Hybrid Optimization algorithms for Image Classification using Deep Learning

Dr. Muralisankar, Mr. S. Vijayananthan, Mr,Arunkumar, Mr. M. Jaweriya Sharieff,
Ms. A.R.Rizwana, Ms. S. Thaslima, Ms. I.Jasheena,
Al – Ameen Engineering College,Erode

Abstract

Big Data was a collection of massive databases that each include a massive quantity of data, primarily unstructured information like social media information, online storage information, IoT, and actual information. This information must be mined. Big data mining is more than just unstructured data mining; it's about forming associations and deducing bigger patterns. We may summarise that big data was not a specific technology utilised for data mining, with 'Hadoop' being one of them. To find the useful information from massive amount of data to organizations, we need to analyze the data. Mastery of data analysis is required to get the information from unstructured data on the web in the form of texts, images, videos or social media posts. This paper presents an overview on Big Data, Advantages and its scope for the future research. Big Data present opportunities as well as challenges to the researchers. An overview on opportunities to healthcare, technology etc. is given. This paper gives an introduction to Hadoop and its components. This paper also concentrates on application of Big Data in Data Mining.

1.Introduction

It refers to the efficient handling of large amount of data that is impossible by using traditional or conventional methods such as relational databases or it is a technique that is required to handle the large amount of data that is generated with advancements in technology and increase in population. Big data helps to store, retrieve and modify these large data sets. For example with the advent of smart technology there is rapid increase in use of mobile phones due to which large amount of data is generated every second, so it is impossible to handle by using traditional methods hence to overcome this problem big data concepts were introduced. most analysts and practitioners currently refer to data sets from 30-50 terabytes(10 12 or 1000 gigabytes per terabyte) to multiple peta-bytes (1015 or 1000 terabytes per peta-byte) as big data. Figure No. 1.1 gives Layered Architecture of Big Data Systems

There are numerous techniques for data collection and aggregation that employ tree structure. The method proposed in [1], which combines Directed Diffusion Protocol and Clustering can analyze, gather, and aggregate information from sensor nodes without relying on the surrounding environment. The majority of data collection methods are concerned with extending network life and conserving energy [2]. In this way, each node is dependent on its position in the tree and sends data in the appropriate time window.[3-5] This approach's node synchronization for receiving and sending data might significantly lower average energy usage. FTEP was a two-level clustering-based distributed and dynamic CH election algorithm. In EEMC [6], CHs at each level are chosen using a likelihood function that takes the excess energy and distance variable into account extremely well. For cluster creation, the entire information is supplied and received by the sink node in this approach. Chiu- Kuo Liang et al. suggested Steiner Points Grid Routing. [7] A new virtual network structure is created instead of the virtual network in GGR to minimize the overall energy usage for data transfer between the sink node and the source code. The concept is to build a virtual grid structure out of square Steiner trees [8].

[9]establish the tranquility and possibility of mobility sink which is used to accept the different adaptability sinks to deal the core of the wireless sensor in different institution to shape the groups in that manner. Hence the adaptability sink examines a single sensor point to bounce, collect data from various places and finally announce the base station.

[10]evaluate sink mobility in the wireless sensor network. He suggests a transverse process depends on scheduling the function. Using this mechanism process, thee interesting way to plan an algorithm was executed to determine the easiest way to remove the obstacles. The sink mobility was determined to walk in this way and gather information from the efficient head of the cluster through the transmission directly. The sink of mobility return back to the region after gathering the information. The result of the simulation proceeds to present the algorithm to extend the lifetime of the network in an effective manner. The mechanical process recommends the procedures to make a smart idea to reduce the difficulties in wireless sensor network transverse with problems.

[11] proposed an essential aggregate information algorithm called routing of congregation chain agent for mobile. This Cluster Chain Mobile chain Agent Routing chooses the value of cluster head to utilize the selected optimal cluster heads. The member of the congregation is used to form the chain and also to transact the information to the mobility sink. The loss of way, power of the signal, and energy level in a residual way are determined to align the way for routing the mobile agent.

[12-13] present a flexible core point with a wireless sensor network to save the meeting point conveniently for the algorithm of mobile routing for energy efficiency to extend the life of association of wireless sensor network. The collecting of cluster head was selected by the importance of waiting to remove the core point of very low essential connection in the sink. [14-`15]present a dense wireless sensor network, where every joint might transfer its message to the correct joints. The path of the joints is long and short hops or using a range of large radio connectionmethods.

2. Proposed method

Researchers assume the network is split into many clusters, each with its cluster-head (CH). The suggested method operates in two stages and works separately in each cluster.

2.1 Flow of Data Packets

The cluster head transmits the data packet to its neighbors during this phase. The following information is included in the data packets:

1. Node position: Each node must be aware of its previous position.
2. Existing Energy: A node's residual energy.
3. Information Label: The data value detected by a node.
4. Hop count: The no. of hops from the CH.

When a node gets a data packet, it deems the sender to be one of its potential parents and saves the data. The packets' node position, existing energy, and information label sections are updated. The hop count is incremented, and the packet is sent to its neighbors. This procedure will continue until the data packet is received by all members of the cluster.

2.2 Data Forwarding and Tree Construction

When all of the nodes have received the data packet, each one chooses a parent to whom it should send its data. The following filters will be used to make this decision: The parent with the shortest hop length from the CH will be chosen first. If there are many nodes with the shortest hop distance, the parent will be chosen from those with the greatest existing energy. All of the aforementioned factors contribute to optimal parent choices. Filter 1 determines the shortest route between a node and the cluster head. Filter 2 extends the grid's life by including the most robust nodes. Filter 3, by examining the nodes' data connection, reduces network overhead.

2.3 Model of Energy

Our energy model is similar to that of [23]. The energy usage for transmitting K bits in this model is equivalent to:

$$E_{TX}(K,d) = E_{elec} \times K + \varepsilon_{amp} \times K \times d^2$$

And the energy required to receive K bits is:

$$E_{RX}(K) = E_{elec} \times K$$

$$\varepsilon_{amp} = 100 \text{ pJ/bit/m}^2 \quad E_{elec} = 50 \text{ nJ/bit}$$

2.4 Using inter-cluster multi-hop routing to send the big coefficients to the BS:

As shown in Figure 1, all sensors are dispersed at random in a circular detecting area with the BS at the centre. Because CHs are picked at random from the sensors with a probability of Nc/N, they are likewise dispersed at random across the region.[16-18] Assume that you have a method for forming a routing tree that connects all of the CHs, with the root being the BS at the center of the sensing region, as described in [20]. A random variable, represented as x, can be used to represent the length between a random BS and the CH. P_{hops} $_{(x)}$ denotes the likelihood of being able to link at a distance x using hops or fewer hops.

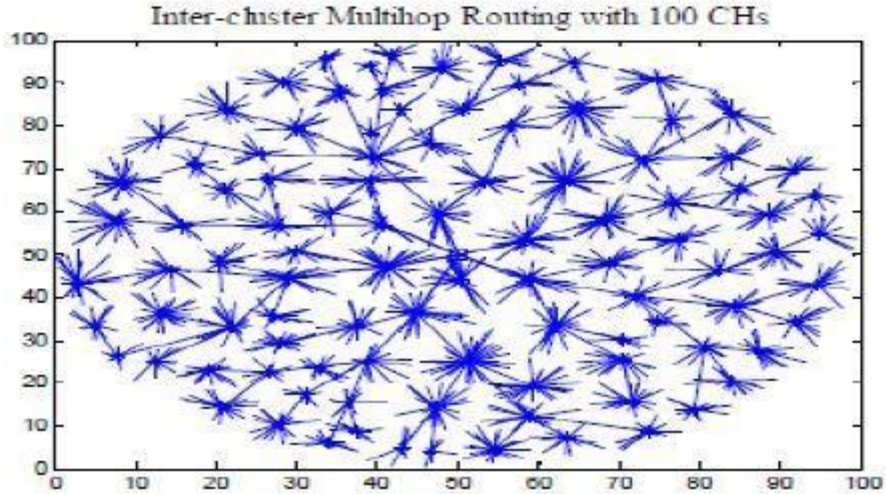

Fig. 1. Network transmissions with inter-cluster multi-hop routing are marked as Phops(x) when the BS at the center has fewer hops.

The average value of the number of hops (hops) is determined as follows in the paper [19].

$$E[hops] = max(hops) - \sum_{n=1}^{max(hops)-1} \frac{P_{hops}(x)}{P_{max(hops)}(x)},$$

The highest number of hops permitted is *max(hops)*. Lastly, calculate the total power used for transmitting K big coefficients from CHs to the BS as follows:

$$P_{toBS} = \left\{ hops_{max} - \sum_{hops=1}^{hops_{max}-1} \frac{P_{hops}(x)}{P_{max(hops)}(x)} \right\} R^2 K,$$

where *R* represents the communication range of the CH. Based on the CH density, this value can be adjusted.

$$P_{total} = (\frac{N}{N_c} - 1)\frac{L^2}{2\pi} + E[hops]\, R^2 K.$$

To put it another way, if the number of CHs decreases, we must raise R to keep all CHs linked as a routing tree. Using multi-hop relaying, we can calculate the overall power usage for data gathering across the whole network. The total transmission power usages are linear functions of the number of massive coefficients K, as shown in Equations (18), (20), and (23).

2.5 Hadoop:

Solution for Big Data Processing Hadoop was developed by Google's MapReduce that is a software framework where an application break down into various parts. Hadoop is an open source project It consists of many small sub projects which belong to the category of infrastructure for distributed computing. Hadoop mainly consists of: • File System (The Hadoop File System) • Programming Paradigm (Map Reduce)

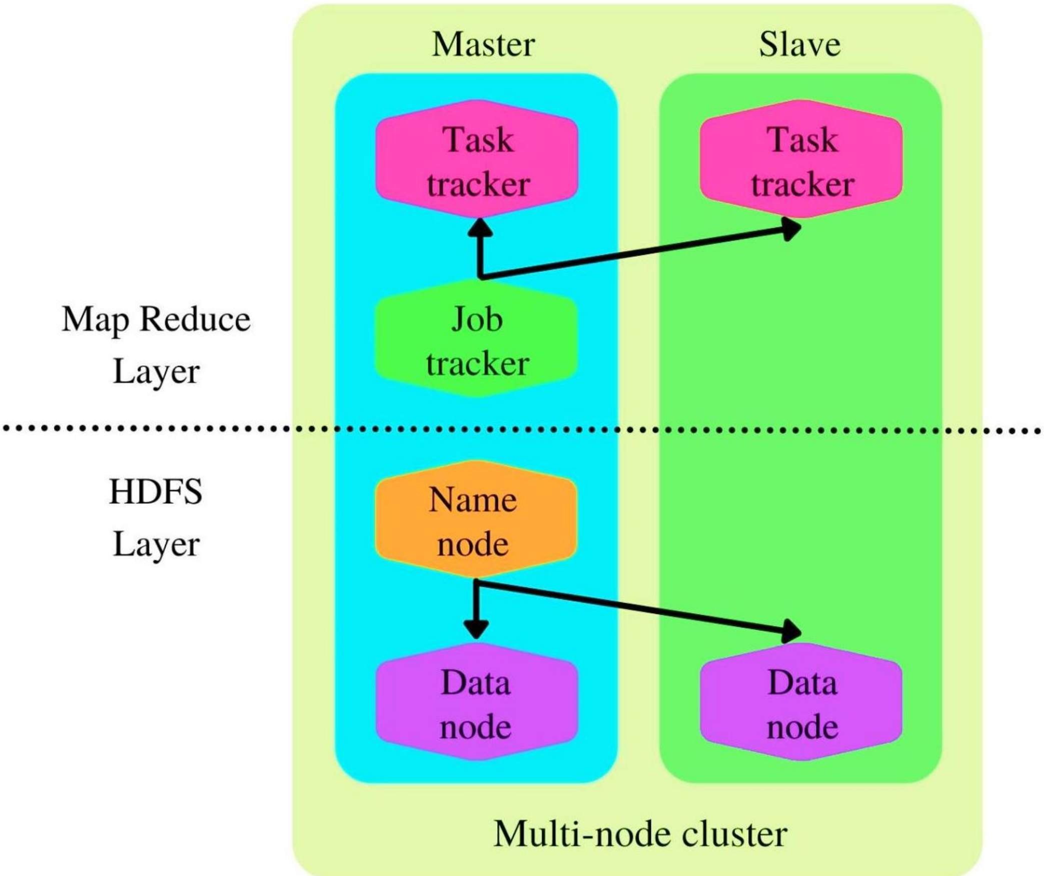

Fig. 2: Hadoop Architecture

2.6 HDFS Architecture

Hadoop includes a fault-tolerant storage system called the Hadoop Distributed File System or HDFS. HDFS is able to store huge amounts of information, scale up incrementally and survive the failure of significant parts of the storage infrastructure without losing data. Hadoop creates clusters of machines and coordinates work among them. Clusters can be built with inexpensive computers. If one fails, Hadoop continues to operate the cluster without losing data or interrupting work, by shifting work to the remaining machines in the cluster. HDFS manages storage on the cluster by breaking incoming files into pieces, called “blocks,” and storing each of the blocks redundantly across the pool of servers. In the common case, HDFS stores three complete copies of each file by copying each piece to three different servers.

2.7 Map Reduce Architecture

The processing pillar in the Hadoop ecosystem is the MapReduce framework. The framework allows the specification of an operation to be applied to a huge data set, divide the problem and data, and run it in parallel. From an analyst’s point of view, this can occur on multiple dimensions. For example, a very large dataset can be reduced into a smaller subset where analytics can be applied. In a traditional data warehousing scenario, this might entail applying an ETL operation on the data to produce something usable by the analyst. In Hadoop, these kinds of operations are written as MapReduce jobs in Java. There are a number of higher level languages like Hive and Pig that make writing these programs easier. The outputs of these jobs can be written back to either HDFS or placed in a traditional data warehouse. There are two functions in MapReduce as follows: • map – the function takes key/value pairs as input and generates an intermediate set of key/value pairs • reduce – the function which merges all the intermediate values associated with the same intermediate key

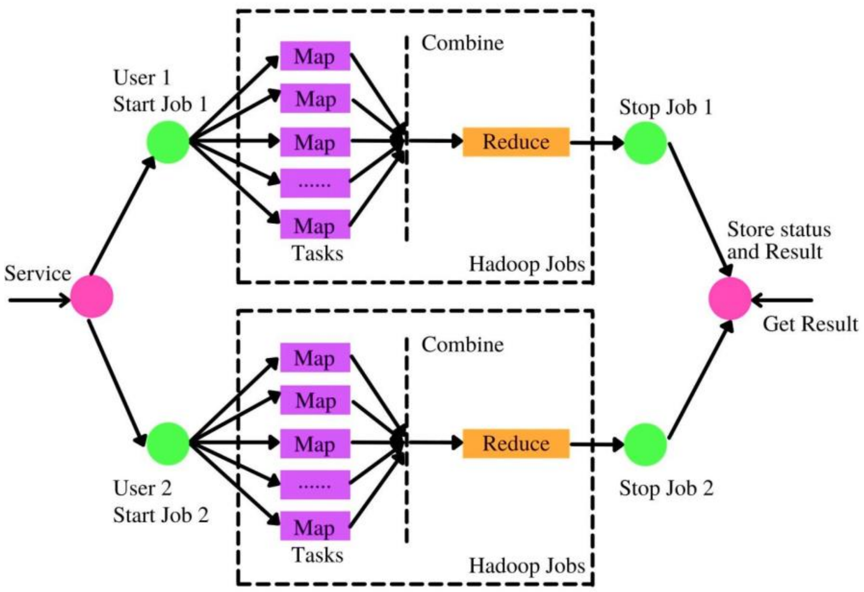

Fig. 3: MapReduce Architecture

2.8 Big Data Advantages and Good Practices

The Big Data has numerous advantages on society, science and technology. Some of the advantages (Marr, 2013) are described below: 1. Understanding and Targeting Customers This is one of the biggest and most publicized areas of big data use today. Here, big data is used to better understand customers and their behaviors and preferences. Companies are keen to expand their traditional data sets with social media data, browser logs as well as text analytics and sensor data to get a more complete picture of their customers. The big objective, in many cases, is to create predictive models.

2. Understanding and Optimizing BusinessProcess Big data is also increasingly used to optimize business processes. Retailers are able to optimize their stock based on predictions generated from social media data, web searchtrends and weather forecasts.HR business processes are also being improved using big data analytics.

3. Improving Science and Research Science and research is currently being transformed by the new possibilities big data brings. Take, for example, CERN, the Swiss nuclear physics lab with its Large Hadron Collider, the world's largest and most powerful particle accelerator. Experiments to unlock the secrets of our universe – how it started and works - generate huge amounts of data. The CERN data centre has 65,000 processors to analyse its 30 petabytes of data. However, it uses the computing powers of thousands of computers distributed across 150 data centres worldwide to analyse the data. Such computing powers can be leveraged to transform so many other areas of science and research.

3. Simulation results

The researcher will look at both sorts of networks in this segment: square sensing areas with a dimension of 100 x 100 and circular sensing areas with a radius of R0 = 50. K-means and LEACH are 2 typical clustering techniques to be contrasted with our study; K-means reduces intra-cluster energy usage, whereas LEACH optimizes power exhaustion between CHs and non-CH sensors. Sensorscope: Sensor Networks for Environment Management signals, both unsorted and sorted, are taken into account.

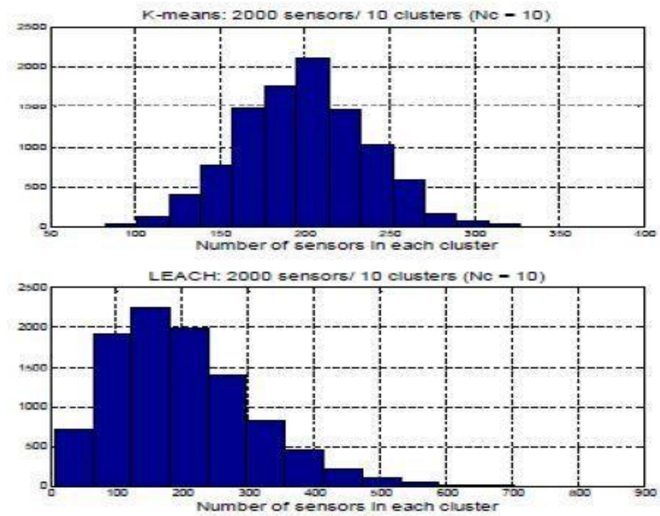

Fig. 4. Histograms from 2 clustering algorithms, Kmeans and LEACH, on a network of 2000 sensors spread across a 100 × 100 square area.

Different values of K are provided by these forms of data, which impact both the transmission cost from the CHs to the BS and the reconstruction inaccuracy at the BS. The network's power usage is shown first, followed by the DCT reduction results. The normalized reconstruction defect is used to calculate the signal retrieval reconstruction error.

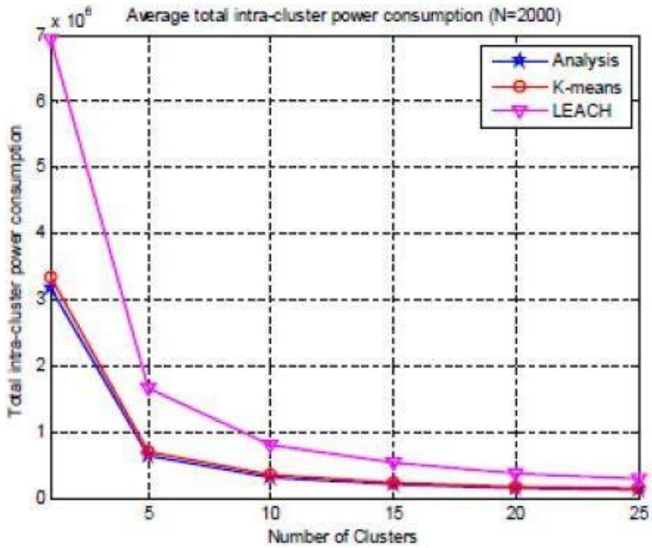

Fig. 5. *Total intra-cluster energy consumption for a network with 2000 sensors spread across a 100 x100 square area.*

Figure 2 shows how 2 clustering methods, Kmeans and LEACH, evaluate the probability of the number of sensors in each group. As can be seen in Figure 3, Kmeans gives a more consistent size for clusters than LEACH, resulting in lower intra-cluster power usage. It is proven that as the clusters get smaller, the intra-cluster used power decreases. Figure 4 displays the overall power required by the network to send data to the BS at the Li = 3L location.

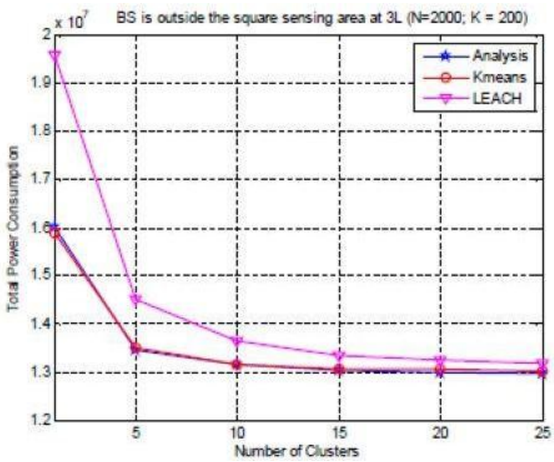

Fig. 6. Total power consumption for all transmissions in the network with 2000 sensors partitioned into a different number of clusters when the BS is outside the sensing area at 3L

Researchers employ a greedy method introduced in [19] to construct a routing tree between CHs to work in the ring sensing region with multi-hop routing. Researcher utilizes multiple broadcast ranges R = [30 50 40 31 20] equivalent to Nc = [20 60 110 250 350] for each network partitioned into multiple numbers of clusters.

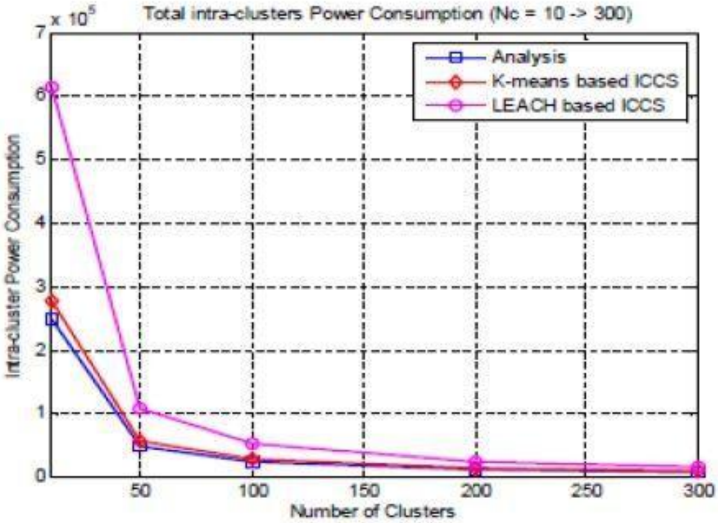

Fig. 7. Total intra-cluster power consumption for the network with 2000 sensors distributed in a circular sensing area with radius R0 = 50

The intra-cluster energy consumptions computed using K-means, LEACH, and the analysis scenario are shown in Figure 5.

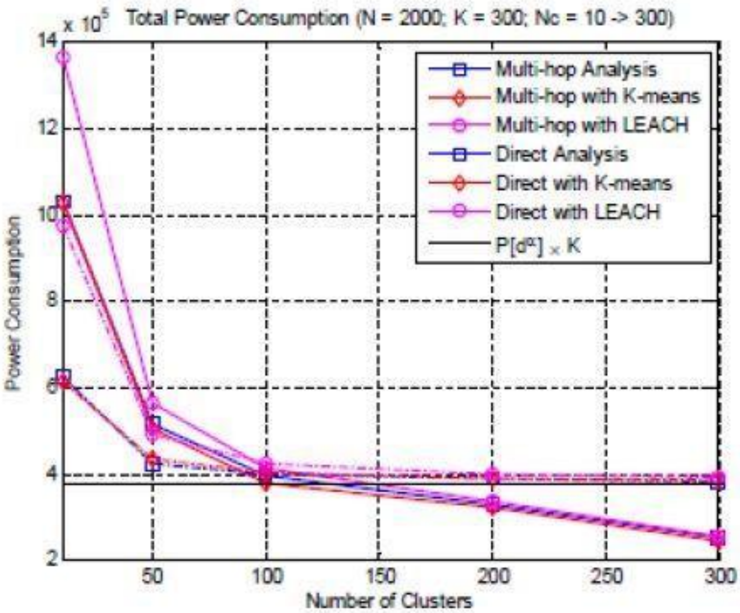

Fig. 8. Total power consumption for all network transmissions using inter-cluster multi-hop to forward K large coefficients the BS in a circular sensing area with radius R0 = 50

Figure 6 analyzes and contrasts the overall power usage of both multi-hop and direct techniques for forwarding coefficients to the BS. When the number of clusters is minimal, it is demonstrated that the direct approach consumes less energy than the multi-hop method. Multi-hop networking methods utilize less electricity than direct forwarding methods when the number of clusters is larger than 140.

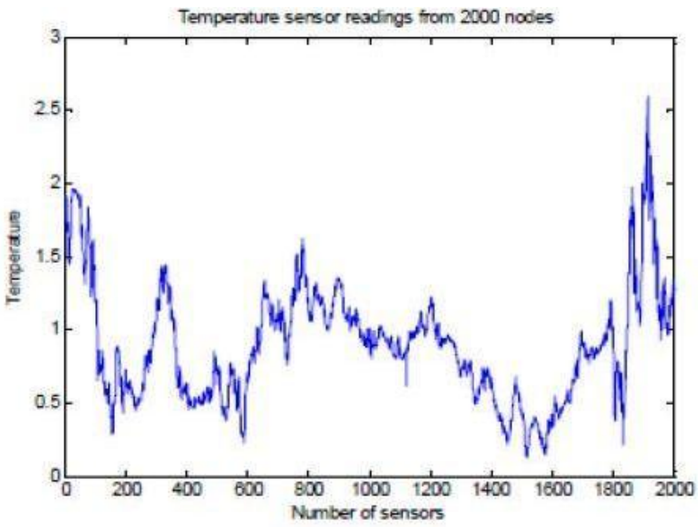

Fig. 9. Unsorted sensory readings collected from 2000 sensors

Figure 7 depicts sensor data that have not been sorted, and Figure 8 depicts their modified coefficients in the DCT sector. In the modified vector, all signal energy was maintained, but it is now concentrated in a limited number of big coefficients. If we just send these K big values to the BS, we will use a lot less power than if we send all of the data.

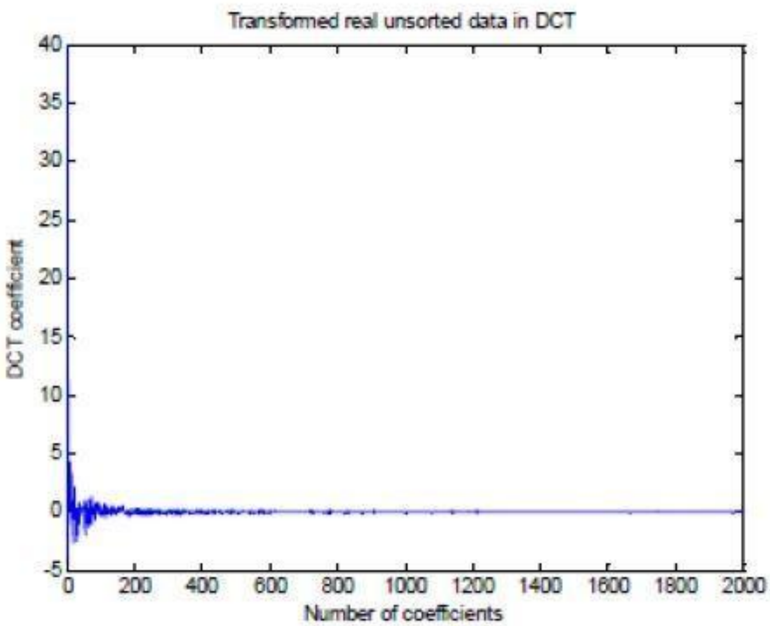

Fig. 10. DCT transformed coefficients from 2000 unsorted sensory readings

In Figure 9, the signals are sorted in increasing order. Figure 10 depicts the DCT converted coefficients. The big coefficients are focused on the lower-numbered factors in this example. Because the values of K are smaller in ordered signals than in unordered signals, the communication cost can be decreased. We opted to arrange sensor data at each cluster for this reason.

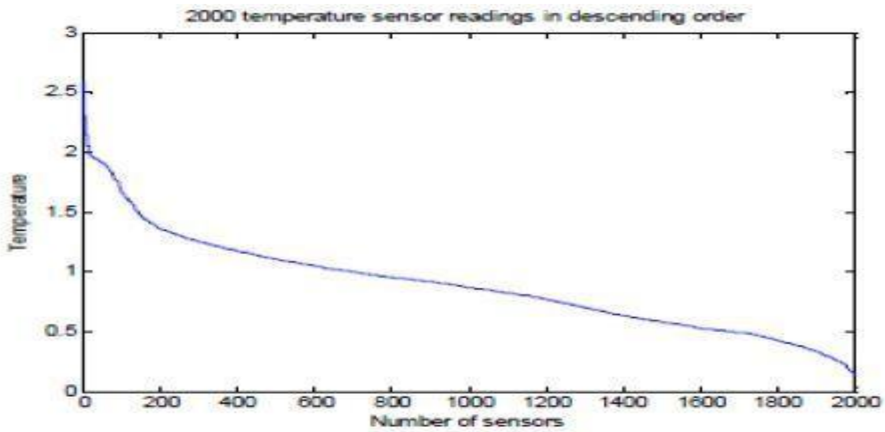

Fig. 11. Descending sorted sensory readings collected from 2000 sensors

Figure 11 illustrates that in various numbers of groups, both forms of sorted data produce the same error function. In ordered data, the big coefficients are concentrated in the smaller numeric coefficients.

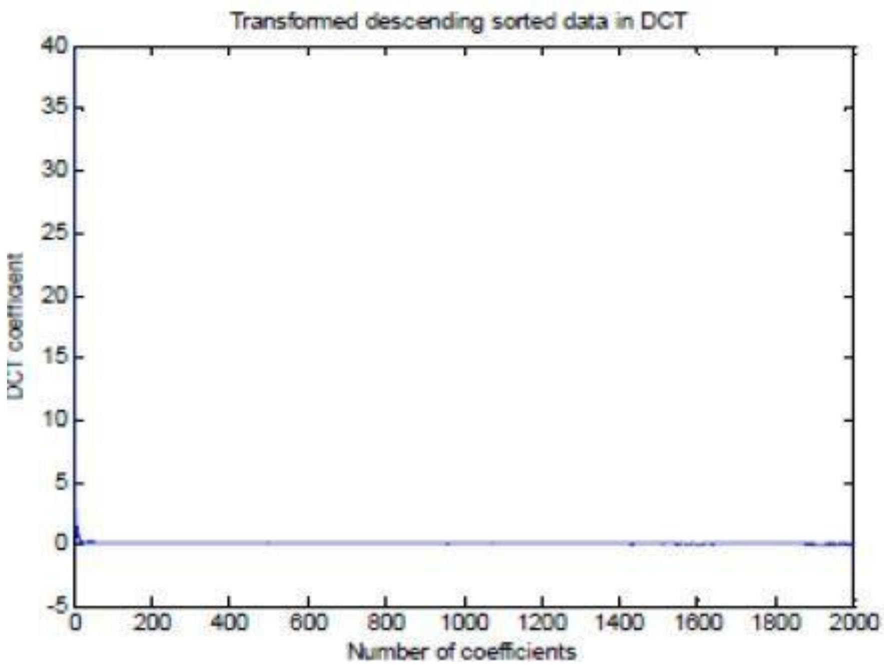

Fig. 12. DCT transformed coefficients from 2000 sorted readings

The ones from ordered signals convey greater signal energy than the ones from unordered signals when the same coefficients are delivered to the BS. That explains why sorted data has lower reconstruction mistakes. Figure 12 illustrates that raising the number of clusters or lowering the overall number of variables K submitted to the BS leads to higher reconstruction losses.

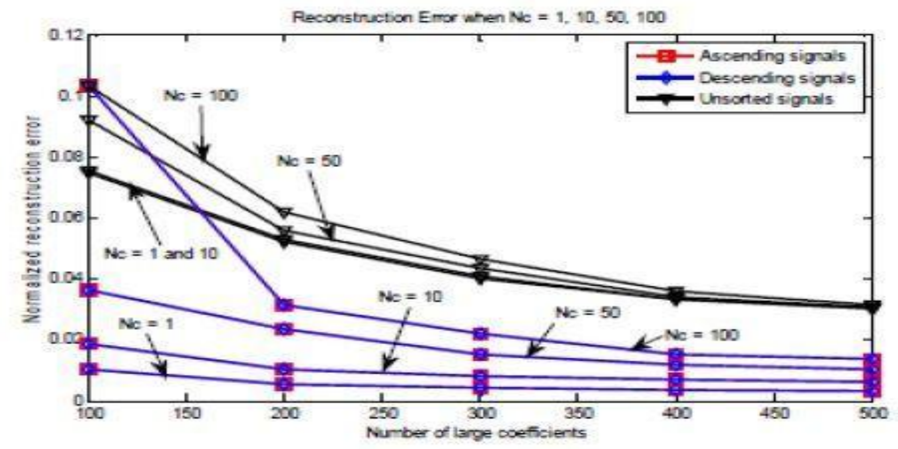

Fig. 13. Reconstruction error vs. the number of large coefficients with varying numbers of clusters,

As the number of clusters grows, we may correct the mistakes by sending more of the bigger DCT coefficients to the BS. DCT compression uses relatively minimal power in a noiseless situation since the network only delivers K big transformed coefficients (KN).

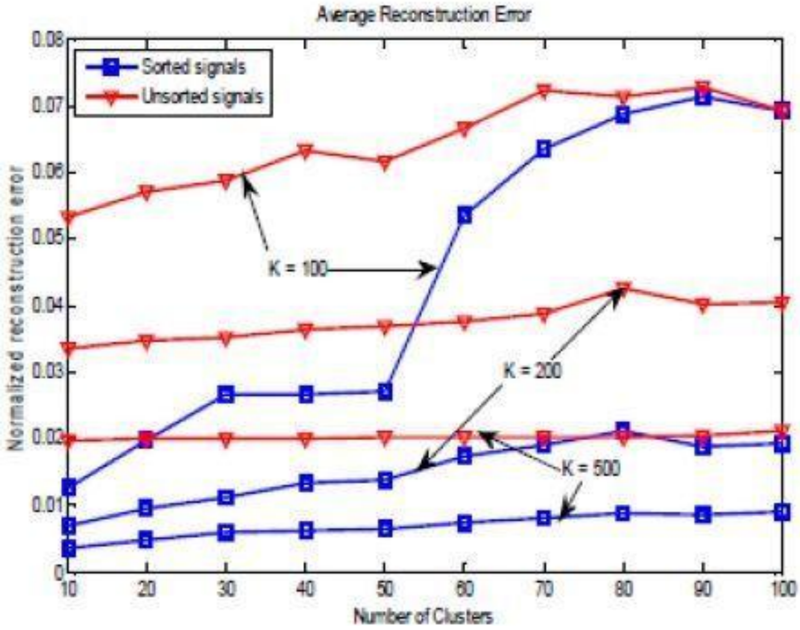

Fig. 14. Reconstruction error versus number of clusters with different numbers of the large coefficients (K)

In signal recovery procedures, K is often only approximately 10% as big as N to fulfill an error target, as illustrated in our simulation findings. Noise is an issue in real-world networks. As demonstrated in Figure 13, DCT compression degrades fast. As the overall number of big coefficients grows, the reconstruction errors grow as well.

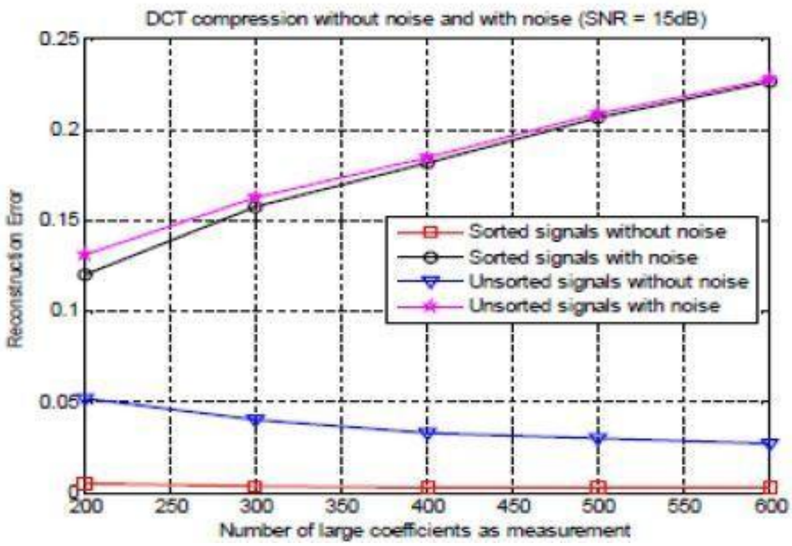

Fig: compression reconstruction error versus the number of large coefficients with noise and noiseless

4. Conclusions:

The origins, workings, and design of Hadoop are all described in this paper. This paper discusses the "3Vs," or Volume, Speed, and Diversity, as well as the issues of big data and technologies like Hadoop, which are designed to process enormous amounts of data via the map-reduce technique.

References

[1] A. Agrawal et al., “GCRP: A grid-cycle routing protocol for wireless sensor network with mobile sink,” AEU-Int. J. Electron. Commun., vol. 94, pp. 1–11, 2018.

[2] F. Ahmed et al., “Mobile sinks assisted geographic and opportunistic routing based interference avoidance for underwater wireless sensor network,” Sensors, vol. 18, no. 4, p. 1062, 2018.

[3] O. Aldabbas et al., “Unmanned ground vehicle for data collection in wireless sensor networks: mobility-aware sink selection,” Open Autom. Control Syst. J., vol. 8, no. 1, 2016.

[4] M. H. Anisi, A. H. Abdullah, and S. A. Razak, “Energy-efficient data collection in wireless sensor networks,” Wireless Sensor Netw., vol. 3, no. 10, pp. 329–335, 2011.

[5] A. M. Brisha, “Classifying Sensors Depending on their IDs to Reduce Power Consumption in Wireless Sensor Networks,” Int. J. Online Eng., vol. 6, no. 2, 2010.

[6] S. Chatterjea and P. Havinga, “A dynamic data aggregation scheme for wireless sensor networks,” in Proc. ProRISC 2003, 14th Workshop on Circuits, Systems, and Signal Processing, 2003.

[7] D. Dhanasekar, “Efficient Energy and Data Collection in Wireless Sensor Networks Using Leach Based Clustering,” unpublished.

[8] H. Dubois-Ferriere, D. Estrin, and M. Vetterli, “Packet combining in sensor networks,” in Proc. 3rd Int. Conf. Embedded Networked Sensor Syst., pp. 102–115, 2005.

[9] M.-F. Horng et al., “An extensible particles swarm optimization for energy-effective cluster management of underwater sensor networks,” in Proc. Int. Conf. Comput. Collective Intell., Berlin, Germany: Springer, pp. 442–452, 2010.

[10] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed diffusion: A scalable and robust communication paradigm for sensor networks,” in Proc. 6th Annu. Int. Conf. Mobile Comput. Netw., pp. 56–67, 2000.

[11] K. Mekkaoui and R. Abdellatif, “Optimal Hop Lengths to Ensure Minimum Energy Consumption in Wireless Sensor Networks,” Int. J. Technol. Diffus., vol. 9, no. 4, pp. 1–18, 2018.[12] M. Krishnan, S. Yun, and Y. M. Jung, “Improved clustering with firefly-

optimization-based mobile data collector for wireless sensor networks," AEU-Int. J. Electron. Commun., vol. 97, pp. 242–251, 2018.

[13] S. Madden, J. Hellerstein, and W. Hong, "In-network query processing in TinyOS," ACM Trans. Database Syst., vol. 30, pp. 122–173, 2003.

[14] S. Madden et al., "TAG: A tiny aggregation service for ad-hoc sensor networks," ACM SIGOPS Oper. Syst. Rev., vol. 36, SI, pp. 131–146, 2002.

[15] A. Rady et al., "Energy-efficient routing protocol based on sink mobility for wireless sensor networks," IET Wireless Sensor Syst., vol. 9, no. 6, pp. 405–415, 2019.

[16] K. Sundus and I. Almomani, "Mobility Effect on the Authenticity of Wireless Sensor Networks," in Proc. IEEE Jordan Int. Joint Conf. Electr. Eng. Inf. Technol. (JEEIT), 2019.

[17] J. Wang et al., "Energy-efficient routing algorithm with mobile sink support for wireless sensor networks," Sensors, vol. 19, no. 7, p. 1494, 2019.

[18] Y. Wu et al., "Energy-efficient wake-up scheduling for data collection and aggregation," IEEE Trans. Parallel Distrib. Syst., vol. 21, no. 2, pp. 275–287, 2009.

[19] Y. Yang and Y. Miao, "A path planning method for the mobile sink in farmland wireless sensor network," in Proc. IEEE 2nd Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC), pp. 838–842, 2017.

[20] Y. Yao and J. Gehrke, "Query Processing in Sensor Networks," in Proc. Conf. Innov. Data Syst. Res. (CIDR), 2003.