Journal of Applied Logic 15 (2016) 130–149

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

Translation of first order formulas into ground formulas via a completion theory

Robert Demolombe*, Luis Fariñas del Cerro, Naji Obeid

Université de Toulouse–IRIT–CNRS, France

ARTICLE INFO

Article history:
Received 16 March 2015
Received in revised form 25 June 2015
Accepted 12 August 2015
Available online 17 February 2016

Keywords:
Modeling in logic
Domain
Independent formulas
Automated deduction
Application to biology

ABSTRACT

A translation technique is presented which transforms a class of First Order Logic formulas, called Restricted formulas, into ground formulas. For the formulas in this class the range of quantified variables is restricted by Domain formulas.

If we have a complete knowledge of the predicates involved in the Domain formulas their extensions can be evaluated with the Relational Algebra and these extensions are used to transform universal (respectively existential) quantifiers into finite conjunctions (respectively disjunctions).

It is assumed that the complete knowledge is represented by Completion Axioms and Unique Name Axioms à la Reiter. These axioms involve the equality predicate. However, the translation allows to remove the equality in the ground formulas and for a large class of formulas their consequences are the same as the initial First Order formulas. This result open the door for the design of efficient deduction techniques.

1. Introduction

The technique presented in this paper was originally motivated by a concrete application where we have been faced to problems of performance when we wanted to derive consequences of a logical model. The solution we found has been generalized and could be applied to many other domains. This solution required a theoretical justification which is the basic contribution of the work which is presented here.

The application is in the field of cancer research where biologists have to analyze the interactions between some proteins that can lead to the apoptosis, that is, the death of a cell that can propagate cancer.

These interactions are represented in [9,20] by a graph whose nodes represent types of proteins and the edges represent properties like: the protein y can activate the protein x (formally represented by the predicate $CA(y, x)$) or the protein z can inhibit the capacity of protein y to activate the protein x (formally represented by the predicate $CICA(z, y, x)$). A given protein may be activated (formally represented by the predicate $A(x)$) or inhibited (formally represented by the predicate $I(x)$).

* Corresponding author.
E-mail address: robert.demolombe@orange.fr (R. Demolombe).

http://dx.doi.org/10.1016/j.jal.2016.02.002
1570-8683/

The following rule is an example of a particular kind of protein interactions:

If protein y is activated and y can activate protein x and there is no protein z such that z is activated and z can inhibit the capacity of y to activate x, then x is activated.

This rule is formally represented in First Order Logic by:

(R) $\forall x \forall y((A(y) \wedge CA(y,x) \wedge \neg\exists z(A(z) \wedge CICA(z,y,x))) \to A(x))$

If we want to derive by abductive reasoning the sufficient conditions about protein activations and protein inhibitions that lead to the activation of a given protein a we have to consider the following instance of this rule:

(R1) $\forall y((A(y) \wedge CA(y,a) \wedge \neg\exists z(A(z) \wedge CICA(z,y,a))) \to A(a)$

The biologists assume that they have a complete knowledge of the graph, that is, a complete knowledge of the extension of the predicate $CA(y,x)$ and of predicate $CICA(z,y,x)$. This knowledge completeness can be formally represented "à la Reiter" by the following formulas:

(C1) $\forall y(CA(y,a) \leftrightarrow y = b \vee y = c)$
(C2) $\forall z(CICA(z,y,a) \leftrightarrow (z = d \wedge y = b) \vee (z = e \wedge y = b) \vee (z = f \wedge y = c))$

We have tried to directly derive answers to queries finding consequences of sentences like (R), (C1) and (C2) plus the axioms of equality. However, due to the properties of equality, the derivations were extremely time consuming.

The reason is that when we are not dealing with equality to derive, for example, $Q(a)$ from $P(a) \to Q(a)$ we just have to try to derive $P(a)$. However, if we have equality we have to try to derive all the atoms of the form $a = t$. If we get: $a = b_1$, $a = b_2$, $\ldots$, $a = b_n$, to derive $Q(a)$ we also have to try to derive: $P(b_1)$, $P(b_2)$, $\ldots$, $P(b_n)$. In addition for each b_i we have to try to derive all the atoms of the form: $b_i = s$. If we get: $b_i = c_{i,1}$, $\ldots$, $b_i = c_{i,n_i}$, to derive $Q(a)$ we also have to try to derive all the atoms of the form: $P(c_{i,k})$. This process must be repeated until the graph of the equality relation has been completely explored. Moreover, due to the properties of equality, this graph necessarily contains loops.

This short explanation allows to understand why derivations with equality may be extremely expensive.

Then, we have tried to remove equality predicate though the price we had to pay was to reduce the set of allowed queries. The benefit was to reduce derivations from formulas with variables to derivations from ground formulas and without equality. Even if we have not analyzed the cost of these derivations in general, in our application we have obtained fast answers (see the orders of magnitude at the end of section 3).

The intuitive idea is that, since equality is only used to express the completeness of the extension of some predicate, we just have to use the completion axioms, in a preliminary step to substitute these predicates by their extensions and then, after elimination of equality, to derive answers to queries.

That can be detailed in the context of this example as follows.

From (R1) and (C1) it can be inferred (R3) by substitution of equivalent formulas and by properties of equality. Indeed, (R1) can be rewritten as:

(R2) $\forall y(CA(y,a) \to ((A(y) \wedge \neg\exists z(A(z) \wedge CICA(z,y,a)))) \to A(a))$

From the property of substitutivity of equality a formula like $\forall y(CA(y,a) \to \phi(y))$ is equivalent to $\phi(b) \wedge \phi(c)$. Therefore we have:

(R3) $(A(b) \wedge \neg\exists z(A(z) \wedge CICA(z,b,a)) \to A(a)) \wedge ((A(c) \wedge \neg\exists z(A(z) \wedge CICA(z,c,a))) \to A(a))$