SOME TECHNIQUES FOR COMPUTATION OF CONTEXTS AND SITUATIONS FROM SENSOR DATA

SYNOPSIS

of

Thesis to be submitted in fulfillment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

Submitted by

SANGEETA YADAV

Department of Computer Science Engineering and Information Technology

JAYPEE INSTITUE OF INFORMATION TECHNOLOGY

A-10, SECTOR-62, NOIDA, INDIA

MARCH 2015

1. Introduction

Wireless Sensor Networks(WSN) are traditionally characterized as distributed ad hoc mesh networks deployed for measurement of physical parameters such as temperature, pressure and humidity etc. , physiological parameters like body temperature, motion, orientation information and many more at unprecedented resolution [1]. Nodes in a WSN are typically tiny devices equipped with modest computational power, memory and battery along with sensing module [2].

Information potential of already existing sensors around us and few enhancements can be utilized for new path breaking applications in health care, security, habitat and climate monitoring [3]. Sensors enable these applications by providing highly detailed and frequent monitoring/ tracking information. The humane application that has been considered in this thesis is sensor based activity determination of monitored person with the purpose of proactive care.

Cameras, microphones and GPS devices are few conventional devices used for round the clock remote monitoring of a person. However, many people are not comfortable with these devices and consider them as breach of their privacy. Use of simple embedded sensors is non-obtrusive and more acceptable. Moreover, video and audio output from conventional monitoring devices require time consuming and complex processing methods [4]. On the other hand, data emanating from sensors is simpler, detailed, lightweight, voluminous but less complex to process.

Wearable sensors and those embedded in ambient surroundings and usable objects can provide desired monitoring information. In recent times, sensors like accelerometers, RFId tags, pressure, humidity, ambient light, temperature sensors and many more have become pervasive in our day to day environments. It is these sensors and their data that has driven research in this thesis.

The problem being addressed here is sequential *aggregation, fusion, analysis and interpretation* of raw sensor data. Solutions were sought with following broad objectives:

1) Transformation of raw sensor data into a form commensurate to that of human perception capabilities through simple on-node operations
2) Achieve robustness against uncertainties of wireless transmission

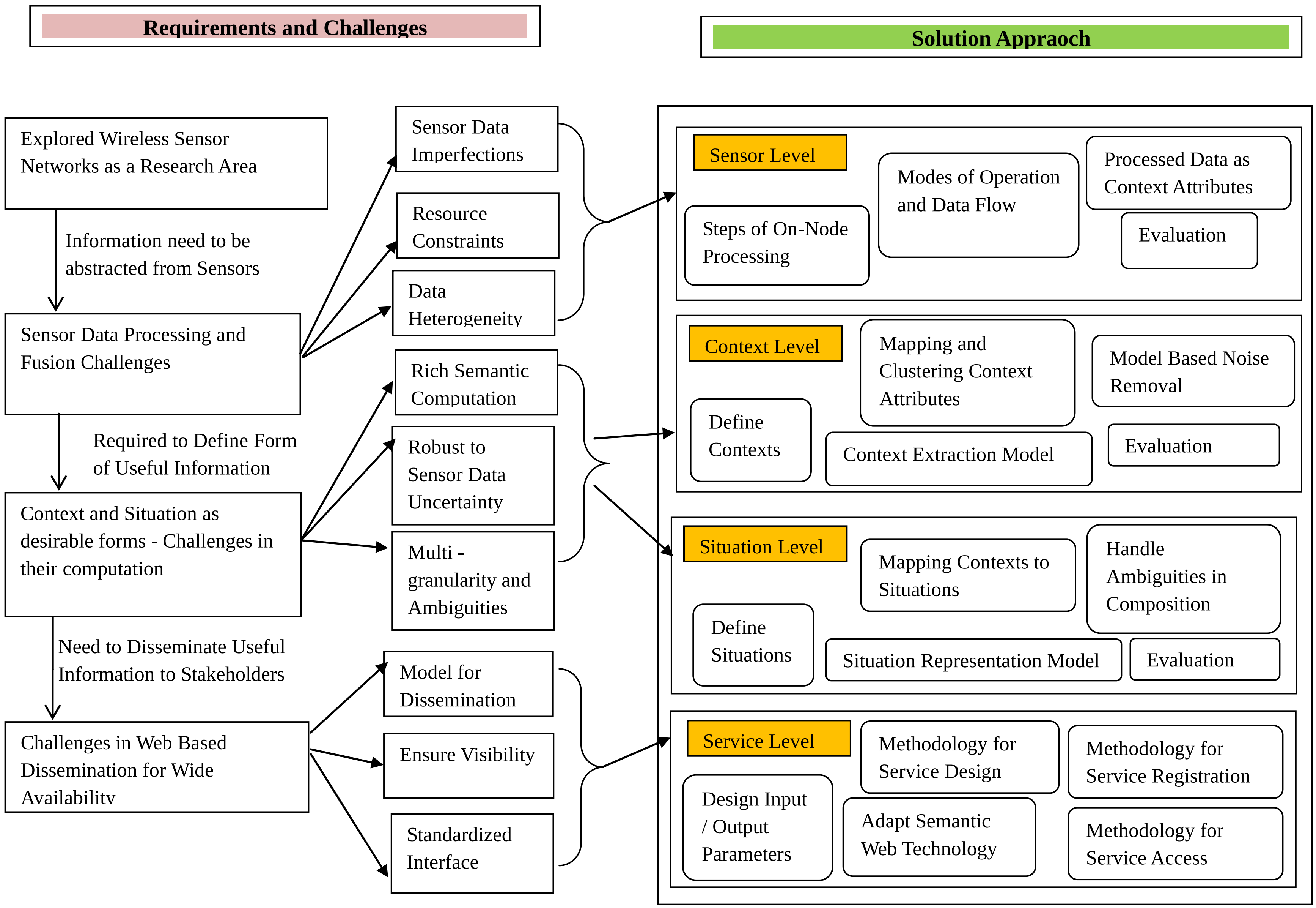

Figure 1 Overview of Requirements, Challenges and Solution Approach Used in Thesis

3) Make user interface independent of number and types of sensors

4) Accommodate ambiguities of real life situation composition

5) Achieve high accuracy of abstraction in an energy efficient manner

2. Background

Existing literature was surveyed to find about the existing research issues in wireless sensor networks in general and data handling in particular. Figure 1 shows an overview of systematically found requirements and related challenges for undertaken problem in left block. Solution approach developed in this thesis is summarized as components of right block of Figure 1.

Wireless Sensor Networks (WSNs) provide ample opportunities for multidisciplinary research. Various aspects of WSN and recent advances were surveyed to get an updated look at the state of art. An interesting finding was that from applications, assumptions, scale of deployment and protocol design point of views, WSNs can broadly be categorized in at least two classes. These have been referred as Homogeneous Wireless Sensor Networks (HoWSN) and Heterogeneous Wireless Sensors Networks (HeWSN) in this thesis. Identification of issues and characteristics of WSNs are not generic enough to be addressed without this categorization. The understanding developed about architectural and functional assumptions in these two broad classes of WSNs is summarized as Table 1.

Table 1 Comparison of Generic Categories of WSNs

Feature	**Homogeneous WSNs (HoWSN)**	**Heterogeneous WSNs (HeWSN)**
Cardinality	Thousands of nodes	Tens of nodes
Topology	Flexible Mesh or Hybrid	Point to point/multipoint to point Topology
Deployment	Self-Organizing - Random placement of nodes	Deterministic - Predefined Placement
Coverage	Broad Geographic Area, Highly Distributed	Usually Indoor or bounded area
Routing	Data Centric , Multi hop as Proximity of sink to every node is not possible	Single hop as Node to Sink Spatial proximity present
Demography	Unattended, Unpredictable Surroundings many times hostile environment	Well understood and user friendly surroundings
Redundancy	Redundant – many similar sensors , with overlapping coverage	Not Redundant due to heterogeneous sensors
Fault Tolerance	Can survive even if certain number of nodes fail, redundancy compensates for failure	Prone to single node failure due to lack of redundancy
Data Fusion	Gateway aggregates raw/aggregated data presented to it	Fused Information – created from low level abstracted data at Gateway
Example Applications	Weather/ Climate Monitoring, Coastline Monitoring, Border Tracking, Smart transport and logistics	Home Sensor Network, Body Sensor Network for health care

In early years of research, WSNs were assumed to be self-organized networks of large number of sensors placed randomly in unattended environments to form large sensing fields. A sizeable portion of research is devoted to protocol development for such WSNs. Research prototypes of these WSNs have been implemented for large scale climate monitoring, military applications and habitat monitoring.

However, some recent applications of interests like Home Sensor Networks and Body Sensor Networks use heterogeneous sensors for multi modal but shorter range sensing coverage. Thus, protocols and processing methods of conventional WSNs doesn't apply to them. Major differences have been presented in Table 1.

Research in this thesis has been devoted to sensor data fusion irrespective of underlying communication and routing protocol. Sensor data handling has to be addressed with respect to uncertainties like loss of data due to breakdown of sensors or obstacles, noisy environments and wireless medium. Constraints of sensors like energy and bandwidth are key factors in data handling approaches for both types of WSNs.

Routing centric aggregation as statistical summary over space and time has been common data handling approach in HoWSNs[5-7]. For systems with multiple types of sensors, new information is generated by fusion of aggregated or raw data of individual sensors. Goal of multi modal sensor fusion is to obtain more meaningful and abstract semantic descriptions than single sensor in reasonable time. State of art sensor fusion models and algorithms were studied for this purpose [8].

Studies in a parallel field of research, that is, "Context Aware Computing" were done for defining forms of sought levels of abstraction. Context Aware Computing aims to leverage automated knowledge of physical and social contexts of user to create smarter applications [9]. Context centric paradigm can be useful in sensor data fusion for obtaining rich semantic model of raw data [10].

Context computation methods fuse sensor data to deduce contexts. The underlying sensors could be similar or heterogeneous. For example, physical context of a person can be obtained by data from many accelerometers worn on various parts of human body defining locomotion. The abstract location context can be determined by fusing data from different sensors like GPS, ambient light and ambient noise etc.

Several machine learning algorithms have been proposed for context computation and reasoning with uncertain sensor data [11]. The algorithms need to be robust to achieve desired accuracy in presence of uncertain sensor data. In this thesis, existing techniques for context computation and inference from sensor data have been explored.

Probabilistic and fuzzy logic based approaches have been used in literature for handling uncertainty of sensor data. Naïve Bayes, Hidden Markov models and Dempster- Shafer theory of evidence are few such approaches [12]. Besides these, decision trees, K- nearest neighbour, support vector machines, artificial neural networks, multi- layer perceptron and random forests has been used for specific context of locomotion detection [13-17]. However, in almost all of these either single sensor or few similar sensors were used as data sources.

Very high accuracy ranging from 77 to 99 % has been claimed for these approaches. However, it is difficult to verify these claims due to non - availability of used data. Linear discriminant analysis, quadratic discriminant analysis, nearest neighbor and nearest centroid classifiers have been used on a publicly available dataset and accuracy ranging from 40% to 86% has been reported [18]. This dataset and its implemented algorithms have been used for comparison of context recognition methods developed in this thesis.

For further analysis and interpretation, extracted contextual information is taken at a higher semantic level which is closer to human understanding [19]. These abstractions have been termed as "Situations" in this work. A situation is said to be occurring if its description matches the set of currently detected contexts. Some situation recognition models proposed in literature consider sensors as input. Thirunarayan et al [20] explained synthesis of high- level, reliable information for situation awareness by querying low-level sensor data. In this research, symbolic contexts have been considered as input to situation recognition and these methods thus need to manipulate symbols and not raw sensor data. Some techniques proposed in literature for same are based on task based recognition, template based matching, event trees and fuzzy inference rules [21 – 22]. A guided clustering algorithm has been proposed to obtain clusters of sensor data to summarize individual events in [22]. The macro-clusters then integrate the information from multiple events. Srivastav et al define framework for abstraction of situations from objects and their usage information [23]. The relational dependencies among objects are modeled as cross-machines called relational Probabilistic Finite State Automata (PFSA) and are modeled using xD-Markov machine construction. These PFSAs are mapped to situations.

Models built using these approaches are too dependent on data which is not good for scalability of methods to situation recognition in different places or for different persons. Besides these, specification based approaches on first order predicate logic and ontologies have also been developed [24 – 25]. It was concluded that for situation computation, data based and expert specification based approaches individually are not suitable for our problem. A hybrid flexible approach for situation computation has been sought in this research.

Dissemination of interpreted data to remote stakeholders in a standardized manner is another problem of interest and semantic web methods were explored for it.

Based on these studies from literature, a generalizable sensor fusion framework that can support mapping of multi- sensor raw data into contextual information hierarchy and has device independent availability has been sought. The research problem has thus been defined as *"Development of computational techniques to extract Contexts and Situations in real time and in energy efficient manner utilizing application specific remotely sensed data"*. The overall focus of this thesis has been to develop methods to process, and produce semantic interpretations of sensor data streams.

3. Methodology

Right half of Figure 1 shows the four level solution approach and tasks done at each level. Each level significantly transforms its input to semantically more useful form, while handling sensor related challenges wherever applicable. A bottom-up methodology has been adopted to implement the levels. It starts with raw sensor data and works towards its transformation to user comprehensible description, that is, *Situations*.

Figure 2 represents more detailed schematic diagram for same. Four levels have been implemented in four sequential stages of computation where each stage outputs data at an enhanced abstraction level. Model parameters for each stage can be learnt only on archived sensor data. Two modes have been implemented, firstly training is done with part of available data to learn and communicate the computed parameters to next logical device and second mode is testing the models with remaining data.

Multiple and heterogeneous sensors worn by user or embedded in surroundings are the main sources of information. Data from these sensors is processed locally by sensors before transmission. Main objective of processing is to reduce the amount of data and transform it from continuous valued to finite valued. Experiments on various processing steps are done in first stage.

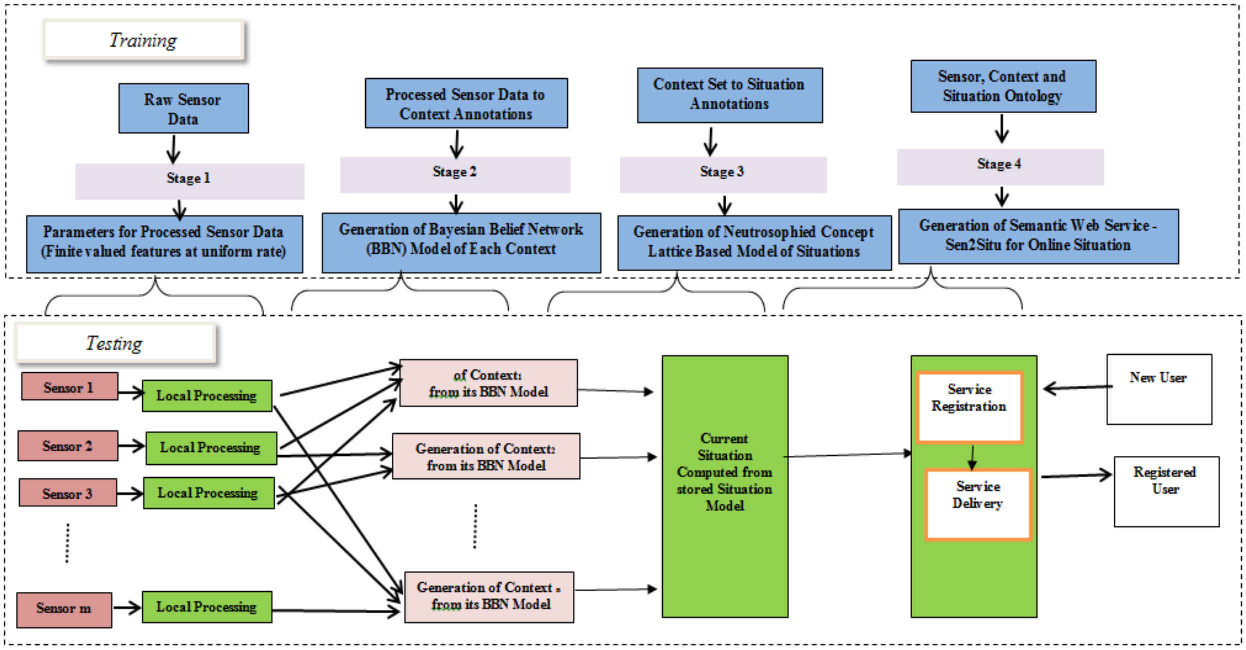

Figure 2 Schematic Representation of Adopted Methodology

Processed data is used in next stage for building supervised Bayesian Belief Network (BBN) based graphical and probabilistic fusion models of various types of sensors to a particular context. For example, fusion of sensors on left and right arms of user and on objects can detect the context "*action on an object*". This model is utilized during testing to compute contexts from sensors. Detailed performance of these models and their comparison with other methods is subject matter of stage 2 processing.

Set of contexts annotated with corresponding situation ground truths has been utilized to create a transparent, user understandable situation computation model based on neutrosophic concept lattice. The built model addresses the challenge of ambiguities in contextual composition of situation. Lattice construction and evaluation has been dealt with in stage 3.

In today's connected world, it would be worthwhile to present our live situation computations over web to interested user. A semantic web service framework to do so has been proposed in last stage.

Algorithms developed at all stages have been tested on benchmark third party real sensor datasets. Computation of contexts and situations has been considered as generic classification problem and standard metrics of time slice accuracy, situation/context wise recall, precision and F-measure have been used for evaluation of results.

Results of proposed methods have been compared with existing approaches, either qualitatively or quantitatively. Most existing methods have been described on closed personal datasets, thus it was not possible to empirically compare their results with ours. In few cases, where datasets were publicly available, methods developed in this work have been quantitatively compared with already reported results. Performance of proposed methods have been found to be better than existing methods in terms of recognition accuracy, time taken to compute the model and addressing challenges of sensor noise and indeterminate description of situations.

The proposed schematic is considered to be scalable towards increase in number of sensors, contexts and situations due to its modular approach. It is also scalable to multiple implementations of similar setups while extensibility to multi user in same location has been identified as future implementation and validation task.

The novelty of this research lies in providing a holistic solution of sensors to situation abstraction and improving AI & machine learning algorithms to adapt to this application. The proposed architecture is expected to have performance improvements over existing system as:

1) Unlike existing studies which partially address the sensor to situation problem, this research has worked on holistic approach of sensor processing, sensor to context computation, context to situation computation and sensor based situation service provisioning.
2) Situation computation, unlike existing black box machine learning methods, is transparent and understandable by user. Besides computation, it also detects the inefficiency of existing sensors to identify the target situations.
3) The stage wise structure of “sensors to processed data”, “processed data to contexts” and the final “contexts to situations” makes it easier to apply different artificial intelligence algorithms according to requirement of that stage.

Work done in four stages is described in more details in next section.

4. Implementation

Input to the proposed sensor to situation transformation is raw sensor data. Stage wise implementation has been done as follows:

Stage 1- Sensor Level

In stage 1, raw sensor data is processed at sensor boards. One of the objectives of processing sensor data is to remove random errors in sensor data due to noise. For continuous monitoring, sensors are deployed at various locations for sensing different parameters of monitored events and/or objects and transmit it continuously to a defined destination. As a result, the total amount of sensor data at receiver becomes quite voluminous.

Transmission and use of this data in its raw form is neither cost effective nor useful. Decreasing cost of transmission and storage requirements is another objective of local processing. Third objective is to transform raw data into a form usable at next machine learning step [26]. It involves conversion of data destined to same receiver to a uniform rate and make it finite valued.

A single method cannot solve all issues and separate techniques have been implemented to address each problem. These steps are smoothing for noise reduction, windowing and feature extraction for reducing quantity of data and discretization for bringing data to a low level of abstraction and reducing bandwidth requirements. Existing standard methods have been utilized at each step.

For smoothing simple moving average, simple moving median and weighted smoothing has been used. Window sizes ranging from duration of smallest gesture to average duration has been considered. A heuristic based on mean of minimum duration of each context has been proposed. Simple time domain features like median, mode, inter quartile range, mean, range and standard deviation have been used due to their low computation requirements. For discretization popular supervised and unsupervised methods like equal frequency, uniform binning, Proportional K Interval discretization and entropy based methods have been implemented. Prior to these transformations, missing and outlier values were removed by replacing them with last available clean value.

Performance of smoothing was determined by RMSE and exponential moving average smoothing performed best out of all. Various window sizes ranging from frequently used 500msec to twice the heuristic function were tested on autocorrelation (ACF) achieved with each size. It was found that for all sensors, the ACF converged for our heuristic size. Extracted features from segment were evaluated using normalized relative approximation error. The experiments revealed mean to be the best feature in terms of approximation. As a result of discretization, the continuous valued sensor data was converted into maximum 55 finite values as against original range of 1990 values in one of the runs. Effect of discretization was also evaluated in stage 2 while doing context classification.

Main contribution of stage 1 has been exhaustive evaluation of candidate methods and parameter combinations at various steps to obtain optimal set of processing parameters. These processing steps with best recognized parameters are to be applied at each sensor node to process locally collected data.

Stage 2- Context Level

As described in methodology, processed sensor data to Contextual Information (CI) computation is done in second stage. Defining contexts in terms of sensors and dealing with data coming from multiple types of sensor modalities where part of the data might be erroneous or missing were few challenges for this stage. Following steps were done for implementation of this stage while meeting the challenges:

1. Defined generic contexts relevant to any underlying application.
2. Developed method for mapping set of available sensors to act as sources for each context by agglomerative clustering approach.
3. Implemented Bayesian Belief Network (BBN) based probabilistic generative models using existing BBN structure learning approaches in conditional independence and score & search categories.
4. Constructed a novel BBN structure search method using evolutionary computing based monkey search meta heuristic. The method was shown to have lesser computation time and better accuracy than one of the best existing algorithm.
5. Implemented mechanisms to use developed models for inference of contexts from complete/incomplete available test sensor data.
6. Demonstrated utilization of developed model to handle erroneous characteristics of received sensor data.

For mapping sensors to contexts, Kiplings method [27] of “Why, When, Where, Who, What and How (W5H1)” investigation about current surroundings has been used. Some relevant interpretations of these questions and their answers have been discussed in Table 2.

Available sensors can be manually categorized to extract context in each of the above category. Hierarchical clustering has also been used to algorithmically identify clusters of correlated sensors and map a cluster to each context. BBN has been chosen for model implementation after qualitatively comparing its suitability with respect to other alternate discriminative models like Neural Networks, Support vector machines and instance based classifiers like decision trees and nearest neighbor classifiers.

Table 2: Interpretation and Relevance of W5H1

Interrogative Word	General Interpretation	Relevance in Sensor based context determination
What	What specifically? What next?	What is being Used? What is being done? Recognizing activities, interactions, harness spatio-temporal relations
Where	Where is someone?	Providing a spatial frame
When	When did something start/end? When did that happen?	Providing a time frame for associations to determine context
Why	Why does that happen? Why not?	Association of actions with other actions, identification of tasks and behaviour patterns
Who	Who is doing this? Who is present?	The actors or subjects being monitored
How	How does it work? How does it relate? How many? How much?	Tracing the information flow through multiple modalities, recognizing expressions, movements, gestures.

For learning BBN model from data, two categories of methods have been proposed in literature. First of these learn independence relations among variables to eliminate edges from a fully connected graph. Representative PC and BNPC algorithms were studied in this category. However, on implementation it was found that these algorithms do not scale well to large number of variables. Thus, other category of methods using score and search paradigm was focused. In these, Bayesian score biased to minimum description length is used to assign a score to a candidate graph.

Exhaustively searching best or optimal graph is a NP hard problem. Existing algorithms thus apply heuristics to simplify the searching procedure of optimal graph. A heuristic method based on bio-inspired evolutionary algorithm of monkey search termed as MS2L (Monkey Search Structure Learner) has been proposed in this work. The proposed method was compared against existing algorithms like K2, Minimum Weighted Spanning Tree (MWST), hill climbing and greedy search. MS2L finds optimal graph in terms of Bayesian information criteria (BIC) score and takes lesser computation time as compared to these algorithms.

After finalizing network structure, conditional probabilities were calculated from given dataset. BBN model is used for inference of contexts by calculating the context value for which the posterior probability distribution is maximized on given sensor data.

Evaluation of MS2L and other approaches described above was done on sensor and non-sensor datasets. Good results were obtained in terms of accuracy and computation time taken to infer the contexts.

Accuracy of classification on non-sensor health related benchmark dataset ranging from 76% to 100% was achieved with proposed model as compared to 57% to 98% with other methods like naïve bayes classifer, hybrid bayes classifier and neural networks.

On benchmark sensor dataset, where locomotion detection of a human being was considered as context and many accelerometer based sensors tied to various body parts were used as source of processed data, accuracy ranging from 65% to 72% was obtained using our method on reduced data (in local processing) and an average 84% for original data rate. This was similar to 84% accuracy using nearest neighbor classification on original data reported in [18]. However, for noisy data the proposed model outperformed the existing methods. On this data the model obtained 80% accuracy as against about 74% with nearest neighbor classifier. BBN is a preferred model due to its better model building and maintenance mechanism as well as comparable performance.

Proposed model has also been used for online processed sensor data cleaning. Missing and outlying values have been replaced by appropriate values considering posterior probabilities calculated by model. It was found through analysis on realistic imperfect datasets that cleaned sensor data results in better performance of context extraction. Context detection model advanced is applicable to any new application with its own set of target contexts and corresponding sensor data.

<u>Stage 3- Situation Level</u>

Once contexts are extracted from sensors data then situation detection mechanisms get isolated from sensor complications. Contexts like locomotion, objects being used, gestures, location of the activity etc. individually represent just one aspect of monitored environment. In 3rd stage of implementation, problem of utilization of extracted contextual information in generating high level complete picture as user situations has been addressed.

A situation has been formalized as conjunction of a set of concurrently occurring contexts. For example, description of a person X's contextual information like location - "in kitchen", "using microwave and freezer," accessing "utensils and cupboards," can be mapped to overall situation "preparing breakfast".

The situation interpretations are currently dominated by either learning based data centric methods or knowledge based methods most of them being specification based blackboard systems. Decision trees, Hidden Markov Models, Support Vector Machines are examples of first category. Variants of rule based systems like fuzzy rules, case based reasoning, ontological reasoning and template based recognition fall under second category [28]. First category of methods doesn't have a transparent inference mechanism while second one cannot handle uncertainty. In this stage, a transparent user understandable situation representation and recognition approach has been developed. The technique also addresses uncertainties associated in defining situation in terms of existing contexts.

The problem of situation recognition has been addressed through following tasks:

1) Developing concept lattice based model to represent situations in terms of Contextual Information (CI).

2) Demonstrate use of model for recognizing situations from available CI using varying training data and in multiple sensor based environments.

3) Enhance model using fuzzy logic and more generalized neutrosophic logic to penalize performance of non-deterministic activities and reward the deterministic ones with purpose of deducing ineffectiveness of existing sensors and contexts to define certain situations.

Formal Concept Analysis (FCA) has been used as basis for construction of flexible and efficient situation computation model to extract situations from contexts. FCA is a method based on applied lattice and order theory, used for creating *concept hierarchy* from a collection of objects and their properties. Each concept in the hierarchy represents the set of objects sharing the same values for a certain set of properties; and each sub-concept in the hierarchy contains subset of objects in the concepts above it [29]. For situation representation and recognition, situations are considered objects and relevant contexts are considered their properties. FCA based model was implemented for third party datasets [30].

Overlapping contexts define situations in our proposed framework. The presence of a context doesn't thus guarantee a specific situation. This indeterminism in situation definition has been dealt with in this stage. Concept lattice structure has been modified using neutrosophic logic to address this issue. Situation inference in such structure is obtained using truth, indeterminacy and falsity membership of each context in each possible situation. Situations are recognized from set of concepts using Jaccard's similarity coefficient. Situation with maximum membership score is output as current situation. Equal weights have been assigned to truthness and indeterminacy for calculation of final membership score.

The developed models were tested on context and situation dataset. Situations in this set were five *activities* of daily routine and five types of *Contexts* for which corresponding ground truth was available. The contexts were locomotion, objects used with left hand, objects used with right hand, actions done by left hand and actions done by right hand [31]. Precision of up to 100% was achievable with fuzzy and neutrosophic concept lattices for correctly distinguishable activities.

Neutrosophic lattice was particularly useful in highlighting indistinguishable activities like *"Cleanup"* as all the objects used during cleanup are also used during other activities. The reported accuracy here went down to as low as 10% due to non-determinism factor, while for conventional and fuzzy concept lattice it was not so. Overall accuracy was also compared with other available results on J48, Hidden naïve bayes and nearest neighbour classifiers. Difference in accuracy was less than ±5% for deterministic activities and up to 10% for "*Cleanup*" activity specifically.

The models also support situations to be outputted in sets of exact, most probable and likely situations if complete CI is not available. Experiments also show the transferability of learnt model to any new application instance without any learning phase. Situation recognition methods described here can be useful in more complex situations like remote border monitoring and livestock tracking etc.

Stage 4- Service Level

After development of computation model, the last link in making sensor data available from sensing fields to end user sitting anywhere in the world has been dealt with [32]. Stage 4 focuses on Internet compatible data access constructs to make sensor data and its interpretations available over web in a standardized manner. Methodology adopted here to create such a service is shown in Figure 3.

To be compatible with recent web standards, it has been proposed to provide sensor data and abstractions based on it, that is, contexts and situations as semantic web based service to all registered users on demand.

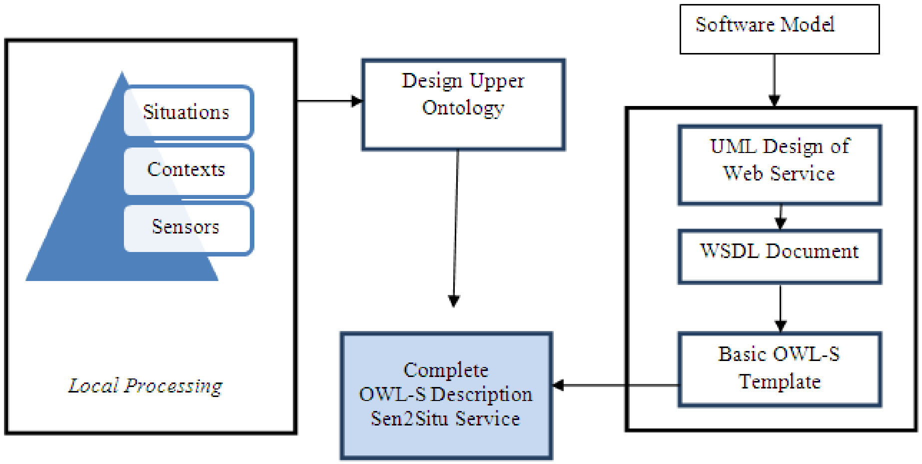

Figure 3: Methodology for Obtaining Sensor Service Description

An upper ontology for organization of concepts related to sensor data, contexts and situations has been suggested. The ontology is then further mapped to a service provider web service. Existing web standards are leveraged and customized to achieve our goal.

Service providing devices like local sink, context gathering devices, end user device and sensors have been modeled using Unified Modeling Language (UML). Formally represented UML models have then been recasted to Web Ontology Language for Services (OWL-S) standard for semantic web through intermediate conversion to Web Service Description Language (WSDL) format. After this the sensor web service is ready to be launched. Procedures for registering to and accessing the developed services have been developed utilizing application layer web standards of Simple Object Access Protocol (SOAP) and Hypertext Transfer Protocol (HTTP). Flow diagrams for both processes of registration for first time user and access for existing user have been described.

5. Research Contributions

The objectives set for this research have been realized as following research contributions:

- Detailed investigation of contemporary research issues in WSNs and Data Handling
- Framework for hierarchical fusion of multi sensor multi rate data as Contexts and Situations
- In depth empirical study of methods of local sensor processing for low level abstraction
- Developed heuristics for online segmentation of sensor data and its evaluation
- Developed method for type and number of context identification from application data
- Development of bio-inspired meta heuristic based algorithm to obtain contexts from uncertain low level abstractions

- Developed a formal concept lattice based model to map and organize situations in terms of available Contextual Information (CI)
- Improved the lattice model to include neutrosophic membership concepts to quantize inherent indeterminism in situation description
- Develop soft inference process to reveal inadequacy of underlying contextual description in definition of target situations
- In the wider context of this problem, the use of fused sensor data as service to remote users has been conceptualized.

6. Conclusions

Sensors of several types are becoming pervasive in our surroundings. These generate streams of data about our interaction with various objects and activities. Computation methods to cater to challenges of processing and useful abstractions from data emanating from different types of sensors have been studied and developed in this thesis. Approach to the research work had been model creation, algorithm designing, analyzing, and performance evaluation on third party datasets using standard metrics. Algorithms have been developed and implemented for sensor processing, context extraction and situation recognition. A service architecture for implementation of remote WSN as a service has been conceptualized.

The results obtained for context classification and situation recognition using techniques proposed in this work were found to be better than some of the existing techniques evaluated. The difference in performance was upto 10-15% in terms of recognition accuracy specifically on noisy data.

Some interesting open problems also surfaced during research. Implementation in diverse real time scenarios and applications and enhancement of existing methods to multi-user cases etc are few to mention. Exposure of proposed methods to these cases will strengthen their validity. Studies with respect to enhanced scale in terms of number of users and coverage has been identified to be of future interest.

This research and its outcomes will be beneficial in tele monitoring applications aimed at providing patient-centered, timely and location - independent healthcare. Simple activities of daily living play a vital role in determining functional abilities and independence of a person.

7. Organization of the Thesis

The thesis has been organized in eight chapters. A brief outline of contents of each of the chapters is as below:

Chapter 1 gives the introduction to the research area and its challenges. Motivation of choosing the research area and objectives set for research has also been described in this chapter. Main contributions made through this research work have been highlighted. The chapter has been concluded with chapter wise outline of thesis.

Chapter 2 provides introduction and review of recent developments in Wireless Sensor Networks. The focus is on issues arisen in past decade. Some generic WSN issues of interest, their popular solutions and recent developments have been discussed. Mainly sensor data handling and usage has been reviewed and difficulties associated with its processing have been identified. Existing sensor solutions have also been discussed in this chapter. The outcome of this chapter has been identification of sensor data categories, compilation of generations of sensor nodes and understanding sensor data problems and processing methods.

Chapter 3 focuses on significance of context and situation awareness in general applications. Mechanisms to make contexts explicit and capture it by utilizing data from various data gathering sources and specifically sensors has been explored in this chapter. Representation, recognition and efficient & autonomous inference of situations are few of the issues addressed at higher level of situation computation. The chapter is concluded with some gaps found in sensor fusion based context and situation awareness.

Chapter 4 is about cleaning and transformation of sensor data on node itself. Objectives of processing sensor data had been to remove the imperfections of data, decrease cost of transmission, reduce storage requirements and increase the usefulness of data. The sensor data is first cleaned by smoothing and removal of missing values and outliers. The cleaned sensor data is then divided into segments of suitable size, with optimal overlapping to extract useful feature per segment. The features are then discretized to obtain an efficient representation. Third party human locomotion dataset has been used for validation and measure effect of each step on quality of data. Most optimal parameters for all steps have been learnt from training data by applying various alternative methods

In **Chapter 5**, mechanisms have been developed to fuse processed data from variety of sensors to deduce more useful information specific to application domain as current "contexts". A generative model based on Bayesian Belief Network based model to quantify relations between processed sensor data and sought contexts have been constructed from training data. Mechanisms to use developed models in extraction of correct contexts from available test sensor data have also been implemented. The model has also been put to use to detect and correct some errors in received sensor data.

In **Chapter 6**, problem of utilization of extracted contextual information in generating high level complete picture as user situations has been addressed. Situations have been formalized as highly abstract representation of current scenarios which are commensurate with human perceptions of environments. A concept lattice based model has been developed to represent situations in terms of Contextual Information. Model has also been enhanced using neutrosophic logic to output multiple situations and their associated confidence, in case of ambiguous CI. Methods to use the models for recognizing situations from available Contextual Information in fast and accurate manner have also been defined here. Good recognition results have been obtained.

Chapter 7 reports the work done in wider context of the problem of making sensor data available from sensing fields to end user sitting anywhere in the world. The chapter focuses on development of Internet compatible data access constructs using recent semantic web standards. An upper ontology for sensor data organization and fetching as per the framework of logical abstractions proposed in this thesis has been proposed. The ontology is then further mapped to a service provider web service. Service registration and usage flows have been described.

Chapter 8 summarizes the conclusions drawn from the work undertaken in this thesis and proposes some future directions in which this work can be enhanced.

References

[1] J.F. Chamberland and V. Veeravalli Venugopal, "Wireless sensors in distributed detection applications." *IEEE Signal Processing Magazine*, Vol. 24, No.3 pp. 16-25, May 2007.

[2] J. Yick,, Biswanath Mukherjee, and Dipak Ghosal, "Wireless sensor network survey," *Computer Networks*, Vol. 52, No. 12 pp. 2292-2330, 2008

[3] S.S. Iyengar and Richard R. Brooks, eds. *Distributed Sensor Networks: Sensor Networking and Applications*, Vol. 26, CRC Press, 2012

[4] A. Bauer and Y. Fischer, "Task-Oriented Situation Recognition", *Situation Awareness I*, Proc. SPIE 7709, 770906, 2010.

[5] W. Feng and Jiangchuan Liu. "Networked wireless sensor data collection: issues, challenges, and approaches." *IEEE Communications Surveys & Tutorials*, Vol. 13, No.4, pp: 673-687, Nov. 2011.

[6] A. Abbasi and Y. Mohamed, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, Vol. 30, No.14, pp. 2826-2841, June 2007.

[7] E. Fasolo et al. "In-network aggregation techniques for wireless sensor networks: a survey," *IEEE Wireless Communications*, Vol. 14, No.2, pp. 70-87, April 2007.

[8] J. Esteban et al. "A review of data fusion models and architectures: towards engineering guidelines." *Neural Computing & Applications* 14.4 (2005): 273-281.

[9] A. K. Dey, "Understanding and using context." *Personal and ubiquitous computing* 5.1 (2001): 4-7.

[10] E. Elnahrawy, "Research Directions in Sensor Data Streams: Solutions and Challenges", In DCIS Technical Report, DCIS-TR-527, Rutgers University, 2003

[11] C. Bettini et al. "A survey of context modelling and reasoning techniques" *Pervasive and Mobile Computing* 6.2 (2010): 161-180.

[12] B. A. Truong, Young-Koo Lee and Sung-Young Lee. "Modeling uncertainty in context-aware computing." *Fourth Annual ACIS IEEE International Conference on Computer and Information Science*, 2005.

[13] E. M.Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors", pp. 158-175, 2004, Springer Berlin Heidelberg.

[14] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data", In *AAAI* Vol. 5, pp. 1541-1546, July 2005.

[15] L. Sun, D. Zhang, B. Guo and S. Li, "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations", In *Ubiquitous intelligence and computing*, pp. 548-562. Springer Berlin Heidelberg. 2010

[16] A. M. Khan, et al. "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer." *IEEE Transactions on Information Technology in Biomedicine*, 14.5, p: 1166-1172, 2010.

[17] J. R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. "Activity recognition using cell phone accelerometers." *ACM SigKDD Explorations Newsletter* 12.2, pp: 74-82, 2011.

[18] H. Sagha et al, "Benchmarking classification techniques using the Opportunity human activity dataset", In *IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2011* pp. 36-40 October, 2011.

[19] U. Meissen et al. "Context-and situation-awareness in information logistics." *Workshop on Current Trends in Database Technology-EDBT*, Springer Berlin Heidelberg, 2005

[20] K. Thirunarayan, C. Henson and A. Sheth, "Situation awareness via Abductive Reasoning from Semantic Sensor data: A preliminary report," *International Symp. on Collaborative Technologies and Systems*, pp.111-118, May 2009.

[21] L. Dahlbom, G. Niklasson and A. Loutfi, "Towards Template-Based Situation Recognition," in *Proc. of Intelligent Sensing, Situation Management, Impact Assessment and Cyber-Sensing*, SPIE Vol. 7352, 2009.

[22] Lu-An Tang, Yu Xiao, Kim Sangkyum et al., "Multidimensional Sensor Data Analysis in Cyber-Physical System: An Atypical Cube Approach," *International Journal of Distributed Sensor Networks*, Vol. 2012, Article ID 724846, 19 pages, 2012. http://dx.doi.org/10.1155/2012/724846

[23] A. Srivastav, W Yicheng, E Hendrick, I Chattopadhyay, A Ray and S Phoha, "Information Fusion for Object &Situation Assessment in Sensor Networks," *In Proceedings of American Control Conference* (ACC), pp.1274-1279, 2011.

[24] O. Brdiczka, L. James, and Patrick Reignier, "Learning situation models in a smart home." IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 39, No., pp. 56-63, 2009.

[25] M. Kokar, Christopher J. Matheus, and Kenneth Baclawski. "Ontology-based situation awareness," Information fusion, Vol. 10, No.1, pp. 83-98, 2009.

[26] D. Figo et al. "Preprocessing techniques for context recognition from accelerometer data." Personal and Ubiquitous Computing, Vol. 14, No.7, pp. 645-662, 2010.

[27] R. Kipling, The elephant's child, Just so stories. In Rudyard Kipling. New York: Exeter Books, 1986.

[28] J. Ye, S. Dobson and S. McKeever, "Situation Identification Techniques in Pervasive Computing: A Review", *Pervasive Mobile Computing*, pp 36-66, Vol. 8, Issue 1, February 2012.

[29] C Carpineto, G Romano, Concept Data Analysis: Theory and Applications, John Wiley, 2004.

[30] T V Kasteren, G Englebienne and B Kröse, "Transferring Knowledge of Activity Recognition Across Sensor Networks", *In Proc. of the 8th international conference on Pervasive Computing*, pp 283-300, 2010.

[31] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition

[32] A. Sheth, C. Henson and S. S. Sahoo, "Semantic Sensor Web," *IEEE Internet Computing*, pp. 7883, July/August, 2008

Authors Publications Out of This Research Work

Journals

1. Sangeeta Mittal, Alok Aggarwal, and S. L. Maskara, " Online Cleaning of Wireless Sensor Data Resulting in Improved Context Extraction" in International Journal of Computer Applications, Vol. 60 Number 15, pp. 24-32, Dec 2012. *(Impact Factor: 0.82) (H5 Index: 24 H5 Median: 35)*

2. Sangeeta Mittal, Krishna Gopal and S.L. Maskara, "A Versatile Lattice Based Model for Situation Recognition from Dynamic Ambient Sensors", International Journal on Smart Sensing and Intelligent Systems, pp. 403-432, Vol. 6, No. 1, Feb 2013. *(Indexed in Scopus) (H5 Index: 5)*

3. Sangeeta Mittal, Alok Aggarwal, and S. L. Maskara, "Contemporary Developments in Wireless Sensor Networks" , International Journal of Modern Education and Computer Science, pp. 1-13, Vol.4, No. 3, April 2012. *(Indexed By DOAJ, EBSCO, Index Copernicus and CrossRef) (H5 Index: 7 H5 Median: 13)*

Conferences

1. Sangeeta Mittal, and S. L. Maskara, "Towards Development of Wireless Sensor Network Services Based on Semantic Web ", in Proceedings of IEEE INDICON 2010, INDIA, Dec 2010 (Indexed in Scopus, EI Compendex, IEEE Explore) *(H5 index -11 , h5 median – 12)*

2. Sangeeta Mittal, and S. L. Maskara, "A Review of Some Bayesian Belief Network Structure Learning Algorithms", in Proceedings of 8th International Conference on Information, Communications and Signal Processing, Singapore, Dec 2011 *(Indexed in Scopus, EI Compendex, IEEE Explore) (H5 Index: 12 H5 Median: 17)*

3. Sangeeta Mittal, Alok Aggarwal, and S. L. Maskara, "Application of Bayesian Belief Networks for Context Extraction from Wireless Sensors Data", in Proceedings of 14th International Conference on Advanced Communications Technology, South Korea, Feb 2012. *(Indexed in Scopus) (H5 Index: 19 H5 Median: 24)*

4. Sangeeta Mittal, Alok Aggarwal, and S. L. Maskara, "Situation recognition in sensor based environments using concept lattices", in Proceedings of the International Information Technology Conference -CUBE '12 held at Pune, pp. 579-584. *(Indexed in Scopus and DBLP) (H5 Index: 6 , H5 Median: 10)*

5. Sangeeta Mittal, Krishna Gopal, and S. L. Maskara, "Preprocessing Methods for Context Extraction from Multivariate Wireless Sensors Data - An Evaluation", *in Proceedings of IEEE INDICON 2013*, INDIA, December 13-15, 2013. *(Indexed in Scopus) (H5 index -11 , h5 median – 12)*

6. Sangeeta Mittal, Krishna Gopal, and S. L. Maskara, "Effect of Choice of Discretization Methods on Context Extraction from Sensor Data–An Empirical Evaluation." 10th International Conference on Distributed Computing and Internet Technology, held at KIIT Bhubneswar, Springer International Publishing, 2014. pp: 146-151. (Indexed in Scopus)

7. Sangeeta Mittal, Krishna Gopal, and S. L. Maskara, " A Novel Bayesian Belief Network Structure Learning Algorithm Based on Bio-Inspired Monkey Search Meta Heuristic", in Proceedings of International Conference on Contemporary Computing, IC3 2014 , JIIT Noida. (Indexed in Scopus) (H5 index -10 , h5 median – 15)

(Sangeeta Yadav)	(Prof. Krishna Gopal)
Research Scholar	Supervisor -1