Research Article

Associated Clustering Strategy for Wireless Sensor Network

Yajuan Sun, Huanzhao Wang, Kewang Zhang, and Xudong Yang

Department of Computer Science and Technology, Xi’an Jiaotong University, Xi’an 710049, China

Correspondence should be addressed to Yajuan Sun; sunshine2225@163.com

Received 13 January 2014; Accepted 28 March 2014; Published 15 April 2014

Academic Editor: Shanshan Zhao

Copyright © 2014 Yajuan Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the soil moisture monitoring problem and propose a WSN associated clustering strategy based on spatiotemporal data correlation, which ensures that the nodes within each cluster can share a good data correlation and consequently makes the cluster head do the data fusion more efficiently. As a result, the energy of each node will be saved and the lifetime of the whole sensor network will be extended. In the associated clustering strategy, the different clusters can be divided by the correlation characteristics of nodes data, which is based on a dynamic model and a correlation characteristics model after the correlation coefficient analysis. Simulation results show that our proposed associated clustering strategy works very well in soil moisture measurement. Moreover, as compared with the traditional random clustering, the associated clustering strategy based on data correlation achieves better performance for each cluster, and will be more efficient in data fusion at the cluster heads.

1. Introduction

The improvement of embedded system design, sensor node design, and low power wireless communication techniques has made the large-scale wireless sensor networks (WSN) become an attractive solution for many applications [1]. WSN has features such as large-scale and high-density, independence from infrastructure, self-organizing, and adaptive network topology, making it exhibit a distinct advantage in many applications. When used in industrial or agricultural fields [2], WSN can provide cost-effective data monitoring and meanwhile bring several remarkable superiorities, such as concealment, ease of deployment, timeliness of data, reliability, and high coverage density.

Clustering [3] is a network management method, which divides the network nodes into several separated subsets according to some certain rules. Cluster head is responsible for collecting the data within its cluster and forwarding it to the base station. Data integration in cluster head can reduce the redundant data in networks and, consequently, save the energy consumption and prolong the network lifetime. Existing clustering strategy, such as LEACH, GAF, TEEN, and PEGASIS, is mostly cluster by the distribution of cluster head, distance between nodes, remaining energy, and network topology. Data correlation within one cluster is rarely a major consideration in clustering. In practical applications, data between nodes are generally correlated [4]. However, traditional random clustering strategy has poor data correlation within cluster, which leads to low data integration and thus generate massive data redundancy in network. Therefore, WSN clustering strategy based on data correlation has become a hot spot. In clustering process, partitioning the associated nodes into one cluster can get efficient data integration in cluster head. Accordingly, reducing the traffic in network greatly saves network resource and energy.

The main contribution of this work is to exploit the spatiotemporal correlation to divide the WSN clusters. The inherent correlation of a specific application is taken to be the basis of clustering. Take soil moisture measurement correlation to divide the WSN clusters. The inherent correlation of a specific application is taken to be the basis of clustering. Take soil moisture measurement application as an example and establish a universal associated clustering strategy, which is also suitable for industrial applications when the soil moisture model is replaced by the particular industrial model. Based on Rodriguez-Iturbe soil moisture model, establish a dynamic model of soil moisture in greenhouse. Provide a clustering strategy based on spatiotemporal data correlation, which makes nodes in a cluster share a good correlation so as to do more efficient data integration, consequently, saving energy and extending the network lifetime.
2. Related Works

Several methods have been developed to analyze the spatial and temporal correlation characteristics in WSN. In [5], variograms are used to analyze spatial correlation. In [6], spatial correlation is used for schedule so as to achieve energy efficient data aggregation. In [7], an error-bounded data compression using data spatial-temporal correlation is provided. However, in the above studies, they mostly get data correlations from a single dimensional relation, spatial or temporal; besides, associated correlation is also nonqualitative. Particularly, spatial correlation is still just equivalent to physical location correlation, but not the data spatial correlation. In [8, 9], linear correlation between n-dimensional random variable and the definition of linear correlation coefficient of multiple variables are studied.

In [10], temporal and distance correlation are both used as clustering basis to reduce traffic in network to achieve the goal of energy saving. Similar clustering method is also mentioned in [11], which only takes the spatial correlation into account. In [12], a novel clustering algorithm based on correlation of sensor data is proposed. Its key is to express the data redundancy of WSN as formalized data correlation, thereby considering clustering from the perspective of data dependency. In [13], Yang et al. studied to define the correlation of soil moisture between different vertical depths by correlation analysis and R-type hierarchical clustering analysis.

In study of soil moisture model, Rodriguez-Iturbe et al. in [14] proposed several typical dynamic stochastic models of soil moisture and the corresponding probability density function of soil moisture. Rodriguez-Iturbe model is the one which considered more completely the dependence of random input and output of soil moisture. Main factors of soil moisture such as rainfall, vegetation, and soil are expressed as quantified model. Further discussions of spatial and temporal variability of soil moisture are proposed in [15]. However, in this model, rainfall was proposed as the main factor of soil moisture, which is difficult to achieve fine-grained soil moisture analysis model. As for the detection of soil moisture in greenhouse, irrigation replaced rainfall to be the main entry of soil moisture. In [16, 17], sprinkling irrigation is analyzed, which is an irrigation method widely used in greenhouse. However, its model only depends on some parameter coefficient.

In this paper, consider the main affect factors of soil moisture in greenhouse; a dynamic soil moisture model is established. Meanwhile, a spatiotemporal correlation characteristics model is proposed after the soil moisture model. Associated clustering strategy based on data correlation is finally raised, which can implement efficient data integration within each cluster and reduce traffic in network, so as to achieve energy balanced efficient WSN.

The remaining of the paper is organized as follows: soil moisture correlation characteristics model is established in Section 3, which is based on a dynamic soil moisture model and an irrigation model. Section 4 proposes an associated clustering strategy using the idea of correlation clustering pedigree chart. Simulation results of clustering algorithm and verification of correlation are deployed in Section 5. Finally, conclusions and further works are mentioned in Section 6.

3. Data Correlation Characteristics

3.1. Soil Moisture Model. Soil moisture is affected by many factors like climate, rainfall, irrigation, soil structure, and so forth. So, it is difficult to establish an accurate soil moisture model by taking all the factors into account. Based on the classic Rodriguez-Iturbe soil moisture dynamic model, we established a soil moisture model specifically for greenhouse, which takes irrigation as a major factor of soil moisture entry. The model gives a complete expression of soil moisture input and output items. Ignoring the effects of terrain, it proposed a quantitative simulation for several major factors such as irrigation, vegetation, evaporation, and leakage; meanwhile, take the upper and lower bounds of soil water capacity into consideration. Soil moisture model can be described by the following water balance equation:

\[
\begin{align*}
\frac{dZ_r}{dt} &= (1 - \phi) I(u, t) - K_z \frac{s(u, t) - s_1}{1 - s_1}, \quad s > s_1, \\
\frac{dZ_r}{dt} &= (1 - \phi) I(u, t) - V \cdot s(u, t), \quad s_0 < s < s_1,
\end{align*}
\]

where \(s_1\) is the upper bounds of soil water capacity and \(s_0\) is the soil drought coefficient. Overflow and leakage are key factors associated with the current soil moisture when the extreme situation \(s > s_1\) occurred. \(S(u, t)\) is the soil moisture level at time \(t\) and location \(u\), \(n\) is the soil porosity and \(Z_r\) is the depth of the root zone, and the product \(nZ_r\) represents the capacity of the soil to store water. \(I(t)\) is the intensity of the irrigation process and \((1 - \phi)\) is the net irrigation coefficient (usually 0.4–0.6), which is determined by the plant species and the condition of vegetation. \(K_z\) is soil hydraulic conductivity. The sum of evaporation transportation, leakage, and runoff losses is associated with the current soil moisture and represented by \(V \cdot S(u, t)\), in which \(V\) is the soil water loss coefficient, depending on vegetation coefficient \(K_v\).

3.2. Irrigation Model. Assume a greenhouse environment using sprinkler irrigation technology, with sprayers uniformly distributed in the crop area. As shown in Figure 1, distance between sprayers is \(D\); each sprayer uniformly watered a circle whose radius is \(R\). Consider that no area is irrigated by 3 sprayers at the same time, and thus we have \(2R > D > \sqrt{3}R\).

In Figure 1, take the lower left point of the monitoring area as the coordinate origin, with sprayers located at \((mD, nD)(m, n = 0, 1, 2, \ldots)\). For a point in the area, its irrigation coverage degree \(C\) is the number of the sprayers whose irrigation range covered the point (distance between the sprayer and \((x, y)\) is smaller than \(R\)). Consequently, the irrigation intensity is expressed as \(C(x, y) \cdot X\), in which \(X\) is irrigation intensity of one single sprayer.
Therefore, irrigation model of point \((x, y)\) is represented as follows:

\[ I(x, y, t) = C \cdot X \cdot t. \] (2)

### 3.3. Correlation Characteristics Model.

Based on the soil moisture model and the irrigation model, we can get spatiotemporal correlation characteristics model of soil moisture, which indicate the relation between each influence factor and the correlation characteristics. Accordingly, it can better guide our clustering strategy when network is deployed.

The correlation coefficient is used to define the linear correlation of two variables. Corresponding, we use it to describe the correlation of soil moisture at two different points. The correlation coefficients \(\text{cor}(S_A(t), S_B(t + h))\) characterize the correlation degree of soil moisture at point \(A\) and point \(B\). The bigger the \(|\text{cor}(S_A(t), S_B(t + h))|\) is, the bigger the correlation degree of soil moisture at point \(A\) and point \(B\) is.

The correlation coefficient is defined as 

\[ \rho = \frac{(\text{cov}(S_A, S_B))/((\sqrt{\text{var}(S_A)} \cdot \sqrt{\text{var}(S_B)})), \] which is calculated from the covariance and square deviation of soil moisture.

The soil moisture model in (1) can be normalized as

\[ \frac{dS_A(t)}{dt} = -a_A S_A(t) + b_A I_A(t), \] (3)

where \(a_A = \|V/(nZ_r)\|_A\) is the water loss coefficient of point \(A\). The net irrigation coefficient is \(b_A = [(1 - \phi)/nZ_r]_A\), associated with the intercepting characteristics of vegetation.

Accordingly, the soil moisture of point \(A\) is represented as

\[ S_A(t) = e^{-at} \left( \int e^{at} b_I(t) dt + c_1 \right) = e^{-at} b_A I_A(t - \omega) d\omega. \] (4)

The expectation of soil moisture at point \(A\) is calculated as

\[ E[S_A(t)] = \frac{b_A}{a_A} E[I(t)] = \frac{b_A \pi R^2}{a_A D^2} \cdot X \cdot t. \] (5)

The covariance of soil moisture at point \(A\) and point \(B\) is

\[ C_{AB}(h) = \int b_A b_B e^{-a_{A}t - a_{B}v} \text{cov} \]

\[ \times (I_A(t - u), I_B(t + h - v)) du dv \]

\[ = b_A b_B \cdot X^2 \]

\[ \times \left( C_A C_B + \frac{b_A b_B}{a_A a_B} \rho^2 \right) \]

\[ \times \left\{ \frac{2e^{-a_{A}h}}{(a_A + a_B)(\rho - a_A)} - \frac{2e^{-h}}{(\rho^2 - a_B^2)} \right\}, \] (6)

where \(C_A\) is the irrigation coverage degree of point \(A\) and \(\rho = (\pi R^2/D^2)\) describes the distribution density of irrigation sprayers.

According to the calculation of correlation coefficient, the spatiotemporal correlation characteristics model of soil moisture is proposed as follows:

\[ \text{cor}(S_A(t), S_B(t + h)) \]

\[ = \frac{n_A n_B - (b_A b_B/a_A a_B) \rho^2}{\sqrt{C_A^2 - ((b_A/a_A) \rho)^2}} \sqrt{C_B^2 - ((b_B/a_B) \rho)^2}} \]

\[ \cdot \sqrt{a_A a_B (\rho + a_A) (\rho + a_B)} \]

\[ \times \left\{ \frac{2e^{-a_{A}h}}{(a_A + a_B)(\rho - a_A)} - \frac{2e^{-h}}{(\rho^2 - a_B^2)} \right\}. \] (7)

This correlation characteristics model can be used to approximate the real data correlation to describe the correlation characteristics of soil moisture.

### 4. Associated Clustering Strategy

According to the correlation characteristics calculated in (7), the associated clustering strategy uses cluster analysis to divide associated sets into clusters.

Clustering is process to partition data into different clusters or classes. Objects within one cluster share a great similarity, while objects between different clusters have a great dissimilarity. Clustering is a statistical analysis technique that divides objects into relatively homogeneous groups (clusters). The clustering analysis process is shown in Figure 2. Data correlation is the basis of clustering process in this paper.

The initialization process of associated set partition algorithm is described as each node is initiated as a single associated set \(G_i\) in which \(V_i\) is node \(i\) and \(N\) is the number of sensor nodes). The correlation characteristics of two associated sets are defined as the minimum correlation characteristics between a random node in \(G_i\) and a random node in \(G_j\). During the process of initialization, using bivariate correlation represents the correlation characteristics of the initiated associated set. The clustering process of associated set partition algorithm is described as follows: select two sets (for example: \(G_m\) and \(G_n\)) with the biggest correlation...
Input: $N$ is the number of sensor nodes, $K$ is the expected number of clusters, $data[N][2]$ is the coordinate data of $N$ sensor nodes.

Output: $G[i]$: Associated set obtained from the associated set partition algorithm.

```
Input: $N$ is the number of sensor nodes, $K$ is the expected number of clusters, data $[N][2]$ is the coordinate data of $N$ sensor nodes.

Output: $G[i]$: Associated set obtained from the associated set partition algorithm.

(1) /* Initialization: initialize associated set and initial correlation between these sets */
(2) FOR $i$ IN [1:1:N]
(3) /* Initialize associated set $G[i]$: nodes set contained by associated set marked with $i$ */
(4) InitGroup($V_i, G[i]$); //Initialize $G[i]$ with single $V_i$
(5) FOR $j$ IN [1:1:i-1]
(6) /* Initialize the correlation $cor[i][j]$ between $G[i]$ and $G[j]$ by using (3) */
(7) InitCor($cor[i][j]$); /* During initialization, correlation between $G[i]$ and $G[j]$ is the correlation between node $V_i$ and $V_j$ */
(8) }
(9) }
(10) /* Clustering process: combine the associated sets until its number is no larger than $K$ */
(11) WHILE ($Group > K$) //Group is number of associated set
(12) {
(13) FindMaxCor($G[m], G[n]$); /* select two set $G[m]$ and $G[n]$ with the biggest correlation among all the associated sets */
(14) Merge($G[m], G[n], G[m]$); /* combine this two sets into one set $G_m$ */
(15) Delete($G[n]$); //delete $G[n]$
(16) /* update the correlation between newly generated $G[m]$ and the other set. $cor[k][m]$ is then smaller one between $cor[m][k]$ and $cor[n][k]$ */
(17) FOR $k$ IN [1:1:N]
(18) IF (Exist($G[k]$) && ($k != m$))
(19) UpdateCor($cor[k][m], min(cor[m][k], cor[n][k])$);
(20) }

Algorithm 1: Associated set partition algorithm.
```

5. Simulation Results

In order to verify the effectiveness and performance of associated clustering strategy, simulation was taken in experimental environment with a monitor area $4 \times 6$ m. 50 sensor nodes were deployed randomly in the area, with location known. For the soil moisture application, sprayers in the greenhouse are distributed with irrigation radius $R = 1$, distance $D = 1.85$. The distribution of nodes, sprayers, and crops in the monitoring area is shown in Figure 3. Take the lower left point of the area as coordinate origin. The crop 1 is planted in the area within $x: [0, 2]$, the coefficient of this area is $a = 0.07, b = 0.001$; The crop 2 is planted in the area within $x: [2.5, 4]$, the coefficient of this area is $a = 0.4, b = 0.002$; The area between crop 1 and crop 2 is open, its coefficient is $a = 0.2, b = 0.01$.

5.1. Effectiveness of Soil Moisture Correlation Characteristics.

An ideal data correlation characteristics model should be able to approximately describe the real data correlation. Model with such character can effectively illustrate the spatiotemporal correlation of soil moisture. Figure 4 is verification of the soil moisture correlation characteristics model.

In Figure 4, take the correlations between node 1 and node 1, 2, ..., 20 as an example and compare the real data correlation and the correlation calculated by correlation model to verify the effectiveness of the soil moisture correlation characteristics model. Each node holds a serial of 5 values as its collected data. The real data correlation is calculated by the corcoef function in Matlab. In Figure 4, calculated
correlation represents the correlation calculated by correlation model, and the real correlation represents the real data
correlation. As the result shown in Figure 4, spatiotemporal
correlation characteristics model in (7) is verified to be
effective to approximate the real data correlation. So, the
model can be used to describe the correlation characteristics
of soil moisture.

5.2. Associated Clustering. According to Algorithm 1 and the
soil moisture correlation characteristics calculated by (7), Use
the randomly generated coordination data of sensor nodes
data[N] [2] and expected cluster number K as input data,
executing associated clustering on the experimental environ-
ment in Figure 3. The result of clustering is in Figure 5.
As known from (7), the main factor of data soil moisture
correlation characteristics contains soil and vegetation coeffi-
cient a and b and irrigation coverage degree C. According
to the experimental environment settings, coefficients a, b, C
can be got from the coordination data. The expected cluster
number K helps to determine the grain of cluster. Therefore,
the result of clustering is associated with node location and
expectation cluster number.

5.3. Comparison with Random Clustering. In order to verify
the associated clustering strategy based on spatiotemporal
data correlation can make nodes within each cluster share a
greater data correlation and select the random clustering pro-
tocol: LEACH clustering algorithm as comparison. Figure 6
is the result of LEACH clustering. In LEACH, cluster head
is randomly generated and the other nodes decide to join a
cluster according to the distance between it and the cluster
head. As a result, LEACH clustering is more likely to cluster
the nodes closer.

In order to compare the data correlation within a cluster,
use multivariate linear relationship $\rho_X = \frac{(n - r + 1 - |R|)}{(n - 1)}$ in [9] to measure the correlation degree
of nodes data within a cluster, in which $\rho_X$ defined the linear
correlation coefficient of n-dimension random variables $X = (x_1, x_2, \ldots, x_n)$. R is the correlation coefficient matrix of X
and $r$ is the rank of $R$. $|R|$ is the largest nonzero subtype.

Take each node data as random variables $X_i$, and then,
we can calculate the multivariate linear relation $\rho_X$ of each
cluster, which is the cluster correlation. For each cluster
generated by associated clustering strategy and LEACH
6. Conclusions and Further Work

In this paper, the clustering problem of WSN is studied. In the application of soil moisture measurement, we established a dynamic soil moisture model, and after correlation coefficient analysis of the model, proposed a soil moisture correlation characteristics model, which is used to represent the correlation of sensor data, also used as the basis of clustering. Finally, an associated clustering strategy based on spatiotemporal correlation characteristics is proposed. As a result, the associated clustering strategy divides nodes with high correlation into a cluster, makes the cluster head do the data fusion more efficiently. Thus the energy of each node will be saved and the lifetime of the whole sensor network will be extended. This associated clustering strategy is also suitable for other industrial or agriculture applications when using a particular industrial model to replace the soil moisture model.

In practical scenarios, the distance between nodes affects the energy consumption in WSN. During the process of data transmission, the longer is the distance, the more energy is consumed. So, in the further works, we can take both associated clustering strategy and distance factor into consideration, in order to optimize clustering strategy of WSN.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No. 61170245 and the National Natural Science Foundation of China under Grant No. 61103242.

References

[1] V. Rajaravivarman, Y. Yang, and T. Yang, “An overview of wireless sensor network and applications,” in Proceedings of the 35th Southeastern Symposium on System Theory, pp. 432–436, March 2003.
[2] X. Yang, "Design and implementation of intelligent urban irrigation system: ZigBee modules for sensor networking," in Proceedings of the IEEE 2nd International Conference on Software Engineering and Service Science (ICSESS '11), pp. 461–464, July 2011.
[3] Q. Bian, Y. Zhang, and Y. Zhao, "Research on clustering routing algorithms in wireless sensor networks," in Proceedings of the International Conference on Intelligent Computation Technology and Automation (ICICTA '10), pp. 1101–1103, May 2010.
[4] A. Masoum, N. Meratnia, and P. J. M. Havinga, "Analysis of the impact of data correlation on adaptive sampling in wireless sensor networks," in Proceedings of the 9th International Conference on Networked Sensing Systems (INSS '12), pp. 1–2, June 2012.
[5] S. Li, "Geostatistical space-time modeling for temperature estimation," in Proceedings of the 1st International Conference on Agro-Geoinformatics (Agro-Geoinformatics '12), pp. 1–5, August 2012.
[6] S. Kondo, A. Kanzaki, T. Hara, and S. Nishio, “Energy-efficient data gathering using sleep scheduling and spatial correlation based on data distribution in wireless sensor networks,” in Proceedings of the International Conference on Network-Based Information Systems (NBiS ’11), pp. 194–201, September 2011.

[7] M.-H. Li, C.-C. Lin, C.-C. Chuang, and R.-I. Chang, “Error-bounded data compression using data, temporal and spatial correlations in wireless sensor networks,” in Proceedings of the 2nd International Conference on Multimedia Information Networking and Security (MINES ’10), pp. 111–115, November 2010.

[8] S.-Z. Zhang and S.-M. Wang, “The correlation coefficient matrix and multi linear correlation analysis,” University Mathematics, vol. 27, no. 1, pp. 195–198, 2011.

[9] J. F. Kun and L. Z. Chun, “Research on linear correlation of multi-dimensional random variables,” University Mathematics, vol. 24, no. 3, pp. 144–147, 2008.

[10] H. Jiang, S. Jin, and C. Wang, “Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor networks,” IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 6, pp. 1064–1071, 2011.

[11] Z. Chen, S. Yang, L. Li, and Z. Xie, “A clustering approximation mechanism based on data spatial correlation in wireless sensor networks,” in Proceedings of the 9th Annual Wireless Telecommunications Symposium (WTS ’10), pp. 1–7, April 2010.

[12] M.-H. Yeo, M. S. Lee, S.-J. Lee, and J.-S. Yoo, “Data correlation-based clustering in sensor networks,” in Proceedings of the International Symposium on Computer Science and its Applications (CSA ’08), pp. 332–337, October 2008.

[13] S. Yang, Y. Wang, and K. Sun, “Soil moisture content sensors placement based on the vertical variety law,” Transactions of the Chinese Society of Agricultural Machinery, vol. 39, no. 5, pp. 104–107, 2008.

[14] I. Rodriguez-Iturbe, V. Isham, D. R. Cox, S. Manfreda, and A. Porporato, “Space-time modeling of soil moisture: stochastic rainfall forcing with heterogeneous vegetation,” Water Resources Research, vol. 42, no. 6, Article ID W06D05, 2006.

[15] V. Isham, D. R. Cox, I. Rodriguez-Iturbe, A. Porporato, and S. Manfreda, “Representation of space-time variability of soil moisture,” Proceedings of the Royal Society A, vol. 461, no. 2064, pp. 4035–4055, 2005.

[16] L. H. Jun and G. S. Hong, “Mathematical modeling of soil moisture infiltration under sprinkler irrigation,” Journal of Irrigation and Drainage, vol. 25, no. 2, pp. 15–19, 2006.

[17] P.-Y. Zhu, J.-C. Song, and G.-S. Zhang, “Calculus model in the design of sprinkler irrigation system,” Journal of Southeast University for Nationalities (Natural Science Edition), vol. 31, no. 4, pp. 495–497, 2005.
Submit your manuscripts at http://www.hindawi.com