ISSN: 2278-0181
ICESMART-2015 Conference Proceedings

Degree Clustering Method and Data Density Correlation for Data Aggregation in WSN

Mrs. Bindu Madavi P.
Dept. of CSE,
T. John Institute of Technology
Bangalore,India

Roopa H. L.
MTech IVSem (CNE),
T.John Institute of Technology
Bangalore,India

***Abstract*— Wireless sensor network is a group of specialized transducers with a Communication infrastructure for monitoring and recording conditions at diverse locations. Sending local representative data to the sink node based on the spatial-correlation of sampled data is called as the data aggregation. The sensor nodes monitor a geographical area and collect sensory information.. To conserve energy this information is aggregated at intermediate sensor nodes by applying a suitable aggregation function on the received data. Aggregation reduces the amount of network traffic which helps to reduce energy consumption on sensor nodes. It however complicates the already existing security challenges for wireless sensor networks In our paper, we point out the problem that the recent spatial correlation models of sensor nodes data are inadequate for measuring the correlation in a critical environment. In addition, the data representative is not consistent when compared with actual data. Hence we propose the data density correlation degree, which is much needed to solve this problem. The proposed method correlation degree is a spatial correlation measurement that measures the correlation between a sensor node's data and its neighboring sensor nodes' data. Based on this correlation degree, a data density correlation degree clustering method is presented in detail so that the representative data have a low distortion on their correlated data in a WSN. We are alsoperforming the simulation experiments with two actual data sets to evaluate the performance of the DDCD clustering method. Results shows that representative data achieved using the proposed method have a lower data distortion than those achieved using the Pearson correlation coefficient based clustering method.**

***Index Terms*—WSN, Data aggregation, data density, in correlation degree, methods of clustering.**

I. INTRODUCTION

A Wireless sensor network (WSN) refers to a group of spatially dispersed and dedicated sensors for monitoring and recording the physical conditions of the environment and organizing the collected data at a central location. WSNs measure environmental conditions like temperature, sound, pollution levels, humidity, wind speed and direction, pressure [1], [2]. The importance of wireless sensor networks arises from their capability for detailed monitoring in remote and inaccessible locations where it is not feasible to install conventional wired infrastructure and also in detecting and accurately evaluating the events in the monitored area with the collected data. For this the sensor nodes are used.. However, this will cause the overlapping of sensor nodes' sensing areas and the spatial redundancy of adjacent sensor nodes' data [3], [4]. If every sensor node conveys collected data to the sink node, the sensor nodes will consume much more energy. To reduce the amount of transmitted data in a WSN, a great number of Correlation-based data aggregation methods have been studied in the literature [5]–[11].

According to the level of sampled data in data aggregation Strategy, data aggregation methods are grouped into three classes:

- Data level aggregation,
- Feature level aggregation and
- Decision level aggregation [12].

Also, based on the aggregation strategy, we can divide the data level aggregation methods into three types:

- In-network query type [5], [13],
- Data compression type [6], [14] and
- Representative type [7], [9], [15], [16].

The first step makes a delay. The second type is of limited usefulness as it is too complex. The third type is sensitive to the correlation measurement of sensor nodes.

The main intention of the representative type is selecting a representative sensor node locally and sending its observation to the destination node. Hence, the relative error between a representative data and its correlated data is a significant index for evaluating the represented performance.

Some researchers have systemically discussed spatial-correlation models based on geographic locations of sensor nodes or statistic features of sensor nodes' data [7]–[9], [17], [18]. The assumption of spatial-correlation models based on sensor nodes' locations is that the close sensor nodes are more correlated than the distant ones. Thereby, the spatial correlation degree function is modeled to be nonnegative and decrease monotonically with the distance between sensor nodes.

Even though he sensor nodes are usually deployed in some harsh environment, with the sensing distortion of sensor nodes, noise between sensor nodes, located terrain of sensor nodes and communication condition uncertain in practice. The neighboring sensor nodes may be uncorrelated. Additionally, the spatial location of sensor node is not accurate in general, making it hard to accurately model the spatial correlation of sensor nodes based on the locations of sensor nodes.

The Pearson Correlation Coefficient (PCC) as represented mathematically below was used to measure the correlation of sensor nodes' data, and it is a kind of spatial correlation model based on sensor nodes' data [19].

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

Although this correlation coefficient could reflect the linear correlation between two sensor nodes' data well, much data needs to be sent to the sink node and it only describes the linear dependence. In other spatial correlation models based on sensor nodes' data [8], [18], statistical features are introduced according to the application of a WSN. However, much rude data should be sent to the sink node, and these models have high computational complexity

In this work, a data density correlation degree (DDCD) was proposed to measure the spatial correlation of sampled data and try to resolve the drawbacks in existing spatial correlation models. With the DDCD clustering method, sensor nodes which are in the same cluster have a high correlation degree, while those belonging to different clusters have a low correlation degree. Furthermore, the time complexity of the DDCD clustering algorithm is $O(n)$. The message complexity is $O(Kn)$. Where the K is the maximum degree of the sensor network topology graph.

The remaining topics of this paper are organized as follows. Section II discusses related work on spatial correlation models in WSN. Section III presents the data density correlation degree to measure the spatial correlation of sensor nodes' data, as well as introduces the DDCD clustering method in detail. In section IV, the accuracy of the representative sensor node in DDCD clustering method is validated by comparing the performance of DDCD clustering method, α-local spatial (α-LS) clustering method [18] and PCC based clustering method. Moreover, the energy consuming of these clustering methods is discussed. Section V presents the conclusion of this study and our future work.

II. RELATED WORK

The existing system on modeling spatial correlation, the spatial correlation models are mainly based on the locations of sensor nodes or statistical features of sensor nodes' data. The spatial correlation model in [7] simulated the transmitting process of data from data source to the sink node. The spatial correlation between two sensor nodes is depicted by a function of the spatial distance between them. Four types of spatial correlation functions are given. To capture the spatial-temporal characteristics of point and field sources in WSN, the spatial-temporal correlation models for point and field sources are theoretically analyzed in [20]. Meanwhile, the spatial-temporal characteristics of point and field sources were analytically derived along with the distortion functions. The correlation degree between two sensor nodes was obtained by the overlapping degree of their sensing areas [21], [22]. This model is very convenient. However, it is difficult to pinpoint the locations of sensor nodes, and the sensor nodes' sensing areas change with their remaining energy. Thus, this type of spatial correlation degree model is not accurate and impractical.

In a real-time environment, the area covered by a WSN is categorized into some irregular parts. The sensor nodes in the same part have a high correlation, while those have low correlation. Along the boundary of two adjacent parts, two close sensor nodes that are in two different parts do not correlate. This practical situation is ignored in [7] and [20]– [22]. In order to solve the drawbacks of spatial correlation models based on the spatial distance between sensor nodes, the correlation of sensor nodes in the data domain was modeled in [8]. Unfortunately, with the model proposed in [8], if two sensor nodes' data are the same at two different time intervals, the correlation degrees in these two time inter-vals will differ. The result doesn't agree with the reality. The definition of the spatial correlated weight considers the average spatial distance deviation between each sensor node's sampled data and that sampled by its neighbors within a predefined communication radius [18].

In order to accurately detect the damage occurs gradually, a semantic clustering model based on fuzzy system was proposed to find out the semantic neighborhood relationship in [24]. At the network starts up process, a physical clustering is done to form a hierarchical physical organization consisted of two levels. The upper level encompasses CHs and the lower level consists of sensor nodes which are subordinated to one of the CHs. When a sensor node's data satisfies a domain rule related to the event monitored by the WSN, this sensor node is called the "candidate". If the data of the "candidate" changes, the "candidate" becomes a semantic neighbor. Then, the CHs utilize the data of all the semantic neighbors which are in the same cluster or in the neighboring clusters to obtain an aggregated data by the fuzzy inference system as described in [24]. The semantic neighbors are correlated to the domain rules of the monitored event, so that this semantic clustering method is suitable for event detection.

III . CL U STE RI N G ME T H O D OVERVIEW

A. Data Density Correlation Degree

In a WSN, if a certain number of neighboring sensor nodes' data are close to a sensor node's data, this sensor node can represent its neighbors in the data domain. This representative sensor node is called the core sensor node

Definition 1: **Core sensor node**. Assume sensor node v has n neighboring sensor nodes. They are respectively v_1, v_2, ..., v_n. The data object of v is D. Its neighboring sensor nodes' data objects are respectively D_1, D_2, ..., D_n. If there are N data objects in D_1, D_2, ..., D_n whose distances to D are less than ε and $minPts \leq N \leq n$ then the sensor node v is called the core sensor node. Where *minPts* is the amount threshold, ε is the data threshold.

Definition 2: **Data density correlation degree**. Let sensor node v has n neighboring sensor nodes which are within the cycle of the communication radius of v. They are v_1, v_2, ...,v_n, respectively. The data object of v is D, and its neighboring sensor nodes' data respectively D_1, D_2, ..., D_n. Among these n data objects, there are N data objects whose distances to D are less than ε, and $minPts \leq N \leq n$. Then the data density correlation degree of sensor node v to the sensor nodes whose data objects are in ε-neighborhood of D

B. Data Density Correlation Degree Clustering Method

In cluster-based networks, to select the representative sensor nodes, we proposed the data density correlation degree (DDCD) clustering method, which will be presented in detail in this section. The WSN is modeled by undirected graph $G = (V, E)$. Where V is the sensor node set consisting of all sensor nodes in the WSN, E is the edge set consisting of all links in the WSN. The antenna of sensor node i ($i \in V$) is an Omni-directional antenna, with a communication radius of $\alpha(i)$. Let $N(i)$ be the set of sensor nodes within the circle of the communication radius of i. In cluster-based data aggregation networks, the data transmission process is that every cluster head sends aggregated data obtained from its member nodes to the sink node by one hop or multi-hops . The DDCD clustering algorithm includes three procedures: the Sensor Type Calculation (STC) procedure, the Local Cluster Construction (LCC) procedure and Global Representative sensor node Selection (GRS) procedure

The following pseudo code is the Sensor Type Calculation algorithm applied to each sensor node Procedure of STC

Input: Data Threshold – ε ; Amount Threshold – *minPts;*
Weights – a_1, a_2, a_3;
Neighboring Sensor nodes set of Sensor node i – $N(i)$.
Output:
Sensor Type – Core sensor nodes or Non-core sensor nodes;
Two Sets of sensor nodes' ID stored in each sensor node
–$NodeSet_{inner}(i)$ and $NodeSet_{outer}(i)$;
Data Density Correlation Degree–$Sim(i)$
(Step 1)
for each $i: i \in V$ pardo { /* Parallel process for each i */
sensor node i is a non-core sensor node, NearNum=0;
$NodeSet(i)_{inner} = \phi$, $NodeSet(i)_{outer} = \phi$
for each j: $j \in N(i)$ {
if ($\|d(i) - d(j)\| \leq \varepsilon$) NearNum −−;
}/*end for*/
if (NearNum $\geq$ *minPts*) {
Sensor node i is a Core sensor node;

$$Sim(i) = a_1(1 - \frac{1}{exp(N - minPts)}) - a_2(1 - \frac{d_\Delta}{\varepsilon}) - a_3(1 - \frac{d}{\varepsilon})$$

else
$Sim(i)$=0;
} /* end if */
}/*end for*/

(Step 2)
for each i: $i \in V$ pardo { /* Parallel process for each i */
if (i is a Core sensor node) {
for each j: $j \in N(i)$ {
if ($\|d(i) - d(j)\| \leq \varepsilon$)
$NodeSet(i)_{inner} = \{j\} \cup NodeSet(i)_{inner}$;
else $NodeSet(i)_{outer} = \{j\} \cup NodeSet(i)_{outer}$;

```
Input:
Sensor Type – Core sensor node or Non-core sensor node;
Two sets of sensor nodes' IDs in each sensor node
        –NodeSet_inner(i) and NodeSet_outer(i);
Data Density Correlation Degree –Sim(i)
Output: ClusterSet= {ClusterSet(i) | i ∈ V} ,
      DDCD_Set = {DDCD_max(i) | i ∈ V}
/*Send Information*/
 (Step 1)
for each i: i ∈ V  pardo { /* Parallel process for each i */
   if (sensor node i is a core sensor node)  {
      for each j:  j ∈ NodeSet(i)_inner  {
sensor node i sends the packet –(i, 1, Sim(i)) to sensor node j
      } /* end for */
      for each j:  j ∈ NodeSet(i)_outer  {
         sensor node i sends the packet—(i, -1, Sim(i)) to sensor
node j
      } /* end for */
   } /* end if */
   if (sensor node i is a non-core sensor node) {
      for each j:  j ∈ N(i) {
         sensor node i sends the packet—(i,0, Sim(i)) to sensor
node j;
      } /* end for */
   } /* end if */
}/*end for*/

/*Receive Information*/
/*Receive_Packet. ID is the first byte of sent packets in Step 1*/
/*Receive_Packet. Relationship is the second byte of sent
packets in Step 1*/
/*Receive_Packet. Sim(i) is the third byte of sent packets in
Step 1*/
(Step 2)
```

Procedure of LCC

With the local clusters achieved by the LCC procedure, we can obtain the global clusters using the GRS procedure. The pseudo code for GRS algorithm is as follows

```
Input:  ClusterSet= {ClusterSet(i) | i ∈ V}
      DDCD_Set = {DDCD_max(i) | i ∈ V}
/* DDCD_max(i) = (ID(i), Sim_max(i)) */
Output: Representative Sensor Nodes
 (Step 1)
(1.1)  for each i: i ∈ V pardo { /* Parallel process for each i */
          if  DDCD_max(i).ID ≠ i {
       for each j:  j ∈ ClusterSet(i) {
        sensor node i sends  DDCD_max(i)  to sensor node j;
            } /* end for */
          } /* end if */
        } /* end for */
(1.2)  IterativeFlag=0;
   for each i: i ∈ V pardo { /* Parallel process for each i */
     MaxSim=max( {DDCD_max(j).Sim_max(j) | j = ClusterSet(i)} )
     MaxID= {j | DDCD_max(j).Sim_max(j) = MaxSim}
       if  DDCD_max(i).ID ≠ MaxID  {
        DDCD_max(i) = (MaxID, MaxSim) ;
            IterationFlag=1;
       }end /*end if*/
   } /* end for*/
(1.3) if IterationFlag==1
          goto (1.2);
   else stop.
```

IV. PE R FOR MA NC E ANA LYS I S

In this section, we selected the PCC based clustering method and the α-LS clustering method to evaluate the DDCD clustering method by comparing their clustering performance. Before performance comparison, we introduced the global average relative error as the performance index. Due to there are several parameters should be confirmed before the DDCD clustering method is performed, the way how to set these parameters is presented as well. In the end, an analysis of energy consuming in clustering process for these three clustering methods was given.

A. Clustering Performance Index

Obviously, if a sensor node's data could represent its correlated sensor nodes' data well, the relative error between representative data and the correlated data should be small. Therefore, we can use the average relative error to measure the concentration of data within a cluster.

Definition 3: **Average relative error within a cluster**. Consider $m+1$ sensor nodes $v_0, v_1, v_2, \ldots,$ v_m which are divided into a cluster. Their data are D_0 , D_1 , D_2 ,. . ., D_m , respectively. D_0 is the representative data. Then the average relative error of D_0 within the cluster is:

$$\bar{E} = \frac{\sum_{i=1}^{m} e_i}{m} \qquad (2)$$

Where $e_i = |D_0 - D_i|$. e_i is the relative error between D_0 and D_i.

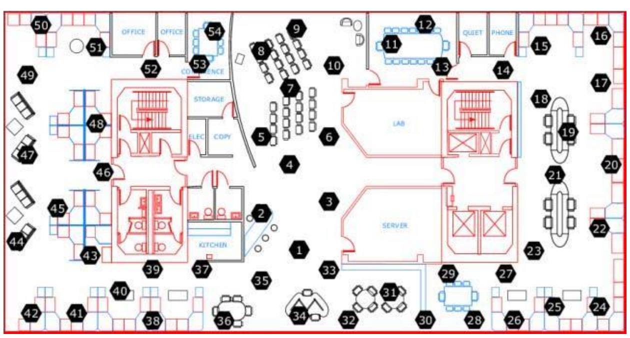
We noticed that if every representative sensor node is a good representation of its cluster members in a WSN, all the average relative errors within the clusters will be small. Thus, the global average relative error could be used to measure the performance of the clustering method. It is shown as Eq.3.

$$\bar{E}_g = \frac{\sum_{i=1}^{k} \bar{E}}{k} \qquad (3)$$

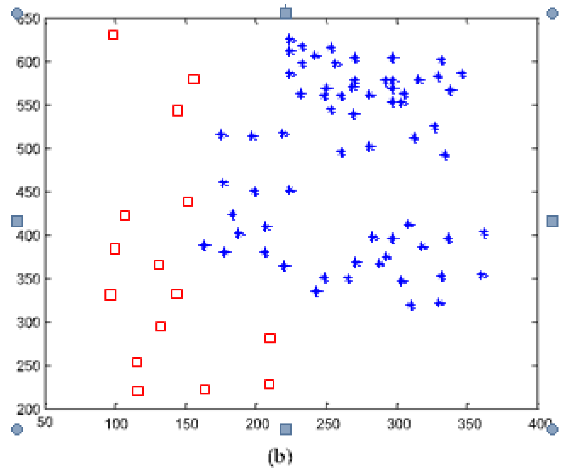
B. Analysis of Parameters in DDCD Clustering Method

In DDCD clustering method, each sensor node obtains data from its neighboring sensor nodes which are within the circle of its communication radius firstly. The communication radius of sensor nodes concerns the number of its neighboring sensor nodes. With the distributions of sensor nodes in the Intel Berkeley Research Lab and LUCE, we will illustrate how we obtain the communication radius for DDCD clustering method. The deployments of sensor nodes are shown in Fig. 2.

In Fig. 2(a), the sensor nodes are almost dispersed uniformly in the Intel Berkeley Research Lab. We are able to get a minimum spatial distance for each sensor node. The minimum spatial distance is the distance between a sensor node and its nearest sensor node. And among all the minimum spatial distances, the maximum value is 5.66 meters. For the connectivity of the network, the communication radius of sensor node is at least 5.66 meters. Thus, in DDCD clustering method with Intel lab data, the communication radius of sensor node is set to 6 meters in our experiments so that the number of neighboring sensor nodes is 4 or 5 for most of sensor nodes. Fig. 2(b) shows the dispersion of sensor nodes in LUCE. In this figure, 15 red squares represent the sensor nodes whose minimum spatial distances are larger than 30 meters, and 65 blue asterisks are that ones whose minimum spatial distances are less than or equal to 30 meters. In our experiments, we just applied clustering method to the blue asterisks sensor nodes in Fig. 2(b) because data aggregation clustering method is designed for the WSN where sensor nodes are densely deployed. Other sensor nodes which are in the sparse area could adjust their communication radii according to their minimum spatial distances. With the analysis on these two cases, we can find that the communication radius is dependent on the deployment of sensor nodes in densely covered area.



(a)



(b)

Fig 2: sensor node distribution in experiment in lab (a) in LUCE (b)

The number of neighboring sensor nodes is 4 or 5 for most of sensor nodes. Fig. 2(b) shows the dispersion of sensor nodes in LUCE. In this figure, 15 red squares represent the sensor nodes whose minimum spatial distances are larger than 30 meters, and 65 blue asterisks are that ones whose minimum spatial distances are less than or equal to 30 meters. In our experiments, we just applied clustering method to the blue asterisks sensor nodes in Fig. 2(b) because data aggregation clustering method is designed for the WSN where sensor nodes are densely deployed. Other sensor nodes which are in the sparse area could adjust their communication radii according to their minimum spatial distances. With the analysis on these two cases, we can find that the communication radius is dependent on the deployment of sensor nodes in densely covered area. In our experiments, the communication radius is equal to or a little larger than the maximum value of minimum spatial distances. For the sensor nodes in the Intel Berkeley Research Lab, the communication radius is set to 6 meters. For those sensor nodes deployed in densely covered area in LUCE, the communication radius is 30 meters. With these communication radii, the number of neighboring sensor nodes is 4 or 5 for most of sensor nodes.

The amount threshold *minPts* is the least amount for a sensor node which is able to represent some neighboring sensor nodes. It means that if a sensor node is able to represent some sensor nodes, there should be at least *minPts* sensor nodes' data in the ε-neighborhood of its data. If we increase the value of *minPts*, the numbers of RSN and ISN will increase in DDCD clustering method and the global average relative error will decrease, and vice versa. Thus, we can adjust the value of *minPts* according to users' requirement on global average relative error, numbers of RSN and ISN. In our experiments, the value of *minPts* is set to 2 because the number of neighboring sensor nodes is just 4 or 5 for most of sensor nodes.

C. Clustering Performance Comparison

Experiment With the Research Lab Data

A set of sensor network data has been collected in the Intel Berkeley Research Lab. 54 sensor nodes measuring temperature and humidity were deployed in the lab and continuously worked for 35 days. In our experiment, we randomly chose one day's data from the collected data, with the temperature averages per two minutes regarded as the sample data. Thereby, every sensor node has 720 samples

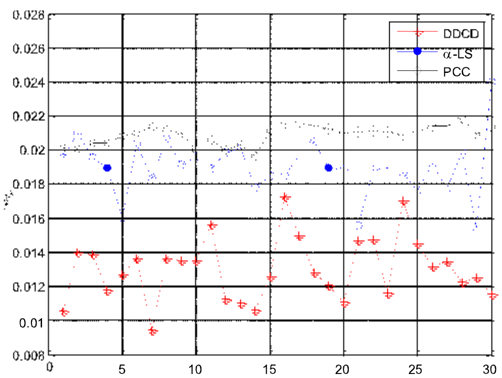


Fig 3: Global average error comparison for diff clustering *D. Clustering Performance Comparison*

Experiment with LUCE Data

In July 2006, the LUCE was carried out on the EPFL campus. This experiment aimed to better understanding micrometeorology and atmospheric transport in the urban environment. In order to cover the heterogeneous areas, 94 sensor nodes are densely deployed. Therefore, the sampled data are temporal and spatially correlated. We chose the data collected on January 1st 2007, and regarded the temperature aver-ages per two minutes as the sample data. At a sample time,

E. Rationality Comparison Experiment With LUCE Data

In practice, when representative sensor nodes and isolated sensor nodes are selected, the sink node will just receive sampled data from these sensor nodes in a time interval. And in this selected time interval, every sensor node's sensed data changes slightly. In order to evaluate the rationality of the DDCD clustering method, we chose three different start time labels. At a start time label, the RSN and ISN are obtained with the DDCD clustering method. And with the collected data of RSN and ISN, we could achieve the global average relative errors at the chosen start time label and the following 19 time labels. Likewise, the α-LS clustering method are performed. In the PCC based clustering method, we obtained the PCC values of neighboring sensor nodes with the data within the first 10 time labels. Therefore, the sensor nodes should transmit 10 rounds sample data to the sink node at the first 10 time labels in PCC based clustering method.

The values of parameters in each clustering method are the same with that in Section IV.D. The start time labels were chosen according to the trends of sample data in the selected time intervals, because the clustering methods mentioned in our experiments are suitable for the sampled data changing gradually. Three different start time labels are selected and corresponding clustering performance results were obtained as shown in Fig. 6. Meanwhile, we achieved the numbers of RSN and ISN for the referred clustering methods in our experiments at different start time labels, as shown in Table I.

From Fig. 6, we can see that the global average relative errors for the DDCD clustering method are always the least among that for the three clustering methods in different time intervals. It means the DDCD clustering method has better accuracy performance in data representation.

In Table I, the numbers of RSN and ISN at the first time label are the least for the DDCD clustering method. At the 71st time label and the 91st time label, each number of ISN for the PCC based clustering method is the least at its time label. And we can easily get that the total number of RSN and ISN for the DDCD clustering method is just a little larger than that for the PCC based clustering method. According

TABLE I
NUMBERS OF RSN AND ISN FOR DIFFERENT CLUSTERING METHODS

Start time label	Clustering method	Number of RSN	Number of ISN
1	α-LS	19	31
	PCC	9	45
	DDCD	6	22
71	α-LS	19	32
	PCC	7	15
	DDCD	6	20
91	α-LS	19	31
	PCC	6	18
	DDCD	6	20

Thus, the energy expenditure is almost the same as that in the DDCD clustering method, while the numbers of RSN and ISN are larger than that of DDCD clustering method according to the results in section IV.C, D and E. More energy is consumed when RSN and ISN send sampled data to the sink node in α-LS clustering method. For the PCC based clustering method, every sensor node has to send several rounds of sampled data to the sink node using an energy-efficient route firstly. Then the sink node transmits the clustering result to each sensor node. The energy consumption is huge in this process.

Therefore, the DDCD clustering method is more energy efficient than the other two clustering methods.

V. CONCLUSION

Our proposed method in this paper are the introduction of the data density correlation degree and the data density cor-relation degree (DDCD) clustering method. more accurate aggregated data can be obtained in cluster- based data aggregation network with the DDCD clustering method, the sensor nodes that have high correlation are divided into the same cluster, allowing produced by the DDCD clustering method. Also, the amount of data conveyed to the sink node can decrease.

The conducted evaluation experiments highlight the clus-tering performance of the DDCD clustering method using two real temperature datasets. The comparative results reveal that the data of RSN can provide more accurate description on the real environmental when compared with the α-LS clustering method and the PCC based clustering method. Meanwhile, the energy consumption in the construction process of clusters was analyzed for these three cluster-ing methods mentioned in our experiments. In summary, the DDCD clustering method is more energy efficient and could obtain better data representation performance than the other two clustering methods. Thus, DDCD clustering method is useful for the application where the sensor nodes are densely deployed and the sampled data change slowly with time.

REFERENCES

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Comput. Netw.*, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] L. M. Oliveira and J. J. Rodrigues, "Wireless sensor networks: A survey on environmental monitoring," *J. Commun.*, vol. 6, no. 2, pp. 1796–2021, 2011.

[3] C. Zhu, C. Zheng, L. Shu, and G. Han, "A survey on coverage and connectivity issues in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 35, no. 2, pp. 619–632, 2012.

[4] G. Fan and S. Jin, "Coverage problem in wireless sensor network: A survey," *J. Netw.*, vol. 5, no. 9, pp. 1033–1040, 2010.

[5] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: A Tiny AGgregation service for ad-hoc sensor networks," *ACM SIGOPS Operating Syst. Rev.*, vol. 36, no. 1, pp. 131–146, 2002.

[6] J. Zheng, P. Wang, and C. Li, "Distributed data aggregation using Slepian-Wolf coding in cluster-based wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2564–2574, Jun. 2010.

[7] M. C. Vuran, Ö. B. Akan, and I. F. Akyildiz, "Spatio-temporal corre-lation: Theory and applications for wireless sensor networks," *Comput. Netw.*, vol. 45, no. 3, pp. 245–259, 2004.

[8] J. Yuan and H. Chen, "The optimized clustering technique based on spatial-correlation in wireless sensor networks," in *Proc. IEEE Youth Conf. Inf., Comput. Telecommun. YC-ICT*, Sep. 2009, pp. 411– 414.

[9] A. Rajeswari and P. Kalaivaani, "Energy efficient routing protocol for wireless sensor networks using spatial correlation based medium access control protocol compared with IEEE 802.11," in *Proc. Int. Conf. PACC*, Jul. 2011, pp. 1–6.

[10] J. N. Al-Karaki, R. Ul-Mustafa, and A. E. Kamal, "Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms," *Comput. Netw.*, vol. 53, no. 7, pp. 945–960, 2009.

[11] C. Hua and T.-S. Yum, "Optimal routing and data aggregation for maximizing lifetime of wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 892–903, Aug. 2008.

[12] S. Iyengar, K. Chakrabarty, and H. Qi, "Introduction to special issue on 'distributed sensor networks for real-time systems with adaptive configuration'," *J. Franklin Inst.*, vol. 338, pp. 651–653, Jan. 2001.

[13] S. Madden, R. Szewczyk, M. J. Franklin, and D. Culler, "Supporting aggregate queries over ad-hoc wireless sensor networks," in *Proc. Mobile 4th IEEE Workshop Comput. Syst. Appl.*, Oct. 2002, pp. 49–58.

[14] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network corre-lated data gathering," in *Proc. IEEE Comput. Commun. Soc. 23rd Annu. Joint Conf. INFOCOM*, Mar. 2004, pp. 2571–2582.

[15] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, pp. 316–329, Apr. 2006.

[16] G. A. Shah and M. Bozyigit, "Exploiting energy-aware spatial correla-tion in wireless sensor networks," in *Proc. 2nd Int. Conf. Commun. Syst. Softw. MiddleWare, COMSWARE*, Jan. 2007, pp. 1–6.

[17] W. Guo, L. Zhai, L. Guo, and J. Shi, *Worm Propagation Control Based on Spatial Correlation in Wireless Sensor Network*. Berlin, Germany: Springer-Verlag, 2012, pp. 68–77.

[18] Y. Ma, Y. Guo, X. Tian, and M. Ghanem, "Distributed clustering- based aggregation algorithm for spatial correlated sensor networks," *IEEE Sensors J.*, vol. 11, no. 3, pp. 641–648, Mar. 2011.

[19] C. Carvalho, D. G. Gomes, N. Agoulmine, and J. N. de Souza, "Improv-ing prediction accuracy for WSN data reduction by applying multivariate Spatio-temporal correlation," *Sensors*, vol. 11, no. 11, pp. 10010–10037, 2011.

[20] M. C. Vuran and O. B. Akan, "Spatio-temporal characteristics of point and field sources in wireless sensor networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2006, pp. 234–239.

[21] N. Li, Y. Liu, F. Wu, and B. Tang, "WSN data distortion analysis and correlation model based on spatial locations," *J. Netw.*, vol. 5, pp. 1442–1449, Dec. 2010.

[22] R. K. Shakya, Y. N. Singh, and N. K. Verma, "A novel spatial correlation model for wireless sensor network applications," in *Proc. 9th Int. Conf. WOCN*, Dec. 2012, pp. 1–6.

[23] F. Bouhafs, M. Merabti, and H. Mokhtar, "A semantic clustering routing protocol for wireless sensor networks," in *Proc. 3rd IEEE Consum. Commun. Netw. Conf.*, Jan. 2006, pp. 351–355.

[24] A. R. Rocha, L. Pirmez, F. C. Delicato, É. Lemos, I. Santos, D. G. Gomes, *et al.*, "WSNs clustering based on semantic neighborhood relationships," *Comput. Netw.*, vol. 56, no. 5, pp. 1627–1645, 2012.

[25] Institute for Nuclear Theory, Seattle, WA, USA. (2004). *Intel Lab Data* [Online]. Available: http://db.csail.mit.edu/labdata/labdata.html

[26] École polytechnique fédérale de Lausanne,Lausanne, witzerland. (2006). *LUCE* [Online]. Available: http: //lcav.epfl.ch/cms/lang/en/pid/86035