Towards Modelling and Reasoning about Uncertain Data of Sensor Measurements for Decision Support in Smart Spaces

Oluwaseun Bamgboye
School of Computing
Edinburgh Napier University
Edinburgh, United Kingdom
O.Bamgboye@napier.ac.uk

Xiaodong Liu
School of Computing
Edinburgh Napier University
Edinburgh, United Kingdom
X.Liu@napier.ac.uk

Peter Cruickshank
School of Computing
Edinburgh Napier University
Edinburgh, United Kingdom
P.Cruickshank@napier.ac.uk

***Abstract*— Smart Spaces currently benefits from Internet of Things (IoT) infrastructures in order to realise its objectives. In many cases, it demonstrates this through certain automated applications that relies on sensor streams that comes with some uncertainties in measurements. However, these sensor data tend to be uncertain or fault-prone due to the faults of the sensor either themselves or the wireless sensor networks. Sometimes, the extreme operating condition of the sensor can be a contributing factor to the uncertainty. The proposed approach provides a software framework that aims at homogenising, annotating and reasoning over these data. The framework consists of four layers that utilizes the semantic process involving a domain ontology and reasoning process to deliver improved quality data streams to applications. This will allow for early detection of missing data points and enhancing the accuracy of decisions and actions in such spaces.**

Keywords*— *Ontology, C-Sparql, Reasoning, Sensor, Data Stream, Smart Space

I. Introduction

Smart Spaces are~~-~~ a part of the smart cities ecosystem, consisting of a number of connected sensor operated devices that are responsible for continuous gathering and exchange of large volume of data in real time. These data provide~~s~~ valuable insights concerning a given phenomenon within the monitored space. A significant issues is that the poor quality of these data has a tendency of influencing the type of outcomes of intelligent applications and decision support systems [1].

Data quality is considered to be fitness for use of a particular data/dataset within a context of interest[2]. Often, the transition of these data from physical quantities (e.g. temperature readings) to digital space (such as software systems) is done with little consideration of poor quality of data resulting from various uncertainties. Existing approaches to addressing this issue places more emphasis on the adoption of statistical techniques and the method of any of the point calibrations [3] built with the sensors by the manufacturer. However, such calibration levels are not adequate in dealing with quality issues relating to data inconsistency, untrustworthiness and missing value/data point [4] of data streams in ~~spatio~~spatial-temporal setting.

The approach proposed by this work introduces a software solution through an architectural framework to resolve the uncertainties that results from sensor failure and other inconsistencies in readings. The principal element of this approach is semantic technology. It considers a specific domain ontology for smart spaces, which adapts from previous sensor ontologies while emphasizing on the sensor measurements. In addition, a non-monotonic reasoning system [5] is included to take advantage of the processing window of the sensor data stream. The contribution of this approach is to improve the accuracy of automated systems and other intelligent applications that heavily relies on sensor readings to form an insight or make certain decisions about such data. It also takes the advantage of improving the discoverability of the sensor measurements or reading during certain automated processing.

The remainder of the paper is as follows: section 2 provides the major objectives for Smart Spaces. Section 3 detailed description and taxonomy of the data quality problems within the smart spaces. List of quality requirements is in section 4 and section 5 contains the description of the proposed framework for the solution of the quality problem. Section 6 presents an overview of the related work. Section 7 concludes and presents the future direction of the work.

II. SMART SPACE OBJECTIVES

Smart Space represents an open pervasive environment that forms one of the major sources of big data as well as member of the smart cities ecosystem. It promotes the availability of essential services through data and knowledge acquisition concerning an environment while improving user experience and decisions [6]. One of the aim of smart space is to achieve a ubiquitous environment with embedded sensors and computers to enhance human activities and decision-making process through strong collaboration of pervasive digital devices and Internet services. Smart space is also required in the development of service construction for end-user applications and computing environments of Internet of Things(IoT) [7]. It is however important to consider the data

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

quality as part of the essential requirement that will guarantee the defined objectives of Smart Spaces.
requires to be semantic driven. In essence it provides the description of the resources in the space with their semantics,

Furthermore, as smart space is a form of service-oriented system that delivers user's need through service construction, it

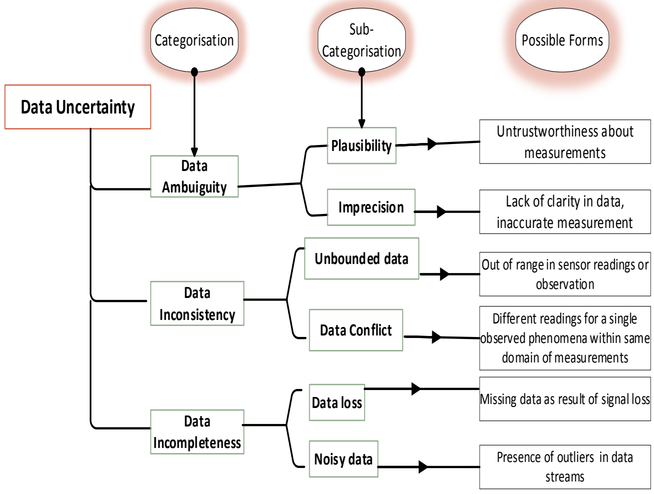

Fig. 1: Taxonomy of Data Uncertainty

III. DATA QUALITY ISSUES

The desire to improve the service construction of smart spaces continues to grow as the cost of sensors continues to remain relatively cheap with assurance of achieving an efficient ubiquitous computing environment. The high volume data produced in this environment is mainly generated from the sensor readings. These sensor readings have inherent uncertainties. A measure of these uncertainties is considered transitive in our proposed taxonomy of data quality problems within smart spaces. In this project, a taxonomy of the uncertain data in smart space is defined as shown in Fig. 1. The taxonomy shows data uncertainty along the dimension of vagueness, inconsistency and incompleteness of sensor measurements. Inconsistency and incompleteness in data is consider to occur at the lower layer of the sensor readings while Vagueness is a quality problem often experienced at the higher contextual level of smart space applications. Earlier study [8] associate quality problems relating to imprecision, inconsistency and noise to random errors. This type of errors exists distinctively in the smart spaces. For example, inconsistency can arise when two sensors of the same type deployed within same space gives different readings. We define this case of problem as Data Conflict in our proposed taxonomy. In a similar perspective, investigation by [9] suggest incompleteness in sensor reading is mostly caused by sensor failure and data loss. The accuracy and reliability of sensor measurement is dependent on the technique for managing the plausibility in sensor reading to improve smart city and smart space experience [8]. However, going by these rising concerns in terms of quality of sensor measurements, the semantic approach promises a holistic solution towards addressing the problems.

IV. DATA QUALITY REQUIREMENTS FOR SMART SPACES

Data remains an essential ingredient for some category of automated systems such as the decision support systems and critical systems that heavily depends on sensor readings in delivering the services. It is therefore important to note that there are some data quality requirements to be satisfied in order to deliver a good quality services. These automated systems must then ensure data meet one or more of the following quality requirements, defined based on review of literatures. It is also considered to be essential as part of the smart space components.

Accessibility refers to the availability and ease of data retrieval within the space despite the heterogeneous or multimodal characteristics of devices and data. The current deployment of sensors in smart space only allows individual sensors to generate their~~its~~ own data format unilaterally without a common access to all multimodal data. It will be necessary to provide an infrastructure for common retrieval of homogeneous sensor data and independent of individual protocol.

Accuracy is the degree at to which sensor reading actually represents the measured phenomena and the extent to which values *v* from sensor readings belongs to a closed interval of $-\beta \leq v \leq \beta$, for the absolute systematic error β. It tells how closely the output or sensor reading from an instrument or device corresponds to its 'true' value. For example, a sensor calibrated at < ± 0.1% of measurement will mean the actual reading will be applied to ± 0.010 units of measurement or less. This means that any variation between the 'true' values is referred to as 'error'

Completeness is the extent to which the sensor readings is not missing and is sufficient in the ratio of the breadth and depth of the data within a particular stream window. For example, an active sensor may suffer from intermittent signal failure thereby registering wrong values for this data point. It is therefore imperative to ensure the missing data points are not delivered as part of complete realistic measurements

Consistency refers to the extent to which data produced by different sensor is available in the same format without undue repetitions for subsequent collaborative or intelligent processing. As an illustration, a multiple number of sensors of the same type and measuring the same phenomena can be placed in the same domain while providing the individual readings to a specific application. It important for each instance of measurement to be uniquely identified prior its consumption by application

Interpretability refers to the degree of the appropriateness and clarity of data in terms of meaning and format. It is noticed not all sensor interprets reading in understandable forms e.g. some sensor interprets measurements as "0" and "1" while interpret using integer and real number values. In addition, it will make more sense if the meaning of this measurement is part of the processing.

Plausibility is the extent to which a given sensor reading is consider true and acceptable or credible for a given measurement. For instance, it may be important to ascertain that a sensor such as temperature sensor is actually giving the reading of the space and not the temperature of human body within the space.

Timeliness represents the difference in the present timestamp of sensor reading and the registering timestamp during measurement. The interval of time between when sensor measures the real-life event and is available for use by other smart space components should be minimal.

V. THE PROPOSED FRAMEWORK

To achieve improved intelligent applications such as decision support systems, we present a four-layer software architectural framework depicted in Figure 2. The framework is designed with the vision of integrating heterogeneous data and IoT compatibility in mind. The framework considers the smart space as an instance of adaptive system capable of gathering and integrating data from different sensors (Mostly embedded in heterogeneous devices) to enrich sensor measurements, make sense of the data and, support decision making systems.

In specific detail, the data abstraction layer consists a number of heterogeneous sensors that produce real time data representing different phenomena of interest such as space temperature, dew point, humidity, pressure etc. These data are processed in conjunction with the modeling and integration layer to generate an enhanced data stream in form of quadruples. The resulting stream need to be aggregated, filtered and reasoned over at the reasoning layer to produce a reliable data stream. The intention to provide support for producing new knowledge about data as well as making more sense of the data to improve quality of decisions and actions by applications.

Data Abstraction Layer consists of the aggregator and the anomaly detection engine. The aggregator collects raw data stream from physical and virtual sensors in the space using the Global Sensor Network (GSN) middleware [11]. The anomaly detection engine takes it from this point using the static knowledge base to perform preliminary data stream filtering on data points with quality related problems.

The Modelling and Integration Layer provides a platform for the interoperability and integration of heterogeneous data and devices in the smart space. It achieves this by implementing a domain ontology, which adapts classes from Semantic Sensor Network (SSN) ontology [12] and Observation and Measurement Schema[13]. The resulting ontology is the Smart Space Uncertainty Model (SmartSUM). Our model also includes specific devices and components common to smart space environment. Further objective of this layer is to provide semantically enriched annotated quadruples such as the one in [14], which is a data stream format enhanced with individual timestamps. This layer also make provision for ontology reuse, which represent one of the objective of the Semantic Web. This data format contains additional property that represent the individual timestamps of each data stream. The layer also provides conceptual knowledge about data streams to the immediate upper layer to enhance the reasoning process.

Reasoning Layer is the other essential component of the framework that performs the non-monotonic reasoning task as required for qualitative automated applications. The idea of the non-monotonic in reasoning is the ability of the system to automatically retract from previous conclusion when the premise changes. It relies on the domain expert knowledge with predefined rules and the semantic annotated data streams from the previous layer. The reasoning engine takes the advantage of the Continuous SPARQL query language [15] to capture the processing window of the semantic streams. This is further processed using Jena rule Language based on domain expert knowledge about identified real life events common to the smart space environment. This layer is capable of performing the complete reasoning functionalities on various aspect of uncertainties as defined in the taxonomy.

The Application Layer contains intelligent or automated application programs that rely on sensor generated data streams in defining certain actions or events within the space. Such tasks are common to applications like the decision support systems, critical systems and other data sensitive applications. These applications will rely on the services provided by the immediate lower layer, which will be available through a common Application Programming Interface (API).

VI. RELATED WORK

There has been a risen interest in providing a common accessibility to sensor data and its observation. This has provided the platform for a large number of applications in various

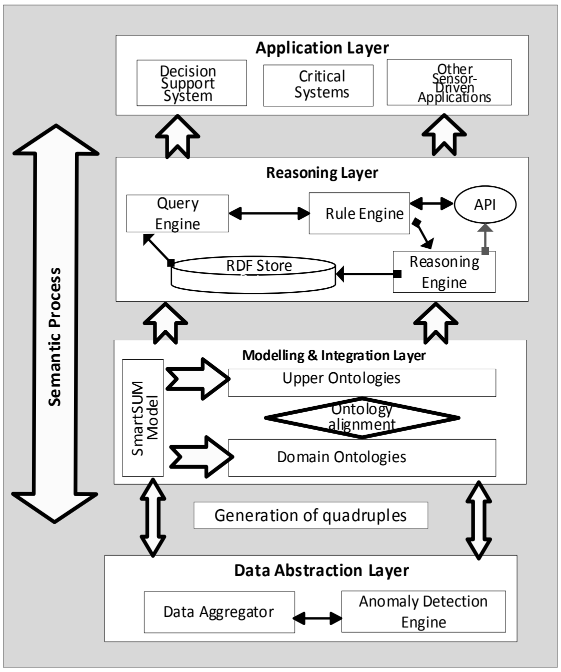

Fig. 2: The Proposed Framework for tackling sensor Data Uncertainty with Semantic Modelling and Reasoning

domains including the Smart Spaces, Geographical Information Systems, Smart health and smart cities as a whole. The Open Geospatial Consortium (OGC) provide a standard for modelling sensor data through the introduction of Sensor Model Language (SensorML) [16], Observation and Measurement Schema [13] and Sensor Web Enablement (SWE) [17]. These approaches to data modelling are built on Extended Mark-up Language (XML). The major limitation of the approach is its inability to support interoperability and provide semantics to the heterogeneous data and devices. It is also considered not suitable for providing support for reasoning task in a knowledge driven application. To achieve the vision of the new pervasive computing and smart applications, the use of semantic technologies is seen to be gaining relevance towards the realization of this vision.

Modelling sensor stream using semantic technologies involving the use of ontology has proven to be realistic in domains such as SSN ontology[12], CSIRO[18], SWAMO[19]. The leading standard among these innovations is the SSN ontology developed by the semantic working group. It contains detail description of sensor types and instruments. However, it does not provide the description of the sensor reading and void of reasoning support for high level IoT applications[20]. Our preliminary investigation~~s~~ shows that most ontology developed for sensor driven applications in the past only focused on specific aspects of domain without careful consideration for smart spaces and uncertainties in their data. Many of the ontologies [21][22][23] presently developed only considers accuracy as being independent of sensor measurement. The term accuracy here remains implicit in terms of data quality and uncertainty perspective. In addition, their perspective of accuracy is condition dependent, which is based on specific requirements of the domain being modelled. Although some of these ontologies are able to capture the real time measurements of the sensor readings, the quality of the data is not within the scope of the models.

Furthermore, the use of SPARQL extensions in querying and reasoning over sensor data stream is one of the recent contributions to the field of semantic technologies [24],[15]. The aspects of reasoning contributed to semantic technology but many of the techniques implemented with this approach remains non-retractable. The World Wide Web Consortium also contributed by introducing some rule languages such as the Semantic Web Rule Language (SWRL), Pellet reasoner, Fact++. Most of the reasoning systems developed only performs reasoning on ontologies but not the data. The initial attempt to address the limitation involve the use of a hybrid approach [25]. In principle, a hybrid approach will require layering ontologies with non-DL rules such as Production rules, Complex Event Processing (CEP), Logic Programming, etc. This method always requires translating the base ontology into a corresponding formalism of the rule system. The effect of this approach is that it can cause loss of information as result of the rules and ontology being treated separately. It is also possible that inferences will not be possible based on the separation of the ontology and the rules. The suggestion to consider the combination of rule language with stream querying was initially put forward in [26]. An initial prototype to realize this vision is called the StreamRule System [5] which specifically targets the Semantic Web. The system integrates Continuous Query Evaluation over Linked Stream (CQELS)[27] with Answer Set Programming (ASP) syntax to achieve the non-monotonic reasoning. The limitation of StreamRule is that the configuration of the system is specified in XML and fail to support the processing of historical data. Furthermore, the attempt to combine the continuous SPARQL with the Semantic Web Rule Language (C-SWRL) [28] to enhance non-monotonic reasoning is currently subject to investigation under the Water Quality Management project. Though the reasoning system is yet to be validated against standard reasoning engine, it is apparently limited in capabilities to the domain of water quality management without consideration for uncertainty in the sensor measurement. It will still be necessary to consider an appropriate non-monotonic system for data quality management and smart spaces. To the best of our knowledge, combining the Jena Rule Language with C-SPARQL is yet to be subject to consideration.

VII. CONCLUSION AND FUTURE WORK

The framework described in the previous section of this paper can effectively contribute to the accuracy of smart space, considered part of the smart city ecosystem. The framework can leverage the real time sensor stream to deliver a very good quality measurement to associated systems such as the decision support system.

In this paper, an identification of uncertainties attributed to sensor readings and measurements in relation to how it influences the quality of decision or actions is clearly established. It follows a transitive taxonomy of quality related problems that can result from the sensor measurements and the effects on applications that consumes the data. The framework presented is currently being developed with initial focus on the development and deployment of a generic ontology for smart spaces. The development is currently following an iterative method with the aim of validating it against a number of use cases.

Semantic technology has recently provided new waves of breakthrough for smart city applications and IoT, which enables possible integration of devices and interoperability of data. In fact, IoT usage or deployment will require more automated systems. However, to enhance the feasibility and accuracy of those systems, a good and standardized framework as the one already illustrated in this work is essential. The future work will consider the introduction of a distributed non-monotonic reasoning engine into the framework. The validation of the reasoning components against previous stream reasoning systems that utilizes C-SPARQL as part of the internal component will need to be established.

REFERENCES

[1] A. Karkouch, H. Mousannif, H. Al Moatassime, and T. Noel, “Data quality in internet of things: A state-of-the-art survey,” *J. Netw. Comput. Appl.*, vol. 73, pp. 57–81, 2016.

[2] L. L. Pipino, Y. W. Lee, and R. Y. Wang, “Data quality assessment,” *Commun. ACM*, vol. 45, no. 4, p. 211, 2002.

[3] B. E. Last and A. M. Utc, “calibrating-sensors,” 2016.

[4] A. Klein and W. Lehner, “Representing Data Quality in Sensor Data Streaming Environments,” *J. Data Inf. Qual.*, vol. 1, no. 2, pp. 1–28, 2009.

[5] A. Mileo and M. Hauswirth, “StreamRule: A Nonmonotonic Stream Reasoning System for the Semantic Web,” in *International Conference on Web Reasoning and Rule Systems*, 2013, vol. 7994,

Formatted: Justified

no. July, p. pp 247-252.

[6] D. G. Korzun, A. M. Kashevnik, and S. I. Balandin, “The Smart-M3 Platform : Experience of Smart Space Application Development for Internet of Things,” *Internet Things, Smart Spaces, Next Gener. Networks Syst. ruSMART 2015,Springer.*, vol. 3, pp. 56–67, 2015.

[7] D. G. Korzun, S. I. Balandin, and A. V. Gurtov, “Deployment of Smart Spaces in Internet of Things : Overview of the Design Challenges Deployment of Smart Spaces in the Internet of,” *Balandin, S., Andreev, S., Koucheryavy, Y. NEW2AN 2013 ruSMART 2013. LNCS, vol. 8121, pp. 48–59. Springer, Heidelb.*, pp. 1–12, 2013.

[8] E. Elnahrawy and B. Nath, “Cleaning and querying noisy sensors,” *Proc. 2nd ACM Int. Conf. Wirel. Sens. networks Appl. - WSNA ’03*, p. 78, 2003.

[9] Y. L. Tan, V. Sehgal, and H. H. Shahri, “Sensoclean: Handling noisy and incomplete data in sensor networks using modeling,” *Main*, pp. 1–18, 2005.

[10] D. Kuemper, T. Iggena, M. Fischer, R. Toenjes, and E. Pulvermueller, “Monitoring data stream reliability in smart city environments,” *2016 IEEE 3rd World Forum Internet Things*, pp. 571–576, 2016.

[11] C. Perera, A. Zaslavsky, P. Christen, A. Salehi, and D. Georgakopoulos, “Capturing sensor data from mobile phones using Global Sensor Network middleware,” *2012 IEEE 23rd Int. Symp. Pers. Indoor Mob. Radio Commun. -*, pp. 24–29, 2012.

[12] M. Compton *et al.*, “Web Semantics : Science , Services and Agents on the World Wide Web The SSN ontology of the W3C semantic sensor network incubator group,” *Web Semant. Sci. Serv. Agents World Wide Web*, vol. 17, pp. 25–32, 2012.

[13] OGC, “Observation ans Measurements,” *http://www.opengeospatial.org/standards/om*, 2008. [Online]. Available: http://www.opengeospatial.org/standards/om. [Accessed: 20-Jul-2007].

[14] G. D’Aniello, M. Gaeta, and F. Orciuoli, “An approach based on semantic stream reasoning to support decision processes in smart cities,” *Telemat. Informatics*, no. July, pp. 0–1, 2017. *In Press*

[15] D. F. Barbieri and P. Milano, “Querying RDF Streams with C-SPARQL,” *ACM Trans. Auton. Adapt. Syst. - Spec. Sect. Form. methods pervasive Comput. pervasive Adapt. self-adaptive Syst. Model. algorithm*, vol. 7, no. 1, 2014.

[16] E. M. Botts and C. A. Robin, “Open Geospatial Consortium Inc . OpenGIS ® Sensor Model Language (SensorML) Implementation Specification Document type : Document subtype : Document stage : Document language : OpenGIS ® Publicly Available Standard Implementation Specification,” *Open Geospatial Consort. Inc.*, 2007.

[17] M. Botts, G. Percivall, C. Reed, and J. Davidson, “OGC (R) Sensor Web Enablement: Overview and High Level Architecture,” *Lect. Notes Comput. Sci.*, vol. 4540, no. December, pp. 175–190, 2007.

[18] M. Compton, H. Neuhaus, K. Taylor, and K. N. Tran, “Reasoning about sensors and compositions,” *CEUR Workshop Proc.*, vol. 522, no. December, pp. 33–48, 2009.

[19] K. J. Witt *et al.*, “Enabling Sensor Webs by Utilizing SWAMO for Autonomous Operations,” in *8th NASA Earth Science Technology Conference.*, 2008.

[20] T. K. Duy, G. Quirchmayr, A. Tjoa, and Hoang Huu Hanh, “A Semantic Data Model for the Interpretation of Environmental Streaming Data,” *Inf. Sci. Technol.*, pp. 1–5, 2017.

[21] R. T. Mcmullen Rick, “The Common Instrument Middleware Architecture (CIMA) Instrument Ontology & Applications,” in *In Proceedings of 2nd Workshop on Formal Ontologies Meet Industry(FOMI 2006), Trento, Italy*, 2006, pp. 655–663.

[22] M. Calder, R. A. Morris, and F. Peri, “Machine reasoning about anomalous sensor data.,” *Ecol. Inform.*, vol. 5, pp. 1–19, 2009.

[23] M. Compton, “The Semantic Sensor Network: Ontology A Generic Language to Describe Sensor Assets.,” in *Proceedings of the AGILE 2009 Pre-Conference Workshop Challenges in Geospatial Data Harmonisation, Hannover, Germany*, 2009.

[24] D. Anicic and P. Fodor, “EP-SPARQL: a unified language for event processing and stream reasoning,” *Proc. 20th …*, pp. 635–644, 2011.

[25] E. Jajaga, L. Ahmedi, and L. Bexheti, “Semantic Web Trends on Reasoning Over Sensor Data,” *8th South East …*, no. July 2015, 2013.

[26] E. Garoufallou, R. J. Hartley, and P. Gaitanou, “An Expert System for Water Quality Monitoring Based on Ontology,” *Commun. Comput. Inf. Sci.*, vol. 544, no. July, 2015.

[27] D. Le-phuoc, “A Native and Adaptive Approach for Linked Stream Data Processing,” in *International Semantic Web Conference*, 2013, vol. 1380, no. February, pp. 370–388.

[28] E. Jajaga and L. Ahmedi, “C-SWRL: SWRL for Reasoning over Stream Data,” *IEEE 11th Int. Conf. Semantic. Computing.*, pp. 395–400, 2017.