Survey of Data Fusion technique for Routing and Performance enhancement in WSN

(Data fusion in WSN- A survey)

Ms. Kanika, Mrs. Priyanka Kamboj
CSE, JMIT RADAUR
Kurukshetra University
Karnal, India

ABSTRACT

Network energy consumption is a critical factor determining the WSN development speed. Various techniques are available for the data generated by the sensor nodes to send it efficiently to base station. Along with this possibility there are many other possibilities that can be taken into consideration and as a result of these techniques and rise in the number of the techniques the wireless sensor network is emerging from the very start, when this field is introduced. In-network data fusion is needed for energy-efficient information flow from a plurality of sensors to a central server or sink. Many techniques has been developed and tested and various techniques succeeded in achieving the ultimate goal of more efficient utilization of the resources. As data is propagated towards the sink, multiple levels of data fusion are likely. The data fusion at various levels should be synchronized in order to fuse data effectively. This paper presents a descriptive and broad review of existing techniques of data fusion. In the survey of the development life cycle of the wireless sensor networks(WSNs) it is felt that the field of data fusion is ignored and remain isolated as compared to other fields researched by the research community. The paper discusses majority of the techniques that are having great impact on the way the research is carried out and provides a path way for the futuristic research in this field.

Keywords—*WSN; energy-efficiency; performance; data fusion*

I. INTRODUCTION

Wireless sensor networks(WSNs) are composed of lots of low-cost, low-powered tiny wireless nodes and can be used for many long-term applications, such as environment monitoring, building protection and supply chain and logistics management[1-3]. For energy saving and environmental protection of the building environment, WSNs can be used to sense and collect environmental parameters such as temperature, humidity, air pressure, which can be used as the input of the automation control system of the building via Internet or industrial data bus. One of the most essential technique in wireless sensor network is known to be Data fusion which is responsible for the integration of unique and various types of data generated by the surrounding sensor nodes. In existence of the Data fusion there is another technique popularly known as data aggregation, which is often confused with the term Data fusion. Data aggregation consist of the collection of all data from the pervasive data sources[4]. Aggregation by its meaning is the summarization of the collected data from the various sources, which in effect reduce the volume of the collected data , ultimately the data received by the sink node is the refined data. This technique finds the advantage where the summarized data is needed and helps the application goal achievement. As a disadvantage of this technique it does not find its place in a wide number of applications that need the data precision. Reduction in the transmitted data results in loss of accuracy. However, fusion concept is having a different path from that of the aggregation phenomenon in WSN. Data fusion in simplest terms is defined to be process of combining complementary data so as to allow interferences from resultant dat, which is not freasible to obtain from the individual measurements.

This paper essentially discusses the uniqueness and distinction for the data fusion technique. Data fusion can be implemented in different protocol layers. The network layer is mainly used for routing and data delivery, the application layer is mainly used for the query, data fusion. For a build environment monitoring system, while lots of environmental data should be sensed and transmitted for real-time acquisition, data fusion strategy in different layers can be used in the process to reduce energy consumption of WSNs, prolong the network life cycle.

Data fusion in network layer: different from the traditional network, WSNs don't care about the specific sensor on a single data, pay more attention to the multi-node cooperative information. collected, such as: temperature monitoring. Temperature distribution are concerned about the area in a specific information, but not limited to specific node value, more is how to transmit this information through the network to gather nodes which makes in the process of data transmission to speed up the convergence of redundant data, and to choose energy efficient routing in the form of multiple hops, reduce conflict of data transmission, improve collection efficiency. Data fusion in the application layer: application layer data fusion technology research is mostly based on query mode of data fusion technology, based on the aggregation of distributed database operation, the user sends a query request to the network using descriptive language query request in the network in a distributed manner. The query results through multiple hops routing returned to the user, handle query requests and return the query result is essentially the process of the data fusion process.

Data fusion technology using computer high-speed computing power and complementary of multi-source data to improve the quality is an integrated information process, which aims to eliminate the contradiction between the data collection of nodes in wireless sensor network, reduce redundancy and uncertainty.

II. DATA FUSION AND ITS DETAILS

Quite often the terms data fusion and data aggregation are used interchangeably but there lies a difference between the two. Data fusion is defined as the use of techniques that combine data from multiple sources and gather this information in order to achieve inferences, correlations, associations which are more efficient and potentially more accurate than if they were achieved by means of a single source [7], whereas, the second term Data aggregation which is a subset of data fusion is just a process of summarizing the data coming from multiple SNs in order to reduce or eliminate redundant data.

Process of data fusion can be centralized or distributed [7]. In centralized data fusion techniques, all the sensed data is sent to a single SN which performs the task of fusion. On the contrary, in distributed techniques, the task of data fusion is not solely the responsibility of a single SN, but the work is distributed among several SNs. In distributed techniques, each SN performs fusion of its own data and data from neighboring nodes. Due to the battery scarcity of a SN, distributed data fusion techniques are preferred over centralized ones. Broadly, there are two prime motives to perform data fusion: one to improve accuracy of data and another to conserve energy.

Some of the applications of data fusion are to detect routing failures, to overcome sensor failures, improve location estimates of nodes, and overcome spatial and temporal coverage problems and to reduce energy consumption in the network. Besides data fusion and data aggregation, there are several other terms of importance such as sensor fusion and multisensor integration. Figure 1 elucidates the relationship between all these terms.

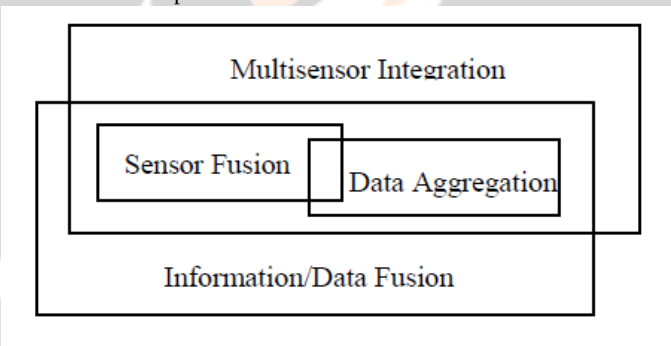

Fig1: Data fusion's different terms

Sensor/Multisensor fusion is used to specify that the data coming from SNs is fused into one representational format. However, Multisensor integration is superset of sensor fusion, where in addition to fusing the data, it makes clear to the system that how this data can be used to accomplish a task by the system and interact with the environment. Data fusion can be categorized based on different parameters. Classification of the same is depicted in Fig. 2.

Data Aggregation

The basic idea of data aggregation is to aggregate data at certain SNs known as aggregators thereby eliminating redundancies and thus, reducing the number of transmissions between SNs. According to [7], "Data aggregation comprises the collection of raw data from pervasive data sources, the flexible, programmable composition of the raw data into less voluminous refined data, and the timely delivery of the refined data to data consumers". Data aggregation reduces load on SNs, which helps in handling the data with priorities more effectively. Data aggregation techniques also have few disadvantages like increased delay, stronger hardware requirements, etc. [8].

Data aggregation shifts focus from address centric routing to data centric routing [9]. In address centric routing, individual SNs send data to the sink via a shortest path, whereas in data centric routing, different SNs send data to sink but routing nodes in between can perform aggregation on the data packets. The task of performing data centric routing with optimal data aggregation is a NP-hard. To reduce the complexity, three suboptimal schemes; Center at Nearest Source, Greedy Incremental Tree and Shortest Path Tree have been proposed [10].

The performance of aggregation methods depends on several factors like the number of source, their position in the network, the topology of the network, etc. Followings are several performance characteristics which are used to design data aggregation algorithms [11],[12]:

- Energy efficiency: Energy is the scarcest resource in WSN and must be consumed efficiently. A data aggregation scheme should be energy efficient. The ultimate goal of any data aggregation scheme is that each sensor possesses and spends the same amount of energy in each data gathering round.

- Network lifetime: Data aggregation schemes should balance the power consumption of all the nodes so that the network lifetime can be increased.

- Scalability: Aggregation protocols must be scalable enough so that they can easily and efficiently work with large number of nodes.

- Latency: Data aggregation mechanism should introduce least latency.

- Data accuracy: Data aggregation protocol should preserve the data accuracy and should also avoid compromised data.

- Overhead: An aggregation algorithm must be designed so as to minimize bandwidth utilization, power consumption and processing requirements.

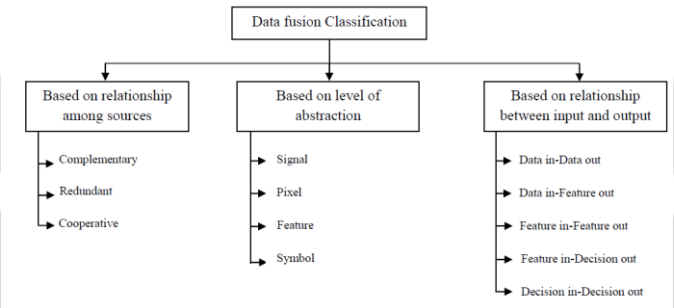

Fig 2: Classification of Data Fusion

III. LITERATURE SURVEY

A sensor network consists of a large number of wireless micro-sensor nodes which are distributed over a certain area. The area may be divided into a number of regions based on the positioning precision requirements, sensing range of the sensors and other application specific requirements. Each micro-sensor node has at least one sensor, a computation unit and a radio transceiver. There are three circumstances that would cause a sensor node to send a report to the sink. First, sensor nodes periodically send reports to the sink and we call this periodical reporting. Second, the sink queries sensors in specific regions for current sensed information and we refer to this as sink inquiry response reports. Third, due to the occurrence of certain events, reports are triggered from sensors in the particular region in which the event occurs; we call these reports event triggered reports. The schemes that proposed can easily apply to the other scenarios as well, with minor modifications. Depending on the target event (application) and the types of sensors deployed (temperature sensors, pressure sensors, motion sensors, etc.), the way in which data is fused may vary. Data fusion models for various types of sensed data may be found in [9].

Zadch L. A. [1] the United States Professor in 1965, proposed fuzzy set theory to seek a kind of processing can't accurately describe the ambiguity problem of rigorous mathematical method. Fuzzy predictive control algorithm is using fuzzy reasoning process of uncertainty information processing and decision-making to improve the performance of predictive control or fuzzy control. Zhao Yaguang [12] applied BP neural network to the data fusion and concluded the data fusion model, a detailed analysis of the basic principles of the BPNN, key issues, strengths and weaknesses as well as the application in WSN data fusion. Combining BP neural network with data fusion, get data fusion model based on the BPNN. Yuan Xia [13] used fuzzy prediction for feature extraction and optimized association of the input and output. In the work [2], BP neural network is usually adopted to solve the problem of intercross sensitivity of pressure sensor to temperature. PSO algorithm is applied to train the weights of neural multi- layer forward neural network. The results of the data fusion show that the stability and accuracy of the sensor are improved greatly. In the work[3]proposed a neural network algorithm based on BP multi-sensor data fusion method for modeling complex situations. In the work[4]A practical, general data fusion scheme was established on the basis of feature extraction and merge of data from multiple sensors. This scheme integrates artificial neural networks for high performance pattern recognition. In the work[5]BP neural network is usually adopted to solve the problem of intercross sensitivity of pressure sensor to temperature. PSO algorithm is applied to train the weights of neural multi- layer forward neural network. The results of the data fusion show that the stability and accuracy of the sensor are improved greatly. A practical, general data fusion scheme was established on the basis of feature extraction and merge of data from multiple sensors. This scheme integrates artificial neural networks for high performance pattern recognition.

Typically, the sink is distant from the area where the sensor nodes reside. The sensor data has to be ultimately relayed to the sink via multiple sensor node relays. One can then visualize the data being transferred via a structure that facilitates the *manyto- one* data transport. In some sense, building such a structure is akin to building a single source multicast tree, the difference is that instead of data propagating from the single source to the multicast group members, the data flows in the opposite direction, i.e., from the members to the sink. A second difference is that en route, data may be fused at various vantage points on the tree. The topology of this *aggregation tree* determines the efficiency with which data may be fused, to a certain extent. While the discussion of algorithms that help to generate and maintain this tree are beyond the scope of this paper, we show the effects of the topology on the fusion process and how our synchronization protocol can function independent of the topology of the tree. Requirement is that each sensor node is aware in contrary another example of data fusion might involve an attempt to estimate the precise location of a target by extrapolating its distance computations from multiple sensors whose co-ordinates are known. Even here, the more the reports the more precise will be the estimated position of the target. of its immediate neighbors; specifically it should know its *parent* i.e., the sensor node to which it sends data (either fused or raw) and its *children*, sensors from whom it receives data. Each *non-leaf* node or *internal node* is responsible for relaying (after possibly performing fusion) data received from its children towards the sink node.

Data Fusion Process Based On SMPSO-BP Algorithm

1 Data Collection

After the forming of the cluster structure of wireless sensor networks, the cluster member nodes can send packets to the cluster head node. In this process, in order to avoid mutual interference of data, member nodes in the same cluster use time division multiplexing (TDM) method to send data to the cluster head node in different time slot.

2 Training of the Network Parameters

The definition of the position vector of the particle cluster nodes: in the three layers structure of BP neural network, Setting R, S1, S2 to be input layer, hidden layer and neurons quantity of output layer respectively, then dimension N of PSO individual unknown vector P is the sum of network weights and thresholds. As shown in the equation $N = S1\times R + S1\times S2 + S1+ S2$

The position vector of particle group individual The definition of the particle cluster nodes velocity vector :

PSO velocity vector of the node as formula(2).

$( ) 2 ,v ,...,v\ i\ i\ N\ V = v$ (2)

where $N = S1\times R + S1\times S2 + S1+ S2$ .In each dimension, there is a maximum limit for the particle velocity Vmax (Vmax> 0).

Here, define the function of updateParticle (), the main role of this function is to update PSO individual position and velocity, in order to obtain optimal BP network parameters. First, thing needed to initialize the particle swarm, including particle swarm scale M, the initial position Xi and velocity Vi of the particle; Second, calculate the fitness value of each particle based on the fitness function and update individual extreme pBest [i] and the global extreme gBest; At the end, decide whether each particle needs to have mutation operation, according to the setting conditions.

3 Data Fusion in Clusters

After each member node in the cluster send data to the cluster head node, following the assigned slot interval, the cluster head node complete the data fusion work, then transport all the data to the remote base station.

IV. FACTS ARRISING FROM SURVEY

The following points are extracted as a conclusion from survey:

- Wireless Sensor Networks would be widely deployed in future mission-critical applications.
- As wireless sensor networks continue to grow and become more common, we expect that further expectations of performance and data services will be required of these wireless sensor network applications.
- One of these advancements includes data fusion technique in routing protocol of wireless sensor network.

V. CONCLUSION

The proposed review is done with the objective of finding the research gap. During the review it is felt that studies with the objective of energy efficiency in data fusion is considerably less. As the outcome of the review, some open challenges comes out. Firstly, the scale of the simulation and system evaluation is done at the small scale. The focus of most of the research is away from the energy efficiency. Most of the prior research is focused towards the data aggregation. Some of the authors made assumptions away from reality. This review provides a

aim to work in the field of Data fusion. Futuristic implementation calls for addressing such issues without which outcomes of studies can not be properly justified with respect to real time environments. Hence , future work will be to introduce a framework that could substantially enhance the performance of the data fusion in WSN.

REFERENCES

[1] Abhijit Bhattacharya, Anurag Kumar. A shortest path tree based algorithm for relay placement in a wireless sensor network and its performance analysis[J]. Computer Networks, 2014, 71: 48-62.

[2] Getsy S Sara,D. Sridharan. Routing in mobile wireless sensor network: a survey[J]. Telecommunication Systems,2014, 57(1): 51-79.

[3] Amal Tiab,Louiza Bouallouche-Medjkoune,F. Jiang,Z. Li,S. Simani. Routing in Industrial Wireless Sensor Networks: ASurvey[J]. Chinese Journal of Engineering,2014,2014.

[4] Akyildiz I F, Su W, Sankarasubramaniam Y. Wireless sensor networks: a survey[J], Computer networks, 2002, 38(4): 393-422.

[5] Mathilde Arnaud, Véronique Cortier, Stéphanie Delaune. Modeling and verifying ad hoc routing protocols[J]. Information and Computation, 2014, 238: 30-67.

[6] Hedetniemi S, Liestman A. A Survey of Gossiping and Broadcasting in Commuication Networks[J], Networks, 1998, 18 (4): 319-349.

[7] Delfín Rupérez Cañas, Luis Javier García Villalba, Ana Lucila Sandoval Orozco, Tai-Hoon Kim. Adaptive routing protocol for mobile ad hoc networks[J]. Computing, 2014, 96(9): 817-827.

[8]] Chee-Yee Chong, Srikanta P K. Sensor Networks: Evolution, Opportunities, and Challenges[J], Proceedings of the IEEE, 2003,91(8): 1247-1256.

[9] Waleed S. Alnumay,Pushpita Chatterjee, Uttam Ghosh. Energy Aware Secure Routing for Wireless Ad Hoc Networks[J]. IETE Journal of Research,2014,60(1): 50-59.

[10] Haas Z, Halpern J Y , Li L. Gossip-based Ad hoc Routing[C], In Proceeding of the IEEE 1NFOCOM. New York :IEEE Communications Society, 2002:1707-1716.