International Journal of Engineering & Technology, 7 (4) (2018) 3583-3590

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.19280
Research paper

Framework for assessing data quality resembles factor in large-scale wireless sensor network

B. Prathiba [1], Dr. K. Jaya Sankar [2], Dr. V. Sumalatha [1]

[1] *Jawaharlal Nehru Technological University, Anantapur, India*
[2] *Vasavi College of Engineering, Hyderabad, India*
**Corresponding author E-mail: balireddyprathibha@gmail.com*

Abstract

The usage of sensors has undergone a significant revolution right from the wireless sensor network (WSN) to Internet-of-Things (IoT). Existing research towards clustering protocols shows that their inclination is more on solving common issues and not more on data redundancy that should be controlled to enhance data quality. The closer relationship between the redundancy and data quality in WSN is less investigated for a practical solution. Therefore, we present a framework called as DQRF, i.e., Data Quality Resembles Factor which is exclusively meant for carrying out clustering operation for increasing data quality. The framework is supported by four sequential algorithms, which are designed to find the node that carries significant redundant information. The idea is to capture global details of all the nodes in the short run of clustering to identify and eliminate all possible errors. The proposed system offers approximately 93% of enhancement in data in contrast to the existing system.

Keywords: *Clustering; Data Quality; Energy; Error Minimization; Redundancy; Wireless Sensor Network.*

1. Introduction

The usage and commercial adoption of Wireless Sensor Network (WSN) is more than two decades old, where it is found increasingly used in habitat monitoring system, forest fire detection, industrial monitoring system and healthcare etc., Fahmy et al. [1] & Emary et al. [2]. With the presence of infrastructure, the sensor nodes capture environmental information and forward this information to the base station. This process is called as data aggregation Lu et al. [3]. At present, various studies give a quick overview of different data aggregation procedures Ranjan et al. [4] & Pantazis et al. [5]. The research work carried out till date essentially focuses on solving energy problems [5], traffic management issues, scheduling and provisioning Nithya et al. [6] and security Modirkhazeni et al. [7], etc. In addition to this, there have been various studies in WSN that has only emphasized on incorporating optimization techniques to improve communication performance among the sensor nodes Fei et al. [8].

The pace of such research work is quite good even today, as there are many problems in WSN that are quite unsolved even with the presence of a series of dedicated research contributions. A closer look in all the 99% methodologies adopted in research contribution is only limited to communication from member node to cluster head that completes at base station Wang et al. [9], Sivagami et al. [10], Rajeswari et al. [11], Sasirekha et al. [12], Nayak et al. [13], Liu et al. [14] and Zhou et al. [15]. Little information is reported in the research work about what happens to the aggregated data when it finally arrives in the base station. Existing studies of analytical base operation used in Internet-of-Things (IoT) explains, that there is a massive generation of voluminous sensory data that is subjected to mining process to extract specific knowledge Kaiwartya et al. [16], Sheng et al. [17], Abidoye et al. [18], Zhu et al. [19] and Bijarbooneh et al. [20]. Interestingly, such sensory data suffers from various problems of heterogeneity, veracity, and velocity if a heterogeneous sensor application is considered. Another significant problem is that the data quality is highly unstructured. As a result it is quite hard to subject it to any specific storage model. Although there are various causes and reasons of such advance problems, one simple cause of all of the issues is the presence of redundant data in the base station.

The initiation of this problem begins right from the clustering process. In existing system of clustering Firdaus et al. [21], Subha et al. [22], Kumari et al. [23], Tiwari et al. [24] and Krishnakumar et al. [25], the algorithms are more inclined on grouping the nodes on the basis of distance from the base station to generate clusters followed by selection of cluster head. The existing research work towards such problems is successful in solving the problem of cluster head selection but is unable to solve the ongoing issues of data redundancy. It is a computational challenge if the assumed network is multi-hop in WSN which calls for a vicious loop of internal communication among the nodes. Hence, at present, we do not find much work where clustering operation is developed to address the issue of data quality. An absence of data quality has various adverse effects on communication.

The first harmful effect is non-reliable data that generates false positive. Such data is non-tolerable for some applications, e.g., healthcare, weather broadcasting, etc., where precision is the sole important factor in data captured from the sensory application. There can be a definition of data quality, where it is mainly related to how unique it could be. Data quality will also mean that each sensor node should be able to forward unique information during the process of data aggregation as well as data fusion process. However, it is very likely that same event could happen in multiple places and information about the same event is captured by multiple sensors. Although such information may be unique for local cluster, it may pose a higher degree of redundancy in the more significant number of clusters. As an aggregated data are

time-stamped, so even same data could be time-stamped in a different way, and they successfully bypass the redundancy check within a base station. This phenomenon also adversely affect when the analysis is carried out in this aggregated data. Hence, data redundancy directly affects data quality and is worth investigating this problem to ensure that data aggregation smoothly takes place and also saves some energy while forwarding the data.

Although our work is limited only till base station, knowing what happens to the data in the base station may give a better realization of some potential protocol to offer better quality of data. Therefore, we introduce a novel clustering mechanism that increases the data quality to a higher level using an analytical modeling approach. We introduced a series of algorithms that is responsible for exploring the point of presence of redundant data in the data domain so that it can be reduced. More exploration of redundancy will lead to more reduction in error. Section 2, discusses the existing literature followed by a discussion of research problems in Section 3, proposed solution in 4. Section 5, discusses algorithm implementation followed by a discussion of result analysis in Section 6. Finally, the conclusive remarks are provided in Section 7.

2. Review of literature

This section briefs various existing research techniques towards data quality problems associated with WSN. Our prior study Prathiba et al. [26] has briefed certain approaches while this section further updates more. Xenekis et al. [27] have presented a weighted clustering approach by introducing a cost-based factor on the selection of cluster head in WSN as an enhancement from LEACH. Study towards uneven clustering technique was proven to be solving the energy problem associated with the sensor nodes especially in the case of heterogeneous networks Zhang et al. [28]. Omari and Fateh et al. [29] have implemented ant colony optimization for enhancing the performance of a hierarchical clustering technique in sensory application. Energy-efficient clustering approaches are also proven to be effective towards multi-hop network formation Liu et al. [30]. Katiyar et al. [31] have presented a clustering technique especially focusing on heterogeneous WSN. The work was focused entirely on incorporating energy efficiency in WSN using multilevel clustering approach. There are also literatures reported to use weighted clustering mechanism for solving energy efficiency issues. The works carried out by Zhang et al. [32] have jointly used hierarchical approach integrated with weighted method to carry out clustering operation. The work carried out by Ambekari and Sirsikar [33] have presented a study where different forms of the clustering techniques have been evaluated to address the energy efficiency problems in WSN.

The mechanism of selection of cluster head was advocated to be one of the most effective approaches for clustering as studied by Belabed et al. [34] and Bouallegue et al. [35]. This approach integrates the clustering methodology with the fountain code. Adoption of similar form of weighted methodology towards enhancing clustering operation was discussed by Kumrawat and Dhawan [36]. The work has discussed a unique optimization mechanism for improving the battery of node. The research work carried out by Ebadi et al. [37] had presented such mechanism where the selection of cluster-head is encouraged for enhancing network lifetime using degree of node and remnant energy. Li et al. [38] have recently implemented the concept of compressive sensing to perform routing operation during clustering in WSN. The complete work has focused on energy dissipation problems and uses sparse sensing in modeling purpose. The analysis of the clustering mechanism for designing a specific performance metric was carried out by Zeb et al. [39].

Inclusion of network concept over clustering method was also found to be an effective part of energy efficiency techniques Chidean et al. [40]. Literature has also contributed increasing number of work towards using weighted clustering mechanism. Discussion of varied clustering schemes was carried out by Tripathy and Chinara et al. [41]. Similar direction of problem addressing was carried out by Hong et al. [42] where the energy modeling is carried out by tree-based topologies. Usage of received signal strength indicator was reported to contribute for energy efficient clustering Wang et al. [43].There are also existing research works focusing on energy-problems, e.g., Aldawsari et al. [44], Baker et al. [45], Baker et al. [46], Baker et al. [47], and Baker et al. [48].

Irrespective of varied clustering schemes, majority of the existing mechanisms of clustering is only focused on solving energy dissipation issue among the sensor nodes in WSN. Few research works have emphasized on solving the data redundancy issues that may exist large-scale network. Consideration of dense network and its adverse effect on data packet is also few to find. More importantly, there is no reported standard clustering model to claim robust data quality in any form of the environmental condition that the node may encounter. This is a significant research gap, which calls for immediate attention of researchers. A thorough study of approaches in the existing system will also show that clustering is more of an iterative process where the similar process occurs for selecting the cluster head on the basis of few parameters. There is a good possibility where multiple set of information, as well as conditions, can be utilized for ensuring that each clustering per stages of data aggregation offers a significant improvement on energy efficiency as well as data quality. The next section outlines the identified problems from the literature.

3. Problem identification

At present, various research works are being carried out towards improving the clustering-based approaches in WSN that also focuses on energy efficiency. However, various problems are found unaddressed. The first significant problem is that none of the existing clustering approaches are found to emphasize on addressing data redundancy issues in the large-scale and dense configuration of WSN. There is a closer relationship between data redundancy as well as energy efficiency, which is less found to be investigated in the existing literature.

Another significant problem found in the existing system is that it adopts highly recursive methods in existing clustering mechanism leading to good energy efficiency but at the cost of redundant data. It is also found that the possibilities of replicated data in multiple clusters in existing clustering are never controlled in post clustering stage. It is only limited to set up the stage. Finally, the exploration of redundant data is not much emphasized from the adjacent node viewpoint, but it is carried out only during a routing process where it does not encapsulate global redundancies.

Therefore, the statement of research problem could be cited as "It is a highly difficult task to design an approach that retains a robust balance between the clustering performances as well as leverage the extent of data quality in large-scale and highly dense WSN."

4. Proposed methodology

This work is an extended version of our prior work Prathiba et al. [49].The core goal of the proposed study is to introduce an analytical model for enhancing the data quality of a sensor node using a novel clustering model. The architecture of the proposed system is shown in figure 2, and this section illustrates it further. The complete work is formed by the situation where the data of a sensor node are quite similar to the data of its adjacent sensors. In such a condition, such adjacent sensor could be represented by that sensor node with respect to the information domain. Such form of the sensor node is termed as Target Node (TN). It can also be mathematically represented. Consider that a sensor node δ (δ_1, δ_2, …..,δ_n) has n-number of adjacent sensor nodes. The information I is retained within the node δ. Therefore, I_1, I_2 ….I_n will represent information of the adjacent sensors. It is also considered that the corresponding distance of all the information (I_1, I_2, ….I_n) to I is less than a certain threshold value and N such that the range of N (number of adjacent node information) is (valued threshold, information threshold). It will mean that if the value of N is consid-

ered higher, then it will offer better chance to the sensor node δ to represent its adjacent sensors whose information retained comes within the range of threshold of information I. It will also mean that the sensor node δ is found to offer presence of higher spatial redundancy between all the adjacent nodes and target nodes. The proposed model focuses on identifying the node that bears redundant information in order to mitigate redundancy in the large network with high node density.

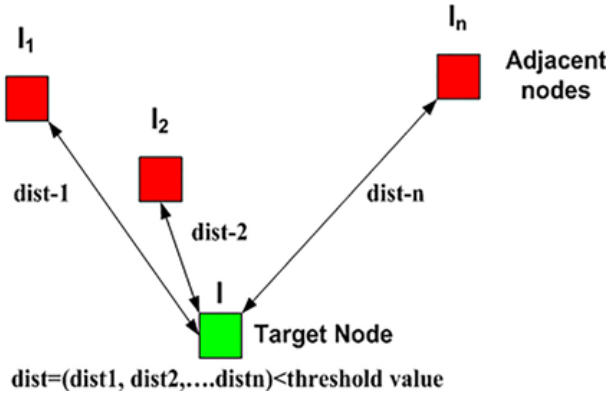

Fig. 1: Pictorial Illustration of Data Distribution.

Consider a node δ has adjacent nodes of n that fall under same transmission region of each other with node δ. The variable I is considered as information of node δ and I_1, I_2 … I_n our information of its adjacent nodes. It is also considered that there exists N number of information which is spatially located within a threshold to I which falls within a limit lower threshold and an upper threshold of n. Therefore, Data Resembles Factor of the proposed system can be mathematically expressed as,

$$DRF = \begin{Bmatrix} o & N < L \\ \Delta & otherwise \end{Bmatrix} \quad (1)$$

In the above expression, L represents the minimal threshold, and Δ represents a function that computes spatial distance considering different weights associated with each distance to represent redundancy score.

$$\Delta \rightarrow f\ (N, L, (h1, h2, h3), d, dm) \quad (2)$$

In the above expression, the variable d represents the spatial distance between information I and all the information retained by adjacent nodes, d_m is mean spatial distance between information of N number and Information I, h_1, h_2, and h_3 are associated weights such that summation of it is a unity (in adherence to probability theory). From the above empirical relationship, it is evident that DRF increases with an increase in the total number of adjacent nodes. Even if the value of d and d_m decrease, DRF will always be increasing. The architecture of the proposed system is as follows:

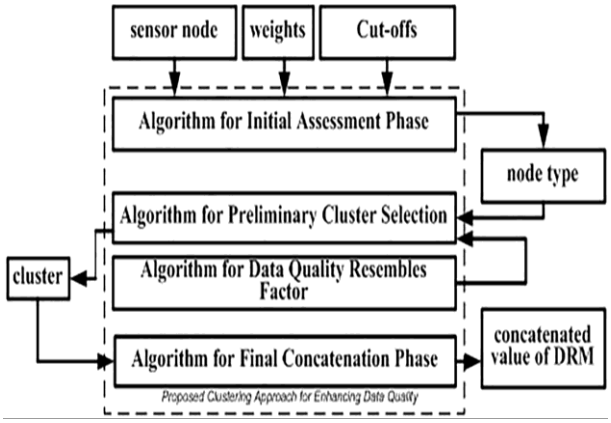

Fig. 2: Architecture of Proposed Study.

The architecture above shows the presence of three different algorithms in order to perform clustering. The main idea is to select a special type of node called as Target Node (TN) which bears a higher resembling factor of a set of its adjacent nodes' data. Therefore, finding one TN finds many feasible nodes that bear redundant information in the data domain. The proposed study also implements an empirical expression for data resembles factor in order to assess the level of redundant information. Simple analytical modeling is carried out by establishing an association between data resembles factor with adjacent nodes, data, distance, etc. For better effectiveness, we introduce two different forms of cut-off value to control data and the number of the nodes. Finally, the algorithm presents a Data Quality Resembles Factor (DQRF) that is responsible for computing the spatial resemble factor of all the aggregated data. The study considers that nodes residing on same clusters could have higher data resembles factor value as compared to that of different clusters. The complete modeling is done using three different forms of nodes, i.e., i) Delegatory Node (DN), ii) Quarantined Node (QN), and iii) Member Node (MN). Both DN and QN extract the raw sensory data and forward to the base station while MN only forwards the fused data aggregated by DN or QN. The algorithm performs a sequential operation where the first algorithm yields the outcome of the node type that acts as an input to the second algorithm. This result in the generation of cluster followed by the final concatenated value of data resembles factor. In the entire process, there is a generation of a lot of sub-clusters that are finally concatenated in order to generate supreme clusters. Therefore, the process results in the highly minimal amount of error formation during the transmission in WSN and thereby it successfully controls the error level as low as possible. The next section elaborates about the algorithm that has been implemented for this purpose.

5. Algorithm implementation

The prime functionality of all the three algorithms is to carry out a useful and novel clustering process to enhance the data quality in WSN. The implementation of the proposed system is carried out using 4 sequential algorithms viz. Algorithm for Initial Assessment Phase (IAP), Algorithm for Preliminary Cluster Selection (PCS), Algorithm for Final Concatenation Phase (FCP), and Algorithm for Data Quality Resembles Factor (DQRF).

5.1. Algorithm for initial assessment phase (IAP)

The core idea of this algorithm implementation is to construct logic for assisting exploration of a Target Node (TN) as well as Non-Target Node (NTN). Identification of TN significantly helps the network to classify all the robust node required to be involved in the clustering mechanism as well as it also assists in making a decision of optimizing the utilization of NTN. The proposed implementation also assumes that identification of TN node is the primary step to be carried out without which the classification remains partial. The complete classification of TN from NTN assists in offering potential information about the degree of data quality retained within these nodes. The steps of this algorithm are as follow:

i) Algorithm for Initial Assessment Phase

Input: n (sensor node), A (area of simulation), h1/h2/h3 (weights), α, β

Output: n_{type} (node type)

Start

1) For i=1:n
2) For j=1: A
3) If (τ≥β)
4) Flag n_{type} → TN
5) Else
6) Flag n_{type} → NTN
7) End

8) Compute γ
9) End
End

Referring figure 3, consider that a source node no is interested to set up its communication with its neighboring nodes n1 and n2. The node no sends a request beacon to both neighboring nodes that further responds back to the node no. According to the proposed system, the response is mainly data resemble factor, node type, and identity of the nodes. The algorithm obtains this information from all the neighboring nodes and then performs a conditional logic (Line-3 of above algorithm) to confirm if the neighboring node is TN or NTN. This process occurs quite faster in the form of response and is considered only for immediate neighboring nodes only.

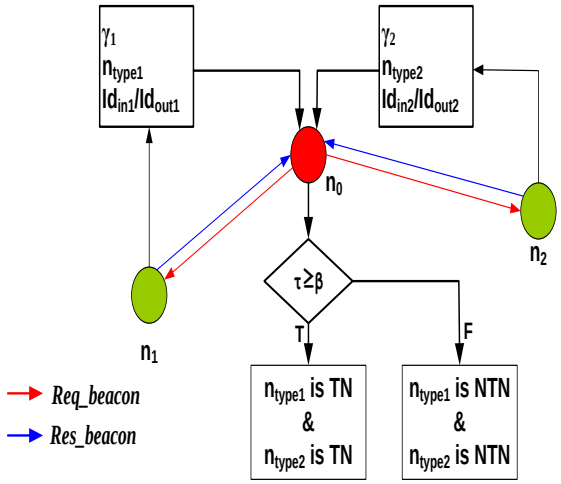

Fig. 3: Internal Processing of Algorithm-1.

The primary responsibility of IAP algorithm is to make an assessment within it to find if it belongs to Target Node (TN) or Non-Target Node (NTN) followed by calculation of data resembles factor γ. Based on this input, this algorithm performs identification of targeted node and data resembles factor value using a mathematical expression (1), non-target nodes are adjacent nodes, it is termed for better classification of algorithm modelling. The study considers three different weights h1, h2, and h3 as they are associated with effective spatial redundancyd1, and dm. There are only three spatial factors considered in the complete modelling of redundancy based on the data points; hence, three weights are considered. Once the node n_i executes IAP algorithm, then different information of node type n_{type}, identity of sensor nodes ID_{in} and ID_{out}, as well as data resembles factor γ is stored within it. The complete decision of the proposed algorithm for flagging the adjacent nodes to be TN/NTN is based on this.

If the adjacent node is found to be TN then this node is considered to be participating in the data aggregation process and is updated in α adjacent node. The variable α represents total number of adjacent nodes that are participating in the data aggregation process. The value of α can be adjusted based on the density requirement of the sensory application. Similarly, ID_{out} captures the other nodes which do not come under the range of α. Computation of data resembles factorγ is shown in the last algorithm. The matrix ID_{in} and ID_{out} are considered to be empty if the adjacent node is found to be NTN. The threshold value α represents threshold-value which is selected under ideal probability logic while β represents the lower value of the threshold data points. Hence, β should be initialized by minimum [2] as standard data points to represent the presence of at least two adjacent nodes to generate redundancy. This algorithm acts as an initial clustering process in order to perform an efficient decision towards identifying TN as well as NTN that finally results in the better form of a node yielding better data quality. The next algorithm initiates clustering as the preliminary stages in WSN.

5.2. Algorithm for preliminary cluster selection (PCS)

This algorithm offers its first level of confirmation for the formation of the clustering process. Figure 4 pictorially illustrates the clustering and internal process of an algorithm. In the initial process, an algorithm selects the already clustered nodes based on its type from IAP algorithm, and constructs a memory by sensor no, where only the adjacent sensor information is retained with respect to their identities. It further offers a confirmation link by checking of the adjacent sensors (n_3, n_5, and n_4) then forwards positive unit information (Figure4 (a)). How, it aborts connectivity if the neighboring sensor forwards negative unit information (say n4 in Figure4 (b)). Therefore, n4 will be termed as Quarantined Node (QN) in this process.

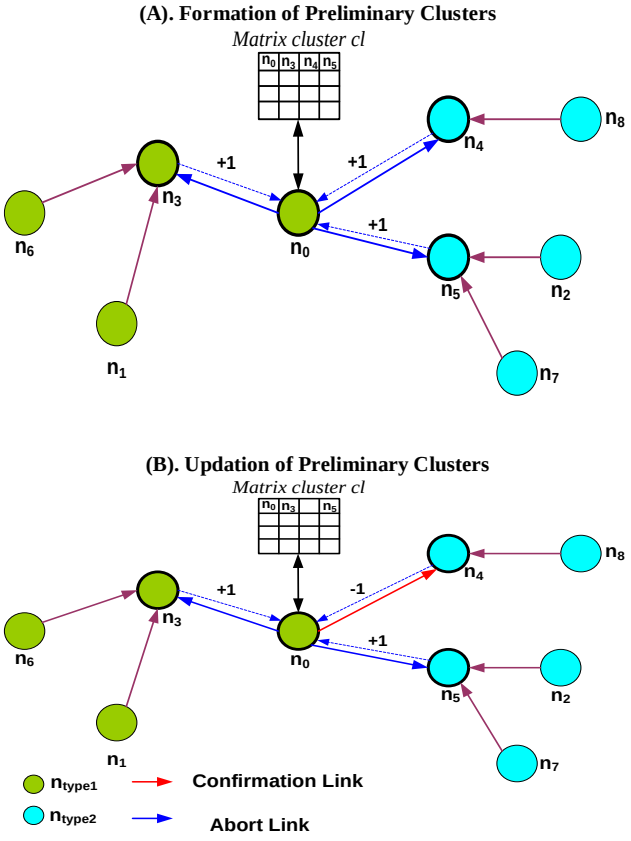

Fig. 4: Internal Processing of Algorithm-2.

ii) Algorithm for Preliminary Cluster Selection

Input: ntype (type of sensor node), ID_{in} / ID_{out} (inter and intra-cluster nodes), γ (DRF)
Output: cl (cluster), ρ (DQRF)
Start
1) For i=1: n
2) If ntype=TN
3) For each j=1: IDin
4) For each j=1: IDout
5) [ni→packet→nj]-1& [ni→packet→nj]+1
6) EndIf
7) If ntype=NTN
8) [ni→packet→nj]0
9) For each j=1: IDin
10) Ifγ(j)> γ(i)
11) ρ= γ(j)
12) End
13) If$_{ni}$(n_{type})=TN
14) Cl(i)→IDin(i)
15) End
16) End
End

All the local clusters that are formed from an Algorithm for Initial Assessment Phase are constructed considering all the sensors n.

The PCS algorithm is responsible for formulating the sub-clusters from the outcome of the prior algorithm, i.e., type of sensor node. The algorithm initially checks for the identified form of the node (TN/NTN) as well as it also checks its identity from ID_{in} and ID_{out} matrix. This algorithm considers two different types of nodes, i.e., the nodes that fall within a cluster IDin and nodes that fall outside of the cluster ID_{out}. This process is formed in order to pick the data that will be required to be transmitted to the respective adjacent nodes. The algorithm then assesses the specific node that resides in the same sub-cluster depending on the information it has received from its adjacent nodes. If n_i is found to be TN, then the node that resides within ID_{in} will be considered as the member node of sub-cluster i. This process is followed by the inclusion of identity information in the form of ID_{in} and cl. Once the node receives positive unit information (+1) from its adjacent node, then the sensor node and adjacent node is considered to reside within the same cluster. The matrix cl stores the identity of the neighbor node. However, upon receiving negative unit information (-1), the node will be considered as Quarantined Node (QN). This step is followed by a comparison of the received information of ρ (this information is extracted from the nodes that reside within matrix ID_{in}) with its value of ρ. Finally, the highest value of ρ, as well as the individual value of it, is stored in the sensor. The outcome is, therefore, the formation of cluster cl that is maintained by all the sensors and is considered to be updated during any communication cycle.

iii) Algorithm for Final Concatenation Phase

After the local clusters have been formulated in Algorithm for Preliminary Cluster Selection, there will be a generation of various numbers of such processes that will pose challenges in the analysis. Hence, all these local clusters are required to be concatenated. At the same time, it is also known that data resembles factor value is reposited on each sensor node that has to be used to perform this concatenation operation. The steps of the algorithm areas follow:

Algorithm for Final Concatenation Phase
Input: cl (cluster)
Output: γMAX (concatenated value of DRM)
Start
1) For i=1: n
2) For j=1: cl
3) If ρ (i).nID≠i
4) ni→ρ (i)→nj
5) End
6) End
7) $\gamma_{MAX} \rightarrow \arg_{max}(\rho_{max}.\gamma|cl)$, nID_max→ (j| $\rho_{max}.\ \gamma = \gamma_{max}$)
End

The above algorithm performs integration of all the sub-clusters formed from previous algorithm steps on the basis of the highest value of ρ with a prime intention to select such a node (called as Delegatory Node DN) whose information maps with a set of its neighbor node information residing within this cluster. Therefore, all the sensor nodes that correspond with the neighboring nodes with a similar map of information are termed as DN. After the simulation round is over, the identity of the DN is stored in all nodes in matrix ρ_{max}. Hence, only the sensor node that is found with matched identity with that of stored ρ_{max} will be allowed to send the packet to the base station. One of the advantages of this algorithm is that, it uses the identity of the entire delegatory node that assists in the faster selection of the clusters with a better classification of more or less presence of redundancy. Hence, making a decision of redundancy identification as well as mitigation becomes quite easier. By opting for shorter as well as energy efficient routes, the sampled information is forwarded that is obtained by checking the similarity between id of sensor nodes with a maximum value of data resembles factor.

iv) Algorithm for Data Quality Resembles Factor

This is the final algorithm that is responsible for performing a selection of final cluster with respect to the inter as well as intra clustering process. The prime objective of this algorithm is also to carry out unique and computationally cost-effective clustering process. It also ensures that in every fresh rounds of data aggregation, unique node selection will be continued for assisting increased data quality. The steps of the algorithm are stated below.

Algorithm for Data Quality Resembles Factor
Input: n, r, α, β, S_a, I_{mat}, A, γ
Output: cl
Start
1) init n, r, α, β
2) Sa→rand(x,y)
3) Imat→ [I-0.5.arb(I)]/max(I)
4) For i=1:n
5) A=find(d≤r)
6) End
7) γ=[h_1, h_2, h_3, β, d_1, d_2]
8) For j=1: A
9) If |Δd|≤ α
10) set inner node
11) Else
12) set outer node
13) If γ ≠0
14) [$ID_{in}ID_{on}$][n_i→msg→n_j]
15) obtaincl→unique (ID_{in})
16) End
End

The explanation of the above algorithmic steps are as follow: The algorithm takes the input of α (cut-off value of data), n(number of nodes),β(minimum number of sensor node),r(communication radius), and S_a (Simulation Area). An algorithm gives the outcome of cl (cluster). A random deployment strategy is adopted in the proposed system (Line-2) where the study considers location of sink to be very far from majority of the sensors. A matrix of data I_{mat} is formed for structuring the information associated with the simulation zone (Line-3). Entire neighboring sensors are considered for the simulation (Line-5). The formation of a variable γ in Line-7 is constructed mathematically as,

$$\gamma = o \tag{3}$$

$$\gamma = h_1.\phi + h_2.\varphi + h_3.\delta \tag{4}$$

The construction of this variable γ is carried out on the basis of dependability attribute h (where h= {h_1, h_2, h_2}.The study implements probability concept for this purpose. The formation of other independent variables ϕ, φ, and δ is considered to be equivalent to empirical normalization over a threshold parameter α and β. Distance-based parameters d_1 and d_2, computed using Euclidean distance formula between two points of data and average distance between two data retained by neighboring sensor and new matrix, are also considered for construction of the variable γ. The algorithm then carries out evaluation of target Node TN and non-target node NTN where TN represent a sensor that retain information about the relevant data with respect to the data corresponding to adjacent nodes. The absolute value of the spatial distance i.e., Δ_d is consider to be lesser as compared to the threshold value of the α value of the neighboring sensor. It also assesses the condition where the nearest score of the similarity match with the consecutive threshold value β (Line-9). The study considers that all the sensors that are associated with this logical condition is represented as a member of new cluster called as inner sensor (Line-10). Otherwise they are rejected and it becomes outer sensors (Line-12). The outer sensors are then transformed to the inner sensor with an aid of similar grouping mechanism until it encapsulates all the sensors within the simulation area. According to the proposed condition, if the value of γ is found to be non-zero (Line-13) then an initialization of intra-cluster communication takes place by the sensors that is continued by the intra-cluster communication as seen in Line-14. A cluster cl is formulated (Line-15) by the proposed system. The next section discusses the result analysis.

6. Result analysis

This section discusses the results being obtained by implementing the algorithm discussed in the prior section. The analysis of the proposed study is considered by simulation of a possible scenario of data aggregation or data fusion. It is because while performing data aggregation, clustering is one essential operation which can encounter nodes with redundant data due to spatial factors of redundancy in data points. Hence, the study has considered the redundancy caused owing to data aggregation and fusion process. The assessment of the proposed implementation has been carried out by using error as a main performance parameter. A closer look into literature shows that there is an usage of various other forms of factors, e.g., trustworthiness, reliability, validity, confidence, authenticity, etc. However, all these parameters are not directly linked with errors and encapsulated information related to application viewpoint and not architecture wise. Therefore, the study considers error factor that offers simpler assessment factor with respect to architecture operation and not application wise. Hence, an error was emphasized in the proposed study. The proposed study considers the calculation of error as a difference between an absolute value of the distance of spatial data point d, and that of a maximum value of data resembles factor. Hence, an error represents a degree of data redundancy. While performing simulation, the proposed study does not consider any form of existing fault model or processing technique as there are no benchmarked works reported in the area of WSN associated with data quality. By using the error computation, the proposed system claims of offering unique data that increase the level of precision in the quality of data. Apart from this, as the proposed system also uses a clustering technique for redundancy management, therefore, the outcome obtained could be claimed to offer cumulatively precise value with faster response time.

This outcome always favors fault tolerance-based operation in WSN as it is compared with frequently existing weight-based clustering approach and hierarchical clustering approach in comparative performance analysis. As the proposed system uses three different forms of weights as well as its algorithm also favors energy efficiency; therefore, it has been compared with weight-based clustering approach as well as hierarchical-based clustering approach. It should be noted that hierarchical-based clustering algorithm offers significant energy efficiency. The complete logic has been designed on the normal 32-bit machine with Windows platform and programmed in Matlab environment. The simulation environment considers testing with 1000 nodes with a transmission radius of 5 meters. The threshold values of α and β are considered to be 0.31 and 2 respectively.

It can be seen that, the focus of an algorithm is towards accomplishing a higher quality of sensor data and hence an error-based parametric evaluation is carried out with respect to an increasing trials of simulation. The proposed study is also compared with an existing system of weight-based clustering approach and hierarchical clustering approach. The reason of adopting these as an existing system is, because they are found frequently used by researchers. The design and implementation of the algorithms are carried out in a repetitive manner in order to find the consistency in the outcome. As the deployment of the node is carried out in a random manner, therefore, it is quite evident that error performance will also undergo significant changes. We performed 50 test iterations to find that there are changes in performance of existing weighted-based clustering approach and hierarchical clustering approach, but there are no significant changes in the curve of the proposed clustering approach. Although, there are changes in the proposed system curve too the changes are negligible.

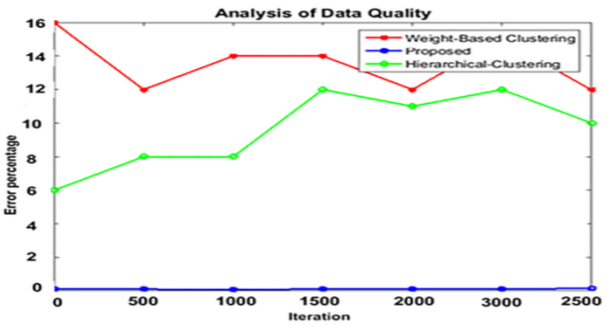

Fig. 5: Comparative Analysis of Error Percentage.

The outcome of the proposed study is shown in figure 5, where it is very clear about the performance outcome. Therefore, we show the magnified version of this outcome by removing the curves of an existing system and considering only the curve of a proposed system. The result indicates that the total variation of error percentage only ranges from 0.116% to 0.130%, which is hugely negligible. We also find that the curve of the proposed system shows good error minimization performance during the 500th iteration and 1000th iterations. After this, there is a little surge of error due to the inclusion of algorithmic steps to recompute the cluster outcomes in two steps. However, such fluctuation does not affect the communication performance in sensor application and instead offers better data quality.

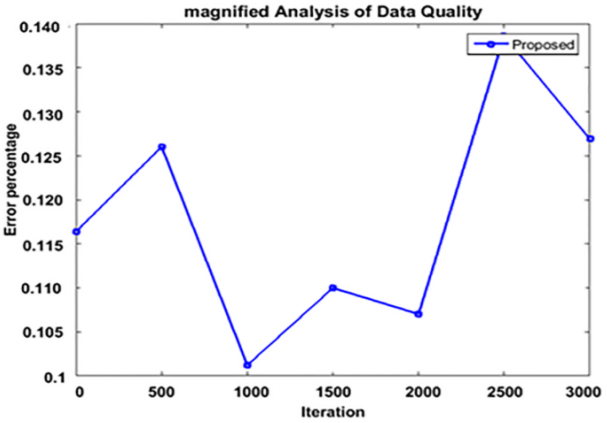

Fig. 6: Magnified Analysis of Proposed Error Percentage.

Referring to figure 6, it can be seen that there is a significant drop of error percentage by the proposed system while carrying out the data aggregation process in presence of redundancy. The beneficial point of weight-based method of clustering is that it offers overall utilization of resources that causes increasing dependencies on multiple set of parameters e.g., mobility, energy, distance, time, etc., and this operation is carried out in a highly iterative manner. Unfortunately, this mechanism suffers a significant pitfall of neglecting redundant data while only emphasizing on resource utilization. Hence, adoption of weight-based clustering method is not recommended for large scale implementation as it potentially lowers down data quality. On the other hand, it has been seen that better energy efficiency is observed for hierarchical-based clustering method. However, such method also overlooks the presence of redundant data that directly reduces the data quality. The reason for better performance of the proposed system to enhance data quality is that it induces the potential for considering fusing large number of data associated with redundancy. An efficient filtering process is carried out where the redundant data is filtered on the basis of the scores of the DRM value. Another reason for superior performance is the usage of probability theory which not only makes the system computation faster but also makes it quite precise and scalable. This technique also leads to generation of maximum number of unique data resulting in higher accuracy. Implemented over normal core-i7 processor; the proposed system consumes very less processing time of 0.6s which is quite faster and almost instantaneous as compared to existing system.

7. Conclusion

Conventional clustering technique in WSN has some good number of research contributions to ensure energy efficiency among the sensor nodes that tend to increase the network lifetime. However, the existing literature has no report of a clustering approach that offers a significantly lower scale of error after clustering is done. Hence, a defective clustering mechanism may lead to the generation of potentially unreliable data that will degrade the performance of the application at an exponentially faster pace. Therefore, the proposed framework offers a series of algorithms that ensure to find a node possessing more information about the redundant data. The contributions of study are viz. i) the proposed study computes data redundancies depending on the effective distance retained among the points of data, which is not only easy but also does not require any extra memory to compute DRF, ii) the algorithm offers higher precision in data content for which reason it can be used for sensor application that demands accuracy, e.g., healthcare, iii) the algorithm is also highly responsive and hence fits well for application, e.g., monitoring of natural calamities, accident event identification, healthcare (patient monitoring), etc. The study outcome is compared with the most frequently deployed clustering models to find that the proposed system offers lower error value showing higher data quality. After the data quality is improved; it is now capable of providing reliability in the communication process. The future research work direction will be to develop a novel algorithm that can ensure the reliability of sensory data processing on an upcoming reconfigurable network of the wireless sensor network. It can also be used for developing a new algorithm of data quality with the aid of a new mathematical model for ensuring energy efficiency and data consistency.

References

[1] H.M.A. Fahmy, Wireless Sensor Networks: Concepts, Applications, Experimentation and Analysis, Springer, 2016. https://doi.org/10.1007/978-981-10-0412-4.

[2] M. M. E. E. Emary, S. Ramakrishnan, Wireless Sensor Networks: From Theory to Applications.CRC Press, 799 (2013).

[3] Y. Lu, P. Kuonen, B. Hirsbrunner, & M. Lin, M. Benefits of data aggregation on energy consumption in wireless sensor networks, IET Communications, 11, 8, (2017) 1216-1223.

[4] R.K. Ranjan, & S.P. Karmore, Survey on secured data aggregation in wireless sensor network.International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, (2015), 1-4.

[5] N.A. Pantazis, S.A. Nikolidakis, & D.D. Vergados, Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey. IEEE Communications Surveys & Tutorials, 15, 2 (2013) 551-591 https://doi.org/10.1109/SURV.2012.062612.00084.

[6] R. Nithya, & N. Mahendran, A Survey: Duty cycle based routing and scheduling in wireless sensor networks, second International Conference on Electronics and Communication Systems (ICECS), Coimbatore, (2015), 813-817.

[7] A. Modirkhazeni, N. Ithnin, & O. Ibrahim, Secure Multipath Routing Protocols in Wireless Sensor Networks: A Security Survey Analysis. Second International Conference on Network Applications, Protocols and Services, Kedah, (2010), 228-233. https://doi.org/10.1109/NETAPPS.2010.48.

[8] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, & L. Hanzo, A Survey of Multi-Objective Optimization in Wireless Sensor Networks: Metrics, Algorithms, and Open Problems. IEEE Communications Surveys Tutorials, 19, 1, (2017) 550-586. https://doi.org/10.1109/COMST.2016.2610578.

[9] H. Wang, D. Xiong, L. Chen, & P. Wang, a Consensus-Based Time Synchronization Scheme with Low Overhead for Clustered Wireless Sensor Networks. IEEE Signal Processing Letters (2018).

[10] L. Sivagami, & J. M. L. Manickam, Cluster-Based MAC Protocol for Collision Avoidance and TDMA Scheduling in Underwater Wireless Sensor Networks. The Computer Journal, 59(10), (2016) 1527-1535. https://doi.org/10.1093/comjnl/bxw049.

[11] K. Rajeswari, & S. Neduncheliyan, Genetic algorithm based fault tolerant clustering in wireless sensor network. IET Communications, 11(12), (2017) 1927-1932. https://doi.org/10.1049/iet-com.2016.1074.

[12] S. Sasirekha, & S. Swamynathan, S,Cluster-chain mobile agent routing algorithm for efficient data aggregation in wireless sensor network. Journal of Communications and Networks, 19(4), (2017) 392-401. https://doi.org/10.1109/JCN.2017.000063.

[13] P. Nayak, & B. Vathasavai, Energy efficient clustering algorithm for multi-hop wireless sensor network using type-2 fuzzy logic. IEEE Sensors Journal, 17(14), (2017) 4492-4499. https://doi.org/10.1109/JSEN.2017.2711432.

[14] X. Liu, J. Li, Z. Dong & F. Xiong, Joint design of energy-efficient clustering and data recovery for wireless sensor networks, IEEE Access, 5, (2017) 3646-3656. https://doi.org/10.1109/ACCESS.2017.2660770.

[15] Y. Zhou, N. Wang & W. Xiang, Clustering hierarchy protocol in wireless sensor networks using an improved PSO algorithm, IEEE Access, 5, (2017) 2241-2253. https://doi.org/10.1109/ACCESS.2016.2633826.

[16] O. Kaiwartya, A.H. Abdullah, Y. Cao, J. Lloret, S. Kumar, R.R. Shah, & S. Prakash, Virtualization in wireless sensor networks: fault tolerant embedding for internet of things. IEEE Internet of Things Journal, 5(2), (2018) 571-580. https://doi.org/10.1109/JIOT.2017.2717704.

[17] Z. Sheng, H. Wang, C. Yin, X. Hu, S. Yang, & V.C. Leung, Lightweight management of resource-constrained sensor devices in internet of things IEEE internet of things journal, 2(5), (2015) 402-411.

[18] A.P. Abidoye, & I.C. Obagbuwa, Models for integrating wireless sensor networks into the Internet of Things, IET Wireless Sensor Systems, 7(3), (2017) 65-72.

[19] J. Zhu, Y. Song, D. Jiang, & H. Song Multi-armed bandit channel access scheme with cognitive radio technology in wireless sensor networks for the internet of things. IEEE access, 4, (2016) 4609-4617. https://doi.org/10.1109/ACCESS.2016.2600633.

[20] F.H. Bijarbooneh, W. Du, E.C. Ngai, X. Fu & J. Liu, Cloud-assisted data fusion and sensor selection for internet of things. IEEE Internet of Things Journal, 3(3), (2016) 257-268. https://doi.org/10.1109/JIOT.2015.2502182.

[21] T. Firdaus & M. Hasan, A survey on clustering algorithms for energy efficiency in wireless sensor network. In Computing for Sustainable Global Development (INDIACom), 3rd International Conference, (2016, March) 759-763.

[22] C.P. Subha, S. Malarkan, & K. Vaithinathan, A survey on energy efficient neural network based clustering models in wireless sensor networks. In Emerging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT), International Conference, (2013, January) 1-6.

[23] P. Kumari, M.P. Singh, & P. Kumar, Survey of clustering algorithms using fuzzy logic in wireless sensor network. In Energy Efficient Technologies for Sustainability (ICEETS), International Conference, (2013, April) 924-928.

[24] T. Tiwari, & N.R. Roy, Heirarchical clustering in heterogeneous wireless sensor networks: A survey. In Computing, Communication & Automation (ICCCA), International Conference (2015) 1385-1390. https://doi.org/10.1109/CCAA.2015.7148596.

[25] A. Krishnakumar, & V. Anuratha, Survey on energy efficient load-balanced clustering algorithm based on variable convergence time for wireless sensor networks. In Advanced Computing and Communication Systems (ICACCS), 3rd International Conference, 1, (2016, January) 1-5.

[26] B. Prathiba, K.J. Sankar, & V. Sumalatha, Enhancing the data quality in wireless sensor networks—a review. In Automatic Control and Dynamic Optimization Techniques (ICACDOT), International Conference, (2016, September) 448-454.

[27] A. Xenakis, F. Foukalas, & G. Stamoulis, Minimum weighted clustering algorithm for wireless sensor networks, In Proceedings of the 19th Panhellenic Conference on Informatics, (2015, October) 255-260. https://doi.org/10.1145/2801948.2801999.

[28] Y. Zhang, W. Xiong, D. Han, W. Chen, & J. Wang, Routing algorithm with uneven clustering for energy heterogeneous wireless sensor networks. Journal of Sensors, (2016). https://doi.org/10.1155/2016/7542907.

[29] M. Omari & W.H. Fateh, Hybrid Hierarchical Clustering Protocol in Wireless Sensor Networks based on Ant Colony Algorithm and MR-LEACH. In Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication, 72, (2015, November).

[30] Z. Liu, W. Xing, Y. Wang & D. Lu, Hierarchical spatial clustering in multihop wireless sensor networks. International Journal of Dis-

tributed Sensor Networks, 9, 11, (2013). https://doi.org/10.1155/2013/528980.
[31] V. Katiyar, N. Chand, & S. Soni, Efficient multilevel clustering for large-scale heterogeneous wireless sensor networks. In Proceedings of the International Conference on Communication, Computing & Security, (2011, February) 1-6.
[32] J. Zhang, J. Chen, Z. Xu, & Y. Liu, LEACH-WM: Weighted and intra-cluster multi-hop energy-efficient algorithm for wireless sensor networks. In Control Conference (CCC), 35th Chinese, (2016, July) 8325-8329
[33] J.S. Ambekari & S. Sirsikar, Comparative Study of Optimal Clustering Techniques in Wireless Sensor Network: A Survey, In Proceedings of the ACM Symposium on Women in Research, (2016, March) 38-44. https://doi.org/10.1145/2909067.2909074.
[34] F. Belabed & R. Bouallegue, Performance evaluation of the optimized weighted clustering algorithm in wireless sensor networks. In Advanced Information Networking and Applications Workshops (WAINA), 31st International Conference, (2017, March) 222-225 https://doi.org/10.1109/WAINA.2017.49.
[35] R. Bouallegue & Belabed, An optimized weight-based clustering algorithm in wireless sensor networks. In Wireless Communications and Mobile Computing Conference (IWCMC), International, (2016, September) 757-762.
[36] M. Kumrawat & M. Dhawan, Optimizing energy consumption in wireless sensor network through distributed weighted clustering algorithm, In Computer, Communication and Control (IC4), International Conference, (2015, September) 1-5.
[37] S. Ebadi, A Multihop Clustering Algorithm for Energy Saving in Wireless Sensor Networks, International Scholarly Research Network ISRN Sensor Networks (2012).
[38] X. Li, X. Tao & G. Mao, Unbalanced expander based compressive data gathering in clustered wireless sensor networks. IEEE Access, 5, (2017) 7553-7566. https://doi.org/10.1109/ACCESS.2017.2696745.
[39] A. Zeb, A.M. Islam, M. Zareei, I-A. Mamoon, N. Mansoor, S. Baharun, & S. Komaki, Clustering analysis in wireless sensor networks: the ambit of performance metrics and schemes taxonomy. International Journal of Distributed Sensor Networks, 12(7), (2016) 4979142. https://doi.org/10.1177/155014774979142.
[40] M.I. Chidean, E. Morgado, M. S-Junquera, J. R-Bargueño, J. Ramos, & A.J. Caamaño, Energy efficiency and quality of data reconstruction through data-coupled clustering for self-organized large-scale WSNs. IEEE sensors journal, 16(12), (2016) 5010-5020 https://doi.org/10.1109/JSEN.2016.2551466.
[41] A.K. Tripathy & S. Chinara, Comparison of residual energy-based clustering algorithms for wireless sensor network, ISRN Sensor Networks, (2012).
[42] Z. Hong, R. Wang, & X. Li, A clustering-tree topology control based on the energy forecast for heterogeneous wireless sensor networks. IEEE/CAA Journal of Automatica Sinica, 3(1), (2016) 68-77
[43] Y. Wang, I.G. Guardiola, & X. Wu, RSSI and LQI data clustering techniques to determine the number of nodes in wireless sensor networks, International Journal of distributed sensor networks, 10(5), (2014) 380-526. https://doi.org/10.1155/2014/380526.
[44] B. Aldawsari, T. Baker & D. England, Trusted energy efficient cloud-based services brokerage platform. Int. J. Intell. Comput. Res, 6, (2015) 630-639.
[45] T. Baker, M. Asim, H. Tawfik, B. Aldawsari & R. Buyya, An energy-aware service composition algorithm for multiple cloud-based IoT applications. Journal of Network and Computer Applications, 89, (2017) 96-108 https://doi.org/10.1016/j.jnca.2017.03.008.
[46] T. Baker, Y. Ngoko, R. T-Calasanz, O.F. Rana, & M. Randles, Energy efficient cloud computing environment via autonomic meta-director framework. In Developments in e-Systems Engineering (DeSE), Sixth International Conference, (2013) 198-203
[47] T. Baker, B. A-Dawsari, H. Tawfik, D. Reid & Y. Ngoko, GreeDi: An energy efficient routing algorithm for big data on cloud, Ad Hoc Networks, 35, (2015) 83-96. https://doi.org/10.1016/j.adhoc.2015.06.008.
[48] T. Baker, D. Lamb, A. T-Bendiab, & D. A-Jumeily, Facilitating Semantic Adaptation of Web Services at Runtime Using a Meta-Data Layer. In Developments in E-systems Engineering (DESE), (2010) 231-236
[49] B. Prathiba, K.J. Sankar & V. Sumalatha, A Novel Clustering Algorithm for Leveraging Data Quality in Wireless Sensor Network.In International Conference on Next Generation Computing Technologies Springer, Singapore (2017, October). 687-694.