Heterogeneous Data Management on Environmental Sensors Using Ontology Mapping

Kaladevi Ramar and T. T Mirnalinee
Department of Computer Science & Engineering, SSN College of Engineering
Chennai, India
Email: kala_ramar@yahoo.com, mirna123@gmail.com

Abstract—**Wireless sensors are used for monitoring environments such as climate change, water quality and disaster management. The critical task is interpretation of the complex multidimensional data generated by sensors and interoperability among various system and handling of heterogeneous data. There is also need to get a common understanding and decision making by various agencies and research institutions. To handle these requirements importance is given to both syntactic and semantic nature of the data. The syntactic part is responsible for the interoperability of the sensor data. Interoperability is addressed by using Open Geospatial Consortium (OGC) Sensor Web Enablement (SWE) framework has developed a suite of specifications enable sensors to be accessible and controllable via the web. Ontologies are widely used as a means for solving the information heterogeneity problems because of their capability to provide explicit meaning to the information. In this paper we propose a Service Oriented Architecture (SOA) model for the interoperability of sensor networks data and semantics of the data. The heterogeneities between various sensor networks are reduced through ontology mapping. The service is available to the users through standard data formats like XML and protocol like SOAP.**

Index Terms—**wireless sensor networks, semantic web, data management, environmental sensors, service oriented architecture (SOA)**

I. Introduction

Wireless sensor networks are composed of a huge number of sensor nodes, each node is providing its data. Each sensor system is meant for one application and various formats of data and interfaces are used for these systems. It leads to heterogeneity of data and it reduces the interoperability among related applications. The issues related with sensors are interoperability of several monitoring systems and heterogeneous data. The challenges are related with abstraction level, quality of sensor data, integration and fusion of data, identification and location of relevant sensor based data sources and rapid development of applications.

The Service Oriented Architecture (SOA) is a flexible system to handle different tasks and provide services. Service Oriented Architecture allows many clients to search and find services they need from multiple providers. Open Geospatial Consortium has developed Sensor Web Enablement, which is a standard for developing sensor webs. This standard provides the resources and the guidance needed when developing a set of sensors interconnected using the web.

Various agencies or systems are providing weather forecasting information. The formats of data, naming of parameters, measurement units are different for different networks. To reach a common understanding and interoperability of data, we need to reduce the heterogeneity of data thereby semantics is provided over the data. Each system or network which provides weather data is represented using an ontology structure. Ontology represents the best efforts of the technical community to unambiguously capture the definitions and interrelationships of concepts in a variety of domains.

The heterogeneity of sensor data is minimized by ontology mapping. Ontology mapping is the process of finding semantic correspondence between similar elements belonging to different ontology. The services like each weather forecasting systems and its interoperability and the ontological representation are available through Service Oriented Architecture's registry Universal Description, Discovery and Integration (UDDI). The user can access the service through Web Services Description Language (WSDL) interface and Simple Object Access Protocol (SOAP). A knowledgebase is a collection of Ontology Web Language (OWL) statement about resources. Then querying language is used to retrieve information from the knowledge base.

The paper is organized as follows; Section II provides brief information about sensor web related works in various domains. Section III explains the proposed system architecture and its components in detail. Section IV concludes the proposed work and gives directions for future use.

II. Related Works in Data Management of Envirnomental Sensors

Amit Sheth. *et al.* [1] proposed Semantic Sensor web (SSW) where sensor data is annotated with semantic metadata to increase interoperability and provide contextual information for situational knowledge. A semantic sensor data would contain spatial, temporal, and thematic information essential for discovering and analyzing sensor data. The SSW framework is tested over weather data using complex queries.

Manuscript received June 11, 2013; revised August 28, 2013.

doi: 10.12720/lnit.1.4.154-158

The NASA's [2] advanced information systems technology program planned to infuse the sensor web concepts into space missions and data systems. The goal is to employ new data acquisition strategies and systems for integrated earth sensing that are responsive to environmental events for both applications and scientific purposes.

Scott M. Fairgrieve *et al.* [3] proposed a sensor web framework PULSENet to provide for the discovery, access, use and control of heterogeneous sensors, their metadata, and their observed data. PULSENet is used an intermediary role with other sensor, processing and decision support system. The system is to integrate a different type of real world sensors over the internet to show the feasibility of a standards based, interoperable sensor web.

Trina Myers *et al.* [4] proposed Smart Environmental Monitoring and Analysis Technologies (SEMAT) project to develop advanced wireless sensor networks to collect, store, process and interpret coastal data. The aim of the project is to build a smart wireless sensor architecture that allow the user to use their existing hardware and new technologies within a multi-scale monitoring system that allow them to interrogate and task aspects of the sensor network according to need and monitoring outcomes. This will allow feed monitoring data into intelligent analytical tools for structuring the results and further decision making.

Oscar Corcho *et al.* [5] focused some of the most important challenges in sensor-based data management and sensor network configuration. The challenges are related to the abstraction level, integration and fusion of data, quality, identification and location of relevant sensor-based data sources, rapid development of applications. The challenges are being addressed by means of semantic-based methods, techniques and technologies.

Surya S. Durbha *et al.* [6] developed a tool for coastal buoys and stations. To address the needs sharing of information in a uniform way and achieve interoperability syntactic standardization and semantic enrichment of the information sources are required. A coastal semantic mapping (COSEM-Map) tool developed as a part of this work facilitates the harmonization of different representations. Both web based and mobile based clients were developed to process syntactic and semantic queries.

Harshal patni *et al.* [7] developed a framework to make the sensor data openly accessible by publishing it on the Linked Open Data (LOD) cloud. This is accomplished by converting raw sensor observations to RDF and linking with other datasets on LOD and allows large amount of sensor data are accessible for utilization and analysis. Linked Data used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the semantic web using URIs and RDF.

Flora Amato *et al.* [8] proposed semantic enriched data model for sensor network interoperability. This paper propose a architecture for the interoperability of the sensor networks, which is based on web service technologies and on a common data model enriched with semantic concepts and annotations. The architecture for risk management is completely based on services available to authorized end users and/or applications that need them.

A Flexible Data and Sensor Planning Service for Virtual Sensors [9] proposed a service oriented framework to schedule, plan and empower diverse sensors and heterogeneous ordering systems. The framework contains Open Geospatial Consortium (OGC) compliant Sensor Planning Service (SPS), a Web Notification Service (WNS), Sensor Observation Service (SOS) and virtual sensors. This paper contains two important technologies namely a flexible SPS middleware and an asynchronous message notification mechanism. The results shows the system has improvements in a uniform planning service for more satellites, a seamless connection with data order system , and a flexible service oriented framework for virtual sensors. The framework's next step is to study how to use Geo Processing Workflow (GPW) technology to integrate all the operations of Data and Sensor Planning Service (DSPS) and how to implement multi satellite sensor collaborative planning at the service level.

The above mentioned systems are focused on semantic annotation of sensor data and semi automated ontology mapping tool for ocean data. In this paper semi automated ontology mapping tool along with Service Oriented Architecture is proposed for weather forecasting data.

III. Proposed Architecture

Service Oriented Architecture [10] is a form of distributed computing where service providers publish their services to a registry where they may be easily discovered by clients who need them. This registry contains information on how to connect to the service and what data the service needs to operate.

Service Oriented Architecture allows many clients to search and find the services they need from multiple providers.

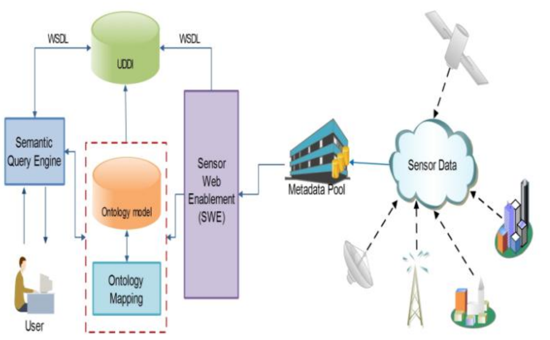

Figure 1. Proposed system architecture.

To manage the data coming from heterogeneous environmental sensors, we propose a system in Fig. 1 based on web services and OGC SWE standards. The sensor data is either received from real time sensors like pressure sensor, humidity sensor or collected from data providing agencies. Fig. 1 shows the general framework for handling efficient management of weather related sensors data. For example weather forecasting is collectively done by temperature, pressure and humidity sensors.

The Meta data pool contains the information about the sensor data. Sensors measuring phenomena are usually in binary or proprietary formats. There is a need of metadata to manage the sensor data. A good semantic sensor network should provide spatial, temporal, and thematic information for analyzing and discovering sensor data.

The service providers need to represent their services in a standard way. The Open Geospatial Consortium has developed Sensor Web Enablement (SWE), which is a standard for developing sensor webs. The communications between the components in the architecture are through Simple Object Access (SOAP) protocol and Web Service Description Language (WSDL) interface. SOAP is a XML based protocol to applications exchanges their information over the web. The reason to use SOAP is language and platform independent. WSDL is a XML based document describes a web service. It also specifies the location of the service and the operations or methods the service exposes. UDDI is used along with SOAP and WSDL to support web services. A UDDI registry contains categorized information about businesses and the services that they offer, and it associates those services with technical specifications of the Web Service. These technical specifications are usually defined using WSDL. A Web Service consumer queries the UDDI registry to find the WSDL descriptions to determine how to use the Web Service. The UDDI specification defines an API based on SOAP messages, with a WSDL description of the registry service.

A. Interoperable Web Services

The central component of the model is the SWE [11]. SWE is an interoperability framework for accessing and utilizing sensors and sensor systems via Internet and Web protocols. The suite supports three languages and four webs service interfaces. These are Sensor Model Language (SensorML), Transducer Model Language (TransducerML) Observations and Measurements (O&M), Sensor Observation Service (SOS), Sensor Alert Service (SAS), Sensor Planning Service (SPS) and Web Notification Service (WNS).

B. Semantics Through Ontologies

The Semantic Web describes the relationships between things and the properties of things. The Semantic Web is an abstract representation of data on the Web.

The central idea of the Semantic Web is to extend the current human-readable web by encoding some of the semantics of resources in a machine accessible form. Computers will be better able to search, process, integrate and present the content of these resources in a meaningful, intelligent manner. The core technological building blocks are ontology languages, flexible storage and querying facilities, reasoning engines, etc.

Data from various forecasting systems use different formats and names. To retrieve knowledge and get interoperability of more sensor networks will be an issue. The difference may in naming scheme, measuring units and temporal measures. For example temperature measurement is specified with the name temp or min temp and max temp or low temp and high temp. Relative Humidity is a name used in some domains and Humidity in other domains. Pressure is mentioned as mean sea level pressure or barometer or barometric pressure in various systems. Pressure is measured either as millibar (mb) or hectopascals (hpa). Similarly wind speed is measured as kilometer per hour ((km/hr) or (km/h)) or knots (Kts) These kind of heterogeneities need to be solved to get an interoperable services.

Ontology [2] is a formal representation of a domain, composed of concepts and named relationships. Ontology provides a common understanding of specific domains that can be communicated between people and application systems. Fig. 2 explains two ontology models for two different forecasting systems. First ontology model belongs to WeatherBug and the second model is belongs to Centre for Development of Advanced Computing (C-DAC) systems. Domain ontology is ontology about a particular domain of interest.

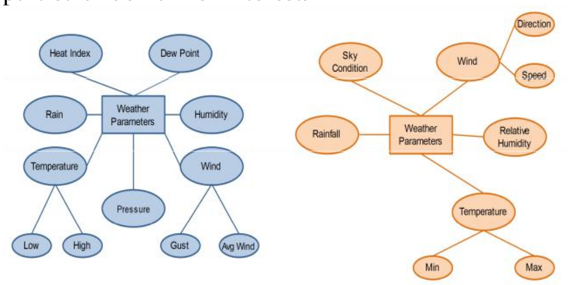

Figure 2. Domain ontology.

C. Ontology Mapping

Ontology mapping is the process of finding semantic correspondence between similar elements belonging to different ontologies [12]. It is also defined as for each entity in ontology1 find a corresponding entity in ontology2. This method improves exchange of knowledge and interoperability among various sources.

The reason to use ontology mapping is to reduce heterogeneities. The mismatches in the representation of the ontology are language level mismatches and ontology level mismatches [13]. In language level mismatches are due to syntax, expressiveness and semantics of the primitives. In ontology level mismatches are the same term describing different concept, different terms describing same concept, different point of view etc.

The commonly used mapping techniques are probabilistic approaches, machine learning, graph based methods, reasoning and theorem proving. Manual mapping is labor-intensive and error-prone process. A Semi-automatic machine learning technique proposed in [14] which create semantic mappings between ontologies by using joint probability distribution function. Machine learning-based methods are used for finding ontology mappings based on the similarity of instances in the ontologies.

D. Ontology Mapping Using Instance Based Algorithm

Ontology specifies a conceptualization of a domain in terms of concepts, attributes, and relations. The concepts in the ontology model are entities of interest in the domain. Each concept in the model is associated with a set of instances. The ontology matching is finding semantic mapping between ontologies. The simple way of matching is one-to-one (1-1) mapping between the elements. For example the concept Humidity is maps to Relative Humidity. Similarly low temperature is maps to minimum temperature and high temperature is maps to maximum temperature.

The instance based algorithm takes two input ontologies and produce mapping pairs from both the inputs. In machine learning approach statistics of the data, instance and values of information are taken into account .The instance is sample data (30 days data) for each concept in the input ontology is to be collected and trained. Fig. 3 shows the flow of instance based mapping technique.

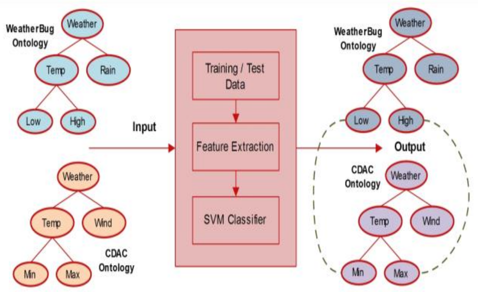

Figure 3. The flow of instance based mapping technique.

The feature extraction task extracts a set of informative features from the input data that are necessary for achieving better classification accuracy. Kernel PCA is a nonlinear form of principal component analysis proposed to efficiently compute the principal components in high-dimensional feature spaces which are non-linearly related to the input variables [15].

KPCA was used first for nonlinear feature extraction then to evaluate its effect on a subsequent classification in combination with learning algorithms such as support vector machines. The combination of KPCA and SVM could identify patterns with more accuracy compared to the combination of PCA and SVM [16].

A support vector machine (SVM) is a supervised Learning technique from the field of machine learning applicable to both classification and regression. The Support vector machine (SVM) is a popular and powerful technique for classifying any kind of data [17]. SVM learning machine seeks for an optimal separating hyper plane that separates objects belonging to different classes. The margin of the hyper plane is called support vectors. The key features of SVMs are the use of kernels, the absence of local minima, the sparseness of the solution and the capacity control obtained by optimizing the margin.

The advantage of incorporating KPCA with SVM is reduced computational burden on SVM. The mapping is performed on the classes, properties, data instances of the ontology's. A decision will be taken based on the output from Support Vector Machine.

Table I shows the mapping results for 5 parameters based on instance based algorithm. CDAC parameters are in the row and WeatherBug parameters are in the column. The value indicates the matching percentage between parameters of two ontologies. The Max Temp in CDAC system has 86% probability matching with Hi-Temp of WeatherBug system.

TABLE I CONCEPTS SIMILARITY IN TWO DIFFERENT ONTOLOGIES

Classes	Max Temp	Min Temp	Rainfall	Wind Speed	Relative Humidity
Hi Temp	0.83	0.43	0.01	0.01	0.03
Lo Temp	0.46	0.86	0.01	0.01	0.02
Rain	0.01	0.01	0.78	0.01	0.01
Avg wind	0.01	0.01	0.01	0.04	0.01
Humidity	0.01	0.01	0.01	0.01	0.62

IV. CONCLUSION AND FUTURE WORK

This paper focused to design ontology based framework using semiautomated machine learning technique to reduce heterogeneities among various weather forecasting systems and improve interoperability between the systems. This work will extended to other forecasting networks to get a common consensus for the research group and other agencies. The same approach is applicable for other sensor domain, which has heterogeneous information.

REFERENCES

[1] A Sheth, C. Henson, and S. Sahoo, "Semantic sensor Web," *IEEE Internet Computing*, pp. 78-83, July/August 2008.

[2] K. Moe, "NASA technology for the earth observation sensor Web," in *Geoscience and Remote Sensing Symposium*, pp. 128-131, 2008.

[3] S. M. Fairgrieve, J. A. Makuch, and S. R. Falke, "PULSENet: An implementation of Sensor Web standards," *International Symposium on Collaborative Technologies and Systems*, pp. 64-75, 2009.

[4] T. Myers, I. Atkinson, and R. Johnstone, "Semantically enabling the SEMAT Project: Extending marine sensor networks for decision support and hypothesis testing," in *Proc. International Conference on Complex Intelligent and Software Intensive Systems*, 2010, pp. 974-979.

[5] O. Corcho, R. Garcia-Castro *et al.*, "Five challenges for the semantic sensor Web," *Semantic Web-Interoperability, Usability, Applicability*, 2010, pp. 121-125.

[6] S. S. Durbha, R. L. King, S. K. Amanchi, S. Bheemireddy, and N. H. Younan, "Standards-based middleware and tools for coastal sensor Web applications," *IEEE Journal of Selected Topics in Applied Earth Science and Remote Sensing*, pp. 451–466, 2010.

[7] H. Patni, C. Henson, and A. Sheth, "Linked sensor data," in *International Symposium on Collaborative Technologies and Systems*, 2010, pp. 362 – 370.

[8] F. Amato, V. Casola, A. Gaglione, and A. Mazzeo, "A semantic enriched data model for sensor network interoperability," *Simulation Modeling Practice and Theory*, pp. 1745-1757, 2011.

[9] Z. Chen, N. Chen, L. Di, and J. Gong, "A flexible data and sensor planning service for virtual sensors based on Web service," *IEEE Sensors Journal*, pp. 1429–1439, June 2011.

[10] [Online]. Available: http://www.wpi.edu/Pubs/E-project/Available/E-project-101407-34528/unrestricted/report.pdf

[11] M. Botts, G. Pecivall, C. Reed, and J. Davidson, "OGC sensor web enablement: Overview and high level architecture," in *Lecture Notes in Computer Science, GeoSensor Networks*, New York: Springer, 2008, pp. 175–190.

[12] A. H. Doan and A. Y. Halevy, "Semantic integration research in the database community: A brief survey," *AI Magazine*, 2005.
[13] [Online]. Available: http://disi.unitn.it/~accord/RelatedWork/Matching/Noy-MappingAlignment-SSSW-05.pdf
[14] A. H. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Ontology matching: A machine learning approach," *Handbook on Ontologies in Information Systems*, S. Staab and R. Studer, Eds. New York: Springer-Velag, 2003.
[15] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computing*, vol. 10, pp. 1299–1319, 1998.
[16] W. Jianning, J. Wang, and L. Liu." Feature extraction via KPCA for classification of gait patterns," *Human Movement Science*, pp. 393-411, 2007.
[17] V. Vapnik, *The Nature of Statistical Learning Theory*, NewYork: Springer-Verlag, 1995.

Kaladevi Ramar Born in India on 26.05.1977. She has received the BE degree in computer science and engineering from Bharadhidasan University, India, in 1998 and the ME degree in computer science and engineering from Anna University, India in 2006. Currently doing her PhD in Anna University India.

She worked as a Lecturer in an affiliated college of Bharadhidasan university from 1999 to 2005. She was a Senior Lecturer from June 2007 to Feb 2010 and also served as an Assistant Professor from mar 2010 to Nov 2010 in an affiliated college of Anna university. Her research interests include heterogeneous data management, sensor data interoperability and semantic web. She has published papers in three international conferences and two of them are listed in IEEE Xplore

Ms. Kaladevi is a life member of ISTE.

