

Research on Node Scheduling Algorithm of Sensor Network based on Support Vector Machine

Xinyi Zhou

College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150000, China
zhouxinyi1207@163.com

Abstract. In the current Internet era with information explosion, SN (Sensor network) has important scientific research value and social and economic value in many fields because of its micro-intelligence and huge and basic perceptual data. Ensuring the coverage quality of the network cannot be at the expense of network connectivity. The decrease of network connectivity is equivalent to the failure of some nodes in the network, and the resulting data collection hole will affect the judgment of end users. In this paper, the scheduling algorithm of SN nodes based on SVM (Support vector machine) is studied. Considering the excellent performance of EN (ensemble learning), a scheduling algorithm of EN-LS-SVM is proposed by introducing bagging and random subspace method. The research results show that the network lifetime of the proposed algorithm is extended by 180s and 13.28% compared with the traditional algorithm on the premise that the network coverage is not less than 80%. The algorithm proposed in this paper can reduce network coverage vulnerabilities, keep network coverage at a high level and ensure network QoS.

Keywords: Support Vector Machine; Sensor Network; Node Scheduling.

1. Introduction

Coverage is the embodiment of SN (Sensor network)'s ability to perceive the physical world, which is often used to describe the standard of monitoring QoS (Quality of Service) [1-2]. How to prolong the network life cycle while maintaining the network coverage has become a key issue in SN research. In the current Internet era with information explosion, SN has important scientific research value and social and economic value in many fields because of its micro-intelligence and huge and basic perceptual data. At present, in most node scheduling schemes, the sleep time of nodes is fixed in the whole network life cycle, which leads to that some active nodes can't wake up in time because of the fixed sleep time when they finish processing data. Based on the research background of SN, this paper deeply analyzes its key technologies, focuses on the scheduling algorithm of SN nodes based on SVM, and analyzes its performance.

2. Problem Description

The coverage control of SN is to optimize the state distribution of network nodes under the condition of considering the performance of network, such as the number of working nodes, coverage, connectivity and load balance, so as to realize the efficient use of network energy [3-4]. SN coverage control mainly includes two aspects: ensuring network coverage quality and network connectivity [5]. Ensuring the coverage quality of the network cannot be at the expense of network connectivity. The decrease of network connectivity is equivalent to the failure of some nodes in the network, and the resulting data collection hole will affect the judgment of end users.

Assume that n wireless sensor nodes with perceived distance r are randomly deployed in the monitoring area A . For each sensor node $i(i = 1, 2, \dots, n)$, the following assumptions are made:

The location information of nodes in the network is unknown; Nodes are randomly deployed in the monitoring area, and the position of nodes will not change after deployment. Network coverage reaches 100% when all nodes are active at the same time. The communication radius of a node is twice the perceived radius (Figure 1).

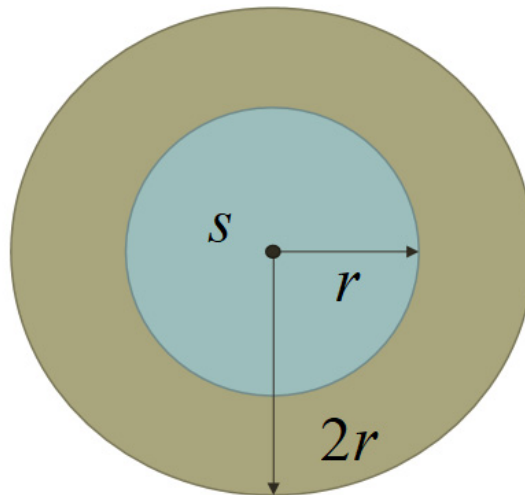


Figure 1. Node coverage area

In order to enhance the monitoring quality of SN and improve the network reliability, sensor nodes are usually deployed in the target area of interest on a large scale and with high density, which leads to the overlapping coverage areas of a large number of nodes in the network. In view of the large number of nodes in the sensor network, the node scheduling algorithm designed should have strong scalability and should be distributed as much as possible. Scheduling nodes in the sensor network can effectively use the energy of nodes and prolong the working time of the network, but at the same time it will bring great computational load to nodes, so we need to compromise these issues.

The definition of network coverage refers to the ratio of the sum of the areas covered by all active nodes to the total area of the monitoring area:

$$A_{Coverage} = \frac{A_{live} \cap A}{A} \times 100\% \quad (1)$$

Where $A_{Coverage}$ is the network coverage and A_{live} is the sum of the coverage areas of active nodes in the current network.

Because of the high density of nodes in three-dimensional environment, the information of neighboring nodes in local area is more abundant than that in two-dimensional, so the strategy of judging redundant coverage of nodes by information such as distance and number between neighboring nodes is more economical and efficient than that based on location information and random probability. Node scheduling algorithm can reduce the energy consumption of nodes and improve the energy utilization rate of the network by making the redundant nodes sleep, which is one of the important mechanisms to realize the balance of network energy consumption and prolong the network lifetime.

3. SN Node Scheduling based on SVM

At present, most of the research on node scheduling is mainly theoretical research, and it is usually simulated and verified in an ideal experimental environment, and the specific practical application is still rare. At present, most node scheduling algorithms assume that the network monitoring area is a regular whole without any obstacles, without considering the influence of environmental complexity on network fault tolerance, connectivity and perceived quality. Is the energy balance research of SN also a key point. The balance of energy consumption of the whole network plays a very important role in maintaining the balance of network data collection and avoiding network shrinkage and holes. Therefore, it is necessary and practical to continue to carry out in-depth research and analysis on the above issues.

The reason why the nodes on the boundary of the monitoring area die first is that the nodes on the deployment boundary are affected by the boundary factors, which makes it difficult for the nodes on the boundary to reach the coverage redundancy condition and thus have no chance to enter the dormant state. The effective perception rate is not simply the superposition of the perception rates of these sensors, because sensing the same target will be considered as repetitive behavior and should be ignored. In SN, the dynamic environment will have a certain impact on task scheduling. Therefore, the dynamic optimization goal of this paper not only needs to find the optimal solution, but also should adjust the scheduling scheme in time when the environment changes.

In the neighbor discovery stage, the sensor nodes in the network broadcast the neighbor discovery message, and their neighbor nodes reply the message to the node, and the node obtains the neighbor node information according to the replied message. Since the sensor node can have multiple sensing targets $g \in G_t, \forall t \in T$, the sensing targets need a certain perception rate f_{stg} and duration c_t on the sensor node $s \in S$. All sensor nodes must meet the following restrictions to ensure multi-task schedulability:

$$\sum_{t \in T} \sum_{g \in G_t} c_t \cdot f_{stg} \leq 1, \forall s \in S \quad (2)$$

Node scheduling technology is a key technology of SN, and its accuracy and anti-interference directly affect the validity of the whole system data [6-7]. In an environment like SN where there are a lot of interference and outliers, it is necessary to require the scheduling algorithm to have strong robustness. As a machine learning technology, SVM (support vector machine) is used for classification, pattern recognition and function estimation [8]. SVM is a binary classification algorithm based on supervised learning, and its core is to find the optimal separation hyperplane that maximizes the interval in the feature space. By using kernel method, linear SVM can be extended to nonlinear SVM [9]. LS-SVM (Least squares SVM) is widely used in various fields because of its superior performance.

Another ingenious feature of LS-SVM is that it adds a relaxation variable to deal with the possible noise problem of sample data, as shown in Figure 2 below. VM allows data points to deviate from the hyperplane to a certain extent, which is the outlier value that can be set in SVM algorithm, corresponding to the length of black implementation in the above figure. With the addition of relaxation variables, SVM not only pursues the best local effect, but also considers the overall situation of sample data distribution.

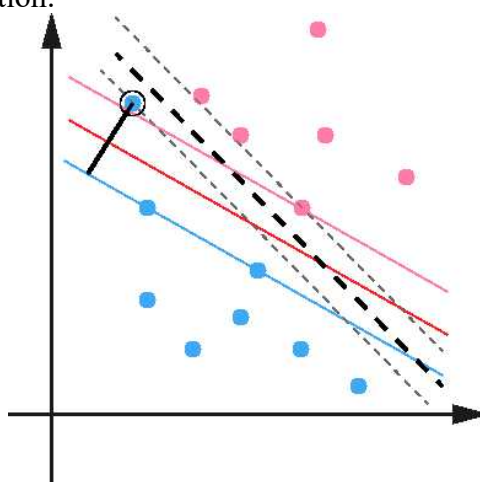


Figure 2. Relaxing variables to process sample data

On the premise that all the training samples in the input space are correctly classified by SVM (that is, the empirical risk is zero), the samples are optimally classified. For nonlinear problems, if we simply classify them in the original space, we can't achieve satisfactory results. Therefore, we can change nonlinear problems into linear problems in a high-dimensional space and realize a dependency

relationship with output, so as to achieve a linear relationship between feature vectors and position coordinates that we want to build in this paper.

EN (ensemble learning) is a machine learning method that has emerged in recent years. The EN method thinks that combining multiple learners can achieve better generalization ability than using a single learner [10]. The construction of EN model includes two steps: firstly, several basic learners are generated in parallel or serial mode. Among them, the parallel method is called bagging and the serial method is called boost. In order to ensure that the EN model has good generalization performance, the performance of each base learner should be as strong as possible, and different base learners should be as different as possible.

Considering the excellent performance of EN, a scheduling algorithm of EN-LS-SVM is proposed by introducing bagging and random subspace method. Different from the traditional bagging regression model, which usually uses the average weight combination method, EN-LS-SVM algorithm uses the weighted combination method, and the weight of the base learners with good performance in the ensemble model should be higher than that of the base learners with poor performance.

For the data set D containing n samples, sample m samples at a time, and the probability that the samples are not drawn after each sampling is $\left(1 - \frac{1}{m}\right)^m$. When it approaches infinity, there are:

$$\lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^m = \frac{1}{e} \approx 0.368 \quad (3)$$

When the sample set contains many redundant or irrelevant features, using random subspace method to construct EN model can effectively improve the generalization ability of learners.

The salient feature of SVM is the mapping of sample space. In fact, when the kernel function processes the input samples, it establishes a mapping from the original space to the high-dimensional space, so that only the inner product operation is needed. $K(x, x_i)$ is called the kernel function, and the function of the kernel function is to replace the inner product operation in the high-dimensional space. Accordingly, the objective function in the previous section is transformed into:

$$\lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^m = \frac{1}{e} \approx 0.368 \quad (4)$$

Support vector regression machine firstly transforms the input space into a high-dimensional space through the nonlinear change defined by the inner product function, and then finds the generalized optimal classification surface in this space.

The EN-LS-SVM scheduling model is obtained as follows:

$$\hat{z}(d) = \sum_{i=1}^N \alpha_i K(d, d_i) + b \quad (5)$$

EN-LS-SVM uses equality instead of SVM's inequality constraints, so that the problem becomes solving a set of linear equations instead of solving quadratic programming, which greatly reduces the computational complexity of the algorithm.

4. Simulation Verification

In this section, the improved SN simulation platform under J-Sim is used to simulate the algorithm proposed in this paper, and the performance is compared with the classical coverage algorithm.

It can be seen from Figure 3 that the greater the density of nodes in the network, the more boundary nodes can be identified, and the accuracy will also increase, indicating that the algorithm is better. The reason is that the higher the density of nodes, the more the number of neighbors that nodes can refer to, and the more accurate the threshold of the minimum number of neighbors judged by redundancy strategy.

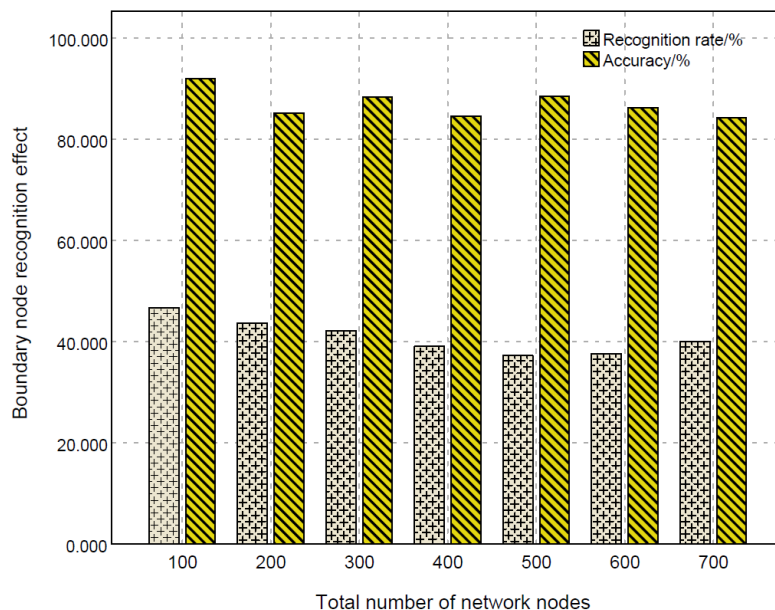


Figure 3. Identification effect of boundary nodes

Because of the randomness of network node deployment and the uncertainty of neighbor node location relationship, the algorithm can't completely identify all border nodes, and the accuracy of identification can't guarantee 100%. Even so, the algorithm still compensates the dormancy probability of most border nodes, and obviously inhibits the boundary contraction problem caused by network aging. The energy consumption balance of the network has also been significantly improved.

Fig. 4 shows the comparison of network coverage with time between the algorithm proposed in this paper and LDAS (lightweight deployment-aware scheduling).

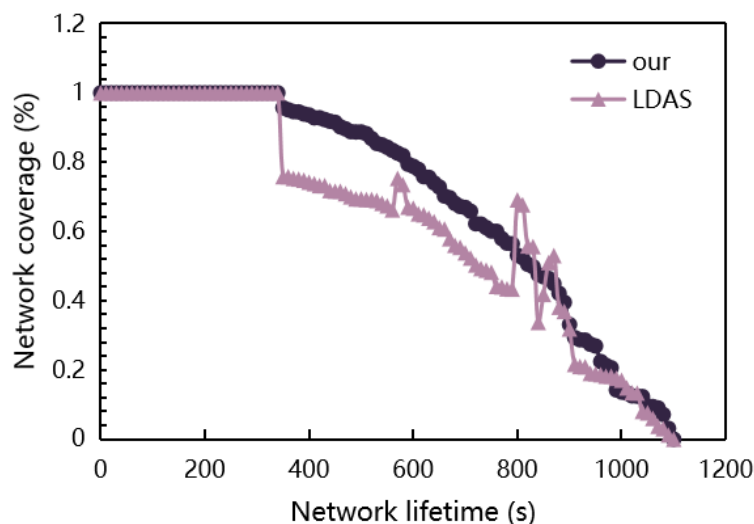


Figure 4. Network coverage comparison chart

It can be seen that the network is available when the network coverage is not less than 80%. On this premise, the network lifetime of the algorithm proposed in this paper is extended by 180s and 13.28% compared with LDAS algorithm.

The backup node strategy is adopted in the algorithm proposed in this paper. When the node fails, the backup node works instead of the failed node, and the coverage loopholes caused by the node failure are covered by the backup node, which reduces the occurrence of network coverage loopholes, keeps the network coverage rate at a high level, and ensures the network QoS.

5. Conclusion

Coverage control of SN is to optimize the state distribution of network nodes under the condition of considering the number of working nodes, coverage, connectivity and load balance of the network, so as to realize the efficient use of network energy. Based on the research background of SN, this paper deeply analyzes its key technologies, focuses on the scheduling algorithm of SN nodes based on SVM, and analyzes its performance. The research results show that the greater the density of nodes in the network, the more boundary nodes can be identified, and the accuracy will also increase, indicating that the algorithm is better. Under the premise that the network coverage is not less than 80%, the network lifetime of the proposed algorithm is extended by 180s and 13.28% compared with the traditional algorithm. The algorithm proposed in this paper can reduce network coverage vulnerabilities, keep network coverage at a high level and ensure network QoS.

References

- [1] Movassagh M, Aghdasi H S. Game theory-based node scheduling as a distributed solution for coverage control in wireless sensor networks[J]. *Engineering Applications of Artificial Intelligence*, 2017, 65 (10): 137-146.
- [2] Junior E C, Vieira L, Vieira M. 3DVS: Node Scheduling in Underwater Sensor Networks using 3D Voronoi Diagrams[J]. *Computer Networks*, 2019, 159(4):73-83.
- [3] Chen H, Li X, Zhao F. A Reinforcement Learning-Based Sleep Scheduling Algorithm for Desired Area Coverage in Solar-Powered Wireless Sensor Networks[J]. *IEEE Sensors Journal*, 2016,7(6):2763-2774.
- [4] Lee S, Kwon B, Lee S, et al. BUCKET: Scheduling of Solar-Powered Sensor Networks via Cross-Layer Optimization[J]. *IEEE Sensors Journal*, 2015, 15(3):1489-1503.
- [5] Aghdasi H S, Abbaspour M. Energy efficient area coverage by evolutionary camera node scheduling algorithms in visual sensor networks[J]. *Soft Computing*, 2016, 20(3):1191-1202.
- [6] Ying X, Zeng Z R, Ou D. An energy efficient hole repair node scheduling algorithm for WSN[J]. *Wireless Networks*, 2015, 23(1):1-14.
- [7] Bakht A J, Motameni H, Mohamadi H. A learning automata-based algorithm for solving the target k-coverage problem in directional sensor networks with adjustable sensing ranges[J]. *Physical communication*, 2020(10):42.
- [8] Karunanithy K, Velusamy B. CSDGP: cluster switched data gathering protocol for mobile wireless sensor networks[J]. *IET Communications*, 2019, 13(18):2973-2985.
- [9] Bahi J, Elghazel W, Guyeux C, et al. Resiliency in Distributed Sensor Networks for Prognostics and Health Management of the Monitoring Targets[J]. *Computer Journal*, 2018, 59(2):275-284.
- [10] Chatterjee P, Member S, IEEE, et al. Load Balanced Coverage with Graded Node Deployment in Wireless Sensor Networks[J]. *IEEE Transactions on Multi-Scale Computing Systems*, 2017, 2(2):1-1.