http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (24): 8576-8582, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.8576.8582

Internet of Things Positioning Model Study on Wireless Sensor Network

Xia Hu, Min Zhou and Qing Xia
School of Management, China University of Mining and Technology,
Xuzhou 221116, China

Abstract: In various wireless sensor networks applications, the sensor node positioning problem is the basis and key technology of IOF (internet of things), which is essential for IOF monitoring activities. Due to the influence of environment and many other factors, the data collected by sensor node is easily generate errors, leading to the existing location algorithm or model positioning effect not so satisfactory. Aiming to this problem, this paper using internet (Geometric) topology structure information, based on the model of Location Estimation-Locality Preserving Canonical Correlation Analysis (LE-LPCCA), through transforming the depiction of signal space and physical space local information, proposed a new IOF positioning model LE-RLPCCA. Compared with the kind methods at present, model LE-RLPCCA has a better robustness of error-tolerance technology, positioning accuracy and stability improved evidently.

Key words: Wireless sensor network, Internet of things, positioning, model

INTRODUCTION

The internet of things is an internal which interconnected objects with identity, perception and intelligent processing capability by communication technology. Wireless sensor network is one of the elements of IOT. In various WSN applications, the positioning information is essential for IOT monitoring activities as being the basis and key technology of IOT application.

Currently, in typical WSN positioning technology, RSSI has been widely used for its positioning technology dispense with extra hardware support and meet the requirement of low power and cost. But based on the instability of positioning results (Wang *et al*., 2010), the applicability is limited on some extent. To solve this problem, researchers have proposed a number of improvement strategies and algorithms. Compared with same methods at present, the wireless sensor network localization model fully use of signal space and physical space topology structure information: Based on Location Estimation-Locality Preserving Canonical Correlation Analysis (LEPCC), which positioning accuracy and stability improved significantly.

All the above introduced positioning algorithms and models when existing with wrong data, its positioning effect will not be satisfactory. Aiming at this problem, this paper on the basis of introducing and reviewing IOT location mechanism structure and LE-LPCCA positioning model based on WSN, illustrating the design idea and location model of algorithm LE-RLPCC detailed, proposing a LE-Robust LPCCA (LE-RLPCCA) positioning model with better robustness. Compared with the experiment results of the existing similar typical methods in the real world, LE-RLPCCA has higher location robustness and stability.

IOT LOCATION MECHANISM BASED ON WSN

In IOT location mechanism based on WSN, using two matched dataset of signal space and physical space is a key process to establish a mapping between these two spaces. In WSN, there is a close correlation between signal dataset and physics location dataset. However, because the data collected easily contain noise, which will lead to this correlation cannot response into the collected raw data evidently.

Aiming at this problem, typical canonical correlation analysis (CCA) (Wang *et al*., 2005) is a modeling approach to establish mapping between two datasets and make their correlation largest. But CCA can only mining the linear relationship between datasets, which is usually more difficult to apply to real world. Literature (Girod and Estrin, 2001).Follow this idea, the use of nonlinear kernel CCA (KCCA) (Girod and Estrin, 2001) to mining the nonlinear relationship between these datasets and thus achieving the required nonlinear mapping, proposing LE-KCCA (Girod and Estrin, 2001) sensor location algorithm. But model KCCA adopt a unified global nonlinear mapping, without taking consideration of the local structure characteristics of internet. On the other hand, the complicated and fickle topology structure of

Corresponding Author: Min Zhou, School of Management, China University of Mining and Technology, Xuzhou 221116, China

IOT also make the choice of kernel parameter in KCCA become a difficulty and the choice whether good or bad is crucial for a effective communication of model KCCA. Therefore, we use LPCCA method (Gu *et al*., 2011) to mining the local topology structure information of signal space and physical space. However, LE-LPCCA model is not for error data, so we develop a LE-PCCA location model with robustness.

LE-LPCCA MODEL BUILDING

In WSN location mechanism based on RSSI, the location of node is usually divided into the OTP (offline training phase) and OLP (Gu *et al*., 2011). During the training phase, through the signal strength and physical coordinates of known nodes to obtain the mapping of signal space and physical space, thus establish positioning model; during the positioning phase, using obtained mapping to evaluate position of unknown nodes. In this part, we mainly review LE-LPCCA model.

In WSN, the signal space and physical space data collected by known sensor can be divided into two datasets, considered a $X=[x_1,x_2,\cdots,x_n]^{p\times n}$ as signal strength collected by known nodes, among them, the dimension of every signal vector $x_i(i=1,2,\cdots,n)$ can be referred as p, p means the number of Access Point (AC), $Y=[y_1,y_2,\cdots,y_n]^{q\times n}$ is the physical coordinates of corresponding nodes. Usually in actual space, coordinates are two-dimensional or three-dimensional, therefore, q = 2 or 3. The primary task to build positioning model is the mapping establishing between two datasets. CCA is a classical approach used to build two datasets mapping, the object is X and Y to seek two set of base vector $w_x \in R^p$ and $w_y \in R^q$ and thus make the biggest correlation between transformed data $w_x{}^T(x_i-\bar{x})$ and $w_y{}^T(y_i-\bar{y})$, among them:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

is the sample average of X and Y, respectively. Through associatively definition and some mathematical derivation, CCA can be expressed as the solution of optimization problem of method (1):

$$\max_{w_x,w_y} w_x^T\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-x_j)(y_i-y_j)^T w_y$$

$$\text{s.t.}\, w_x^T\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-x_j)(x_i-x_j)^T w_x=1$$

$$w_y^T\sum_{i=1}^{n}\sum_{j=1}^{n}(y_i-y_j)(y_i-y_j)^T w_y=1 \quad (1)$$

Through solve optimization problem method (1), we can get w_x and w_y, them make the data transformation shaped like $w_x^T(x_i-\bar{x})$, the correlation of two transformed datasets is biggest. However, if build a mapping between signal space to physical space and structure positioning model based on CCA in WSN, only mining the linear correlation phenomenon between two datasets and without the usage of network local structure information needs to be done. To compensate for this problem, when LE-LPCCA algorithm (Gu *et al*., 2011) in the phase of building mapping, to introduce network local structure information in CCA, to transform original global nonlinear problem into some local linear problem to calculate typical related issues of each small domain and solve these sub problem, thus achieve the purpose of solving nonlinear problem through local linear method. LE-LPCCA model first give the definition of neighbor nodes in WSN.

Definition 1: In signal space, $ne^{(i)}$ means the node symbol set has the similar signal strength collected by node i, means ordinate set in local neighbor sample of X. In physical space, $ne^{(i)}$ means ordinate set of neighbor node close to i node. Among them, local neighbor is divided by K-neighbor definition: if x_j (y_j) is the k-neighbor sample of x_i (y_i), then we consider x_j (y_j) is the local neighbor of x_i (y_i). According to above definition, the similarity matrix of WSN topology structure in signal space and physical space is:

$$S_X=\{S_{ij}^X\}_{i,j=1}^{n}$$

and:

$$S_Y=\{S_{ij}^Y\}_{i,j=1}^{n}$$

and the matrix factors are:

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

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

Parameter t_x set as the average distant:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|s_i - s_j\right\|^2 / n(n-1)$$

in signal space, parameter t_y also used for similar process. Therefore, we can see that the bigger S^X_{ij} (or S^Y_{ij}) is, the closer distance between x_i and x_j (or y_i and y_j), if both x_i and x_j (or y_i and y_j) are not in the scope within domain range, then the similarity is 0. And S^X_{ij} (or S^Y_{ij}) rely on the layout of sensor nodes, so S^X_{ij} (or S^Y_{ij}) can various from network topology structure, thus show a highly flexibility.

With the above definition of similar matrix, the typical correlation of network local domain can be defined as:

$$w_x^T \bullet \sum_{j=1}^{n} S^X_{ij}(x_i - x_j)S^Y_{ij}(y_i - y_j)^T \bullet w_y$$

Therefore, a local network nonlinear question can be decomposed n local linear sub problem. And in turn the combination of this sub problem can referred as similarity of original problem. So after considering local distribution characteristic of data, LPCCA can be described as followed optimization problem:

$$\max_{w_x, w_y} w_x^T \sum_{i=1}^{n}\sum_{j=1}^{n} S^X_{ij}(x_i, x_j)S^Y_{ij}(y_i, y_j)^T w_y$$

$$\text{s.t.}\, w_x^T \sum_{i=1}^{n}\sum_{j=1}^{n} S^{X^2}_{ij}(x_i - x_j)(x_i - x_j)^T w_x = 1$$

$$w_y^T \sum_{i=1}^{n}\sum_{j=1}^{n} S^{Y^2}_{ij}(y_i - y_j)(y_i - y_j)^T w_y = 1 \quad (3)$$

The procedure of (3) is similar to CCA.

DISCUSSION AND CONCLUSIONSLE-RLPCCA LOCALIZATION MODEL PROPOSED

Geometric representation of wrong data: In complex or unbelievable network environment, influenced by many factors such as network attacks, hardware errors, environmental obstacles, the data during the transmission and localization process are easily generate distortion or errors, different with the simple errors caused by general network, this error data will seriously affect the results.

This error data with sever deviation are also referred to as "outliers", which features are making erroneous data in the geometry spatial distribution keep away from its normal data point, thus with lower density of distribution (Sun and Chen, 2007), so the motivation of localization is in the same dataset, the point with lower density of distribution will make the influence lighter, thus realize a better robustness.

RLPCCA model description: Through the depiction of LE-LPCCA, we get a definition of similarity matrix of:

$$S_X = \left\{S^X_{ij}\right\}^n_{i,j=1}$$

And:

$$S_Y = \left\{S^Y_{ij}\right\}^n_{i,j=1}$$

Consequently, we can obtain:

$$D^X_i = \sum_{j=1}^{n} S^X_{ij}$$

And:

$$D^Y_i = \sum_{j=1}^{n} S^Y_{ij}$$

among them, D^X_i depict located area intensity of node j = 1 in signal space. If D^X_i (D^Y_i)is bigger, means area intensity of node i is higher. Further, we can depict the density based on node topology structure:

$$M^X_i = \frac{D^X_i}{\sum_{j=1}^{n} D^X_j};\; M^Y_i = \frac{D^Y_i}{\sum_{j=1}^{n} D^Y_j} \quad (4)$$

Obviously, when the $M^X_i(M^Y_i)$ is bigger, it means the density of node i in signal (physical) space is bigger. Thus we can obtain density matrix:

$$M_X = \text{diag}\left[M^X_1, M^X_2, \cdots, M^X_n\right]$$

of signal space and density matrix:

$$M_Y = \text{diag}\left[M^X_1, M^X_2, \cdots, M^X_n\right]$$

of physical space. Substituting M_i^X and M_i^Y into algorithm 3, replacing S_{ij}^X and S_{ij}^Y, respectively, we can obtain follows optimization problem:

$$\max_{w_x, w_y} w_x^T \sum_{i=1}^{n}\sum_{j=1}^{n} M^X_i(x_i, x_j)M^Y_i(y_i, y_j)^T w_y$$

$$s.t.w_x^T\sum_{i=1}^{n}\sum_{j=1}^{n}M_i^{X^2}(x_i-x_j)(x_i-x_j)^Tw_x=1$$

$$w_y^T\sum_{i=1}^{n}\sum_{j=1}^{n}M_i^{Y^2}(y_i-y_j)(y_i-y_j)^Tw_y=1 \quad (5)$$

Through further expansion and consolidation, we can get:

$$\max_{w_x,w_y} w_x^TXM_{XX}X^Tw_y$$

$$s.t.w_x^TXM_{XX}X^Tw_x=1$$

$$w_Y^TYM_{YY}Y^Tw_y=1 \quad (6)$$

These symbols such as:

$$M_{XY}=DM_{XY}-M_X{}^\circ M_Y$$

$$M_{XX}=DM_{XX}-M_X{}^\circ M_X$$

$$M_{YY}=DM_{YY}-M_Y{}^\circ M_Y$$

means operator (t^A, $B\in R^{n\times n}$, $(A^\circ B)_{ij}=A_{ij}B_{ij}$, A_{ij} means the $_{ij}$ component of A); DM_{XY} (DM_{XX}, DM_{YY}) is the diagonal matrix of size n×n, among them, the i diagonal component equal to the summation of i line(because of its symmetry, or the i line) of $M_X{}^\circ M_Y$ $(M_Y{}^\circ M_Y, M_X{}^\circ M_X)$.

Using Lagrange multiplier method, solving optimization problem (Girod and Estrin, 2001), can easily get generalized Eigen value equation of RLPCCA as follows:

$$\begin{pmatrix}0 & XM_{XY}Y^T\\ YM_{YX}X^T & 0\end{pmatrix}\begin{pmatrix}w_x\\ w_y\end{pmatrix}=\lambda\begin{pmatrix}XM_{XX}X^T & 0\\ 0 & YM_{YY}Y^T\end{pmatrix}\begin{pmatrix}w_x\\ w_y\end{pmatrix}$$

Solving (7), after getting base vector set (w_x, w_y),we can transform original data such as $w^T{}_xx$, $w^T{}_yy$. Reference[9] provides a solving method of CCA on the usage of Singular Value Decomposition (SVD) technology. This study also uses SVD to solve RLPCCA. Mak:

$$H=(XM_{XX}X^T)^{-\frac{1}{2}}(XM_{XX}X^T)(YM_{YY}Y^T)^{-\frac{1}{2}}$$

$$u=(XM_{XX}X^T)^{\frac{1}{2}}w_x$$

$$v=(YM_{YY}Y^T)^{\frac{1}{2}}w_y$$

The algorithm (7) can be marshaled as:

$$\begin{cases}HH^Tu=\lambda^2u\\ H^THv=\lambda^2v\end{cases} \quad (8)$$

Algorithm (8) actually corresponding to the SVD decomposition of matrix H, assuming:

$$H=UDV^T=\sum_{i=1}^{d}\lambda_iu_iv_i^T$$

is the decomposition solution of matrix H, the i component of diagonal matrix D exactly equal to λ_i, u_i and v_i is the i line of matrix U and V, respectively, corresponding singular value λ_i,we can get:

$$\begin{cases}w_{xi}=(XM_{XX}X^T)^{-\frac{1}{2}}u_i\\ w_{yi}=(YM_{YY}Y^T)^{-\frac{1}{2}}v_i\end{cases} \quad (9)$$

Using algorithm (9) can get the i (i = 1, 2,..., d) pair base vector of RLPCCA problem, here d≤min(p, q). Using SVD decomposition to solve RLPCCA has the characteristic of computational stability andλ_t, w_{xi}, w_{yi}, can be obtained at once. Now suppose:

$$W_X=[w_{x1},w_{x2},\cdots,w_{xd}]$$

And:

$$W_Y=[w_{y1},w_{y2},\cdots,w_{yd}]$$

LE-RLPCCA localization algorithm: Known Signal strength matrix X and corresponding physical coordinate matrix Y. The object of localization model is to evaluate the position coordinates of y_g through signal vector:

$$x_g=(x_{g1},x_{g2},\cdots,x_{gn})^T$$

received by unknown node g. In section 4.2, we have discussed establish a mapping method form signal space to physical space detailed on the usage of RLPCCA in training phase; the next task is how to local unknown nodes fast in positioning phase: due to the characteristic of RLPCCA algorithm is to make the biggest correlation between signal space and physical space of transformed data and still keep original space localization topology structure information. Therefore, the points with neighboring coordinates still keep adjacency in signal space after changing. According to this, after making data transformation to signal strength indicator collected by unknown nodes, we can find the nearest known nodes from K, so the physical coordinate $\left[y^1, y^2, \cdots, y^K\right]^{q\times K}$ of K node must be in y_g nearby. Finally we evaluate the physical coordinate of g through cancroids method:

$$y_g = \frac{y^1 + y^2 + \cdots + y^K}{K} \tag{10}$$

EXPERIMENTS

Experiments description: Experimental data are collected in real network environment, deploying on a ancient city wall, the sensor nodes are relatively uniformity, which distributed on a area of 99 square meters wall and have 4 AP nodes (including non-coplanar). At the same time, the equipment also has a frequency bandwidth of 24 GHz IEEE 802.11b wireless networks. All data collection is obtained through IBM laptop linking to an external wireless USB network adapter. In this experiment we have collected 299 samples in all, of which 60% of randomly selected data samples used for training, training the mapping from signal space to physical space; the remaining 40% are used for test samples, to calculate positioning accuracy. To compare experiments, we also run the LE-LPCCA (Gu *et al*., 2011) and LE-KCCA (Girod and Estrin, 2001) model. Experimental tool used MATLAB 7.0. Positioning accuracy is expressed by the average error, the smaller average error is, the higher algorithm positioning accuracy is.

In LE-RLPCCA algorithm, the main parameters are neutralizing with (10).In the literature (Gu *et al*., 2011) have discussed the parameters choice of LE-LPCCA model detailed and this experiment is similar to its choice. So, here we no longer introduce in detail but simply select k = 3 and K = 4.

Analysis of experimental results: In order to verify the positioning effect and robustness of LE-RLPCCA algorithm under wrong data, the training data can be divided into three forms: (1) Signal strength data is incorrect; (2) Physical coordinate of known nodes are incorrect; (3) Both two set of dataset are incorrect, among them, to (1) and (2) these two situations, randomly select 0% to 50 of training data as error data and set error

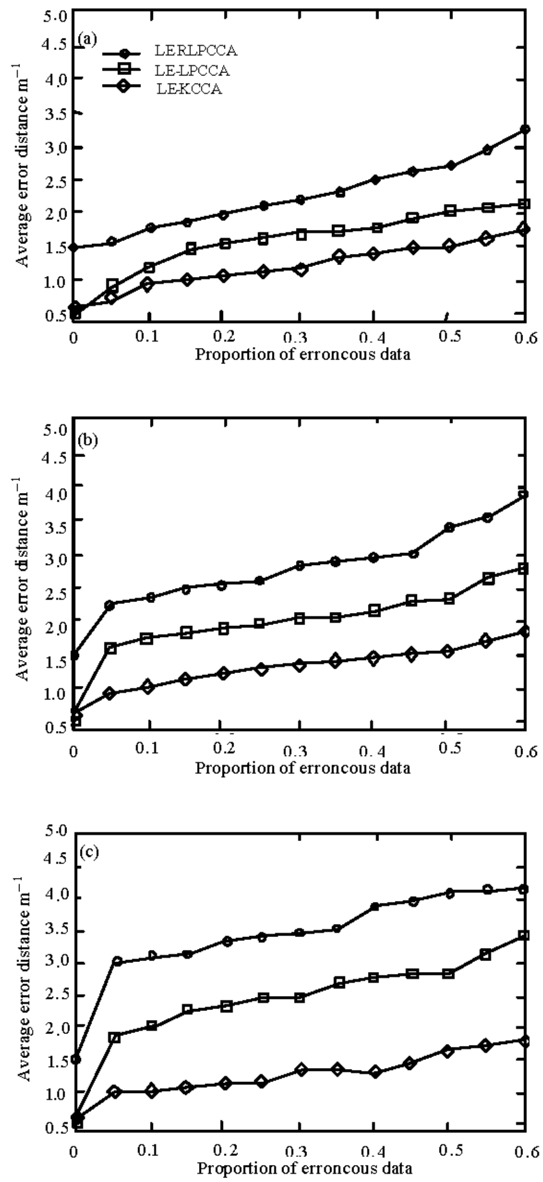

Fig. 1(a-c): Relationship between the portion of error data and error signal intensity

value three, six and ten times of normal value respectively; to the (3) situation, assumption error data occupied 30% of all training data and set when error value is three, six, ten times than normal value, the error value of incorrect physical coordinate is one-ten times than the normal value. Positioning average error shows as:

As it shows in Fig. 1-3, we can see that LE-LPCCA model based on the network topology structure

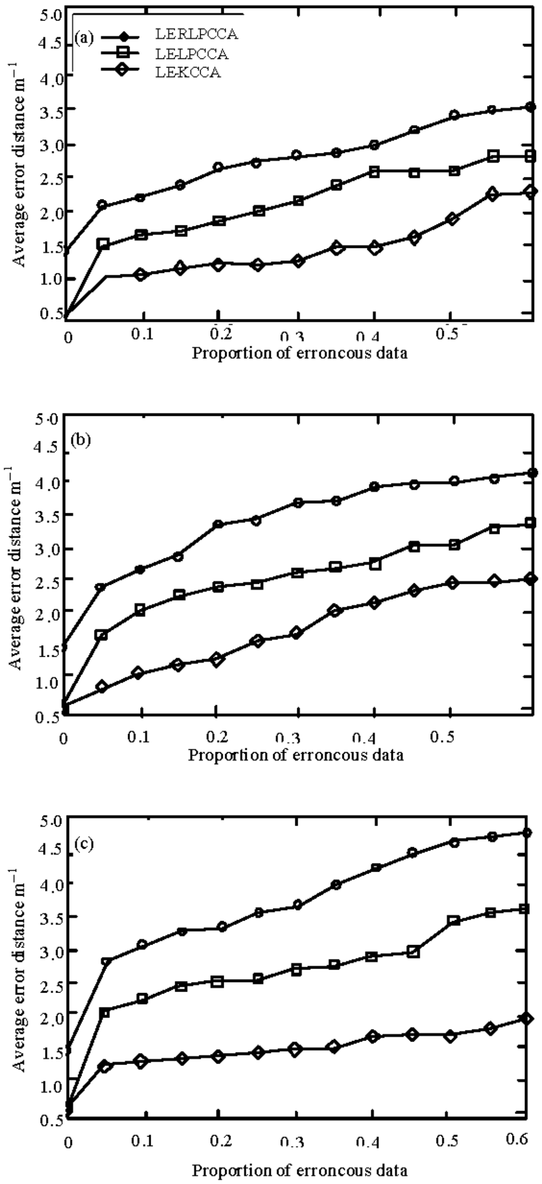

Fig. 2(a-c): Relationship between the portion of error data and positioning mean error under misaddress coordinate

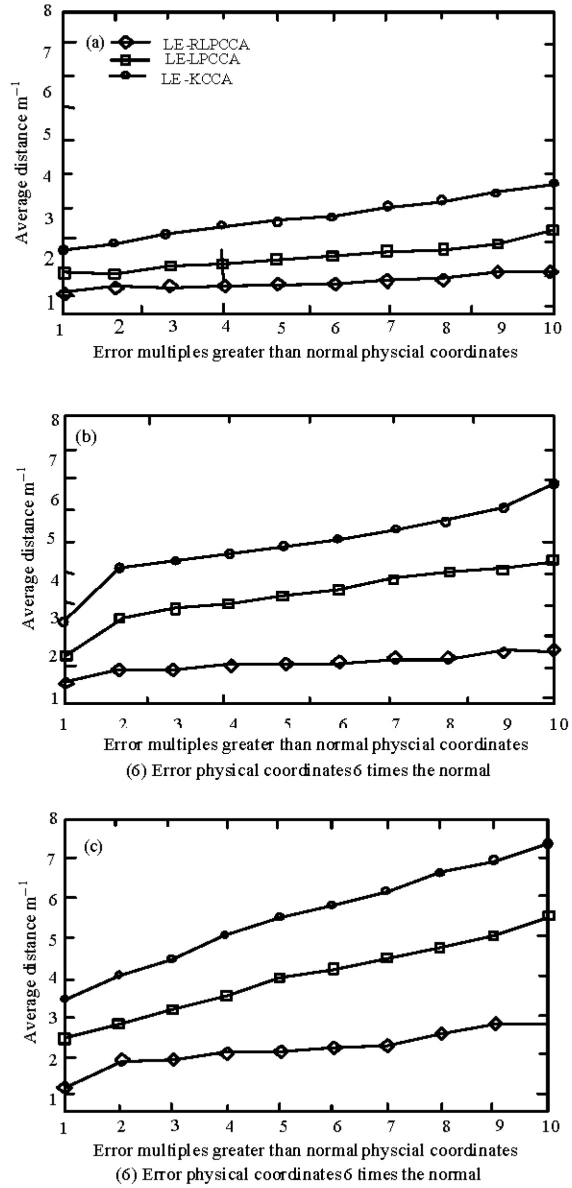

Fig. 3(a-c): Relationship between error physical coordinate value increases and positioning mean error

information, so the error data mainly act on corresponding local domain space, while when executing LE-KCCA, the error data will influence whole n network space, so reveals a worse robustness. Compared with similar studies, LE-RLPCCA model in the scene of existing error data have higher positioning accuracy and steadily performance, which is suitable for large-scale IOT positioning evaluation.

ACKNOWLEDGMENTS

This study was supported by a grant from the Fundamental Research Funds for the Central University (No. 2013XK01) and the Fundamental Research Funds for the Central University (No. JG101487).

REFERENCES

Girod, L. and D. Estrin, 2001. Robust range estimation using acoustic and multimodal sensing. Proceedings of the IEEE/RSI International Conference on Intelligent Robots and Systems, Oct. 29-Nov. 3, IEEE Xplore Press, Maui, Hawaii, USA., pp: 1312-1320.

Gu, J.J., S.C. Chen and Z. Yi, 2011. Maintaining canonical correlation analysis with local wireless sensor network node positioning. J. Software, 11: 2883-2891.

Sun, T.K. and S.C. Chen, 2007. Locality preserving CCA with applications to data visualization and pose estimation. Image Vision Comput., 25: 531-543.

Wang, C.Q., J.M. Chen and Y. Sun, 2010. Sensor network localization using kernel spectral regression. Wireless Commun. Mobile Comput., 10: 1045-1054.

Wang, F.B., S. Long and F.Y. Ren, 2005. Self-positioning system and algorithm in wireless sensor network. J. Software, 16: 1067-1080.