International Journal of Scientific Engineering and Technology ISSN : 2277-1581
Volume No.4 Issue No5, pp: 335-337 01 May. 2015

A New Approach in Wireless Sensor Network Fingerprinting Localization by Affinity Propagation and K-Nearest Neighbor Method

Pejman abdollahzadeh karegar
Ramin shaghaghi kandovan
Email: pejmanabd1365@gmail.com

Abstract:* Due to the increasing development of wireless sensor network, localization using fingerprint has particular importance, Which includes database called Receive Signal Strength (RSS) vectors as the basis of indoor positioning, In this paper a clustering method called affinity propagation used for clustering datapoints, And then in the online phase, coarse localization and fine localization is utilized , in coarse localization an adaptive method is used with exemplars, eventually in fine localization the similarity of nearest neighbor and k-nearest neighbors is applied. Finally, the simulation results show the improvement of our theory compared with the methods of without clustering.

Keywords: localization using fingerprint, affinity propagation, coarse localization, fine positioning

I. Introduction

In recent decade, the indoor WLAN positioning technology has caught significant attention by a variety of universities and research institutes[1]. time of arrival (ToA), angle of arrival (AoA) and received signal strength (RSS) are the three most representative measurements for the position estimation. Compared to the ToA and AoA measurements, the RSS can be more easily measured without any additional special hardware devices in current open public WLAN networks. However, the most significant challenge of the RSS readings is about the irregular variations of RSS due to the variable radio channel attenuation, signal shadowing, multi-path interference and even the variations of indoor temperature.[2]

There are two phase, offline phase, and online phase. In the offline phase, we build a database of RSS fingerprint, from all APs, for our test bed. A number of signal strength samples, which is called the fingerprint, are collected for all APs Using portable devices. fingerprints can divide all of the datapoints to uniform units, each fingerprint includes RSS vectors from each cells of floor, and contains as a tuple <x, y, RSS_i> , where <x ,y> indicates the physical coordinate, RSS_{ij} indicates average RSS from AP[a] i in RP[b] j. these database is used to locate the user in the online phase.

Once the database of fingerprints exists, a device calculates position by recording a fingerprint and "matching" to the database. This usually consists of measuring a "distance" between the recorded fingerprint and each RP fingerprint in the database. We will refer to this distance as the "vector distance" which has units related to dBm (as opposed to "geometric distance" in meters between the TP and an RP).[10]

New accurate and scalable positioning algorithm is estimated to locate the user's position with low computation cost in a public WLAN environment. this algorithm consists of two steps: (1) the coarse positioning step is used to obtain the cluster which the user belongs to; and (2) the fine positioning step is utilized to calculate the accurate coordinates of the user. One effective solution for localization is the k-nearest neighbor (kNN) algorithm to estimate the mobile user's position at the centroid of the K closest neighbors. The closest neighbors are defined as the RPs which have the smallest RSS distance to the on-line new collected RSS readings.[3]

The paper is organized as follows. Sections 2 addresses the detailed steps of the off-line affinity propagation, in section 3 on-line cluster matching-based coarse positioning and fine positioning is investigated respectively. The performance of our proposed algorithm is verified in Section 4. Finally, Section 5 concludes this paper.

II. Clustering by affinity propagation

Clustering data based on a measure of similarity is a critical step in scientific data analysis and in engineering systems. A common approach is to use data to learn a set of centers such that the sum of squared errors between data points and their nearest centers is small. When the centers are selected from actual data points, they are called "exemplars." The popular k-centers clustering technique begins with an initial set of randomly selected exemplars and iteratively refines this set so as to decrease the sum of squared errors. k-centers clustering is quite sensitive to the initial selection of exemplars, so it is usually rerun many times with different initializations in an attempt to find a good solution.[1] We are going to investigate a different method and consider all datapoints in our database as a potential exemplar at the same time. We adopt a method that will send real-value messages throughout the network until a set of exemplars and corresponding clusters emerges. each similarity is set to a negative squared error (Euclidean distance) as:

$$s(i,k) = -\|\psi_i - \psi_k\|^2 \quad (1)$$

the value of each similarity between points build a matrix with L*L dimension. L is total number of Reference Points. There are

[a] Access Point

[b] Reference point

two kind of messages are recursively transmitted between RPs, the "Responsibility message" $r(i,k)$ sent from RP i to candidate exemplar RP k, describes the indicant fitness for RP k to be served as the exemplar for point i, taking into account other potential exemplars k' for RP i .
The "availability" a(i,k), transmitted from candidate cluster center point k to point i, reflects the indicant fitness for how well it would be for point i to select point k as its candidate cluster center, taking into account the support from other points that cluster center point k should be an exemplar for them, is define as:

$$r(i,k) = s(i,k) - \max\{a(i,k') + s(i,k')\} \qquad (2)$$

$$a(i,k) = \min\{0, r(k,k) + \sum_{i' \neq \{i,k\}} \max\{0, r(i',k)\}\} \qquad (3)$$

This algorithm also takes an input value $S(k,k)$ for each data point, So that the points that have higher value would be more likely to be an exemplar, these values are known as preference. Number of clusters is strongly influenced by the correct choice of the preference values, if we assume that all of the data points has the same probability as exemplar, preferences should be equal for each RPs. it will be achieved different clusters by changing this value. we have assumed in our research that the preferences are median of all similarity values.

III. fine and coarse localization in Online phase

In the online phase, we sampled the RSS value in our test points, and build an RSS measurement vector:

$$\psi_r = [\psi_{1,r}, \psi_{2,r}, \dots, \psi_{M,r}] \qquad (4)$$

where $\{\psi_{k,r}, k = 1, \dots, M\}$, In order to reduce the space into subsets each of our clusters, we utilize coarse localization, and we reduce the whole area of data point, by cluster matching.
It is defined similarity function as the negative of Euclidean distance of the online RSS vector ψ_r to the RSS vector of each exemplars as:

$$s(i,k) = -\|\psi_r - \psi_k\|^2 \qquad (5)$$

ψ_j is the RSS vector of kth exemplars. The clusters with the largest similarity values are chosen as the desired clusters.
Selecting wrong clusters at this stage is the main source of error as the maximum localization error. So all of these methods tend to reduce the probability of error and thus reduce the maximum positioning error. At the end, our database can be limited into subset of $|\widetilde{N}| = C$,that C is the number of datapoints in the selected group.

III. A Fine localization using K-nearest neighbor

If we assume online RSS vector as:
$\psi_r = [\psi_{1,r}, \psi_{2,r}, \dots, \psi_{M,r}]$, where $\{\psi_{k,r}, k = 1, \dots, M\}$, so that r is the number of test points, which in our study is 96 points. these RSS vectors from the online phase and RSS of each selected cluster members are compared as the Euclidean distance which can be obtained:

$$D = \sqrt{\sum_r \sum_j (\psi_r - \psi_j)^2} \qquad (6)$$

The reference point at which has a minimum distance with RSS vector of test point, has been selected as desired location. During sampling stage, some of Aps may be present in our database, but at another time, they may not be in our fingerprint, We will ignore this transient Aps, in this article. Another technique in deterministic location is considering the average distance to K nearest neighbors (KNN) in the signal space, which is called K-nearest neighbor method. simulation demonstrate that the positioning accuracy of this method is better than nearest neighbor method by using affinity propagation clustering. In this Method ,the results are sorted from nearest neighbors, then select first K location, then averaged corresponding coordinates. This method will average all K coordinates , and each coordinate is weighted equally.

IV.Experimental Results and Analysis

collecting data was done at the time interval of one week on different days and hours, The distances between each RPs have considered 1.5 and 2 m. And the sampling for each point was collected for 60 seconds and was measured at a sampling rate 0.5 ms. A total of 120 samples were considered for each point. A number of 99 RPs and 96 TPs are intended.
Sampling was performed by Net strumbler software, And also try to operate sampling during quiet days of first floor of engineering university of Shahr-e-Rey, that Human barriers influence less in our research results. Simulation results are shown in Figures 1, 2 and 3, Figure 1 displays the effect of changing the value of k on positioning mean error for clustered and non-clustered mode. Figures 2 and 3 are shown plots of cumulative distribution function in terms of errors for k-nearest neighbor and nearest neighbor methods. Ultimately, it will be a comprehensive comparison between these two methods that was performed according to these two diagrams.
As you can see in Figure 1, the mean error is displayed in terms of k. As shown in Figure 1, the best case for mean positioning error is k-nearest neighbor for clustering mode and by affinity propagation with k = 2, that corresponding value is 1.99 m, Which is improved 1.7 meters in comparison with the K-nearest neighbor without clustering. As is known, the impact of clustering algorithm is a completely Specified in accuracy of indoor localization.

As shown in Figures 2 and 3, priority of KNN is shown in comparison with NNS method, which Figure 3 is considered for KNN and the case of k = 2. For example, the NNS can achieved the probability of 40%, for error of less than 2 meters, but for KNN method it can be obtained with probability of 55% for the same error. Which KNN method represents an improvement compared to the NNS.

ISSN : 2277-1581
01 May. 2015

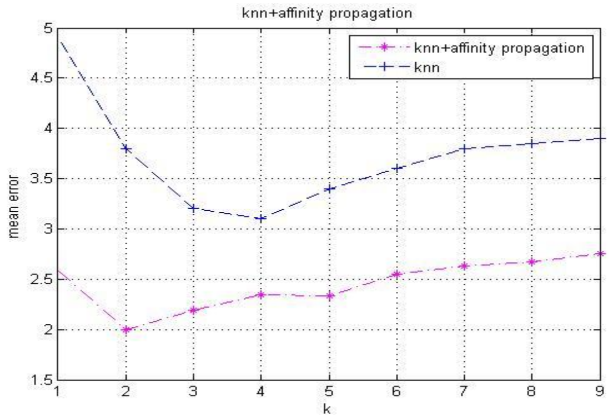

Figure 1 - The effect of changing the value of k on average positioning error for clustered and non-clustered modes

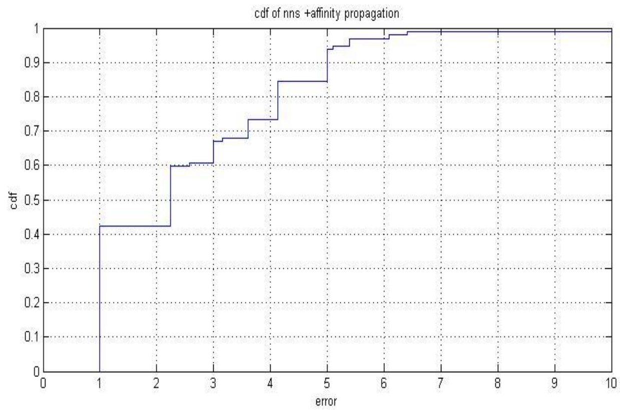

Figure 2- CDF for the nearest neighbor in terms of error

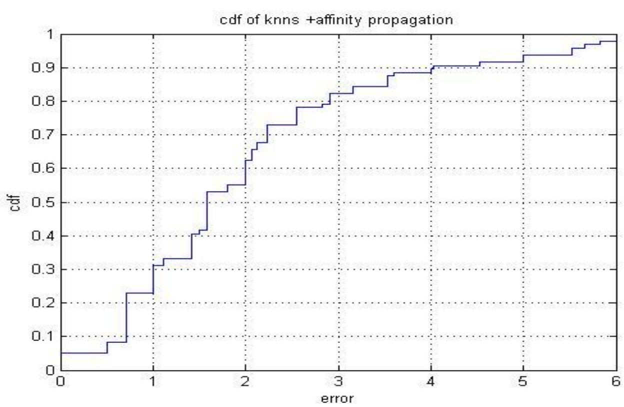

Figure 3- CDF Diagram of error based on KNN and for K = 2

V.Conclusion

According to the results obtained from the simulations, it can be easily seen the effect of clustering in localization accuracy and reducing the mean error location, It is also seen that with clustering method the complexity computation was reduced to a subset of each clusters. Then in online phase simulation was done for two methods as K-nearest neighbor and nearest neighbor, and CDF Diagrams in terms of error based on KNN and NNS were investigated, and a comparison between these two methods of fine localization was done.

References

i. Zhou M, Wong AKS, Tian ZS, Zhang VY, Yu X, Luo X (2013) Adaptive mobility mapping for people tracking using unlabelled Wi-Fi shotgun reads. IEEE Comm Lett 17(1):87–90

ii. ZhouM, Tian Z, Xu K, Yu X,Wu H (2013) Theoretical entropy assessment of fingerprint-based Wi-Fi localization accuracy. Expert Syst Appl 40(15):6136–6149

iii. Khodayari S, Maleki M, Hamedi E (2010) A RSS-based fingerprinting method for positioning based on historical data. Proc IEEE SPECTS 306–31

iv. Brendan j.Frwy and Delbert Dueck ,cluster by Passing Messages Between Data Points, ,2007, pages 972-976

v. Brett Dawes,Kwan-Wu Chin ,A comparison of deterministic and probabilistic methods for indoor localization ,Pages 442-451,2011

vi. v.k.Jain,Shashikala Tapaswi, Perfomance Analysis of Received Signal Strength Fingerprinting Based Distributed Location Estimation System for Indoor WLAN,springer,pages 113-127 2013

vii. Vaidyanathan Ramaduranu ,Localization in Wireless sensor Networks, A Probabilistic Approach ,IEEE,2008

viii. Chan Feng,Received-signal-strength-Based Indoor Positioning Using Compressive Sensing ,IEEE Transations On Mobile Computing, 2012

ix. seyed A,Reza Zekavat ,Handbook of Position Location, pages 530-570, 2010

x. Vahideh Moghtadaiee, Andrew G.Dempster ,Wifi Fingerprinting Signal Strength Error Modeling for Short Distances, IEEE Transations On Mobile Computing,2012