Study on Entropy Correlation in Wireless Sensor Networks for Energy Efficiency

Nguyen Thi Thanh Nga*[*]*, Nguyen Kim Khanh, Ngo Hong Son
Hanoi University of Science and Technology, No. 1, Dai Co Viet, Hai Ba Trung, Hanoi, Viet Nam
Received: January 25, 2018; Accepted: May 25, 2018

Abstract

Correlation characteristic can bring many significant potential advantages for the development of efficient communication protocols for wireless sensor networks. To exploit the correlation in WSNs, it is necessary to build the correlation model. However, most of the present correlation models only consider the linear and distance dependence correlation or computation complexity. This paper presents a novel entropy correlation model with less computation complexity that could be applied practically. Moreover, two energy efficient aggregation schemes including on-off scheme which offers an efficient way to choose representative nodes in a cluster with permitted distortion and compression scheme which reduces in-network message length suitable to high correlation data are also presented in this paper using the proposed correlation models.

Keywords: Entropy correlation coefficient, Correlation model, Compression, Representative node, Distortion

1. Introduction

In recent years, the advanced development in micro-electro-mechanical systems (MEMS) and the wireless communications have enabled the wide deployment of wireless sensor networks (WSN) which expand sensing capabilities in space and time that can satisfy requests from various modern applications. Because of low-cost, small in size, and no-replace battery powered characteristics of sensor nodes, energy conservation is commonly recognized as the key challenge in designing and operating the network.

In typical WSNs applications, sensors are required spatially dense deployment in order to achieve satisfactory coverage [1]. As a consequence, multiple sensors will record information about a single event in the sensing field, i.e. these sensed data are correlated with each other. The existence of correlation characteristic can bring many significant potential advantages for the development of efficient communication protocols well-suited for the WSNs paradigm [2, 3].

To exploit the correlation in WSNs, it is necessary to build the correlation model. There have been many research efforts to study correlation model in WSNs. In [3], correlated nodes are supposed to observe the same source, and the observed data is the sum of a correlated version of the source and observed noise. The correlation model is distance's dependence and could be classified into four groups including Spherical, Power exponential, Rational quadratic and Matérn. In [4], the correlation model is proposed such as correlation coefficient is a function of distance among nodes. Other research papers consider the correlation as the similarity of sensed data [5]. Some papers define the correlation model in different ways such as linear predictive model [6], node weight [7], data density correlation degree [8].

All the above models consider only the linear correlation between data and distance based. To solve more general correlation relation, entropy-based correlation models are considered [9- 12]. In [9], the joint entropy of a group of nodes are calculated using real data set and then a distance based joint entropy function is built by approximation to the calculated joint entropy. Distance-based joint entropy models are proposed in [9, 10]. In [11], instead of calculating directly from real data, entropy correlation coefficient is chosen to be Pearson linear correlation coefficient to reduce the computation complexity but reduce the generality of using entropy. In [12], joint entropy is calculated from real data and then the joint entropy of a node set is approximated by an exponential function of a number of nodes in the set. The advantage of this model is a distance-independent model, but the disadvantage is the complexity in determination correlation among nodes. Joint entropy values of all possible node groups have to be calculated in order to select correlated nodes.

To overcome these above difficulties, the concept of correlation ratio, similarly to entropy correlation coefficient, is used in [13], but the correlation model is

[*] Corresponding author: Tel.: (+84) 904.567.424
Email: ngantt@soict.hust.edu.vn

not yet established. In addition, the mathematical explanation of choosing correlation region has not been done. The research in [14] continues the work in [13] by presents a model of evaluated joint entropy bases on single nodes and entropy correlation coefficient of a pair of nodes. Theoretical and practical validation has been done in this research, however, applications of the correlation model have not been shown. The application of the correlation model is initially presented in [15]. The correlation characteristic is used to group correlation nodes into clusters and compression aggregation is done, in order to save energy. This paper extends the work in [15] by developing and evaluating of two aggregation types for WSNs including data compression and representative types using the entropy correlation model.

2. Joint entropy evaluation

2.1. Entropy concept

To measure the correlation among sets of data, we first consider the concept of entropy and mutual information.

If a random variable X takes on values in a set $X = \{x_1, x_2, \ldots x_n\}$, and is defined by a probability distribution $P(X)$, then the entropy $H(X)$ of the random variable X is written as:

$$H(X) = -\sum_{x \in X} P(x)\log_2 P(x) \quad (1)$$

Joint entropy is the entropy of a joint probability distribution or a multi-valued random variable. If X and Y are jointly distributed according to $P(x, y)$, the joint entropy $H(X,Y)$ is:

$$H(X,Y) = \sum_{x \in X}\sum_{y \in Y} P(x,y)\log_2 P(x,y) \quad (2)$$

Mutual information is a quantity that measures a relationship between two random variables which are sampled simultaneously. The formal definition of the mutual information $I(X,Y)$ of two random variables X and Y, whose joint distribution is defined by $P(X,Y)$ is given by:

$$I(X,Y) = -\sum_{x \in X}\sum_{y \in Y} P(x,y)\log_2 \frac{P(x,y)}{P(x)P(y)} \quad (3)$$

The relation between mutual information and entropy is given by:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \quad (4)$$

The normalized measures of mutual information called entropy correlation coefficient [16] that is given as follows:

$$\rho(X,Y) = 2\frac{I(X,Y)}{H(X)+H(Y)} = 2 - 2\frac{H(X,Y)}{H(X)+H(Y)} \quad (5)$$

The entropy correlation coefficient ρ varies from 0 to 1, depending on the correlation between two nodes. The larger the value of ρ, the higher the correlation is.

2.2. Joint entropy approximation

Supposing that there is a set of N data $\{X_1, X_2, \ldots, X_N\}$ with the entropy of each data, $H(X_i)$, and entropy correlation coefficient, $\rho_{ij} = \rho(X_i, X_j)$, with any $1 \le i \ne j \le N$ satisfies the following conditions:

$$H_{min} \le H(X_i) \le H_{max} \quad (6)$$

$$\rho_{min} \le \rho_{ij} \le \rho_{max} \quad (7)$$

The joint entropy is estimated based on the idea of hierarchical clustering [17] as follows:

a. Joint entropy upper bound

With a group that has only one node, the entropy of one node is limited by equation (6):

$$H_1 = H(X_i) \le k_1 H_{max} \quad (8)$$

where $k_1 = 1$.

With a group of two nodes X_i and X_j, from the definition of entropy correlation coefficient in equation (5) we have:

$$H_2 = H(X_i, X_j) = \frac{2 - \rho(X_i, X_j)}{2}\left(H(X_i) + H(X_j)\right)$$

In addition,

$H(X_i), H(X_j) \le H_{max}$, and $\rho(X_i, X_j) = \rho_{ij} \ge \rho_{min}$

Then

$$H_2 \le \frac{2-\rho_{min}}{2}(2H_{max}) = (2 - \rho_{min})H_{max}$$

$$\text{or } H_2 \le k_2 H_{max} = bH_{max} \quad (9)$$

where $k_2 = b = 2 - \rho_{min}$.

The coefficient k_2 can also be rewritten as follows:

$$k_2 = \frac{b}{2}.2 = \frac{b}{2}(k_1 + 1)$$

With a group of three nodes X_i, X_j and X_k, at first, two nodes X_i and X_j are merged to create a new cluster represented by node X_{ij} with $H(X_{ij}) = H(X_i, X_j) \le k_2 H_{max}$. According to hierarchical clustering [17, 10], the correlation coefficient between one cluster and another cluster can be obtained by the greatest/shortest/average correlation coefficient from any member of one cluster to any member of the other cluster. Therefore:

$$\rho(X_{ij}, X_k) = \min\{\rho(X_i, X_k), \rho(X_j, X_k)\} \ge \rho_{min}$$

Then

$$H_3 = H(X_i, X_j, X_k) = H(X_{ij}, X_k)$$
$$= \frac{2-\rho(X_{ij}, X_k)}{2}\left(H(X_{ij}) + H(X_k)\right)$$
$$\leq \frac{2-\rho_{min}}{2}(k_2 H_{max} + H_{max})$$
$$= \frac{b}{2}(k_2 + 1)H_{max} = k_3 H_{max} \quad (10)$$

where $k_3 = \frac{b}{2}(k_2 + 1)$

Similarly, joint entropy H_m of a group with m nodes could be considered to be the joint entropy of a sub-cluster with *(m-1)* nodes and the remaining node. The entropy of the sub-cluster is joint entropy of *(m-1)* nodes and the entropy correlation coefficient between the sub-cluster and the main node is the greatest/shortest/average correlation coefficient from any member of the sub-cluster to the remaining node. Thus:

$$H_m \leq \frac{2-\rho_{min}}{2}(k_{m-1} H_{max} + H_{max})$$
$$= \frac{b}{2}(k_{m-1} + 1)H_{max}$$
$$= k_m H_{max} \quad (11)$$

where $k_m = \frac{b}{2}(k_{m-1} + 1)$

From recurrence relation of k_m, the general formula to calculate k_m can be obtained as follows ($m>2$):

$$k_m = 2\left(\frac{b}{2}\right)^{m-1} + \left(\frac{b}{2}\right)^{m-2} + \cdots + \left(\frac{b}{2}\right)^2 + \frac{b}{2} \quad (12)$$

Or in the more compact way (in case $b \neq 2$):

$$k_m = \frac{\left(\frac{b}{2}\right)^m - 1}{\frac{b}{2} - 1} + \left(\frac{b}{2}\right)^{m-1} - 1 \quad (13)$$

b. Joint entropy lower bound

Lower bound of the joint entropy of a group with m node could be determined in a similar way to the upper bound. The results are as follows:

With a group that has only one node, we have:

$$H_1 = H(X_i) \geq l_1 H_{min} \quad (14)$$

where $l_1 = 1$.

With a group of m nodes ($m \geq 2$)

$$H_m \geq l_m H_{min} \quad (15)$$

where $l_m = \frac{c}{2}(l_{m-1} + 1)$

From the recurrent relation of l_m the general formula to calculate l_m can be obtained as follows ($m>2$):

$$l_m = 2\left(\frac{c}{2}\right)^{m-1} + \left(\frac{c}{2}\right)^{m-2} + \cdots + \left(\frac{c}{2}\right)^2 + \frac{c}{2} \quad (16)$$

Or in the more compact way (in case $c \neq 2$):

$$l_m = \frac{\left(\frac{c}{2}\right)^m - 1}{\frac{c}{2} - 1} + \left(\frac{c}{2}\right)^{m-1} - 1 \quad (17)$$

3. Correlation region definition and correlation clustering algorithm

3.1. Correlation Region Definition

As mentioned in [4], sensor nodes in a correlation region can record information of a single event in the sensor field, i.e. these sensed data have a correlation with each other. Because the sensed data is taken from the same event, the number of bits to represent sensed data should be not so different, i.e. the entropy of sensed data is similar. On the other hand, the entropy correlation coefficient of all pairs in this region is also similar. Therefore, we can define a correlation region as follows:

Definition 1*:* A correlation region is a region where the sensed data of all nodes have similar entropy value and entropy correlation coefficients between all pairs of nodes are similar.

- $H_0 = H(X_1) = H(X_2) = \cdots = H(X_m)$
- $\rho_0 = \rho_{ij} = \rho(X_i, X_j),\ \forall i \neq j$

In practice, it is difficult to obtain the similarity between two entropies or entropy correlation coefficient. Then, the correlation region could be defined by a more practical way as bellows.

Definition 2*:* A group of m nodes $\{X_1, X_2, \ldots, X_m\}$ is in a correlation region if:

- $H_0 \leq H(X_1), H(X_2), \ldots, H(X_m) \leq H_0 + \Delta H$
- $\rho_0 \leq \rho_{ij} = \rho(X_i, X_j),\ \forall i \neq j$

where ΔH is small enough.

H_0 is called "base entropy" and ρ_0 is called "correlation level" of the region. The higher the correlation level, the more correlation of the region is.

With this definition, it is seen that the upper and lower bounds of joint entropy are quite similar and therefore we can estimate the joint entropy of the m nodes $\{X_1, X_2, \ldots, X_m\}$ by the following equation:

$$H(X_1, X_2, \ldots, X_m) = k_m H_0 \quad (18)$$

where k_m is calculated by using equation (12) or (13) with $b = 2-\rho_0$. This joint entropy formula is called correlation model and will be used in the next sections.

According to [12], nodes in a correlation group will share much information among them. Therefore, the joint entropy will not increase much when the number of nodes in the group increases. In other words, the joint entropy will go to "saturation" state when the number of nodes increases. The faster the approaching saturation state, the more correlation among nodes is. And as shown in Fig.1, the proposed joint entropy calculation (18) completely satisfies the above property. This validates the proposed correlation model.

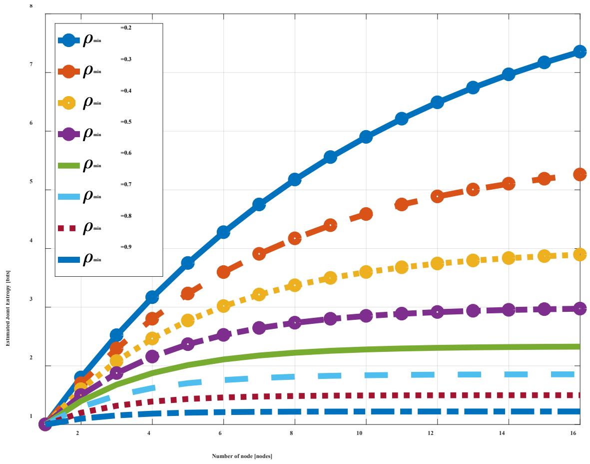

Fig. 1. Estimated joint entropy with different values of entropy correlation coefficients using upper bound function (with H_{max} = *1[bits]*).

3.2. Correlation Clustering Algorithm

Using the definition of correlation region, a sensor field could be divided into correlation regions with specified base entropy and correlation level. The clustering process is described in Fig.2.

In the step (*) of the algorithm, the base entropy, and correlation level is chosen such that they can cover all possible values of entropy and entropy correlation coefficient in the network. The value of entropy correlation coefficient should be chosen from high to low.

In the step (**) of the algorithm, if there is more than one node satisfy the condition $0 < C(X_i) = \max\{C(X_j),\ X_j \in G\}$, the node that has maximum entropy value will be removed.

In comparison to entropy-based clustering in [12], the proposed algorithm is simpler with less computation. In the proposed method, the only entropy of single node and joint entropy of node pairs are calculated. While in [12], joint entropy values of all possible node groups have to be calculated.

4. Entropy Correlation with Data Aggregation

According to the aggregation strategy, data level aggregation methods are divided into three types: in-network query type, data compression type, and representative type [8]. Correlation is appropriated with data compression type and representative type. This section considers the applications of the proposed entropy correlation model to data aggregation in WSNs, including compression aggregation and representative aggregation.

```
BEGIN
REPEAT
    Choose H0, ρ0, ΔH; (*)
    Initialize new group G = ∅;
    FOR each node Xi in the network and not
    belong to any group
        IF  H0 ≤ H(Xi) ≤ H0 + ΔH
            Add Xi into G
        ENDIF
    ENDFOR
    REPEAT
        FOR each node Xi in G
            Calculate C(Xi)= number of node Xj
    that ρ(Xi,Xj) < ρ0
        ENDFOR
        FOR each node in G
            IF  0 < C(Xi)  = max{C(Xj), Xj∈G}
             Remove Xi from G (**)
            ENDIF
        ENDFOR
    UNTIL max{C(Xj), Xj∈G}=0
UNTIL all nodes are grouped
END
```

Fig. 2. Correlation-based clustering algorithm.

4.1. Compression Aggregation

There are many compression techniques, which could be applied in wireless sensor networks [2]. In this paper, three qualitatively different routing schemes including Distributed Source Coding (DSC) [18], Routing Driven Compression (RDC) [19], and Compression Driven Routing (CDR) [20] will be evaluated to choose the most appropriate lossless compression approach. These schemes are simplified models of schemes which have been proposed literately in [9].

To compare and evaluate different routing-plus-compression, we focus on energy expenditure. In WSNs, energy expenditure is mainly from data transmission that is proportional to the amount of transfer data. Therefore, entropy that quantifies the amount information could be considered as the representation of energy expenditure.

Let's consider the arrangement of sensor nodes in a grid, where only $(2n - 1)$ nodes are in the first column are sources. There are n_1 hops on the shortest path between the sources and the sink. The paths taken by data and the intermediate aggregation of three considered schemes can be seen in [9].

In DSC, the sensor nodes have knowledge about their correlations, and they can compress data to avoid transmitting redundant information. In this case, ideally, each source can send exactly the right amount of uncorrelated data to the sink along the shortest possible path without the need of intermediate compression. The energy expenditure (E) for this scheme is calculated by:

$$E_{DSC} = n_1 H(X_1) + n_1 H(X_2|X_1) + \cdots + n_1 H(X_{2n-1}|X_1, X_2, \ldots X_{2n-2})$$

$$E_{DSC} = n_1 H_{2n-1} \quad (19)$$

In RDC) the sensor nodes do not have any knowledge about their correlations and send data along the shortest paths to the sink while allowing for opportunistic compression wherever the paths overlap. The energy expenditure (E) for this scheme in this considered scenario can be derived as:

$$E_{RDC} = (n_1 - n)H_{2n-1} + 2H_1 \sum_{i=1}^{n-1} i + \sum_{j=1}^{n-1} H_{2j-1} \quad (20)$$

In CDR, nodes have no knowledge of the correlations, but the data is compressed close to the sources and initially routed so as to allow for maximum possible compression at each hop. This leads to the collection of data removed of all redundancy at a central source from where it is sent to the sink along the shortest possible path. The energy expenditure (E) for this scheme in this considered scenario could be derived as:

$$E_{CDR} = n_1 H_{2n-1} + 2\sum_{i=1}^{n} H_i \quad (21)$$

Using the estimated joint entropy model (18) for the above expressions, we can quantify the performance of each scheme. Figure 3 shows the energy expenditure for the DSC, RDC and CDR schemes as a function of the entropy correlation coefficient (with $n = n_1 = 50$, $H_0 = 1[bit]$).

It can be found that DSC scheme has lowest energy expenditure because the number of transmitted bits is smallest among lossless compression schemes. The higher the correlation level, the smaller the energy usage is. For RDC scheme, the correlation level does not affect much of the energy usage of the scheme. For CDR scheme, the energy usage is high with small correlation level, but it reduces very fast when the correlation level increase, and approaches DSC scheme in high correlation level area.

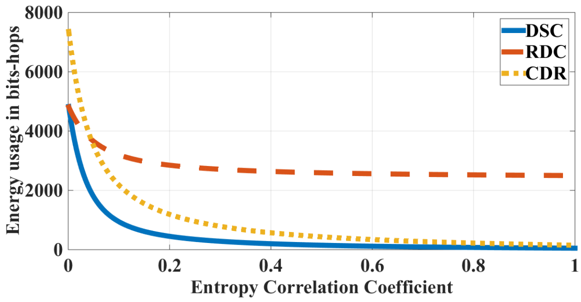

Fig. 3. Energy expenditures for the DSC, RDC and CDR schemes with respect to entropy correlation. **coefficient**

From this result, DSC and CDR are appropriate for wireless sensor network with high correlation characteristics of the environment. However, DSC scheme is still difficult to implement in practice while CDR could be implemented easily by local compression. Therefore, CDR is highly recommended to be the compression approach for wireless sensor network with high correlation characteristics.

4.2. Representation Aggregation

In a correlation region with high enough correlation level, it may not be necessary for every sensor node in a correlation group to transmit its data to the sink; instead, a smaller number of sensor measurements might be adequate to communicate the event features to the sink within a certain reliability/fidelity level. These working sensors are called representative nodes of the region/group. In this case, to evaluate the reliability/fidelity level, the distortion function is used.

a. Distortion function

This research uses entropy correlation concept to evaluate the correlation, therefore the predefined entropy distortion function proposed in [10] is used. The distortion function is defined as followings: Supposing that there are a total number of N sensor nodes in the considered area, and denote their observed data as $\{X_1, X_2 \ldots X_N\}$. The joint entropy of all these N sensors, $H(X_1, X_2 \ldots X_N)$, is the maximum amount of information that can be gained for the area of interest. If a subset of these sensors denoted as $\{X_{i1}, X_{i2} \ldots X_{iM}\}$, are selected to report their observed data to the sink, the information gained at the sink is $H(X_{i1}, X_{i2}, \ldots, X_{iM})$. The distortion function is defined as the ratio of the decrease in the amount of information to the maximum amount of information, given by:

$$D = \frac{H(X_1, X_2, \ldots, X_N) - H(X_{i1}, X_{i2}, \ldots, X_{iM})}{H(X_1, X_2, \ldots, X_N)} \quad (22)$$

The value of D satisfies $0 \le D \le 1$. It can be interpreted as the percentage of information loss due to network resource constraints.

Using the estimated joint entropy equation (18), the distortion can be calculated as:

$$D = 1 - \frac{k_{iM}}{k_N} \qquad (23)$$

It is realized that the distortion is depending on the entropy correlation coefficient of the group, the number of representative nodes and the total number of nodes in the group. Therefore, with the desired distortion, the number of representative nodes can be determined for a correlation group with known correlation coefficient and the total number of nodes in the group.

b. Number of representative nodes

The number of representative node in a correlation group is determined based on entropy correlation coefficient and the total number of nodes in the group. Since the representative based aggregation is more effective with high correlation region, in this paper, the region with high correlation level ($\rho_0 \geq 0.5$) is considered. Because of high correlation level, the joint entropy will go to a saturation value when the number of nodes increases. For that reason, with the same value of distortion, the number representative nodes do not depend on the number of node in the area if this number is high enough. In this case ($\rho_0 \geq 0.5$), the estimated joint entropy goes to saturation state when the number of nodes in group reaches 14 nodes as shown in Figure 1. For that reason, we only need to consider the total number $N \leq 20$.

The relation between the distortion and the number of representative nodes with different values of the entropy correlation coefficients are shown in Figure 4 (in case N=15).

It can be found that the distortion becomes smaller when the entropy correlation coefficient is higher, and the number of representatives is higher. Now we choose a value of distortion, for example, $D = 0.1$, to determine the number of representative nodes with different values of a total number of nodes in the group. Table 1, shows the number of representative nodes in cases of $D = 0.1$.

It is found that to obtain the same distortion, the number of representative nodes are not so different with a different number of total nodes of the correlation group. It only depends on the correlation coefficient. Thus, the number of representative nodes can be determined easily from correlation level and desired distortion using the above conclusions.

Table 1. Number of representative nodes with distortion D = 0.1

ρ_0	0.5	0.6	0.7	0.8	0.9
N=10	7	6	5	4	2
N=15	8	6	5	4	2
N=20	8	6	5	4	2

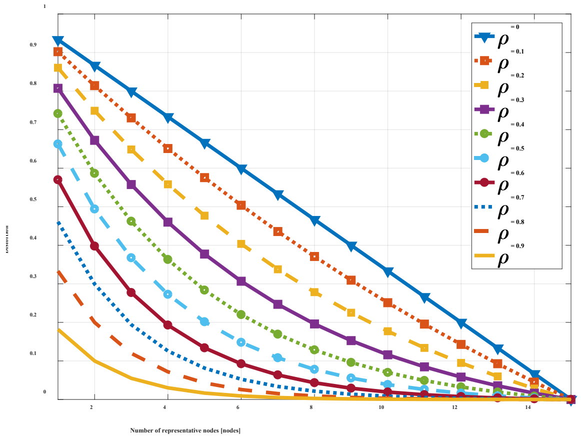

Fig. 4. The relation between distortion and the number of representative nodes in case of N = 15 nodes

c. Selection of the representative nodes

```
R = RepSelection ({X1, X2, …, XN}, iM)
BEGIN
        C = {X1, X2, …, XN}; //corr. node set
        Initialize new group R = ∅;
    FOR Xi, Xj ∈ C
            Calculate ρ(Xi,Xj);
        Calculate ρ̄(Xi) = (1/N) Σ_{j=1}^{N} ρ(Xi,Xj);
    ENDFOR
    FOR k = 1 to iM
        Find Xi = arg min_{Xi∈C} (ρ̄(Xi));
            Add Xi to R;
            Remove Xi from C;
    ENDFOR
        Return R;
END
```

Fig. 5. Maximizing obtained information based representative node selection algorithm

After knowing the number of representative nodes, it is necessary to select these nodes in the group. The purpose of selection is to maximize the obtained information from representative nodes, i.e. the representative nodes are least correlated with all other nodes in the group. This can be done by calculating the average entropy correlation coefficient of one node with all other nodes in the nodes and choose *iM* nodes with least values of average to be representative nodes where *iM* is a number of representative nodes. The selection algorithm of representative nodes is shown in Figure. 5.

The above selection purpose is to maximize the total information. The selection can also be based on different purposes such as maximizing coverage (total covered areas by representative nodes is maximum) or energy balancing (the nodes with most values of energy are chosen to be presentative nodes).

5. Conclusions and further study

The paper has introduced a practical model to estimate the joint entropy of a group of data using only entropy of each data and entropy correlation coefficient of each pair of data. This model is used to define a correlation region and a clustering algorithm is proposed.

After clustering the network into correlation region, the data compression and representative aggregation strategies are deployed to take the advantage of correlation characteristic. Using the estimated joint entropy model, some routings with compression schemes have been evaluated and Compression Driven Routing scheme is the most appropriate scheme for Data Compression Aggregation in Wireless Sensor Network with correlation characteristics.

Moreover, the estimated joint entropy model is also used to establish the distortion function for representative aggregation. Then the number and selection of representative nodes are also presented.

In the future, a complete routing protocol with data compression and representative aggregation strategies will be developed and implemented to validate the results of this paper in practice.

Acknowledgments

This research is funded by the Hanoi University of Science and Technology (HUST) under project number **T2017-PC-081**

References

[1] I. F. Akyildiz, et. al. Wireless Sensor Networks: A Survey. *Computer Networks (Elsevier) Journal*, vol. 38, no. 4, (March 2002), 393-422.

[2] Srisooksai, et al. Practical data compression in wireless sensor networks: A survey. *Journal of Network and Computer Applications* 35.1 (2012), 37-59.

[3] M. C. Vuran, et. al. “Spatio-temporal correlation: Theory and applications for wireless sensor networks,” *Comput. Netw.*, vol. 45, no. 3, pp. 245–259, 2004.

[4] Shakya, Rajeev K., et. al. "Generic correlation model for wireless sensor network applications." *IET Wireless Sensor Systems* 3.4 (2013): 266-276.

[5] Liu, Chong, et. al. "An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation" *IEEE Trans. on parallel and distributed systems* 18.7 (2007).

[6] Gupta, Himanshu, et al. "Efficient gathering of correlated data in sensor networks." *ACM Tran. on Sensor Networks (TOSN)*4.1 (2008)

[7] Ma, Yajie, et al. "Distributed clustering-based aggregation algorithm for spatial correlated sensor networks." *IEEE Sensors Journal* 11.3 (2011): 641-648.

[8] Yuan, Fei, et. al. "Data density correlation degree clustering method for data aggregation in WSN." *IEEE Sensors Journal* 14.4 (2014): 1089-1098.

[9] Pattem, et. al. "The impact of spatial correlation on routing with compression in wireless sensor networks." *ACM Trans. on Sensor Networks (TOSN)* 4.4 (2008): 24.

[10] Dai, Rui, and et. al. "A spatial correlation model for visual information in wireless multimedia sensor networks." *IEEE Trans. on Multi.* 11.6 (2009): 1148-1159.

[11] Wang, Fan, et al. "Energy-efficient clustering using correlation and random update based on data change rate for wireless sensor networks." *IEEE Sensors Journal* (2016): 1-1.

[12] D. Maeda, et.al. Efficient Clustering Scheme Considering Non-uniform Correlation Distribution for Ubiquitous Sensor Networks. *IEICE Trans. on Fund. of Electronics, Comm. and Computer Sciences* E90-A (7), (2007), 1344-1352.

[13] N. T. T. Nga, et. al. Correlation-based Clustering for Energy Saving in Wireless Sensor Network. *Journal of Science and Technology Technical Universities, No. 115 (2016), pp. 51-57*

[14] N. T. T. Nga, et. al. Entropy-based Correlation Clustering for Wireless Sensor Networks in Multi-Correlated Regional Environments. *IEIE Trans. on Smart Proc. and Comput., vol. 5, no. 2, (2016), 85-93.*

[15] N. T. T. Nga, et. al. Entropy correlation and its impact on routing with compression in wireless sensor network. *SoICT 2016: 235-242*

[16] Thomas M. Cover et. al. Elements of Information Theory. *Copyright@1991 John Wiley & Sons, Inc. Chapter 2 pp. 13-49.*

[17] A.K. Jain, et. al. *Data Clustering: A Review*, ACM Computing Surveys, Vol.31, No.3, (Sept. 1999).

[18] Pradhan, et. al. Distributed source coding using syndromes (DISCUS): Design and construction. *IEEE Trans. on Information Theory* 49.3 (2003), 626-643.

[19] Intanagonwiwat, et al. Impact of network density on data aggregation in wireless sensor networks. *Distributed Computing Systems, 2002. Proceedings*. 22nd International Conference on. IEEE.

[20] Anna Scaglione et. al. On the interdependence of routing and data compression in multi-hop sensor networks. *MobiCom '02.* ACM, New York, NY, USA.