Locality Preserving Canonical Correlation Analysis Distributed Localization Algorithm for Wireless Sensor Networks

Taotao Huang, Jingjing Gu* and Yi Zhuang

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing
TaotaoHuang@nuaa.edu.cn, gujingjing@nuaa.edu.cn

Abstract

Localization is essential for wireless sensor networks. The state-of-the-art methods mainly adopt low accurate signal strength to perform localization, which are suffering from low localization accuracy and high variance. The machine learning methods are introduced to confront the low data quality challenges and provide considerable localization accuracy and other advantages. However, these series of methods also bear some drawbacks such as high training cost and high energy consumption. To this end, learning from our previous algorithm LE-LPCCA (Location Estimation-Locality Preserving-Canonical Correlation Analysis), we proposed an improved version, called, LE-DLPCCA (LE-Distributed-LPCCA), which greatly reduces the training cost and energy consumption. Specifically, LE-DLPCCA employs a clustering algorithm based on energy equilibrium. The training process, which maps the signal space into physical space, is conducted in a distributed manner for each cluster. Then, in the positioning phase, the unknown node estimates the distances from the most similar anchor nodes through the mapping and perform the localization of the unknown nodes through the maximum likelihood method. Demonstrated by multiple simulations, LE-DLPCCA algorithm is in high accuracy, fast localization model efficiency and low average energy consumption.

Keywords: wireless sensor network; clustering algorithm; energy equilibrium; LE-LPCCA; maximum likelihood method; distributed

1. Introduction

Wireless sensor network is composed of a large number of cheap sensor nodes deployed in monitoring area. The sensor nodes form a multiple hops self-organizing network system through wireless communication. WSN is cooperative to perceive, collect and process the information of perceptive objects in the network coverage. It has the characteristics of large scale, self-organization, dynamic, reliability and data centric. As the microprocessor volume decreases and performance improves, the application prospect of wireless sensor network is rather extensive, its application mainly concentrated on environmental monitoring, medical care, military and other fields [1-2]. One of the basic functions of the wireless sensor network is to collect the location of the sensors. Sensor nodes without location information are meaningless. Therefore, localization becomes a hot technology in the Wireless Sensor Network.

At present, the localization algorithm of WSN generally contains the range-based localization algorithm and the free range-based localization algorithm [3]. The range-based localization algorithm depends on measuring the actual distance and the angle between adjacent nodes. It mainly include RSSI [4], TOA [5], AOA [6], TDOA [7] ,*etc*. The free range-based localization algorithm is largely based on the connection information of the neighbor nodes and the specific coordinate events. It mainly include

Centroid algorithm [8], Convex programming algorithm, DV-Hop algorithm [9], APIT algorithm [10] and so on.

The above localization algorithm has been widely used. Due to the hardware equipment, the external environment and other factors, they result in a larger error, many algorithms have been improved. ZANCA *et al*. present an indoor localization algorithm [11] based on RSSI in wireless sensor networks, which needs no additional hardware equipment, low power consumption and low cost. However, it lacks of stability and the localization accuracy is poor. LEE *et al*. propose a wireless sensor network localization algorithm [12] based on improved AOA algorithm in Ad-Hoc network, which requires additional hardware, vulnerable to the impact of the environment and the result is rather unstable. CHEN *et al*. use global beacon nodes of the average hop distance [13] to calculate the distance between unknown nodes and beacon nodes, and adopt hyperbolic localization method to solve for the coordinate of the unknown node. Xu *et al*. put forward a APIT based centroid localization algorithm [14] to overcoming the localization error caused by the uneven and sparse beacon nodes, which improves localization accuracy. But this algorithm applies only to sparse anchor nodes in wireless sensor network. It is not suitable to the general wireless sensor networks.

Overall, the above localization algorithms are mainly unstable, low localization accuracy and high energy consumption. In order to solve these problems, researchers have provided a lot of improvement strategies and algorithms. In recent years, modeling and localization by machine learning has become one of the research focuses [1-2], which regard the essence of the wireless sensor localization process as a machine learning process. Researchers In-depth mine the available data in wireless sensor networks such as signal strength, physical coordinates, network topology information. Then we establish a localization model for wireless sensor networks to estimate the physical coordinates of the unknown nodes.

In this paper, we propose an improved machine learning localization algorithm and construct a model to improve the positioning accuracy, enhance the speed of modeling and reduce the average energy consumption. The proposal utilizes the clustering algorithm to divide the monitoring area into many subdomains and establish the mapping between the signal space and physical space. In addition, we adopt the maximum likelihood method to estimate the physical coordinates of the unknown nodes. Finally, the performance of the algorithm is verified and compared with other methods. The simulation results show that the proposed algorithm outperforms the other two methods in terms of accuracy, speed of modeling and the energy consumption.

LE-DLPCCA makes the following contributions:

a) The algorithm can provide a more accurate and stable method to estimate the physical coordinates of the unknown nodes.

b) The cluster head nodes of subdomain perceive the invasion of unknown nodes, which awaken the other cluster nodes. So it can reduce the energy consumption of sensor nodes.

c) The algorithm can establish the whole model by combination of sub models, so we can increase the speed of modeling.

Outline Section II briefly introduces the related work of machine learning localization algorithm. In Section III, the LE-DLPCCA localization algorithm is described in details. In Section IV, exemplifies it with the simulation to illustrate the efficiency of this algorithm and make the conclusion in the Section V.

2. Related Work

2.1. Localization Algorithm Based on Machine Learning

Machine learning is applied to positioning in the wireless network since 2005. Nowadays it becomes one of the highlights in the localization field. Common machine learning tools mainly include: 1) the positioning algorithm based on classification method; 2) the positioning algorithm based on the linear regression; 3) the positioning algorithm based on the core technology; 4) the positioning algorithm based on manifold; and 5) the positioning algorithm based on Semi-Supervised.

As for the first class, the classification method is mainly used to identify, locate and detect the error nodes. Xin *et al* pose a novel distributed algorithm [15], which transform the localization problem into a binary classification problem. As for the second class, the linear regression is an approach for fitting the curve; find a suitable function and location model. Frank *et al*. present a novel pseudo-3D localization algorithm [16] using linear regression. As for the third class, the core technology mainly selects a suitable kernel function to mine the hidden information. KEJI *et al*. put forward a support vector machine(SVM) localization algorithm [17]. The proposed method divides the network area into a plurality of equal compartments, each cell representing the machine learning algorithm in a category.

As for the fourth class, manifold learning is mainly used to mine the local information of the network. Gu *et al*. come up with the LE-LPCCA localization algorithm [18], which combine the CCA algorithm with the model. Through the analysis of the topology of wireless sensor network, with locality preserving canonical correlation analysis model, it establishes the signal space to real physical space mapping. As for the fifth class, Semi-Supervised is mainly suitable for the network of less beacon nodes. Jeffrey *et al*. perform a novel approach [19], which combines collaborative filtering with graph-based semi-supervised learning to learn both mobile user's locations and the locations of access points. CHEN *et al*. present a semi-supervised Laplacian regularized least squares algorithm [20]. With the help of semi-supervised manifold localization algorithm, they establish the mapping relationship between signal strength and pair-wise distance between nodes.

The localization algorithm based on machine learning reaches good positioning accuracy and stability to some extent. However, it requires a lot of time to build a learning model and wastes much energy in machine learning. If the model needs the high accuracy, it may be more complex. Since it has high requirement for dimension calculation and frequent iteration, it is especially not suitable for large networks. Aiming at solving the problem, we try to adopt a distributed way. The LE-DLPCCA proposed in this paper is to start with these problems.

2.2. Brief Review of LE-LPCCA

In the localization algorithm based on machine learning, the main procedure consists of two stages: training phase and positioning stage. In the training phase, this paper mainly establishes a mapping between signal strength and physical coordinates. In the positioning stage, locates the unknown nodes through the mapping of the training phase. Compared with other similar algorithms, the LE-LPCCA algorithm has higher positioning accuracy and more stable performance. Here, we mainly review the LE-LPCCA localization model.

In large scale wireless sensor networks, assuming that there is a set of sensor nodes deployed in the open field, including m anchor nodes, they transmit signals to unknown nodes. The unknown node receives a set of signal strength matrix $X = [x_1, x_2, ..., x_n]$ and the corresponding beacon node physical coordinate matrix $Y = [y_1, y_2, ..., y_n]$, thereby obtaining a set of training set $\{(x_i, y_i)\}$.

Definition 1: In the signal space, $ne(i)$ is expressed as the index set of ith local nearest neighbors, x_i is the signal set of the local nearest neighbor. In the physical space, $ne(i)$ represents the index set of ith local nearest neighbors. x_i is the coordinate set of the local nearest neighbor. The local nearest neighbor is defined by the k-nearest neighbor. If it is the k- nearest neighbor, it is called the local neighbor.

According to the definition, the similarity matrix of the topological structure is $S_X = \{S_{ij}^X\}_{i,j=1}^n$ and $S_Y = \{S_{ij}^Y\}_{i,j=1}^n$ in the wireless sensor network between signal space and physical space. S_{ij}^X and S_{ij}^Y are illustrated as follows:

$$S_{ij}^X = \begin{cases} \exp(-\|x_i - x_j\|^2 / t_x), & j \in ne(i) / i \in ne(j) \\ 0 & else \end{cases} \tag{1}$$

$$S_{ij}^Y = \begin{cases} \exp(-\|y_i - y_j\|^2 / t_y), & j \in ne(i) / i \in ne(j) \\ 0 & else \end{cases} \tag{2}$$

Here, $t_x(t_y)$ is a parameter, and we set $t_x = \sum_{i=1}^n \sum_{j=1}^n \|x_i - x_j\|^2 / n(n-1)$ $(t_y = \sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 / n(n-1))$. The bigger the S_{ij}^X is, the closer the distance between x_i and x_j is. S_{ij}^Y has the same characteristics as S_{ij}^X. If x_i and x_j are not in the neighborhood, the similarity is zero. y is the same as x. It can be seen that S_{ij}^X or S_{ij}^Y depend on sensor node layout. Therefore, S_{ij}^X or S_{ij}^Y can be changed with the network topology, which shows a high degree of flexibility.

The optimization problem by deducing and simplifying the LE-LPCCA as follows:

$$\max_{w_x, w_y} w_x^T X S_{XY} Y^T w_y$$
$$s.t. \quad w_x^T X S_{XX} X^T w_x = 1 \tag{3}$$
$$w_y^T Y S_{YY} Y^T w_y = 1$$

Where set $S_{XY} = D_{XY} - C_X \circ C_Y$, $S_{XX} = D_{XX} - C_X \circ C_X$, $S_{YY} = D_{YY} - C_Y \circ C_Y$. Symbol $\circ$ is operator, for the matrix of the same size $A, B \in R^{n \times n}$, $(A \circ B)_{ij} = A_{ij} B_{ij}$, A_{ij} is the element of the row i and column j of the matrix A. $D_{XX}(D_{YY}, D_{XY})$ is a $n \times n$ diagonal matrix. The ith of the diagonal elements equal to the sum of the i columns (or i rows) of $S_X \circ S_X (S_Y \circ S_Y, S_X \circ S_Y)$ matrix elements.

Finally, the optimization equations can be solved by using the Lagrange multiplier method and elimination method, and the mapping matrix is w_x and w_y.

LE-LPCCA wireless sensor localization algorithm does not need to go through the center of the processing. The positioning phase avoids the error caused by RSSI. Thereby, the positioning accuracy can be higher. However, it belongs to the centralized positioning method, which has high sensor energy consumption and iterative complex in modeling. Thus, it is not suitable for large-scale wireless sensor network. LE-DLPCCA wireless sensor localization algorithm based on learning algorithm is introduced in this paper. We cluster the positioning area, reduce the number of iterative training phase model and cost down the energy consumption of sensor.

3. LE-DLPCCA Algorithm

In this paper, the basic idea of LE-DLPCCA algorithm is to divide the area into several sub clusters and locate the unknown nodes in a distributed manner. First of all, we put forward a clustering algorithm based on energy equilibrium to dividing the monitoring area into many subdomains and electing the cluster head nodes. Then, we train the mapping between the signal space and physical space in each subdomain, and afterwards, unknown node estimates distance from the biggest similarity anchor nodes through the mapping. Finally, the cluster head nodes of subdomain perceive the invasion of unknown nodes, awaken the other cluster nodes and locate the unknown nodes through the maximum likelihood method.

3.1. Clustering Algorithm Based on Energy Balance

Clustering algorithm based on energy balance is introduced in this paper. Therefore, we can reduce the energy consumption of every beacon node. First of all, the whole wireless sensor network is divided into a number of square virtual sub region by using the GAF algorithm[21]. The sensor nodes judge node whether in the hotspots according to theorem 1.

Theorem 1: The X node coordinates are assumed as $(0,0)$, the hotspot is $M(x,y)$. If coordinates of any sensor nodes is $q(x,y)$ relative to the node X, wireless sensors possess the same radius of communication L, if and only if formula 4, so judge that sensor node $q(x,y)$ is in the hotspots.

$$\begin{cases} R \le L \\ r^2 + (2r)^2 \le L^2 \end{cases} \Rightarrow r \le \frac{L}{\sqrt{5}} \tag{4}$$

Where r is the length of grid side, R is the furthest distance of adjacent grid and L is the communication radius between sensor nodes. The virtual grid partition by GAF algorithm is shown in Figure1.
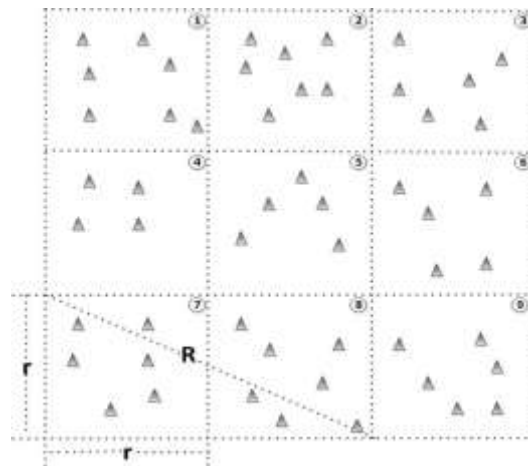

Figure 1. The Virtual Grid Partition by GAF Algorithm

Then, according to the restriction of self-energy, rewards and punishments signal generating based on surrounded environment. LE-DLPCCA algorithm elect the optimal cluster head node by way of competitive, the process of which is shown as follows.

Step1. The cluster node set T_{Sleep} as sleeping time. The size of T_{Sleep} and their self-energy must satisfy the following type.

$$T_{Sleep} = \frac{t_{balance}}{(E_s - E_{min})^\delta} \tag{5}$$

Where T_{Sleep} is the sleep time of the beacon node, E_s is the beacon node residual energy, E_{min} is the lowest energy of the beacon node work. δ is attenuation coefficient, which is influenced by regional environment beacon node layout. $t_{balance}$ is the factor of time balance.

If nodes receive the statement information of the cluster head node, transfer to the Step4 statement, or become a temporary cluster head and transfer to the Step2 Executive.

Step2. If the cluster node is elected as temporary cluster head node, sending cluster head statement information to cluster all nodes and receive the confirming information. If declarative information of other temporary cluster head nodes is received, we perform a comparison of energy. If the temporary cluster head node energy is low, sending a confirmation for cluster member acknowledgment message to the other cluster head nodes, then go to Step4, or enter the next step.

Step3. Temporary cluster heads receive confirmation information of the cluster members, elected as the cluster head node, and the nodes of the cluster members is bound to join.

Step4. If the cluster node receives the cluster head information statement, apply to join the cluster, and then enter a dormant state.

3.2. Construction of the LE-DLPCCA Model

In LE-DLPCCA model, the local structure of the data information is introduced into the CCA (Canonical Correlation Analysis). The original global nonlinear problem is transformed into several local linear problems, and then on these sub problems, we obtain multiple sets of sample projection base vectors through optimization calculation. Therefore, we can solve the nonlinear problem by the LE-DLPCCA model.

Due to the clustering in the wireless sensor network, we can construct a global model by combination of sub models. Compared with the global model, the data matrix of each sub model is rather smaller. We can improve the speed of modeling. The global model is shown as follow:

$$M = \begin{bmatrix} MAX_{1,1} & MAX_{1,2} & \dots & MAX_{1,n} \\ MAX_{2,1} & MAX_{2,2} & \dots & MAX_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ MAX_{n,1} & MAX_{n,2} & \dots & MAX_{n,n} \end{bmatrix} \tag{6}$$

According to the above description, each sub model may have the following optimization problems.

$$MAX_{i,j} = \begin{cases} \max_{w_x, w_y} w_{x(i,j)}^T X_{(i,j)} S_{XY(i,j)} Y_{(i,j)}^T w_{y(i,j)} \\ s.t. \quad w_{x(i,j)}^T X_{(i,j)} S_{XX(i,j)} X_{(i,j)}^T w_{x(i,j)} = 1 \\ \quad\quad w_{y(i,j)}^T Y_{(i,j)} S_{YY(i,j)} Y_{(i,j)}^T w_{y(i,j)} = 1 \end{cases} \tag{7}$$

Here we set each sub model $S_{XY} = D_{XY} - C_X \circ C_Y$, $S_{XX} = D_{XX} - C_X \circ C_X$, $S_{YY} = D_{YY} - C_Y \circ C_Y$. Symbol $\circ$ is operator, for the matrix of the same size $A, B \in R^{n \times n}$, $(A \circ B)_{ij} = A_{ij} B_{ij}$, A_{ij} is the element of the row i and column j of the matrix A. $D_{XX}(D_{YY}, D_{XY})$ is a $n \times n$ diagonal matrix. The ith of the diagonal elements equal to the sum of the i columns (or i rows) of $S_X \circ S_X (S_Y \circ S_Y, S_X \circ S_Y)$ matrix elements.

To solve the optimization problem within each cluster, the Lagrange function can be defined.

$$L = w_x^T X S_{XY} Y^T w_y - \frac{\lambda_1}{2} (w_x^T X S_{XX} Y^T w_x - 1) - \frac{\lambda_2}{2} (w_y^T X S_{YY} Y^T w_y - 1) \tag{8}$$

The equation (8) derivative, simplified, easy to get the generalized eigenvalue equation:

$$\begin{pmatrix} & XS_{XY}Y^T \\ YS_{YX}X^T & \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} XS_{XX}X^T & \\ & YS_{YY}Y^T \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix} \qquad (9)$$

Solve (9) to obtain each base vectors $(w_{x(i,j)}, w_{y(i,j)})$ within the cluster, carry out the linear change of signal intensity matrix and physical space matrix; obtain the signal strength or physical coordinates corresponding to the value.

Global base vectors obtained for:

$$(W_X, W_Y) = \begin{bmatrix} (w_{x(1,1)}, w_{y(1,1)}) & (w_{x(1,2)}, w_{y(1,2)}) & \cdots & (w_{x(1,n)}, w_{y(1,n)}) \\ (w_{x(2,1)}, w_{y(2,1)}) & (w_{x(2,2)}, w_{y(2,2)}) & \cdots & (w_{x(2,n)}, w_{y(2,n)}) \\ \vdots & \vdots & \ddots & \vdots \\ (w_{x(n,1)}, w_{y(n,1)}) & (w_{x(n,2)}, w_{y(n,2)}) & \cdots & (w_{x(n,n)}, w_{y(n,n)}) \end{bmatrix} \qquad (10)$$

3.3. The Estimation of Target Position

The above sub-section describes how to build the global model to create a suitable equation for the local topology structure between the signal and physical spaces. Specific steps of our algorithm are provided in the last paragraph of this section. Next, we will turn to the positioning phase.

In the positioning phase, sub cluster head nodes are aware of unknown nodes $g(u,v)$, wake up the cluster nodes and obtain data $\{X_{(i,j)}, Y_{(i,j)}\}$. We make the linear change for the known data $\{X_{(i,j)}, Y_{(i,j)}\}$, $\{Q^{X_{(i,j)}}, Q^{Y_{(i,j)}}\} = \{W_{X(i,j)}^T * X_{(i,j)}, W_{Y(i,j)}^T * Y_{(i,j)}\}$, so that the correlation between $Q^{X_{(i,j)}}$ and $Q^{Y_{(i,j)}}$ can reach the maximum. Then we look for K nearest neighbor nodes in the vector $Q^{X_{(i,j)}}$ through the calculation of Euclidean distance. From the characteristics of topology, the closer the data in signal space is, the shorter the corresponding distance of the physical space is. Therefore, we can find the physical coordinates of k neighboring nodes in the $Q^{Y_{(i,j)}}$ and obtain the corresponding coordinates $(u_1, v_1), (u_2, v_2), ..., (u_k, v_k)$.

In this paper, the point we final concern is, how to estimate the coordinate of the unknown node $g(u,v)$. Our algorithm adopts DV-Distance method [22] to solve it. Firstly, we calculate a correction ratio of polyline-distance between unknown node g and known node k based on $\lambda_{gi} = \sum L_{ji} / \sum D_{ji}, j = i+1, j = i$. Secondly, we compute the amended distance between itself and the known node based on $L_{gi} = \lambda_{gi} \times D_{gi}$. Finally, we use the maximum likelihood method to find the location of unknown node $g(u,v)$. The distributed localization is shown in Figure 2.
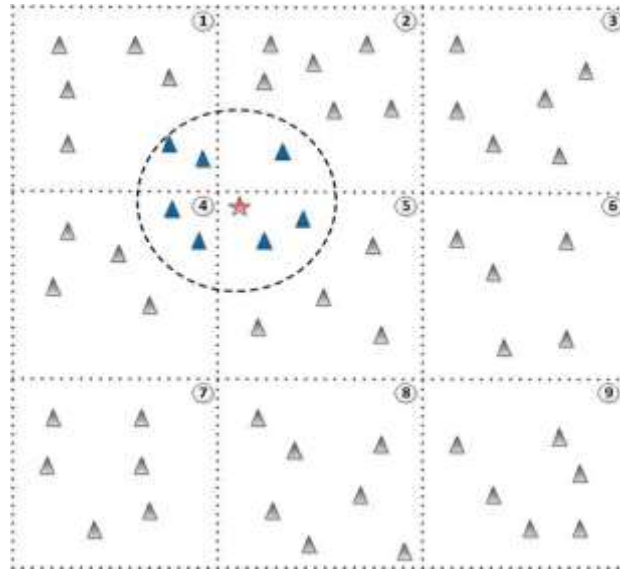
Figure 2. Distributed Localization in Wireless Sensor Network

Aiming at solve the location of $g(w, v)$, we can work it out by equation 11, which is shown as follows.

$$\begin{cases} (u_1 - w)^2 + (v_1 - v)^2 = L_{g1}^2 \\ \vdots \\ (u_k - w)^2 + (v_k - v)^2 = L_{gk}^2 \end{cases} \tag{11}$$

We transform the equation 11 into a matrix $(AW = b)$.

A is converted:

$$A = \begin{bmatrix} 2(u_1 - u_k) & 2(v_1 - v_k) \\ \vdots \\ 2(u_{k-1} - u_k) & 2(v_{k-1} - v_k) \end{bmatrix} \tag{12}$$

b is converted:

$$b = \begin{bmatrix} u_1^2 - u_k^2 + v_1^2 - v_k^2 + L_{gk}^2 - L_{g1}^2 \\ \vdots \\ u_{k-1}^2 - u_k^2 + v_{k-1}^2 - v_k^2 + L_{gk}^2 - L_{g(k-1)}^2 \end{bmatrix} \tag{13}$$

In following type 14, g is the physical coordinates of the unknown node.

$$g = \begin{bmatrix} u \\ v \end{bmatrix} \tag{14}$$

According to the standard minimum variance method, we compute the approximate physical coordinates of the unknown node $g(u, v)$, as shown in (15):

$$g = (A^T A)^{-1} A^T b \tag{15}$$

LE-DLPCCA specific steps are illustrated as follows, where Step1~Step3 is the training stage and Step4~Step7 is the positioning stage.

Step1. Calculate similarity matrix between S_{ij}^X and S_{ij}^Y (equation 1 and 2).

Step2. Through the GAF algorithm, the whole monitoring area is divided into clusters, and the cluster head node is elected by the energy balance algorithm.

Step3. Calculate the global mapping relation (W_X, W_Y) through performing typical correlation analysis on cluster signal strength and physical coordinates (equation 6 ~10).

Step4. The head node in the sub cluster perceives the unknown node, and the other nodes are awakened to get the data $\{X_{(i,j)}, Y_{(i,j)}\}$.

Step5. We do the linear variation for known data $\{X_{(i,j)}, Y_{(i,j)}\}$, $\{Q^{X_{(i,j)}}, Q^{Y_{(i,j)}}\} = \{W_{X(i,j)}^T * X_{(i,j)}, W_{Y(i,j)}^T * Y_{(i,j)}\}$, so that the correlation between $Q^{X_{(i,j)}}$ and $Q^{Y_{(i,j)}}$ reaches the maximum.

Step6. Find several nearest neighbor nodes in $Q^{X_{(i,j)}}$ by Euclidean distance calculation.

Step7. Calculate the physical coordinates of unknown nodes by the maximum likelihood method.

4. Simulation and Analysis

In order to verify the performance of the proposed LE-DLPCCA algorithm, two groups of data are used to validate the algorithm. The first group of data is supported by the Hong Kong University of science and technology department of computer science[23]. The experiment is made of Cross-bow MICA2 sensor nodes, including 8 AP nodes, which are deployed in different locations of the college building. Therefore, each sensor node receives signal strength of 8-dimensional vector. The device is an IEEE 802.11b wireless network with a 2.4GHz frequency bandwidth. All the data was collected by an IBM laptop connected to an external wireless USB network adapter. The second group of data is generated by simulation experiment in MATLAB2013a. In the simulation, all the sensor nodes are randomly deployed in the area of 500m * 500m. The communication radius of the sensor node is set to 50 meters.

The data set provided by the Hong Kong University of Science and Technology comprises two sets of data. The first set of data is the signal strength value of the 8 AP nodes collected by the mobile node at different time and different locations, the second set of data is the corresponding physical coordinate values; we collected a total of 2999 samples of data. We will use part of the data as training data, learning positioning algorithm. The rest are used as test data to calculate the positioning accuracy. In the second simulation, based on the RSSI signal propagation model, we generate two sets of data, including signal strength and physical coordinates. This paper collected a total of 3000 samples of data. The experimental mode is similar to the first group of data.

Figure 3 shows the effect of distribution the LE-DLPCCA algorithm. On the left is the physical coordinate distribution of the node, and the right is the signal strength value of the unknown node in the ellipse. The red dots are beacon nodes and the blue circles are unknown nodes.
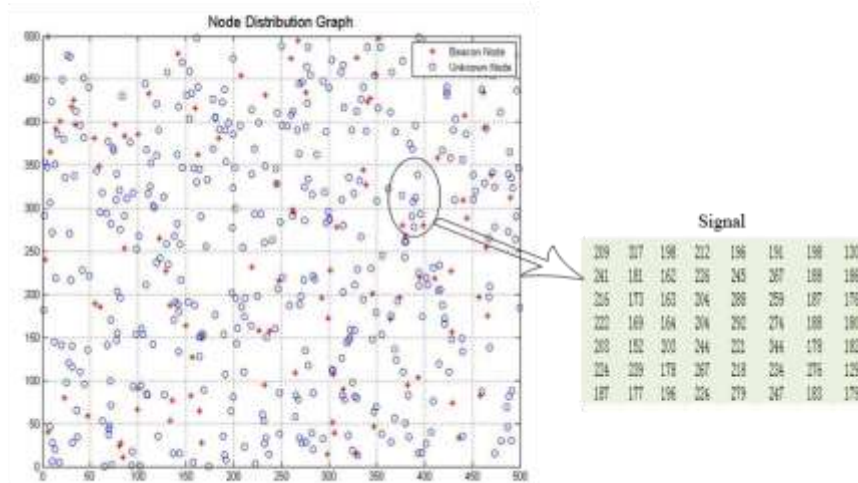
Figure 3. Distribution of Nodes in the LE-DLPCCA Algorithm

In two experiments, the LE-DLPCCA algorithm is compared with another two algorithms, namely, LE-KCCA algorithm and LE-LPCCA algorithm. LE-DLPCCA is the extension of the LE-LPCCA, which considers computational overhead and the energy consumption. LE-KCCA algorithm selected Gaussian kernel as kernel function, and kernel parameter was 0.05. The experimental environment was MATLAB2013a. In order to verify the validity of the algorithm, the experiment was run 50 times in two datasets, and the average was taken as the evaluation standard.

4.1. Comparison of Positioning Accuracy

Figure 4 shows the localization error of three kinds of localization algorithms. The training samples proportion is 30%, 40%, 50%, 60%, 70%, 80%. As the deployment of the sensor needs to be in accordance with local logistics conditions, the selection of two sets of data training samples is completely random.

The LE-DLPCCA algorithm performs better than the other two algorithms. In our scheme, when the training sample proportion comes up to 70%, the positioning accuracy is about 86%, and has a stable positioning effect. As the training sample proportion increases, the localization result is more stable. The LE-LPCCA algorithm has a high positioning accuracy when the proportion is large, which is about 85%. However, when the proportion is small, the positioning accuracy is poor compared to the LE-DLPCCA algorithm, because LE-LPCCA algorithm finally finds the unknown node through the centroid algorithm in the positioning phase. Nevertheless, LE-DLPCCA adopts a maximum likelihood method to improve the positioning accuracy. For the LE-KCCA algorithm, when the training samples proportion is 60%, the positioning result is better and the positioning accuracy is 80%. But with the increase of proportion, the positioning accuracy is reduced by about 1%. Mainly due to random selection of training samples, distribution cannot be fully guaranteed to be Gaussian distribution, and the selection of kernel and kernel parameters is not appropriate, affecting the stability of positioning results.
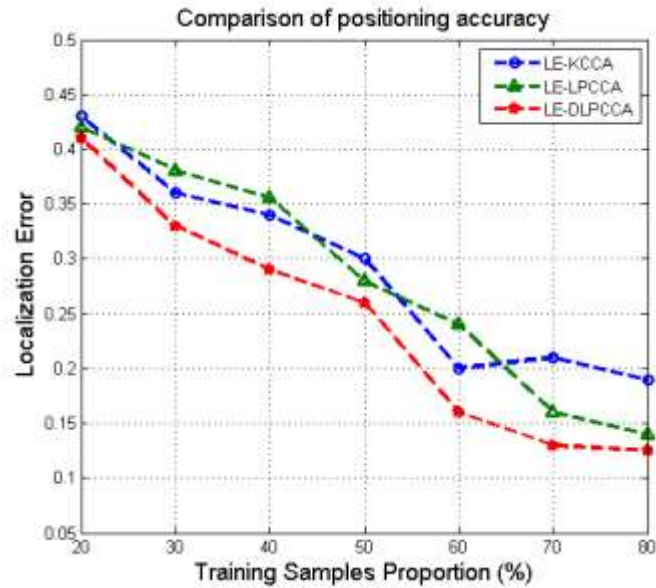
Figure 4. Localization Error under Different Training Samples Proportion

4.2. Comparison of Computational Overhead

In wireless sensor networks, localization algorithm based on machine learning is divided into the training stage and the positioning stage, the efficiency of modeling in the training stage is the key to machine learning algorithm.

In the figure 5, considering the increase of the training sample set, the influence on the model establishment time is considered. When the number of training sets reaches 3000, the computational overhead of LE-DLPCCA decreases by 8 times. In the process of modeling, the time cost is mainly to solve the generalized eigenvalue equation. The computation time of the similarity matrices and the positioning time are negligible. The dimension of the LE-KCCA and LE-LPCCA matrices are both 3000. But the dimension of LE-DLPCCA is only 250. Because we adopt a clustering to build sub models, which instead of a global model, the computational overhead performance of the modeling increases exponentially.
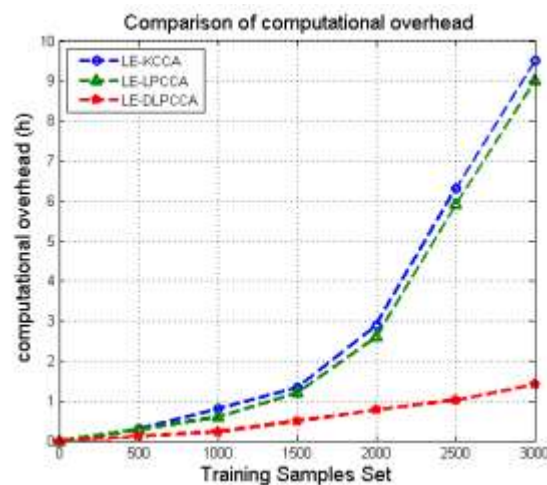

Figure 5. Computational Overhead of Model Building

4.3. Comparison of Positioning Energy Consumption

Because of the limited energy of the node, reducing the energy loss of the node has great significance to improve the connectivity and prolong the survival time of the network.

In experiment 3, since HKUST's dataset cannot determine the energy cost, we only simulate the second set of data sets. In Figure 6, as the number of acquisitions increases, the LE-KCCA and LE-DLPCCA algorithms are approximately linearly increasing in energy consumption. In contrast, LE-DLPCCA algorithm consumes less power, saving an average of 3 times the energy. Because in the positioning phase, only part of the members are in a wake-up state, we can reduce the energy consumption of other sensor nodes and improve the survival rate of the entire wireless sensor network nodes.

The average energy consumption is defined as

$$E_{avr} = \frac{\sum_{i=1}^{n}(E - E_s)}{n} \times 100\% \qquad E_s \text{ is residual energy} \qquad (16)$$
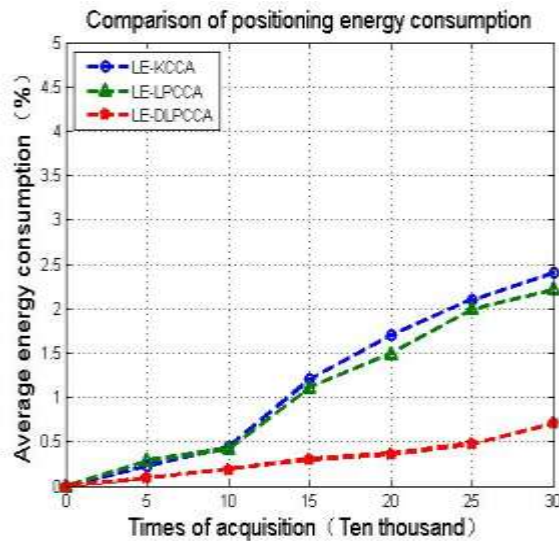

Figure 6. Positioning Energy Consumption

5. Conclusion

In this paper, we propose the LE-DLPCCA algorithm. The basic idea of LE-DLPCCA algorithm is to divide the area into several sub clusters and introduce the LE-LPCCA algorithm in each sub clusters to elect the beacon nodes in the wireless sensor network. LE-DLPCCA has many advantages, such as considering the local spatial information, distributed network, mining of the mapping between the two sets of data. The experiments have verified the efficiency of the proposed algorithm in positioning accuracy, the time of model building and energy consumption. The work has a wide application prospect.

The localization algorithm based on machine learning is easily affected by the external environment. In the training phase, the influence of temperature and humidity parameters on the model is not considered. This will be the subject of my future work.

Acknowledgments

References

[1] M. V. Ramesh, "Design, development, and deployment of a wireless sensor network for detection of landslides", Ad Hoc Networks, vol. 13, **(2014)**, pp. 2-18.

[2] J. Yick, B. Mukherjee and D. Ghosal, "Wireless sensor network survey", Computer networks, vol. 52, no. 12, **(2008)**, pp. 2292-330.

[3] F. B. Wang, L. Shi and Ren, "Self-localization systems and algorithms for wireless sensor networks", Ruan Jian Xue Bao(J Softw), vol. 16, no. 5, **(2005)**, pp. 857-68.

[4] S. Mahfouz, F. M. Chehade and P. Honeine, "Non-parametric and semi-parametric RSSI/distance modeling for target tracking in wireless sensor networks".

[5] J. Cheon, H. Hwang and D. Kim, "IEEE 802.15. 4 ZigBee-Based Time-of-Arrival Estimation for Wireless Sensor Networks", Sensors, vol. 16, no. 2, **(2016)**, pp. 203.

[6] S. Mohapatra, S. Behera and C. Tripathy, "A comparative view of AoA estimation in WSN positioning", Computational Intelligence in Data Mining. Springer, vol. 3, **(2015)**, pp. 123-33.

[7] S. I. Lopes, J. M. Vieira and J. Reis, "Accurate smartphone indoor positioning using a WSN infrastructure and non-invasive audio for TDoA estimation", Pervasive and Mobile Computing, vol. 20, **(2015)**, pp. 29-46.

[8] J. Rui and Y. Zhen, "An improved centroid localization algorithm based on iterative computation for wireless sensor network", ACTA PHYSICA SINICA, vol. 65, no. 3, **(2016)**.

[9] X. Li, L. Yan and W. Pan, "Optimization of DV-hop localization algorithm in hybrid optical wireless sensor networks", Journal of Heuristics, vol. 21, no. 2, **(2015)**, pp. 177-95.

[10] X. Li, L. Chen and J. Wang, "Fuzzy system and Improved APIT (FIAPIT) combined range-free localization method for WSN", KSII Transactions on Internet and Information Systems (TIIS), vol. 9, no. 7, **(2015)**, pp. 2414-34.

[11] G. Zanca, F. Zorzi and A. Zanella, "Experimental comparison of RSSI-based localization algorithms for indoor wireless sensor networks", proceedings of the Proceedings of the workshop on Real-world wireless sensor networks, F, ACM, **(2008)**.

[12] S. Y. Lee, J. W. Park and L. Barolli, "A localization algorithm based on AOA for ad-hoc sensor networks", Mobile information systems, vol. 8, no. 1, **(2012)**, pp. 61-72.

[13] H. Chen, K. Sezaki and P. Deng, "An improved DV-Hop localization algorithm for wireless sensor networks; proceedings of the Industrial Electronics and Applications", 2008 ICIEA 2008 3rd IEEE Conference on, F, IEEE, **(2008)**.

[14] X. L. Xu, F. Q. Zhang and S. B. Li, "Study on wireless sensor networks centroid algorithms based on APIT", Transducer and Microsystem Technologies, vol. 7, no. 18, **(2011)**.

[15] X. Shi, A. Yin and Q. Zhang, "Localization in wireless sensor networks based on K-Nearest Neighbo", Yi Qi Yi Biao Xue Bao/chinese Journal of Scientific Instrument, vol. 35, no. 10, **(2014)**, pp. 2238-47.

[16] F. Vanheel, J. Verhaevert and E. Laermans, "Pseudo- 3D RSSI- based WSN localization algorithm using linear regression", Wireless Communications and Mobile Computing, vol. 15, no. 9, **(2015)**, pp. 1342-54.

[17] K. Mao C. F., F. Ye, P. Wang and Q. Chen, "Node localization algorithm in wireless sensor networks based on SVM", Journal of Computer Research and Development, **(2014)**, pp. 2427-36.

[18] J. J. Gu, S. C. Chen and Y. Zhuang, "Localization in Wireless Sensor Network Using Locality Preserving Canonical Correlation Analysis", Journal of Software, vol. 11, no. 17, **(2010)**.

[19] J. J. Pan, S. J. Pan and J. Yin, "Tracking mobile users in wireless networks via semi-supervised colocalization", IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 3, **(2012)**, pp. 587-600.

[20] J. Chen, C. Wang and Y. Sun, "Semi-supervised Laplacian regularized least squares algorithm for localization in wireless sensor networks", Computer Networks, vol. 55, no. 10, **(2011)**, pp. 2481-91.

[21] S. Sen, D. Karmakar and S. Setua, "An power efficient algorithm for distributed ad-hoc cluster based Wireless Sensor Networks", proceedings of the Computer, Communication, Control and Information Technology (C3IT), 2015 Third International Conference on, F, IEEE, **(2015)**.

[22] Q. Wang, Y. Zhuang and G. U. J. Jing, "Target Positioning Algorithm Based on WSN in Perimeter Intrusion Detection", Computer Engineering, vol. 39, no. 9, **(2013)**, pp. 39-44.

[23] J. Gu and S. Chen, "Manifold- based canonical correlation analysis for wireless sensor network localization", Wireless Communications and Mobile Computing, vol. 12, no. 15, **(2012)**, pp. 1389-404.

Authors

Taotao Huang, he received the B.S. degree in Software Engineering from ANHUI University of Technology in 2013. Now he is a M.S. student in the College of Computer Science and Technology at Nanjing University of Aeronautics and Astronautics. The main research interests are wireless sensor network and machine learning.

Jingjing Gu, she received Ph. D. degree from College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics. Now she is an Associate Professor of the College of Computer Science and Technology at Nanjing University of Aeronautics and Astronautics. The main research fields are data mining and wireless sensor network.

Yi Zhuang, she was born in 1957. She graduated from the Department of Computer Science, Nanjing University of Aeronautics and Astronautics in 1981. Now she is a professor and Ph. D. supervisor of the College of Computer Science and Technology at Nanjing University of Aeronautics and Astronautics. Her research interests include network and distributed computing, dependable computing, information security.