TBox Acquisition and Information Theory

Jordi Alvarez
TALP Research Center
Universitat Politècnica de Catalunya
Mòdul C6, Campus Nord, 08034 Barcelona
jalvarez@talp.upc.es

Abstract

This paper explores the application of information theory notions to description logics. The goal of the paper is to present a formal framework in which the problem of theory learning or TBox acquisition from a set of models could be defined and boarded.

TBox acquisition is formalized using model theory as a process of intensional knowledge acquisition from a set of interpretations. Closed world semantics is used in order to build ABoxes from interpretations. Interpretation entropy is then shown to be related to the number of assertions in a given ABox representing it. From this point, the amount of information provided by a TBox can be defined. And finally, this leads to TBox acquisition formalization.

1 Introduction

Some work has been done in the intersection of description logics and machine learning. Nevertheless, most of the work known to the author has been directed towards the computation of the least common subsumer (LCS) introduced in [3]; and its usage in more complex learning tasks. For example, see [4, 7]. While [8, 5] try to acquire a whole theory (using the LCS computation as a subtask).

This means that most of this work has been done in the framework of concept learning; in which we have a set of positive examples (instances of the concept to be learned), and (possibly) a set of negative examples (objects that are not instances of the concept to be learned).

In this paper I see the learning task from a different perspective: that of TBox acquisition. The problem is not a concept to be learned from a set of examples (the training set); but a theory to be acquired from a set of models.

In doing so, we are leaving the supervised "concept driven" approach and going towards an unsupervised TBox acquisition approach.

Anyway, concept learning framework could also have been used for our purposes: the theory can be seen as a concept, and a model as a theory instance. In this way, we would still be doing supervised (concept driven) learning[1]. But I prefer to provide a formal framework that is the same one that is being used to provide a formal base for DL deductive reasoning tasks.

2 Description Logics Base

The DL framework used all along this paper is the usual one (see [6] for a detailed explanation). A knowledge base Σ is defined as a pair constituted by a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$ ($\Sigma =< \mathcal{T}, \mathcal{A} >$). $\mathcal{T}$ contains the intensional knowledge and is constituted by a set of axioms. All axioms are supposed to be general concept inclusion axioms ($C \leq D$, where C and D are terms from the concept language) in which cycles may occur.

No remarks are done here in what refers to the expressivity of the concept language as the theory introduced in this paper is almost concept-language independent[2]. From now on, the concept language will be referred as $\mathcal{L}$.

A *conceptual interpretation* $\mathcal{I} = < \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} >$ is defined as a non-empty set $\Delta^{\mathcal{I}}$ (the domain); and a function $\cdot^{\mathcal{I}}$ that maps every concept in $\mathcal{L}$ into a subset of $\Delta^{\mathcal{I}}$, and every role expression in $\mathcal{LR}$ into a subset of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$[3].

Semantics for our "generic" $\mathcal{L}$ (and also for $\mathcal{T}$) are defined using the conceptual interpretation and the so called descriptive semantics, following the approach in [2] (in which cyclic TBoxes are given descriptive semantics).

For our convenience, we define:

$$Models(\mathcal{T}) \doteq \{\mathcal{I} | \mathcal{I} \text{ is a model for } \mathcal{T}\} \tag{1}$$

The extensional knowledge is provided as a set of assertions that constitute the ABox. Assertions provide information about "concrete objects" called individuals, and can be concept membership assertions ($C(a)$, where $C \in \mathcal{L}$ and a

[1]Can we conclude that unsupervised learning is in fact a special case of supervised one? At least, we can conclude that the difference between supervised and unsupervised learning is only a matter of point of view!

[2]For reasons that will be clear later, it is required that the concept language contains concept name negation.

[3]The term *conceptual interpretation* has been introduced in order to distinguish it from what will later be called *specific interpretation*. The motivation to distinguish two slightly different notions of interpretation will be clear later, when the process of learning a theory from a set of models is defined. From now on, the term interpretation will be used when there is no ambiguity or when it is not important to distinguish which notion of interpretation are we using.

is an individual) or role membership assertions ($R(a,b)$, where R is a role and a,b are individuals).

The notion of conceptual interpretation $\mathcal{I} =< \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} >$ is extended to that of *specific interpretation* in order to cope with individuals. In a specific interpretation, the function $\cdot^{\mathcal{I}}$ also maps every individual for which $\mathcal{A}$ provides information to an element of $\Delta^{\mathcal{I}}$.

For purposes that will become clearer later we define a *simple assertion* as a concept membership assertion of the form $A(a)$, or a role membership assertion of the form $S(a,b)$, where A is a concept name, S is a role name, and a, b are individuals.

ABox semantics can also be checked at [6].

3 Theory Learning

In this section we will formally define the process of theory learning from a pair $< \mathcal{T}_0, \mathcal{M} >$; where $\mathcal{T}_0$ is an initial theory we have, and $\mathcal{M} = \{\mathcal{I}_1, \mathcal{I}_2, \ldots \mathcal{I}_n\}$ is a set of conceptual interpretations. $\mathcal{T}_0$ is a subset of the theory to be learned $\mathcal{T}^*$ and is commonly referred as *background knowledge*, and the conceptual interpretations are supposed to be models of $\mathcal{T}^*$.

The reasoning procedures performed by DL systems can be seen as the process of determining the set of models of a given knowledge base; and then check some properties over those models (that corresponds to the obtained result). In the same way, theory learning can be seen as the process of determining a TBox $\mathcal{T}^*$ from a subset of its models.

Example Let's consider an example: an adaptation of the arch learning task introduced by Winston. Figure 1 shows several arches, each one of them could correspond to a conceptual interpretation in $\mathcal{M}$. Elements in the domain correspond to physical objects. The conceptual interpretations contain a set of concepts and roles that are represented in a graphical way in figure 1.

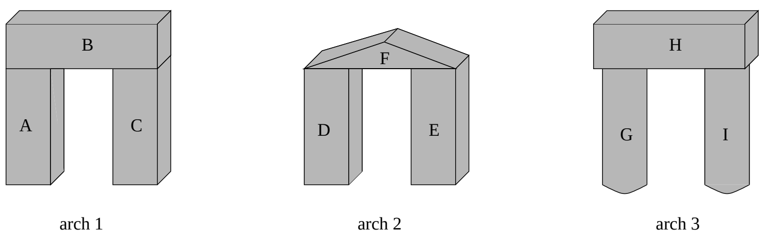

Figure 1: Three examples of arch. Each one of them corresponds to a model of the theory to be learned.

The set of used concepts are: *on_floor* (the object is on the floor), *flat_top* (the object has a flat top), *flat_bottom* (it has a flat bottom), *brick* (it is a brick), *wedge* (it is a wedge), and *cylinder* (it is a standing cylinder); and the set of roles: *supports* (relates an object to another object supported by the first one), and *touches* (relates two objects that are in contact and are at the same level (one object is not supported by the other)).

The theory to be learned ($\mathcal{T}^*$) corresponds to a theory such that $\mathcal{M}$ is a subset of $Models(\mathcal{T}^*)$, and additionally, it contains as much redundant information in $\mathcal{M}$ as it can be captured with the used concept language. Then, if $\mathcal{M}$ is representative enough for the whole set of archs, and the concept language being used is expressive enough, $\mathcal{T}^*$ would correspond to an axiomatization for arches. The concept language used in this example is $\mathcal{ALCN}$.

The initial theory $\mathcal{T}_0$ could correspond to a set of axioms trying to capture physical world laws restricted to our representation environment. Both $\mathcal{T}_0$ and a possible $\mathcal{T}^*$ are shown in table 1[4]

	$\mathcal{T}_0$ **axioms**
Φ_1	$\neg\ on\_floor \leq \exists supports^{-} \top$
Φ_2	$\exists supports^{-} \top \leq \neg\ on\_floor$
Φ_3	$\exists supports \top \leq flat\_top \sqcap flat\_bottom$
Φ_4	$brick \leq flat\_top \sqcap flat\_bottom$
Φ_5	$wedge \leq \neg\ flat\_top \sqcap flat\_bottom$
Φ_6	$cylinder \leq flat\_top \sqcap flat\_bottom$
Φ_7	$brick \leq \neg\ cylinder$
Φ_8	$cylinder \leq \neg\ brick$

	$\mathcal{T}^* \setminus \mathcal{T}_0$ **axioms**
Φ_9	$\top \leq \neg \exists touches \top$
Φ_{10}	$\neg\ on\_floor \leq \neg \exists supports \top$
Φ_{11}	$\neg\ on\_floor \leq\ \geq_2 supports^{-} \sqcap \leq_2 supports^{-}$
Φ_{12}	$on\_floor \leq \exists supports \top$

Table 1: $\mathcal{T}_0$ and a possible $\mathcal{T}^*$ for the arch learning problem.

3.1 Model Representation

In order to be able to learn from conceptual interpretations, every interpretation $\mathcal{I}$ should be represented in some way. As the interpretation provides extensional knowledge about the domain $\Delta^{\mathcal{I}}$, an easy way to represent an interpretation $\mathcal{I}$ is through an ABox $\mathcal{A}$.

A new ABox semantics will be defined in order to obtain an ABox that strictly represents $\mathcal{I}$. In order to do so, we first define an *equivalence relation* between conceptual interpretations.

We say two conceptual interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ are *equivalent* if there exists a bijection τ between $\Delta^{\mathcal{I}_1}$ and $\Delta^{\mathcal{I}_2}$ such that:

1. $C^{\mathcal{I}_1} = \{t | \tau(t) \in C^{\mathcal{I}_2}\}$ for every $C \in \mathcal{L}$, and
2. $R^{\mathcal{I}_1} = \{(s,t) | (\tau(s), \tau(t)) \in R^{\mathcal{I}_2}\}$ for every $R \in \mathcal{LR}$, and

[4]For the sake of simplicity, the arch model being used is quite simple (two supporting objects always of the same height and a supported object). In the same direction, $\mathcal{T}_0$ is also very simplistic, allowing some situations that cannot occur in our gravitatory physical world: two objects supporting each other. Or one object supporting itself, for example.

The representation $\mathcal{A}$ of an interpretation $\mathcal{I}$ should be such that $\Sigma =< \varnothing, \mathcal{A} >$ will have as models all *specific interpretations* that are $\mathcal{A}$-extensions of *conceptual interpretations equivalent* to $\mathcal{I}$[5]; and no other interpretation should be a model of $\mathcal{A}$. This assures us that only the domain objects may change in $\mathcal{A}$ models; but the behaviour of objects in the domain and the relations among those objects are completely defined by $\mathcal{A}$. τ relates objects with the same behavior.

For computational and storage reasons, we define a new semantics called *closed-world semantics*. Closed-world semantics uses *close world assumption* in order to avoid the need to have *negative* concept assertions. We say that a specific interpretation $\mathcal{I} =< \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} >$ is a *closed-world model* of $\Sigma =< \mathcal{T}, \mathcal{A} >$ if:

1. $\mathcal{I}$ is a model of Σ, and
2. For every concept name $A \in \mathcal{L}$, and for every individual a that appears in $\mathcal{A}$:

   $a^{\mathcal{I}} \notin A^{\mathcal{I}}$ if $< \mathcal{T}, \mathcal{A} >$ has a (standard) model $\mathcal{I}_0$ in which $a^{\mathcal{I}_0} \notin A^{\mathcal{I}_0}$
3. For every role name $P \in \mathcal{L}$, and for every pair of individuals a, b in $\mathcal{A}$:

   $(a, b) \notin P^{\mathcal{I}}$ if $< \mathcal{T}, \mathcal{A} >$ has a model $\mathcal{I}_0$ such that $(a^{\mathcal{I}_0}, b^{\mathcal{I}_0}) \notin P^{\mathcal{I}_0}$

The second and third points are only defined for concept names because the rest of the work is done by the semantics of $\mathcal{L}$ constructors.

Note that with this definition, consistent knowledge bases may have no closed-world model. Take for example, $\Sigma =< \mathcal{T}, \mathcal{A} >$, such that $\mathcal{T} = \{\neg A \leq B\}$, and $\mathcal{A} = \{R(a, b)\}$. Then, Σ has a model $\mathcal{I}_0$ in which $a^{\mathcal{I}_0} \notin A^{\mathcal{I}_0}$; and another one $\mathcal{I}_1$ in which $a^{\mathcal{I}_1} \notin B^{\mathcal{I}_1}$. But it has no model $\mathcal{I}$ in which $a^{\mathcal{I}} \in \neg A^{\mathcal{I}} \wedge a^{\mathcal{I}} \in \neg B^{\mathcal{I}}$.

From this point on, we will use the term *closed-world ABox* for ABoxes for which we assume closed-world semantics; and *closed-world knowledge bases* for knowledge bases with a closed-world ABox. And we will refer to the set of models of a closed-world knowledge base $\Sigma =< \mathcal{T}, \mathcal{A} >$ as $Models_{cw}(\Sigma)$.

Taking all this into account, we can compute the representation of a conceptual interpretation $\mathcal{I}$ as a closed-world ABox $\mathcal{A}$ in the following way:

1. For every object o_i in $\Delta^{\mathcal{I}}$ we define a new individual a_i. We also define a function $\cdot^{\mathcal{R}}$ that maps every object o_i to its corresponding individual a_i.
2. For every object $o \in \Delta^{\mathcal{I}}$ and every concept name $A \in \mathcal{L}$ add the assertion $A(o^{\mathcal{R}})$ if $o \in A^{\mathcal{I}}$.

[5]The term $\mathcal{A}$-extension is used here in the sense that a conceptual interpretation is converted to a concrete interpretation by adding a mapping from $\mathcal{A}$ individuals to $\Delta^{\mathcal{I}}$.

3. For every pair of objects $(o_1, o_2) \in \Delta^{\mathcal{I}}$ and every role name $S \in \mathcal{LR}$ add the assertion $R(o_1^{\mathcal{R}}, o_2^{\mathcal{R}})$ if $(o_1, o_2) \in R^{\mathcal{I}}$.

We will refer to the closed-world ABox computed from a conceptual interpretation $\mathcal{I}$ using this procedure as: *Representation($\mathcal{I}$)*[6]. Note that $Representation(\mathcal{I})$ is constituted only by simple assertions; and that $\mathcal{I} \in Models_{cw}(\mathcal{A})$, since $\mathcal{A}$ is being built from $\mathcal{I}$. Table 2 shows the closed-world ABox resulting from applying this procedure to figure 1 first arch.

3.2 Quantifying Description Logics Information

The next step in order to formalize the notion of theory learning is the quantification of the amount of information a given theory $\mathcal{T}$ is able to predict in $Models(\mathcal{T})$. Then, if we have *learned* $\mathcal{T}$ from $< \mathcal{T}_0, \mathcal{M} >$ we will be able to state how much redundant information in $\mathcal{M}$ elements is represented in $\mathcal{T}$.

Before proceeding, a new kind of ABox assertion is defined: *role non-membership assertion* $\neg\, R(a, b)$, where $R \in \mathcal{LR}$, and a, b are individuals. We extend also the notion of simple assertion to this new kind of assertions. An interpretation $\mathcal{I}$ satisfies the assertion $\neg\, R(a, b)$ iff $(a, b) \notin R^{\mathcal{I}}$. Then, we define an *extended ABox* as an ABox allowing role-nonmembership assertions; and an extended simple ABox as an extended ABox containing only simple assertions.

As will be shown (and used) later, any interpretation can be represented by a set of boolean variables. Then, the introduction of this kind of assertion allows to establish a correspondence between assertions and boolean variable values (every assertion corresponds to a boolean variable value, and every boolean variable value has a corresponding assertion)[7]. Instead, there was the choice to introduce role negation in the concept language; but different to concept negation, concept languages with role negation are not widely used. So, I have prefered not to impose such an expressivity restriction in the concept language. This new type of assertions and extended ABoxes will only be used in this section in order to quantify description logics information.

Then, we define the *ExtendedRepresentation* of a conceptual interpretation $\mathcal{I}$ as the only extended simple ABox $\mathcal{A}_E$ such that $Models(< \varnothing, \mathcal{A}_E >) = Models_{cw}(Representation(\mathcal{I}))$. Table 2 shows the *ExtendedRepresentation* of the first arch of figure 1.

The *entropy* of a conceptual interpretation $\mathcal{I}$, $entropy(\mathcal{I})$, is defined as the number of assertions in $ExtendedRepresentation(\mathcal{I})$.

The entropy of $\mathcal{I}$ has a direct interpretation as the number of boolean variables necessary to represent $\mathcal{I}$. If no knowledge is available, for every concept

[6] The given algorithm is restricted to interpretations with a finite domain.

[7] Note for example that in order to represent the first arch in figure 1 when no intensional knowledge is available it is necessary to use role non-membership assertions in order to state that there is no pair of objects that touches laterally (see also table 2).

Representation(arch 1)
$on\_floor(a)$
$on\_floor(c)$
$brick(a)$
$brick(b)$
$brick(c)$
$flat\_bottom(a)$
$flat\_bottom(b)$
$flat\_bottom(c)$
$flat\_top(a)$
$flat\_top(b)$
$flat\_top(c)$
$supports(a,b)$
$supports(c,b)$

ExtendedRepresentation(arch 1)		
$on\_floor(a)$	$\neg on\_floor(b)$	$on\_floor(c)$
$brick(a)$	$brick(b)$	$brick(c)$
$flat\_bottom(a)$	$flat\_bottom(b)$	$flat\_bottom(c)$
$flat\_top(a)$	$flat\_top(b)$	$flat\_top(c)$
$\neg wedge(a)$	$\neg wedge(b)$	$\neg wedge(c)$
$\neg cylinder(a)$	$\neg cylinder(b)$	$\neg cylinder(c)$
$\neg supports(a,a)$	$supports(a,b)$	$\neg supports(a,c)$
$\neg supports(b,a)$	$\neg supports(b,b)$	$\neg supports(b,c)$
$\neg supports(c,a)$	$supports(c,b)$	$\neg supports(c,c)$
$\neg touches(a,a)$	$\neg touches(a,b)$	$\neg touches(a,c)$
$\neg touches(b,a)$	$\neg touches(b,b)$	$\neg touches(b,c)$
$\neg touches(c,a)$	$\neg touches(c,b)$	$\neg touches(c,c)$

Table 2: This table shows the closed-world representation, and the extended representation of arch 1 in figure 1.

name A, and every element $d \in \Delta^{\mathcal{I}}$ we need a boolean random variable stating whether $d \in A^{\mathcal{I}}$ (true) or whether $d \notin A$ (false). Additionally, for every role name R, and every element pair $(d_1, d_2) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ we need a boolean random variable stating whether $(d_1, d_2) \in R^{\mathcal{I}}$ or not. When no knowledge is available (and then we suppose that the probability for true is the same that the probability for false, and is equal to $\frac{1}{2}$), the entropy for any of these boolean random variables is $-\frac{1}{2} * log\frac{1}{2} - \frac{1}{2} * log\frac{1}{2} = 1$.

Every assertion in $\mathcal{A}_E$ represents the value of one of these boolean variables. So, the notion of entropy that has been introduced here is a direct adaptation of Shannon measure.

Example The entropy of the arch 1 of figure 1 is 36, as it can be checked by counting the number of assertions in its extended representation (see table 2). In fact, for this simple arch model, it can easily be seen that every arch model can be represented with 36 boolean variables, since the number of domain elements will always be three, and there are six different concept names and two different role names.

The minimum expression of a given extended simple ABox $\mathcal{A}_E$ with respect to a TBox $\mathcal{T}$, $Minimum_{\mathcal{T}}(\mathcal{A})$, is defined as an extended simple ABox $\mathcal{A}_0$ such that $Models(< \mathcal{T}, \mathcal{A}_0 >) = Models(< \mathcal{T}, \mathcal{A} >)$ and $|\mathcal{A}_0| = minimum(\{|\mathcal{A}_E| : Models(\mathcal{A}) = Models(\mathcal{A}_E)\}$[8].

$Minimum_{\mathcal{T}}(\mathcal{A})$ determines the minimum number of assertions that can be used to represent the extensional knowledge in $\mathcal{A}$ given the intensional knowledge in $\mathcal{T}$.

Then, we can define the amount of *redundant information* in an extended simple ABox $\mathcal{A}$ with respect to a TBox $\mathcal{T}$ as:

$Redundant_{\mathcal{T}}(\mathcal{A}) = |\mathcal{A}| - |Minimum_{\mathcal{T}}(\mathcal{A})|$.

[8]Note that there can be several $\mathcal{A}_E$ that minimize $|\mathcal{A}_E|$. $Minimum_{\mathcal{T}}(\mathcal{A})$ is one of them.

Example Going on with our arch example; and given the initial TBox $\mathcal{T}_0$ in table 1, we can compute $Minimum_{\mathcal{T}_0}(ExtendedRepresentation(arch\ 1))$. The result is shown in table 3. Then, it follows that $Redundant_{\mathcal{T}_0}(er_{arch1}) = 19$.

$Minimum_{\mathcal{T}_0}(ExtendedRepresentation(arch1))$					
$on\_floor(a)$	$on\_floor(c)$	$brick(a)$	$brick(b)$	$brick(c)$	$supports(a,b)$
$\neg supports(b,b)$	$supports(c,b)$	$\neg touches(a,a)$	$\neg touches(a,b)$	$\neg touches(a,c)$	$\neg touches(b,a)$
$\neg touches(b,b)$	$\neg touches(b,c)$	$\neg touches(c,a)$	$\neg touches(c,b)$	$\neg touches(c,c)$	

Table 3: This table shows $Minimum_{\mathcal{T}_0}(ExtendedRepresentation(arch\ 1))$ (min_{arch1}). It contains only 17 assertions from the 36 initially present in $ExtendedRepresentation(arch\ 1)$ (er_{arch1}). The set of models for $< \mathcal{T}_0, min_{arch1} >$ is equal to the set of models for $< \mathcal{T}_0, er_{arch1} >$. That is, the assertions in $er_{arch1} \setminus min_{arch1}$ can be infered from $< \mathcal{T}_0, min_{arch1} >$.

Now we are able to talk about the amount of information encoded in a TBox $\mathcal{T}$ from a model theoretic point of view. For this purpose, we define:

- The amount of information provided by $\mathcal{T}$ with respect to a model $\mathcal{I}$:

  $Info_{\mathcal{I}}(\mathcal{T}) = Redundant_{\mathcal{T}}(ExtendedRepresentation(\mathcal{I}))$

- The amount of information provided by $\mathcal{T}$ with respect to a set of conceptual interpretations $\mathcal{M} = \{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n\}$ that are models of $\mathcal{T}$:

  $Info_{\mathcal{M}}(\mathcal{T}) = \frac{\sum_{\mathcal{I} \in \mathcal{M}} Info_{\mathcal{I}}(\mathcal{T})}{n}$

3.3 Learning a TBox

Let's go back to our problem. We want to learn a TBox $\mathcal{T}^*$ from a pair $< \mathcal{T}_0, \mathcal{M} >$ constituted by an initial theory $\mathcal{T}_0$ such that $\mathcal{T}_0 \subseteq \mathcal{T}^*$, and a set of $\mathcal{T}^*$-models $\mathcal{M} = \{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n\}$.

Then, we define the process of learning a TBox $\mathcal{T}^*$ from a pair $< \mathcal{T}_0, \mathcal{M} >$ as the process of acquiring a set of axioms $\{\Phi_1, \Phi_2, \ldots, \Phi_m\}$ such that:

1. $\mathcal{T}^* = \mathcal{T}_0 \cup \{\Phi_1, \Phi_2, \ldots, \Phi_m\}$
2. $\mathcal{M} \subseteq Models(\mathcal{T}^*)$
3. $\forall\ TBox\ \mathcal{T} : \mathcal{T}^* \subset \mathcal{T} \wedge \mathcal{M} \subseteq Models(\mathcal{T}) \Rightarrow Info_{\mathcal{M}}(\mathcal{T}) = Info_{\mathcal{M}}(\mathcal{T}^*)$
4. $\forall\ TBox\ \mathcal{T} : \mathcal{T}_0 \subseteq \mathcal{T} \wedge \mathcal{T} \subset \mathcal{T}^* \Rightarrow Info_{\mathcal{M}}(\mathcal{T}) < Info_{\mathcal{M}}(\mathcal{T}^*)$

The first condition defines $\mathcal{T}^*$ as the union of $\mathcal{T}_0$ and the set of acquired axioms. The second one states that the set of conceptual interpretations from which we learn should be models of $\mathcal{T}^*$. The third point states that all TBoxes including $\mathcal{T}^*$ and having $\mathcal{M}$ as models provide no new information. And finally the fourth point states that no subset of $\mathcal{T}^*$ that has $\mathcal{T}_0$ as a subset provides as much information as $\mathcal{T}^*$ provides.

This definition does not imply uniqueness. It only states clearly from a model theoretic point of view, which conditions must hold $\mathcal{T}^*$.

In order to talk about axioms individually, the following definitions are introduced:

- $\mathcal{T}$ is *consistent* with respect to $\mathcal{M} = \{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n\}$ iff $\mathcal{M} \subseteq Models(\mathcal{T})$.
- An axiom Φ is *consistent* with respect to a pair $< \mathcal{T}, \mathcal{M} >$ iff $\mathcal{T} \cup \{\Phi\}$ is consistent with respect to $\mathcal{M}$.
- An axiom Φ is *interesting* with respect to a pair $< \mathcal{T}, \mathcal{M} >$ iff Φ is consistent with respect to $< \mathcal{T}, \mathcal{M} >$, and $Info_{\mathcal{M}}(\mathcal{T} \cup \{\Phi\}) > Info_{\mathcal{M}}(\mathcal{T})$.

The two latter definitions can also be extended to sets of axioms $\{\Phi_1, \Phi_2, \ldots, \Phi_m\}$.

Up to here we have defined the problem of learning a theory $\mathcal{T}^*$ from a set of models $\mathcal{M}$. Nevertheless, we have not said a word about procedures to find $\mathcal{T}^*$ or, if not possible, to find a theory *as similar as possible* to $\mathcal{T}^*$. In any case, the procedure has to perform a search in the space of TBoxes. This is, of course, intractable.

A proposal of a general procedure for approximate theory learning can be found in [1]. The procedure has been defined in a probabilistic extension of the model presented here as an incremental procedure that introduces new axioms to $\mathcal{T}_0$ one by one.

It deals with the inherent intractability of the problem by the introduction of a pair of heuristic measures based on a simplified view of the information theory stuff introduced in this paper adapted for the probabilistic environment. Unfortunately, no theoretical bounds have been obtained for the *similarity* of the TBox obtained as the result of the procedure and $\mathcal{T}^*$. Nevertheless, the procedure has been successfully used in the adaptation of some typical Inductive Logic Programming benchmarks.

Example Given the initial theory in table 1 ($\mathcal{T}_0$), and arch 1 from figure 1; let's check whether axiom Φ_9 is interesting with respect to $< \mathcal{T}_0, \{arch\ 1\} >$.

$Minimum_{\mathcal{T}_0 \cup \{\Phi_9\}}(er_{arch1})$ corresponds to $\{on\_floor(a), on\_floor(c), brick(a),$ $brick(b), brick(c), supports(a,b), supports(c,b), \neg\ supports(b,b)\}$. The amount of redundant information is then 28, strictly greater than it was for $\mathcal{T}_0$. So, Φ_9 is interesting for $< \mathcal{T}_0, \{arch\ 1\} >$.

Now, doing the same for Φ_{10} with respect to $< \mathcal{T}_0 \cup \{\Phi_9\}, \{arch\ 1\} >$ we obtain that $Minimum_{\mathcal{T}_0 \cup \{\Phi_9, \Phi_{10}\}}(er_{arch1})$ corresponds to $\{brick(a), brick(b),$ $brick(c), supports(a,b), supports(c,b), \neg\ supports(b,b)\}$. So, Φ_{10} is interesting for $< \mathcal{T}_0 \cup \{\Phi_9\}, \{arch\ 1\} >$.

4 Conclusions

This paper provides a formalization for the problem of TBox acquisition from a set of models. The formalization uses model theory, the same formal base used by description logics reasoning machinery. Information theory notions have been integrated into description logics: the paper defines the amount of information provided by a TBox, and from this point, defines the properties a learned TBox should hold.

5 Acknowledgements

Thanks to the reviewers, who have provided quite valuable comments on the presented work.

References

[1] Jordi Alvarez. A formal framework for theory learning using description logics. Technical Report LSI-00-R, Universitat Politcnica de Catalunya, 2000.

[2] Martin Buchheit, Francesco M. Donini, and Andrea Schaerf. Decidable reasoning in terminological knowledge representation systems. *Journal of Artificial Intelligence Research*, 1:109–138, 1993.

[3] W. Cohen, A. Borgida, and H. Hirsh. Computing least common subsumers in description logics. In *Proceedings of AAAI-92*, pages 754–760, 1992.

[4] W. Cohen and H. Hirsh. The learnability of description logics with equality constraints. *Machine Learning*, 17(2–3):169–199, 1994.

[5] William W. Cohen and Haym Hirsh. Learning the CLASSIC description logic: Theoretical and experimental results. In John Doyle, E. Sandewall, and P. Torasso, editors, *Proceedings of 4th International Conference on Principles of Knowledge Representation and Reasoning*, pages 121–133, Bonn, 1994. Morgan Kaufmann.

[6] F.M. Donini, M. Lenzerini, D. Nardi, and A. Schaerf. Reasoning in description logics. In G. Brewka, editor, *Studies in Logic, Language and Information*, pages 193–238. CLSI Publications, 1996.

[7] M. Frazier and L. Pitt. CLASSIC learning. *Machine Learning*, 25:151–193, 1996.

[8] J. Kietz and K. Morik. A polynomial approach to the constructive induction of structural knowledge. *Machine Learning*, 14(2):193–217, 1994.