Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-8, Issue-3, March 2026
DOI: 10.53469/jrse.2026.08(03).05

A Hierarchical Indoor Localization Method Based on Location Clustering and Intra-Class WKNN

Shi Dong, Shenglei Pei*, Lin Tan

School of Intelligent Science and Engineering, Qinghai Minzu University, Xining 810007, Qinghai, China
**Correspondence Author*

Abstract: ***Fingerprint localization methods based on Wireless Sensor Networks (Wi-Fi) have attracted extensive attention due to the widespread availability of infrastructure. However, traditional approaches are limited by significant signal fluctuations, complex and variable environmental conditions, and multipath effects, making it difficult to further improve localization accuracy. To address these challenges, this paper integrates the K-means clustering algorithm and proposes a region-aware improved Weighted K-Nearest Neighbors (WKNN) hierarchical localization algorithm, achieving a synergistic enhancement of both localization accuracy and time efficiency. Specifically, in the offline phase, the proposed method first clusters reference points based on their positional information at the location level, forming multiple spatially correlated sub-regions, and trains a dedicated WKNN model for each sub-region. In the online localization phase, coarse clustering is first applied to determine the region to which the target point belongs, after which fine-grained sample-level WKNN localization is performed only within that region. This two-stage processing mechanism effectively narrows the search space, reduces the probability of mismatches caused by similar RSSI vectors across different regions, and improves matching efficiency. Experimental results show that compared with the traditional WKNN algorithm, the proposed method reduces localization error by 4.48% and decreases real-time testing time overhead by 76.68%, which fully demonstrates the effectiveness of the proposed algorithm.***

Keywords: K-Means Clustering Algorithm, Hierarchical Localization, Wireless Sensor Network, Area Perception, Improved Weighted K-Nearest Neighbor Algorithm.

1. Introduction

With the continuous development of wireless communication technology, the demand for location-based services has shown a significant growth trend [1]. However, in complex indoor scenarios, due to interference from obstacles such as buildings, the positioning performance of the Global Positioning System (GPS) is often unsatisfactory, and even ineffective indoor positioning can not be achieved [2] [3]. In view of this, in order to provide stable, reliable and accurate indoor positioning and navigation services, indoor positioning systems have made substantial progress in the past few decades.

Currently, there are many technologies that can be applied to indoor positioning, such as Wi-Fi [4], Bluetooth [5], radio frequency identification (RFID) [6], ultra-wideband (UWB), etc. Each technology has unique characteristics and advantages. Among them, Wi-Fi-based indoor positioning technology has attracted widespread attention from researchers due to its significant advantages such as convenient deployment, low hardware cost, wide signal coverage and strong penetration ability. Wi-Fi-based indoor positioning technology can be divided into positioning methods based on channel state information (CSI) and positioning methods based on received signal strength indicator (RSSI) according to the data type used [7] [8]. Related studies show that Wi-Fi indoor positioning can achieve high positioning accuracy using CSI, but its equipment procurement cost is relatively high. In contrast, RSSI data can be easily obtained by users through mobile phones. In view of this, this paper selects Wi-Fi-based RSSI for indoor positioning research.

RSSI-based indoor positioning can be technically divided into two types: methods based on signal propagation models and methods based on fingerprint recognition. Methods based on signal propagation models utilize the propagation characteristics of wireless signals in indoor environments. By measuring the signal strength received at the receiver and combining it with a pre-established signal propagation model, the distance between the transmitter and receiver is estimated, thus achieving positioning. The drawback is that positioning accuracy is greatly affected by the environment, and the model parameters are difficult to determine accurately. Fingerprint-based RSSI indoor positioning utilizes the mapping relationship between wireless signal characteristics and location to achieve positioning. It does not rely on a precise signal propagation model but instead collects and stores signal characteristic data from different locations in advance. During actual positioning, the real-time measured signal characteristics are matched with data in a fingerprint database to determine the target location. This method offers higher positioning accuracy and stronger adaptability to environmental changes.

Currently, fingerprint-based indoor positioning methods are mainly divided into two categories: deep learning-based indoor fingerprint positioning and traditional indoor fingerprint matching [9]. Traditional fingerprint positioning methods include K-nearest neighbors (KNN) [10] [11], weighted K-nearest neighbors and support vector machine (SVM) [12] [13] [14]. K-nearest neighbors are simple to implement and do not require complex training, but they have the defects of strong sensitivity to the selection of K value and the "curse of dimensionality" of high-dimensional data. Although weighted K-nearest neighbors improve the robustness of positioning and the environmental adaptability through distance weighting, they are still sensitive to outliers.

Therefore, this study proposes a region-aware indoor positioning algorithm. Unlike traditional algorithms that directly search for the optimal solution, this study hierarchically divides the target area into sub-regions through location-level clustering before positioning, and trains a

WKNN for each sub-region. During testing, the algorithm first matches the region where the test sample is located, and then positions the target area. The algorithm can better adapt to local environmental features, thereby reducing the error accumulation of the global model due to environmental complexity. At the same time, signal feature clustering makes the region division more reasonable and can capture the differences in signal distribution in different regions.

2. Related Work

In recent years, researchers have been actively studying various technologies for indoor positioning systems, focusing on improving accuracy and robustness. Yuangbo Wang et al. [15] addressed the problem that most weighted K-nearest neighbor (WKNN) algorithms use the Euclidean distance of received signal strength (RSS) as the distance metric for fingerprint matching, but the RSS Euclidean distance is not consistent with the actual location distance. They analyzed the relationship between RSS similarity and location distance and proposed a new WKNN algorithm based on signal similarity and spatial location [16]. Rong Zhou considered the impact of signal fluctuations on positioning accuracy and integrated the empirical mode decomposition threshold smoothing method (EMDT) with the improved weighted K-nearest neighbor (WKNN) algorithm, naming it EMDT-WKNN [17]. Bai used the K-means (K-Means) algorithm to cluster reference points (RPs), thereby automatically dividing the entire target area into multiple regions. They used the mean of RSS measurements to characterize each reference point and divided the regions accordingly [18]. Dongdeok Kim 's KNN-based method relies on distance metrics to select nearest neighbors, but traditional metrics often fail to capture the complexity of indoor environments and have limitations in identifying nonlinear relationships.

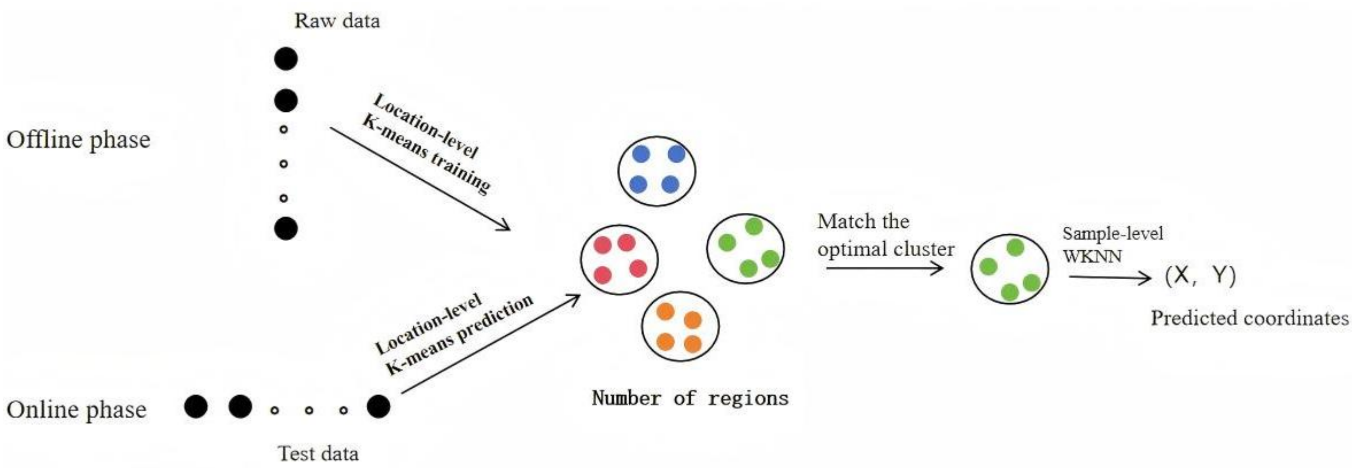

Figure 1: System Overall Structure Diagram

To address the limitations, a novel Wi-Fi fingerprint positioning method based on WKNN was proposed, which incorporates distance metric learning [19]. Baofeng Wang considered that different wireless access points (APs) have different states and that each AP contributes differently to RSS similarity. Based on RSS differences, a priority weight was designed to further optimize the ability of fingerprint distance to distinguish physical distance. The above strategy was integrated into the classic weighted K-nearest neighbor algorithm to propose a new indoor fingerprint positioning technology [20]. Hamada Rizk fused Received Signal Strength Indication (RSSI) and Round Trip Time (RTT) measurements to propose a hybrid deep learning indoor positioning system that integrates fingerprint recognition and time measurement technologies—RRLoc.

3. The Proposed Method

This study employs a hierarchical positioning design to improve positioning accuracy and efficiency. In the offline phase, the target area is first divided into several sub-regions through location-level clustering, and then a dedicated WKNN positioning model is trained on the sample-level data within each sub-region. In the online positioning phase, newly arrived test samples are first identified to their respective sub-regions through region matching, and then the corresponding WKNN model for that region is invoked for accurate location estimation. The specific structure is shown in Figure 1.

3.1 Feature Clustering

Clustering algorithms are unsupervised learning methods that automatically group data into multiple clusters by calculating the similarity (or distance) between samples, resulting in higher similarity among samples within the same cluster and lower similarity among samples from different clusters. Each cluster represents a potential data category or pattern. Typical clustering methods include partition-based K-Means, density-based DBSCAN, and hierarchical clustering, which are widely used in anomaly detection, image compression, and indoor positioning. K-Means, due to its high computational efficiency and simple implementation, is often used in conjunction with WKNN and is the choice for this paper.

Means is a classic partition-based unsupervised clustering algorithm. Its core objective is to iteratively optimize the data to divide it into K clusters, such that samples within a cluster have high similarity (close proximity) and samples between clusters have low similarity (far proximity). This process can be divided into two steps:

Assigning samples to clusters: Assign each sample to the cluster containing the nearest cluster center (mean point).

The mean point of each cluster is recalculated and used as the new cluster center.

Through continuous iteration, the algorithm gradually converges to a local optimum.

Suppose that n RPs collect signals from m routers in space. Then, based on the location-level sample set $X = \{x_1, x_2, x_3, \ldots, x_n\}$, where x_n represents the signal collected from the m routers at the n location point. Next, the cluster number K is set, and iterative sample allocation and cluster center $\mu_i^{t+1} = \frac{1}{|C_i^t|}\sum_{x \in C_i^t} x$ re performed. This ultimately results in cluster partitioning $C = \{C_1, C_2, C_{3,\ldots,} C_k\}$, where C_i is the sample set of the i cluster, and $\cup_{i=1}^{k} C_i = X$.

In K-means-based clustering algorithms, minimizing the sum of squares within a cluster is typically used as the objective function. Minimizing the sum of squares within a cluster can be expressed as:

$$J(\quad) = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2$$

$\|x - \mu_i\|^2$ It is the squared Euclidean distance from the sample x to the cluster center μ_i.

are collected from each RP m in space n, and location-level clustering is performed. By dividing the signal distribution into logical regions, computational complexity is effectively reduced (e.g., narrowing the online matching range) and positioning errors caused by cross-regional signal interference are minimized.

3.2 Weighted KNN

In indoor positioning scenarios, while traditional KNN algorithms classify or regress by selecting the K nearest samples to the target point, their "equal treatment" voting mechanism for neighboring samples has significant shortcomings. Samples farther from the target point may have larger deviations due to noise or signal attenuation, yet they still have the same weight as nearest neighbors, making the positioning results susceptible to interference and resulting in a high misclassification rate in boundary areas. To address this issue, Weighted KNN (WKNN) improves upon traditional KNN by introducing a distance-weighted mechanism. Its core idea is to dynamically allocate weights based on the spatial distance between samples and the target point, so that samples closer to the target point contribute more to the prediction results. This strategy not only enhances the representativeness of local data distribution but also effectively suppresses the influence of distant noisy samples, thereby significantly improving positioning accuracy and robustness in indoor environments with non-uniform signal distribution.

Its working principle can be divided into three steps. First, given test data x, find the nearest K samples (neighborhoods) from the training set, denoted as $N_k(x) = \{x_i, x_2, x_3, \ldots x_K\}$, and use Euclidean distance as the distance metric. Next, calculate the distance weights. For each neighborhood sample, calculate its x_i distance d to the target point x, and convert the distance into weights using a weighting function w_i. The weighting function uses inverse distance weighting, expressed as:

$$w_i = \frac{1}{d(x, x_i) + \varepsilon}$$

This ε is a local minimum value, intended to prevent the denominator from being zero. Finally, a weighted average of the neighboring sample labels is calculated as the prediction result.

3.3 Three-level Positioning

The hierarchical approach of clustering before localization stems from a systematic optimization of traditional single-location methods, which suffer from low computational efficiency and poor adaptability in complex indoor environments. Its core logic involves adaptively dividing the indoor space into relatively consistent local regions using clustering algorithms, forming a two-stage framework of "coarse localization (regional judgment) + fine localization (local optimization)." In the clustering stage, unsupervised learning automatically discovers environmental structures, rapidly narrowing the search range and isolating the influence of anomalous samples. In the fine localization stage, a dedicated model is used within the target region to improve accuracy. This hierarchical strategy not only reduces the global O(n) computational complexity to a local O(k) (where k is the number of samples in the region), but also enables the model to specifically capture local signal patterns by decoupling environmental heterogeneity.

Assuming that for n each location point (RP), m the RSSI generated by each access point (AP) is analyzed k using a set of RSSI values, and these values are used as group samples for each location point k, then the proposed method establishes a hierarchical localization system. In terms of region formation, a location-level clustering algorithm is used to divide the overall space into different regions based on the average features of the RSSI signal, replacing the overall features. Localization within these regions is performed using sample-level localization. This establishes a coarse-grained and fine-grained hierarchical localization system.

4. Experimental Results and Analysis

4.1 Experimental Scenario

To measure the performance of the proposed solution, this experiment was conducted in [location missing].

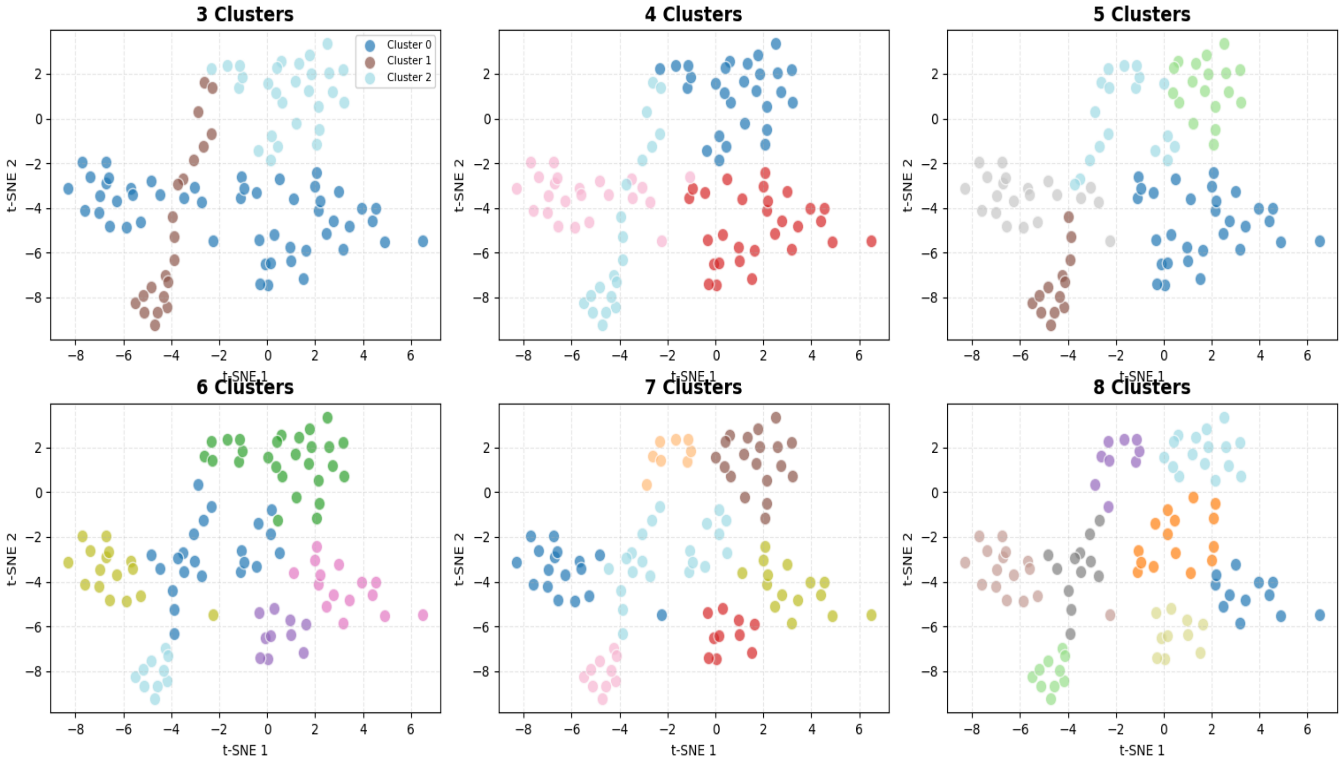

Figure 2: Location-based clustering effect

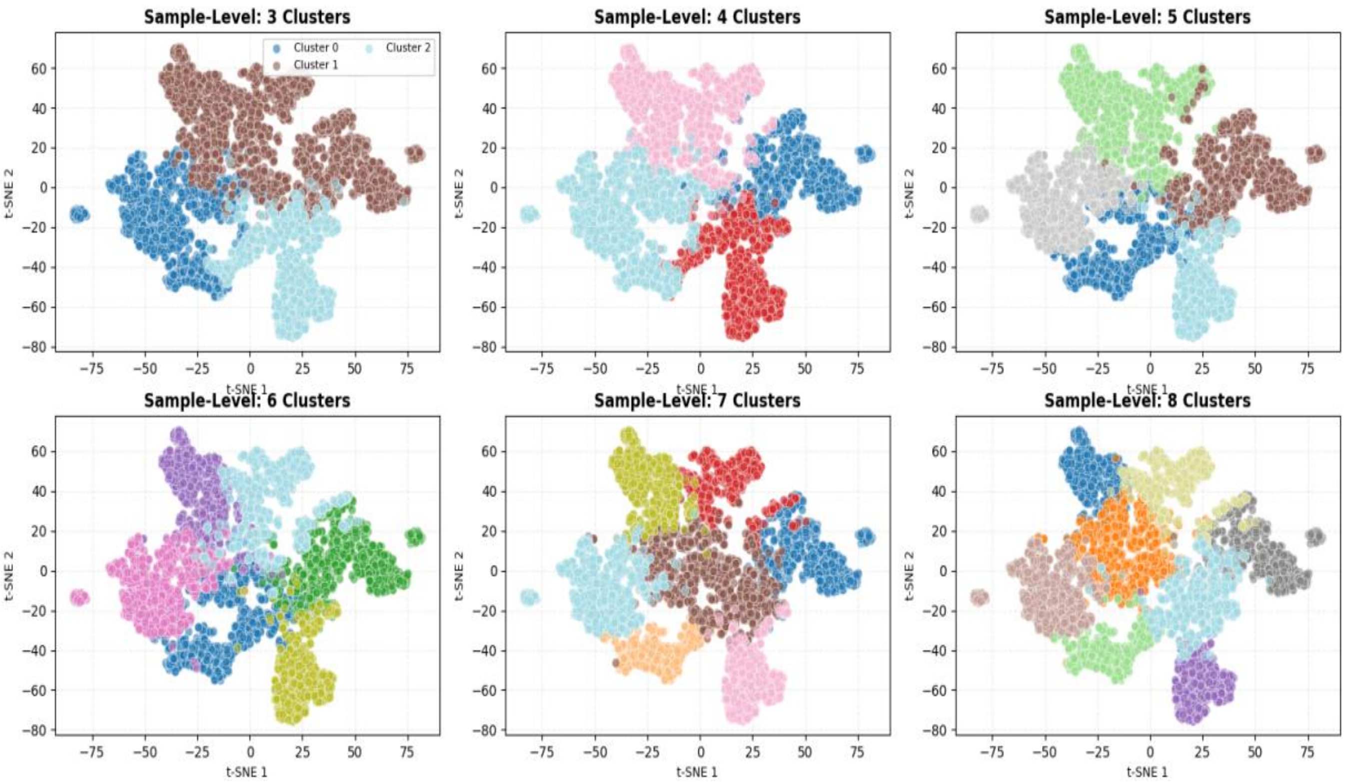

Figure 3: Results based on sample-level clustering

A 100 ×-meter simulation scenario was conducted, and a logarithmic path loss model was used for modeling. During offline training, 50 sets of RSSI data were collected from 100 evenly spaced locations across 6 routers for training. During testing, 20 locations were randomly selected in space, and the same 50 sets of data were collected for testing. Specific model parameters are shown in Table 1.

When evaluating the performance of the model, the mean absolute error (MAE) is used as one of the evaluation metrics.

Table 1: Simulation Experiment Parameter Settings

Parameter name	Parameter value
Simulation area	100 meters ×100 meters
Signal propagation model	Logarithmic path loss model
Number of routers	6
Distance between adjacent points in the training set	1 meter
Number of samples at the same location	50

$$MAE = \frac{1}{N}\sum\left|y_i - \hat{y}_i\right|$$

Where N is the number of data points, y_i is the i-th actual value, $\hat{y}_i$ and is the i-th predicted value.

4.2 Comparison of Sample-Level and Location-Level Clustering Results

Different clustering strategies affect the final localization performance, and the choice of clustering parameters directly determines the quality of clustering. Figure 2 shows K the clustering results of different values when using location-based clustering.

As shown in the figure, location-based clustering generally achieves better results, with clear boundaries between different categories. Figure 3, however, illustrates the performance of sample-based clustering. Since sample-based clustering treats all 50 data points collected at each location as independent samples, while this method preserves more of the original information about the location points, even when these data are divided into different clusters, the boundaries between the clusters are difficult to distinguish clearly. Therefore, sample-based clustering performs worse than location-based clustering.

4 3 Comparison of WKNN Results Based on Sample-level and Location-level

After completing the location-based clustering operation on the dataset, the matching clusters for each sample in the test set were first determined. Subsequently, a comparative analysis was conducted on the results obtained by the sample-based weighted K-nearest neighbor (WKNN) localization method and the location-based WKNN localization method within the cluster. At the same time, location-based and sample-based global WKNN without clustering were added for comparison.

Figure 4: Sample-level WKNN localization results with different cluster numbers

Figure 4 shows the average positioning error results for using sample-based WKNN within clusters and global sample-based WKNN without clustering. The optimal positioning error for sample-based WKNN within clusters is 1.344 meters, while the global WKNN positioning error is 1.407 meters, representing a 4.48% improvement over the traditional sample-based WKNN. This fully demonstrates the effectiveness of clustering.

Figure 5 shows the average localization error of using location-based WKNN within clusters and global location-based WKNN without clustering. The optimal localization error using location-based WKNN within clusters is 1.574 meters, while the global WKNN localization error is 1.516 meters, representing a 3.83% improvement compared to the traditional location-based WKNN. This fully demonstrates the effectiveness of the proposed method. It also shows that for traditional WKNN, sample-level weighting is superior to location-level weighting.

Figure 5: Location-level WKNN localization results for different cluster numbers

4.4 Comparison of Localization Results with Different K values

Different K values determine the number of clusters. We K compared the localization error when the value is 5 with that of the traditional WKNN, and the results are shown in Figure 6.

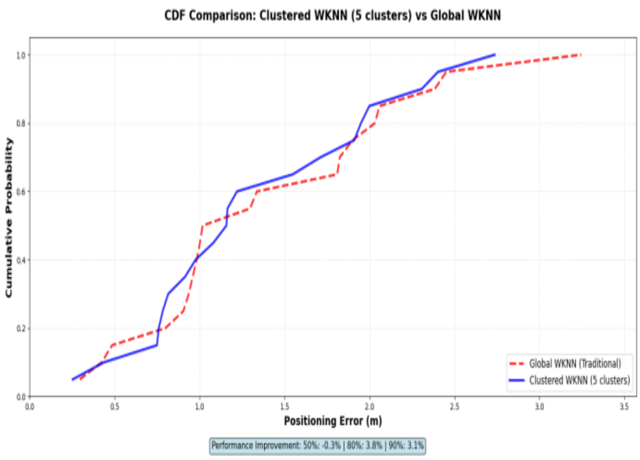

Figure 6: Comparison of optimal cluster partitioning and traditional WKNN

Compared to traditional methods, our method achieves a point error of less than 2 meters in over 80% of cases, which traditional methods cannot.

4.5 Comparison of Time Spending Results

Table 2 compares the time cost of the proposed method with that of the traditional sample-level WKNN.

Table 2: Comparison of Time Spending

	The method in this article	Traditional WKNN
Offline phase	0.0339 seconds	0.0018 seconds
Online phase	0.1658 seconds	0.7110 seconds

As shown in Table 2, although the offline training time of the proposed method is longer than that of the traditional WKNN, the real-time testing time is significantly shorter. Compared to the traditional WKNN, the proposed method achieves further improvements in both accuracy and time efficiency.

5. Conclusion

This paper proposes a hierarchical indoor positioning method based on location clustering and intra-cluster WKNN. This method constructs a two-level positioning framework by fusing location-level K-means clustering with sample-level improved weighted K-nearest neighbors (WKNN), thereby significantly improving time efficiency while maintaining positioning accuracy. Experiments show that compared with the traditional WKNN method, this method achieves further improvements in both positioning accuracy and time efficiency.

References

[1] Hoang MT, Yuen B, Dong X, et al. Recurrent neural networks for accurate RSSI indoor localization [J]. IEEE Internet of Things Journal, 2019, 6(6): 10639-10651.
[2] Van ĩ cek P, Beutler G, Kleusberg A, et al. DIPOP differential positioning package for the global positioning system [J]. 2023.
[3] Sun G, Chen J, Guo W, et al. Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs [J]. IEEE signal processing magazine, 2005, 22(4): 12-23.
[4] Maduranga M W P, Tilwari V, Abeysekera R. Improved RSSI indoor localization in IoT systems with machine learning algorithms [J]. Signals, 2023, 4(4): 651-668.
[5] Janczak D, Walendziuk W, Sadowski M, et al. Indoor positioning based on Bluetooth RSSI mean estimator for the evacuation supervision system application [J]. IEEE Access, 2024, 12: 55843-55858.
[6] Zhang D, Yang L T, Chen M, et al. Real-time locating systems using active RFID for Internet of Things [J]. IEEE Systems journal, 2014, 10(3): 1226-1235.
[7] Zhang B, Sifaou H, Li G Y. CSI-fingerprinting indoor localization via attention-augmented residual convolutional neural network [J]. IEEE Transactions on Wireless Communications, 2023, 22(8): 5583-5597.
[8] Nkabiti K P, Chen Y. Application of solely self-attention mechanism in CSI-fingerprinting-based indoor localization [J]. Neural Computing and Applications, 2021, 33(15): 9185-9198.
[9] Abdullah A A, Mohammed N S, Khanzadi M, et al. In-depth Analysis on Machine Learning Approaches: Techniques, Applications, and Trends [J]. ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY, 2025, 13(1): 190-202.
[10] Zheng X, Cheng R, Wang Y. Rssi-knn: A rssi indoor localization approach with knn [C]//2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA). IEEE, 2023: 600-604.
[11] Nair A P, Manoj A K. Indoor Navigation System Using Wi-Fi RSSI [C]//2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT). IEEE, 2024, 1: 954-959.
[12] Hu J, Hu C. A WiFi indoor location tracking algorithm based on improved weighted K nearest neighbors and Kalman filter [J]. IEEE Access, 2023, 11: 32907-32918.
[13] Chen B, Ma J, Zhang L, et al. An improved weighted KNN fingerprint positioning algorithm [J]. Wireless Networks, 2024, 30(6): 6011-6022.
[14] Bi J, Zhao M, Yao G, et al. PSOSVRPos: WiFi indoor positioning using SVR optimized by PSO [J]. Expert Systems with Applications, 2023, 222: 119778.
[15] Wang B, Gan X, Liu X, et al. A novel weighted KNN algorithm based on RSS similarity and position distance for Wi-Fi fingerprint positioning [J]. IEEE Access, 2020, 8: 30591-30602.
[16] Zhou R, Yang Y, Chen P. An RSS transform—Based WKNN for indoor positioning [J]. Sensors, 2021, 21(17): 5685.
[17] Bai S, Wu T. Analysis of K-Means algorithm on fingerprint based indoor localization system [C]//2013 5th IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications. IEEE, 2013: 44-48.
[18] Park J H, Kim D, Suh Y J. WKNN-based Wi-Fi fingerprinting with deep distance metric learning via siamese triplet network for indoor positioning [J]. Electronics, 2024, 13(22): 4448.
[19] Wang B, Li Q, Liu J, et al. A WKNN Indoor Fingerprint Localization Technique Based on Improved Discrimination Capability of RSS Similarity [J]. Sensors, 2024, 24(14): 4586.
[20] Rizk H, Elmogy A, Yamaguchi H. A robust and accurate indoor localization using learning-based fusion of Wi-Fi RTT and RSSI [J]. Sensors, 2022, 22(7): 2700.