❒   3822

Framework for cost-effective analytical modelling for sensory data over cloud environment

Manujakshi B.C.[1], K B Ramesh[2]
[1]Department of Computer Science and Engineering, Presidency University, India
[2]Department of Electronics & Instrumentation Engineering, RV College of Engineering, India

Article Info	ABSTRACT
	In order to offer sensory data as a service over the cloud, it is necessary to execute a cost-effective and yet precise data analytical logic within the sensing units. However, it is quite questionable as such forms of analytical operation are quite resource dependent which cannot be offered by the resource constraint sensory units. Therefore, the proposed paper introduces a novel approach of performing cost-effective data analytical method in order to extract knowledge from big data over the cloud. The proposed study uses a novel concept of the frequent pattern along with a tree-based approach in order to develop an analytical model for carrying out the mining operation in the large-scale sensor deployment over the cloud environment. Using a simulation-based approach over the mathematical model, the proposed model exhibit reduced mining duration, controlled energy dissipation, and highly optimized memory demands for all the resource constraint nodes.

Corresponding Author:

Manujakshi B.C.,
Assistant Professor, Department of Computer Science and Engineering,
Presidency University, Bengaluru, Karnataka, India.
Email: manujakshibc2014@gmail.com

1.   INTRODUCTION

The utilization of the sensing units has been witnessed among the commercial users from more than a decade in the form of Wireless Sensor Network (WSN) [1]. The conventional approach of WSN calls for performing two basic operations i.e. data fusion and data aggregation [2]. It is already known that sensor applications are normally deployed in the scenario which is hazardous or physically challenging for humans to reach. Hence, the robustness in the sensing performance really matters in WSN. There are various essential concepts that state a sensor network encounters various forms of challenges e.g. security problems [3], routing problems [4], traffic management problem [5], and energy problems [6]. Out of all these set of problems, energy is one of the most challenging problems in WSN to be encountered with. It is believed that energy factor is directly proportionally to the data forwarding performance in WSN [7]. Therefore, there has been various significant research works towards addressing the problem of energy constraint in WSN [8]. However, such solutions are not researched enough when WSN is integrated with various upcoming applications over a reconfigurable network. At present, sensors are considered an integral component of Internet-of-Things (IoT) where sensors are connected with multiple other forms of networking devices where the various medium of communication can be utilized [9]. Hence, the challenging problem in this aspect is that energy demands for the sensor nodes are very much different in WSN and in IoT [10].

At present, there are few research work towards the direction of solving energy problems in IoT even, but there is again a potential research gap and that gap is: existing system of IoT do not typically consider the standard sensor node configuration for which reason it is nearly vague to understand how to solve the energy problems. Another aspect to observe in this situation is that the data forwarding process is the prime reason for energy depletion. It is also known that IoT is meant for not only aggregating the data but

also performs a predictive operation on the sensory data. There are multiple advantages in this aspect viz. i) the precision as well as accuracy of prediction increases if the analytical operation is performed right after the data is fused within a sensor and before the fused data is forwarded to the sink, ii) the size of the analyzed data is always lesser than aggregated data which will consume less channel-based resources as well as lesser computational demands to process it. However, there is still an unsolved question evolved for this process which is how to design such analytics which can be executed over resource constraint sensor nodes. The memories of the sensor nodes are quite limited and so is the processing capability. In order to execute any analytical protocol over a sensor node, there is a need of sufficient memory as the majority of he predictive schemes in IoT are highly iterative and it depends upon the scale of the data that are required to be analytically processed to the datacenter or warehouses. With the evolution of big data, the complexity of the structuredness of the data is consistently kept on increasing. Although there are existing software frameworks for distributed storage, they are least capable of performing extraction of specific knowledge which leads to the emergence of a novel method of analytics.

Therefore, this paper introduces a novel and cost-effective architectural approach based on frequent patterns in order to perform an analytical operation over sensory data. The proposed system makes use of graph-based tree structure in order to construct the proposed data analytical concept along with novel usage of frequent pattern and a different variant of estimates.

This section is in continuation of the review of existing literature from our prior work [11]. Although, there are various studies already carried out towards applying sophisticated data analytics approaches over the sensory data collected in the IoT ecosystem. The correlation-based analysis is one of the simplest forms of shortlisting or grouping the similar forms of data and hence considered as one of the cost-effective mechanism of data analytics. The work carried out by Bertrand and Moonen [12] has presented a mechanism to compute the correlation factor on the distributed scale using a tree-based approach. The complete analysis of the data is carried out on the basis of a correlational factor also considering the temporal attribute along with it. However, such forms of data also have a higher complexity. The work carried out by Parwez et al. [13] where the complex data from mobile data has been considered for evaluating the anomaly pattern associated with the data. The authors have used unsupervised learning clustering scheme.

A similar direction of the work has also been carried out by Rahman et al. [14] using time-series analysis. Rehman et al. [15] have presented a discussion of big data analytics where the authors have presented the significance of concentric computing approach in the presented system. A similar direction of the research work has been carried out by Sun et al. [16] with respective to the discussion of analytical operations over the internet of things and big data analytics. The recent works of Wang et al. [17] have presented a discussion of analytical operation over energy-based data. Big data analytical approach was also discussed by Yue et al. [18] towards constructing the case of specific event detection over social networking application specific to geographical location.

Most recently, Ramos et al. [19] have presented a work where sensory data has been extracted from the interface of the mixed signal using non-linear quantization method. Sharma and Wang [20] have developed a framework that can perform mining operation on edge and cloud over its respective data. WSN, when integrated with IoT, offers massive processing of different forms of data. However, if the data bears higher size and complexity (e.g. multimedia data) than processing it along with structurization becomes the most complex task. This problem was addressed by Cao et al. [21] where a unique analytical operation has been carried out using a self-optimizing approach over multimedia contents. The optimization was carried out on the basis of context in order to ensure that it offers resource friendly operation applicable for a practical environmental situation. Based on the context, the model prototyping is carried out that is capable of fine-tuning the energy demands of the sensors in the IoT environment. However, the self-optimization principle doesn't offer cost-effective complex and unstructured data handling mechanism apart from video data. Such a problem was also investigated by other researchers where a system-on-chip is basically designed in order to support the complex processing of data and other heavy algorithms. The authors have also used a machine learning approach for carrying out feature extraction and training.apart from this fog computing was also found to be involved in performing an analytical operation in the existing system. A unique fog computing based modeling was carried out using Raspberry Pi in order to assess the performance of different data analytical model (He et al. [22]). The system assists in performing large scale of analysis work over multiple workstations in order to offer a greater deal of scalability. The study outcome was found to offer reducing job computation time with increasing complexity. It is also claimed to offer better service and resource management.

Applying analytic-based approach has also been carried out over sensitive data e.g. healthcare data. Unfortunately, such data grows so massively that retention of their privacy is an utmost concern. Applying normal security algorithms cannot be carried out owing to the complications associated with it. A light-weight analytical approach has been presented by Gong et al. [23] using a predictive approach.

The study has encoded privacy to the training data while used regression model for performing prediction over partitioned data. The study also showed that analytics can be highly helpful for leveraging the security standards too which is one of the troubleshooting problems over the distributed and dynamic ecosystem of the cloud environment. Further studies towards healthcare data are carried out by Li et al. [24] which has discussed the extraction of medical information within the ambulatory vehicles. The model developed offers consistent monitoring of the health factor of the driver in a vehicle. Real-time sensor-based data extraction of health statistics has been carried out in order to assess the event. Study towards predictive scheme has been also promoted by Yildirim et al. [25] considering the case study of wind turbines. The authors have presented an integrated framework that is responsible for performing maintenance over the data collected from wind farms. The paper has also presented a scheduling framework for assessing the performance of dynamic models.

Usage of open source in framing up data analytics was reported to be used for developing a monitoring module for real-time damage with respect to the structural health aspect. Liao et al. [26] have developed a model for data analytics that is capable of extracting intelligence from infrastructure along with power efficiency in its internal operation. The authors have used real-time sensors in order to perform the experiment where the outcome is assessed by compression and delay mainly. Munoz et al. [27] have presented a joint modeling of analytics using internet of thing, software-defined network over cloud environment. An experimental approach has been presented by Munoz et al. [27] which is mainly meant for controlling bottleneck condition within the traffic scenario. Deployment of the analytical operation was also shown to assists in solving classification problems (Otero et al. [28]). The study also presents a unique decision-making system for performing prediction over cloud environment. Analysis of the traffic-related data was also carried out by Shao et al. [29] offering solution towards bottleneck traffic situation. Studies towards advanced analytics have been carried out by Ivanov et al. [30] where precision farming is advocated for performing an analytical operation in WSN. Chandrakala and Rao [31] have presented migration of VM to enhance the security of the cloud computing. Rhioui and Oumnad [32] have presented IoTs survey for calculating human activities from all over. Sindhu et al. [33] have presented a new incorporated structure to make sure superior data quality in big data analytics on cloud surroundings. The paper has also presented a mechanism of sophisticated farm management and discussed various practical challenges associated with it. The authors have presented a prototyping approach for experimenting their concept. Therefore, there is various work that has been carried out towards developing analytics over the cloud. The next section discusses the problem that is identified from the review.

We have performed a rough investigation of the mining techniques implemented exclusively for the wireless sensor network. It was explored that the recent trend of knowledge extraction is more or less into the identification of any unique pattern from the complex data. However, with overcoming of the data complexity over cloud using SDaaS, the analysis could be done more effectively with lesser response time. However, the problem still resides as existing techniques call for observing only the frequent pattern from the occurrence of an event. The meaning of this occurrence is-that although sensor sense all the data but it only forwards the data which has significant information of occurrence of an event. It is done in order to avoid the communication overhead. The significant research gap is-till date the frequent pattern approaches are never accurate for exploring the potential pattern of data as it can only highlight the epochs present in the database that contains the frequent patterns. Such problems are very dangerous for healthcare application as well as an application that monitors the climatic condition, especially where sensors are used to monitor the health statistics. Finally, such incomplete knowledge gets accumulated in Hadoop that leads to unwanted cost and expenditure of cloud services to storage and perform error-prone analysis of data. It is explored that such problems lead to following issues:

−   Limitation of Cloud-based Control: It is not possible for a cloud to track all the sensor data from one point of the base station. Hence, if the tracking of data is done from multiple points of the base station that frequent patterns can be extracted but the significant pattern cannot be extracted if the large network with its dynamicity is not considered.

−   Incompatibility of existing mining: If sensors collaborate with cloud architecture that it is essential that knowledge discovery (or mining) of the data also to be done in same way. Unfortunately, the existing mining approaches were never testified for energy consumption. It is still unsolved. Moreover, the big question mark is non-applicability of conventional mining technique if the sensory data is massive and highly unstructured.

−   The split opinion of tree-based approach: Studies towards tree-based approaches are less and more criticized irrespective of its benefits. Till date tree-based approach was utilized only for routing, however, it can be also utilized for constructing the topology for data analytics too, which is not the much-researched topic.

-   More Inclined towards specific event processing; Majority of the existing research work is essentially focused on specific forms of event detection and doesn't consider the exponential higher level of challenges associated with the data. Hence, the existing structure of data cannot be directly subjected to present data analytical approach.

Therefore, the problem statement is "" To develop a simple framework for carrying out sophisticated analytical operation over a distributed cloud environment." In order to address this problem, it is necessary to evolve up with a definitive architecture that is discussed in the next section.

2.   SYSTEM ARCHITECTURE

The proposed study is a continuation of our prior implementation of [34] and [35]. This framework will be more focused towards extraction of knowledge discovery from the complex sensory data. Now that after implementation of our prior prototype SDaaS, a framework is presented that can actually generate a stream of sensor data and use HBase to perform proper management of structured data. The framework also has a user-centric analytical module in SDaaS. However, this part of the study will be the continuation of SDaaS and will incorporate some of the novel and potential mining features for a better degree of knowledge extraction.

This part of the study will consider adopting an analytical approach completely. Figure 1 shows the indicative scheme to be adopted. The prime objective will be to extract a significant pattern that elicits the hidden relationship of patterns among the data collected from sensors. The mathematical modeling will comprise various variables e.g. the number of sensors, time slots, formulations of frequent patterns, etc. We will also formulate a condition of an epoch for analyzing the frequent patterns. The outcome of SDaaS framework will be considered as an input for this part of the mining technique. Hence, it is just an extension of SDaaS framework for enhancing mining of knowledge from complex sensory data.
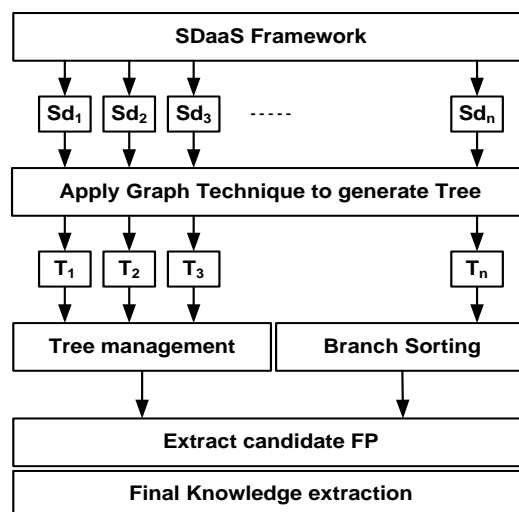

Figure 1 Adopted methodology

A novel graph theory has been developed in order to construct the tree for every data being generated by the SDaaS framework. The aim is to explore the relationship among the tree branches. The generated tree $(T_1, T_2, \ldots.. T_n)$ will be subjected to further two operations i.e. i) tree management and ii) branch sorting. Tree management is all about arranging the tree structure for facilitating traversals operations in the tree while branch sorting will pertain to sorting operation so that the generated tree could be easily subjected to the mining operation. A novel algorithm will be developed that can apply a pattern mining approach in order to extract all candidate frequent patterns from the entire tree. The advantage of this process will be that it presents a technique of sophisticated data mining technique for a very large area considering the complexity for both homogeneous and heterogeneous types of the network. We will also investigate the possibilities of incorporating distributed mining approaches as well as parallel mining approaches so that sensor data could be truly used as a cloud service directly. The discussion of adopted methodology is further supported by rationale and its respective contribution:

a. The rationale of the adopted methodology

It is quite evident there similar event-based information can be captured by different forms of sensing device which will impose a bigger challenge in processing the analytical operation as the data is absolutely not unique and is highly redundant. Such set of equivalent information among the sensing devices or set of such devices can be easily represented by frequent patterns which are highly essential for the real-world analysis of bigger data stream. Technically, such forms of frequent patterns extract a higher extent of the temporal relationship existing among different sensing units in the IoT environment. Hence, if there is any significant event than all the information of the connected sensing units can be extracted and they can be utilized for performing certain actuation forms of task for the sensors. Such information obtained is highly important in order to perform management of the IoT resources. However, at present, there are no such approaches ever evolved in order to extract such unique forms of frequent patterns from the data stream of the sensing units. Apart from this, it is a highly computationally expensive task in order to extract such forms of unique frequent patterns from the higher number of sensing units over the cloud environment. Therefore, these methodology offers a highly distributed and yet well-synchronized connectivity among the sensing nodes in order to extract essential mined information.

b. The contribution of the adopted methodology

Following are the contribution of the proposed system:
- A unique form of frequent-pattern based approach is introduced to establish the potential relationship among the nodes as well as their respective data.
- A unique tree structure has been presented that is capable of formulating a unique communication process governing a unique flow of analytical data within WSN.
- A unique and novel analytical model is introduced that performs the computation of different forms of sensory estimates for better granularity in the mining process.
- The proposed system has also a supportability of single-hop as well as multi-hop routing operation and hence is significantly utilized in the present-day application of WSN.
- A highly simplified computational modeling is carried out which offers better technical adoption in presence of near real-world sensing demands.

Therefore, the proposed system is meant for addressing the problem associated with performing an analytical operation over the sophisticated and complex sensory data over the cloud environment.

3. MODEL DISCUSSION

The complete analytical process of the proposed system is designed on the basis of a novel model that uses the concept of frequent patterns as the backbone. The model harnesses the potential concept of simplicity in using frequent patterns and addresses the scalability problem associated with it by introducing a unique concept of epoch management and tree-based mining. The core ideology of the proposed model design is to ensure the introduction of the tree-based structure for the topology construction in such a way that it doesn't only assists in routing data but it offers more granularities in forwarding more error-free data. The proposed model also offers significant benefits towards extracting all possible forms of relationship just by using the novel idea of frequent patter approach.

a. Model parameters

The proposed model considers that there are M set of sensory motes where $M=\{m_1, m_2, \ldots., m_n\}$, where n represents a total number of sensors. As the sensor perform data collection on the basis of its predefined time slots t (t_1, t_2, ….t_a) so the study considers that effective current time (difference of t_{s+1} and t_s, where $s\epsilon[1, a-1]$) is empirically represented as δ, which a is the size of the time slot. The proposed model also represents pattern α as a set of specific k number of sensors i.e. $\alpha=\{m_1, m_2, \ldots.m_k\}$. A tuple $\beta(\beta_t, \gamma)$ is a representation of an epoch which is required for constructing a sensory data repository matrix db. Therefore, this matrix db is a collection of the finite number of epoch considering γ as a specific pattern corresponding to the event that has been captured by the sensor within the defined time slot. The variable β_t is considered as the time slot of an event. The modeling aspects consider to carry out an analytical operation using frequent patterns concept applicable over sophisticated data in distributed sensory applications. Therefore, a condition is set where a supportability of a specific pattern σ is only considered to be valid by an epoch $\beta(\beta_t, \gamma)$ if,

$$\sigma \subseteq \gamma \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1)$$

The occurrence of the specific pattern σ over the repository *db* is a number of the epochs that are found within *db* itself. It can be mathematically represented as,

$$f(\sigma, \text{db}) = |\{\beta(\beta_t, \gamma)| \sigma \subseteq \gamma |\} \tag{2}$$

For better precision, the conditional logic is designed to highlights that a specific pattern σ can be only called as a frequent pattern if $f(\sigma, db)$>minimal cut-off support.

b. Implementation strategy

The complete implementation strategy is all about constructing a novel tree for a given sensory repository *db* after reading all the data within it in one shot. The implementation strategy also considers that there is a non-redundant route between two nodes that initiating from the root node of the tree. The cumulative information of the sensory nodes corresponding to all repository *db* can be extracted from this tree. The weighted sensory estimated score is numerically higher than or equal to the cumulative weighted sensory estimates of all the children nodes.

c. System implementation

The complete system is implemented on the mechanism of graph theory, which is further divided into two phase viz. graph construction phase and mining phase. The proposed system uses certain essential attributes that are used for assessing the quality of the mining operation on the basis of the estimates obtained from each epoch for individual sensors.

- Sensory Estimates (SE): Basically, this attribute represents an active number of eventual reading captured by sensor mote for a given epoch.
- Epoch Sensory Estimates (eSE): This attribute represents the cumulative sensory estimate corresponding to a given epoch.
- Cumulative Sensory Estimate (cSE): This attribute depicts the cumulative value of epoch sensory estimate corresponding to all the epoch in sensory repository db.
- Share Estimate (ShE): The share estimate attribute is computed as an estimated score of the group of sensors for a given epoch divided by the cumulative sensory estimate for the sensory repository.
- Minimum Share: The minimum share is basically a kind of cut-off value which considers that the sharing score of pattern specific to an event within a timeslot is more than or equal to minimum share.

The tree-based mining approach is implemented that takes the input of n (total motes), m (motes) which after processing offer an output of ordered tree. The steps involved in this algorithm are:

Algorithm for Tree-based Mining
Input: n, m
Output: ordered tree
Start
1. init
2. **For** i=1:n
3.   construct a tree(r_{node}, c_{node}, h_{mat})
4.   set epoch→SE_1
5.   sort(m)$_n$
6.   sort(wSE) & reorganize tree
7. **End**
End

The design of the proposed implementation consists of r_{node} as the root node, c_{node} is a child node, and h_{mat} is a matrix to list headers. The complete algorithm implementation is carried out in two distinct stages i.e. *mapping* stage and *re-ordering* tree stage. The first stage of mapping is carried out by organizing the motes in a specific order on the basis of their identifier. For this purpose, the tree is constructed by incorporating all the epoch in the entire sensory repository sequentially in order to obtain a final tree. The proposed system also constructs the matrix h_{mat} for retaining the list of headers of the sensor. This operation is carried out for maintaining the order of the motes as well as stores the weighted sensory estimates (wSE) associated with the sensor. In order to maintain the traversal feature of the tree, the model also maintains neighboring points. In the preliminary stages of implementation, the tree structure is usually empty and it initiates with the root node and finally, the tree is constructed using all the defined epoch. The final stage of implementation is basically associated with the re-ordering process. This implementation stages basically targets memory reduction and offers a faster process in the analytical operation. The stage

---

begins with ordering the list of elements in h_{mat} in decreasing order with respect to wSE where merge and sorting process can be utilized followed by further re-ordering the tree structure on the basis of the new value. A branch sorting mechanism can be used for the purpose of sorting operation of the tree. In this tree construction process, all the links among the sensor nodes are basically sorted using this branch sorting algorithm. The analytical process is carried out by assessing the growth in the pattern in order to extract the knowledge of all the sensory data from the constructed table. The proposed system utilizes the pattern-growth [36] concept in order to perform the analytical operation with an aid of weighted sensory estimates. The complete mechanism is also carried out using both single hop as well as the multi-hop scheme of data aggregation. It will mean that proposed system offers execution of its analytical operation in both single and multihop network for offering its supportability to all upcoming routing schemes that work on the principle of distributed mining approach over cloud environment.

The operation of the parallel process of analytical operation is as follows: For this purpose, the proposed system first takes the input of repository *db* and performs segregation of the *db* based on n number of sensors. The outcome of this analytical operation will be *frequent patterns* of sensory data as the *knowledge*. The process considers the local repository system for each location and performs the sorting of all the epoch β based on their specific identifier. It then includes the value of epoch β into the designated tree structure followed by the updating operation of the header table. The process then forwards the primary sensory estimates depending upon its epoch and the weight value. These values are forwarded to the root node. The sensory estimated based on secondary epoch and weight factor is constructed by the root node. The secondary sensory estimates in this routing table are then re-ordered in decreasing order and then it is forwarded to the primary location of the repository. The final tree is reconstructed on the basis of the newly acquired order followed by the identification of the similar value of the sensory estimates with epoch and weight over all the routes and then they are integrated to the single sensor mote. The next step of the analytical operation is carried out only if the partitioned repository is processed. The user obtains the minimum share and then all the sensor mote checks if epoch based sensory estimates are found more than the minimum cut-off value of the sensory estimates. In the positive case, the sensor mote is included in the upcoming pattern list followed by an iterative analytical operation that prefixes the mote by constructing a revised tree structure. All the upcoming patterns are added with the sensor mote in the upcoming list and then the upcoming patterns are forwarded to the root node from the entire primary node. Finally, all the frequent patterns are obtained from the root node which not only reduces the time of analysis but 90% cut short the process of mining operation when performed from the root node.

4.   RESULT DISCUSSION

The analysis of the proposed system is carried out in MATLAB using a simulation-based approach. Normal system using 4GB RAM and windows OS is used for the simulation study. One of the essential factors to be considered during the analysis of the result is the data as the complete process of the mining operation is carried out on this. The proposed system considers the mote configuration based on MEMSIC nodes. The proposed system is simulated in order to generate maximized tuples of sensory data against any specific environmental data. The analysis was also carried out that data aggregation has been carried out in the presence of inferior quality of the wireless medium from the transmitting source.

The analysis of the proposed outcome also assumed that if there are any forms of missed/skipped sensory data than it is considered as undetected an event that assists in generating skipped sensory reading. A synthetic data is generated that actually doesn't offer any form of information of the share data corresponding to the all the items equivalent to all the unit transactions. In order to connect/map all the items with the exact eventual data, the proposed system adds an arbitrary number associated with all the items. A simulation environment of 100x100 m^2 is constructed in MATLAB with 50 sensor nodes being distributed randomly over it. As the proposed system is designed over the concept of IoT; therefore, a gateway node is positioned in the center of the simulation area that is meant for assisting in the translational services of different types of routing operation in WSN. The analysis is carried out with respect to two scenarios where the data aggregation is carried out using a single hop communication system as well as distributed (or multihop) communication system. Figure 2 and Figure 3 highlight the simulation process of data aggregation.

The analysis of the outcome obtained from the simulation study:
a.   Analysis of time required for analytical operation

Processing time plays an essential role in offering clarity towards a faster response rate of the mining process. As the proposed system is implemented over the distributed tree structure, it is essential to understand that *how fast the mining operation is possible*? Otherwise, the applicability of the proposed mining operation is difficult to understand. Figure 4 highlights the comparative analysis of the proposed system with existing frequent pattern algorithm over increasing hypothetical data. The outcome shows that

although processing time increases for both approaches, it is comparatively better for the proposed system in contrast to the existing approach. The prime reason behind this outcome is that existing frequently pattern just records the occurrences of items which also exponentially grow with the increase in traffic. However, the proposed system obtains highly time-based unique data, which not only reduces the size of patterns but also offers faster mining operation.
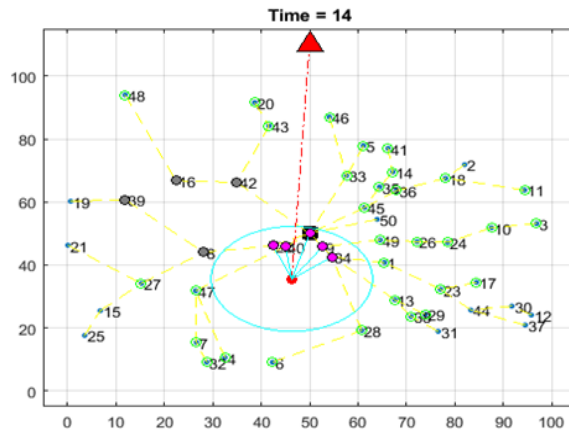

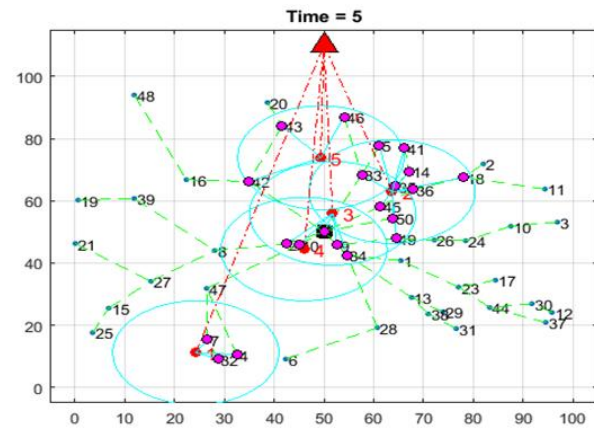
Figure 2 Single hop data aggregation
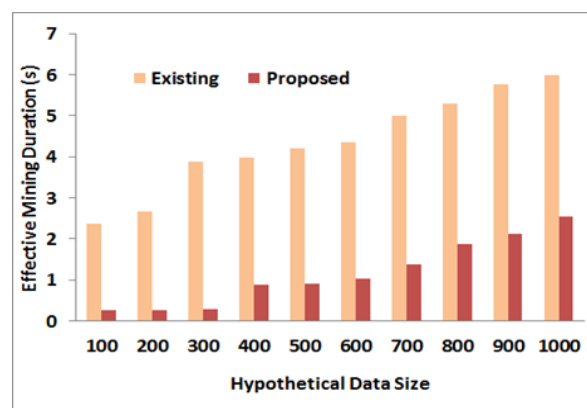

Figure 3 Distributed data aggregation

Figure 4 Comparative analysis of mining time

b. Analysis of energy depletion for analytical operation

Energy is one of the most practical parameters in order to judge the effectiveness of any form of processing being used within the sensor mote. For practical operation, it is always expected that a sensor mote should not deplete energy at a very faster rate in order to sustain a better form of network stability and lifetime. Figure 5 highlights that the proposed system offers highly reduced energy consumption compared to existing frequent pattern based mining approach. The prime reason behind this is existing frequent patterns retains a maximum number of patterns which increases the size of the tree if the incoming data is considered as a bit stream. This causes excessive depletion of transmission energy towards aggregation followed by mining in the cloud environment. However, proposed address this problem by introducing a tree construction and re-ordering where the mining operation is carried out over root node with all updates causing maximum energy saving.

c. Analysis of memory consumption for analytical operation

The proposed system offers reduced memory consumption with each node programmed with 1000 KB of internal memory. Figure 6 highlights that existing frequent patterns-based mining approach consumes higher memory owing to the increase of data size over the tree, whereas the proposed system maintains the mined data only in the root node.
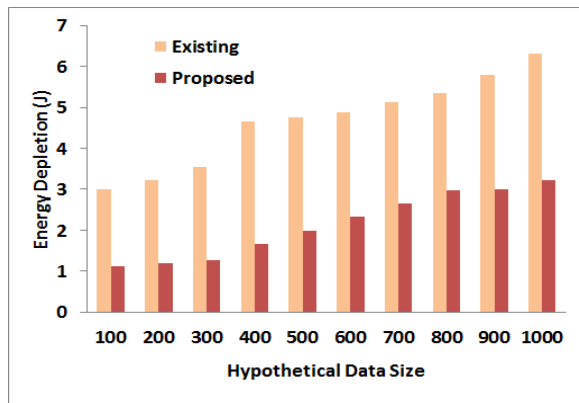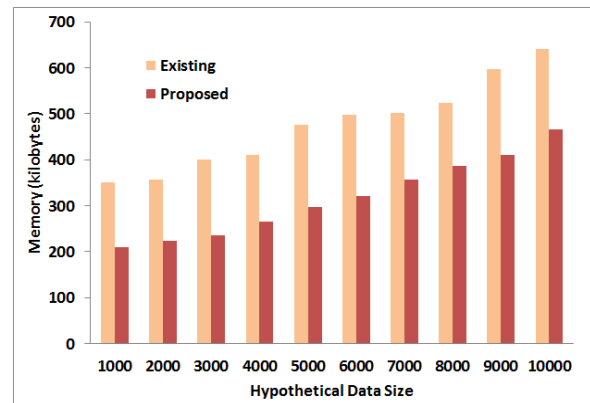
Figure 5. Comparative analysis of energy depletion

Figure 6. Comparative analysis of memory

5.   CONCLUSION

The core idea of the proposed paper is to showcase that there is a higher deal of complexity in order to plan a design of an architecture performing an analytical operation over large sensory data on the cloud environment. Studies and approaches using the existing system are more focused on case-specific mining while the number of studies of knowledge extraction exists for the sensor network. However, significant research towards knowledge extraction of sensory data over IoT environment is just in the nascent stage and need more explorative process. The proposed system argues that frequent pattern can be treated as knowledge but they cannot be directly used in such complex and distributed environment. Therefore, the paper introduces a significant novelty in the existing system of frequent patterns and introduces various sensory estimates to it in order to offer more granularities in the analytical outcome. The study also uses tree-based topology in order to make a structured of data management associated with mining outcome. The proposed model is simulated in MATLAB over normal system configuration to find that it offers reduced energy consumption, reduced delay, and better memory utilization.

REFERENCES

[1]   Udoh, Emmanuel, "Evolving Developments in Grid and Cloud Computing: Advancing Research: Advancing Research", *IGI Global*, pp. 383, 2012
[2]   Ahmed Abdelgawad, Magdy Bayoumi, "Resource-Aware Data Fusion Algorithms for Wireless Sensor Networks", *Springer Science & Business Media*, pp. 108, 2012
[3]   Javier López, Jianying Zhou, "Wireless Sensor Network Security" IOS Press, Computers, pp. 313, 2018
[4]   Zaman, Noor, "Wireless Sensor Networks and Energy Efficiency: Protocols, Routing and Management: Protocols, Routing and Management", *IGI Global*, pp. 655, 2012
[5]   olger Karl, Andreas Willig, "Protocols and Architectures for Wireless Sensor Networks", *John Wiley & Sons*, pp. 497, 2007
[6]   Holger Karl, Andreas Willig, "Protocols and Architectures for Wireless Sensor Networks", *John Wiley& Sons, Technology & Engineering*, pp. 497, 2007
[7]   Mohammad S. Obaidat, Sudip Misra, "Principles of Wireless Sensor Networks", Cambridge University Press, pp. 415, 2014
[8]   Frode Eika Sandnes, Yan Zhang, Chunming Rong, Laurence Tianruo Yang, "Ubiquitous Intelligence and Computing" *5th International Conference, UIC 2008, Proceedings*", Springer Oslo, Norway, June 23-25, pp. 763, 2008
[9]   David Hanes, Gonzalo Salgueiro, Patrick Grossetete, Robert Barton, Jerome Henry, "IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things", *Cisco Press*, pp. 576, 2017
[10] Christopher Siu, "IoT and Low-Power Wireless: Circuits, Architectures, and Techniques", *CRC Press, Technology & Engineering*, pp. 394, 2018
[11] Conti, Francesco, et al. "An IoT endpoint system-on-chip for secure and energy-efficient near-sensor analytics", *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol 64, no. 9, pp. 2481-2494, 2017.
[12] Bertrand, Alexander, and Marc Moonen. "Distributed canonical correlation analysis in wireless sensor networks with application to distributed blind source separation." *IEEE Transactions on Signal Processing,* vol. 63, No. 18, pp. 4800-4813, 2015.
[13] Parwez, Md Salik, Danda B. Rawat, and Moses Garuba. "Big data analytics for user-activity analysis and user-anomaly detection in the mobile wireless network." *IEEE Transactions on Industrial Informatics,* vol.13, no. 4, pp. 2058-2065, 2017.

[14] Rahman, Ashfaqur, John McCulloch, and Quazi Mamun. "Prediction with uncertainty: a novel framework for analyzing sensor data streams." *IEEE Sensors Journal,* vol 15, no. 1, pp. 382-386, 2015.

[15] Rehman, Ur Muhammad Habib, et al. "Big Data Analytics in Industrial IoT Using a Concentric Computing Model." *IEEE Communications Magazine*, vol 56, no 2, pp 37-43, 2018.

[16] Sun, Yunchuan, et al. "Internet of things and big data analytics for smart and connected communities." *IEEE access*, vol 4, pp. 766-773, 2016.

[17] Wang, Ying, Ashish Pandharipande, and Peter Fuhrmann. "Energy Data Analytics for Nonintrusive Lighting Asset Monitoring and Energy Disaggregation." *IEEE Sensors Journal,* vol 18, no 7, pp. 2934-2943, 2018.

[18] Yue, Peng, et al. "An SDI approach for big data analytics: The case on sensor web event detection and geoprocessing workflow." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol 8, no. 10, pp. 4720-4728, 2015.

[19] Pena-Ramos, Juan-Carlos, et al. "A Fully Configurable Non-Linear Mixed-Signal Interface for Multi-Sensor Analytics." *IEEE Journal of Solid-State Circuits,* vol 53, no 11, pp. 3140-3149, 2018.

[20] Sharma, Shree Krishna, and Xianbin Wang. "Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks." *IEEE Access,* vol. 5, no. 99, pp. 4621-4635, 2017.

[21] Cao, Ningyuan, et al. "Self-optimizing IoT wireless video sensor node with in-situ data analytics and context-driven energy-aware real-time adaptation." *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 64, no. 9, pp. 2470-2480, 2017.

[22] He, Jianhua, et al. "Multitier fog computing with large-scale iot data analytics for smart cities" *IEEE Internet of Things Journal,* vol 5, no. 2, pp. 677-686, 2018.

[23] Gong, Yanmin, Yuguang Fang, and Yuanxiong Guo. "Private data analytics on biomedical sensing data via distributed computation" *IEEE/ACM transactions on computational biology and bioinformatics,* vol. 13, no. 3, pp. 431-444, 2016.

[24] Li, Xian, Hui Huang, and Ye Sun. "DriTri: An in-vehicle wireless sensor network platform for daily health monitoring" *SENSORS, IEEE*, 2016.

[25] Yildirim, Murat, Nagi Z. Gebraeel, and Xu Andy Sun, "Integrated Predictive Analytics and Optimization for Opportunistic Maintenance and Operations in Wind Farms" *IEEE Transactions on Power Systems*, vol, 32, no. 6, pp. 4319-4328, 2017.

[26] Liao, Yizheng, et al, "Snowfort: An open source wireless sensor network for data analytics in infrastructure and environmental monitoring", *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4253-4263, 2014.

[27] Muñoz, Raül, et al., "Integration of IoT, Transport SDN, and Edge/Cloud Computing for Dynamic Distribution of IoT Analytics and Efficient Use of Network Resources," *Journal of Lightwave Technology,* vol. 36, no. 7, pp. 1420-1428, 2018.

[28] Otero, Carlos E., et al. "A wireless sensor networks' analytics system for predicting performance in on-demand deployments," *IEEE Systems Journal*, vol. 9, no. 4: 1344-1353, 2015.

[29] Shao, Yun, et al. "Heuristic optimization for reliable data congestion analytics in crowdsourced eHealth networks," *IEEE Access,* vol. 4: pp. 9174-9183, 2016.

[30] Ivanov, Stepan, Kriti Bhargava, and William Donnelly, "Precision farming: Sensor analytics," *IEEE Intelligent systems*, vol. 30, no. 4, pp 76-80, 2015.

[31] Chandrakala, N., and B. Thirumala Rao. "Migration of Virtual Machine to improve the Security in Cloud Computing," *International Journal of Electrical & Computer Engineering*, vol. 8, no. 1, pp. 210-219, 2018.

[32] Rghioui, Amine, and Abdelmajid Oumnad. "Internet of Things: Surveys for Measuring Human Activities from Everywhere" *International Journal of Electrical & Computer Engineering*, vol 7, no. 5, pp. 2474-2482 , 2017.

[33] Sindhu, C. S., and Nagaratna P. Hegde. "A Novel Integrated Framework to Ensure Better Data Quality in Big Data Analytics over Cloud Environment" *International Journal of Electrical & Computer Engineering*, vol. 7, no. 5, pp. 2798-2805 2017.

[34] Manujakshi, B. C., and K. B. Ramesh. "SDaaS: framework of sensor data as a service for leveraging services in Internet of Things." *In International Conference on Emerging Research in Computing, Information, Communication, and Applications*, pp. 351-363. Springer, Singapore, 2016.

[35] Manujakshi, B. C., and K. B. Ramesh. "A Novel Experimental Prototype for Assessing IoT Performance on Real-Time Analytics." In *Computer Science Online Conference*, pp. 46-55. Springer, Cham, 2018.

[36] Rashid, Md Mamunur, Iqbal Gondal, and JoarderKamruzzaman. "Share-frequent sensor patterns mining from wireless sensor network data." *IEEE Transactions on Parallel and Distributed Systems,* vol 26, no. 12, pp.3471-3484, 2015.

BIOGRAPHIES OF AUTHORS

Manujakshi B. C. is working as assistant professor in the department of computer science and engineering, presidency university, Bengaluru, Karnataka, India. completed B. E in computer science and engineering from SJMIT, Chitradurga, Karnataka, India, and M. Tech in computer science and engineering from UBDTCE, Davangere, Karnataka. Area of interest are Big data and cloud computing technique.

K. B. Ramesh, Associate Professor and Head, Department of Electronics and Instrumentation Engg.R V college of Engineering, Bengaluru, Karnataka, India. Hehas completed PhD in Computer Science and Engineering from Kuvempu University. He hasaround twenty-threeyears (23) of teaching experience in E&I Engg. His major research area is in Computer Science and Engineering and minorresearch area is in Biomedical Engineering/Bioinformatics.