

An Improved Dv-Hop Localization Algorithm Based On Machine Learning For Wireless Sensor Network

Rulin Dou and Xiaoyong YAN*

*College of Information Technology, Jingling Institute of Technology, Nanjing
211169, China(s)*

*drl@jit.edu.cn, xiaoyong_yan@jit.edu.cn**

Abstract

The DV-Hop localization method has a series of superiorities, such as high distributiveness and expandability, which is perfectly fit for large-scale deployment. Moreover, it may lead to reasonable positioning accuracy merely in isotropous dense network. In practical environment, most scenes are anisotropic, with unevenly distributed nodes. In this paper, kernel PCR method is applied to collect and utilize the correlation between hop count and real distance, so as to build an optimal relationship model, converting hop count information between nodes into the value of real distance, so that DV-Hop method may be applicable to different environment. Compared with existing similar and typical methods, the method proposed in this paper has higher environment adaptability, as well as higher positioning accuracy and stability.

Keywords: *Wireless Sensor Network; Range -Free Localization; Kernel Principal Component regression*

1. Introduction

Wireless sensor network(WSN)[1-2] refers to a new type network integrating communication, imbedding, sensor and other technologies altogether, which is able to sensing, collecting and processing the information of objects in the deployment area, as well as to converge the sensed information to information center via random self-organizing wireless communication network by multi-hop relay. It is different from traditional network, with a series of distinctive features, such as self-organizing and dynamic nature, centering on data, *etc.* Currently, the technology has been widely applied in military, environment monitoring and prediction, healthcare, and other fields [3-5]. Among all these applications, it is an important factor to determine the occurrence position of event. Show by related literatures[6], approximately 80% context sensing information is correlated with node position.

It is relatively a more simple method to determine the position of nodes by outfitting GPS to each node. However, owing to high cost, in large-scale WSN, it is actually impractical to deploy GPS on all nodes. By contrast, a more economical way is to deploy GPS on a small number of sensor nodes, so as to transform them into known nodes. On this basis, localization model or algorithm may be figured out according to location information of these nodes, so as to determine the location of unknown nodes. At present, plenty of researchers have put forward many models or algorithms for WSNnode localization. These algorithms may be divided into range-based algorithm and range-free algorithm[7,8]. The range-based algorithm has to measure the real distance or direction of adjacent nodes, so as to calculate the location of unknown nodes. This algorithm may help to improve the accuracy of distance. Yet, it

proposes higher requirement on nodes' hardware configuration, consumes more energy, and is easily affected by temperature, barrier and other environmental factors. As for this, it is not applicable to large-scale application. The range-free algorithm needs not to measure the real distance or direction between nodes. Instead, it determines the position according to network connectivity and other relevant information. Such algorithm is free from extra hardware support, leading to smaller communication consumption. Although the localization accuracy is low, compared with the rough precision of WSN, it is quite sufficient. As for this, range-free method is being increasingly noticed by researchers.

2. Background

DV-Hop algorithm [9,10] is a typical range-free location estimation algorithm, with high distributiveness and expandability. It is one of the distributive localization methods proposed based on distance vector routing and GPS localization ideology. The localization accuracy of DV-Hop algorithm mainly relies on the accuracy of estimated average distance of each hop. Compared with the real distance between nodes, there might be certain error. Moreover, topological structure of network may as well affect the localization accuracy. As for this, generally, DV-Hop algorithm is only applicable to isotropous dense network. The major reason leads to low localization accuracy of DV-Hop is fuzziness of hop distance. In other words, nodes with the same hop count may still come about different distances. Similarly, the same distance is not always created by the same hop count. Especially, in network with unevenly distributed nodes, average hop distance calculated with DV-Hop algorithm is hard to describe the average distance between nodes. Such uneven distribution is mainly resulted by anisotropic property of network topology.

For a WSN, it is assumed that there is a mapping function $f_h: \mathbb{R}^{2d} \rightarrow \mathbb{R}$, which describing the real distance and hop count mapping relationship between node pairs. Assuming that the measured distance from node c_i to node c_j is described as $h_{ij} = f_h(c_i, c_j)$, while the real Euclidean distance between nodes is d_{ij} , if $h_{ij} = f_h(c_i, c_j) = g_d(d_{ij})$, where $g_d: \mathbb{R} \rightarrow \mathbb{R}$ then such WSN is referred to as isotropy; or else, it is called anisotropy[3]. Isotropic network is quite rare in true environment. Instead, most existing network are anisotropic[11]. Anisotropy is mainly resulted by barriers, uneven distribution, and failure of certain nodes, which leads to significant holes in node deployment area. Figure 1 shows isotropous and anisotropic network topology.

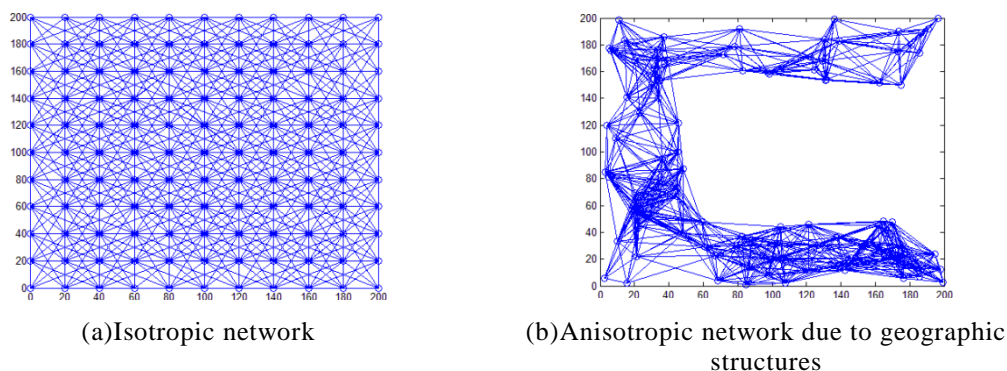


Figure 1. Sensor Network Topologies

In order to minimize localization performance reduction caused by anisotropy of network topology, Doherty and other scholars [12] took point-to-point communication links between nodes as geometrical limiting condition of node location. Moreover, such link relationship was further described as a set of convex set limiting condition, acquiring global optimal solution with semi-definite programming and linear programming methods, so as to figure out nodes' location. Convex programming is a centralized localization algorithm, with high computation cost. In order to improve the efficiency, beacon nodes shall be deployed on network borders. Or else, estimation of nodes' location may deviate towards the center of the network. Because of these problems, optimal localization method may not be taken as a feasible location estimation strategy.

In recent years, learning machine method [13-15] is adopted to excavate knowledge concealed in data collection, so as to figure out the localization model. This has become a development tendency of researches on localization of sensor network. For range-free method, learning machine method may be used to calculate the similarity between nodes based on their connectivity, i.e. training a prediction model according to the similarity or dissimilarity between nodes. On this basis, nodes' relative coordinates or absolute coordinates may be predicted with the prediction model. As the prediction model preserves topological information of network to the largest extent, the influence of anisotropy of network topology is greatly reduced[16-18]. In addition, localization algorithm based on learning is able to tolerate certain measurement noise. Some algorithms are even insensitive to measurement noise. As for this, there is no high requirement on applying which measurement technology to compute similarity or dissimilarity between nodes.

MDS-MAP(P)[19] is a typical learning-based localization method. Based on the original centralized MDS-MAP algorithm, a distributive localization strategy MDS-MAP(P) was developed. Drawing support from local sub-graph connectivity information, the method figures out relative coordinate sub-graph within its range with MDS-MAP method. Finally, the sub-graph is merged into a global graph. As MDS-MAP is not directly applied in the whole network, the performance of MDS-MAP(P) in anisotropic network topology is greatly improved. Such divide-and-conquer method improves localization accuracy to some extent. However, Method MDS-MAP is complicated in computation, and high in communication volume. Moreover, it is affected by the size of local area. Furthermore, when merging sub-graph into global graph, the method may as well be affected by accumulative error. HCQ(hop-count quantization) algorithm[20] is mentioned in literatures. The algorithm divides neighbor nodes within one hop's distance into three disjoint sets according to their practical hop count. Drawing support from certain calculation methods, hop count between nodes is corrected to a non-integer value. On this basis, multi-dimensional calculation method is adopted to figure out the location of unknown nodes according to the precise hop count. In order to avoid the size problem of local area, and to reduce the dependency of network deployment condition, Hyuk *et al.* proposed the PDM(Proximity Distance Map) algorithm [18]. The PDM correlated collected real distance with hop distance, so as to build an optimal linear conversion matrix T . Drawing support from the matrix T , hop distance between nodes is converted into estimated distance, so as to compensate measurement error caused by uneven distribution of nodes. In addition, in the conversion process, TSVD(Truncated Singular Value Decomposition) is applied for truncation to reduce the influence caused by noise and co-linearity. However, the fuzziness between hop distance and distance is not a linear relationship, but a nonlinear

relationship. For nonlinear problems, linear method is unable to correctly describe nonlinear structure in data. Researchers have found that, to build model with kernel trick [22, 23] is an effective solution. As is shown in Figure 2, a certain kernel function is adopted to map the raw data into proper high-dimensional feature space. As for this, nonlinear problem hard to be solved in the input space is converted into linear problem in feature space.

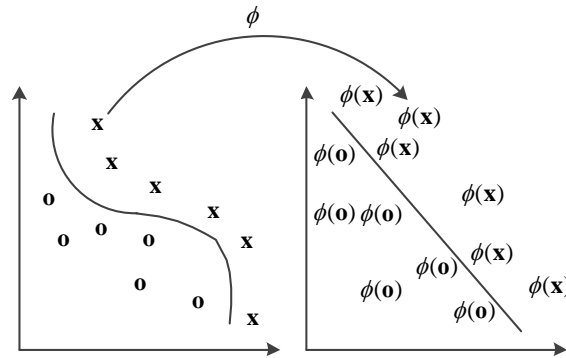


Figure 2. The Kernel Function ϕ Transform the Data into a Higher Dimensional Feature Space to Make it Possible to Perform the Linear Separation

Jaehun et al. [16, 17] put to use Kernel-based SVM(Support Vector Machine) regression method, which perfectly solves the hop-distance ambiguity problem. The algorithm perfectly solved the nonlinear relationship between hop-count and real distance, further improving the localization accuracy. SVM [23,24] is a learning algorithm developed by *Vapnik* based on statistical learning theory. It minimizes practical risk with structural risk minimization principle, with high generalization ability, and may be used to process high-dimensional small sample data. SVM training algorithm has to solve a linear constraint quadratic programming problem. However, *Jaehun* only considered utilizing kernel method to make nonlinear data linearly detachable, and that SVM regression (SVR) method based on traditional approach was adopted. SVR is traditionally used with only one output, which repeatedly constructs the relationship between distance and hop distance node by node, increasing the computation volume. In addition, in order to avoid collinearity problem in regressive process, traditional SVR method normalizes parameters manually, so that estimation method is hard to adapt to localization scale variation.

PCR(Principal component Regression) [25,26] is an approach similar to TSVD. It is actually a multiple regression method, putting to use truncation method to reduce the influence of noise and co-linearity. Yet, PCR is still different from TSVD. The major differences include: PCR employs PCA(Principal component Analysis) [26] as the pioneer to seek for linear regressor. After data centralization, PCA method measure data variance by identifying the so-called principal axis. Such principal axis shows the direction of maximum variance following the descending order of significance. After abandoning main direction when certain variances are lower than a certain threshold, PCR method is endowed with noise reduction ability, so as to reduce regressive variance's influence on regression prediction precision. As for this, PCR has high prediction ability. In addition, before PCA computation, data shall be standardized, *i.e.*,

being subtracted by their respective mean value, and being divided by their respective standard deviation, so as to eliminate data submergence caused by dimension difference. For range-free localization algorithm, data standardization is helpful in solving the inconsistency between hop-count and real distance, so as to make predicted value closer to practical prediction direction. Inspired by kernel method, Schölkopf *et al.*, [27] put forward KPCA(Kernel Principal component analysis). In high-dimensional space, samples have better linear divisibility, so that compared with PCA, KPCA has better identification performance. Based on KPCA, Rosipal *et al.*, [28] promoted linear PCR to nonlinear KPCR(Kernel Principal component Rgression). Similarly, KPCR has higher prediction precision than linear PCR. Owing to the superiorities of PCA and KPCA, they have been successfully applied in range-based localization approaches, with high localization accuracy achieved. However, they are seldom applied in range-free localization approaches, especially KPCR-based method.

The paper is devoted to research on range-free DV-Hoplocalization problems. In order to reduce the influence of network anisotropy range-free localization accuracy, a KPCR-based range-free localization method is raised, referred to as KPCR-DVHop. The algorithm reserves all advantages of the original DV-Hop method, which is able to effectively identify anisotropic network. The computation efficiency is higher than latest SVR-based localization method.

The rest of the paper is organized as fellows. Section 3 proposes the model and presents the localization scheme of KPCR-DVHop. Simulations are provided in Section 4 and 6 concludes the paper.

3. Relevant Theory Review

KPCR [25,29]is realized based on KPCA. Drawing support from KPCA method, the dimensionality is reduced. New variables obtained are taken as independent variables in multiple regressions, for regression calculation and analysis. It may as well be considered that, KPCR is to map data into feature space, and to perform PCR in the space. As for this, KPCA is the key to successful locazation. KPCA is an effective method to extract raw sample. It adopts kernel trick to stealthily project raw data into high-dimensional feature space, and then to realize data feature extraction in this high-dimensional space. In the following, we are to introduce the invention of kernel method, and then to discuss KPCA-related problems.

3.1. Kernel Trick

Kernel trick is an effective method to extract nonlinear feature of raw sample. It adopts kernel method to stealthily project raw data into high-dimensional feature space \mathcal{F} , and then to realize data feature extraction in this high-dimensional space.

In the year of 1995, Vapnik[24]put forward a theory based on statistical learning theory rule in small sample circumstance, which was an important development and supplement to traditional statistical learning, providing theoretical framework for learning theory and method with limited samples. With the theory as the foundation, a new universal learning algorithm is developed: SVM. Compared with previous approaches, SVM has plenty of theoretical and practical advantages. Initially, Support Vectors (SV) is used to solve mode identification problems, with the aim to find decision-making rules with better generalization performance. As a matter of fact, SV is a sub-set of training set. Optimal classification of SV is equal to classification of training set. At present, statistical learning theory and SVM have been taken as research hotspots by scholars in machine learning community.

Definition 1: For any two samples in the raw input space $x, y \in \Delta$, assuming that $\phi : \Delta \rightarrow \mathcal{F}$ is a mapping from a non-linearly detachable raw input space Δ to a linearly detachable high-dimensional feature space \mathcal{F} , if there is function κ complying with:

$$\kappa(x, y) = (\phi(x), \phi(y)) \quad (1)$$

Function κ is hereby referred to as inner product function or kernel function.

In general, kernel function has the following two features:

(1) Symmetry

$$\kappa(x, y) = \kappa(y, x) \quad (2)$$

(2) Satisfying Cauchy-Schwarz inequation

$$(\kappa(x, y))^2 \leq \kappa(x, x) \cdot \kappa(y, y) \quad (3)$$

The below two lemmas gives necessary and sufficient condition as a kernel function:

Firstly, kernel function is known to be symmetrical. Moreover, for any real vector $V = (v_1, \dots, v_m)^T$, there is:

$$V^T K V = \left\| \sum_{i=1}^m v_i \phi(x_i) \right\|^2 \geq 0 \quad (4)$$

Where, $K = (\kappa(x_i, x_j))_{i,j=1,\dots,m}$ is a matrix with element as $m \times m$; $x_i \in \Delta (i = 1, \dots, m)$; m refers to the number of samples. Hereby, the following lemmas are true:

Lemma 1: Assuming $\kappa(x, y)$ that is a real symmetric function in a finite space, $\kappa(x, y)$ is kernel function, if and only if $K = (\kappa(x_i, x_j))_{i,j=1,\dots,m}$ is positive semi-definite matrix.

More generally, according to Hilbert-Schmidt theory $\kappa(x, y)$ might be any symmetrical function satisfying the below ordinary condition:

Lemma 2 (Mercer theorem)[23]: Symmetrical Function $\kappa(x, y)$ under L_2 shall be ensured to be expanded as Positive Coefficient $\alpha_k > 0$:

$$\kappa(x, y) = \sum_{k=1}^{\infty} \alpha_k \phi_k(x) \phi_k(y) \quad (5)$$

In other words, $\kappa(x, y)$ describes an inner product in a certain feature space, the necessary and sufficient condition shall be:

$$\int g^2(u) du < \infty \quad (6)$$

For all $g \neq 0$, the below conditions shall be satisfied:

$$\iint \kappa(x, y) g(u) g(v) du dv > 0 \quad (7)$$

At present, frequently studied kernel functions are divided into three categories[22]. All of them may be matched up with existing methods.

(1) Inner product function adopting polynomial form, i.e.:

$$\kappa(x, y) = ((x \cdot y) + c)^q \quad (8)$$

(2) Inner product function adopting radial basis (RB), i.e.:

$$\kappa(x, y) = \exp\left(-|x - y|^2 / 2\sigma^2\right) \quad (9)$$

(3) Sigmoid inner product function, such as:

$$\kappa(x, y) = \tanh(v(x, y) + c) \quad (10)$$

In the aforementioned three commonly applied kernel functions, Parameter q, σ, v, c , are constant. Selection of these constants is based on experience. Currently, selection of these parameters is lack of an effective and universal standard. As is known to all, selecting an

appropriate parameter is quite significant to the solution of problem. Model selection technology provides a principle for selection of kernel parameter. Gaussian kernel function reserves the distance similarity of input space. In this paper, Gaussian kernel function is adopted to calculate the similarity between nodes.

3.2. Kernel Principal Component Analysis

The basic ideology of KPCA: firstly, nonlinear mapping is adopted to project raw sample nonlinearly detachable from input space into a linearly detachable high-dimensional (or even infinite-dimensional) feature space. On this basis, principal component analysis is to be performed in this new space. In order to avoid linearly un-detachable problem, kernel technology in SVM is introduced, *i.e.*, to replace inner product computation of sample in feature space with kernel function satisfying Mercer condition. New, the principal component analysis algorithm is described as follows:

A group of training data $X = [x_1, \dots, x_m]$ in raw space is given, while the corresponding covariance matrix shall be:

$$S_t^\phi = \frac{1}{m} \sum_{i=1}^m (\phi(x_i) - m_0^\phi) (\phi(x_i) - m_0^\phi)^T \quad (11)$$

Where, $\phi(x_i)$ refers to corresponding data of x_i in feature space via mapping function ϕ ; m_0^ϕ is the mean value of all samples in the high dimensional space \mathcal{F} , with the expression form shown in Equation 12.

$$m_0^\phi = \frac{1}{m} \sum_{i=1}^m \phi(x_i) \quad (12)$$

Similar to the extraction process of linear PCA feature, our purpose is to extract feature vector corresponding to non-zero eigenvalue, so as to constitute the projection space. Yet, it is quite complicate to realize data centralization with the above formula. Schölkopf put forward a solution method assuming that the data had already been centralized. As for this, covariance matrix in high-dimensional space may be figured out via the below formula:

$$S_t^\phi = \frac{1}{m} \sum_{i=1}^m \phi(x_i) \phi(x_i)^T \quad (13)$$

Order:

$$Q = [\phi(x_1), \dots, \phi(x_m)] \quad (14)$$

At this point, covariance matrix S_t^ϕ may be converted to the below form:

$$S_t^\phi = \frac{1}{m} Q Q^T \quad (15)$$

We firstly define the below $m \times m$ kernel matrix $K = Q^T Q$. Elements in the matrix may be acquired via kernel trick:

$$K_{ij} = \phi(x_i)^T \phi(x_j) = (\phi(x_i) \cdot \phi(x_j)) = \kappa(x_i, x_j) \quad (16)$$

Through calculation, feature vector v_1, v_2, \dots, v_m corresponding to the first m largest non-zero eigenvalue $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ in matrix K may be figured out. As for this, orthogonal eigenvector $\varphi_1, \dots, \varphi_l$ corresponding to the first non-zero eigenvalue $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l$ in covariance matrix S_t^ϕ shall be:

$$\varphi_i = \frac{1}{\sqrt{\lambda_i}} Q v_i \quad i = 1, 2, \dots, l \quad (17)$$

Now, the problem is to acquire the eigenvector in Formula (18). Actually, if matrix K is centralized, we may get:

$$\tilde{K} = K - 1_m \cdot K - K \cdot 1_m + 1_m \cdot K \cdot 1_m \quad (18)$$

Where, 1_m is unit matrix with the element as $1/m$.

If having obtained eigenvector of \tilde{K} , it is possible to obtain eigenvector ϕ_1, \dots, ϕ_m of S_t^ϕ . As for this, for a new sample x , projection realized by KPCA is shown below:

$$y = P^T \phi(x) \quad (19)$$

Where $P = [\phi_1, \dots, \phi_m]$

Putting Formula (15) and Formula (18) into the above Equation (19), we may obtain:

$$\begin{aligned} y = P^T \phi(x) &= \left[\frac{v_1}{\sqrt{\lambda_1}}, \dots, \frac{v_m}{\sqrt{\lambda_m}} \right]^T Q \phi(x) \\ &= \left[\frac{v_1}{\sqrt{\lambda_1}}, \dots, \frac{v_m}{\sqrt{\lambda_m}} \right]^T [\kappa(x_1, x), \dots, \kappa(x_m, x)]^T \quad i = 1, 2, \dots, m \end{aligned} \quad (20)$$

As for this, in the new feature space, KPCA has all mathematical feature of PCA, as well as other unique superiorities:

(1) When principal component of the same amount is adopted, KPCA leads to better identification performance than PCA.

(2) KPCA may further improve the identification performance by providing more components than linear situation. In other words, KPCA is able to extract components exceeding the dimensionality of raw input space. Assuming that the number of sample m is larger than the dimensionality d of raw input space, linear PCA may only figure out non-zero eigenvalue. By contrast, KPCA is able to extract non-zero eigenvalue, which is impossible to linear PCA.

(3) Computation complexity of KPCA may not increase along with the rapid growth of dimensionality converted space. It is only related with the dimensionality of raw input space, while irrelevant with the dimensionality of converted space.

(4) Different from other nonlinear PCA, the essence of KPCA is to figure out the eigenvalue and eigenvector of matrix $K = (\kappa(x_i, x_j))_{i,j=1,\dots,m}$ (m refers to the number of training sample), which does not involve nonlinear optimization.

Owing to the above superiorities, KPCA is perfect applied in mode identification, data compression and other related fields. However, we know that, matrix K in feature space \mathcal{F} is $am \times m$ matrix. When the number of training sample m is large, the computation efficiency of KPCA will be greatly improved. Currently, plenty of literatures have made explanation on this problem via different angles. In addition, *Suykens* put to use the ideology of least square SVM classifier, explaining PCA and KPCA from the angle of constrained optimization, providing us with a brand new knowledge on PCA and KPCA. Traditional PCA is lack of probability model structure, while such structure is quite significant to mixed model and BA decision-making. On the other hand, traditional PCA is only able to extract second order information. *Zhou* studied KPCA from the angle of probability, combining probability PCA (PPCA) with KPCA, and put forward the method of KPPCA, which overcame the two disadvantages of PCA. At present, kernel principle component analysis has become an effective method for kernel-based feature extraction.

4. Localization Algorithm with Kernel Principal Regression

4.1. Problem Statement

The paper is mainly designed to study location estimation of n sensor nodes in two-dimensional area. Assuming that there are n nodes deployed in the area, listed as X_1, \dots, X_n . ID of these nodes is separately $1, \dots, n$. The real coordinates of node $X_i (i \in n)$ shall be $c_i (c_i \in \mathbb{R}^2)$. Assuming that the first $m (m \ll n)$ nodes in n are beacon nodes, while the rest $n - m$ nodes are unknown nodes, their coordinates have to be estimated via localization algorithm, so as to make estimated coordinates closer to the real coordinates of unknown nodes.

DV-Hop localization algorithm [9,30] is range-free algorithm proposed by *Dragos Niculescu et al. of Rutgers University*. Its localization principle: firstly, typical distance vector exchange protocol is applied to figure out the hop count all nodes from known nodes. Having obtained the location of known nodes and the hop count, known nodes will calculate network average hop distance via $\phi_i = \sum_{j=1}^m d_{ij} / \sum_{j=1}^m h_{ij}$. Where, d_{ij} and h_{ij} are separate beacon node i to beacon node. When receiving correction value, unknown nodes will be able to figure out its distance $\hat{d}_{si} = \phi_i h_{si}$ from known nodes according to hop count h_{si} and correction value \hat{d}_{si} . When unknown nodes have acquired the distance to three or more known nodes, trilateration localization shall be performed in the third phase. As for this, two kind of information is mainly considered in this paper: hop count and real Euclidean distance between nodes. Assuming that measured data and real data collected by sensor node $i (i \in m)$ may be separately described by two groups of data sets, denoting $\mathbf{h}_i = [h_{i1}, h_{i2}, \dots, h_{im}]^T$ as the minimum hop count to m known nodes, while $\mathbf{d}_i = [d_{i1}, d_{i2}, \dots, d_{im}]^T$ as the real distance between corresponding nodes. After a period of time, two data matrixes may be obtained between beacon nodes, i.e. minimum hop count matrix $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m]$ and real Euclidean distance matrix $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m]$.

KPCR-DV Hop method put forward in this paper is improved based on PDM method. According to PDM method, there is certain relationship between hop count and real Euclidean distance of beacon nodes. However, affected by environment, such relationship is nonlinear. According to kernel learning principle, data is projected into high-dimensional feature space \mathcal{F} (with the dimensionality as $M, M \leq \infty$), while the originally non-detachable data becomes detachable. Drawing support from mapping function Φ , dimensionality of hop count matrix \mathbf{H} is increased to feature space, obtaining $\Phi(\mathbf{H}) = (\phi(\mathbf{h}_1), \dots, \phi(\mathbf{h}_m))^T$. Assuming that the data has been centralized, i.e. $\sum_{i=1}^m \phi(\mathbf{h}_i) = 0$; Then, the relationship between hop count and real distance shall be described as:

$$\mathbf{D} = \Phi(\mathbf{H})\boldsymbol{\eta} + \boldsymbol{\epsilon} \quad (21)$$

Where, $\boldsymbol{\eta} = (\eta_1 \eta_2 \dots \eta_M)^T$ refers to regression coefficient vector; $\boldsymbol{\epsilon}$ refers to random error vector in feature space.

In order to figure out the optimal relationship between hop count and real distance, the equation has to figure out the optimal estimation value $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$. As for this, $\|\boldsymbol{\eta}\|^2$ shall be minimized. When $\|\boldsymbol{\eta}\|^2$ is minimized, we may get:

$$\Phi^T \Phi \hat{\boldsymbol{\eta}} = \Phi^T \cdot \mathbf{D} \quad (22)$$

For Equation (22), there might be certain linear correlation between some vectors and other vectors in the same feature space. At the moment, their contribution to regression is limited. In addition, there might be noise in feature space as well. As for this, estimated regression coefficient may be acquired by calculation. Yet, the given model parameter is quite sensitive to sample data. In order to reduce the dimensionality for variable calculation, an effective way to reduce regression calculation is to perform principal analysis on independent

variable, using the obtained principal component to execute regression, so as to eliminate the influence caused by correlation between independent variables.

4.2. Localization Algorithm With KPCR

In learning-based WSN localization mechanism, estimation process of nodes is often comprised by two phases: offline training phase and online localization phase[31]. In offline training phase, the mapping relation is figured out according to the measured data (hop count) and distance of known nodes, so as to build localization model. In online localization phase, the obtained mapping is applied to estimate the location of unknown nodes. As KPCR is regression method based on KPCA, when having acquired the first l principal components, Equation (21),(22) regression coefficient vector may be obtained.

$$\hat{\boldsymbol{\eta}} = \mathbf{P}_l(\mathbf{T}_l^T \mathbf{T})^{-1} \mathbf{T}_l^T \mathbf{D} \quad (23)$$

Where, $\mathbf{P}_l = \Phi^T \mathbf{A}_l$, while the value of $\mathbf{A}_l = [\varphi_1, \dots, \varphi_l]$, φ_i is shown in Equation 17. Equation 23 may as well be described as:

$$\hat{\boldsymbol{\eta}} = \Phi^T \mathbf{A}_l (\mathbf{T}_l^T \mathbf{T})^{-1} \mathbf{T}_l^T \mathbf{D} \quad (24)$$

In the training phase, KPCR method may be used to figure out the relationship between hop count and distance of known nodes (Equation 24). At the moment, according to the model built in training phase, it is possible to figure out the prediction equation for the distance from unknown nodes to known nodes.

$$\begin{aligned} \hat{\mathbf{D}} &= \Phi \hat{\boldsymbol{\eta}} \\ &= \Phi \Phi^T \mathbf{A}_l (\mathbf{T}_l^T \mathbf{T})^{-1} \mathbf{T}_l^T \mathbf{D} \\ &= \mathbf{K} \mathbf{A}_l (\mathbf{T}_l^T \mathbf{T})^{-1} \mathbf{T}_l^T \mathbf{D} \end{aligned} \quad (25)$$

From Equation (25), we may find that, in each principal component analysis, it is only necessary to extract the first l eigenvectors in kernel matrix \mathbf{K} , so as to figure out corresponding \mathbf{A}_l and there is no need to calculate \mathbf{P}_l and $\hat{\boldsymbol{\eta}}$ Equation 25 may as well be described as:

$$f(\mathbf{h}) = \sum_{i=1}^m \eta_i \mathbf{K}(\mathbf{h}, \mathbf{h}_i) \quad (26)$$

In online localization phase, when unknown nodes have obtained the hop count to known nodes, the prediction model (Equation 25 or 26) will be applied to estimate the distance to known nodes. Before the estimation, kernel function $\mathbf{K}(\mathbf{h}, \mathbf{h}_i)$ shall be centralized with the below method:

$$\hat{\mathbf{K}} = (\hat{\mathbf{K}} - (\mathbf{1}/m) \mathbf{1}_{n-m} \mathbf{1}_m^T \bar{\mathbf{K}}) (\mathbf{I}_m - (\mathbf{1}/m) \mathbf{1}_m \mathbf{1}_m^T) \quad (27)$$

In the equation, $\bar{\mathbf{K}}$ refers to kernel matrix on building of the model, $\hat{\mathbf{K}}$ stands for kernel matrix obtained with the measured distance between unknown nodes and known nodes, $\mathbf{1}_{n-m}$ is full $\mathbf{1}$ column vector of $n - m$ dimensionalities. When the hop count between unknown nodes and known nodes is put into Equation (25-27) to estimate corresponding physical distance of unknown nodes, trilateration or polygon method will be applied to estimate the coordinate location of unknown nodes.

5. Performance Evaluation

One of the important features of range-free localization method is that, it is quite fit for large-scale deployment. This requires thousands of sensor nodes, while in labs, it is difficult to realize such large-scale real network. As for this, in researches on large-scale range-free

node localization algorithm, software simulation is often applied to estimate the advantages and disadvantages of localization algorithm.

In this section, the performance of KPCR-DVHop algorithm is to be analyzed and assessed via simulation experiment. Matlab2013b software is employed to analyze and compare methods proposed in this paper. In the experiment, all nodes are evenly distributed in two-dimensional space. In order to reduce the one-sidedness of single experiment, each deployment goes through 50 simulations while nodes in each experiment are randomly re-distributed in the experiment area. Mean value of 50RMS(Root Mean Squares)[32] is taken as the assessment basis.

In order to assess the performance of methods proposed in this paper, nodes are assumed to be randomly or regularly deployed in the monitoring area. In addition, in order to evaluate the adaptability of the proposed methods to network topology anisotropy, obstruction is added in the aforementioned two deployment strategies, i.e. assuming that there is a large obstruction in the deployment area, impeding the direct communication between nodes. Such area is of C-shape. In allusion to different network topology structure, nodes are re-deployed in the same area for several times, assessing the average localization error. This experiment also compare our method with three previous methods: (1)The classic DV-Hop method proposed in[9]; (2) PDM proposed in[18];and (3) SVR proposed in [16]in two group experiments.Furthermore, for fairness, in PDM localization, we denoted abandoning threshold in TSVD as 3, i.e. abandoning eigenvectors with eigenvalue less than or equal to 3. There is certain relationship between kernel parameter σ and the distance between training samples. In the experiment, we assume σ as 50 times of the average distance between sample nodes. Configuration of C and ϵ in SVR method uses for reference related literature[33], while C is also configured based on σ according to related literature[34].

5.1. Localization Results with Regularly Deployed Sensors

The main purpose of regular deployment of nodes in the monitoring area is to investigate the influence of beacon node collinearity on localization accuracy. C-shaped deployment is designed to test the influence of none-line-of-sight problem on localization accuracy.

In the experiment, nodes are regularly deployed in $195\text{units} \times 195\text{units}$, while the side length of grid is 15units . Without obstruction, have totally 196 nodes. Obstruction $150\text{units} \times 75\text{units}$ is deployed in C-shaped area, while the number of nodes is changed to 156. 10-20 nodes are selected from these nodes as beacon nodes, while location of these beacon nodes is assumed having been given. Fig.3 shows the final localizationresult of 12 beacon nodes in a square area. Circles refer to unknown nodes, while squares refer to beacon nodes. Straight lines connect the real coordinates and estimated coordinates of unknown nodes. Fig.3(a) shows a certain localization result with regularly deployed 12 beacon nodes, and the RMS is 3.75. Fig.3(b) shows a certain localization result with regularly deployed 12 beacon nodes in C-shaped area, while the R is 4.005.

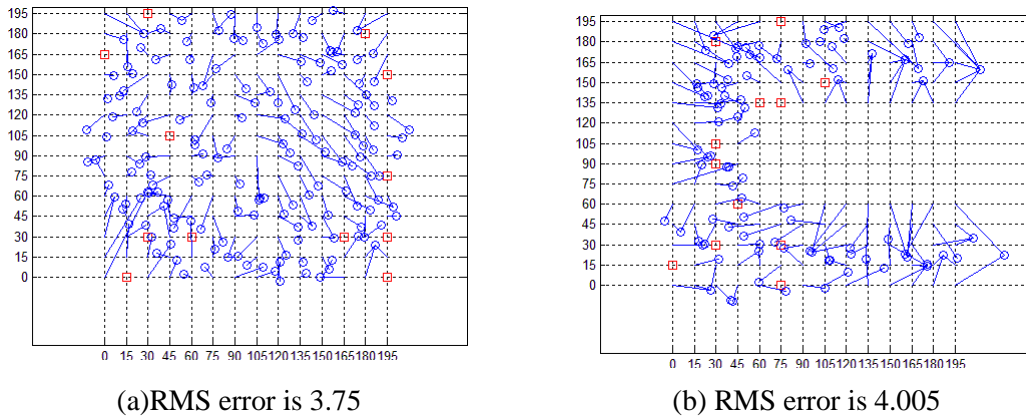


Figure 3. Localization Results with Regularly Deployed in 2D Environment

Figure 4(a-b) shows RMS variation curve of 4 localization methods, with the number of beacon nodes varying from 10 to 20. Figure a is RMS variation curve for regular deployment in square area. Figure b is R variation curve for regular deployment in C-shaped area. It may be easily seen from Figure 4(a-b) that, RMS value of DV-Hop is the highest, while the curve fluctuates up and down. RMS value of C-shaped area is higher than square area. There are mainly three reasons leading to the result:

1. DV-Hop algorithm is a localization method using for reference distance vector routing, replacing straight line distance between nodes with hop distance between nodes. As for this, the localization performance is high only when nodes are evenly distributed.
2. The final estimation of DV-Hop relies on LS(Least Squares). LS is an unbiased estimation method, quite sensitive to ill-posed problem. For location estimation algorithm, co-linearity or approximate co-linearity of beacon nodes for localization may lead to major estimation error. Increment of beacon nodes also improved the probability of beacon node co-linearity or approximate co-linearity.
3. DV-Hop method takes advantage of LS to figure out the regression coefficient between hop distance and coordinate. LS method applied has not yet performed standardization operation on data. As for this, direct computation between data of different dimensionality may inevitably lead to data submergence, and the estimation value is inevitably unstable.

Because of the aforementioned three problems, as for DV-Hop algorithm, in regular deployment, the RMS value of square area is larger than 4, while the RMS value of C-shaped area is larger than 6 or even close to 7.

From Fig.4, we may find that, in allusion to the rest improved DV-Hop methods, in regular deployment, the RMS value of square area varies from 3 to 4, while the RMS value of C-shaped area is most between 5 and 4. Where, the RMS descending tendency of RMS value of PDM method is obviously smaller than that of DV-Hop algorithm along with the increment of beacon nodes. PDM method makes use of T to construct the relationship between hop count and distance. To some extent, TSVD compensates distance estimation in network with unevenly distributed network. However, TSVD is a linear method, which is difficult to realize optimal conversion matrix in nonlinear hop count and distance relationship. In addition, P also ignores the dimensionality factor between hop count and distance. The method based on

SVR, as well as KPCr-DVHop method proposed in this paper decreases along with the decrement of beacon nodes, while the RMS value is smaller than that of PDM method. Localization method based on SVR and KPCR-DVHop method are based on kernel learning. To some extent, the two methods may effectively differentiate the fuzzy relationship between hop count and distance. SVR method partition data by constructing optimal classification hyper-plane, which is a multi-input and single-output method. However, for range-free localization methods, there is multi-value to multi-value relationship between hop count and distance. When adopting SVR, correlation between multiple values is not considered, while parameter adjustment is quite complicated. KPCR is a multiple regression method, and its regression takes into consideration the correlation between parameters, while parameter adjustment is quite simple. As for this, localization performance of such KPCR-DVHop method is better than SVR.

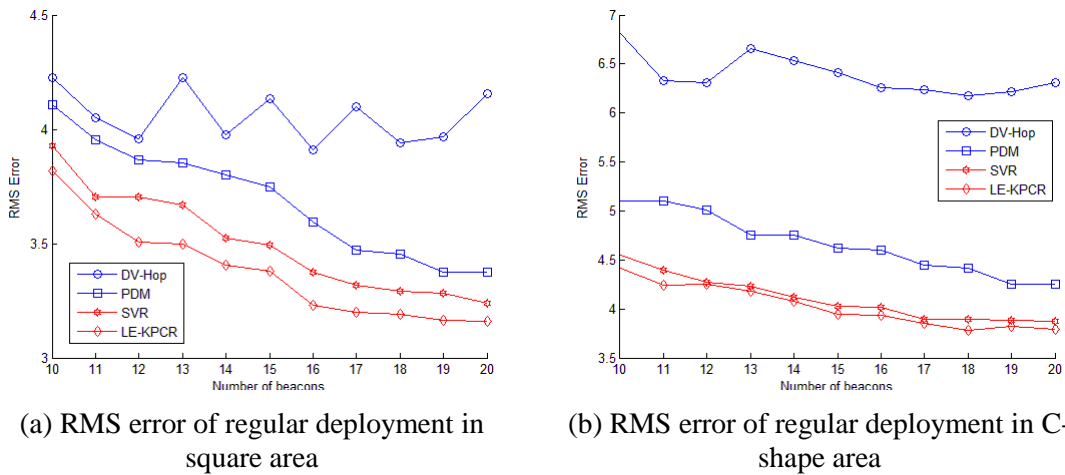


Figure 4. Simulation Results of the Regularly Deployed Sensor Network with Different Number of Beacons

5.2. Localization Results with Randomly Deployed Sensors

As random deployment is more close to practical situation, the experiment is mainly designed to test whether the algorithm is applicable to different real situations. Similarly, random deployment experiment also includes two situations, C-shaped area with obstruction and square area without obstruction. In the two experiments, 200 nodes are randomly deployed in 200units × 200units monitoring area. Similar to regular deployment, compare DV-Hop, PDM and SVR three localization methods, so as to evaluate RMS variation of the four algorithms along with the change of node quantity. In situation with obstruction, an 150units × 75units object is placed in the deployment area, so as to manually impede communication between nodes in this area. 10 to 20 beacon nodes are selected from these nodes as beacon nodes, and their location is assumed to be known.

Figure 5(a) shows the final localization result of 12 beacon nodes in a square area. Figure 5(b) shows a certain localization result with regularly deployed 12 beacon nodes, and the RMS is about 3.371; Figure 5(c) shows a certain localization result with regularly deployed 12 beacon nodes in C-shaped area, while the RMS is about 4.05.

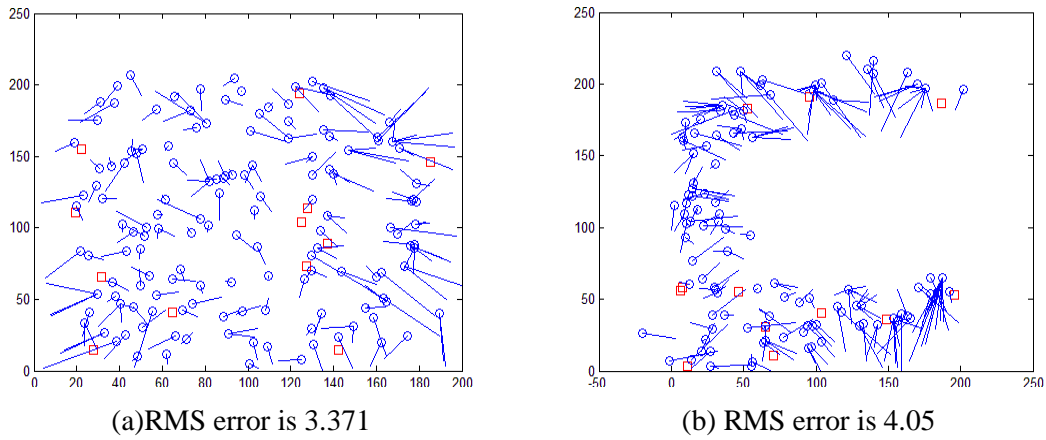


Figure 5. Localization Results with Randomly Deployed in 2D Environment

Figure 6(a-b) shows RMS variation curve of four localization methods, with the number of beacon nodes varying from 10 to 20. Figure 6(a) is RMS variation curve for random deployment in square area. Figure 6(b) is RMS variation curve for random deployment in C-shaped area. Similarly, nonlinear relationship between hop count and distance, beacon node collinearity and unification of different dimensionality are not considered, so that the localization result of DV-Hop is poor and unstable. Especially in C-shaped area, network topology anisotropy even enlarged the RMS of DV-Hop. The rest three methods are more or less improved than DV-Hop algorithm. Compared with Figure 4, their RMS values are more closer, decreasing gradually with the increment of beacon node quantity. SVR method, also based on kernel method, is superior to linear PDM method. KPCR-DVHop, taking into consideration data correlation, is superior to SVR.

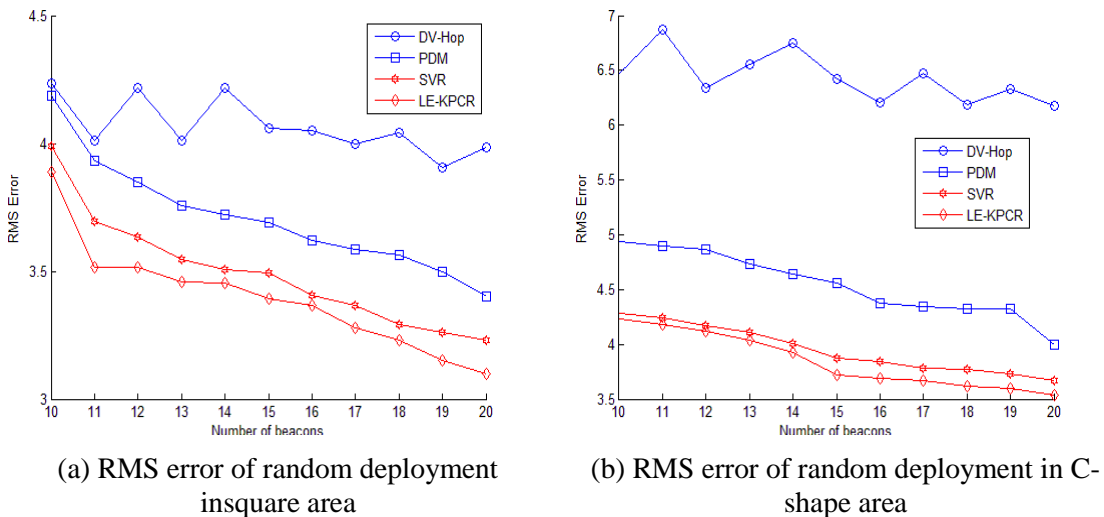


Figure 4 Simulation Results of the Randomly Deployed Sensor Network with Different Number of Beacons

6. Conclusion

In this paper, a KPCR-DVHop range-free localization model is proposed. The algorithm fully makes use of hop count and real distance information of known nodes in WSN. Moreover, kernel PCR method is applied to measure covariance between variables. On the basis of reserving the required identification information, the algorithm is able to effectively eliminate the correlation between information, as well as to describe samples with few features. On the basis of eliminating noise between variables, the method is able to avoid ill-posed problem, which may play an effective role in WSN environment with major possible errors in data mining, so as to overcome issues of existing algorithm or model, such as low localization accuracy, and poor robustness, etc.

Compared with similar researches, KPCR-DVHop range-free localization model shows high localization accuracy and stable performed in different network topology situations.

Acknowledgements

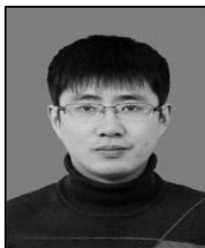
The paper is sponsored by Provincial University Natural Science Research Foundation of Jiangsu Education Department (12KJD510006,13KJD520004); Natural Science Foundation of Jiangsu (BK2012082); Doctoral Scientific Research Startup Foundation of Jinling Institute of Technology (JIT-B-201411); Educational Reform Foundation of Jinling Institute of Technology (2010JGXM-02-2).

References

- [1] I. Emary and S. Ramakrishnan, "Wireless Sensor Networks: From Theory to Applications", CRC Press, United Kingdom, (2013).
- [2] S. Khan, A. S. Pathan and N. A. Alrajeh, "Wireless Sensor Networks Current Status and Future Trends", CRC Press, United Kingdom, (2013).
- [3] S. Nikolettseas and J. Rolim, "Theoretical Aspects of Distributed Computing in Sensor Networks", Springer, New York, (2011).
- [4] P. Morreale, F. Qi and P. Croft, "A green wireless sensor network for environmental monitoring and risk identification", International Journal of Sensor Networks, vol. 10, (2011), pp. 73-82.
- [5] M. Bocca, L. M. Eriksson, A. Mahmood and R. Jäntti, *et.al.*, "A Synchronized Wireless Sensor Network for Experimental Modal Analysis in Structural Health Monitoring", Computer-Aided Civil and Infrastructure Engineering, vol. 26, (2011), pp. 483-499.
- [6] G. Cabri, L. Mamei and M. Zambonelli, "Location dependent services for mobile users", IEEE Transaction on Systems Man and Cybernetics, part A, vol. 33, no. 6, (2003), pp. 667-681.
- [7] Z. Chen, F. Xia, T. Huang and H. Wang, "A localization method for the Internet of Things", The Journal of Supercomputing, vol. 63, no. 3, (2013), pp. 657-674.
- [8] Y. Liu and Z. Yang, "Location, Localization and Localizability Location-awareness Technology for Wireless Networks", Springer, (2011).
- [9] D. Niculescu and B. Nath, "Based Positioning in Ad Hoc Networks", Telecommunication Systems, vol. 22, no. 1-4, (2003), pp. 267-280.
- [10] G. Wu, S. Wang, B. Wang and Y. Dong, *et.al.*, "A novel range-free localization based on regulated neighborhood distance for wireless ad hoc and sensor networks", Computer Networks, vol. 56, (2012), pp. 3581-3593.
- [11] G. Zhou, T. He, S. Krishnamurthy and J. A. Stankovic, "Models and Solutions for Radio Irregularity in Wireless Sensor Networks", ACM Transactions on Sensor Networks (TOSN), vol. 2, no. 2, (2006), pp. 221-262.
- [12] L. Doherty, K. Pister and L. E. Ghaoui, "Convex Position Estimation in Wireless Sensor Networks Proc. of the IEEE INFOCOM", (2001), pp. 1655-1663.
- [13] J. Gu, S. Chen and T. Sun, "Localization with Incompletely Paired Data in Complex Wireless Sensor Network", Wireless Communications, IEEE Transactions, vol. 10, no. 9, (2011), pp. 2841-2849.

- [14] J. J. Pan, S. J. Pan, J. J. Yin and M. Q. Yang, *et.al.*, "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.
- [15] W. Kim, J. Park, H. J. Kim and C. G. Park, "A multi-class classification approach for target localization in wireless sensor networks", *Journal of Mechanical Science and Technology*, vol. 28, no. 1, (2014), pp. 323-329.
- [16] J. Lee, W. Chung and E. E. Kim, "A new kernelized approach to wireless sensor network localization", *Information Sciences*, vol. 243, (2013), pp. 20-38.
- [17] J. Lee, B. Choi and E. Kim, "Novel Range-Free Localization Based on Multidimensional Support Vector Regression Trained in the Primal Space", *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, vol. 24, no. 7, (2013), pp. 1099-1113.
- [18] H. Lim and J. C. Hou, "Localization for Anisotropic Sensor Networks", *IEEE INFOCOM, 24th Annual Joint Conference of the IEEE Computer and Communications Societies, Proceedings IEEE*, (2005), pp. 138-149.
- [19] Y. Shang and W. Ruml, "Improved MDS-Based Localization", *INFOCOM*, (2004), pp. 2640-2651.
- [20] D. Ma, M. Er, J. Bangwang and H. B. Lim, "Range-free wireless sensor networks localization based on hop-count quantization", *Telecommunication Systems*, vol. 50, no. 3, (2012), pp. 199-213.
- [21] L. Hogben, "Handbook of Linear Algebra", CRC Press, United Kingdom, (2007).
- [22] J. Shawe-Taylor and N. Cristianini, "Kernel Methods for Pattern Analysis", Cambridge University Press, (2004).
- [23] B. Scholkopf and A. J. Smola, "Learning with Kernels", MIT Press, Cambridge, (2002).
- [24] V. N. Vapnik, "The Nature of Statistical Learning Theory", Springer, New York, (1995).
- [25] A. Wibowo and Y. Yamamoto, "A note on kernel principal component regression", *Computational Mathematics and Modeling*, vol. 23, no. 3, (2012), pp. 350-367.
- [26] I. Jolliffe, "Principal Component Analysis", Springer, New York, (2002).
- [27] B. Schölkopf, A. Smola and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem", *Neural Computation*, vol. 10, no. 5, (1998), pp. 1299-1319.
- [28] R. Rosipal, M. Girolami, L.J. Trejo and A. Cichocki, "Kernel PCA for Feature Extraction and De-Noising in Non-linear Regression", *Neural Computing & Applications*, vol. 10, no. 3, (2001), pp. 231-243.
- [29] A. Wibowo, "Nonlinear Predictions in Regression Models Based on Kernel Method", University of Tsukuba, Tsukuba, (2009).
- [30] D. Niculescu and B. Nath, "Ad-hoc positioning system(APS)using AoA", *IEEE INFOCOM*, (2003), pp. 1734-1743.
- [31] A. Hossain, Y. Jin, W.S. Soh and H.N. Van, "SSD: A Robust RF Location Fingerprint Addressing Mobile Devices'Heterogeneity", *Mobile Computing, IEEE Transactions*, vol. 12, no. 1, (2013), pp. 65 – 77.
- [32] Y. Liu, Y.H. Hu and Q. Pan, "Distributed, Robust Acoustic Source Localization in a Wireless Sensor Network", *Signal Processing, IEEE Transactions*, vol. 60, no. 8, (2012), pp. 4350 – 4359.
- [33] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression", *Neural Networks*, vol. 17, (2004), pp. 113-126.
- [34] S. Keerthi and C.J. Lin, "Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel", *Neural computation*, vol. 15, no. 7, (2003), pp. 1667-1689.

Author



Rulin Dou is a technician at the School of Information Technology, Jinling Institute of Technology, China. He received the MS Degree in computer application technology from Nanjing University of Science and Technology in 2011. His current interests are in the areas of wireless sensor networks, machine learning, computer network and pattern recognition.