JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

An Efficient Data Compression Model for Wireless Sensor Networks Using Deep Learning Technique

[1]R. Logambigai, [2]S. Suba, [3]A. Kannan

[1,2]Department of Mathematics,
[1,2]College of Engineering Guindy, Anna University, Chennai, India
[3]School of Computing Science and Engineering,
[3]VIT- Vellore Campus, Vellore, India

Abstract : Wireless Sensor Networks (WSNs), are essential for real - world application areas. The WSN sensor nodes have a finite amount of battery life. The life expectancy of the detector nodes is increased by making efficient use of battery capacity. The majority of the energy has been used in the transmission of a substantial volume of data that the sensor nodes have gleaned. One efficient way to decrease communication energy consumption in WSNs is data compression. Compaction Model using Denoising Autoencoder (CM-DAE), an innovative model to compress data and lower communication energy consumption, has been proposed in this work using cluster-based WSN. In the proposed work, data pruning is done by cluster member nodes using a deep-learning approach. For the purpose of transmitting sensed information to the base station, the cluster head nodes compress the data using neural networks. Using data from real sensors, the proposed model is examined and analyzed with the frameworks. The experimental results demonstrate that the suggested model outperforms other existing schemes in terms of energy savings, lifetime of the network, and ratio of delivered packets.

***IndexTerms* - Multilayer Feedforward, Compaction Model, Wireless Sensor Networks (WSN), Information Gathering.**

I. INTRODUCTION

Wireless sensor networks (WSNs) are composed of a large number of small, inexpensive sensors to perform a distributed task. The deducting, processing, and information transfer are all parts of how sensor nodes in WSNs operate. The sink can carry out more complicated operations than the regular nodes and is typically connected to a power source. Wireless sensor networks face difficulties with coverage, scalability, low latency, data collection, and other aspects. The tiny batteries that typically power the sensor nodes cannot be changed or recharged.

In a wireless sensor network, many sensors are useful for collecting and communicating information on their own. The network experiences uneven energy consumption as a result of data switch from the origin sensor to the base station. Because neighboring sensors at the drain are most frequently used for data transfer to the sink, they quickly run out of supplies, which leads to the sink getting out of control. The life expectancy of the network is shortened as a result.

Numerous WSN issues consideration must be given in order to produce a highly energy-efficient WSN. Data transmission and collection are one such issue. Sensor nodes only use the majority of their energy during the gathering and transmission of data. The data collection method describes how sensor nodes function to collect information and send it to the drain.

In a cluster-based WSN, all the sensor nodes are grouped into clusters. A distinct node acts as the cluster head in each cluster and the other nodes serve as members of the cluster. Through the group head, the member nodes' sensed data are transmitted to the base station. Various data compression techniques are used in WSN to lower communication energy consumption. Additionally, current neural network-based compression techniques enhance compression while ignoring the network's computation time and WSN application potential. It's possible that the sensor nodes' data collection is redundant and sparse. A decrease in transmission energy is required to lengthen the lifespan of the sensor nodes.

The reduction of energy use during data collection and transmission is the main goal of this paper. Our earlier research on fuzzy logic-based cluster-based networks [1] has been taken into account in this paper. In this study, fuzzy rules are used to form clusters and choose cluster heads. Data is gathered from the cluster members by the cluster head node and placed in the sink through additional cluster head nodes.

While the cluster members in [1] transmit all collected data to the cluster head, the cluster head uses more energy than the cluster members. Moreover, the cluster head gathers information from the other cluster nodes and sends redundant information to other cluster head nodes in order to reach the sink. By sending all of the data through nodes, the majority of the energy is wasted in this situation. A new model called the Compaction Model using a Denoising Autoencoder (CM-DAE) has been proposed to address the aforementioned problems and reducing the sensor node's transmission energy will increase the network lifetime. The proposed approach aims to reduce the amount of data and compress it before transmission by using a machine-learning technique.

The following is a list of this paper's main contributions:

(1) To categorize similar and dissimilar data readings, the cluster member node uses machine learning techniques to reduce energy consumption. As a result, fewer data readings are required before transmission to the group head.

(2) The neural network approach is used by the cluster head node to reduce the amount of data that the group member sensed.

The article's remaining sections are structured as follows. Some of the pertinent work is covered in Section 2. Furthermore, the suggested approach is demonstrated in Section 3. In Section 4, the experimental design and results are discussed. The summary of this proposed work is presented in Section 5.

II. RELATED WORK

Numerous plans and techniques have been created recently to address the problem of data collection and transmission in wireless sensor networks. It is effective to use data compression to reduce the amount of data and communication energy used by WSNs. An artificial intelligence technique that excels at mathematical modeling is called machine learning (ML).

In a WSN cluster-based routing, the authors of [2] proposed a machine learning system called a capsule network and data pruning technique in order to improve throughput, prolong the life of the network, and effectively use energy. To increase system lifetime and energy consumption in cluster-based wireless sensor networks, an effective method of cluster formation using fuzzy rules is proposed in [1]. A novel cluster formation method utilizing neuro-fuzzy rules was presented in [3], and a new routing protocol has been proposed to perform routing effectively for IoT-based WSNs.

Using a grid-based routing algorithm, authors in [4] suggested extending the lifespan of WSNs and conserving the energy of sensor nodes. In [5], the authors introduced a brand-new routing method for sending data from isolated nodes in a cluster-based network to the sink. This method is referred to as Routing with Isolated Nodes. Their work enhances the collection of data from the remaining sensors after cluster formation.

Authors of [6] conducted a survey on ML-based WSN algorithms. They listed benefits, drawbacks, and network lifetime-affecting factors in their survey. Additionally, for power generation, mobile sink scheduling, synchronization, and traffic engineering. Identified input values and the associated output values are provided for learning in supervised learning. This information helps to forecast the outcome of unknown inputs. These kinds of algorithms have been used to address a range of WSN-related problems, such as node localization, high dimensionality reduction, information similarity. A lack of output class labels occurs in unsupervised learning. In order to categorize the input samples into different classes, this type of learning finds similarities among them. The probabilistic Bayes theorem serves as the foundation for the classifier known as Gaussian Naive Bayes. Contrary to how attributes of a class are each independently linked to the parent class, this classifier does not link one class to another. The authors of [7] provided a survey of the various threats that use machine learning techniques to attack both IoT and WSNs. They divided attacks that target the IoT system and WSN into different types in their work. Additionally, they presented the critical attacks for IoT and WSN. Additionally, they conducted a survey on various efforts using machine learning approaches to secure WSNs and IoT. To maximize the energy of WLAN, the authors of [8] proposed a context-aware approach to classifying network traffic using machine learning classifiers. The authors of [9] took a two-level approach to data collection, one at the node level and the other at the group head level. In order to provide aggregated data to the base station, they performed data aggregation using the exponential moving average at the node level and the Euclidean distance function at the cluster head level.

To reduce data transmission, the authors of [10] suggested a mechanism called Zoom-In Zoom-Out (ZIZO). For data compression, the sensor level and cluster head level are the two levels on which the ZIZO mechanism functions. At the sensor level, they combined the data and sent it to the second level using the index-bit encoding compression technique. The cluster head optimizes the sensors' sampling rate in the second level using a process called sampling rate adjustment.

Convolutional Neural Networks (CNNs) have recognized to be incredibly effective in a diversity of fields, which has encouraged the widespread use of ML in a variety of fields. Some data generation models, such as the Restricted Boltzmann Machine (RBM) and Variational Autoencoder (VAE), are created by ML[11]. The authors of [12] conducted a survey on the machine learning techniques applied to WSN problems. They have discussed about the benefits and drawbacks of significant machine learning algorithms in their work. The definitions that capture the fundamental nature of machine learning were also discussed by the authors. The Stacked RBM-AE compression model for sensor nodes data compaction was proposed by the authors of [13] after researching the use of RBM and autoencoder in combination. They use a two-layer structure for their model, with the encoding layer being used to compress sensed data and the decoding layer being used to reconstruct data.

Researchers from [14] investigated various methods of data collection and compared their results. To forecast the effectiveness and dependability of various data-gathering strategies, they developed an analytical model. They came to the conclusion that there is no one best data collection method that works for all applications. The authors of [15] researched secure data aggregation techniques and covered each one's advantages and disadvantages. They have categorized the secure data aggregation methods in their study based on the network framework, configuration, cryptographic techniques, encryption algorithm, security mechanisms, and other elements.

In [16], authors have presented an approach for training the deterministic auto-encoders. A penalty term has been added to the classical cost function and the authors achieved better results than regularized auto-encoders and denoising auto-encoders. The authors in [17] developed a variational autoencoder to model images, inscriptions, and captions. They used Deep Generative

Deconvolutional Network as a image decoder and Convolutional Neural Network as an image encoder. They also used a Bayesian support vector machine for labels and a recurrent neural network for captions.

In [18], authors have introduced a technique for optimizing the regularized objective function within each minibatch which does not decrease the convergence rate when the minibatch size gets increased. According to the authors of [19], a learning process that repeatedly modifies the network's connection weights reduces the amount of variance between the input and output vectors. A study on neurons and weight snipping in Artificial Neural Networks was provided by the authors of [20], and they compared various nodes and weight pruning algorithms. The authors have deduced from the experimental results that node pruning provides a better solution than a non-pruned network. In [21], authors have studied various current pruning methods which enhance the neural network design. They also compared the performance of different pruning methods based on sensitivity analysis and mutual information on actual datasets.

The network-collected sensed data expand exponentially over time. The data has exponentially increased and demonstrates key big data traits like high volume and high variety. The difficulty of data collection is further increased by the wireless sensor nodes' constrained computational and communication capabilities. A new economic model called the Compaction Model using Denoising Autoencoder (CM-DAE) is proposed to address the challenges associated with data collection and to make sensor nodes more energy efficient. This model has been trained to recreate the latently compact sensor values at the sink node.

III. Proposed Work

The proposed Compaction Model using Denoising Autoencoder (CM-DAE) model is given in Figure 1. In this model, machine learning techniques and neural network techniques are used to reduce the amount of data transmission between group member nodes and group head nodes. Moreover, the reduced amount of data is compressed and delivered to the base station through the network's cluster head nodes. In this work, by reducing the amount of data and by compression the communication energy cost is minimized.

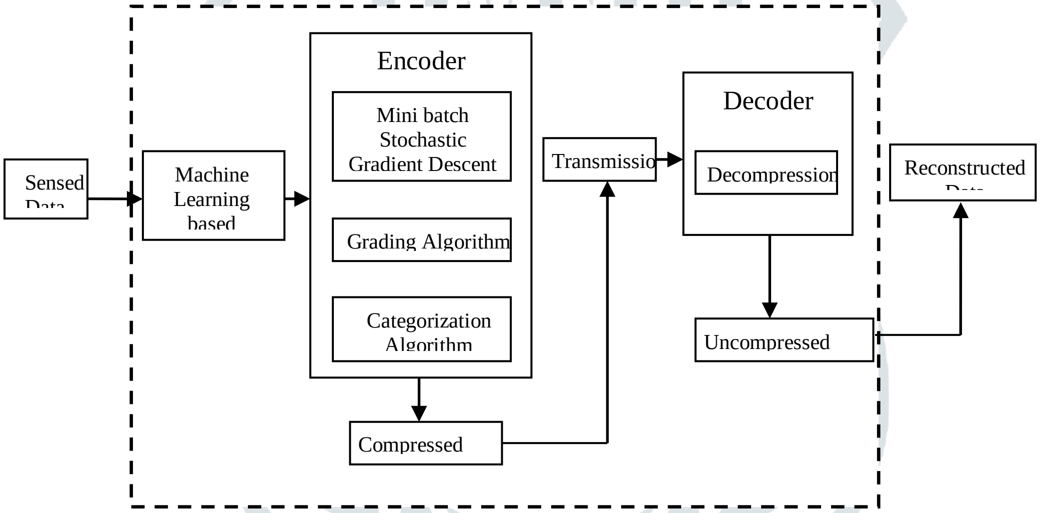

Figure 1. Typical CM-DAE Model

3.1 Machine Learning based classification

The cluster member nodes which use the proposed model categorize the sensed data on a regular basis. The machine learning model is created after the machine learning classifier trains and evaluates the samples. As a result, the model is capable of identifying fresh samples and correctly classifying them. The cluster member node sends the group head all the similar data as one data stream and all the disparate data in its original form. Thus, only a small amount of data is transmitted by each cluster member node, saving energy for the individual sensor nodes.

3.2 Denoising Autoencoder

A neural network that has been trained to try to replicate its input in its output is an autoencoder. The Denoising Autoencoder compresses the data by intentionally tampering with the input data, by introducing noise, or by concealing some of the input values. Denoising Autoencoder (DAE) is trained in this study using sample data that has been categorized. The DAE is instructed to decode the strong feature values from its deeply corrupt format. Figure 2 depicts the encoder and decoder components of the DAE model.

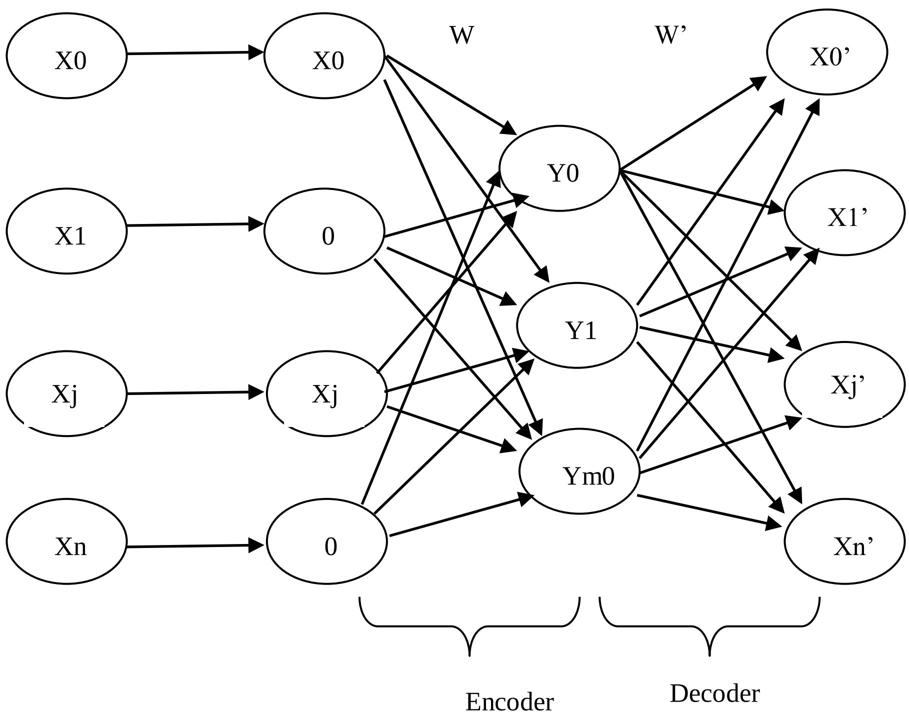

Figure 2. Typical DAE Model

The source data, let's say x is normalized and altered to take on a deeply corrupt form., say x', during the encoding process. This is done by masking some of the x's input entries to 0 up to a corruption level. The more sensed data that is corrupted is set to 0, the larger the. As a result, in this encoder part, the DAE uses a Weight Matrix with the formula $y = f(Wx')$ compress the corrupted data. Then, using, $x' = f(W'y)$ the compressed data is reconstructed in the decoder section of DAE.

The non-linear sigmoid activation function which is used to train the DAE model is given in equation (1).

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

The mini-batch stochastic gradient descent algorithm [18] is repeatedly run in order to lower the error in the network. It reduces the discrepancy between the originally sensed data and the data that were reconstructed. In cluster head nodes, the trained DAE is used to wrap and then recreate the sensed data. The sensed data is compressed in each cluster head node using the weighing matrix W, which can be thought of as the measurement matrix for data. After gathering all of the compact data, the weighing matrix W' can be used to reconstruct the original sensed data in the sink node and is known as the data reconstruction matrix.

3.3 Grading based on denoising autoencoder

The grading algorithm given in algorithm 3.1 is used for dividing the network's entire set of data into many associated groups in a tree-like structure.

Algorithm 3.1 Grading algorithm

Step 1*: Check whether all nodes are grouped or not*
Step 2*: if a node ' j ' n[j]=not grouped and if r[i] > f(r[i])*
Step 3: *parent = j*
Step 4: *Sensed value at root node r[i] = Sensed value at 'j'*
Step 5*: Grouping is broadcasted to all other nodes*
Step 6: *if a node say node i is not yet grouped after some random waiting time then r[i] = i*
Step 7: *parent = 'i'*
Step 8*: Sensed value at root node r[i] = Sensed value at ' i '*
Step 9: *Grouping is broadcasted to all nodes*
Step 10*: Construct a tree structure*

Here *r[i]* is the root node and *f(r[i])* is the sensed value at the root node of *i*. The output of this algorithm is a tree structure consisting of data in each neuron. Once the tree-like structure is constructed, then the sensor node group the compressed and uncompressed data using categorization based on the grading algorithm. The categorization based on the grading algorithm for the DAE model is given in algorithm 3.2. This algorithm distributes the processing of each sensor and compresses the data collected. Each leaf node in the suggested algorithm transmits a label and data as a tuple to its parent. (In the event that the label is set to 0, the sensed data is sent in an uncompressed format. In the absence of that, a compact form of the sensed data is transmitted. This algorithm's objective is to send compressed sensed information to the base station. The sink node then uses equation (2) to compile the compact data from the entire sensor system.

$$z = \sum_{i=1}^{n} x_i W_i \quad (2)$$

Further, the sink node can reassemble the original data from the compact sensed information using equation (3) given below.

$$\bar{x} = f(W' f(z)) \quad (3)$$

Algorithm 3.2 Categorization Algorithm-based grading algorithm

Step 1:*Every sensor node starts by sending a label and data tuple to its parent*

Step 2: *If the label is set to 0, the sensed data are sent in uncompressed format;*

Step 3: *Sensing values are sent in the compressed format if the label is set to 1;*

Step 4:*In the case of internal nodes, there are two cases involved.*

Case 1: *if (k>m-1)*

k= number of children node

m = number of hidden layers

Sensor node 'i' transmits in the uncompressed form to parent node p[i] the sensed data ' x_i *' gathered from all of its children nodes.*

Case 2: *Else, sensor node 'i' computes the compressed data of all its child nodes to p[i] using,* $\Sigma\ (X_i + W_i)$

Step 5: *Collect the compressed data from all nodes at the sink and reconstruct the data.*

IV. Results and Discussions

A set of data from the real world has been used to assess how well the proposed model performs. The Intel Berkeley Research Laboratory WSN dataset[22] has been used during the tests. From February 28 to April 5 of 2004, 54 Mica2Dot sensor nodes were monitored for temperature, humidity, and light intensity on a 30-second cycle. Several tests have been conducted to assess the effectiveness of the proposed scheme. The proposed model CM-DAE has been modeled using Keras and Tensor flow. The proposed model's effectiveness is measured against that of the Stacked RBM-AE[13], CBN-VAE[24], and DCDA[23] models. The suggested algorithms are tested, and the test results are displayed.

From the original temperature dataset, for training, 70% of the dataset was used, 25% for test results, and 5% for confirmation. In DAE the number of hidden layers n is an important parameter. This represents the main feature of the input data. It is obvious that the DAE can provide a higher compression rate the smaller the value of n. As a result, it is comparable to the signal's sparsity.

Using Signal-to-Noise Ratio (SNR) metric given in equation (4) CM-DAE's data reconstruction performance is calculated without considering the input data size.

$$\mathrm{SNR} = 10\log_{10}\frac{\|x\|_2^2}{\|x-\hat{x}\|_2^2} \quad (4)$$

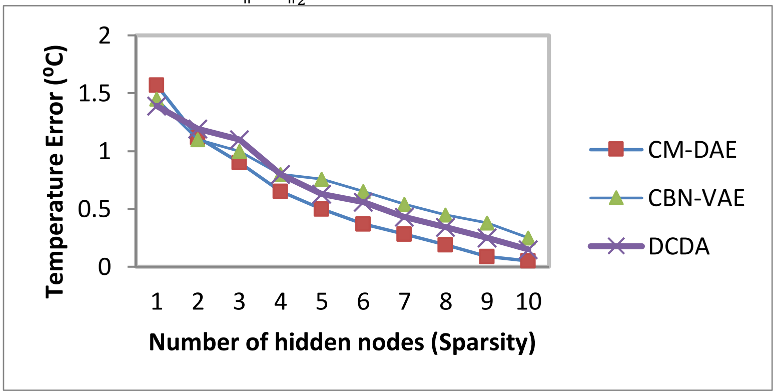

Figure 3. Average reconstruction rates of various models

Figure 3 shows a result of comparing the average reconstruction performance of various models. From the figure it is clear that CM-DAE performs better than the CBN-VAE and DCDA model. Additionally, the DAE's SNR increased and then gradually stabilized as the number of hidden nodes increased. Data compression rates can be raised by lowering the number of hidden nodes.

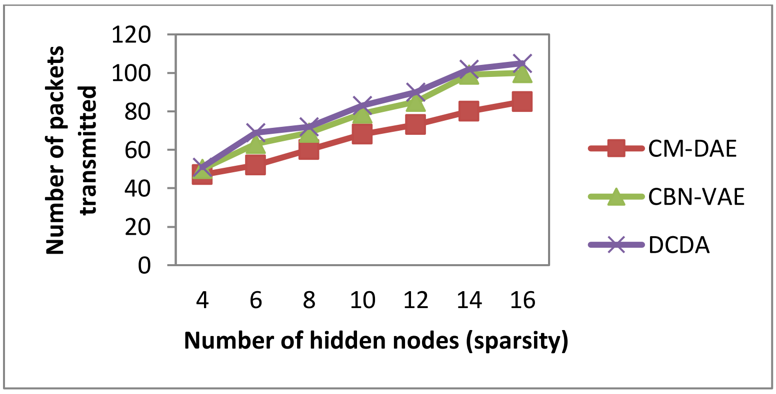

Figure 4. Number of transmitted data of CM-DAE and other various models

Figure 4 shows the quantity of data sent from various model sensor nodes to sinks. This figure makes it clear that our suggested CM-DAE compression model better than other models. It is evident that more data are transmitted is directly proportional to the quantity of hidden nodes. Furthermore, the number of transmitted data of CM-DAE is lower than the DCDA and CBN-DAE, shows that the compression performance is high in our proposed model.

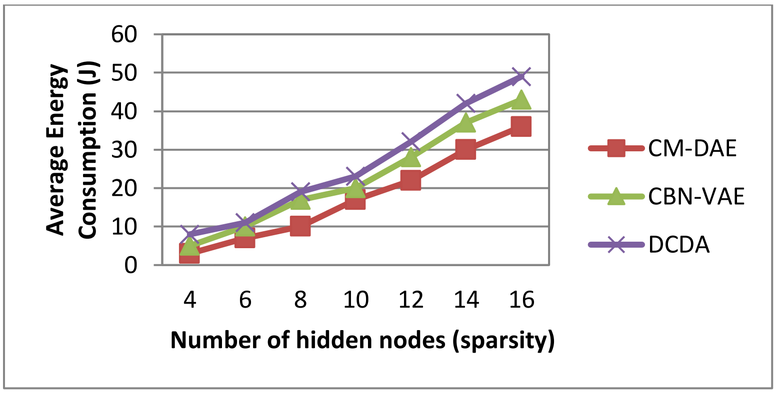

Figure 5 Average Energy Consumption of CM-DAE, CBN-VAE and DCDA

Figure 5 gives the average energy consumption of CM-DAE, CBN-VAE and DCDA models. From the figure, it is clear that our proposed model consumes less energy than other two models. In the proposed model, as the quantity of hidden nodes grows, it compresses more number of values which in turn reduces the transmission energy of the sensor nodes.

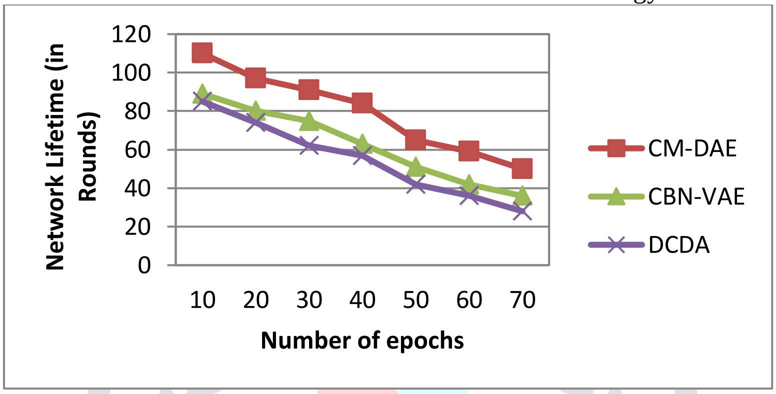

Figure 6 Network Lifetimes of CM-DAE, CBN-VAE and DCDA

The network lifetime of the existing CBN-VAE and DCDA model are compared with proposed CM-DAE model in Figure 6 by running number of epochs on the given temperature data. The figure makes it abundantly clear that our suggested CM-DAE model improves the lifetime of the network than CBN-VAE and DCDA model by compressing efficiently and increases the lifetime of the sensor nodes by lowering the transmission energy of the sensor nodes.

V. Conclusion

In this paper, CM-DAE a compaction model for WSN has been suggested. In this cluster-based WSN, the sensed data are classified into similar and dissimilar data in the cluster member nodes. The sensed data in the cluster head are used to train a DAE., along with its encoder and decoder sections are then utilized for both compressing the sensed data and reconstructing the original data. An additional categorization system based on trees has been developed in order to collect the compressed data in cooperative and hybrid modes. The results of the experiment show that the proposed model is superior to other existing models in terms of data compression rate, average energy consumption, data transmission rate and data reconstruction rate. Future works in this compression approach can be the use of advance deep learning principles for efficient data collection in WSN.

REFERENCES

[1] Logambigai, R., & Kannan, A. 2016, Fuzzy logic based unequal clustering for wireless sensor networks. Wireless Networks, 22(3): 945-957.

[2] Umamaheswari, S. 2020, Capsule network-based data pruning in wireless sensor networks. International Journal of Communication Systems, 33(5): 1-12.

[3] Thangaramya, K., Kulothungan, K., Logambigai, R., Selvi, M., Ganapathy, S., & Kannan, A. 2019. Energy aware cluster and neuro-fuzzy based routing algorithm for wireless sensor networks in IoT. Computer Networks, 151: 211-223.

[4] Logambigai, R., Ganapathy, S., & Kannan, A. 2018. Energy–efficient grid–based routing algorithm using intelligent fuzzy rules for wireless sensor networks. Computers & Electrical Engineering, 68:62-75.

[5] Logambigai, R., Ganapathy, S., & Kannan, A. 2016. Cluster based routing with isolated nodes in WSN. Int. J. Res. Appl. Sci. Eng. Technol.(IJRASET), 4(3):1-4.

[6] Kumar, D. P., Amgoth, T., & Annavarapu, C. S. R. 2019. Machine learning algorithms for wireless sensor networks: A survey. Information Fusion, 49:1-25.

[7] Mamdouh, M., Elrukhsi, M. A., & Khattab, A. 2018 August. Securing the internet of things and wireless sensor networks via machine learning: A survey. IEEE, International Conference on Computer and Applications (ICCA):215-218.

[8] Saeed, A., & Kolberg, M. 2018. Towards optimizing WLANs power saving: Novel context-aware network traffic classification based on a machine learning approach. IEEE Access, 7 : 3122-3135.

[9] Jan, S. R., Khan, R., Khan, F., Jan, M. A., Alshehri, M. D., Balasubramaniam, V., & Sehdev, P. S. 2021. Marginal and average weight-enabled data aggregation mechanism for the resource-constrained networks. Computer Communications, 174:101-108.

[10] El Sayed, A., Harb, H., Ruiz, M., & Velasco, L. 2020. ZIZO: A Zoom-In Zoom-Out Mechanism for Minimizing Redundancy and Saving Energy in Wireless Sensor Networks. IEEE Sensors Journal, 21(3): 3452-3462.

[11] Salakhutdinov, R., Mnih, A., & Hinton, G. 2007, June. Restricted Boltzmann machines for collaborative filtering. In Proceedings of the 24th international conference on Machine learning:791-798.

[12] Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H. P. 2014. Machine learning in wireless sensor networks: Algorithms,strategies, and applications. IEEE Communications Surveys & Tutorials, 16(4): 1996-2018.

[13] Liu, J., Chen, F., & Wang, D. 2018. Data compression based on stacked RBM-AE model for wireless sensor networks. Sensors, 18(12): 4273.

[14] Campobello, G., Segreto, A., & Serrano, S. 2016. Data gathering techniques for wireless sensor networks: A comparison. International Journal of Distributed Sensor Networks, 12(3):1-17.

[15] Yousefpoor, M. S., Yousefpoor, E., Barati, H., Barati, A., Movaghar, A., & Hosseinzadeh, M. 2021. Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: A comprehensive review. Journal of Network and Computer Applications, vol. 190 (103118):1-42.

[16] Salah, R., Vincent, P., & Muller, X. 2011. Contractive auto-encoders: Explicit invariance during feature extraction. In Proc. of the 28th International Conference on Machine Learning,:833-840.

[17] Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A., & Carin, L. 2016. Variational autoencoder for deep learning of images, labels and captions. In Proc. of the 30th Annual Conference on Advances in neural information processing systems, : 1-9.

[18] Li, M., Zhang, T., Chen, Y., & Smola, A. J. 2014, August. Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, :661-670.

[19] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986. Learning representations by back-propagating errors. Nature, 323: 533-536.

[20] Bondarenko, A., Borisov, A., & Alekseeva, L. 2015, June. Neurons vs weights pruning in artificial neural networks. InProceedings of the International Scientific and Practical Conference on ENVIRONMENT. TECHNOLOGIES. RESOURCES, 3: 22-28.

[21] Augasta, M., & Kathirvalavakumar, T. 2013. Pruning algorithms of neural networks—a comparative study. Open Computer Science, 3(3):105-115.

[22] Intel Lab Data, http://db.csail.mit.edu/labdata/labdata.html, 2017.

[23] Li, G., Peng, S., Wang, C., Niu, J., & Yuan, Y. 2018, An energy-efficient data collection scheme using denoising autoencoder in wireless sensor networks, Tsinghua Science and Technology, 24(1):86-96.

[24] Liu, J., Chen, F., Yan, J., & Wang, D. 2019. CBN-VAE: A Data Compression Model withEfficient Convolutional Structure for Wireless Sensor Networks. Sensors, 19(16):1-22.