Detection of Outdated Sensors in Wireless Network via a New Protocol

Hasanian Ali Thuwaib
Kut University College, Department of Law, Alkut, Wasit, Iraq.
E-mail: hasaneen.A.dweeb@alkutcollege.edu.iq

Ridhab Sami Abd-Ali
Computer Techniques Engineering Department, Al-Mustaqbal University College, Babylon, Iraq.
E-mail: RidhabSami@mustaqbal-college.edu.iq

Safaa Hadi Abdula Ali Altai
Education Directorate of Muthanna Province, Muthanna Province, Iraq.

Received March 16, 2021; Accepted June 30, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V18SI04/WEB18146

Abstract

A novel method is proposed using the nonlinear mapping with kernel functions to correctly locate the outdated sensors in a wireless sensor network (WSN). Such detection system used Cornell regression and solved via the vector support regression (VSR) plus multi-dimensional backup vector regression (MBVSR). The developed method was simplistic and effective without the need of any additional hardware for any measurement. It required only the vicinity and information of location from the anchor nodes to detect the outdated sensors. It was achieved in three stages including the measurements, kernel regression, and stepping stage. First step measured the proximity information from a given grid. The relationships between the proximity and geographic distance among the sensors’ nodes were generated in the kernel regression stage. For the stepping phase, every sensor node found its location in the distributed way via the kernel regression. Simulation results showed the robustness and high efficiency of the proposed scheme.

Keywords

WSN, Sensors, SVR, MSVR, Nonlinear Mapping, Kernel Function.

Introduction

For environmental monitoring, a wireless sensor network (WSN) have been developed, prototyped, and incorporated into the information system which became fully operational since May 2016 (Sousa et al, 2017). The WSNs have various applications in tracking of
objects (Li et al., 2011; Sahoo et al., 2013), distance managing (Pogkas et al., 2005), and intelligent home devices (Lee et al., 2007). Usually, the WSN is consisted of several sensors and comparatively fewer anchor nodes. The sensor nodules being very tiny their positions are usually unidentified. Conversely, the anchor nodes possess extra hardware and their positions are recognized in advance. For most of the applications, the knowledge about the locations of these sensor nodules in a known network is significant. The easiest resolution for positioning these nodes is to connect a global positioning system (GPS) (Hofmann-Wellenhof et al., 1993) for all the sensor nodes. However, this is highly expensive and tedious because the sensor nodes need inadequate resources. In essence, the location information of all the sensor nodes is prerequisite for various applications of the WSNs, (Khedr et al., 2011; Chung et al., 2011). In addition, the geographic positions recognition of these sensor nodules in the WSNs is given from the minimum position information of the anchor nodules and some measurement between the nodes.

Over the past few decades, numerous studies have been conducted to develop robust algorithms for accurately locating the sensors in various WSNs (Bulusu et al., 2001; He et al., 2003; Langendoen et al., 2003; Lim et al., 2005; Liu et al., 2012; Mao et al., 2007; Nagpal et al., 2003; Niculescu et al., 2003; Savvides et al., 2001; Shang et al., 2004; Shilton et al., 2008) Generally, these algorithms are divided into two classes based on the interval and non-interval types. The range-based and location-based algorithms are used for the interval measurements that include the distance and angle estimation from the anchor nodes. The range-based and location-based algorithms yield relatively accurate estimates without any rigorous procedures wherein the software are relatively expensive (Savvides et al., 2001). Conversely, the unplanned positioning algorithms only use the connectivity data between sensors and thus no extra hardware are needed by them for the measurement.

For the large-scale applications of the WSNs involving hundreds and thousands of sensors, unadjusted algorithms are preferred because the addition of extra hardware to all sensors is essential. However, the inaccuracy of the location estimation is the main disadvantage of an unadjusted algorithm wherein the deficiency of measuring the distance between the nodes is its main limitation. Despite some efforts, an unplanned positioning algorithm for accurate location estimation of the sensors in the WSNs is far from being achieved. Based on these factors, the present work exploited the machine learning techniques for detecting the outdated sensors in the WSNs.

Present article is structured in the following way: Section 2 described earlier works concerning the unplanned location algorithms in WSN. Section 3 emphasized the essence of the proposed new methods for locating the sensors. Section 4 presented the simulation
results, data analyses and compared the performance of the newly developed system with the existing advanced methods. Section 5 concluded the paper.

Related Works

Earlier, many studies have been performed on the unplanned location in the WSNs (Chuang et al, 2002; Sanchez-Fernandez et al, 2004). The Distance Vector-Hop (DV-Hop) approach was proposed to calculate the mean hop-distance between the sensor nodes (Niculescu et al, 2003). Bulusu et al, have proposed a coarse-grained graphene-based algorithm called the non-intermittent proximity and center of gravity positioning (CLA) algorithm (Bulusu et al, 2001). In this scheme, anchor nodes were distributed in their locations and every sensor node calculated the location according to the attached anchor nodes. Later, the approximate Point-In-Triangulation (APIT) locating algorithm was introduced (He et al, 2003). This algorithm could split the location by separating the surroundings into triangular areas amid the anchor nodes, whereas the probable region was based on the existence of every sensor node within or exterior to the triangle.

Liu et al, have introduced an improved localization algorithm (ILA) depending on the APIT and vertical binary feature (Liu et al, 2012). In ILA, the sensor modules were categorized into one of the 4 groups. The others were the normal node, lateral node, hypo-isolated node and distinct node wherein each node was classified by different rule-based classifications. Lim and co-workers (Lim et al, 2005) proposed a proximity distance map (PDM) algorithm for managing the anisotropic network. In this algorithm, the locators were considered as an embedded problem that mapped linearly the geographical separation to the neighborhood measure thereby resolved the issue through a linear map matrix. To be more specific, the PDM created a linear conversion matrix that could be mapped from the neighborhood vector to the distance one. For outside settings, (Zhong et al, 2011) introduced a regulated signature distance (RSD) that could easily be embedded within the connection-based and location-based algorithms for improving the positioning accuracy (Zhong et al, 2011).

For the precise detection of the location in the WSNs, several kernel-based methods have been developed using the classifications or regression approaches. (Pan et al, 2006) developed a correlation focal analysis-based location algorithm. (Brunato et al, 2005; Tran et al, 2008) utilized the Support Vector Machine (SVM) to locate the obsolete nodes. A WSN-based SVM-based algorithm was devised to determine the sensors location accurately (Tran et al, 2008). A kernel regression algorithm was introduced by (Kuh et al, 2006) that used a recursive least-square algorithm to resolve the location-related issue. A
matrix regression method was presented (Honeine et al, 2008) following the kernel method. Various SVR algorithms were proposed via the received signal strength (RSS).

Ibrahim and et al, have used the Optimized Network Engineering Tools (OPNET) Modeler simulator and demonstrated multiple ways of detecting the intrusions through showing and recognizing the variations of the IP address distributions of the data packets in the simulation (Ibrahim and et al, 2020). C. A WSN was installed on the coastal region using 802.15.4 to achieve the mobile connectivity among the feasible nodes (Flores-Cortés et al, 2017). Based on the SVR location multidimensional presentation an algorithm was developed (Lee et al, 2013). In addition, kernel-based algorithms provided perfectly accurate results for the detection. All the aforementioned methods except the one proposed by (Tran et al, 2008) used the distance measurement like RSS amid sensor and anchor nodes to resolve the positioning issues. Based on these factors, we presumed that each sensor must be linked to nearly all the sensor and anchor nodes in a specified WSN for the fruitful SVR model. The proposed novel kernel-based approach without any distance measure was shown to outperform the existing state-of-the-art techniques. However, this hypothesis may not be practical for several real-world applications.

Sensors Location Detection in WSN using Kernel Regression

Most of the previously developed methods did not use the general or topological information from all WSNs. It rather only used the local information about the sensor node, thus reducing the location function of the anisotropic WSNs. To overcome this limitation, the LSVR plus multi-dimension backup SVR (MSVR) methods were introduced for the accurate location detection as described hereunder.

1. LSVR is a one dimension location regression approach that uses the connection measurements and estimation of sensor node locations by calculating the arrival time of each receiver (multilateration).
2. LMSVR is a two-dimension location regression scheme that uses the connection measurement to estimate the sensors node locations.

Figure 1 illustrates the overall function of the developed kernelized approach for solving the WSN location problem which can further be applied to the ML theory. Majority of the SVRs being one-dimensional (Musicant et al, 2004) their extensions are not straightforward.
Figure 1 Overall function of the proposed kernelized method

Locations Detection via LSVR

The LSVR is achieved through 3 steps as follows:

**Step 1:** It is the measurement stage wherein every anchor node \((S_i \in A)\) and \(S_j \in \Sigma STS\) sensor node exchanges each hop data where every anchor node sends its position information to other anchor nodes.

**Step 2:** It is the SVR stage where every anchor node assesses its SVRi spacing using the SVR model and transmits it to all the sensor nodes \((S_j \in \Sigma TS)\). Thus, a total of M model spacing \((SVR1, SVR2, \ldots, SVRM)\) are allocated to all sensor nodes \(S_j\).

**Step 3:** It is the staging step where each sensor node \((S_j)\) is calculated from all the anchor nodes depending on the distance model \((SVR1, SVR2, \ldots, SVRM)\) that are presented in the SVR stage. Ultimately, the position is estimated by this method for calculating the arrival time of each recipient in a distributed manner. Now, it is customary to describe briefly each stage.

1. **Measurement Step**

Consider the neighborhood vector between \(S_i \in A\) and other anchors via:

\[
p_i = [p_{i,1} \ldots p_{i,m}]^T \in Z^M
\]

where \(p_{i,j} = p(S_i, S_j)\), \(S_i, S_j \in A\) and \(p_i\) for \(i = 0\); \(z\) is the set of integer numbers. The total information of the proximity between the anchor nodes was displayed using:

\[
p = [p_1 \ldots p_M] \in Z^{M \times M}
\]

Similarly,

\[
d_i = [d_{i,1} \ldots d_{i,m}]^T \in R^M
\]
Consider the geographic distance vector amid anchor nodes ($S_i \in A$) and additional anchors where $d_{i,j} = d(S_i, S_j)$, $S_i, S_j \in A$ and $d_i, i = 0$. The general distance among the anchor nodes yields:

$$D = [d_1 \ldots d_M] \in R^{M \times M}$$

Consider $\bar{p}_k$ as the neighborhood vector between the $k^{th}$ node of the Sk²ili sensor and all anchor nodes ($S_i \in A$). Then,

$$\bar{p}_k = [\bar{p}_{k,1} \ldots \bar{p}_{k,M}]^T \in Z^M$$

Thus,

$$\bar{p}_{k,i} = p(S_k, S_i), S_k \in \Sigma, S_i \in A$$

Initially, each anchor node ($S_i \in A$) sends a hi-res message to all additional nodes with both sensor and anchor ($S_i \in S$) nodes so that every anchor node ($S_i \in A$) and the sensor node (Sk²) in the proximity vector knows $p_i$ and $p \rightarrow _k$. Every anchor node directs an INFO message that includes the neighborhood vector $p_i$ and position information (pos($S_i$)) to the other anchor node $S_j \in A$, so that every node makes anchor (P and D). Communication expenses for the measurements step $M(M+N-1)$ and $M(M-1)$ message is the INFO one.

2. **Learning Step of SVR**

In the learning stage, every anchor node ($S_i \in A$) creates a kernel regression model to calculate the distance to each other node. This means that every node gathers the anchor ($S_i$) and M data pair of learning. For example, $(p_j, d_{i,j}) \in Z^M \times R$ ($j = 1, 2, \ldots, M$) and anchor ($S_i$) to all nodes include the anchor $S_j \in A$. Then the anchor node ($S_i$) gives the train the largest nonlinear function written as:

$$f_i(p_i)w_i^T \varphi(p_i) + b_i$$

3. **Locations**

Each node of the Skesionice sensor starts the location after receiving the SV from the SVR model and the adjacent vectors of the anchor nodes, which is given by:

$$\{(\hat{\alpha}_1, \hat{\alpha}_1^*, b_1, p_1), (\hat{\alpha}_2, \hat{\alpha}_2^*, b_2, p_2), \ldots, (\hat{\alpha}_M, \hat{\alpha}_M^*, b_M, p_M)\}$$
The estimated distance $d_k$ can be determined using the relation:

$$
\tilde{d}_k = [\tilde{d}_{k,1} ... \tilde{d}_{k,M}]^T = [f_1(\tilde{p}_k) ... f_M(\tilde{p}_k)]^T,
$$

$$
= \left[ \sum_{j=1}^{M} (\hat{a}_{1,j} - \hat{a}_{1,j}^*)k(\tilde{p}_k, p_j) + b_1 ... \sum_{j=1}^{M} (\hat{a}_{M,j} - \hat{a}_{M,j}^*)k(\tilde{p}_k, p_j) + b_M \right]^T
$$

After estimating the distance ($\tilde{d}_k$), the anchor nodes are evaluated and the arrival time of the receiver is calculated to determine the position of the sensor nodes $S_k$.

**Location Detection via LMSVR**

The second approach is a multidimensional version of the kernel regression for locating the sensor node accurately. In the LSVR approach, a one-dimensional distance regression model is constructed based on the vicinity measurement and the location of a $S_k^2$SciT sensor node is approximated by relating the calculated arrival time of the waves to each receiver and the estimated distance of the anchor nodes. An anchor node is presumed to be used as a hole (empty) node that teaches the MSVR model and thus must be the utmost useful knot. The LMSVR is achieved in 3 stages given by:

**Stage 1:** It is the measurement step where every anchor node ($S_i \in A$) and $S_j^2\Sigma\Sigma\Sigma TS$ sensor node exchanges the information on the number of hops. In addition, every anchor node ($S_i$) transfers its location information to one of the anchor nodes $S_k^2A$.

**Stage 2:** It is the MSVR stage in which the $S_k^2A$ nodes evaluates the position model via the developed MSVR approach and releases it to entirely $S_j^2\Sigma\Sigma\Sigma$ sensor nodes.

**Stage 3:** It is the positioning stage wherein each sensor node $S_j^2\Sigma\Sigma\Sigma TS$ evaluates its location via the MSV model provided by each well node at the earlier MSVR stage.

**Simulation Results and Discussion**

Based on the above-mentioned new approach using the nonlinear mapping with kernel functions to precisely locate the outdated sensors in a WSN simulation was performed. The SVR and MSVR methods were used to generate the data. The obtained computational results were analyzed and discussed in this section. In addition, we compare the efficiency of the new detection scheme with some of the existing three techniques reported in the literature. These methods used the same conditions such as the connection information among all nodes and locations of the anchor nodes (no further measures like RSSI, TOA, and AOA were needed), and the comparable ranges of all the anchor and sensors nodes. In the simulation, the anchor sensors and nodes were distributed over an area of (100 × 100
square) in the presence of two types of networks such as isotropic (where the nodes were distributed unevenly with identical density and connectivity over the entire networks) and anisotropic grids (wherein the nodes were non-uniform). To construct an isotropic network, the following probability density function was used (with $D_x=100$ and $D_y=100$):

$$f(x, y) = \frac{1}{D_xD_y} \text{ for } 0 \leq x \leq D_x \text{ and } 0 \leq y \leq D_y$$

Figure 2 and 3 illustrates an isotropic and anisotropic grid architecture used in the present simulation. Three different simulation cases were implemented as described below.

**Case I:** A total of 3 WSNs with changed node density were taken: (1) a dispersed network having anchor and sensor nodes of 21 and 70, respectively, (2) an intermediary network having anchor and sensor nodes of 21 and 210, respectively, and (3) a dense network having anchor and sensor nodes of 100 and 400, respectively. The ranges of communication (R) for all nodes were kept constant.

**Case II:** The network extent was constant at 300, and the ratios of the anchor nodes were varied from 10-30% to evaluate the performance modification in the performance wherein the communication interval was kept constant.

**Case III:** A more realistic environment was selected. For all nodes, the communication ranges (R) were fixed to an average value and the irregular radio propagation model was applied. The fixed network size was 300 and the ratios of the anchor nodes were varied from 10-30% to estimate the alteration in the performance.

![Figure 2 The isotropic network structure used in the present experiment](http://www.webology.org)
Figure 3 The anisotropic network structure used in the present experiment

Figure 4 compares the obtained simulation results for the location error of an isotropic network under different anchor populations. It was observed that the denser network produces minimum location error regardless of the location method, indicating a clear balance amid the node densities and location efficiency. Nonetheless, the saturation of the location accuracy showed that the network with unnecessary high densities can be evaded. The newly introduced LMSVR technique consumed higher time compared to the existing approaches, particularly for the dense networks which were attributed to the rapid increase in the MSVR calculation as a learning set or an enhancement of the amount of anchor nodes. The positioning function for the network is shown using the box plot.
Figure 4: Errors in the location of an isotropic network when compared with different population of the anchor nodes

Figure 5 compares the obtained simulation results for the location error of an anisotropic network under different anchor populations. In this case the number of nodes was kept constant to 300 and the anchor ratio was changed from 10 to 30%. The positioning function for the network is shown using the box plot.
Both Figures 4 (a)-(e) and 5 (a)-(e) display the network location results for the corresponding anchor node ratios of 10, 15, 20, 25, and 30%. The boxes elucidate the outcomes obtained using the DV-hop, PDM, LSVM, LSVR and LMSVR approaches. Clearly, the location accuracy depended on the amount of anchor nodes for the kernelized LSVM, LSVR, and LMSVR location algorithms. Nonetheless, with too much decrease in the amount of the anchor nodes, the proposed scheme revealed relatively improved performance compared to the existing advanced techniques.
Conclusion

We developed a new kernelized regression approach to identify outdated sensors in a given WSN accurately and efficiently. The problem of locating the outdated sensors in the network was modeled as a kernel regression problem. This newly developed model was simulated for three different cases and the results were compared with the existing methods reported in the literatures. The performance of the newly developed scheme was evaluated and the results were validated. The proposed scheme with nonlinear mapping and embedded kernel function revealed better performance that the existing methods. Unlike the existing sensor location detection scheme, only the modest proximity information between sensor nodes was employed for measurement in the new method in which every sensor node assessed its position in a distributed fashion. The comparative assessment demonstrated that the proposed method depending on the mean location error and compatibility outperformed the existing state-of-the-art techniques.

References


