parameter is added to the inputs). Finally, in this research, 30 input parameters were used to learn and train the network. Since signal strength is dependent on location in the interior space and spatial data has properties such as spatial autocorrelation and spatial non-stationary of a place that makes it difficult to work with them. So, for the system learning GWR algorithm is used. In this research, the data structure for machine learning in GWR-GA network is used according to Table 1.

Input						Output
Time	Direction	$RSSI_1$	$RSSI_2$	…	$RSSI_{28}$	X
						Y

Table 1. The structure of input and output data for GWR-GA algorithm

According to Figure 1, the research method is divided into two phases, offline and online, which in the offline phase, the RSSI, time and direction data in fingerprint point are collected at first. Then the effective parameters are selected using the genetic algorithm while GWR network training with minimum error. Then a proper model is created. In the online phase, the data needed is collected by the user's application, then the inputs are entered into the network based on the GWR model, and the local location of the user is displayed in the user's map.

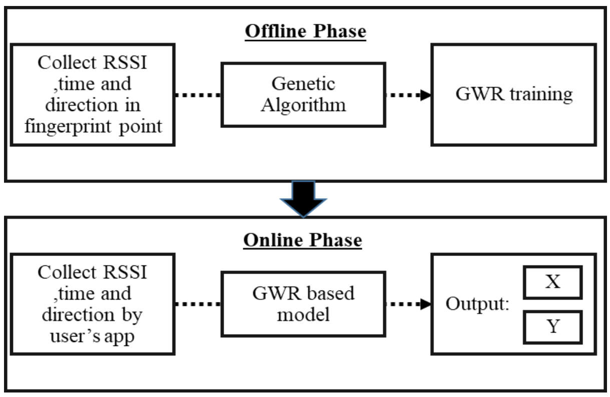

Figure 1. Process flowchart of the proposed method

Finally, this study aims to compare the proposed algorithm (GWR-GA) with other machine learning methods such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), GWR and Linear Regression model.

2.1 Finger Printing

A large number of the indoor RSS-based WLAN positioning systems use the fingerprinting technique for acquiring the RSS and position relation to estimate user's position accurately. The location fingerprinting is separated into two phases: offline and online phase. The offline phase of the location fingerprinting consists of measuring RSS or the other non-geometric features from different APs by WLAN-enabled mobile device at reference points to create a database of location fingerprints. This database is called as the radio map on the server. Furthermore, the training phase of the proposed model is performed in offline stage. RSS readings are collected at the regular period of time and particular orientation at each reference point. In the online phase of location fingerprinting, the device collects RSS readings in real-time to estimate the location of device.

2.2 Geographic Weighted Regression

Brunsdon et al introduced GWR method (Bivand et al., 2017; Brunsdon et al., 1998) and this method is an outgrowth of ordinary least squares regression (OLS). This method consider the spatial dependency of observation as the weight matrix. The regression coefficients are estimated locally at spatially referenced data points with GWR method because of environment homogeneity and non-stationarity. The equation of GWR is shown in the following (Brunsdon et al., 1998):

$$y = \beta_0(u,v) + \sum_{j=1}^{p} \beta_j(u,v)X_j + \varepsilon \qquad (1)$$

where y is the dependent variable, Xj is the independent variable, p is the number of independent variables, ε is the residual of the model, and B_j is the coefficients of regression that are a function of observation point (u,v). The GWR is the weighted adjustment and the coefficients of regression can be computed by (Brunsdon et al., 1998):

$$\hat{\beta}(u,v) = (X^T W(u,v) X)^{-1} X^T W(u,v) y \qquad (2)$$

where W matrix entails the geographical weights and this matrix is diagonal matrix as follows (Brunsdon et al., 1998):

$$\begin{bmatrix} W_1(u,v) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & W_n(u,v) \end{bmatrix} \qquad (3)$$

It is vitally important to note that determination of the geographical weights is necessary for GWR method. Therefore, several kernels have been considered. In this paper, Gaussian kernel is chosen as follows [15]:

$$W_{ij} = \varphi\left(\frac{d_{ij}}{\sigma h}\right) \qquad (4)$$

where W_{ij} is the geographic weight of observation j at the point i, and φ is the normal standard distribution function. d_{ij} is the distance between two points i and j, σ is the standard deviation for d_{ij} for each point and h is the bandwidth. d_{ij} is the Euclidean distance in a Cartesian coordinate system and also it is the orthodromic distance in a spherical coordinate system. Selecting appropriate bandwidth is crucial to determine the geographic weights due to the fact that GWR model approach OLS model as long as this bandwidth is too large, and if too small bandwidth is selected, the variance grow up (Charlton and Fotheringham, 2009).

There are various methods for determining the optimal bandwidth. One of the mentioned method is the Cross Validation method which can be computed by (Brunsdon et al., 1998):

$$\sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(h)]^2 \qquad (6)$$

where n is the number of observation, y_i is the observation i, and y_i is the estimated value for the observation i computed by the other observations. Moreover, y_i is a function of bandwidth and if bandwidth minimizes the function, it will be regarded as the optimal bandwidth.

The bandwidth is more efficient than the kind of kernel in the fitness process. Bandwidth can be selected by two methods which are Fixed bandwidth and Unfixed (changeable) bandwidth (Charlton and Fotheringham, 2009). If data are distributed regularly, fixed bandwidth will be used. Unfixed bandwidth is considered in the cases that data are almost irregular and have clustered distribution. Therefore, bandwidth decreases in the high density area and vice versa. The minimum and maximum of observation points in search band is one criterion for this change. In addition, bandwidth can be alter in a way that the fixed number of observations would stay on each band.

2.3 Genetic algorithm for determining optimal factors

A number of factors should be chosen to enhance the performance of GWR. In this paper, genetic algorithm is considered for determining optimal factors. This algorithm takes a correlation between factors into account to opt for the best collection of factors. John Holland introduced genteic algorithms (Delavar, 2004; Pahlavani et al., 2006; Bajpai and Kumar, 2010; Holland, 1992; Mitchell, 1998) in 1960 and these algorithms are inspired by the process of Darwin’s evolutionary ideas of natural selection and belong to the larger part of evolutionary algorithms (EA). Genetic algorithms are adaptive heuristic search algorithms and are used to solve optimizations problems (Nahr et al., 2014). In genetic algorithm, the population of the solutions based on the given problem is considered. Selection, crossover, and mutation are central operators of the simplest genetic algorithm(Mitchell, 1998).

2.4 Validation

After each of the methods is implemented and modeled, the results should be validated and compared. Therefore, in this research, three parameters of the coefficient of determination, time complexity and root mean square error using cross-validation method were used to evaluate the results. The determination coefficient shows the correlation between the observed values and the calculated values which is always between 0 and 1, the value of one represents a complete correlation between the observed values and the calculated values, and the zero value represents the independence of the observed values and the values calculated. The coefficient of determination and the root mean square error have been calculated using Equations (7) and (8) (Delavar et al., 2019; Junninen et al., 2004; Li and Heap, 2011).

$$R^2 = \left[\frac{1}{N}\frac{\sum_{i=1}^{N}[(P_i - \bar{P})(O_i - \bar{O})]}{\sigma_p \sigma_o}\right]^2 \tag{7}$$

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}[P_i - O_i]^2\right)^{1/2} \tag{8}$$

where N is the number of observations, O_i is the observed parameter, Pi is the calculated parameter, $\bar{O}$ is the mean of the observation parameter, $\bar{P}$ is the average calculation parameter, σ_o is the standard deviation of the observations and σ_p is the standard deviation of the calculation (Delavar et al., 2019).

3. EXPERIMENTAL RESULTS

In this study, 28 access points were used to collect RSSI for each point, the map of the study area is shown in Figure 2. As can be seen in this study, 67 regular points with a distance of 80 cm were used (first dataset), by selecting 27 points from this set of points, we reach points 160 cm apart (second dataset). The region's coordinate system is locally defined.

The RSS readings are collected at four common directions namely North, East, South, and West and in two periods of time, i.e. Morning and Evening, for each references point. Generally, 536 points was collected in the mentioned study area.

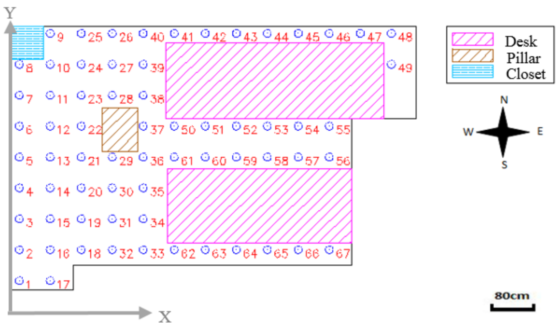

Figure 2. The plan of selected study area

In this study, ASUS laptop was used for collecting the RSS readings. The profile of this laptop is described in Table 2. The proper C# code was implemented in this laptop for collecting RSS readings in references points and inserting them with their directions and their time intervals to MySQL database. Then, this information saved within this database. In Figure 3, by pressing 'insert to DB' click in mentioned C# code, the RSS in one reference point at one direction in one time interval was collected and inserted to MySQL database.

Model	Asus, x550c
CPU	Intel Core i5 3337U 1.80GHz up to 2.70 GHz
Memory	6 GB RAM
Platform	Windows 10
Wireless networking support	802.11n

Table 2. The profile of Asus laptop was used for collecting the RSS readings

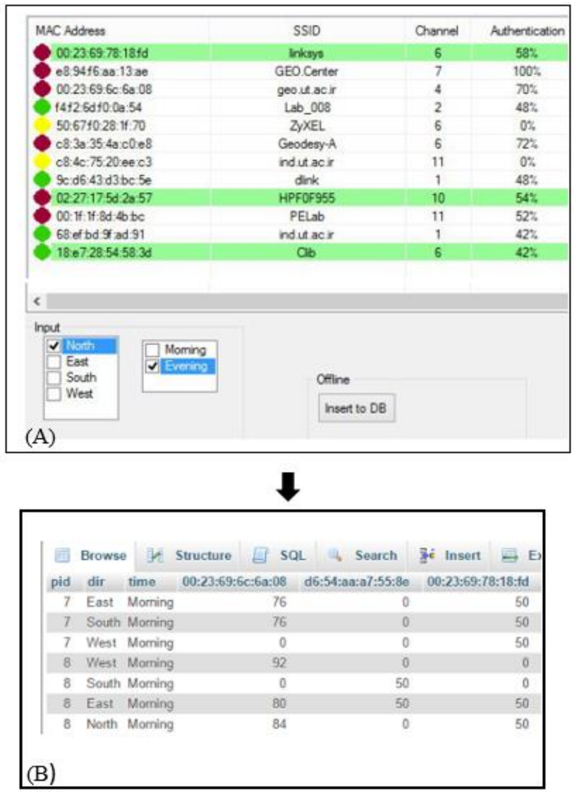

Figure 3. A) Proposed C# code for collecting, B) Insertion of RSS to MySQL database

After collecting the required database in offline phase in the two mentioned datasets (first and second dataset), ANN, SVR, Linear Regression, GWR and GWR-GA methods are used for each dataset to learn the network. The results are described in Table 3 and Figure 4 show the correlation between the actual and prediction data in the GWR-GA method.

Method	First dataset			Second dataset		
	RMSE(cm)	R^2	Time(s)	RMSE(cm)	R^2	Time(s)
ANN	141.4	0.84	12.7	193.12	0.79	9.8
SVR	142.6	0.73	0.36	152.4	0.68	0.27
Linear Regression	171.2	0.6	6.32	184	0.54	1.81
GWR	2.6	0.99	56	15.08	0.98	43
GWR-GA	1.76	0.99	41	12.4	0.98	38

Table 3. Implementation results of ANN, SVR, Linear Regression, GWR and GWR-GA methods for the online phase of indoor positioning

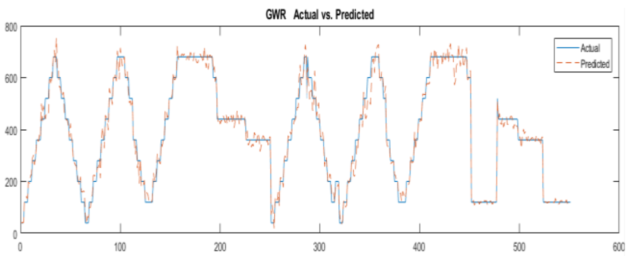

Figure 4. Compare actual data with predicted data in GWR-GA model

According to Table 3, the results show that the GWR method for fingerprinting network learning has less error and more R^2 coefficient than ANN, SVR and Linear Regression, but the time

complexity of this method is longer than other methods. On the other hand, the genetic algorithm can identify effective access points for fingerprint network learning, so according to Table 3, improved the error and time complexity of GWR method.
In this study, two datasets with different intervals of checkpoints were used which the results show if the distance between checkpoints is reduced fingerprint network learning will be done with less error.

4. CONCLUSION

Because of the high frequency utilization of the Wi-Fi signal in Indoor space, when using this system, we may deal with noise distortion error at Indoor Positioning, which has been reduced to some extent by considering the time parameter and the signal measurement signal. In general, this research has achieved the following results:
- To determine indoor positioning by fingerprint method, GWR method has better accuracy than neural network methods. But there is more time complexity.
- Combining Genetic Algorithm and GWR reduces learning error.
- If the distance between checkpoints is increased, the learning of the fingerprint network will occur with greater error.
-The high number of access points does not increase the accuracy of position determination, but depends on the structure of their placement in the environment.

The GWR method has been used for accurate spatial and nonlinear spatial mapping, which has increased the accuracy of Indoor Positioning. The innovation of this research has been the use of GWR algorithm as well as the improvement of this method by using genetic algorithm. Suggestions for this research are as follows:
- Using the GWRT algorithm, which takes into account the time, in addition to the location,
- Using and comparing other optimization algorithms.
- Enlarging the study area and testing it with data at different distances.
- Evaluating this research on other hardware and operating systems.
- Using methods other than the fingerprint method. Such as TDOA, AOA, DOA, etc.
- Using noise removal filters.

REFERENCES

Bajpai, P., Kumar, M., 2010. Genetic algorithm–an approach to solve global optimization problems. Indian Journal of computer science and engineering 1, 199–206.

Bivand, R., Yu, D., Nakaya, T., Garcia-Lopez, M.-A., Bivand, M.R., 2017. Package 'spgwr.' R software package.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician) 47, 431–443.

Charlton, M., Fotheringham, A.S., 2009. Geographically weighted regression. White paper, National Centre for Geocomputation National University of Ireland Maynooth.

Chen, C.-S., 2012. Artificial neural network for location estimation in wireless communication systems. Sensors 12, 2798–2817.

Delavar, M.R., Gholami, A., Shiran, G.R., Rashidi, Y., Nakhaeizadeh, G.R., Fedra, K., Hatefi Afshar, S., 2019. A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran. ISPRS International Journal of Geo-Information 8, 99.

Delavar, M.R., Samadzadegan, F., Pahlavani, P., 2004. A GIS-Assisted optimal urban route finding approach based on genetic algorithms. International archives of photogrammetry remote sensing and spatial information sciences, 35(Part 2), pp.305-308.

Guo, D., Zhang, Y., Xiang, Q., Li, Z., 2014. Improved radio frequency identification indoor localization method via radial basis function neural network. Mathematical Problems in Engineering 2014.

Holland, J.H., 1992. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press.

Hwang, R.-C., Hsu, P.-T., Cheng, J., Chen, C.-Y., Chang, C.-Y., Huang, H.-C., 2011. The indoor positioning technique based on neural networks, in: 2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, pp. 1–4.

Jan, S.-S., Yeh, S.-J., Liu, Y.-W., 2015. Received signal strength database interpolation by Kriging for a Wi-Fi indoor positioning system. Sensors 15, 21377–21393.

Jekabsons, G., Zuravlyov, V., 2010. Refining Wi-Fi based indoor positioning, in: Proceedings of 4th International Scientific Conference Applied Information and Communication Technologies (AICT), Jelgava, Latvia. pp. 87–95.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., Kolehmainen, M., 2004. Methods for imputation of missing values in air quality data sets. Atmospheric Environment 38, 2895–2907.

Kehua, M., YaoDong, C., Xiao, M., 2011. An indoor positioning technology based on GA-BP Neural Network, in: 2011 6th International Conference on Computer Science & Education (ICCSE). IEEE, pp. 305–309.

Li, J., Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecological Informatics 6, 228–241. https://doi.org/10.1016/j.ecoinf.2010.12.003

Mitchell, M., 1998. An introduction to genetic algorithms. MIT press.

Nahr, S.T., Pahlavani, P., Hasanlou, M., 2014. Different optimal band selection of hyperspectral images using a continuous genetic algorithm. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(2), p.249.

Pahlavani, P., Gholami, A., Azimi, S., 2017. AN INDOOR POSITIONING TECHNIQUE BASED ON A FEED-FORWARD ARTIFICIAL NEURAL NETWORK USING LEVENBERG-MARQUARDT LEARNING METHODJ.

ISPRS. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42.

Pahlavani, P., Samadzadegan, F., Delavar, M.R., 2006, September. A GIS-based approach for urban multi-criteria quasi optimized route guidance by considering unspecified site satisfaction. In International Conference on Geographic Information Science (pp. 287-303). Springer, Berlin, Heidelberg.

Subedi, S., Gang, H.-S., Ko, N.Y., Hwang, S.-S., Pyun, J.-Y., 2019. Improving Indoor Fingerprinting Positioning With Affinity Propagation Clustering and Weighted Centroid Fingerprint. IEEE Access 7, 31738–31750.

Zou, H., Huang, B., Lu, X., Jiang, H., Xie, L., 2015. A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine. IEEE Transactions on Wireless Communications 15, 1252–1266.

Zou, H., Jin, M., Jiang, H., Xie, L., Spanos, C.J., 2017. WinIPS: WiFi-based non-intrusive indoor positioning system with online radio map construction and adaptation. IEEE Transactions on Wireless Communications 16, 8118–8130.