A Multi-classifier Approach for Fuzzy KNN Based WIFI Indoor Localization

Yuanfeng Du[1,2,3](✉) and Dongkai Yang[1,2,3]

[1] Beihang University, Beijing, China
yfdu1989@163.com
[2] Shandong Orientation Electronic Technology Co. Ltd., Jinan, China
[3] Postdoctoral Workstation of Jining Hi-Tech Industry Development Zone, Jining, China

Abstract. WIFI fingerprint positioning technology has been widely studied and developed for a long time and lots of experiment systems have been established. However, the time-varying and nonlinear features of the WIFI signal impede the development in the application level. The performance of existing positioning systems could be very unstable due to the signal varying. Our object is to combine the fuzzy technology and multi-classifier approach to improve the system accuracy and robustness. The proposed method adopts the fuzzy integral to fusion the results obtained from different fuzzy K-nearest neighbor (KNN) classifiers generated by adaBoost. Experiment results demonstrate that our approach improves the average positioning errors and their standard deviations by 21% and 26% separately compared to the traditional KNN algorithm.

Keywords: WIFI fingerprint · Fuzzy · Multi-classifier · AdaBoost

1 Introduction

As the development of information technology, the location based service (LBS) has become more and more important for people's daily lives. The positioning technology based on global navigation satellite system (GNSS) has achieved a great success in outdoor scenarios. However, the complicated environment and the multipath effect cause great difficulties in indoor positioning. Now the research in WIFI fingerprint positioning technology is very popular to promote the development of indoor applications.

Precedent works have considered many algorithms to deal with the WIFI signal fluctuation, such as the widely used KNN, Bayesian [1]. However, how to describe the signal distribution in the area and combat the interference effectively are still big problems. The deep learning algorithm is used in [2] to obtain better accuracy, with higher cost.

Some researchers tried to solve the problem by using the fuzzy technology, which has a strong non-linear characteristic and an anti-interference ability. An effective calculation of Euclidean distance is undertaken by means of fuzzy logic methods. The systems comprising a better understanding of the fuzzy inference technology could provide an improvement of accuracy [3]. Another solution proposed is the use of fuzzy

S. Han et al. (Eds.): AICON 2019, LNICST 287, pp. 393–401, 2019.
https://doi.org/10.1007/978-3-030-22971-9_34

rule-based classification, which divides the RSSI into three linguistic terms, low, medium and high [4–7]. Though the above mentioned fuzzy approaches have made some progresses, they only adopt fuzzy logic without applying detailed fuzzy measurement. On the other hand, some basic multi-classifier technologies, such as bagging [8] and Bayesian fusion [9], are adopted by some studies to improve the positioning system robustness, which provide good research directions.

In this paper, our object is to combine the fuzzy technology and multi-classifier approaches to improve the system accuracy and robustness. We adopt the improved fuzzy KNN algorithm as the basic classifier, which establishes fingerprint database with the average value, the upper quartile, the median and lower quartile of the signal sequence in each grid. Then, the adaBoost algorithm is proposed to obtain sub-classifiers. At last, positioning result is obtained by the fuzzy integral fusion approach. The traditional AP selection and K-means clustering method are also adopted to reduce the computational complexity.

The following parts of this paper are organized as follows: Sect. 2 presents the fuzzy KNN based fingerprint algorithm. Section 3 delineates our proposed system architecture, including the adaBoost algorithm and fuzzy integral fusion approach. Experiment procedures and simulation results in Sect. 4 demonstrates the efficiency of our approach. This is followed by the conclusion in Sect. 5.

2 Fuzzy KNN Based Fingerprinting Algorithm

The widely used KNN algorithm is very practical, and just the average value of the received signal strength (RSS) vector namely S needs to be computed. In the online stage, the real-time RSS vector M is compared with S through Euclidean distance to find the k nearest neighbor nodes. The final positioning result can be obtained as following. l_i is the location of the ith nearest node and l is the positioning result.

$$l = \frac{1}{k}\sum_{i}^{k} l_i \tag{1}$$

Some improved algorithms such as the weighted KNN have been proposed, but there is always only the average feature of each grid for comparison. In this paper, the Fuzzy KNN algorithm is proposed to provide several typical features for comparison. What's more, instead of the absolute weights based on the distances, the relative fuzzy membership degree for each grid is obtained.

During the offline phase, m samples are collected in c grids and the RSS sequence is $R = (\mathrm{R}_1, \mathrm{R}_2, \ldots, \mathrm{R}_c)$, $\mathrm{c} = 1, 2, 3\ldots, \mathrm{L}$. L is the number of grids. In grid i, the average value q_{i1}, the upper quartile q_{i2}, the median q_{i3} and lower quartile q_{i4} of the signal sequence $(\mathrm{R}_i : \mathrm{R}_{i1}, \mathrm{R}_{i2}, \ldots, \mathrm{R}_{im})$ are calculated and stored as the offline fingerprint database.

Assume the RSS vector R_i expands from small to large, the value of $(q_{i1}, q_{i2}, q_{i3}, q_{i3})$ changes as following.

$$\left(\frac{1}{m}\sum_{j=1}^{m} R_{i,j},\ R_{i,\left[\frac{(m+1)}{4}\right]},\ R_{i,\left[\frac{2(m+1)}{4}\right]},\ R_{i,\left[\frac{3(m+1)}{4}\right]}\right) \tag{2}$$

Where [ ] is the bracket function.

During the online phase, the similarity degree sequence $(t_{i1}, t_{i2}, t_{i3}, \ldots, t_{ic})$ $i = 1, 2, 3, 4$ sorted by average value, upper quartile, the median and the lower quartile are calculated based on Euclidean distance separately. Then, the final membership degree for each grid $D = (\mu_1, \mu_2, \ldots, \mu_c)$ can be obtained as follows. The gird with the largest membership degree will be considered as the positioning results.

$$\begin{aligned} \mu_j &= \sum_{i=1}^{4}\sum_{s=1}^{c} T_{sj}^{i} \qquad j = 1, 2, 3, \ldots, c \\ T_{sj}^{i} &= \begin{cases} \frac{1}{\sum_{q=1}^{K} q}(K - s + 1) & \mathit{if}\ t_{is} = j \\ 0 & \mathit{else} \end{cases} \end{aligned} \tag{3}$$

T_{sj}^{i} is the member of the jth grid of the ith element. K is an adjustable parameter, which means that K grids are chosen in each similarity degree sequence.

3 System Architecture

Based on the basic fuzzy KNN fingerprinting algorithm, the detailed system architecture is shown in Fig. 1. After collecting the fingerprint database, K-means clustering and AP selection are performed, both of which have been proven to be effective in previous papers. As this paper mainly discusses the fingerprint matching approaches, the simple and traditional K-means clustering and maximum selection methods are adopted to make the positioning system executable.

For each cluster, several Fuzzy KNN classifiers are trained based on AdaBoost algorithm. Then, the confusion matrix of each KNN classifier can be obtained by cross validation. After the classifier subset is selected according to our selection criteria, fuzzy measure is trained and stored.

During the online phase, a smoothing filter is adopted to reducing the signal fluctuation. After the cluster ID is decided, the corresponding sub KNN fuzzy classifiers are used to calculate the membership degree of each grid. At last, the positioning result is obtained by fuzzy Integral Fusion.

3.1 AdaBoost Algorithm for Positioning

AdaBoost algorithm trains the same type of basic classifiers based on different training set and combines them together to constitute a stronger classifier, which is the final strong classifier [10]. According to the theoretical proof, as long as each basic classifier's performance is better than random guessing, the error rate of the fusion

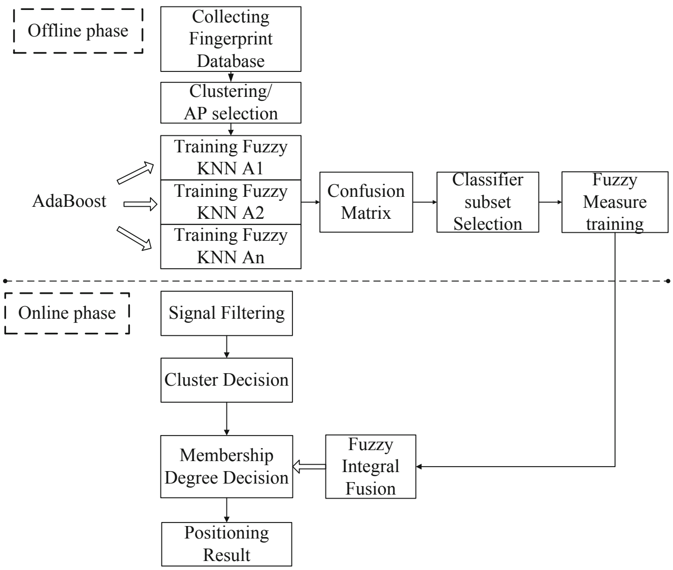

Fig. 1. Detailed system architecture

classifier would tend to zero. And our WIFI positioning classifiers proposed satisfy the assumption. The pseudo code of our method is presented in Fig. 2.

Input: Fingerprint database $R = (R_1, R_2, \ldots, R_c)$ $c = 1, 2, \ldots, L$
The expected classifier number $t_{\max}$
The initial weight of each fingerprint $W_1(i) = 1/(m \times L)$,
$i = 1, 2, \ldots, m \times L$

1. $t \leftarrow 0$
2. do $t \leftarrow t + 1$
3. Training the Fuzzy knn C_t based on R and $W_t(i)$
4. Calculate the confusion matrix $M^{(t)}$ *of* C_t
5. $E_t \leftarrow$ calculate the training error of c_t based on $M^{(t)}$
6. The weight update cofficient $a_t \leftarrow \frac{1}{2}\ln[(1 - E_t)/E_t]$
7. $W_{t+1}(i) \leftarrow \frac{W_t(i)}{Z_t} \times \begin{cases} e^{-a_t} & \text{if the positioning error} \leq 3m \\ e^{a_t} & \text{if the positioning error} > 3m \end{cases}$

Z_t is the normalization coefficient

8. until $t = t_{\max}$

Output: C_t and $M^{(t)}$ $k = 1, \ldots, t_{\max}$

Fig. 2. AdaBoost algorithm for positioning

During the loop, the fuzzy KNN should be trained based on R and $W_t(i)$. Firstly, the vector R_i is sorted from small to large according to the RSS value. Then, the average value $q_{i1}^{(t)}$ is calculated.

$$q_{i1}^{(t)} = \sum_{j=1}^{m} R_{i,j} \times W_t(j) \tag{4}$$

Consequently, the offline database of each sub fuzzy KNN is $(q_{i1}^{(t)}, R_{i,r_{0.25}}, R_{i,r_{0.5}}, R_{i,r_{0.75}})$ with r_φ calculated as followed.

$$\begin{aligned} r_\varphi &= P\left(r \,|\, \sum_{i=1}^{r-1} W_t(i) < \varphi \,\&\&r|\, \sum_{i=1}^{r} W_t(i) > \varphi \right) \\ \varphi &= 0.25,\ 0.5,\ 0.75 \end{aligned} \tag{5}$$

The confusion matrix in step 4 is constructed by the cross-validation of the fingerprint database. As there are L grids in the positioning system, the $M^{(t)}$ will be a $L * L$ matrix in which the entry $M_{i,j}^{(t)}$ presents the number of the instances collected in location l_i and assigned to location l_j by the fuzzy KNN classifier.

To meet the target of reducing the positioning error as much as possible, we adjust the fingerprint weight according to whether the error is greater than 3 meters or not.

3.2 Classifier Subset Selection

As proposed in [11], the improvement of multi-classifier approach depends largely on the characteristic that each classifier does not get involved in the same mistake in decision making. Therefore, it is essential to make the correlation of the sub classifiers smaller during the subset selection to provide better improvement of fusion classifier.

In this paper, the generalized diversity (GD) is adopted as the selection criteria [12]. The classifier subset with the largest GD value will be chosen.

$$GD = 1 - \frac{p(2)}{p(1)} \tag{6}$$

$$p(1) = \sum_{t=1}^{T} \frac{t}{T} p_t \quad p(2) = \sum_{t=1}^{T} \frac{t}{T}\frac{t-1}{T-1} p_t \tag{7}$$

$$p_t = \frac{\sum_{i=2}^{L-1} \sum_{j=i-1}^{i+1} M_{i,j}^{(t)}}{\sum_{i=2}^{L-1} \sum_{j=i}^{L} M_{i,j}^{(t)}} \tag{8}$$

p_t is the probability of the tth sub classifier which has the positioning result with error larger than 3 m. T is the number of the selected classifier subsets from $t_{\max}$.

3.3 Fuzzy Measure Training and Integral Fusion

As the decision error distribution of each classifier changes as grids vary, we calculate the fuzzy measure for each grid $g_\lambda^{(i)}$ based on the confusion matrix $M^{(t)}$.

$$g_i^j = \frac{M_{ii}^j}{\sum_{k=1}^{L} M_{ki}^j} \tag{9}$$

$$\lambda_i + 1 = \prod_{j=1}^{L} (1 + \lambda_i g_i^j) \tag{10}$$

Where g_i^j stands for the fuzzy measurement of C_j in grid l_i. Then, λ_i which is the parameter of $g_\lambda^{(i)}$ can be obtained. $Cl^{(j)}$ stands for the jth sub classifier. Consequently, $g_\lambda^{(i)}$ can be obtained and the fuzzy measure training process is completed.

$$\begin{aligned} &g_\lambda^{(i)}(A_j) = g_i^j + g_\lambda^{(i)}(A_{j-1}) + \lambda_i g_i^j g_\lambda^{(j)}(A_{j-1}) \\ &A_j = \{Cl^{(1)},\, Cl^{(2)}, \ldots, Cl^{(j)}\} \end{aligned} \tag{11}$$

For the integral fusion process, the membership degrees $(\mu_i^{(1)}, \mu_i^{(2)}, \ldots, \mu_i^{(T)})$ are decided by each fuzzy KNN in the online phase as demonstrated in Sect. 2. And the positioning result can be obtained based on the Choquet fuzzy integral fusion. $(c)\int$ stands for the Choquet integral value of each grid [13]. Therefore, the grid with the largest integral value will be considered as the positioning result.

$$(c)\int f(x)dg = \sum_{i=1}^{n} f(x)[g(A_i) - g(A_{i+1}))]\ f(x_1) \leq f(x_2) \leq \ldots \leq f(x_n) \tag{12}$$

$$Result = \arg\max_{i=1}^{L} (c) \int (\mu_i^{(1)},\, \mu_i^{(2)},\, \ldots,\, \mu_i^{(T)}) dg_\lambda^{(i)} \tag{13}$$

4 Experiment and Simulation Results

To evaluate the performance of our proposed method, an experiment is carried out on the 5th floor of LaiFuShi piazza in Shanghai with a surface of approximately 80 m by 50 m (see Fig. 3). There are more than 300 IEEE 802.11 WLAN APs in this scenario. During the experiment, we have collected 80 samples of fingerprints from 79 different locations (see dark squares in Fig. 3, respectively), of which 60 samples are used for training and the rest 20 are for testing. Each location is almost 3 m away from each other and MI 2A is used as the mobile terminal, with Android 4.1.1 system.

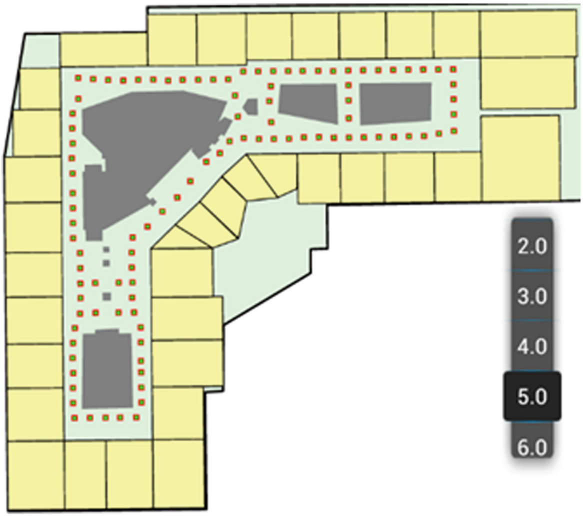

Fig. 3. Experiment scenario

Considering the compromise between system complexity and positioning performance, we adopt the number of selected sub classifier $T = 3$. The results in Table 1 show that only 4.76% of the test points have all the errors of classifiers larger than 3 m. The good diversity of the three classifiers implies high fusion performance.

Table 1. Diversity of the sub fuzzy KNN

Number of sub classifiers with error > 3 m	Zero	One	Two	Three
Number of points	1327	149	104	79
Percentage/%	79.99	8.98	6.27	4.76

The positioning accuracy is compared between the sub fuzzy KNN classifier and the proposed fuzzy integral fusion classifier in the experiment. As shown in Fig. 4, the percentage of test points without error improves from 59.9% to 64.2% and the largest error reduced from 40 m to 32 m, which proves the enhancement of the proposed fusion classifier over all the three sub fuzzy KNN classifiers.

We present some popular positioning matching algorithms for comparison, such as the traditional KNN method, Gauss Bayesian probability algorithm and Kernel-based algorithm. Furthermore, the Voting and Bayesian fusion methods are also evaluated.

As the test and training points are collected in the same locations, many positioning errors are zero and the average error of all the positioning methods are less than 3 m. The evaluation results shown in Table 2 reveal that all the fusion methods, including

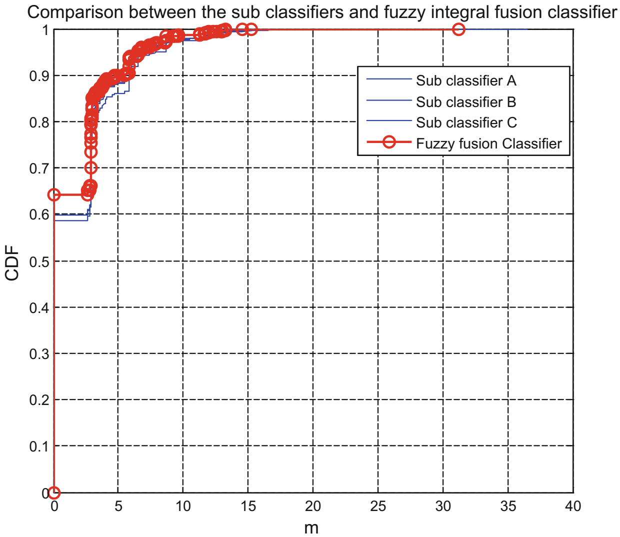

Fig. 4. Comparison of the sub classifier and fuzzy integral fusion classifier

the Voting, Bayesian Fusion and the proposed approach, outperforms the single classifier in terms of the average error distances and their standard deviations. Furthermore, the proposed fuzzy integral fusion approach obtains additional performance improvement.

Table 2. Summary positioning result (meter)

Positioning method	Average error	Standard deviation	90% error
Traditional KNN	1.9725	3.421	6.2
Gauss	2.5083	4.615	6.7
Kernel	1.9222	3.353	5.8
Voting	1.7028	2.81	5.8
Bayesian fusion	1.6809	2.794	5.0
Proposed approach	1.5591	2.598	4.8

5 Conclusion

In this paper, we have proposed a novel multi-classifier approach for fuzzy KNN based WIFI indoor location system. We introduce a new fuzzy KNN classifier based on membership degree. Subsequently, we propose the AdaBoost algorithm together with fuzzy integrate fusion approach in the field of indoor positioning. Practical experiment shows good diversity of selected sub classifiers and the effectiveness of fusion classifier approaches. The proposed approach improves the average positioning errors and their standard deviations by 21% and 26% respectively compared to the traditional KNN algorithm.

References

1. Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**(6), 1067–1080 (2007)
2. Shao, W., Luo, H., Zhao, F., Ma, Y., Zhao, Z., Crivello, A.: Indoor positioning based on fingerprint-image and deep learning. Access IEEE **6**, 74699–74712 (2018)
3. Teuber, A., Eissfeller, B.: WLAN indoor positioning based on Euclidean distances and fuzzy logic. In: Proceedings of the 3rd Workshop on Positioning, Navigation and Communication, Munich, pp. 159–168 (2006)
4. Garcia-Valverde, T., Garcia-Sola, A., Hagras, H.: An adaptive learning fuzzy logic system for indoor localization using Wi-Fi in ambient intelligence environment. In: IEEE World Congress on Computational Intelligence, pp. 25–32 (2012)
5. Li, Y., Liu, X.: Fuzzy logic and neural network based indoor fingerprint positioning algorithms in wifi. Int. J. Comput. Intell. Appl. **13**, 2282–2290 (2014)
6. Alonso, J.M., Ocaña, M., Sotelo, M.A., Bergasa, L.M., Magdalena, L.: WiFi localization system using fuzzy rule-based classification. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2009. LNCS, vol. 5717, pp. 383–390. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04772-5_50
7. Torteeka, P., Chundi, X.: Indoor positioning based on Wi-Fi fingerprint technique using fuzzy k-nearest neighbor. In: 11th International Bhurban Conference on Applied Sciences & Technology, pp. 461–465 (2014)
8. Trawinski, K., Alonso, J.M., Hernandez, N.: A multiclassifier approach for topology-based WiFi indoor localization. Soft Comput. **17**, 1817–1831 (2013)
9. Shin, J., Jung, S.H., Yoon, G., et al.: Electrical Engineering and Applied Computing, pp. 135–147. Springer, Dordrecht (2011). https://doi.org/10.1007/978-94-007-1192-1
10. Riccardi, A., Fernandez-Navarro, F., Carloni, S.: Cost-sensitive AdaBoost algorithm for ordinal regression based on extreme learning machine. IEEE Trans. Cybern. **44**, 1898–1909 (2014)
11. Kuncheve, L.I.: Combining pattern classifiers: methods and algorithms. Technometrics **47**, 517–518 (2005)
12. Partidge, D.: Network generalization differences quantified. Neural Netw. **9**, 263–271 (1996)
13. Yang, R., Wang, Z.Y., Heng, P.A., et al.: Fuzzy numbers and fuzzification of the Choquet integral. Fuzzy Sets Syst. **153**, 95–113 (2005)