

# Energy Conserving Techniques of Data Mining for Wireless Sensor Networks—A Review



Pragati Patil Bedekar, Atul Raut, and Abhimanyu Dutonde

**Abstract** Now a day the application of Data mining in many areas has been tremendously increased. Data management and its processing is becoming the active research area for research community. In this paper the major concern is to study and analyse the impact and performance of various data mining techniques over sensor networks (wireless), as the characteristics of sensor nodes and its wireless nature are the primary concerns. The volume and rate for data generation is huge, which is variable in nature; therefore, it is very much essential to design and implement data mining methods for Wireless Sensor Network. This paper mainly focused on to the comprehensive survey of existing data mining techniques, wherein the various limitations and its probable solutions are highlighted in detailed. A transition of traditional mining techniques to newly introduced research is also analysed in this paper. Detailed working and description of compression algorithm has been stated in this paper.

**Keywords** Data mining · Energy conservation · System efficiency · Wireless sensor network

## 1 Introduction

To extract useful or meaningful information from the large databases; the concept of data mining is used. Therefore, it relates to the meaningful or relevant information from the warehouse.

---

P. P. Bedekar (✉)

Department of Computer Science and Engineering, Abha Gaikwad-Patil College of Engineering, Nagpur, India

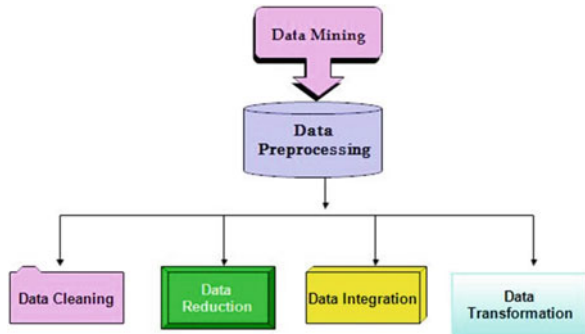
A. Raut

Department of Information Technology, Jawaharlal Darda Institute of Engineering and Technology, Yavatmal, India

A. Dutonde

Department of Information Technology, Tulsiramji Gaikwad-Patil College of Engineering and Technology, Nagpur, India

**Fig. 1** Data processing techniques of data mining



One of the broader areas of Data Mining is Data Pre-processing which includes techniques like

- **Data Cleaning** This technique is used to remove noise and correct inconsistencies in data.
- **Data Integration** This technique is used to combine data from various sources into single warehouse.
- **Data Reduction** Data reduction can be observed by Data aggregation, clustering methods.
- **Data Transformation** Changing format, values, data (transforming data from one form to another) (Fig. 1).

Reliable and real time monitoring plays a vital role in many applications like field monitoring, health and habitant monitoring, tracking of objects, analysis of data etc. Practical monitoring of above mentioned applications is very impotent because maximum time it deals with the fast rate of dynamic and huge amount of data streams which can be geographically distributed all over network. This huge amount of variable data is then analysed, processed and transformed using mining techniques. In this paper existing and revised data mining techniques has been studied.

## 2 Overview of Methods Used

Abdelmoghith and Mouftah [1] The proposed work is focused on the compression of data which is being sent between sink and sensor node. The mathematical results proved that the data reduction is up to 70% and energy saving is up to 37%. The data compression algorithm, S-LZW, LEC, Huffman, ND-Encode, RQT etc., has been studied and analyzed, but the proposed MBC has been proved to be the energy efficient scheme. Mamurjon and Ahn [2] The authors have discussed about Data Collection methods in WSN. Generally gathering sensed data from each node is a tedious job. In general, base node starts from node nearby base station, wherein it has travel all over the network for collecting data from each node, return back

and upload the data to base node. This entire process is energy and time consuming. Hence the author has introduced the concept of mobile sink, where in the mobile sink used to visit every cluster in the network, return back and send sensed information to sink node. The introduced method has decreased energy consumption of nodes. Mudgule et al. [3] have done review on different data reduction techniques like adaptive filter, tree based method, cluster technique, data prediction method, etc. It has been observed that to reduce energy consumption the data reduction technique of Data Mining is important. Data reduction is used to remove unnecessary data while transmission. In this technique, it reduces the repeated data, its processing and analyzed it before transmission and hence as the amount of data reduce it leads to energy preservation in WSN.

### 3 Classification of Compression Techniques

Two algorithms are discussed in this paper that is compression and reconstruction algorithm [4]. Compression algorithm takes  $X$  as input and generates a compressed  $X_c$  whereas reconstruction algorithm executes on compressed representation  $X_c$  and gives output as the reconstruction  $Y$ .

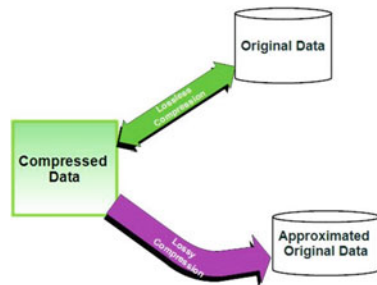
Based on requirement of reconstruction, a data compression technique has been classified into two broader classes as shown in Fig. 2.

1. Lossless compression scheme, reconstruction of original data from compressed data, in which  $Y$  is identical to  $X$  and
2. Lossy compression scheme, provide higher compression than aforesaid compression but allow  $Y$  to be different from  $X$ .

The ultimate aim is to reduce energy consumption of the Network, which can be achieved by using compression techniques. It deals with conversion of data strings of characters into another representation which may have same data in small length as much as possible. The main objective is efficient coding and minimize bandwidth required for transmission.

The main objective of this review is to show performance of different lossless and lossy compression techniques and their comparative study. This technique helps to

Fig. 2 Working of compression technique



save battery power and data storage capacity or transference capacity. It has been studied that more energy is utilized in transmission rather than the processing of information therefore the sending compressed data can work out at some extent. Following is the review on different algorithms/schemes used for data compression.

### 3.1 String Based Compression Techniques

**LZW** is a general technique used for compression of data. This technique is a universal lossless data compression algorithm. It is also called as dictionary based algorithm [5–10]. In the year 1977, A. Lempel and J. Ziv published this algorithm.

Welch [11], Singh and Singh [12], Saidani et al. [13] The LZW algorithm: it never send character, and it tries to send codes for already known strings. Each time a new code is sent, a new code is attached to the dictionary. Working of algorithm is as follow:

- a. Reading of incoming text (no analysis will be performed here)
- b. Check and verify whether the string has been read before or not
- c. If no in step b, then read string and add it to the dictionary
- d. Compression of text will be done if the new code is written instead of string.

The start dictionary contains standard character set, that is, ASCII Codes. Two control codes namely 256 and 257 are attached to this dictionary. The 256 is used as an EOF (end of file) character, and the 257 is used as an EOD (end of dictionary) character. Example,

Let us consider the string c a b b a c a b b a.

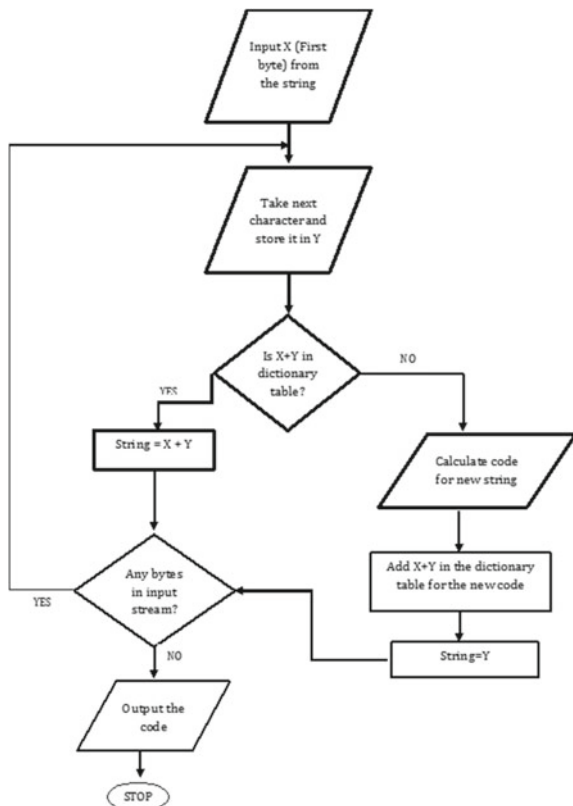
Index	Dictionary	Step	X	y	x+y	Code	
1	A	1		c	C		
2	B	2	C	a	Ca	3	
3	C	3	A	b	Ab	1	
4	Ca	4	B	b	Bb	2	
5	Ab	5	B	a	Ba	2	
6	Bb	6	A	c	Ac	1	
7	Ba	7	C	a	Ca		String Matched
8	Ac	8	Ca	b	Cab	4	
9	Cab	9	B	b	Bb		
10	Bba	10	Bb	a	Bba	6	String Matched
11		11	A		A	1	

At the start, the dictionary contains all possible individual characters and x is empty; y contains the next character in the char stream. Now, let us find out whether the string  $x + y$  is present in the dictionary or not. If it is not found in the dictionary then add  $x + y$  string to the dictionary.

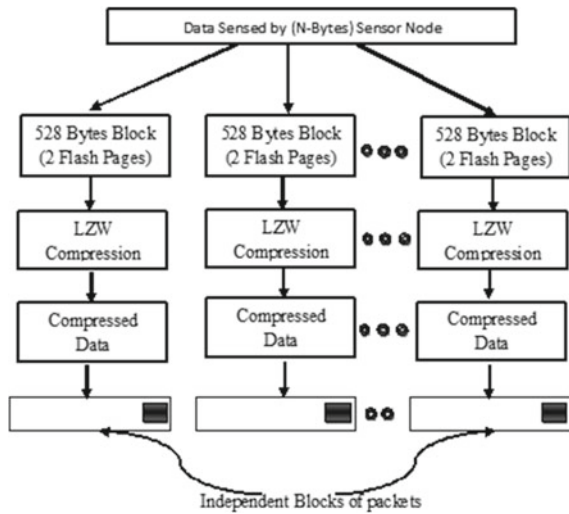
Hence after applying concept of LZW algorithm over the above text the code 3 1 2 2 1 4 6 1 is produced as output.

**S-LZW** here **S** stands for Sensor. S-LZW is a dictionary/string based lossless compression algorithm which is used to support data compression in WSN. It is stated that to transmit plain text data, more energy is required as compared to compressed text. Compressed form data gives more reliability comparatively. It has been observed that in transmission process of data, packet loss occur in WSN at more rate; hence, S-LZW algorithm follows the strategy to divide the incoming data stream into small and independent blocks as shown in the figure. While doing this the algorithm can assure that if packet loss happened in transmission process then it only affect the packet loss following packets in its own block and therefore rest of blocks will be secure enough. With reference to [8] as resources are limited in WSN, therefore a dictionary of size 512 entries is adopted to fit in small memory of a sensor node. At the same time, the algorithm said to compressed data into the block of 528 Bytes as shown in the figure which helps to improve performance. Now the next step is to divide the compressed data into many blocks of independent packets (Figs. 3 and 4).

**Fig. 3** Flowchart for LZW algorithm



**Fig. 4** Working of S-LZW algorithm



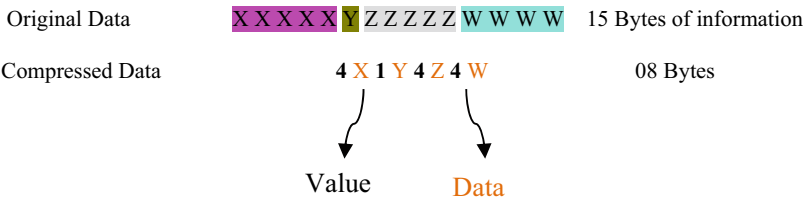
### 3.2 RLE and K-RLE

Run Length Encoding (RLE). As mentioned in S-LZW algorithm, transmission and reception of data consume more power. In order to reduce overhead of energy consumption while data transmission which is also called as in-network processing technique, enhanced compression techniques are used—RLE is one of them. First we will see the working of RLE in detail [9].

RLE (lossless) is a very easy technique which is used to compressed digital data. The data is stored as single data in the form of Value and Count. It is more benefited when data carries the sequences of same data value occurrences in the original runs or data stream; hence, instead of storing same original run, the data will be stored in terms of data value and count. This method is needed to reduce the amount of data needed for storage and transmission.

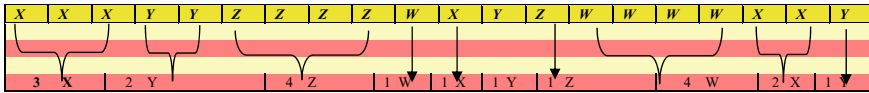
General definition is *d* is a data item occurs at *n* consecutive times in the input stream, replace *n* occurrences with single pair *nd*.

Example: Consider the following original run length.



RLE is used where we expect the consecutive sequence of same runs. In the given example, similar occurrences of X, Y, Z and W are counted and compressed as Count followed by data or vice versa.

- With this example we can see the clear difference of using LZW and RLE.



**K-RLE** means RLE with K-Precisions. In this technique if data items  $d$ ,  $d + K$ , or  $d - K$  occurs  $n$  consecutive times in the input stream then replace the  $n$  occurrences with the single pair  $nd$  [8].

### 3.3 Coding by Ordering

Coding by ordering is a method used to minimize amount of data need to send for further communication. The main objective is to increase network lifetime by using two schemes mainly a. Data Aggregation b. Data Compression [7].

Step 1. The controller breaks down the space into different regions also known as cuboids and transmit the interested data packets to each region.

Step 2. After step 1, maximum nodes transmit sensed data to the controller after the specified intervals.

Step 3. As almost all nodes send data at the specified time, then it will become easier to combine all the readings into a single packet.

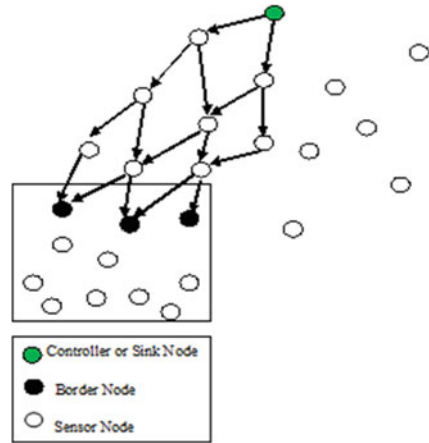
Step 4. One data packet with only one adder will travel from the region to the controller (Fig. 5).

Data Funneling (routing algorithm) has been introduced [7] to reduce the communication overhead while sending the set of sensed data by the number of sensor nodes to the single destination called as sink node. The proposed routing algorithm not only benefited in energy conservation but also used to reduce the data packet collision in the network.

For the compression of data ‘Coding by Ordering’ scheme has been used which losslessly compressed data encode learning information in the sequence or order.

The nodes (nodes) disbursed in large region will send data to the border node. Border node is point where all data from the region (send by  $n$  nodes) will be collected. In this case sensor nodes will compute the communication cost for border node only instead of sink (controller) node. As border nodes needs to communicate with controller so it will compute its communication cost.

**Fig. 5** Coding by order



Border node is the responsible one to place each node’s packet into a big super-packet containing sensed information from environment then send only one big-packet to control node. The packet (node) contains the NODE ID and node ID contains its position and payload. The Border node contains the nodes ID convey belonging of payload assign to which particular node. The authority to choose ordering of packets to super packets is managed by border node, and it will not affect application as the all data reached at controller at the same time.

Example: Assume that there are four nodes in the region with ID’s 1, 2, 3, and 4. Each of these sensor ranging from 1 to 4 sensed some number values let us say {0, ..., 5}. It is stated earlier that the border node has freedom, in the below example the data packets of node 4 has been suppressed by border node, and now it will select possible permutations combination for the remaining 3 nodes that is  $3! = 6$  which is shown in below table [7] (Table 1).

**Table 1** Mapping from permutations to integers

Permutations	Integer
123	0
132	1
213	2
231	3
312	4
321	5



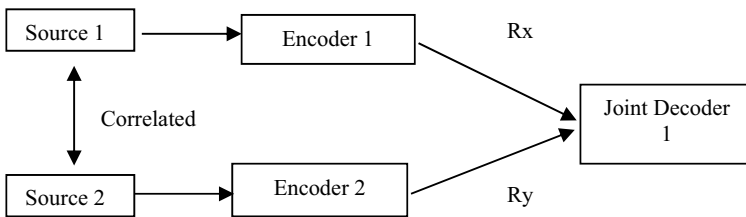
### 3.4 Pipelined In-Network Compression

The authors' motivation behind this compression technique is to save energy required for the communication in wireless network. Thereby reducing the repetitive data stream the said objective has been achieved by Pipelined In-Network compression method.

1. Sensed data by node (mote) is buffered at own node itself
2. Delay for some appropriate time satisfying End-to-End latency-bound
3. Remove same or repeated data from the buffered data.

### 3.5 Distributed Compression

Distributed Compression as the name suggests it is used to do compression in a distributed manner. Here without the communication between n motes or sensor nodes the data redundancies can be removed in a distributed manner.



As it is clear from the above figure, information can jointly compress correlated observations, even with no communication between nodes.

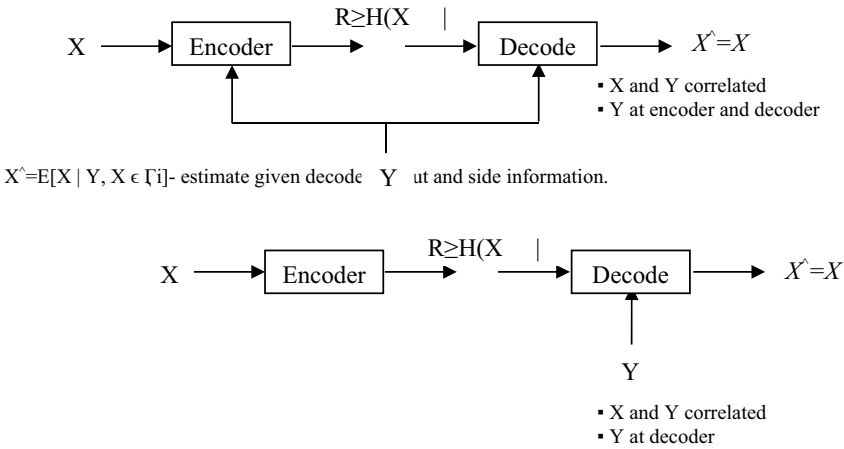
- Source coding with side information-discrete alphabets

X and Y are two data sets likely equal with length of three binary data. Hamming Distance between X and Y is 1 at most.

$$X = [0\ 1\ 0], Y = [0\ 1\ 0], [0\ 1\ 1], [0\ 0\ 0], [1\ 1\ 0].$$

X can be compressed to two bits. If both decoder and encoder know Y, what will happen to the compression rate of X? Then if Y is known to the decoder.

According to this method, since the Decoder know Y and X is only one Hamming Distance apart from Y, it is not sufficient to differentiate  $X = 111$  from  $X = 0000$  [6].



For same reason,  $X = 001$  and  $110$ ,  $X = 010$  and  $101$ , and  $X = 011$  and  $110$  do not need to distinguished from each other. These set of two  $X$  values are grouped as four coset and assigned four different binary index numbers:

coset 1 =  $(000, 111)$ : 00, coset 2 =  $(001, 110)$ : 01, coset 3 =  $(010, 101)$ : 10, coset 4 =  $(011, 100)$ : 11.

If input is  $X = 010$  and  $Y = 110$ , Decoder received  $Y = 110$  as a side information from  $Y$  and  $X = 10$  as a partial information from  $X$ . Then at the Decoder,  $X = 010$  is selected from coset 2 since  $110$  has a hamming distance of 2 from  $110$ . Whether  $X$  know  $Y$  or not,  $X$  can still reduce or compress three bits information into two bits.

This method is particularly used in discrete sources and continuous sources. The same method useful for lossless and lossy compression schemes.

## 4 Conclusion

The major concern to study compression techniques is the constrained of battery of sensor nodes. However, many researchers has proposed the use of compression techniques of data mining to extend the life of sensor networks. The studies stated that comparatively lossless data compression methods are easy to implement and understand, also gives better compression efficiency. This paper talks about the comprehensive survey of various compression algorithms and its detailed working. The compression algorithm of both categories lossless and lossy has been studied and analysed here. The main objective to study different types of compression algorithm is to understand how energy can be saved with the help of it. Data gathering, its processing, aggregation and compression have become the interested field of research.

With the use of basic data mining techniques, in future the paper will focus on various types of aggregation methods, fusion techniques, funnelling techniques,

storage, analysis, and knowledge extraction, and compression of huge volume of data is also the key concern in future.

## References

1. Abdelmoghith, E. M., & Mouftah, H. T. (2013). A data mining approach to energy efficiency in wireless sensor networks. In *International Symposium on Personal, Indoor and Mobile Radio Communication: Mobile and Wireless Networks* (pp. 2612–2626).
2. Mamurjon, D., & Ahn, B. (2013). A novel data gathering method for large wireless sensor networks. In *ELSEVIER-International Conference on Electronic Engineering and Computer Science. IERI Procedia*, 4, 288–294.
3. Mudgule, C. B., Nagaraj, U., & Ganjewar, P. D. (2014). Data compression in wireless sensor network: A survey. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(11), 60–63.
4. Preethi, A., & Ramakrishnan, A. (2015). A survey on data compression techniques in wireless sensor network. *International Journal of Electrical and Computer Engineering*, 2(3).
5. Vidhyapriya, R., & Vanathi, P. (2009). Energy efficient data compression in wireless sensor networks. *The International Arab Journal of Information Technology*, 6(3), 297–303.
6. Sadler, C. M., & Martonosi, M. (2006). Data compression algorithms for energy—constrained devices in delay tolerant networks (pp. 265–278). ACM
7. Kimura, N., & Latifi, S. (2005). A survey on data compression in wireless sensor networks. In *Proceedings of the International Conference on Information Technology: Coding and Computing* (pp. 8–13).
8. Alwadi, M., & Chetty, G. (2015). Sensor selection scheme in temperature wireless sensor network. *International Journal of Wireless & Mobile Networks (IJWMN)*, 7(3), 47–53.
9. Kusuma, J., Doherty, L., & Ramchandran, K. (2001). Distributed compression for sensor networks. In *Proceedings of the IEEE Conference on Image Processing, Thessalonki, Greece, October 2001* (pp. 82–85).
10. Capo-Chichi, E. P., & Friedt, J.-M. (2009). K-RLE: A new data compression algorithm for wireless sensor network. In *Third International Conference on Sensor Technologies and Applications* (pp. 502–507).
11. Welch, T. A. (1984). A technique for high-performance data compression. *IEEE*, 8–18.
12. Singh, N., & Singh, H. (2019). Energy efficient NK-RLE data compression scheme. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(6), 4014–4021.
13. Saidani, A., Jianwen, X., & Mansouri, D. (2020). A lossless compression approach based on delta encoding and T-RLE in WSNs. *Hindawi Wireless Communications and Mobile Computing*, 2–10.