http://www.eab.org.tr

Educational Research Association
The International Journal of
Educational Researchers 2017, 8(1): 19-25
ISSN: 1308-9501

http://ijer.eab.org.tr

WSN Using Collective Scheme of Energy Efficient Grouping and Evidence Retrieval

S. P. Santhoshkumar [1]
M. Praveenkumar[2]
S. Chandramohan[3]

Abstract

The two-layer network structure has been widely adopted in wireless sensor networks (WSNs) for managing sensor nodes. In such a structure, the low layer nodes communicate with their cluster head, followed by the cluster-head nodes communicating with the base station operating in either a one hop or a multi-hop manner. The main focus of node-clustering algorithms is minimizing energy consumption due to strictly limited resources in WSNs. Also, WSNs are data intensive networks with the capability of providing users with accurate data. Unfortunately, data missing is common in WSNs. In this paper, we propose a novel joint design of sensor nodes clustering and data recovery, where the WSNs is organized in a two-layer manner with our developed clustering algorithm, and then the missing data is recovered based on this two-layer structure. Furthermore, in the proposed clustering algorithm, we take both the energy-efficiency and data forecasting accuracy into consideration and investigate the tradeoff between them. This is based on the key observation that the high energy-efficiency of the network can be achieved by reducing the distances among the nodes in a cluster, while the accuracy of the forecasting results can be improved by increasing the correlation of the data stream among the nodes in a cluster. Simulation results demonstrate that our joint design outperforms the existing algorithms in terms of energy consumption and forecasting accuracy.

Keywords: Wireless sensor networks, energy efficiency, node clustering, data forecasting

[1]Assistant Professor, Dept. of CSE, Rathinam Technical Campus, Coimbatore, India
[2] Assistant Professor, Dept. of IT, Rathinam Technical Campus, Coimbatore, India
[3] Assistant Professor, Dept. of IT, Rathinam Technical Campus, Coimbatore, India

Introduction

The development of electronic and sensor technologies, wireless sensor network (WSN) becomes a popular network architecture for current and future wireless communications. Particularly in recent years, WSNs have been widely applied to various practical scenarios, such as intelligent transportation, health-care monitoring, industrial manufacture, robotics, and so on [1]. Furthermore, WSNs will be prevailing with the emergences of intelligent applications, e.g., Smart City [2], Wearable Computing Devices [3], Tactile Internet [4], etc. A major responsibility of the WSNs is accurately sensing and collecting useful data, for example, the measurements of air quality, humidity, biomedical and chemical information, and yielding sensed big data for further analysis [5]. At the same time, cloud-computing enabled technologies, e.g., Cloud-RAN [6] and Fog-RAN [7], provide the WSNs with the leverages of computation, communication and storage resources [8], as well as a promising method to manage, process and preserve the privacy of massive aggregated data [9].

A wireless sensor node consists of multiple modules, including battery, data process units, storage, transmitter/receiver pair, and one or several sensor devices. These sensor nodes collect the information about the surrounding environment and forward it to the base station through a one-hop or multi-hop manner. As such, WSNs serve as bridges between the physical world and human societies, resulting in a cyber-physical system [10]. However, due to limited resources, sensor nodes shall cooperate with each other to carry out complicated tasks [1–3]. For example, mobile crowd-sensing has proved to be an effective and efficient way to collect and process environmental data [4], as well as reconstruct the spatial field of a physical quantity (e.g., traffic condition) [5,6]. Apart from data transmissions, WSNs need to efficiently eliminate the redundant data, query the necessary data, fuse the correlated data and recover the missing data [7].

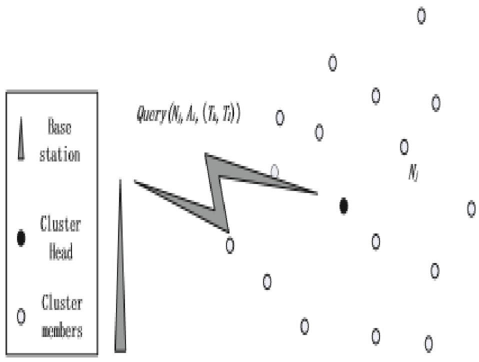

Fig. 1. A typical scenario of query the useful readings.

As shown in Fig. 1, the base station sends a query message to a specific cluster head to request the readings of a node. After receiving the query message, the cluster head communicates with its members to obtain the readings, and then transmits them to the base station. Each node’s readings may include several attributes, such as temperature, humidity, wind speed, and so on [10]. However, it is likely that the base station cannot successfully obtain the desired data due to the hostile environment of the communications. In this case, forecasting of missing data is needed. An intuitive method is collecting all correlated data to the base station and then forecasting the data in a centralized manner through forecasting algorithms. At the same time, intensive data transmissions and processing will cause a large amount of energy consumption at sensor nodes. As WSNs are usually battery-driven with limited power supply, battery lifetime is a vital factor for long-term operations of sensor nodes. Generally, there are two strategies to extend battery life. One is to charge the battery from other energy sources, such as energy harvesting and power transfer [2]. The other one is to develop protocols for efficiently managing energy consumptions, which has been widely studied as a hot topic in the academia society [9]. In particular, a hybrid two-layer structure has been proposed to deal with the energy efficiency issue, where the sensor nodes are divided into multiple clusters and each cluster is managed by its cluster head. This hybrid distributed method can achieve a good tradeoff between the fully centralized and distributed approaches. Furthermore, the hybrid approach consists of two modules, including a node clustering module and a missing data recovery module. In conventional wisdom, node-clustering algorithms only focus on decreasing energy consumption problem given that energy is very strictly limited in WSNs. Meanwhile, WSNs are data-centered networks aiming at providing users with accurate data. A well-developed clustering algorithm should take both energy-efficiency and recovery-accuracy into account simultaneously. In this paper, we propose a distributed data recovery scheme to address the above mentioned issue. The WSNs are assumed to be managed in a two-layer structure, i.e., the network is divided into multiple clusters and the cluster heads play a role as a bridge between the sensor nodes and the base station. To be specific, we first define both the spatial Euclidean distances between nodes [5] and the distance between the reading series generated by the nodes. Then we propose a clustering algorithm, where only the nodes having similar readings and small pairwise distances can be dispatched to the same cluster. Within each cluster, the forecasting process is conducted by each cluster head. We note that over-fitting problem imposes negative effects on the data forecasting accuracy. To address this problem, we further develop a lightweight missing data forecasting algorithm for the WSNs under the case of strictly limited resources. In the simulations, we compare the proposed scheme with the existing approaches in terms of energy consumption and forecasting accuracy. The proposed scheme is the first of its kind that integrates node clustering and missing data forecasting into a unified framework. Simulation results show that the proposed scheme performs much better in terms of forecasting accuracy compared with the existing approaches, e.g., MUlti-SequenCe LEast Squares (MUSCLES) forecasting algorithm [6].

The major contributions are summarized as follows.

1. We identify the significance of achieving high energy efficiency and forecasting accuracy to design massive connected sensor devices networks.
2. A unified framework with the novel two-layer approach is proposed to improve the forecast accuracy and energy efficiency simultaneously. This is achieved by the sensor nodes clustering phase, followed by the missing data forecasting phase.
3. We present a clustering algorithm based on the similarities in the Euclidean distances and aggregated data. In particular, a specific approach is proposed to address the outliers in the process of clustering.
4. A lightweight missing data forecasting algorithm is developed to address the over-fitting issue, thereby significantly improving the forecasting accuracy.
5. Simulation results will demonstrate the effectiveness of the proposed algorithms to design energy efficient and missing data forecasting accurate WSNs, compared with the state-of-the-art algorithms.

The remainder of the paper is organized as follows. In Section II, we summarize the related work in two aspects, i.e., node clustering algorithms and data forecasting techniques.

RELATED WORK

In this section, we review the related works in two aspects, i.e., clustering protocols and data forecasting algorithms.

Nodes Clustering Approaches

Nodes clustering problem in WSNs has been intensively investigated in the literatures and many classic approaches have been proposed. In particular, linked cluster algorithm (LCA) [10, 11] is one of the earliest clustering algorithms, which requires no central controller and is fully distributed. In LCA, the cluster heads form a backbone network and they connect with all the sensor nodes in its cluster directly. This structure is very flexible to implement a wide variety of routing strategies and can be used to avoid the problem of hidden terminals. The hierarchical control clustering algorithm proposed in [7] treats the network as a graph. A cluster is defined as a subset of vertices whose included graph is connected. Finally, a multi-tier hierarchical cluster structure is formed and it satisfies several constrains simultaneously. In the process of clustering WSNs, the authors in [8] argue that it was very unwise to ignore the geographical information of the sensor nodes, especially for a large WSN. They then propose a novel clustering algorithm, which used geographical radius of cluster instead of logical radius. Another classic clustering algorithm, namely, LEACH, is proposed in [9]. LEACH forms clusters based on the received signal strength and then the cluster heads serve as a bridge between the cluster members and the base station. Various applications of LEACH illustrate that it can always produce relatively good results in terms of energy efficiency and data transmission quality.

Data Forecasting Algorithms

Data forecasting techniques have been widely used in WSNs to reduce data transmission and improve the energy-efficiency [10, 11]. The authors in [10] propose an on-mote filtering approach relying on a local multi-step assessment of sensor data with forecasting and assessing value of information. Simulation results showed that the proposed approach reduces the number of data transmissions and the energy consumption significantly. In [3], the authors discuss the implementation of an Artificial Neural Network (ANN) algorithm in a low cost system-on-chip and develop an autonomous intelligent wireless sensor network. Additionally, there are several common missing data forecasting algorithms proposed in the field of time series mining. Specifically, the simplest approach to forecasting the missing data is Yesterday, in which, we replace the missing data with the nearest previous data. The major disadvantage of this method is that the forecasting error accumulates with the increasing of continuous missing data in which situation yesterday becomes malfunctioning. Auto-regression based approaches are also very popular and they have been used in various scenarios. They forecast the missing data of a time series by first mining the pattern underneath the time series and then using the pattern to forecast the missing data. Similar to Yesterday, the forecasting accuracy decreases significantly with the increasing of the missing data. Different from Yesterday and auto-regression based approaches, MUSCLES [6] makes full use of the high correlations between the co-evolving time series and construct a relation between them through linear mathematic tools. Simulation results illustrate that MUSCLES outperforms Yesterday and auto-regression based approaches in terms of forecasting accuracy. However, in MUSCLES, it is hard to define the correlations between the time series and the overfitting problem is ignored.

NODE CLUSTERING ALGORITHM BASED ON LOCATION AND DATA CORRELATIONS

In this section, we will present a novel node clustering algorithm by considering both the locations of all the nodes and data correlations between each pair of data streams generated by the nodes. We first assume that all the nodes in the network are located in a plain area and each node has a standard radio radius r that can be adaptively changed by the nodes in some exceptional cases. Each node in the network has the capability of severing as a cluster head and the nodes take turns to be a cluster head considering that the cluster head consumes much more energy compared with the other nodes. The clusters of all the nodes need to be reconstructed when the lasting time of a round exceeds a threshold or some cluster heads cannot serve as a cluster head anymore because of limited resources.

Cluster Heads Selection

Considering that most distributed approaches offer no guarantee about the number and distribution of the cluster heads, in this paper, we designed a centralized clustering algorithm to choose the cluster heads. In the initial of each round, all the nodes first transmit their IDs, location information and the residual energy to the base station. After receiving all the information about the nodes, the base station first selects the top half nodes that have more residual energy as the candidates of the cluster heads to balance the energy load among all the nodes and prolong the life time of the network. We assume that the communication energy scales exactly with squared distance and our goal is minimizing the amount of energy consumption forallthenon-cluster nodes to communicate with their nearest cluster head. Note that the non-cluster head nodes may not select the nearest cluster head node as its cluster head and the clusters forming process will be discussed in Section III-B.

Clusters Forming

We first introduce the definition of trend closeness among the readings of nodes, which is a common measurement of the correlations between the time series.

Definition 1 (Spatial Closeness between Sensor Nodes):

Under the assumption that all the nodes are located in a plain regime, the spatial distance of two nodes dist is defined as the Euclidean Distance. Sensor node np is spatial close to nq if np is at worst d far from nq. Parameter d is set by the users of the network and naturally not too much larger than the radio radius r, otherwise, the nodes in a cluster cannot communicate with each other very well. The spatial closeness between sensor nodes has important affection on the selection of cluster heads.

Definition 2 (Trend Closeness between Readings of Nodes):

Each node in the network generates readings about the surrounding environment which can be treated as a time series and in most cases the readings are strongly correlated between neighboring nodes. For each node, we treat its readings as a time series and the time series is infinite. In this paper, we define the trend closeness between two time series as the Dynamic Time Wrapping (DTW) which is more robust than the Euclidean Distance. In this paper, for convenience, we compute the trend closeness based on the latest ten readings of the nodes rather than all the historical readings of the nodes.

MISSING DATA FORECASTING

In this section, we will develop the missing data forecasting algorithms for the given clusters, calculated by the algorithms presented in the previous section. Intuitively, the base station can collect all the time series from the network and then recover the missing values based on conventional data forecasting approaches, such as MUSCLES [6]. However, this method is impractical considering that energy is strictly limited and the jamming problem in the network. Therefore, we will design a distributed framework to forecast the missing values based on the node clustering approach proposed previously.

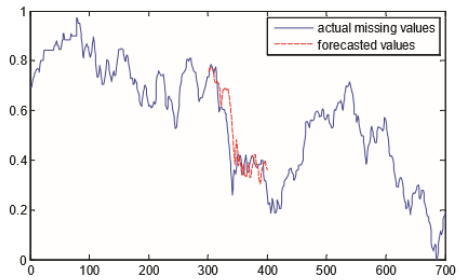

Fig. 2. The actual missing values and the forecasted

A. The Over-fitting Problem in Time Series Forecasting

In this section, we conduct a detailed experiment to present the over-fitting problem in time series forecasting field.

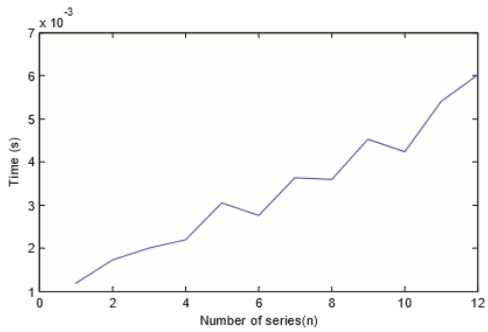

Fig. 3. Running time of forecasting process with different number of series n.

This is the first clear illustration that too many uncorrelated time series will decrease the forecasting accuracy significantly. Therefore, we need to select several similar time series to forecast the

Each time series is composed of 700 data points and we need to measure the distances between each pair of time series xi and xj quantitively by Euclidean distance, i.e.,

$$\text{Dist (xi,xj)} = \sum_{k=1}^{N}(\text{xi}(k) - \text{xj}(k))2)1/2 \quad (1)$$

As an example, the actual missing values and the forecasted results for time series i when n = 6 is presented in Fig. 3. We can find that the change trends between them are very similar which proves that the correlated time series can be used to forecast missing values. We employ mean absolute error (MAE) to measure the forecasting accuracy quantifiably and MAE is defined as

$$\text{MAE} = \sum_{t=1}^{N}(Ft - Ot/N) \quad (2)$$

where Ft is the forecast value for the t^{th} missing value, Ot is the real value of the t^{th} missing value, and N is the total number of all the missing values.

SIMULATION AND RESULT

In this section, we evaluate the performance of the proposed approach and compare it with MUSCLES [6], Yesterday and Auto-regression in terms of energy-efficiency and forecasting accuracy.

We first illustrate the setup of the simulation including the network structure, generation of the nodes' readings, several measurements of the simulation results and some other default parameters. Then, the simulation results under the setup are detailed presented and discussed

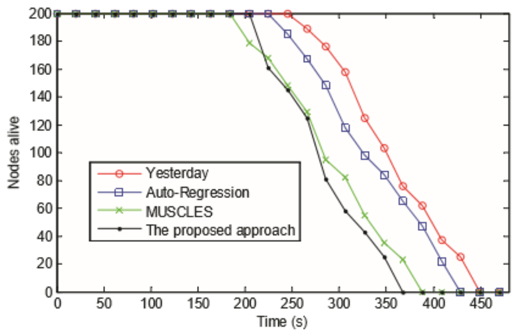

Fig 4. Number of nodes alive over time

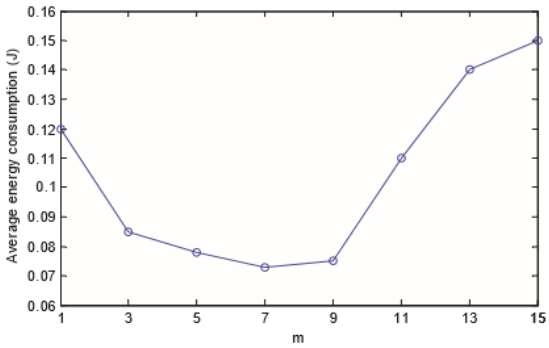

Fig 5. Average energy consumption with different m.

CONCLUSION AND FUTURE RESEARCH

In this paper, we propose a novel clustering algorithm for WSNs, which takes both the energy-efficiency problem and data correlations between the nodes into the unified consideration. The nodes can be assigned to a same cluster only when they have both close space distances and data correlations. Unfortunately, there are always some outliers that their readings are uncorrelated with most of their neighbors and it is impractical to generate some clusters for the outliers only. Therefore, after generating the clusters for most of the nodes, the outliers are assigned to the existing clusters based on the distances between the outliers and the cluster heads. On one hand, close distances between the nodes make it energy-efficient for the nodes in a cluster to communicate with each other.

On the other hand, high data correlations make it accurate for the cluster head to forecast the missing data. In addition, we design a distributed missing values forecasting approach to decrease data transmission in the network. Different from the traditional forecasting approaches, a pre-processing stage is integrated into the framework. Only the high correlated data streams of the nodes are used to forecast the missing data with each other and the uncorrelated data streams are ignored. For the future work, we will focus on reducing the energy consumption further to execute more in-network data processing and prolong the lifetime of the networks. Besides, it is interesting to design a pairwise-nodes correlation measurement monitoring system which is executed by the cluster head. In the system, the pairwise correlations need to be updated in real time which is particular important for real time forecasting.

REFERENCES

[1] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5G era: Enablers, architecture, and business models," IEEE Journal on Selected Areas in Communications, vol. 34, pp. 510–527, Mar. 2016.

[2] A. Zanella, N. Bui, A. Castellani, and L. Vangelista, "Internet of things for smart cities," Internet of Things Journal IEEE, vol. 1, pp. 22–32, Feb. 2014.

[3] S.CiraniandM.Picone,"Wearablecomputingfortheinternetofthings," It Professional, vol. 17, pp. 35–41, Sep. 2015. [4] M. Simsek, A. Aijaz, M. Dohler, and J. Sachs, "5G-enabled tactile internet," IEEE Journal on Selected Areas in Communications, vol. 34, pp. 460–473, Mar. 2016.

[5] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," National Science Review, vol. 1, no. 2, pp. 293–314, 2014.

[6] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," IEEE Transactions on Wireless Communications, vol. 13, pp. 2809–2823, May 2014.

[7] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: issues and challenges," IEEE Network, vol. 30, pp. 46– 53, Jul. 2016.

[8] Y. Shi, J. Zhang, K. B. Letaief, and B. Bai, "Large-scale convex optimization for ultra-dense Cloud-RAN," IEEE Wireless Communications, vol. 22, pp. 84–91, Jun. 2015.

[9] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," IEEE Internet of Things Journal, vol. PP, no. 99, pp. 1–1, 2016.

[10] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber physical systems," Proceedings of the IEEE, vol. 100, pp. 13–28, Jan. 2012.

[11] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," IEEE Transactions on Signal Processing, vol. 63, pp. 4729–4743, Sep. 2015.

[12] K. Yang, Y. Shi, and Z. Ding, "Low-rank matrix completion for mobile edge caching in Fog-RAN via riemannian optimization," in IEEE Global Communications Conference. (GOLBECOM),(Washington, DC, USA), 2016.