Localization Algorithm Based on Classification for Wireless Sensor Networks

Keyu Zhuang

Abstract— Node localization in wireless sensor networks is one of the key technologies as it plays a critical role in many applications. By the analysis of localization algorithm based on support vector machine, and using the classification function of SVM and combined with the RSSI(*Received Signal Strength Indicator*) location algorithm, a new localization algorithm based classification is proposed in the paper. The location area is firstly divided into densely populated regions and sparse regions by SVM, and then different algorithms are used to realize the localization, this can save computation and improve the accuracy. Simulation results show that the algorithm works well, and the algorithm ensures the algorithm complexity in the case of increased positioning accuracy.

Keywords— wireless sensor networks, localization, SVM, RSSI.

I. INTRODUCTION

POSITIONING technology is one of the relatively important technologies in wireless sensor networks. At present, various localization algorithms of nodes are suitable for different applications [1, 2,]. In addition, it is difficult for most of the localization algorithms to solve the boundary problems and blind spots problems in the positioning process. In order to solve these two problems, literature [3] proposed a wireless sensor networks node locating method based on SVM (LSVM). From the simulation results [3], the algorithm can achieve rapid positioning, and very obviously solve the boundary problem in node positioning. Small training volume and good classification effect is used ingeniously by LSVM algorithm and applied to the wireless sensor networks node self-positioning, which achieves good results and effectively solve to the boundary problem and blind spot problem. However, the algorithm only realizes the coarse-grained positioning, that is, it only roughly estimates the position information of the node and the error is relatively large, which is not suitable for situations with high positioning accuracy. In this paper, we use the actual RSSI (Received Signal Strength Indicator) measured by the anchor node to determine the signal loss model. Combined with the LSVM localization algorithm, a node localization algorithm based on classification is proposed and simulated.

Keyu Zhuang is with the School of Automation and Electrical Engineering. Qingdao University of Science & Technology, Qingdao 266042, Shandong, China (corresponding author; e-mail: zhuangkeyu@163.com).

II. CLASSIFICATION PRINCIPLE OF SVM

Suppose the elements in vector space X are divided into two categories: G and -G. Assuming that each data element x has eigenvectors $\vec{x}$ in the feature space $\overline{X} \subseteq \Re^n$, k data points are given, which are called "training points" and correspond to the values $y_1, y_2, \cdots, y_k$ (If $x_i \in G$,corresponding $y_i = 1$. If otherwise, y_i=-1), it should be verified whether the new data x belongs to G

Support vector machine (SVM) can solve the classification problem well. For limited space, the calculation steps are as follows:

Define a kernel function K: $X \times X \to \Re$,which must be a symmetric function, and the $k \times k$ order matrix $[K(x_i, x_j)]_{i,j=1}^k$ must be semi-positive definite (with non-negative eigenvalue). This function gets the maximum value

$$W(\alpha) = \sum_{i=1}^{k} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{k} y_i y_j \alpha_i \alpha_j K(x_i x_j) \quad (1)$$

And meet

$$\sum_{i=1}^{k} y_i \alpha_i = 0 \,, 0 \le \alpha_i \le C, i \in [1, k] \quad (2)$$

Assuming $\{\alpha_1^*, \alpha_2^*, \cdots, \alpha_k^*\}$ is the root of the equation, choose $b = b^*$, for all i values that satisfy $0 < \alpha_i^* < C$, $y_i h_K(x_i) = 1$and the training data points corresponding to (i, α_i^*) are called support vectors. The rules for classifying data points x are: if $x \in G$, then sign function $sign(h_K(x)) = 1$, and

$$h_K(x) = \sum_{i=1 \to k, x_i} \alpha_i^* y_i K(x, x_i) + b^* \quad (3)$$

Statistical learning theory states that as long as a function $K(x_i, x_j)$satisfies the Mercer condition, it corresponds to the inner product in a transformation space $\overline{X}$ .The function $h_K(\cdot)$ represents classifying the training points in space X (G and -G) into hyperplanes in the feature space.

ISSN: 1998-4464

III. Location Algorithm Based on SVM

A. Network Model

Suppose there are N nodes $\{S_1, S_2, \cdots S_N\}$ distributed in two-dimensional area $[0, D]\times[0, D]$ (D> 0). Any node has communication radius r, and each node has the same communication radius. If two nodes can communicate directly with each other, then we think that the distance between them is less than the communication radius r. Supposing the number of anchor nodes is k, $\{S_i\}(i=1\rightarrow k)$, anchor node location is known, then we must determine the location information of remaining node ($\{S_j\}(i=k+1\rightarrow N)$) .

B. SVM Model

$(X(S_i), Y(S_i))$ is used to represent the actual position of node Si and $h(S_i, Y_i)$ represents the single-hop distance of the shortest path between nodes Si and Sj. Each node Si represents a vector $s_i =< h(S_i, S_1), h(S_i, S_2), \ldots, h(S_i, S_k) >$ .SVM training point is the anchor node $\{S_i\}(i=1\rightarrow k)$. Radial Basis Function (BF) is chosen as the kernel function.

$$K(S_i, S_j) = e^{-\gamma\|si-sj\|_2^2} \tag{4}$$

Thereinto, $\|.\|_2$ is norm of I_2 , $\gamma > 0$ is a constant, which is calculated during training.

The following is the process that divides unknown nodes into $M-1 = 2^m - 1$ classes in two directions:

（1） x direction is divided into M-1 categories: each category cx_i contains nodes $x \ge iD/M$.

（2） y direction is divided into M-1 categories $\{cy_1, cy_2, \ldots, cy_{M-1}\}$: each category cy_i contains nodes $y \ge iD/M$.

From the coordinate , each cx_i class is located to the right of the vertical line $x = iD/M$, and the cy_i class is located above the horizontal line $y = jD/M$. Therefore, if the SVM learns that the node S is in the class cx_i and not in the class cx_{i+1}, in the class cy_j but not in the class cy_{j+1}, it can be determined that the node S is in the region $[iD/M, (i+1)D/M]\times[jD/M, (j+1)D/M]$. Assuming that the node is in the center of the region, and if there is no error in classification, the biggest error is $D/\sqrt{2}M$.

Algorithm Agreement

First, analyze the classification in the x-axis direction, the x class constitutes a binary tree, as shown in Figure 4-3, each tree node represents an x class, and the two branches represent whether they belong to the class(0 represents not belonging to, 1 represents belonging to).So that the relationship between the subclass and parent class is very clear, the result is the class sorted in order $cx_1 \rightarrow cx_2 \rightarrow \ldots \rightarrow cx_{M-1}$. According to this decision tree, each node S can use the following algorithm to estimate the position in the x-axis direction:

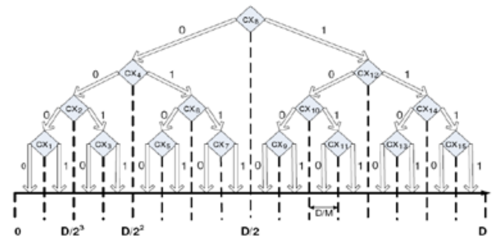

Fig.1. x side axial classification

X-axis positioning:

（1） $i = M/2$ （The root of the decision tree $cx_{M/2}$ ）

（2）If (SVM judges S does not belong to the class cx_i)

a) If cx_i is a leaf class (the bottom of the tree), return $x'(S) = (i-1/2)D/M$

b）Otherwise turn left to subclass cx_j and make i = j

3）Otherwise

a） If cx_i is leaf class (the bottom of the tree), return $x'(S) = (i+1/2)D/M$

b）Otherwise turn to the right subclass cx_t and make i = t

4）Turn to step 2）

By the same token, the decision tree about the y-axis direction can also be deduced in this way, so that the estimated position $y'(S)$ in the y-direction can be obtained. By using this positioning method, each node needs to access the decision tree nodes. Parameter M determines the positioning accuracy

Concrete positioning to achieve it includes three steps: training phase; broadcasting phase; positioning phase.

Training phase: Suppose node a is chosen as the head node. The head node first conducts SVM training, and then it becomes the most important node. The head node may be a base station or sink node.

Training phase is carried out between the anchor nodes, the information exchange can adopt single-propagation routing algorithm. First, each anchor node sends a Hello message to other anchor nodes. After the end, the anchor node can calculate the hops distance to other anchor nodes. Then each anchor node sends an INFO message to the head node. The message includes its own location information and the number of hops between other anchor nodes. After the head node knows this information, it proceeds to the next step. The head node performs SVM training on all 2M-2 classes ($cx_1, cx_2, \ldots, cx_{M-1}$, $cy_1, cy_2, \ldots, cy_{M-1}$), and support corresponding b^* vectors Si for each class when calculating the sum of.

（2）Broadcasting stage

At this stage, the head node broadcast SVM model information to all the nodes first. In this way, any node S obtains all the conditions except for the hop distance in Equation 4-3. At this time, each anchor node (except the head node) sends a Hello message to the entire network, so that any node knows its own hop count and distance from the anchor node after receiving the Hello message. In this way, the positioning conditions are all satisfied.

ISSN: 1998-4464

（3）Positioning stage

When the unknown node starts positioning after it receives changing SVM model, estimate their position according to the positioning algorithm described earlier.

IV. Node Location Algorithm Based on Classification

LSVM algorithm ingeniously uses features of small training amount and good classification effect of support vector machine (SVM), and applies it to the wireless sensor networks node self-localization, which achieves good results and give a better solution to the boundary problem. However, the algorithm only realizes the coarse-grained positioning, that is, it only roughly estimates the position information of the node, and the and error is relatively, which is not suitable for situations requiring high positioning accuracy.

Based on the advantages of LSVM positioning, the signal loss model is determined by using the actual RSSI (Received Signal Strength Indicator) measured by the anchor node. Combined with the LSVM positioning algorithm, this chapter designs the node localization algorithm based on classification as follows:

(1) Each anchor node sends an INFO message to the head node (sink node), which includes its own ID information and location information. In this way, the anchor distribution information of the entire wireless sensor networks is very clear.

(2) The unknown node only receives anchor node information in the line-of-sight range and records the anchor node location information received by itself.

(3) According to the location distribution information of anchor nodes, wireless sensor networks are divided into several small anchor areas according to the distribution density of anchor nodes. Thereinto, the anchor node with larger density is the first batch of anchor area. Anchor nodes with the highest connectivity in this type of area are selected as support vectors to train SVM, assuming a total of m. The anchor nodes with relatively small density in the area are all used as support vectors to train SVMs, assuming there are n nodes in total. This greatly reduces the number of training samples in the LSVM algorithm, which has a total of (m + n) training samples.

(4) After the SVM training ends, the unknown node classifies itself by SVM according to its recorded anchor nodes within the line-of-sight range. Note that the classification here does not distinguish between the x direction or y direction, which belongs to a vector class. After classification, unknown nodes can be divided into (m + n) categories.

(5) After the SVM training ends, the unknown node classifies itself according to its recorded anchor nodes within the line-of-sight range by SVM. Note that the classification here does not distinguish between the x direction or y direction, which belongs to a vector class. After classification, unknown nodes can be divided into (m + n) categories.

(6) Locate unknown nodes that belong to the area with high anchor node density.

The density of anchor nodes in this area is large, and all of them can be positioned and refined. The received signal strength indicator (RSSI) method is used in this algorithm for refinement [4-5]. Radio propagation path loss has a significant impact on the positioning accuracy of the RSSI positioning algorithm. The propagation path loss model commonly used in complex environments is the Shadowing model [6]:

$$\frac{P_r(d_0)}{P_r(d)} = (\frac{d}{d_0})^{\beta} \tag{5}$$

The Shadowing model can predict the average energy Pr (d) received when the distance is d. It uses a distance d_0 close to the center as a reference. Thereinto, β is the Passloss index, and its specific value depends on the actual environment and is usually the empirical value measured from the site. The impact of environmental factors on the Passloss values is non-growth, especially in the indoor environment, the Passloss index of obstacles is two or three times of the sight line. If we use a uniform Passloss index value to locate in an environment with obstacles, the positioning error is very large, and the existing positioning algorithm is difficult to meet the precise positioning requirements under such environment. Therefore, in this thesis, Passloss index is actually measured by anchor node's own position to improve the positioning accuracy. The specific process is as follows:

a. Anchor nodes periodically send their own information: node ID, its own location information.

b. After receiving the information, the ordinary node records only the average RSSI of the same anchor node.

c. After receiving the information of m anchor nodes exceeding the threshold, the ordinary nodes sort the anchor nodes according to their RSSI average from largest to smallest.

d. RSSI value correction calculation is carried out by the first m of the maximum of RSSI value :Based on the above analysis of the shadowing model of radio propagation path loss, the first three anchor nodes with a large RSSI value are selected to form the following triangular set.

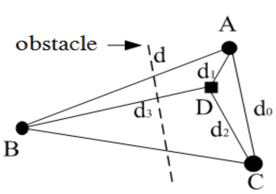

Fig.2. Algorithm sketch

As shown in Fig. 1, assuming that A, B and C are anchor nodes with known positions, D is an unknown position node, and the broken line is an obstacle, the Passloss index β can be derived according to the formula of Passloss (11)

$$\beta = \frac{\log(\Pr(do)) - \log(\Pr(d))}{\log(d) - \log(do)} \tag{6}$$

The Passloss exponent β reflects the consumption index of environmental energy to radio energy. As the number of obstacles increases, the relative value increases. As a result, the speed of average energy received decreases as the distance increases. According to Figure 1 and Equation (13), the locations of anchor nodes A, B and C are known and D is an unknown node. We select the first three anchor nodes with the largest average RSSI received by D as the RSSI correction nodes. Since the location of the anchor node is known, all the parameters in Eq. (13) are known and the value of β can be

ISSN: 1998-4464

found. This eliminates the error caused by the huge difference in radio energy consumption with obstacle and sight distance. After the β value is determined, the RSSI method is used to estimate the distance from the node to the hair node. After the information of the party distance exceeds 3, the location of the node is determined by trilateral or multilateral methods.

After the above six-step positioning calculation, the anchor node distribution-intensive regional positioning has been achieved. However, the positioning is not implemented in the area where the anchor nodes are distributed less densely or there is no anchor node in the one-hop communication range. which is called dead zone effect or boundary effect. For this part of the node, the algorithm carries out the following settlements.

(1) An anchor node in the area sends an INFO message to an unknown node in the area, and the message includes a node ID and its own location information.

(2) After receiving the information, the unknown node in the area records the information. If the node receives more than three anchor nodes within the line-of-sight range, the environment adaptive positioning algorithm is used to determine its position.

(3) If an unknown node receives only two anchor node information in the line-of-sight range, the node considers itself to be located at the center between two anchor nodes. The location of such nodes can be corrected by using the RSSI method and can depend on the situation.

(4) If the unknown node receives only one anchor node information within the line-of-sight range, the node considers itself to be located outside the bounds of the anchor node within the line-of-sight, that is, the outer boundary of wireless sensor networks. If the unknown node does not receive the anchor node information in the range of one hop, it uses the anchor node information that can be received in the two-hop or multi-hop range to locate its own position outside the boundary of the anchor node. The position is then distributed, and the RSSI method can be used to correct it to improve accuracy.

Through the above positioning calculation process, as long as it is the nodes in the Unicom area, it can estimate their position, the blind spot effect and the border effect will be reduced correspondingly.

V. Comparative Analysis of Algorithm Simulation

In order to verify the validity of the algorithm, we use MATLAB to simulate the localization algorithm. The simulation environment is assumed to be plane area of [0,100] ×[0,100] m. The anchor nodes and the localization nodes are randomly distributed. Suppose there are 200 to be determined Bit nodes, select two sets of simulation results in which the number of the anchor nodes is 25,45 .The location simulation of each group was carried out more than 10 times, the error statistics adopt the average of multiple simulations. Respectively LSVM positioning algorithm and simulation comparison of node positioning algorithm based on classification, the following is the simulation results:

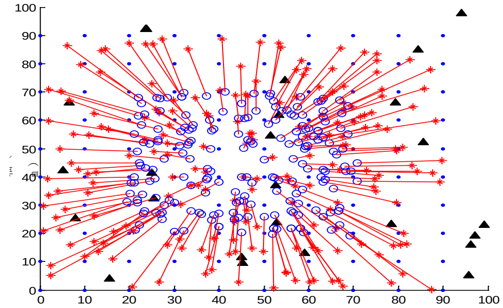

Fig.3. Simulation results map of LSVM algorithm node positioning (The number of anchor nodes=25)

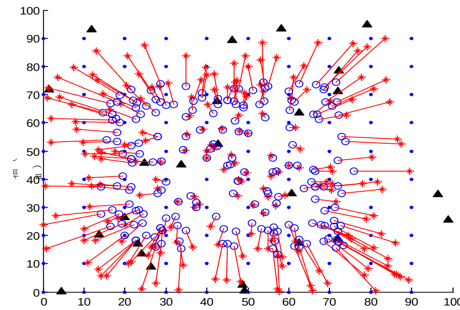

Fig.4. Simulation results map based on node location of Classification algorithm (The number of anchor nodes=25).

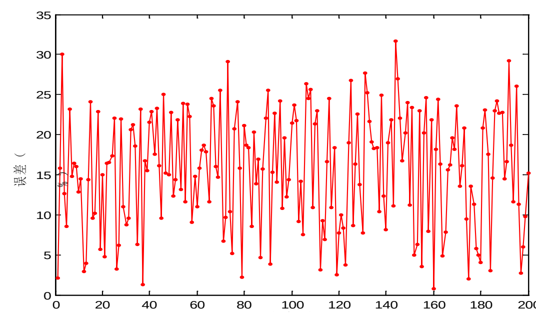

Fig.5. Error map of LSVM algorithm Node positioning (The number of anchor nodes=25).

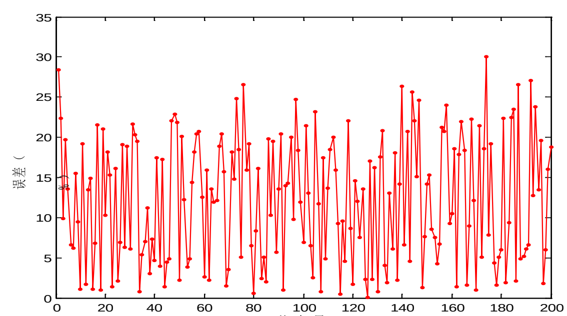

Fig.6. Error map based on Classification algorithm node location (The number of anchor nodes=25).

ISSN: 1998-4464

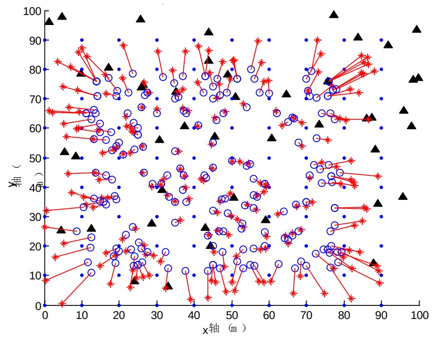

Fig.7. Results distribution of LSVM algorithm node positioning (The number of anchor nodes=45).

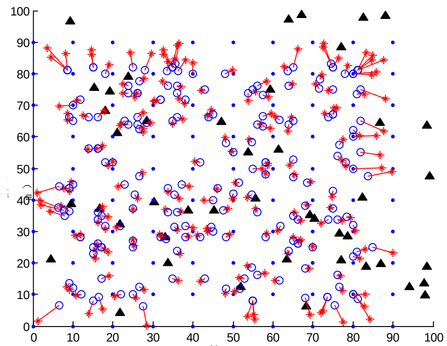

Fig.8. Results distribution based on node location of Classification algorithm(The number of anchor nodes=45)

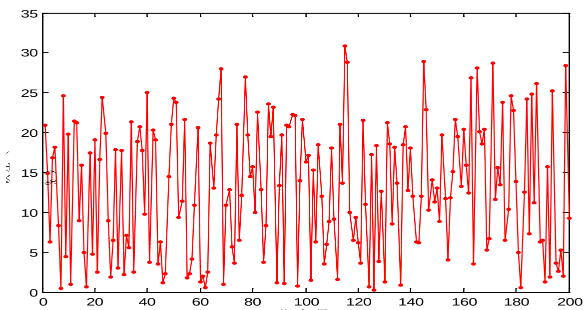

Fig.9. Error distribution of LSVM algorithm Node positioning (The number of anchor nodes=45).

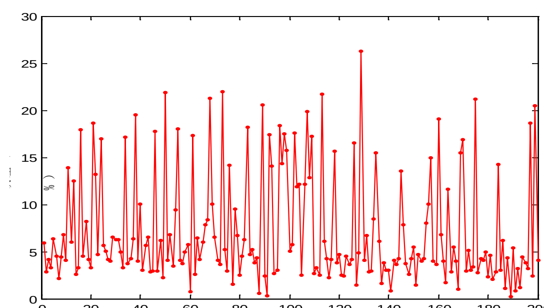

Fig.10. Results distribution based on node location of Classification algorithm (The number of anchor nodes=45).

From the above positioning simulation figure, we can see that the positioning accuracy of the classification-based localization algorithm designed in this chapter is higher than that of the LSVM localization algorithm, and the positioning error is reduced accordingly. The positioning error comparison results are shown in Fig. 10. After statistics, based on the classification algorithm, bit error is about 45% less than the LSVM algorithm.

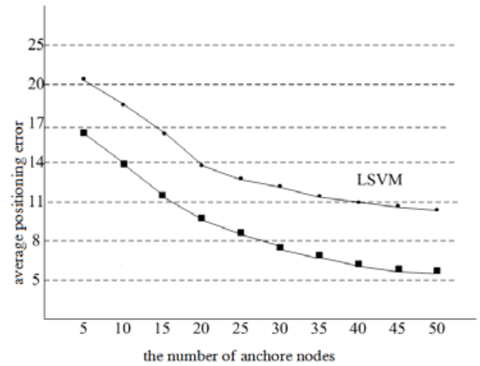

Fig.11. Comparison of positioning error of two algorithms.

VI. Conclusion

Node location is one of the key technologies in wireless sensor networks since it plays a critical role in many applications. If the users cannot obtain the accurate location information effectively, the related applications cannot be accomplished well. Based on the analysis of the SVM-based localization algorithm, this paper uses the SVM classification function and combines it with the RSSI node localization algorithm, and proposes a node localization algorithm based on classification. The algorithm firstly uses the location information of anchor nodes to classify the sensor network distribution areas. For the anchor nodes with densely distributed areas, the anchor nodes with the highest connectivity are selected as support vectors, which greatly reduces the number of training samples in LSVM node location algorithm and reduces Computation. The node density area of anchor nodes is located and refined by using the environment adaptive node localization algorithm, and the regions with fewer anchor nodes are classified again. The nodes with direct connection to the three anchor nodes adopt the environment adaptive positioning algorithm and the node directly connected without the anchor node adopts the LSVM positioning algorithm, which saves the computation amount and improves the positioning accuracy. The simulation results show that the algorithm works well and obviously improves the positioning efficiency.

References

[1] Shi Long, Wang Fubao, Duan Weijun et al, “Wireless sensor networks Range-Free self-localization mechanism and algorithm”, *Computer Engineering and Applications*, vol. 40, no.9, pp.127–130, 2004.

[2] Wang Fubao, Shi Long, Ren Fengyuan, “Self-localization system and algorithm in wireless sensor networks”, *Software Journal*, vol. 16, no.5, pp.1148–1157, 2005.

[3] Duc A. Tran , Thinh Nguyen, “Localization In Wireless Sensor Networks based on Support Vector Machines”, National Scinence Foundation under Grant. 2008.

[4] An Yong, Jiang Ting, Zhou Zheng, et al., “A centroid location algorithm for wireless sensor networks”, *Computer Engineering and Applications*, vol. 43, no.20, pp.136–138, 2007.

ISSN: 1998-4464

[5] Chen Weike, Li Wenfeng, Shou Heng et al., "Weighted centroid localization algorithm of wireless sensor networks based on RSSI", *Journal of Wuhan University of Technology*, vol. 30, no.2, pp.265–268, 2006.
[6] Wang Yang, "Research on positioning technology of wireless sensor networks", University of Science and Technology of China, Hefei, 2007.
[7] Guangjie Han, Huihui Xu, Trung Q. Duong, Jinfang Jiang, Takahiro Hara, "Localization algorithms of Wireless Sensor Networks: a survey", *Telecommunication Systems*, vol. 52, no. 4, pp. 2419–2436.
[8] Baoming Sun, Yan Guo, Ning Li, Dagang Fang, "Multiple Target Counting and Localization Using Variational Bayesian EM Algorithm in Wireless Sensor Networks", *IEEE Transactions on Communications*, vol.65, no. 7, pp. 2985 – 2998.

Keyu Zhuang was born on Feb. 24, 1975**.** He received the Master's degree in Control theory and Control Engineering from Guangxi University. Currently, he is a teacher at Qingdao University of Science and Technology, China. His major research interests include sensor technology.

ISSN: 1998-4464