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 (11): 2251-2255, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.2251.2255

Localization of Wireless Sensor Networks with a Mobile Beacon

[1,2]Jiang Xu and [1]Huan-Yan Qian
[1]School of Computer Science and Engineering,
Nanjing University of Science and Technolog, Nanjing, 210094, China
[2]School of Computer Science and Engineer,
Changshu Institute of Science and Technology, Changshu, 215500, China

Abstract: A vast majority of localization techniques proposed for sensor networks are based on triangulation methods in Euclidean geometry. This study proposes a novel approach known as MSVM that utilizes a mobile beacon to implement location estimation on wireless sensor networks. MSVM is based on Support Vector Machine(SVM).In the approach, the problem of localization is converted into classification, which takes advantage of the signal information generated by mobile beacon between virtual beacon nodes and unknown nodes to compute the location. In this way, expensive beacon nodes with GPS function can be avoided, saving investment on wireless sensor nodes. Shown by the simulation test, the approach has quite high localization precision.

Key words: WSNs, localization, mobile beacon, SVM

INTRODUCTION

Node location information acquisition is one of the important research fields for Wireless Sensor Networks (WSNs). Nodes often endowed with limited computing, processing, wireless communication and storage ability, which are able to sense surrounding environment in a closed range. Between nodes, a network is self-organized to transmit data mutually via wireless communication and to release the data to observers. After years of study and development, researchers proposed many node localization algorithms. According to the principle of whether the practical distance between nodes is measured in the localization process, these location algorithms can be divided as range-based localization algorithm and range-free localization algorithm. Localization approaches based on range include Time Difference of Arrival (TDOA) (Priyantha, 2005), Angle of Arrival (AOA) (Niculescu and Nath, 2003) and Received Signal Strength Indication (RSSI) (Whitehouse, 2002). They all need additional hardware support, which are quite costly ordinarily and consume plenty of energy. Range-free location approaches include centroid algorithm (Bulusu *et al*., 2000), DV-Hop algorithm (Niculescu and Nath, 2001) and Approximate Point-in-triangulation Test (APIT) (He *et al*., 2003). In recent years, some scholars put forward a localization method based on machine learning (Brunato and Battiti, 2005; Nguyen *et al*., 2005; Pan *et al*., 2006; Tran and Nguyen, 2008) which training and learn an estimation model in the deployed area, so as to estimate the location information of unknown nodes in the area.

Support Vector Machine (SVM) (Cortes and Vapnik, 1995) is a classification method established based on statistical learning theory and is designed to solve small sample problems. Its major contents include kernel function and support vector. Support vector is obtained by trainings on training samples. The study has proposed a novel wireless sensor networks node localization approach-MSVM. The approach utilizes a mobile beacon to march on along planned route, and then realizes localization by drawing support from SVM-based machine learning algorithm. As for the problem is localization, the geographic area deployed with wireless sensor nodes is divided into several grids and then to classify these nodes into these grids. In this way, the problem of node localization is simply converted to be the problem of classification. Training data needed by SVM derives from the communication signal vector between auxiliary mobile beacon nodes and fixed nodes. In this study, it is assumed that nodes in the area are able to communicate directly via signal. Thus, the algorithm is not applicable to large scale network.

Corresponding Author: Jiang Xu, School of Computer Science and Engineering, Nanjing University of Science and Technolog, Nanjing, 210094, China

SVM CLASSIFICATION

Assuming that there are n training data samples $x_i(i = 1, 2, ...n, x_i \in R^n)$ with the class mark separately as y_1, y_2, ..., y_n and the samples are then divided into two classes: H or ¬H (non-H). If $x_i \in H$, then $y_i = 1$; otherwise $y_i = -1$.On this basis, whether the new data sample x belongs to H or does not belong to H needs to be predicted. SVM can be used to solve this problem. Ordinarily, the steps proceed as follows:

- Define a kernel function K(x, x) and the function shall meet Mercer condition (Smola and Scholkopf, 2002), The central station then solves the following optimization problem:

$$\min \frac{1}{2} \| w \|^2 \tag{1}$$

subject to:

$$y_i[(w^T \cdot x_i + b)] \geq 1, i = 1, 2, \cdots, n$$

- Through computation, the Wolfe dual problem to the original problem is figured out:

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

subject to:

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \cdots, n$$

- Assuming that the optimal solution to this quadratic programming problem $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_n^*)^T$ has been figured out and:

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i,\ b^* = y_j - \sum_{i=1}^{n} y_i \alpha_i^* K(x_i, x_j)$$

in which the subscript $j \in \{j | \alpha^*_j > 0\}$. Hereby, the decision function is worked out:

$$\begin{gathered} f(x) = \mathrm{sgn}(g(x)) \\ g(x) = (w^* \cdot x) + b^* = \sum_{i=1}^{n} \alpha_i^* y_i K(x, x_i) + b^* \end{gathered} \tag{3}$$

For the data sample x, $x \in H$, iff f(x) = 1. Thus, far, by draw support from a kernel function K(x, x), the input space is projected to a high-dimensional space to figure out a hyper-plane g(x) and to maximize the geometrical interval of training sample set in allusion to the hyper-plane. In the study, Radial Basis Function is adopted as the kernel function:

$$K(x, x') = \exp\left(-\frac{\| x - x' \|^2}{\sigma^2}\right) \tag{4}$$

MSVM

Problem description and model: It is supposed that m unknown wireless sensor network nodes S_1, S_2, ..., S_m are deployed in a two dimensional area $[0, C]\times[0, C](C>0]$. Then, the two dimensional area is divided into T×T grids shown in Fig. 1 (In the following part of the study, the classification along axis X is mainly discussed, while the situation of axis Y is similar). The objective is to estimate the two-dimensional coordinates of these unknown nodes.

Define 2 class sets and each class set contains T classes:

- Define T classes along axis X $\{SVM_{x1}, SVM_{x2}, ..., SVM_{xT}\}$ and each class SVM_{xk} contains all nodes when $X \geq iC/T$
- Define T classes along axis Y $\{SVM_{y1}, SVM_{y2}, ..., SVM_{yT}\}$, and each class SVM_{yk}contains all nodes when $Y \geq iC/T$

ss(Sj, S)(j = 1, 2, ..., m) represents the strength of signal (RSS) received by node S from each node S_j; define signal vector $SS_j(ss(s_1, s_j), ss(s_2, s_j), ..., ss(s_m, s_j))$, in which y_j refers to class mark in allusion to each class. In this study, a mobile node S_b is employed to assist the localization process, so as to make S_b march on (lingering for a certain period at the center of each grid) along the green arrow shown in Fig. 1. As mobile beacon is equipped with GPS, the mobile beacon S_b hereby is able to get n virtual beacon node S_{bi} and location information. These virtual beacons receive RSS from all unknown nodes in the network, so as to figure out the signal vector $SS_{bi}(SS_{bi} = (ss(s_1, s_{bi}), ss(s_2, s_{bi}), ..., ss(s_m, s_{bi})))$, as well as the class mark $y_{bi}(y_{bi} \in \{1, 2, ..., T\})$ corresponding to each class SVM_{xk}. The data information will then be used as SVM training data.

If an unknown node is estimated belonging to SVM_{xk}, while not belonging to SVM_{xk+1}, then it may be considered that the horizontal coordinate of the unknown node shall be (k+1/2)C/T. If the classification estimation is correct, the maximum localization error of unknown node shall be $C/(\sqrt{2}T)$.

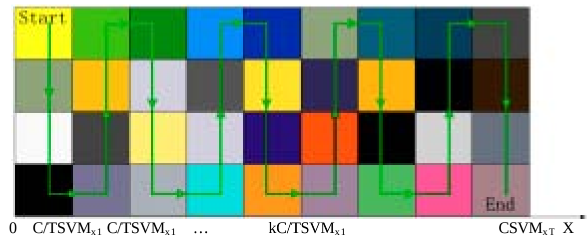

Fig. 1: The two dimensional area is divided into T×T grids. Green arrow indicates the moving direction of mobile beacon

CLASSIFICATION STRATEGIES AND ALGORITHM DESCRIPTION

It can be seen from the above discussion that, classification problem in this study belongs to multi-class classification. Frequently used multi-class classification methods include one vs. rest, one vs. one and binary decision tree. In accordance with the feature of training sample, this study proposes a method of multi-class vs. multi-class. The concrete procedures: firstly, establishing T two-class classification machine and each classification is to divide samples at left and right of the class. On this basis, the attribution of unknown nodes is to be judged. By analyzing training data obtained in the above discussion, the problem of quite unbalanced positive and negative data (i.e., data set skew) emerges. As for this, penalty coefficient C, C_+ for positive sample and C_- for negative sample. Thus, Formula 1 is transformed to be:

$$\min \frac{1}{2}\|w\|^2 + C_+\sum_{y_i=1}\varepsilon_i + C_-\sum_{y_i=-1}\varepsilon_i \tag{5}$$

subject to:

$$y_i[(w^T \cdot x_i + b)] \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, i = 1, 2, \cdots, n$$

The dual problem shall be:

$$\max W(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \tag{6}$$

subject to:

$$\sum_{i=1}^{n}\alpha_i y_i = 0, i = 1, 2, \cdots, n \tag{7}$$

$$\begin{aligned} &0 \leq \alpha_i \leq C_+, y_i = 1 \\ &0 \leq \alpha_i \leq C_-, y_i = -1 \end{aligned} \tag{8}$$

Detailed computation steps are listed as follows:

- Assuming that P = $\{(x_1, y_1),...,(x_l, y_l)\}$ is a given training set and x_h = SS_{bi} ∈ R^m, y_h ∈{1, ..., T}, i = 1, 2, ..., n, h = 2, 2, ..., l, h>1
- For k = 1, 2, ..., T, the following computation is performed: taking training samples at the right side of SVM_{xk} in Fig. 1 as positive samples, while at the left side as negative samples. Figure out the decision function with Formula 6:

$$\begin{aligned} &f^k(x) = \operatorname{sgn}(g^k(x)) \\ &g^k(x) = \sum_{i=1}^{n} y_i\alpha_i^k K(x, x_i) + b^k) \end{aligned} \tag{9}$$

- Judging whether input x belongs to SVM_{xk}.

SVM-based MSVM localization algorithm is comprised by two phases: training and localization.

Training phase: As was described in the above, after mobile beacon acquired SVM training data, run SVM training program to compute $SVMx_i$ α^* of each class and the corresponding b*. On this basis, mobile beacon will then transmit the data to the whole network and all unknown nodes will store the data information, so as to work out the decision function for each class.

Localization phase: Mobile beacon exit the network and communication between unknown nodes is established. Each node will get signal vector SSj = $(ss(x_1, x_j), ss(x_2, x_j),..., ss(s_m, s_j))$, which is brought into Formula 9. If unknown node S_j belongs to SVM_{xk}, while does not belong to SVM_{xk+1}, the horizontal coordinate (axis X) of S_j can hereby be estimated, which is (k+1/2)C/T. The longitudinal coordinate of S_j can be calculated in the same way.

SIMULATION

Assuming that in the two-dimensional area R^2, a wireless sensor network $\{S_1, S_2,...,S_n\}$ comprised by m unknown nodes is randomly deployed. The node ID are

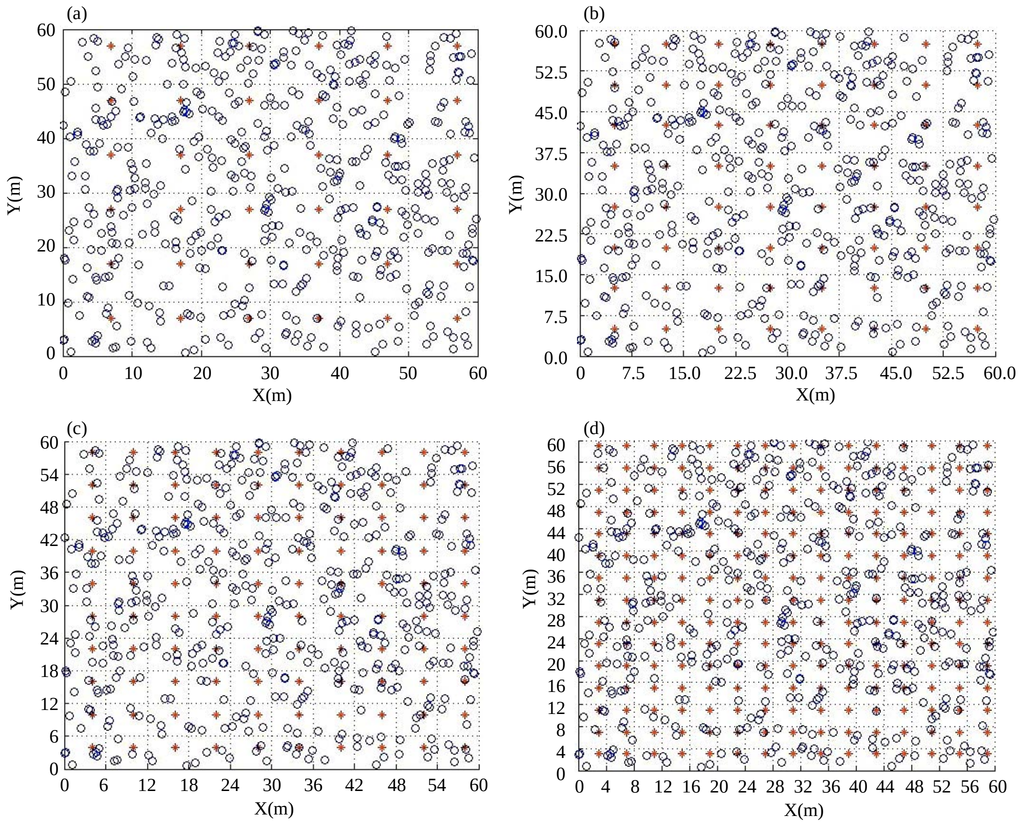

Fig. 2(a-d): Node Distribution when T = 6, 8, 10, 12, 15, n = 500. Blue "o" stands for unknown nodes, red "*" stands for generated virtual beacon nodes, (a) Node Distribution when T = 6, n = 500, (b) Node Distribution when T = 8, n = 500, (c)Node Distribution when T = 10, n = 500 (d) Node Distribution when T = 15, n = 500

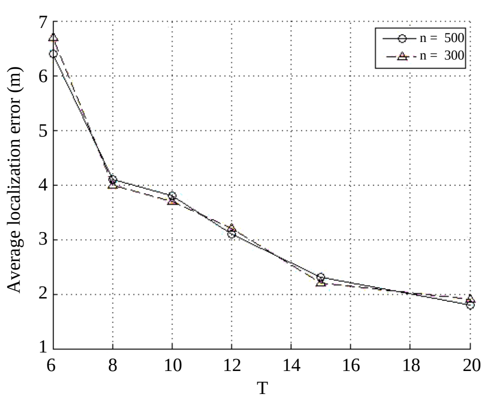

Fig. 3: Average Localization Error when T = 6, 8, 10, 12, 15, 20

separately 1, 2, ..., n and communication radius r. Divide the two-dimensional area as T×T square grids. The objective of node localization is to compute the estimated location value $(\hat{x}_i, \hat{y}_i)$ of unknown nodes, so as to make $(\hat{x}_i, \hat{y}_i)$ close to the real coordinates (x_i, y_i) of unknown nodes as far as possible.

Relevant test parameters are configured as follows: all nodes are deployed under the two-dimensional environment of 60×60 m, with the communication radius to be r = 60 m and unknown node quantity of n = 500 and n = 300. Set T = 6, T = 8, T = 10, T = 12, T = 15 and T = 20 (i.e., the side lengths of square grids are 10, 7.5, 6, 5, 4 and 3 m). In this study, algorithm is realized by MATLAB and invoking libsvm software. Parameter σ in Formula 4 and parameter C_+, C_- in Formula 8 are automatically determined by the mechanisms of libsvm. Node distribution when n = 500 is shown in Fig. 2. The study employs Average Localization Error (ALE) to measure the performance of the algorithm. The formula is described as follows:

$$\text{ALE} = \frac{\sum_{i=1}^{n}\sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{n} \tag{10}$$

In the formula, $(\hat{x}_i, \hat{y}_i)$ stands for the estimated coordinate location of the ith node, while n refers to the quantity of unknown nodes.

It can be seen from Fig. 3 that, with the quantity increases, average localization error decreases. However, when $T \geq 15$, the descending tendency of average localization error is not obvious. This is mainly caused by classification error. In the meanwhile, in the figure, the difference of average localization error between m = 300 and m = 500 is not significant, which proves that, when the network is filled with dense nodes, the localization effect of MSVM is quite stable and is insensitive to node distribution and quantity.

CONCLUSION

Wireless sensor network nodes normally have limited energy and beacon nodes are ordinarily of high cost. In allusion to the fact, the study takes advantage of mobile beacon to generate virtual beacon. Moreover, inspired by learning algorithm, the study proposes SVM-based MSVM algorithm to estimate the location of node, with high localization precision achieved. Yet, MSVM is can only be used with dense wireless sensor networks and has to establish direct communication between nodes. For this reason, the approach is not applicable to node localization of large scale networks. The next work is to research and develop machine learning based approach, which can be applied in large scale networks. Moreover, the approach can get quite satisfactory localization effect when nodes are sparsely distributed.

REFERENCES

Brunato, M. and R. Battiti, 2005. Statistical learning theory for location fingerprinting in wireless LANs. Comput. Networks, 47: 825-845.

Bulusu, N., J. Heidemann and D. Estrin, 2000. Gps-less low cost outdoor localization for very small devices. IEEE Personal Commun. Magazine, 7: 28-34.

Cortes, C. and V. Vapnik, 1995. Support-vector networks. Mach. Learn., 20: 273-297.

He, T., C. Huang, B.M. Blum, J.A. Stankovic and T. Abdelzaher, 2003. Range-free localization schemes in large scale sensor networks. Proceedings of the 9th Annual International Conference on Mobile Computing and Networking, September 14-19, 2003, San Diego, California, pp: 81-95.

Nguyen, X., M.I. Jordan and B. Sinopoli, 2005. A kernel-based learning approach to ad hoc sensor network localization. ACM Trans. Sensor Networks, 1: 134-152.

Niculescu, D. and B. Nath, 2001. Ad-hoc positioning system. Proc. Conf. IEEE Global Telecommunicat., 5: 2926-2931.

Niculescu, D. and B. Nath, 2003. Ad hoc positioning system (APS) using AOA. Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies, March 30-April 3, 2003, Rutgers University, Piscataway, pp: 1734-1743.

Pan, J.J., J.T. Kwok and Y. Chen, 2006. Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing. IEEE Trans. Knowl. Data Eng., 18: 1181-1193.

Priyantha. N.B., 2005. The cricket indoor location system. Ph.D. Thesis, Massachusetts Institute of Technology, Massachusetts, USA.

Smola, A.J. and B. Scholkopf, 2002. Learning with Kernels-Support Vector Machines, Regularization, Optimization and Beyond. 1st Edn., MIT Press, Cambridge.

Tran, D.A. and T. Nguyen, 2008. Localization in wireless sensor networks based on support vector machines. IEEE Trans. Parallel Distrib. Syst., 19: 981-994.

Whitehouse, C., 2002. The design of calamari: An ad hoc localization system for sensor networks. M.Sc. Thesis, University of California at Berkeley.