SSLLE: SEMI-SUPERVISED LOCALLY LINEAR EMBEDDING BASED LOCALIZATION METHOD FOR INDOOR WIRELESS NETWORKS

Vinod Kumar Jain, Shashikala Tapaswi and Anupam Shukla
Atal Bihari Vajpayee, Indian Institute of Information Technology & Management, Gwalior, (M.P.), India

Keywords: Location aware services, User location and tracking, Wireless LANs, Dimensional reduction techniques, Locally Linear Embedding(LLE), Semi-supervised learning.

Abstract: Due to vast applications of mobile devices and local area wireless networks, location based services are popularized and location information use has become important . The paper proposes a method based on Semi-supervised Locally Linear Embedding for localization in indoor wireless networks. Previous methods for location estimation in indoor wireless networks require a large amount of labeled data for learning the radio map. However labeled instances are often difficult, expensive, or time consuming to obtain, as they require great efforts, meanwhile unlabeled data may be relatively easy to collect. So the use of semi-supervised learning is more feasible. In the experiment 101 access points (APs) have been deployed so the Received Signal Strength (RSS) vector received by the mobile station has large dimensions (i.e.101). First we have used Locally Linear Embedding, a dimensional reduction technique to reduce the dimensions of data, and then we have used semi-supervised learning algorithm to learn the radio map. The algorithm performs nonlinear mapping between the received signal strengths from nearby access points and the user's location. It is shown that the proposed scheme is easy in training and implementation. Experimental results are presented to demonstrate the feasibility of the proposed SSLLE algorithm.

1 INTRODUCTION

For mobile devices in Wireless Networks, Location Estimation Systems have become very popular in recent years. These systems provide a new layer of automation called automatic object location detection. Real world applications depending on such automation are many. A variety of applications and services such as enhanced-911, improved fraud detection, location sensitive billing, intelligent transport systems and improved traffic management can be examples for cellular networks (Gezici, 2008). On the other hand, applications such as inventory tracking, location detection of products stored in a warehouse, location detection of medical personnel or equipment in a hospital, intruder detection, and patient monitoring etc. are the example of short range networks (Liu et al., 2007). These applications depend heavily on the underlying location estimation techniques.

In order to estimate the user location, a system needs to measure a quantity that is a function of distance. Moreover, the system needs one or more reference points to measure the distance. In case of the Global Positioning System (GPS)(Burrell and Kao, 2011; Enge and Misra, 1999) the reference points are the satellites and the measured quantity is the time of arrival of the satellite signal to the GPS receiver, which is directly proportional to the distance between the satellite and the GPS receiver. In case of Wireless Local Area Network (WLAN) location determination systems, the reference points are the access points and the measured quantity is the signal strength, which decays logarithmically with distance in free space. Unfortunately, in indoor environments, the wireless channel is very noisy and the radio frequency (RF) signal can suffer from reflection, diffraction, and multipath effect, which makes the signal strength a complicated function of distance. To overcome this problem, WLAN location determination systems tabulate this function by sampling it at selected locations in the area of interest. This tabulation has been known in literature as the radio map (Bahl and Padmanabhan, 2000; Bahl et al., 2000; Battiti et al., 2002; Ding et al., 2008; Youssef and Agrawala, 2004; Youssef and Agrawala, 2008) which captures the signature of each access point at certain points in the area of in-

Jain V., Tapaswi S. and Shukla A..
SSLLE: SEMI-SUPERVISED LOCALLY LINEAR EMBEDDING BASED LOCALIZATION METHOD FOR INDOOR WIRELESS NETWORKS.
DOI: 10.5220/0003676801380146
In *Proceedings of the International Conference on Neural Computation Theory and Applications* (NCTA-2011), pages 138-146
ISBN: 978-989-8425-84-3

terest. Different WLAN location determination techniques differs in the way, that how they construct the radio map and what algorithm is used to compare a received signal strength vector with the stored radio map in the location determination phase.

Although the RSS data sets collected for location estimation has high dimensionality and contains several features, but it may be described as a function of only a few underlying parameters. That is, the data points actually belong to a low-dimensional manifold that is embedded in a high-dimensional space. Our method uses the Locally Linear Embedding Algorithm (Roweis and Saul, 2000) to get the low dimension manifold from high dimensional data.

The collection of training data with labels is an arduous work. Traditional classifiers use only labeled data (feature / label pairs) to train. Labeled instances however are often difficult, expensive, or time consuming to obtain, as they require the efforts of experienced human annotators. Meanwhile unlabeled data may be relatively easy to collect, but there has been few ways to use them. Semi-supervised learning addresses this problem by using large amount of unlabeled data, together with the labeled data, to build better classifiers. Because semi-supervised learning requires less human effort and gives higher accuracy, it is of great interest both in theory and in practice.

In order to meet the requirement of higher accuracy and lower calibration effort, we have proposed a semi-supervised locally linear embedding method for location estimation in indoor wireless networks. The key feature of the proposed method is that advantages of locally linear embedding and semi-supervised learning are integrated to improve the localization efficiency.

Experimental results are presented to demonstrate the feasibility of the proposed scheme. The rest of the paper is organized as following. In section 2 we review some existing work for indoor location estimation in wireless networks. In section 3, we describe the Semi-Supervised Locally Linear Embedding Algorithm for location estimation. Section 4 describes the detailed method of localization. Experimental results are presented in section 5 and finally the paper is concluded in section 6.

2 RELATED WORK

All previous techniques have been proposed for location estimation in wireless networks can be divided into three categories:

2.1 Triangulation

Triangulation is a geometric method based on the geometric properties of triangles to estimate the target location. It is very sensitive to wireless signal propagation. It has two derivations: lateration and angulation.

Lateration determine the location of mobile device by measuring its distances from multiple reference points. Received Signal Strengths (RSS), time of arrival (TOA) or time difference of arrival (TDOA) are usually measured; to derive the distance by computing the attenuation of the emitted signal strength or by multiplying the radio signal velocity and the travel time (Liu et al., 2007).

Angulation estimates the location of mobile device by computing angles relative to multiple reference points. In Angle of Arrival (AOA), the location of the desired mobile device can be determined by the intersection of several pairs of angle direction lines, each formed by the circular radius from a base station or a beacon station to the mobile target(Liu et al., 2007).

2.2 Vicinity

It measures closeness to a known set of locations. Vicinity algorithms provide symbolic relative location information. It depends upon a dense grid of antennas, each having a well-known position. When a mobile device is in the range of a single antenna, it is considered to be co-located with it. When mobile device is in the range of more than one antenna, it is considered to be co-located with the one that receives the strongest signal. This method is relatively simple to implement. It can be implemented over different types of physical media(Liu et al., 2007).

2.3 Radio Finger-printing

It examines a view from a particular vantage point. RF-based finger-printing refers to the type of algorithms that first collect features (fingerprints) of a scene and then estimate the location of an object by matching online measurements with the closest a priori location fingerprints. RSS-based location fingerprinting is commonly used in Radio Finger-Printing methods. (Chen, 2005; Krishnan et al., 2004; Youssef and Agrawala, 2004; Youssef and Agrawala, 2008). There are two stages for location fingerprinting: offline stage and online stage (or run-time stage). During the offline stage, a site survey is performed in an environment. The location coordinates/labels and

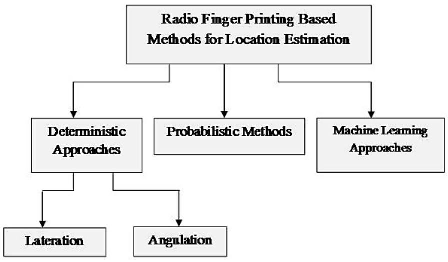


Figure 1: Radio Fingerprinting based methods for Location Estimation.

respective signal strengths from nearby Base Stations/Access Points are collected. During the online stage, a location positioning technique uses the currently observed signal strengths and previously collected information to figure out an estimated location. The main challenge to the techniques based on location fingerprinting is that the received signal strength could be affected by diffraction, reflection, and scattering in the propagation indoor environments(Bahl and Padmanabhan, 2000; Bahl et al., 2000; Battiti et al., 2002).

Radio Finger Printing techniques, which are also known as location fingerprinting, can be categorized into three broad categories: deterministic techniques, probabilistic techniques, and machine learning based techniques as shown in Figure 1. *Deterministic techniques*, represent the signal strength of an access point at a location by a scalar value, for example, the mean value, and use non-probabilistic approaches to estimate the user location. For example, in the Radar (Bahl et al., 2000) system the authors use nearest neighborhood techniques to infer the user location. The accuracy of RADAR is about three meters with fifty percent probability. K. Pehlavan et al. also used kNN (k-nearest neighbour) technique and achieved 2.8 meter distance error (Pahlavan et al., 2002). On the other hand, *probabilistic techniques*, store information about the signal strength distributions from the access points in the radio map and use probabilistic techniques to estimate the user location. For example, the Horus (Youssef and Agrawala, 2004; Youssef and Agrawala, 2008) system from the University of Maryland uses the stored radio map to find the location that has the maximum probability given the received signal strength vector. Probabilistic approaches like Bayesian networks based solutions achieve better performance but they are computationally exhaustive and difficult to scale.

In a heterogeneous environment, e.g. inside a building or in a variegated urban geometry, the received signal strength is a very complex function of the distance, the geometry, and the materials. The complexity of the inverse problem (to derive the position from the signals) and the lack of complete information, motivate to consider flexible models based on machine learning approaches (i.e. artificial neural networks, genetic algorithms, fuzzy systems) (Ahmad et al., 2008; Battiti et al., 2002; Chen, 2005; Ding et al., 2008; Gupta et al., 2009; Yang et al., 2010; Youssef and Agrawala, 2008). Battiti et al. (Battiti et al., 2002) have employed neural networks for this problem. Battiti et al. used feed forward back propagation network that takes RSS of 3 Wireless Access Points (AP) to cover 624 square meter area, and reported median estimation distance error of 1.75 meter. This model assumes that the signals of all the access points are available at every location all the time. Practically, this approach has limited applicability because in real life scenario some AP may not be visible (not in range) at all the locations for all the time (Ahmad et al., 2008). The benefit of machine learning based methods are that they do not need ad-hoc infrastructure in addition to the wireless LAN, while the flexible modeling and learning capabilities of machine learning approaches achieve lower errors in determining the position, and are scalable to incremental improvements. A user needs only a map of the working space and some identified locations to train a system.

3 SSLLE: SEMI-SUPERVISED LOCALLY LINEAR EMBEDDING

The RSS dataset collected for location estimation has high dimensionality and contains several features but it may be described as a function of only a few underlying parameters. Therefore for computing efficiency, dimensional reduction technique is used to find out the low dimensional manifold that is embedded in a high dimensional space. We use Locally Linear Embedding (LLE) (Roweis and Saul, 2000; Saul and Roweis, 2003) algorithm which computes low dimensional embedding of high dimensional data so that neighborhood information is preserved. LLE does not estimate the pair-wise distances between widely separated data points. Unlike Principle Component Analysis (PCA), Multi Dimensional Scaling (MDS), LLE recovers nonlinear structure from locally linear fits. LLE attempts to discover nonlinear structure in high dimensional data by exploiting the local symmetries of linear reconstructions. Notably, LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations, though capable of generating highly nonlinear embeddings, do not involve local minima. LLE maps high-dimensional inputs into a low dimensional "description" space with as many coordinates as observed modes of variability.

Suppose the data consist of N real-valued vectors $\overrightarrow{X_i}$ each of dimensionality D, sampled from some underlying manifold. LLE expects each data point and its neighbors to lie on or close to a locally linear patch of the manifold. Characterize the local geometry of these patches by linear coefficients that reconstruct each data point from its neighbors. Reconstruction errors are measured by the cost function

$$\varepsilon(W) = \sum_i |\overrightarrow{X_i} - \sum_j W_{ij}\overrightarrow{X_j}|^2 \tag{1}$$

which adds up the squared distances between all the data points and their reconstructions. Minimize the cost function subject to two constraints:

First, each data point $\overrightarrow{X_i}$ is reconstructed only from its neighbors, enforcing $W_{ij} = 0$ if $\overrightarrow{X_j}$ does not belong to the set of neighbors of $\overrightarrow{X_i}$, second that the rows of the weight matrix sum to one: $\sum_j W_{ij} = 1$.

Consider a particular data point $\overrightarrow{x}$ with k nearest neighbors $\overrightarrow{\eta_j}$ and reconstruction weights w_j that sum to one. The reconstruction $|\overrightarrow{X} - \sum_{j=1}^{k} W_j\overrightarrow{\eta_j}|^2$ is minimized:

$$|\overrightarrow{X} - \sum_{j=1}^{k} W_j\overrightarrow{\eta_j}|^2 = |\sum_{j=1}^{k} W_j(\overrightarrow{X} - \overrightarrow{\eta_j})|^2 = \sum_{jk} W_j W_k G_{jk}$$

we have introduced the "local" Gram matrix $G_{jk} = (\overrightarrow{X} - \overrightarrow{\eta_j})(\overrightarrow{X} - \overrightarrow{\eta_k})$.

By construction, this Gram matrix is symmetric and semipositive definite. The reconstruction error can be minimized analytically using a Lagrange multiplier to enforce the constraint that $\sum_j W_{ij} = 1$. Then, compute the reconstruction weights:

$$w_j = \frac{\sum_k G_{jk}^{-1}}{\sum_{lm} G_{lm}^{-1}}$$

If the correlation matrix G is nearly singular, it can be conditioned (before inversion) by adding a small multiple of the identity matrix. This amounts to penalizing large weights that exploit correlations beyond some level of precision in the data sampling process. The constrained weights that minimize these reconstruction errors obey an important symmetry: for any particular data point, they are invariant to rotations, rescaling, and translations of that data point and its neighbors. Suppose the data lie on or near a smooth nonlinear manifold of lower dimensionality $d \ll D$. By design, the reconstruction weights W_{ij} reflect intrinsic geometric properties of the data that are invariant to exactly such transformations.

Each high-dimensional observation $\overrightarrow{X_i}$ is mapped to a low-dimensional vector $\overrightarrow{Y_i}$ representing global internal coordinates on the manifold. This is done by choosing d-dimensional coordinate $\overrightarrow{Y_i}$ to minimize the embedding cost function

$$\Phi(Y) = \sum_i |\overrightarrow{Y_i} - \sum_j W_{ij}\overrightarrow{Y_j}|^2 \tag{2}$$

This cost function, like the previous one, is based on locally linear reconstruction errors, but here the weights W_{ij} are fixed, while optimizing the coordinates $\overrightarrow{Y_i}$. Subject to constraints that make the problem well-posed, it can be minimized by solving a sparse $N \times N$ eigen value problem, whose bottom d nonzero eigenvectors provide an ordered set of orthogonal coordinates centered on the origin.

The embedding vectors $\overrightarrow{Y_i}$ are found by minimizing the cost function $\Phi(Y) = \sum_i |\overrightarrow{Y_i} - \sum_j W_{ij}\overrightarrow{Y_j}|^2$ over $\overrightarrow{Y_i}$ with fixed weights W_{ij}. To avoid degenerate solutions, we constrain the embedding vectors to have unit covariance, with outer products that satisfy $\frac{1}{N}\sum_i \overrightarrow{Y_i} \otimes \overrightarrow{Y_i} = I$ where $I = d \times d$. Now the cost defines a quadratic form $\Phi(Y) = \sum_{ij} M_{ij}(\overrightarrow{Y_i}\overrightarrow{Y_j})$ involving inner products of the embedding vectors and the symmetric $N \times N$ matrix $M_{ij} = \delta_{ij} - W_{ij} - W_{ji} + \sum_k W_{ki}W_{kj}$ where δ is 1 if $i = j$ and 0 otherwise.

The optimal embedding, up to a global rotation of the embedding space, is found by computing the bottom *d+1* eigenvectors of this matrix. For such implementations of LLE, the algorithm has only one free parameter: the number of neighbors, K.

The output received after performing the LLE is low dimensional data set. Let the dataset consists of N instances, where the first l instances are the labeled data, and the rest of them are the unlabeled data. The *i-th* data is given as $(X^{(i)}, L^{(i)})$ where $X^{(i)} \in R^k$ is the vector of the Received Signal Strength (RSS) values from the WiFi Access Points (APs), and $L^{(i)} \in \{1,2,3,......c\}$ is the location label assigned to the RSS vector. As for the unobserved RSS values, we filled them by -100, since all RSS values are in the range of $[-100, 0]$ and unobserved RSS value implies that it was too weak to detect. The objective of this task is to predict the location labels of the unlabeled data, $L^{(l+1)}$..........L^N since they are not given.

Let $f^{(i)}(c) \in [0,1]$ indicate the probability with which the location label of the *i-th* instance is c. For the labeled data $i \leq l$, the following holds

$$f^{(i)}(c) = \begin{cases} 1 & \text{if } c = L^{(i)} \\ 0 & \text{(otherwise)} \end{cases} \tag{3}$$

The task is to predict $f(i)(L^{(i)})$ for $i > l$ and $\forall c$, with which we obtain the prediction $\hat{c}(i)$ for $i > l$ by

$$\hat{c}(i) = argmax_c f^{(i)}(c) \tag{4}$$

In the label propagation framework, we try to minimize the discrepancies of the label distributions among neighbourhood instances, which is defined as,

$$\sum_{ij} w^{(i,j)} \sum_{c} (f_i^{(c)} - f_j^{(c)})^2$$

here w_{ij} is a constant called the affinity indicating the similarity between the i^{th} instance and j^{th} instance, which we will define later. It is easy to see that the solution of the above optimization problem satisfies

$$f_i^{(c)} = \frac{\sum_j W^{(ij)} f_i^{(c)}}{\sum_j W^{(ij)}} \quad (5)$$

for $\forall i > l$ and $\forall c$. Therefore, instead of solving the large optimization problem directly, we can iteratively apply Eq. 5 to make local updates of predictions until convergence.

Our definition of the affinity tries to imply that two instances are similar if their RSS vectors are similar. For the affinity between two RSS vectors $X^{(i)}$ and $X^{(j)}$, we used a heat-kernel like function

$$W_x^{(i,j)} = exp\left(-\frac{\left\|X^{(i)} - X^{(j)}\right\|_q^q}{\sigma}\right), \quad (6)$$

where σ is a scale parameter. Also, $\|.\|_q$ is the q-norm which is defined as

$$\|X\|_q = \left(\sum_d |X_d|^q\right)^{\frac{1}{q}},$$

and we set q = 0.5 based on the observation that this choice performed well in our preliminary analysis using the nearest neighbor classifier for the labeled data (Kashima et al., 2007; Yang et al., 2008).

4 THE DETAILED LOCATION ESTIMATION METHOD

4.1 Signal Characteristics

In the experimental testbed, there are a total of 101 APs deployed. Some APs are deployed within the same floor of testbed, others are deployed on other floors or in neighboring buildings. Since, many APs are detected occasionally therefore mobile device didn't receive RSS from all the APs on every location at a particular moment. Figure 2 shows the probability distribution of signal strength received from an AP at a particular fixed location. It illustrates that the signal strength of AP's varies with time and the probability distribution of RSS from AP is Gaussian. Figure 3 shows the number of locations covered by each AP and Figure 4 shows the number of AP's detected at each location space.

4.2 The Detailed Location Estimation Algorithm

The proposed location algorithm which is based on semi-supervised locally linear embedding learning algorithm has two phases: an offline radio map training phase , and an online location estimation phase.

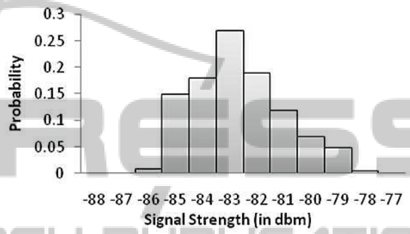


Figure 2: Probability distribution of RSS from a AP at fixed location with time.

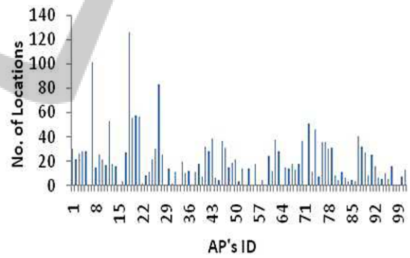


Figure 3: Number of Locations covered by each AP.

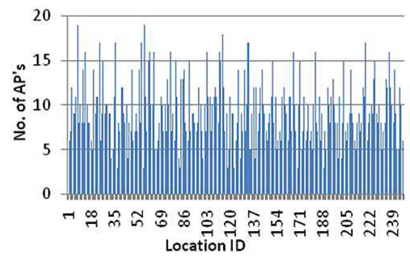
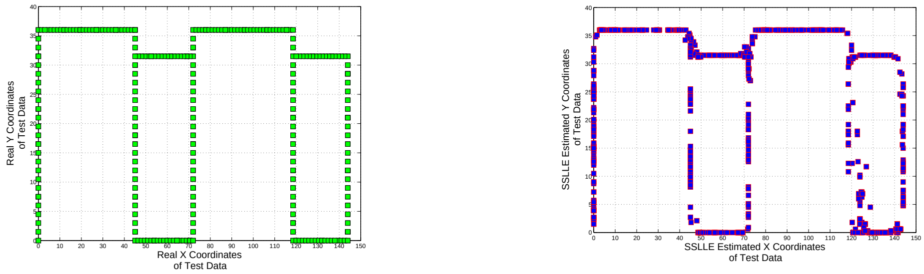


Figure 4: Number of AP's detected at each location.

(1) Offline Radio Map Training Phase. At first in the offline radio map training phase the labeled (rss signal and location id pair) and unlabeled rss signal data by mobile device at various locations is collected. Since all the access points are not visible at every location, so for unobserved rss values we filled them with -100. Let the whole dataset consists of N instances, where the first l instances are



(a) Real Locations of Test Data

(b) SSLLE Estimated Locations of Test Data

Figure 5: Comparison of Real and SSLLE Estimated Locations.

the labeled data, and the rest of them are the unlabeled data. The i^{th} data is given as $(X^{(i)}, L^{(i)})$ where $X^{(i)} \in R^k$ is the vector of the received signal strength (RSS) values from the WiFi Access Points (APs), and $L^{(i)} \in \{1,2,3,.....c\}$ is the location label assigned to the RSS vector. As there are 101 access points deployed in the testbed , the data set has very high dimension, so then we use locally linear embedding algorithm to reduce the dimensionality of data, and finally semi-supervised multiclass label propagation algorithm is applied for training the radio map.

The detailed steps for offline radio map training phase are as follows:

Step 1: Compute the K nearest neighbors of each high dimensional data point, $\overrightarrow{X_i}$.

Step 2: Compute the weights W_{ij} that best reconstruct each data point $\overrightarrow{X_i}$ from its neighbors, minimizing the cost in Eq. 1 by constrained linear fits.

Step 3: Compute the low dimensional embedding vectors $\overrightarrow{Y_i}$ best reconstructed by the weights W_{ij}, minimizing the quadratic form in Eq. 2 by its bottom nonzero eigenvectors.

Step 4: Initialize $f^{(i)}$ by using Eq. 3 for the labeled instances $i \leq l$.

Step 5: Compute the affinities $w_x^{(i,j)}$ between all pairs of instances by using Eq. 6.

Step 6: Continue the following steps, 6(a) and 6(b), until convergence.
(a) Select $i > l$ uniformly at random.
(b) Update $f^{(i)}(c)$ for $\forall c$ by using Eq. 5.

Step 7: Location label prediction for $i > l$ (i.e. unlabeled data) by using Eq. 4.

(2) Online Location Estimation Phase. In the second phase the captured signal vector by the target mobile device from various access points is used to estimate the location of target mobile device. For this purpose the mapping function which is estimated in offline training phase is used.

The detailed steps are as follows:

Step 1: The mobile device captured the signal strength from various detected access points at its location, and then form the rss vector $\overrightarrow{X}$ by filling -100 for unobserved AP^s.

Step 2: Use the semi-supervised locally linear embedding algorithm to estimate the location as described in the offline radio map training phase.

5 EXPERIMENTAL RESULTS

We have used the ICDM DMC'07 dataset for the experiment and to evaluate our proposed algorithm (Qiang Yang, 2007). IEEE ICDM Data Mining Contest (IEEE ICDM DMC'07) published the realistic public benchmark data for indoor location estimation from radio signal strength received by a client device from various WiFi AP^s. They collected the data sets in an academic building in the Hongkong University of Science and Technology, consisting of an area of 145.5m 37.5m. Locations were divided into 247 grids, each of which has a size of about 1.5m 1.5m. There were 101 wireless access points (AP^s) deployed in the building. A person holding a wireless client device walks around a building floor. The client device (which can be a Personal Digital Assistant (PDA) or Laptop) is equipped with a wireless card that can receive signals from wireless access points (AP^s) which are visible out of 101 surrounding wireless access points (AP^s). Each of these AP^s is identifiable with a unique ID. Since, collecting the (RSS values, Location Label) pairs as training data in a large building are very costly, because humans need to take a mobile device and walk through the building to collect the RSS values and mark down the ground locations. Therefore some data are given without labels; that is for those data only the RSS values are given. In addition, a collection of partially labeled user traces are given, which corresponds to a sequence of RSS values collected as a user continuously walks around a building. In the experiment we

Table 1: Performance of Different Methods.

Methods	Min Error Distance (in meters)	Max Error Distance (in meters)	Mean Error Distance (in meters)	Standard Deviation Error Distance (in meters)
SSLLE	0	21.6	3.7033	3.4567
RADAR	0	36.837	6.1374	6.442
Semi-supervised Method used in (Kashima et al., 2007)	0	138.2	19.348	27.859

use total 5333 samples (labeled and unlabeled both) for training the mapping function, and use 2137 samples of test data to test the performance of method. Based on the collection of signal strength values (RSS values), a semi-supervised locally linear embedding algorithm running on the client device tries to figure out the current location of the user. The comparison between SSLLE estimated locations and real locations are shown in Figure 5, and some of the results obtained from SSLLE out of 2137 test results are reported in Table 2.

As shown in the Table 1. the proposed algorithm performs well. The mean error is 3.7033m, and the standard deviation of error is 3.4567m. We also performed experiment to compare the results with the method used in (Kashima et al., 2007) and with the RADAR (Ahmad et al., 2008). The detailed experimental results are summarized in Table 1. As can be seen, the proposed algorithm has a better performance than the others. The proposed algorithm has the smallest mean, standard deviation, minimum and maximum error distances.

6 CONCLUSIONS

In this paper, we describe a semi-supervised locally linear embedding algorithm to estimate the location of mobile device in indoor wireless networks. In this method locally linear embedding is used to reduce the dimensionality of data. In offline radiomap training phase a mapping function is learned between the signal space and physical space using labeled and unlabeled data. And then in location determination phase, we use this function to estimate the location of mobile device. SSLLE uses only a small amount of labeled data and a large amount of unlabeled data. So it greatly reduce the calibration effort since, collecting the (RSS values, Location Label) pairs as training data in a large building are very costly and difficult. The experimental results shows that SSLLE outperforms in terms of mean, standard deviation , minimum, and maximum of error distances in comparison to benchmark methods RADAR (Ahmad et al., 2008), and semi-supervised method used in (Kashima et al., 2007).

ACKNOWLEDGEMENTS

This work was supported by Atal Bihari Vajpayee Indian Institute of Information Technology & Management , Gwalior (M.P.)- India.

REFERENCES

Ahmad, U., Gavrilov, A., Lee, Y., and Lee, S. (2008). Context-aware, self-scaling Fuzzy ArtMap for received signal strength based location systems. *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, 12(7):699–713.

Bahl, P. and Padmanabhan, V. (2000). RADAR: An in-building RF-based user location and tracking system. In *Proceedings of 19th IEEE Annual Joint Conference of Computer and Communications Societies, INFOCOM 2000.*, volume 2, pages 775–784.

Bahl, P., Padmanabhan, V., and Balachandran, A. (2000). Enhancements to the RADAR user location and tracking system. *Microsoft Research*, (MSR-TR-2000-12):13.

Battiti, R., Nhat, T., and Villani, A. (2002). Location-aware computing: a neural network model for determining location in wireless LANs. *Technical Report DIT-02-083, Ingegneria e Scienza dell'Informazione, University of Trento.*

Burrell, G. and Kao, D. M. (2011). About GPS. Retrieved from: http://www.garmin.com/aboutGPS/, Access date: 20.04.2011.

Chen, C. (2005). Hybrid location estimation and tracking system for mobile devices. In *Proceedings of 61st IEEE Vehicular Technology Conference, VTC 2005.*, volume 4, pages 2648–2652.

Ding, X., Li, H., Li, F., and Wu, J. (2008). A novel infrastructure WLAN locating method based on neural network. In *Proceedings of 4th Asian Conference on Internet Engineering*, pages 47–55.

Enge, P. and Misra, P. (1999). Scanning the Special Issue/Technology on the Global Positioning System. In *Proceedings of IEEE, Special Issue on GPS*, volume 87, pages 3–15.

Table 2: Comparison of Real and SSLLE estimated coordinates, and the error distances for some of the data instances used for testing the SSLLE Method.

Data Instance No.	Real X Coordinates	Real Y Coordiantes	SSLLE Estimated X Coordinates	SSLLE Estimated Y Coordiantes	Error Distance (in Meters)
1	139.5	31.5	138.3	31.5	1.2
2	139.5	31.5	138.3	31.5	1.2
3	139.5	31.5	138.3	31.5	1.2
4	88.5	36	86.1	36	2.4
5	88.5	36	88.2	36	0.3
6	88.5	36	88.2	36	0.3
7	88.5	36	88.2	36	0.3
8	91.5	36	92.1	36	0.6
9	91.5	36	90	36	1.5
10	91.5	36	90	36	1.5
11	94.5	36	92.4	36	2.1
12	96	36	93.6	36	2.4
13	36	36	34.8	36	1.2
14	36	36	34.8	36	1.2
15	36	36	34.8	36	1.2
16	37.5	36	35.1	36	2.4
17	37.5	36	35.1	36	2.4
18	37.5	36	35.1	36	2.4
19	37.5	36	37.2	36	0.3
20	37.5	36	37.2	36	0.3
21	37.5	36	37.2	36	0.3
22	37.5	36	37.2	36	0.3
23	37.5	36	38.1	36	0.6
24	37.5	36	38.1	36	0.6
25	37.5	36	38.1	36	0.6
26	39	36	38.7	36	0.3
27	39	36	38.7	36	0.3
28	39	36	38.7	36	0.3
29	39	36	38.7	36	0.3
30	39	36	39.6	36	0.6
31	40.5	36	39.6	36	0.9
32	40.5	36	39	36	1.5
33	40.5	36	39	36	1.5
34	40.5	36	39.3	36	1.2
35	40.5	36	39.3	36	1.2
36	40.5	36	39.3	36	1.2
37	42	36	39.6	36	2.4
38	42	36	39.6	36	2.4
39	42	36	39.6	36	2.4
40	42	36	40.8	36	1.2
41	42	36	40.8	36	1.2
42	42	36	40.8	36	1.2
43	42	36	42	36	0
44	42	36	42	36	0
45	42	36	42	36	0
46	43.5	36	42	36	1.5
47	43.5	36	42	36	1.5
48	43.5	36	42	36	1.5
49	43.5	36	42.9	36	0.6
50	43.5	36	42.9	36	0.6

Gezici, S. (2008). A survey on wireless position estimation. *Wireless Personal Communications*, 44(3):263–282.

Gupta, A., Tapaswi, S., and Jain, V. (2009). Recurrent Grid Based Voting Approach for Location Estimation in Wireless Sensor Networks. In *Proceedings of Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, UIC-ATC'09.*, pages 263–267.

Kashima, H., Suzuki, S., Hido, S., Tsuboi, Y., Takahashi, T., Ide, T., Takahashi, R., and Tajima, A. (2007). A Semi-supervised Approach to Indoor Location Estimation. Retrieved from: www.geocities.co.jp/Technopolis/5893/publication/ICDMDMC2007-1.pdf, Access Date: 20.04.2011.

Krishnan, P., Krishnakumar, A., Ju, W.-H., Mallows, C., and Gamt, S. (2004). A system for lease: location estimation assisted by stationary emitters for indoor rf wireless networks. In *Proceedings of 23rd IEEE Annual Joint Conference of Computer and Communications Societies, INFOCOM 2004.*, volume 2, pages 1001 – 1011.

Liu, H., Darabi, H., Banerjee, P., and Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews.*, 37(6):1067–1080.

Pahlavan, K., Li, X., and Makela, J. (2002). Indoor geolocation science and technology. *IEEE Communications Magazine*, 40(2):112 –118.

Qiang Yang, Sinno Jialin Pan, V. W. Z. (2007). IEEE ICDM Data Mining Contest 2007. Retrieved from: http://www.cse.ust.hk/ qyang/ICDMDMC07/, Access date: 20.04.2011.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.

Saul, L. and Roweis, S. (2003). Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119–155.

Yang, B., Xu, J., Yang, J., and Li, M. (2010). Localization algorithm in wireless sensor networks based on semi-supervised manifold learning and its application. *Cluster Computing*, 13(4):435–446.

Yang, Q., Pan, S., and Zheng, V. (2008). Estimating location using wi-fi. *IEEE Intelligent Systems*, 23(1):8–13.

Youssef, M. and Agrawala, A. (2004). Handling samples correlation in the horus system. In *Proceedings of 23rd IEEE Annual Joint Conference of Computer and Communications Societies, INFOCOM 2004.*, pages 1023–1031.

Youssef, M. and Agrawala, A. (2008). The Horus location determination system. *Wireless Networks*, 14(3):357–374.