Postprint: Research on MD5-KNN-Based Wi-Fi Indoor Positioning Algorithm

Authors: Miao Yunlong, Lu Yanhui, Yin Feng, Yang Shouyi

Date: 2018-05-24T00:00:00+00:00

Abstract

To mitigate the impact of fingerprint data volume and Access Point (AP) quantity in RSSI fingerprint databases on the computational efficiency of the KNN algorithm, a Wi-Fi indoor positioning algorithm based on MD5-KNN is proposed to optimize RSSI fingerprint databases constructed for large-scale venues. During the offline phase, each fingerprint in the RSSI fingerprint database is converted into an MD5 sequence represented by 32 hexadecimal digits. In the online phase, the positioning time required by this algorithm is independent of the number of APs and does not increase linearly with the number of fingerprints, thereby reducing both positioning time and computational overhead. Furthermore, the algorithm adaptively determines an appropriate K value, effectively resolving the issue of manual K value configuration inherent in traditional RSSI-KNN algorithms. Experimental results demonstrate that the proposed algorithm significantly enhances both the positioning accuracy and efficiency of Wi-Fi-based indoor positioning technology.

Full Text

Research on Wi-Fi Indoor Positioning Algorithm Based on MD5-KNN

Miao Yunlong[1], Lu Yanhui[1], Yin Feng[2], Yang Shouyi[1] [1]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China [2]Shenzhen Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China

Abstract: To reduce the impact of fingerprint data volume and access point (AP) quantity in RSSI fingerprint databases on the computational efficiency of KNN algorithms, this paper proposes a Wi-Fi indoor positioning algorithm based on MD5-KNN that optimizes RSSI fingerprint databases constructed for large venues. During the offline phase, each fingerprint in the RSSI database is converted into a 32-bit hexadecimal MD5 sequence. During the online phase, the algorithm' s positioning time becomes independent of AP quantity and does not increase linearly with fingerprint quantity, thereby reducing positioning time and computational overhead. Simultaneously, the algorithm adaptively determines the optimal K value, effectively solving the problem of manual K value configuration required by traditional RSSI-KNN algorithms. Experimental results demonstrate that the proposed algorithm significantly improves both positioning accuracy and efficiency of Wi-Fi-based indoor positioning technology.

Keywords: wireless signal; indoor positioning; received signal strength; message digest algorithm; computational time complexity; K-nearest neighbor algorithm

0 Introduction

In recent years, with advances in wireless communication technology and improved performance of mobile terminal devices, user demand for location-based services (LBS) has grown substantially. Global positioning systems based on satellite signals—including the American GPS system, Russian GLONASS system, Chinese Beidou system, and European Galileo system—have satisfied outdoor positioning requirements with continuously improving accuracy. However, due to complex indoor environments and the inability of satellite signals to penetrate building walls, there remains a lack of mature, widely-adopted indoor positioning technologies with high precision. Consequently, research on indoor positioning has become a major focus.

Existing indoor positioning technologies primarily include Bluetooth, infrared, ultra-wideband (UWB), RFID, ZigBee, ultrasonic, and Wi-Fi. Among these, Bluetooth devices are compact but suffer from poor stability in complex environments and short transmission distances. Infrared technology requires line-of-sight between detectors and targets, which is impractical in complex indoor settings. UWB technology requires additional blind nodes and exhibits high power consumption. While RFID technology offers high positioning accuracy, its anti-interference capability is weak. Furthermore, except for Wi-Fi, other indoor positioning technologies lack well-established infrastructure in existing public spaces, and constructing such infrastructure requires substantial time and economic investment. With the development of network communication technology and increasing demand for higher wireless speeds, Wi-Fi access points (APs) have become widely deployed, making Wi-Fi signals available almost everywhere within buildings. This has made Wi-Fi-based indoor positioning technology increasingly advantageous and a significant research area.

Wi-Fi indoor positioning algorithms are mainly divided into two categories: those based on received signal strength (RSS) and those based on ranging models. Ranging model-based algorithms require prior estimation of indoor channel environments and channel modeling. However, due to the complex and variable nature of indoor environments, ranging model-based methods cannot adapt to new environments. In contrast, RSSI fingerprint-based indoor positioning algorithms offer outstanding advantages in adapting to complex indoor environments. This algorithm involves two phases: offline training (database generation) and online positioning. During the offline phase, RSSI values from surrounding APs are collected at reference points along with their corresponding positions to form fingerprint information, which is used to construct an RSSI fingerprint database. During the online positioning phase, the RSSI fingerprint information received at the target location is measured and compared with fingerprints in the database to find matching entries and estimate the user' s current position. Fingerprint-based positioning algorithms mainly include K-nearest neighbor (KNN) and weighted K-nearest neighbor (WKNN). These algorithms have lower computational complexity, faster operation, and easier implementation compared to other Wi-Fi-based positioning methods. However, with the proliferation of Wi-Fi technology, the number of Wi-Fi hotspots in large venues can reach hundreds or even thousands, and the number of positioning reference points also increases. The growing fingerprint and hotspot quantities result in increasingly large RSSI fingerprint databases constructed during the offline phase, significantly impacting the positioning efficiency of KNN and WKNN algorithms. Additionally, since WKNN algorithms cannot adaptively obtain effective K values for real-time positioning points and require manual configuration, positioning accuracy cannot be guaranteed. Therefore, an urgent solution is needed to shorten positioning time, ensure positioning efficiency when fingerprint databases become large, and adaptively determine K values to improve positioning accuracy.

To address these issues, this paper proposes an improved Wi-Fi indoor positioning algorithm based on MD5-KNN. This algorithm effectively solves the problem of reduced positioning efficiency caused by increasing fingerprint quantities and hotspot numbers, reduces the correlation between computational time complexity and fingerprint database size, and improves positioning efficiency. Simultaneously, it can adaptively determine the K value, effectively enhancing positioning accuracy.

1 MD5-KNN Indoor Positioning Algorithm

1.1 RSSI-Based KNN Indoor Positioning Algorithm

The Wi-Fi fingerprint-based indoor positioning process consists of two phases: offline training and online positioning. The basic principle is illustrated in [Figure 1: see original paper]. During the offline training phase, a fingerprint database is constructed using RSSI fingerprint information from various reference points. During the online positioning phase, the target location is estimated through the KNN algorithm.

The RSSI fingerprint database model is shown in , where the fingerprint information at reference point RPj can be expressed as [(xj, yj), RSSIj1, ···, RSSIji,

···, RSSIjM]. Here, RSSIji represents the received signal strength from the ith AP at the jth reference point, and (xj, yj) represents the coordinates of reference point RPj.

Assuming a target point Pi with coordinates (xi, yi) receives RSSI fingerprint information from M APs as [RSSI' i1, RSSI' i2, ···, RSSI' iM], the distance between this fingerprint and the jth fingerprint in the database is:

$$dist = \left(\sum_{m=1}^{M} |RSSI'_{im} - RSSI_{jm}|^{u} \right)^{1/u} \tag{1}$$

where M represents the total number of APs in each fingerprint entry, and u denotes the distance type: when u=1, dist represents Manhattan distance; when u=2, dist represents Euclidean distance. This paper adopts Euclidean distance.

The KNN algorithm first calculates distances between the target point' s fingerprint and all fingerprints in the database, then selects K fingerprints with the smallest dist values as reference coordinates. Assuming the K matching coordinates obtained by KNN are (xk, yk) where k [1, K], the estimated coordinates of the target point in KNN are:

$$(x_i, y_i) = \left(\frac{1}{K} \sum_{k=1}^{K} x_k, \frac{1}{K} \sum_{k=1}^{K} y_k \right) \tag{2}$$

In WKNN, the estimated coordinates of the target point are:

$$(x_i, y_i) = \left(\sum_{k=1}^{K} W_k x_k, \sum_{k=1}^{K} W_k y_k \right) \tag{3}$$

where the weighting coefficient Wk is:

$$W_k = \frac{1/dist_k}{\sum_{k=1}^{K} 1/dist_k} \tag{4}$$

and distk is the Euclidean distance between the kth reference point and the target point. The positioning error between the actual and estimated positions is:

$$Loc_{err} = \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \tag{5}$$

In RSSI-KNN algorithms, the K value must be manually set before positioning. As shown in equations (2) and (5), positioning error is affected by the K value, and selecting the optimal K is crucial for improving accuracy. Manual K value configuration in each positioning process makes it difficult to obtain the optimal K, resulting in decreased positioning accuracy. Additionally, the K value selection process reduces positioning efficiency. The proposed MD5-KNN algorithm effectively avoids these problems by adaptively determining the optimal K value, as detailed in Section 1.2.

1.2 MD5-KNN Indoor Positioning Algorithm

As shown in equation (1), in traditional KNN and WKNN algorithms, when the number of APs in the fingerprint database (M) increases, the computational time complexity for calculating dist also increases. Similarly, when the fingerprint database grows (N increases), the complexity of obtaining K matching fingerprints increases, reducing positioning efficiency. To address this, we improve the RSSI-KNN indoor positioning algorithm, with the improved principle illustrated in [Figure 2: see original paper].

The improved algorithm adds a conversion component from RSSI fingerprint information to MD5 information on top of the original algorithm. The MD5 algorithm (Message-Digest Algorithm 5) is a widely used hash algorithm in computing that can compress arbitrary-length data into a 128-bit MD5 message. This algorithm is irreversible—user location information cannot be reverse-engineered from the MD5 sequence, effectively protecting user privacy. The pseudocode for converting fingerprint information to MD5 sequences is shown in Algorithm 1.

Algorithm 1: MD5-KNN Indoor Positioning Algorithm - Offline Training Phase

Initialize: RSSI threshold values Thmax, Thmin
Input: Reference point RSSI fingerprints
Output: MD5-based fingerprint database

```
For each RSSI in the fingerprint:
    If RSSI < Thmax and RSSI > Thmin:
        Mark this RSSI as 1
    Else:
        Mark this RSSI as 0
End for
For converted 01 fingerprint sequences:
    Generate MD5 sequence
End for
```

In Algorithm 1, Thmax is the maximum detectable RSSI value, and Thmin is the minimum detectable RSSI value. '0' indicates that the corresponding

AP was not detected or the signal was unstable, while ‘1’ indicates that the AP was detected with a stable signal. The 128-bit information is generated as a 32-bit hexadecimal sequence, converting each fingerprint entry into a 32-bit hexadecimal MD5 sequence. The structure of the constructed MD5-based fingerprint database is shown in .

During the online phase, the number of APs detectable by users during positioning is limited. To ensure consistency between the target fingerprint and database fingerprints, the target point’ s fingerprint data is preprocessed: the fingerprint is padded to length M, with undetected AP RSSI values represented as NULL. The online positioning phase pseudocode is shown in Algorithm 2.

Algorithm 2: MD5-KNN Indoor Positioning Algorithm - Online Positioning Phase

```
Initialize: K = 0
Input: MD5 fingerprint information  i of target point Pi
Output: Estimated position (xi, yi) of target point

For each MD5 fingerprint  j in the MD5 fingerprint database:
    If  i =  j:
        K = K + 1
End for
Estimate position using equation (2)
```

Unlike the RSSI-KNN algorithm in Section 1.1 that requires manual K value setting, the MD5-KNN algorithm automatically determines the optimal K value based on the principle that “nearby locations detect similar AP hotspots”[12]. By counting the number of MD5 fingerprints in the database that match the target point’ s MD5 sequence, the algorithm obtains the optimal K value, avoiding the manual configuration problem of RSSI-KNN.

2 Experimental Results and Analysis

To evaluate the performance of the proposed MD5-KNN algorithm, we selected the first floor of the core teaching building at Zhengzhou University as the experimental environment for RSSI fingerprint collection. The floor covers nearly 2000 m^2 and contains multiple classrooms, cement walls, and wooden doors. The building already has campus Wi-Fi deployed without requiring connection. A Meizu MX6 smartphone was used for RSSI information collection. To reduce the impact of signal instability during offline training, we collected 20 sets of data at each reference point, applied mean filtering, and set the RSSI detection interval to 0.8 seconds. The Wi-Fi fingerprint database constructed during the offline phase contains 3000 fingerprint entries, each including RSSI information from 200 APs. We selected 80 additional locations in the positioning area as target points and applied both RSSI-KNN and MD5-KNN algorithms for position estimation, with positioning performed every minute for 30 trials each.

2.1 Positioning Error Analysis

The resulting positioning error probability density functions are shown in [Figure 3: see original paper], where Figure 3(a) shows the distribution for RSSI-KNN and Figure 3(b) shows the distribution for MD5-KNN. The results indicate that RSSI-KNN positioning errors are relatively dispersed, primarily concentrated between 1-8 m, while the proposed MD5-KNN algorithm exhibits a narrower error range and improved positioning accuracy, primarily concentrated between 1-4 m.

The cumulative distribution functions (CDF) of positioning errors for both algorithms are shown in [Figure 4: see original paper]. The results demonstrate that RSSI-KNN achieves approximately 80% positioning accuracy within 2.5 m error, while MD5-KNN improves this to 90%. Moreover, the CDF of the proposed algorithm approaches 1 more rapidly, indicating superior positioning performance.

To further validate the impact of K value (number of neighbors) on positioning error for both algorithms, we varied the K value during positioning at a target point, with results shown in [Figure 5: see original paper]. The figure shows that RSSI-KNN is significantly affected by K value, with different K values yielding different positioning accuracies. The minimum positioning error of 3.1 m occurs at K=5, which represents the optimal K value. In contrast, MD5-KNN automatically determines the optimal number of neighbors (optimal K) based on MD5 sequence matching, achieving the lowest positioning error.

2.2 Algorithm Complexity Performance Analysis

The computational time complexity of the RSSI-KNN positioning algorithm is expressed as:

$$O_1 = k_1 \times M \times N \tag{6}$$

The computational time complexity of the proposed MD5-KNN positioning algorithm is:

$$O_2 = k_2 \times 32 \times N \tag{7}$$

where k and k represent single-operation times for RSSI-KNN and MD5-KNN respectively, with k > k ; N is the number of fingerprints in the database; and M is the total number of APs detectable during database construction. In MD5-KNN, each RSSI fingerprint is converted into a fixed 32-bit hexadecimal MD5 sequence, requiring only 32-character comparisons during matching. Therefore, MD5-KNN’ s computational time complexity (CTC) is affected only by fingerprint quantity, whereas RSSI-KNN is affected by both fingerprint quantity and AP quantity.

[Figure 6: see original paper] shows the computational time complexity versus fingerprint quantity for both algorithms when AP quantity is fixed. Figure 6(a) compares complexity with more than 32 APs, while Figure 6(b) compares complexity with fewer than 32 APs. The results show that when AP quantity exceeds 32, MD5-KNN requires significantly less time than RSSI-KNN for the same fingerprint quantity. When AP quantity is below 32, MD5-KNN is less efficient than RSSI-KNN due to the fixed 32-byte sequence conversion.

[Figure 7: see original paper] shows the computational time complexity versus AP quantity for both algorithms when fingerprint quantity is fixed. Since MD5-KNN’s complexity is independent of AP quantity while RSSI-KNN’s complexity has a linear relationship with AP quantity, and since large venues like shopping malls have far more than 32 APs, the feasibility and practicality of MD5-KNN are clearly demonstrated.

Actual positioning efficiency during simulation is shown in . When fingerprint quantity increases sixfold (from 500 to 3000), MD5-KNN positioning time increases by only 2.3 times, while RSSI-KNN positioning time increases by nearly 13 times. In large shopping malls, fingerprint databases contain far more than 3000 entries and AP quantities far exceed 32, giving the proposed MD5-KNN algorithm significant efficiency advantages.

3 Conclusion

This paper proposes an MD5-KNN-based indoor positioning algorithm that improves upon RSSI-KNN. Simulation results show that compared with RSSI-KNN, the proposed algorithm achieves significant improvements in positioning accuracy and efficiency, along with enhanced stability. In large venues such as shopping malls, where large areas and numerous hotspots result in increased fingerprint entries and AP quantities, the requirements for positioning accuracy and efficiency become more demanding, further demonstrating the superiority and practicality of the proposed algorithm. Additionally, due to the irreversibility of the MD5 algorithm, RSSI fingerprint information cannot be reverse-engineered from MD5 sequences, indirectly protecting user location privacy.

References

[1] Zhang Chong, Yu Hongyang. Indoor matching location algorithm based on WIFI signal strength [C]// Proc of the 10th International Conference on Communications, Circuits and Systems. Piscataway, NJ: IEEE Press, 2013.

[2] Chen Zhenghua, Zhu Qinchang, Soh Y C. Smartphone Inertial Sensor-Based Indoor Localization and Tracking With iBeacon Corrections [J]. IEEE Trans on

Industrial Informatics, 2016, 12(4): 1540-1549.

[3] Wang Yuan, Lin Jinguo. Discussion on several common indoor positioning techniques [J]. China Instrumentation, 2011(2): 54-57.

[4] Gunawan M, Li Binghao, Gallagher T, et al. A new method to generate and maintain a Wi-Fi fingerprinting database automatically by using RFID [C]// Proc of IPIN. Piscataway, NJ: IEEE Press, 2013: 1-6.

[5] Hou Yuying, Sum Guoyue, Fan Binwen. The indoor wireless location technology research based on WiFi [C]// Proc of the 10th International Conference on Natural Computation. Piscataway, NJ: IEEE Press, 2014.

[6] Zhu Hao, Gu Zonghai, Su Jin, et al. A RSSI location algorithm based on the centroid of point of intersection and its optimization [J]. Journal of Zhengzhou University: Engineering Science, 2010, 31(6): 43-46.

[7] Zhao Linsheng, Wang Hongpeng, Li Pengpeng, et al. An improved Wi-Fi indoor localization method combining channel state information and received signal strength [C]// Proc of the 37th Chinese Control Conference. Piscataway, NJ: IEEE Press, 2017: 8964-8969.

[8] Zegeye W K, Amsalu S B, Astatke Y, et al. Wi-Fi RSS fingerprinting indoor localization for mobile devices [C]// Proc of the 7th Ubiquitous Computing, Electronics & Mobile Communication Conference. Piscataway, NJ: IEEE, 2016: 1-6.

[9] Liu Hui, Darabi H, Banerjee P, et al. Survey of Wireless Indoor Positioning Techniques and Systems [J]. IEEE Trans on Systems Man & Cybernetics Part C, 2007, 37(6): 1067-1080.

[10] Ge Xingbin, Qu Zhiyi. Optimization WIFI indoor positioning KNN algorithm location-based fingerprint [C]// Proc of the 7th IEEE International Conference on Software Engineering and Service Science. Piscataway, NJ: IEEE, 2017: 135-137.

[11] Wang Lei, Zhou Hui, Jiang Guoping, et al. WiFi-based self-adaptive matching and preprocessing WKNN algorithm [J]. Signal Processing, 2015, 31(9): 1067-1074.

[12] Yao Zhifeng. Design and implementation of indoor localization system based on multi-information fusion by using WiFi and inertial sensors [D]. Guangzhou: South China University of Technology, 2016.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.