Expert Systems with Applications 41 (2014) 7914–7926

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An ontology-based framework to support intelligent data analysis of sensor measurements

Fernando Roda [a], Estanislao Musulin [a,b,*]

[a] Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas, Ocampo y Esmeralda, S2000EZP Rosario, Argentina
[b] Universidad Nacional de Rosario, FCEyA, Pellegrini 250, S2000BTP Rosario, Argentina

ARTICLE INFO

Article history:
Available online 27 June 2014

Keywords:
Intelligent data analysis
Temporal abstraction
Temporal reasoning
Description Logic
Ontology
Semantic sensor web

ABSTRACT

In the past years, the large availability of sensed data highlighted the need of computer-aided systems that perform intelligent data analysis (IDA) over the obtained data streams. Temporal abstractions (TAs) are key to interpret the principle encoded within the data, but their usefulness depends on an efficient management of domain knowledge. In this article, an ontology-based framework for IDA is presented. It is based on a knowledge model composed by two existing ontologies (Semantic Sensor Network ontology (SSN), SWRL Temporal Ontology (SWRLTO)) and a new developed one: the Temporal Abstractions Ontology (TAO). SSN conceptualizes sensor measurements, thus enabling a full integration with semantic sensor web (SSW) technologies. SWRLTO provides temporal modeling and reasoning. TAO has been designed to capture the semantic of TAs. These ontologies have been aligned through DOLCE Ultra-Lite (DUL) upper ontology, boosting the integration with other domains. The resulting knowledge model has a modular design that facilitates the integration, exchange and reuse of its constitutive parts. The framework is sketched in a chemical plant case study. It is shown how complex temporal patterns that combine several variables and representation schemes can be used to infer process states and/or conditions.

1. Introduction

Sensing and communication technologies, such as "wireless sensor networks", have witnessed explosive growth in the recent past. These technologies are empowering information systems from many domains such as health care, industrial control systems and environmental monitoring, to collect and store large volumes of data (Molina & Flores, 2012). In addition, the recently developed sensor web technologies enable sensor measurements from all kind of sources to be available for sharing through web services (Sheth, Henson, & Sahoo, 2008; Ye et al., 2012). However, although these data is a very valuable asset for process analysis and supervision, it is usually not properly exploited. Therefore the need for computer-aided systems that extract useful knowledge from that large amount of available data becomes evident.

Intelligent data analysis (IDA) is and emergent research field that aims at feeling the gap between data generation and data comprehension, providing an efficient mean of matching raw data to process knowledge Lavrač, Keravnou, and Zupan (1997). IDA is directed toward application of knowledge for data interpretation. It takes advantage from different tools such as statistics, pattern recognition, data mining, machine learning and data abstraction; to discover the principles that are encoded within the sensed data.

The choice of a particular data representation has a large impact on efficiency and simplicity of IDA tasks (Lin, Keogh, Lonardi, & Chiu, 2003). Qualitative representation and reasoning has proven to be an excellent practice to embrace dynamic processes complexity since it relies on abstracted views of signals behavior, instead on just raw sensor outputs. These views are temporal abstractions (TAs) that interpret past and present states and trends that are relevant for the given set of goals. TAs are interval based representations having a wide range of complexity, from relatively simple level shifts to trend compound abstractions based on combinations of more primitive abstractions (Calbimonte, Yan, Jeung, Corcho, & Aberer, 2012; Cheung & Stephanopoulos, 1990; Janusz & Venkatasubramanian, 1991; Lin et al., 2003; Meléndez & Colomer, 2001; Molina & Flores, 2012; Shahar, 1994). One important advantage of using qualitative representation (QR) for IDA is that it enables artificial intelligence symbol-based reasoning, which brings a transparent way of capturing the process condition. TAs can be interpreted by matching against predefined templates

* Corresponding author at: Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas, Ocampo y Esmeralda, S2000EZP Rosario, Argentina.
E-mail addresses: roda@cifasis-conicet.gov.ar (F. Roda), musulin@cifasis-conicet.gov.ar (E. Musulin).

http://dx.doi.org/10.1016/j.eswa.2014.06.033
0957-4174/

or guidelines (Haimowitz & Kohane, 1996; Seyfang & Miksch, 2004), or reasoned within a higher context (Bellazzi, Larizza, Magni, Montani, & Stefanelli, 2000).

Very different domains such as medicine (Stacey & McGregor, 2007) and Process System Eng. (PSE), have shown an special interest in IDA solutions. Works from these areas present different but complementary approaches for the analysis and interpretation of sensor measurements. Qualitative Trend Analysis (QTA) is an outstanding method widely studied in PSE field (Cheung & Stephanopoulos, 1990; Gamero, Melendez, & Colomer, 2011; Janusz & Venkatasubramanian, 1991; Maurya, Paritosh, Rengaswamy, & Venkatasubramanian, 2010; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003; Villez, Venkatasubramanian, & Rengaswamy, 2013; Villez et al., 2013). QTA can identify a set of basic trends in the measured variable by looking at it derivatives signs. We call these kind of method *shape-based* since they depend on the shape of the observed signals. Note that these methods do not make use of domain knowledge in the abstraction process.

However, in medicine, IDA approaches rely more on heuristic knowledge (Esfandiary, Babavalian, Moghadam, & Tabar, 2014; Stacey & McGregor, 2007) that is captured by different representation schemes ranging from elaborated ontologies (Shahar, 1994; Shahar & Musen, 1996) to more simple point schemadata (Seyfang & Miksch, 2004; Seyfang et al., 2001). These techniques involve taking patient raw time-stamped data and using domain knowledge to generate temporal abstractions; such as "severe anemia for 3 weeks in the context of administering the drug AZT". Shahar (1994) presented one of the first works in that matter; the author calls it the "Knowledge-based Temporal abstraction theory (KBTA)". The core of KBTA is a set of five inference mechanisms (Temporal Context Formation, Contemporaneous Abstraction, Temporal Inference, Temporal Interpolation, Temporal Pattern Matching) supported by an ontology with 5 concepts (Primitive parameters, Events, Contexts, Abstract Parameters, and Patterns). KBTA was implemented in a computer program called RÉSUMÉ. It uses a frame-based languages to formalize the ontology concepts, and inference mechanisms are encoded with external rule-based tools. A shortcoming of this approach is that the knowledge specification and the inference mechanisms are decoupled. Furthermore, these formalisms are not sufficiently expressive to represent complex domain knowledge and temporal entities. In RÉSUMÉ, inferences are also bounded by constraints such as the *closed-word* and the *unique-name* assumptions that may lead to wrong deductions.

O'Connor, Hernandez, and Das (2011a) deal with these issues by means of semantic web technologies, which take advantage of Description Logic (DL) reasoning (Krötzsch, Simancik, & Horrocks, 2012). In particular, the authors make use of OWL[1] (Web Ontology Language) and SWRL[2] to integrate the specification and querying components of the KBTA methods. They showed how SWRL built-in functions are suitable to implement the five KBTA inference mechanisms, including a method for adding TAs on line. However, besides the OWL based implementation, the authors do not introduce changes to the original lightweight ontology of KBTA.

As shown, the above IDA approaches are developed to work with specific TA schemes and therefore they are not flexible enough to support different time series representations and abstraction levels. On the other hands, none of these efforts have considered an important goal of IDA: to extract knowledge from different data sources.

To address this issues, in this work an ontology-based framework to support intelligence data analysis of sensed data is presented. It relies in a novel knowledge model composed by four ontologies: Temporal Abstractions Ontology (TAO), Semantic Sensor Network ontology (SSN) (Compton et al., 2012), SWRL Temporal Ontology (SWRLTO) (O'Connor & Das, 2011) and DOLCE Ultra-Lite (DUL) (Masolo, Borgo, Gangemi, Guarino, & Oltramari, 2003). The framework uses temporal reasoning to search and classify qualitative temporal patterns, that help to infer the process state or condition.

The knowledge model is able to manage several TA schemes; both, knowledge-based temporal abstractions and shaped-based temporal abstractions are supported. This not only brings flexibility, but also enhances the analytical skills of the tool, as dynamic properties can be analyzed by combining different abstract views of the process.

Another important feature of the proposed framework is the integration with SSN that enables a full compatibility with sensor web technologies. Through it, the knowledge base have access to information about sensors and sensor observations from all kind of sources.

DOLCE Ultra-Lite has been employed as the upper-level ontology of the proposed framework. As a featured foundational ontology, DOLCE eases the understanding of the model and boosts future integrations with a large amounts of domains ontologies that are based on it.

Finally, we present a set of ontological correspondences for a full semantic alignment between the aforementioned ontologies.

The paper is organized as follows. Section 2 describes the main components, features and requirement of the proposed IDA framework. Section 3 introduces the ontologies used. In Section 4 these ontologies are semantically aligned to form a novel knowledge model. In Section 5, the framework is illustrated in a simple example in the PSE domain. Finally, in Section 6 concluding remarks are given.

2. Ontology-based framework for IDA

Any IDA system must cope with three interrelated issues: data validation, data representation and data interpretation. In this work, these issues were met by a novel IDA framework that takes advantage of the latest semantic technologies.

Fig. 1 depicts the work-flow in the proposed framework. The IDA process starts with the sensing tasks over the dynamic system under study. It is usually implemented by automatic sensor devices, but it can also be achieved manually. In each observation, a sensor measures a system property and provides an estimated value, a time stamp and some contextual data such as a measurement quality estimation. These observation records are validated and stored in large repositories (usually implemented as time series databases). Validation involves some data pre-processing tasks such as noise reduction, outlier detection and the rectification of input errors. To properly evaluate the stored measurements, the corresponding sensors metadata must be attached to observation records. This information should include sensors precision, operation range, Eng. units, etc. Data acquisition and validation tasks are high domain-dependent and therefore they must be specifically designed for each application environment. For this reason, no implementation guidelines are proposed here for these tasks. Instead, a reusable formal knowledge model for data representation is presented (see Section 3). It provides a foundational structure for the knowledge base (KB) where all the acquired knowledge is stored. As depicted in the figure, the conceptualization is composed by a set of three main ontologies:

1. An ontology of measurements, incorporated by semantically annotating the observation records.

[1] http://www.w3.org/TR/owl-features/.
[2] http://www.w3.org/Submission/SWRL/.