The Protégé-Owl SWRLTab and Temporal Data Mining in Surgery

Guenter Tusch[1,2], Martin O'Connor[1], Timothy Redmond[1], Ravi Shankar[1], and Amar Das[1]
[1] Stanford Medical Informatics, Stanford University, Stanford, CA, USA
[2]Grand Valley State University, Allendale, MI, USA
tuschg@gvsu.edu,{martin.oconnor,tredmond,rshankar,das}@stanford.edu

Abstract. Modern researchers in medicine have access to large clinical databases. Mining those resources often includes exploration of temporal data. In liver transplantation, a researcher might be interested in patients that exhibit an unusual time pattern of clinical tests. Standard query languages like SQL do not support this kind of queries very well due to an insufficient time model. A very flexible approach is Knowledge-based Temporal Abstraction, which has been implemented in a number of proprietary systems. For increased availability in clinical research, we extended the knowledge-based temporal abstraction framework by creating an open-source platform, SPOT. It supports the statistical package R and knowledge representation standards (OWL, SWRL) using the open source Semantic Web tool Protégé-OWL.

1 Introduction

Modern researchers in medicine have access to the large clinical databases that have become more readily available recently. One important aspect of data mining those resources is the exploration of temporal data. For example, in liver transplantation, where a wealth of parameters is obtained from continuously monitored patients, a researcher might be interested to select patients or patient episodes that exhibit an unusual pattern of potential complications of the transplanted organ, each following a typical pattern in time. Standard query languages like SQL are not well suited for this kind of research because of an insufficient time model. However, a very useful and flexible approach is Knowledge-based Temporal Abstraction (KBTA), which has been implemented in a number of proprietary systems. Here time-stamped data points are transformed into an interval-based representation that can utilize Allen's approach of temporal relationships [1]. To make KBTA more readily available for clinical research, we developed SPOT, an implementation using open source and standardized tools: the Web Ontology Language (OWL; `http://www.w3.org/TR/owl-features`), the Semantic Web Rule Language (SWRL; `http://www.daml.org/2003/11/swrl`), Protégé plug-ins; `http://protege.stanford.edu/`, and open source statistical software (R; `http://www.r-project.org/`).

2 Application in Liver Transplantation

Liver transplantation is a complex and challenging surgical procedure that is followed by a complex intensive care and clinical monitoring schedule. A wealth of parameters is obtained from the patients. Potential hepatic complications follow a typical pattern in time. For instance, "acute rejection" can be characterized by increasing AST and ALT (liver enzymes) values [2], which decrease as soon as the rejection therapy is started. If there is no response to the therapy or a negative biopsy, it is not considered rejection. The phase with increasing enzymes may vary in length and range of values, or even only one enzyme may be elevated, same for the phase of decreasing values, but still the time pattern holds. In this example a few typical issues are addressed. First, clinical data come with different time granularities, hourly, daily, monthly or yearly. Second, clinical concepts can be expressed in terms of phases or intervals, e.g. increasing or decreasing enzymes, rejection therapy over 3 days, which can be consecutive or overlapping, and establish a typical pattern in time. Third, it is not so much single parameter values but the relationship of intervals that establishes the clinical concept. These aspects are captured in the valid time model as used in temporal database research and in temporal abstraction.

Temporal abstraction is a comprehensive approach to deal with time-oriented data in medicine. Time-stamped data points are transformed into an interval-based representation of data that can utilize Allen's temporal relationships [1]. Here in general a functional approach is used that maps raw data into higher-level concepts like "states" (e.g. peak, rejection therapy) or "trends" (e.g. increasing, decreasing). The goal is to represent complex medical concepts by these primitives using Allen's time relationships. Shahar [3] introduces Knowledge-Based Temporal Abstraction (KBTA). The KBTA method has a formal model of input and output entities, their relations, and the domain-specific properties that are associated with these entities - called the KBTA ontology. Shahar describes four different output types, state, gradient, rate, and pattern abstraction. States could be low or high bilirubin test levels of the transplanted liver, gradients increasing or decreasing enzymes, rates could be slow or fast, and a pattern periodic.

There are different implementations of the KBTA method all over the world. Almost all of them focus on describing individual patient courses for clinical therapeutic purposes. An overview is found in [4].

3 Implementation and Application to Liver Transplantation

For the implementation we use the open source Semantic Web tool Protégé-OWL and the SWRLTab[5] using temporal built-ins (`http://protege.cim3.net/cgi-bin/wiki.pl?SWRLTemporalBuiltIns`) and extensions built-ins (`http://protege.cim3.net/cgi-bin/wiki.pl?SWRLExtensionsBuiltIns`). Ontologies are used in OWL to formally specify meaning of annotations by providing a vocabulary of terms. New terms can be formed by combining existing ones. The Protégé-OWL plug-in allows to easily building ontologies that are backed by OWL code. OWL ontologies roughly contain three types of resources: *Classes* represent concepts from the knowledge domain (e.g., the class patient), *individuals* are specific instances of classes (e.g., the patient transplanted today), and *properties* determine the values allowed to each individual (e.g., the specific Patient has an ID).

SWRL allows users to write rules that can be expressed in terms of OWL concepts and that can reason about OWL individuals. In the application, we use a temporal ontology implementing the valid time model, and a hierarchical patient ontology with classes: Patient (has) Procedure (has) Interval/Event (has) Valid Time.

In the sequel we describe a concept of bilirubin increase from a high bilirubin level (see [2] – a sub concept of "acute rejection"), abbreviated by "High_Bili_Increase". Bilirubin is a product that results from the breakdown of hemoglobin in the liver. When bilirubin levels are high, it is a sign that the liver is incapable of adequately removing bilirubin in a timely manner. The concept is described in terms of SWRL code in figure 1, using SWRL built-ins and extensions. Each concept creates a "result interval", i.e., an interval instance in the valid time model, using the SWRL extension `createOWLThing`. This allows for modularization of time concepts, i.e., breaking down concepts into (potentially in other clinical domains re-usable) sub concepts.

Explanation	**SWRL - Code**

init variables to capture the underlying ontology	`Patient(?p) ∧`
	`hasProcedure(?p, ?proc) ∧`
	`    hasTest(?proc, ?test) ∧`
	`    hasTestName(?test, ?testName) ∧`

bilirubin increase →?tVT	`    swrlb:equal(?testName, "BILIRUBIN") ∧`
	`    HasOutputType(?test, ?testType) ∧`
	`    swrlb:equal(?testType, "INCREASE") ∧`
	`    temporal:hasValidTime(?test, ?tVT) ∧`

bilirubin high →?tVT2	`hasTest(?proc, ?test2) ∧`
	`hasTestName(?test2, ?testName2) ∧`
	`swrlb:equal(?testName2, "BILIRUBIN") ∧`
	`HasOutputType(?test2, ?testType2) ∧`
	`swrlb:equal(?testType2, "HIGH") ∧`

overlaps intervals ?tVT and ?tVT2	`temporal:hasValidTime(?test2, ?tVT2) ∧` `    temporal:overlaps(?tVT, ?tVT2, "days")∧`

save ?startTime and ?finishTime variables	`temporal:hasStartTime(?tVT, ?startTime) ∧`
	`temporal:hasFinishTime(?tVT, ?finishTime)∧`
	`swrlx:createOWLThing(?hbVT, ?proc)`

create result interval and attach to procedure	`->  temporal:ValidPeriod(?hbVT) ∧`
	`    temporal:hasStartTime(?hbVT,?startTime)∧`
	`    temporal:hasFinishTime(?hbVT,?finishTime)∧`
	`    hasHighBiliIncrease(?proc, ?hbVT)`

Fig. 1. SWRL Code for the concept "High_Bili_Increase"
(`?tVT`, `?tVT2`, and `?hbVT` are interval instances in the valid time model, `temporal:` denotes temporal built-ins based on Allen's temporal relationships, `swrlx:` denotes SWRL extensions.)

4 The Architecture

An overview over the SPOT architecture is depicted in figure 2. The researcher (user) can define clinical concepts, e.g., rejection, and then search for patients or episodes with that concept in the clinical database. Two steps are necessary to accomplish this: Training the system to learn concepts from a subset of the clinical database, and searching for the learned concepts in the entire database. The user has to perform the following tasks in order to train the system: Estimation of intervals from a learning sample, e.g., learning thresholds for a running average of the bilirubin parameter values to model an "increasing bilirubin" interval, building of high level concepts (Temporal Abstraction) (implemented in Protégé-OWL/SWRLTab), and validation of the generated intervals. The statistical tasks are written in S, an interactive environment for data analysis and at the same time a statistical programming language. An open source tool based on S is called R, which is used here for the implementation. R is under the GNU General Public License.

The user might go through that process several times until the classification error for the clinical concepts he/she models is sufficiently small. Adjustments can be made by changing thresholds or adding additional constraints to the SWRL concepts. Finally, the learned abstractions are submitted to the original database.

Besides using the time stamped data from the clinical database, the user needs to identify intervals, only one parameter at a time (e.g., bilirubin). Several different non-overlapping intervals are allowed, i.e. mark as "increasing", "decreasing", "high", etc. The interval value is attached to the time-stamped parameter value.

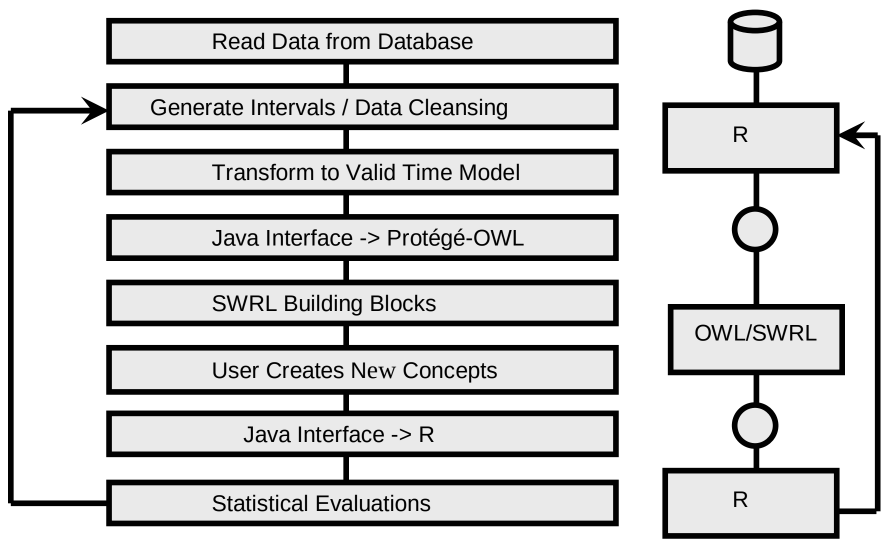

Fig. 2. The SPOT (***S*** – ***P***rotégé – ***O***WL/SWRL – ***T***emporal Abstraction) architecture.

5 Conclusion and Future Aspects

This report shows that SPOT is a feasible approach to use the Protégé-Owl SWRLTab and open source software. The next step is the development of a GUI using the R and Protégé APIs for easy access and manipulation by the user, especially creating macro like constructs that generate part of the SWRL code automatically.

Acknowledgments. We acknowledge the valuable input and support by Tania Tudorache, PhD, Jeremy Miller, PhD, Craig Webb, PhD and Mark Musen, MD, PhD.

References

1. Allen, J.F.: Towards a general theory of action and time. Artif. Intell. 23(2) (1984) 123-154.
2. Tusch, G.: Tools for Modeling Time-Oriented Concepts. Applications for Retrieval of Liver Transplantation Data. In: Talmon, J., Fox, J. (eds): Knowledge Based Systems in Medicine: Methods, Applications and Evaluation. Springer Verlag, Berlin (1991) 36-48.
3. Shahar, Y., Musen, M.A.: Knowledge-based temporal abstraction in clinical domains. Artif. Intell. Med. 8(3) (1996) 267-98.
4. Augusto, J.C.: Temporal reasoning for decision support in medicine. Artif. Intell. Med. 33 (2005) 1-24.
5. O'Connor, M. J., Knublauch, H., Tu, S. W., Grossof, B., Dean, M., Grosso, W. E., Musen, M. A.: Supporting Rule System Interoperability on the Semantic Web with SWRL. Fourth International Semantic Web Conference (ISWC2005), Galway, Ireland, 2005.