Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-300

Comparing Drools and ontology reasoning approaches for telecardiology decision support

Pascal VAN HILLE[a,1], Julie JACQUES[b], Julien TAILLARD[b], Arnaud ROSIER[a], David DELERUE[b], Anita BURGUN[a], Olivier DAMERON[a]
[a]*Inserm U936, Université de Rennes 1, 35043 Rennes, France*
[b]*ALICANTE SA, 50 rue Philippe de Girard, 59113 SECLIN, France*

Abstract. Implantable cardioverter defibrillators can generate numerous alerts. Automatically classifying these alerts according to their severity hinges on the CHA2DS2VASc score. It requires some reasoning capabilities for interpreting the patient's data. We compared two approaches for implementing the reasoning module. One is based on the Drools engine, and the other is based on semantic web formalisms. Both were valid approaches with correct performances. For a broader domain, their limitations are the number and complexity of Drools rules and the performances of ontology-based reasoning, which suggests using the ontology for automatically generating a part of the Drools rules.

Keywords. Ontology, Drools, rules, telecardiology, clinical decision support system, knowledge management

Introduction

Patients suffering from heart failure are increasingly treated with Implantable Cardioverter Defibrillators (ICD) and benefit from home monitoring. Some automatic triage of the alerts according to their emergency level is instrumental to keep up with this overwhelming flow of alerts (from zero most of the time up to as many as twenty alerts per patient per day; with an estimation of 500.000 new patients every year) efficiently. However, this is an intrinsically difficult task because the risk associated with an alert depends on multiple interdependent factors such as the patient’s medical history, his current pathologies and his current treatment. Most of these alerts are related to episodes of atrial fibrillation (AF).

The Akenaton project [1] introduces an automatic classification of AF alerts according to their severity. It hinges on the CHA2DS2VASc score, which evaluation requires domain knowledge for reasoning about the patient's clinical context.

We compared two approaches for implementing the reasoning module. The first approach is based on Drools, a business rule management system. The second approach is based on the Web Ontology Language (OWL) and the Semantic Web Rule Language (SWRL).

[1] Corresponding Author. Pascal Van Hille. Inserm U936, Université de Rennes 1, 35043 Rennes, France

1. Methods

1.1. The Akenaton prototype

To determine the severity level of the alerts, the Akenaton prototype (i) reads AF alerts issued by the patient's ICD and compute AF duration, (ii) extracts relevant data from patient's medical record with natural language processing [2], (iii) integrates alert data and patient clinical context, (iv) reasons on these data in order to determine the patient's CHA2DS2VASc score and Vitamin K antagonists (VKA) status and (v) computes the alert severity level according to AF duration, the CHA2DS2VASc score and the VKA status.

This comparison study focuses on CHA2DS2VASc score determination. We compared DROOLS and ontology reasoning modules following different axis including how the knowledge is modeled, where it is stored, the use of norms and standards, scalability and performances.

1.2. CHA2DS2VASc score

CHA2DS2VASc is a new recommendation of the European Society of Cardiology for determining stroke risk for patients with non-valvular fibrillation. The higher the CHA2DS2VASc score, the higher the risk of thrombo-embolism [3]. Table 1 presents the criteria required to compute the CHA2DS2VASc score. Reconciling the patient's data with the CHA2DS2VASc criteria consists in interpreting the data according to some domain-specific knowledge, typically represented in ontologies.

Table 1. CHA2DS2VASc score assignation points. The patient's score is the sum of the points

Criteria	Points	Criteria	Points
C: Congestive heart failure	1	S2: Stroke	2
H: Hypertension	1	V: Vascular disease	1
A2: Age>=75	2	A: 65<=Age<75	1
D: Diabetes mellitus	1	Sc: Sex category	1

1.3. Set of patients

Correctness of the score and performances were evaluated on a set of 62 patients from the Department of Cardiology at Rennes University Hospital that have a Paradym ICD.

1.4. Drools-based reasoning module

This module computes the CHA2DS2VASc score using 14 Drools Expert 5.0.1 rules and 176 concepts identified with the help of a cardiologist expert (AR). Rules are coded using a domain-specific language (DSL) composed of 52 instructions (thus rules can be expressed in English, as showed in Figure 1). DSL translates rules to Java, hiding Java complexity and allowing the definition of rule templates making maintenance and rule addition easier. Data used for reasoning comes from the database. Some occasional Java-coded post-processing was added such as the determination of the patient's age from his date of birth.

One rule assigns a value to each CHA2DS2VASc criterion (e.g.: 1 point for diabetes). For each of these criteria, a specific rule represents the condition to be verified, and updates its score, like rule “CHA2DS2VASc Diabetes” in Figure 1. The remaining rules reconcile the potential granularity gap between patients data and criteria, like rule "non-insulin dependent diabetes mellitus" that allows patients with non-insulin dependent diabetes to activate rule "CHA2DS2VASc Diabetes".

```
rule "non-insulin dependent diabetes mellitus"
      when  Patient  is  diagnosed  with  non-insulin  dependent
diabetes mellitus
      Then Patient is diagnosed with diabetes
end
--
rule "CHA2DS2VASc Diabetes"
      when Patient is diagnosed with diabetes
      Then Patient s CHADS2 criteria "D" is checked
end
```

Figure 1. Syntax and samples of drools rules

The Drools-based reasoning module is able to trace rule calls and display them as a tree which can be used as an explanation for the score. Several rules can trigger at the same time, or activate again every time an update is done in patient's file. We had to defined a rule order and to prevent CHA2DS2VASc score to compute more than once.

1.5. Ontology-based reasoning module

The ontology-based reasoning module was created in Java with OpenJDK (1.6), OWL API for OWL 2.0 (3.2.0) and Pellet OWL 2 Reasoner for Java (2.3.0). The reasoning is based on a domain ontology in description logic with SWRL rules created for the project. Protégé was used as editor for editing concepts, properties and rules. This ontology contained 840 classes defined formally with OWL DL and describing diseases, medicines and equipment from the field of telecardiology [4]. Data from each patient record were represented as instances of the ontology.

The CHA2DS2VASc score was computed using the best of 10 modeling strategies bases on Java, OWL and SWRL [5]. Only criteria D, C, H, S2, and V which potentially require reference to domain knowledge to be reconciled with patient data were evaluated with the ontology. Age and sex criteria were evaluated directly in the Java program. The sum of points is also computed directly in Java. Five SWRL rules (one per criterion to be evaluated with the ontology) were created. The Pellet reasoner was used to perform the inference. For example, the diabetes mellitus criterion is evaluated with the rule: "dia-criterion-score(?region), diabetes-mellitus(?d), patient(?p), is-affected-by(?p, ?d), quality-quale(?p, ?region) -> has-for-integer-value(?region, 1)". If an instance of "Type 2 diabetes mellitus" exists with the relation "is-affected-by" to the patient instance, because there is a subsumption relation between this concept and "diabetes mellitus" (Figure 2), the rule set to 1 data property "has-for-integer-value" of "dia-criterion-score" instance.

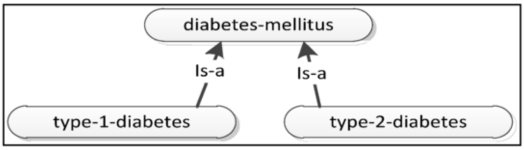

Figure 2. Diabetes mellitus ontological representation

2. Results

Table 2 compares the DROOLS and the OWL+SWRL approaches. It shows that both approaches require few primitives, that the DROOLS approach is more difficult to maintain and to extend, and that both approaches have performance of the same order of magnitude.

Table 2. Drools and ontology approach comparison synthesis

Axis	Drools approach	Ontology approach
Knowledge modeling	14 rules with 52 DSL instructions.	5 rules with 11 specifics concepts. The ontology contains 840 concepts, 185 object properties and 6 data properties.
Rules location	Rules are stored in the Java code but can be created separately with DSL.	Ontology (OWL + SWRL).
Modeling language expressiveness	Flexible and customizable (DSL). Java expressiveness. Can be complex if there are many inter-dependent rules.	Semantics and granularity handled by the ontology. Not suited for calculation. Horn clause (no disjunction).
Capacity to explain reasoning	Drools has debugging possibilities, so DSL files can be customized to provide reasoning explanation.	The pellet-explanation library was not used, thus only program trace available. However, it is possible to find which criteria were met and to see which classes were instantiated.
Knowledge editing tools	Eclipse plugin for Drools offer a suitable development environment for rules manipulation.	Protégé offer a suitable interface for concepts and properties manipulation. In contrast, SWRL rules editor is very basic.
Norms, standards	Rule language is specific to Drool.	W3C recommendations
Reusability	Ad-hoc development. Cannot be reused.	Ontology can be shared.
mean time for score determination	300 ms	150 ms

3. Discussion

The reasoning underlying the CHA2DS2VASc score determination involves both subsumption and classification. Subsumption reconciles the patient's data which are usually precise with the criteria which are more general. Classification is necessary to infer whether a pathology is a peripheral artery disease. These characteristics require advanced reasoning capabilities that favor the OWL and SWRL combination over DROOLS.

The reasoning underlying the CHA2DS2VASc score determination relies on precise and explicit criteria. Moreover, the possibility to justify the decision is important. Both reasons motivate our rules and ontology-based approach. Machine learning approaches introduce a unacceptable risk of false positive and false negative and may lack explanation capabilities.

In both approaches reasoning time is short enough for not being a discriminating factor. This aspect could be reconsidered if a large number of alerts should be classified in real time. The performance impact of adding a large number of rules has not been measured and should also be evaluated.

With a finer level of granularity, the number of rules can be much more important using Drools than using the ontology. This is because SWRL is able to take subsumption relation into account. With Drools, 3 rules would be necessary to take into account "diabetes mellitus" and all its subconcepts where only one must be created in the ontology. An important number of rules can result in a maintenance issue.

Whether the results obtained in this study are applicable to a broader domain, or one with a finer granularity remains to be studied. The Drools engine performances would probably scale up, but the number and the complexity of the resulting Drools rules would make them difficult to develop and to maintain. The ontology-based approach would benefit from modularity and from possible reuse of existing resources, but the performances would probably decrease. A direction worth exploring would consist in using the ontology for automatically generating a part of the Drools rules.

Conclusion

Drools and Ontology are two valid approaches with comparable performances for determining the CHA2DS2VASc score using domain knowledge.

Reasoning based on ontology has the advantage of being simpler to implement and to maintain since the granularity of information is managed in the ontology and not by rules. Reasoning based on Drools provides greater expressiveness than OWL + SWRL.

Acknowledgments

The AKENATON project is funded by the French Agence Nationale pour la Recherche (ANR -07-TecSan- 001).

References

[1] Burgun A, Rosier A, Temal L, Dameron, Mabo P, Zweigenbaum P, Beuscart R, Delerue D, and Henry C. Supporting medical decision in telecardiology: a patient-centered ontology-based approach. In Proceedings of the Medinfo 2010 conference; 2010

[2] Grouin C, Deléger L, Rosier A, Temal L, Dameron O, Van Hille P, Burgun A, Zweigenbaum P. Automatic computation of CH2DS2-VASc score: Information extraction from clinical texts for thromboembolism risk assessment. In: AMIA Annu Symp Proc; 2011.

[3] Camm AJ, Kirchhof P, Lip GYH, Schotten U, Savelieva I, Ernst S, et al. Guidelines for the management of atrial fibrillation. Europace. 2010; 12(10):1360.

[4] Temal L, Rosier A, Dameron O, Burgun A. Modeling cardiac rhythm and heart rate using BFO and DOLCE. Dans: International Conference on Biomedical Ontology. University at Buffalo, NY July 24-26, 2009

[5] Dameron O, van Hille P, Temal L, Rosier A, Deléger L, Grouin C, et al. Comparison of OWL and SWRL-Based Ontology Modeling Strategies for the Determination of Pacemaker Alerts Severity. Dans: Proceedings of the American Medical Informatics Association Conference AMIA; 2011.