Wi-Fi fingerprint Localization Using RSSI-probability Radio Map

Yue Guo, Zhiqing Huang and Jun Lei

School of Software Engineering, Beijing University of Technology, Beijing, 100124, China Beijing Engineering Research Center for IoT Software and Systems, Beijing, 100124, China

Keywords: Wi-Fi Localization, RSSI Probability Distribution, Similarity Measurement, Bhattacharyya distance.

Abstract. Location-Based Service has moved into indoor scene by mature Wi-Fi network infrastructure and intelligent handheld terminal device development in decade years. Traditional indoor localization systems is designed by RSSI fingerprint mechanism to estimate user location. This paper indicated the limitation of pure RSSI by analysing the RSSI performance in experiment and proposed an RSSI-Pro transformation algorithm for radio map construction. To verify the desired accuracy an actual indoor localization system called LocNeedle is designed to implement the method of this article. LocNeedle adopts Bhattacharyya distance as similarity measurement metric to find the target user location, which can achieve that the 95% of localization error distance is under 2.5 meters while 70% is under 1 meter.

Introduction

Recently LBS (Location-Based Service) has attracted numerous enterprises, colleges and other creative fans to launch specific application for different areas [1], such as entertainment [2], medical treatment and intelligent traffic[3]. As the fundamental component, the accuracy of location information could be the most significant issue in a specific LBS system or application [4]. The survey [5] also indicates that besides of mature deployment of Wi-Fi network infrastructure, the development of intelligent handheld terminal drives the indoor localization and navigation research to become a fabulous focus point. At the meantime, the business value of indoor localization system also promotes the research progress. Wi-Fi fingerprint localization approach spends less special antenna configuration cost and network deployment work than geometric methods, such as ToA, TDoA, AoA [6, 7]. And it provides higher localization accuracy and less computational load than radio signal strength loss model approach [8] without access point (AP) location. Therefor it is widely used in implemented applications.

Most of Wi-Fi fingerprint localization mechanism adopt the Received Signal Strength Indicator (RSSI) as the variable to estimate the target position. As a generic radio signal technology measurement, RSSI presents the received signal power level in dBm form and can be computed by the receiver. This article indicated the limitation of pure RSSI by analyzing the RSSI performance in section 3.1 and proposed an RSSI-Probability transformation algorithm for radio map construction in section 3.3. This article adopts Bhattacharyya distance as similarity measurement metric to find the reference point (RP) which is nearest to current user location. To verify the desired accuracy an actual indoor localization system called LocNeedle is designed to implement the method of this article in Section 4.

Related Work

The Wi-Fi fingerprinting localization contains two phases, offline training and online operating. In offline phase the radio map can be established by collecting RSSI reference and in the online phase the matching algorithm is used to calculate the difference of RSSI sample and references in radio map to determine the target location.

The early platform based on the fingerprint localization mechanism is RADAR [9] which was implemented by Microsoft Research. It adopt RSSIs from multiple transmitters with different

directions to build radio map in the offline phase and used the Nearest Neighbor (NN) and the K-Nearest Neighbor (KNN) as matching algorithm in the online phase. The 50% estimation error of RADAR is under 2.5 meters and 90% is under 5.9 meters. Another system called Ekahau [10] solved target location estimation as machine learning problem by modeling the RSSI set with location tag attached and probability reckoning, which controlled the location accuracy in 3 meters. The 50% estimation error of Ekahau is under 1.45 meters and 90% is under 2.76 meters. LiFS [11] used multiple dimensional scaling algorithm to map high dimension RSSI vector into low dimension space and treated reference point plan with the identical process. The low dimension mapping result of these two process were paired by cosine formula to build the radio map, which is a novel approach in radio map construction. And as online matching algorithm the NN is adopt to estimate the target location with 90% of location accuracy under 5.88 meter.

The key points of Wi-Fi fingerprint localization could be focused on approaches in radio map construction and high efficiency matching algorithms in online phase. The radio map construction can be implemented by using pure RSSI distribution or pure RSSI with attached another indicator. However which type of variable is chosen the pure data collection task is still tough in real environment. The online matching algorithms are different for different types of radio map. The classic matching algorithm is KNN which is easily to implement.

Methodolgy

RSSI Limitation

The raw RSSI data is the fundamental source for signal analysis whose result can cause which type of similarity measurement to be adopt. The handheld terminals took data collection task and in LocNeedle the application record Wi-Fi RSSI by accessing Wi-Fi model to scan signal level of each detected APs. But the scanning interval of offline and online phase are different. It is significant that the radio map is built completely. Thus as the signal propagation model eqn (1) shown, the RSSI is measured power related by distance between transmitter and receiver, path loss exponent γ and interference noise X_{σ}

$$P_R|_{dBm} = P_R(d_0)|_{dBm} - 10\gamma \log_{10} \frac{d}{d_0} + X_\sigma|_{dB}$$
(1)

To verify the performance of RSSI value, we collected the RSSI data in 4 reference points to extract 48 samples of one AP. Actually in real environment the RSSI values are unstable because of shadowing, refraction, body interference or other actual path loss, which can be presented in Fig.1.



Fig. 1: Time varying RSSI

Thus the simultaneous RSSI value or a tiny time slot average value could not be fingerprints for online matching. We used histogram estimation approach to process the raw RSSI data. For instance, the RSSI probability distribution of AP (00:3a:98:67:e0:71) is shown in Fig.2. The probability distribution is just Gaussian-Like because of non-infinite sampling.

According to the RSSI probability distribution analysis of total samples, the RSSIs of single AP and single reference point are varying in an interval. The probability distributions of two neighbor reference point is distinguished, owing to the difference of AP deployed location. Fig.3 shows the example of two APs in two neighbor reference points.



AP with RSSIs

Fig. 3: Different Probability Distribution of single AP for neighbor RPs

We found that based on the probability performance which is the probability distributions of different AP RSSI is much stable than simultaneous RSSI value or averaged RSSI in a time slot. Furthermore the calculation of probability is based on the amount of samples which means that the probability distribution approximates the real distribution when more samples are collected. **Architecture**.

According to the generic workflow of Wi-Fi fingerprint localization, LocNeedle architecture includes two phases, which is shown in Fig.4, nevertheless the fingerprint process is improved.



Fig. 4: Architecture of LocNeedle

In Fig.4 the Pretreatment model of offline phase means that raw RSSI data from one reference point would be added the same class tag as the RSSI-Pro transformation input. The RSSI-Pro transformation means that the distinct RSSI of each AP in single reference point would be counted to figure out its appearance probability in raw RSSI data of this reference point. The radio map of LocNeedle is composed by the RSSI-Pro pairs with class tag. In online phase user terminal sends request for current location to server. The query contains raw RSSI samples in a time interval. The pretreatment model of online phase does the probability calculation of each RSSI value of detected APs. The output of pretreatment model inputs the NN matching process model to estimate target location.

RSSI-Pro Radio Map Construction Algorithm

The probability distribution of RSSI is used to build the radio map of LocNeedle. Each record of raw RSSI data (donated by R_{rrd}) was attached a class tag for identification before probability

calculation. Actually, there is a group of raw RSSI record with same class tag, which means these raw RSSI record belong to the same offline reference point. The form of class tag is (x, y, floor) where x and y are virtual mark not the real geographic coordinate. The example record of after offline pretreatment is shown in Table 1.

	Table 1. Example of prefeated RSSI feedful									
id	tag	BSSID	SSID	RSSI	GenTime					
853	(15,814,8)	00:3a:98:6 7:e0:71	bjut_teacher	-65	2014/11/5 20:18:58					

Table 1: Example of pretreated RSSI record

The preprocessed data is the input of RSSI-Pro transformation model and RSSI-Pro transformation procedure is illustrated as Fig. 5. The probability of AP i detected by RSSI s_j over the total L samples can be calculated by:

 $P_i(s_j) = \frac{N_i(s_j)}{L}$, where $N_i(s_j)$ is the number of samples for which RSSI of AP i is s_j and $s_i \in [s_{min}, s_{max}]$.

R _{prd} donates the set of pretreated raw data
R _{class} donates the set of class tag
R _{temp} donates the temporary RSSI record
R _{pro} donates the set of signal probability
R'pro donates the set of sample number of each RSSI
/*Initialization*/
Generate R _{class} by discriminating R _{prd}
/*RSSI-Pro Transformation*/
for $i \leftarrow 1$ to R_{class} . size() then do
set the current class tag K_i
find R _{temp} with K _i from R _{prd}
calculate the size of R_{temp} donated as N
/*RSSI counter*/
use hash map HM_{temp} to store subset of R_{temp} with different BSSID key
for key: HM _{temp} .keySet() then do
count the sample number for each RSSI
create object O _{rssi-pro} to record K _i , current BSSID, RSSI and amount
add $O_{rssi-pro}$ into R'_{pro}
end for
/*RSSI-Pro calculator*/
for $O_{rssi-pro}$: R'_{pro} then do
update the probability property of $O_{rssi-pro}$ with the result of amount over N
add $O_{rssi-pro}$ into R_{pro}
end for
end for

Fig.5: RSSI-Pro	Transformation	procedure
-----------------	----------------	-----------

The core of RSSI-Pro transformation is RSSI-Pro mapping which convert the RSSI with class tag to RSSI-Pro pair with same class tag. The workflow of RSSI-Pro transformation is shown as Fig.6. The RSSI-Pro transformation output the RSSI-Pro fingerprint record which are stored in database to build radio map.

$$\begin{array}{c} \mathsf{Raw} \; \mathsf{RSSI} \; \mathsf{Set} \\ \{R_{prd} | K_i, \mathrm{K}_i \in R_{class}\} & \longrightarrow \end{array} \begin{array}{c} \mathsf{RSSI-Rro} \\ \mathsf{Mapping} \end{array} \xrightarrow{\mathsf{RSSI-Rro}} \{R_{pro} | K_i, \mathrm{K}_i \in R_{class}\} \end{array}$$

Fig.6: RSSI-Pro transformation workflow

In the RSSI-Pro transformation all pretreated RSSI records are involved to generate RSSI-Pro radio map which is extended from traditional radio map with RSSI average value. The difference is figured out in Fig.7.

			Reference Point	AP_ID	RSSI	Probability
			RP1	AP ₁	RSSI1	Pro ₁
				AP ₁	$RSSI_2$	Pro ₂
				AP ₂	$RSSI_1$	Pro ₁
Reference Point	Multiple RSSI value Fingerprint			AP ₂	$RSSI_2$	Pro ₂
RP1	(RSSIm(AP1) RSSIm(AP1) RSSIm(AP1) ···· }			AP ₃	RSSI1	Prol
1411	(recording (r m 1), recording (r m 1), recording (r m 5), 3					
RP ₂	$\{RSSI_{avg}(AP_1), RSSI_{avg}(AP_2), RSSI_{avg}(AP_3), \cdots \}$		RP ₂	AP ₁	RSSI1	Pro ₁
RP3	$\{RSSI_{avg}(AP_1), RSSI_{avg}(AP_2), RSSI_{avg}(AP_3), \cdots \}$		_	AP ₁	RSSI2	Pro ₂
				AP ₂	RSSI1	Pro ₁
		·		AP ₂	$RSSI_2$	Pro ₂
				AP ₃	RSSI1	Pro ₁

Fig.7: Radio map Difference

NN Matching Algorithm

The RSSI measurement sequence in localization estimation query is received in online phase. The NN matching algorithm is used to calculate the distances between user location and each reference point and find the nearest neighbor reference point as the current user location. Generally, the Bhattacharyya distance [12] is known measurement of two probability distribution bound or overlap. In this case, the RSSI measurement sequence is calculated into the RSSI-Pro form which is processed by identical procedure of offline phase, represented as $\{M_i(s), i \in O_l^m\}$.

The set O_l^m keeps the m strongest detectable APs for each offline reference point l in user location. For each AP the correlation coefficient, a.k.a. Bhattacharyya coefficient, is B_i defined as $B_i = \sum_{s \in [s_{min}, s_{max}]} \sqrt{P_i(s) \cdot M_i(s)}$, which is valid when the s value is received in online query. Then the average Bhattacharyya distance between user location and each offline reference point d_l can be defined as:

$$d_{l} = \begin{cases} -\ln\frac{1}{m}\sum_{i\in O_{l}^{m}}\sum_{s\in[s_{\min},s_{\max}]}\sqrt{P_{i}^{l}(s)\cdot M_{i}(s)}, & \text{if }\sum_{s\in[s_{\min},s_{\max}]}\sqrt{P_{i}^{l}(s)\cdot M_{i}(s)} > 0 \\ -\infty & \text{, otherwise} \end{cases}$$
(2)

As the result of Bhattacharyya distance calculation, d_l is negative correlation with the probability distribution similarity, in other words, the reference point with the smallest Bhattacharyya distance is the most similar with current user location. Through the NN matching algorithm the final user location is estimated as the nearest reference point.



Fig.8: building layout

Result and Discussion Test bed Establishment The test bed of this paper is on one floor of teaching building with 22*63 m2 and 19 APs (Access Point). Five of these APs with both 2.4GHz and 5GHz are evenly deployed and other fourteen APs were deployed randomly in different room. A part of building layout is shown in Fig.8. The area covered by red point is test range of experiment.

In the experiment the reference points are distributed with 0.5m away from each other. For offline phase the application scanned signal level in 8 directions with 10 records of each direction at one reference point and then added unique class tag to each record. For online phase the request contained 10 samples in any user desired direction. The handheld terminals with Android 4.4.2 were used in raw RSSI data collection and for avoiding different sensor sensitivity of different vendor all raw RSSI data were collected by identical vendor handheld terminals.

Performance Comparison with other System

The performance metrics of indoor localization system are accuracy and precision which can be evaluated by the cumulative distributive function (CDF) of distance error. In this section LocNeedle implemented by the RSSI-Pro radio map is compared to two other systems, RADAR and LocNeedle implemented by RSSI average radio map. The CDF is the cumulative probability of each error distance calibration and Fig.9 shows the difference between CDF of each system. Intuitively, for LocNeedle implemented by RSSI-Pro radio map, 70% of error distance is under 1 m and the maximum distance error is 4.5 meter.



Fig.9: CDF of localization error for performance comparison

The distance error is contributed by fingerprints space limitation and random RSSI fluctuation in online phase. But the performance of LocNeedle is acceptable by compared to traditional system. Although the probability distribution method is much more stable than the average value method for location estimation and current localization accuracy of LocNeedle is acceptable, the computation complexity of LocNeedle is relatively higher than traditional system, especially when the radio map grows rapidly.

Conclusions and Further Work

In this paper the LocNeedle is implemented by two radio map generation algorithm. The first one is RSSI average value, which is improved classic radio map construction. By analyzing the performance of RSSI average value for localization effect, the RSSI fluctuation process is significant for both offline and online phase of localization system. Thus the second one, pair of RSSI and probability, is proposed to weaken the signal undulation. And for each radio map generation there is specific measurement method in similarity measurement field. The LocNeedle is designed to calculate Euclidean distance between user location and each reference point to find k-nearest neighbor reference point record. For the RSSI-Pro radio map the LocNeedle is designed to calculate Bhattacharyya distance to estimate the user location. Finally the localization effect of LocNeedle is compared to traditional localization system to verify the feasibility of proposed algorithm. In further work, RSSI probability distribution performance analysis is to find the bottleneck of incremental amount of samples to keep accuracy of location estimation but reduce computing cost. Furthermore we concentrate on a self-adapting mechanism for replying the radio

map extension and making LocNeedle more robust and desirable to achieve competitive location accuracy.

Acknowledgements

The research work was supported by the Beijing University of Technology Fundamental Research Funding Project (No.025000514314004). And the test environment was supported by Beijing Engineering Research Center for IoT Software and Systems.

References

- [1] Zhai, K., B. Jiang and W.K. Chan, Prioritizing Test Cases for Regression Testing of Location-Based Services: Metrics, Techniques, and Case Study. *IEEE Transactions on Services Computing*, 7(1): pp. 54-67, 2014.
- [2] Lanir, J., Kuflik, T., Dim, E., Wecker, A. J., & Stock, O., The Influence of a Location-Aware Mobile Guide on Museum Visitors' Behavior. *INTERACTING WITH COMPUTERS*, 25(6): pp. 443-460, 2013.
- [3] Dikaiakos, M. D., Florides, A., Nadeem, T., & Iftode, L., Location-aware services over vehicular ad-hoc networks using car-to-car communication. *IEEE Journal on Selected Areas in Communications*, 25(8): pp. 1590-1602, 2007.
- [4] Liu, Y., Yang, Z., Wang, X., & Jian, L., Location, Localization, and Localizability. *Journal of* Computer *Science and Technology*, 25(2): pp. 274-297, 2010.
- [5] Gu, Y.Y., A. Lo and I. Niemegeers, A Survey of Indoor Positioning Systems for Wireless Personal Networks. *IEEE Communications Surveys and Tutorials*, 11(1): pp. 13-32, 2009.
- [6] Shin, D.H. and T.K. Sung, Comparisons of error characteristics between TOA and TDOA positioning. *IEEE Transactions on Aerospace and Electronic Systems*, 38(1): pp. 307-311, 2002.
- [7] Bishop, A.N., et al., Exploiting geometry for improved hybrid AOA/TDOA-based localization. *Signal Processing*, 88(7): pp. 1775-1791, 2008.
- [8] Hatami A, Pahlavan K. Comparative Statistical Analysis of Indoor Positioning Using Empirical Data and Indoor Radio Channel Models. *In Proceedings of the 3rd IEEE Consumer Communications and Networking Conference*, 1018-102, 2006.
- [9] Bahl, P. and V.N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. in INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. 2000.
- [10]Roos, T., et al., A Probabilistic Approach to WLAN User Location Estimation. *International Journal of Wireless Information Networks*, 9(3): pp. 155-164, 2002.
- [11] Yang, Z., C.S. Wu and Y.H. Liu, Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention. MOBICOM 12: PROCEEDINGS OF THE 18TH ANNUAL INTERNATIONAL CONFERENCE ON MOBILE COMPUTING AND NETWORKING, pp. 269-280, 2012.
- [12]Cha S H. Comprehensive survey on distance/similarity measures between probability density functions. *City*, 1(2):pp.1, 2007.