



# On Knowledge Dependence in Weighted Description Logic

Pietro Galliani<sup>1</sup>, Oliver Kutz<sup>1</sup>, Daniele Porello<sup>2</sup>,  
Guendalina Righetti<sup>1</sup>, and Nicolas Troquard<sup>1</sup>

<sup>1</sup> CORE Conceptual and Cognitive Modelling Group  
KRDB Research Centre for Knowledge and Data  
Free University of Bozen-Bolzano, Italy  
firstname.lastname@unibz.it

<sup>2</sup> LOA Laboratory for Applied Ontology, ISTC-CNR  
Trento, Italy  
daniele.porello@cnr.it

## Abstract

We study a family of operators (called ‘Tooth’ operators) that combine Description Logic concepts via weighted sums. These operators are intended to capture the notion of instances satisfying “enough” of the concept descriptions given. We examine two variants of these operators: the “knowledge-independent” one, that evaluates the concepts with respect to the current interpretation, and the “knowledge-dependent” one that instead evaluates them with respect to a specified knowledge base, comparing and contrasting their properties. We furthermore discuss the connections between these operators and linear classification models.

## 1 Introduction

In this work, we continue the study, began in [15] and [16], of threshold (“Tooth”) operators in the context of Description Logic and related formalisms.

These operators allow us to define a concept in terms of a *weighted sum* of sub-concepts: for instance, as we will see, a concept of the form  $\nabla_{\frac{1}{2}, \frac{1}{2}, \frac{1}{2}}^1(C_1, C_2, C_3)$  is true of some individual  $d$  if and only if at least two of  $C_1$ ,  $C_2$ , and  $C_3$  are true of  $d$ , that is, if and only if<sup>1</sup>

$$1/2 \cdot \left\{ \begin{array}{ll} 1 & \text{if } C_1(d) \\ 0 & \text{otherwise} \end{array} \right\} + 1/2 \cdot \left\{ \begin{array}{ll} 1 & \text{if } C_2(d) \\ 0 & \text{otherwise} \end{array} \right\} + 1/2 \cdot \left\{ \begin{array}{ll} 1 & \text{if } C_3(d) \\ 0 & \text{otherwise} \end{array} \right\} \geq 1.$$

*Weighted Threshold Operators* of this type have been studied extensively in the context of propositional logic and circuit complexity theory (see e.g. [10, 4, 8, 9]), with a particular focus on the *monotone* case in which all the weights assigned to the subformulas are positive. In that context, these operators have been seen to be closely related to several important open problems in complexity theory; in this work, instead, we will focus on their possible application to Knowledge Representation in Description Logic.

<sup>1</sup>By slight abuse of the usual DL notation, here and in the remainder of the paper we write expressions of the form  $C(d)$  for  $d \in C^I$  (where  $I$  is a DL interpretation) and  $C(b)$  for the ABox statement  $b : C$  ( $b$  a name).

Tooth operators are simple, conceptually clear connectives that—because of their obvious connections with linear classification models—provide a natural and straightforward link between knowledge representation and statistical learning.

In what follows, we will introduce two varieties of Tooth operators: the “knowledge-independent” ones, that evaluate subconcepts with respect to the current interpretation, and the “knowledge-dependent” ones that instead, evaluate subconcepts with respect to a specified knowledge base. Then, we will study some of the basic properties of these operators. Finally, we will discuss the application of Tooth Operators to the problem of learning concepts from data.

## 2 On Weighted Concept Combination(s)

A class of  $m$ -ary operators, denoted by the symbol  $\mathbb{W}$  (pronounced ‘tooth’), were presented in [15]. These operators are meant to specify new concepts in terms of *weighted combinations* of sub-concepts: indeed, each such operator  $i$ ) takes a list  $C_1 \dots C_n$  of concept descriptions,  $ii$ ) associates a vector of weights  $w_1 \dots w_n$  to them, and  $iii$ ) returns a complex concept that applies to those instances for which, by summing up the weights of the satisfied concepts, a certain threshold  $t$  is met.

In this work, we will focus on the logic  $\mathcal{ALC}(\mathbb{W}^{\mathbb{R}}, \mathbb{W}_{\mathcal{K}}^{\mathbb{R}})$  (or, for brevity,  $\mathcal{ALC}(\mathbb{W}, \mathbb{W}_{\mathcal{K}})$ ) obtained by adding such operators to  $\mathcal{ALC}$  and allowing weights and thresholds to take any real value  $r \in \mathbb{R}$ ; but most of our results and observations can be generalized easily to most other varieties of Description Logics. We refer the reader to [3] for an introduction to Description Logic and  $\mathcal{ALC}$ . We assume henceforth we are given fixed sets of atomic concepts  $N_C$ , atomic roles  $N_R$ , and individual names  $N_{\text{Ind}}$ .

For  $A \in N_C$ ,  $R \in N_R$ ,  $t \in \mathbb{R}$ ,  $m \in \mathbb{Z}_{>0}$ ,  $\vec{w} \in \mathbb{R}^m$  the set of  $\mathcal{ALC}(\mathbb{W}, \mathbb{W}_{\mathcal{K}})$  concepts is described by the grammar:

$$C ::= \perp \mid \top \mid A \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \forall R.C \mid \exists R.C \mid \mathbb{W}_{\vec{w}}^t(\underbrace{C, \dots, C}_m) \mid \mathbb{W}_{\mathcal{K}, \vec{w}}^t(\underbrace{C, \dots, C}_m)$$

where, in the last expression,  $\mathcal{K}$  ranges over all the *knowledge bases* over the above language, i.e., all the sets of expressions of the form  $C_1 \sqsubseteq C_2$ ,  $C_1(a)$  or  $R(a, b)$ , where  $C_1$  and  $C_2$  are  $\mathcal{ALC}(\mathbb{W}, \mathbb{W}_{\mathcal{K}})$  concepts,  $R \in N_R$ , and  $a$  and  $b$  range over the individual names  $N_{\text{Ind}}$ .<sup>2</sup>

To better visualise the weights that an operator associates to the concepts, we sometimes use the notation  $\mathbb{W}^t(C_1 : w_1, \dots, C_m : w_m)$  instead of  $\mathbb{W}_{\vec{w}}^t(C_1, \dots, C_m)$ ; and likewise, we will write  $\mathbb{W}_{\mathcal{K}}^t(C_1 : w_1, \dots, C_m : w_m)$  instead of  $\mathbb{W}_{\mathcal{K}, \vec{w}}^t(C_1, \dots, C_m)$ .

The semantics of the first new type of expressions – which we will call *Knowledge-Independent Tooth Operators* – is obtained by extending the definition of the semantics of  $\mathcal{ALC}$  as follows:

**Definition 1** (Values of Individuals in Knowledge-Independent Tooth Operators).

Let  $\mathbb{C} = \mathbb{W}^t((C_i : w_i)_{i=1..m})$  be a knowledge-independent tooth operator, let  $I$  be an interpretation, and let  $d \in \Delta^I$  be an individual in the domain of  $I$ . Then

$$v_{\mathbb{C}}^I(d) = \sum_{i \in \{1, \dots, m\}} \{w_i \mid d \in C_i^I\}. \quad (1)$$

**Definition 2** (Semantics of Knowledge-Independent Tooth Operators).

Let  $\mathbb{C} = \mathbb{W}^t((C_i : w_i)_{i=1..n})$  be a knowledge-independent tooth operator and let  $I$  be an interpretation. Then

$$\mathbb{C}^I = \{d \in \Delta^I \mid v_{\mathbb{C}}^I(d) \geq t\}. \quad (2)$$

<sup>2</sup>It is also customary to write  $C_1 \equiv C_2$  as a shorthand for the two subsumptions  $C_1 \sqsubseteq C_2$  and  $C_2 \sqsubseteq C_1$ .

We continue with an example:

**Example 1.** Consider the concept defined by the tooth expression

$$\mathbb{W}^{1.4}(\text{HasTrunk} : 1.0, \text{HasAntlers} : -0.5, \text{HasTusks} : 0.3, \text{IsBig} : 0.4, \text{IsGrey} : 0.4) \quad (3)$$

where HasTrunk, HasAntlers, HasTusks, IsBig and IsGrey are concepts of our language. Then (3) can be seen as an attempt to define the notion of elephant in terms of those other concepts by assigning positive values to qualities that increase the “elephantness” of the subject (e.g. having a trunk, or being big) and negative ones to qualities that decrease it (e.g. having antlers).

Now, let  $I$  be an interpretation containing three individuals Dumbo, Babar and Edgar and such that  $\text{HasTrunk}^I = \{\text{Dumbo}^I, \text{Babar}^I\}$ ;  $\text{HasAntlers}^I = \{\text{Edgar}^I\}$ ;  $\text{HasTusks}^I = \{\text{Babar}^I\}$ ;  $\text{IsBig}^I = \{\text{Babar}^I, \text{Edgar}^I\}$ ; and  $\text{IsGrey}^I = \{\text{Dumbo}^I, \text{Babar}^I\}$ . Then, if we write Elephant for (3), we have that

- $I \models \text{Elephant}(\text{Dumbo})$ : the score of the tooth expression for  $\text{Dumbo}^I$  is  $1.0 (\text{HasTrunk}) + 0.4 (\text{IsGrey}) \geq 1.4$ .
- $I \models \text{Elephant}(\text{Babar})$ , with score  $1.0 + 0.3 + 0.4 + 0.4 = 2.1 \geq 1.4$ .
- $I \not\models \text{Elephant}(\text{Edgar})$ : indeed, the score of the tooth expression for  $\text{Edgar}^I$  is  $-0.5 + 0.4 = -0.1 < 1.4$ .

As it is not difficult to see, for example by considering all possible combinations of attributes, Elephant is equivalent to the Disjunctive Normal Form Boolean combination of concepts

$$\begin{aligned} & (\text{HasTrunk} \sqcap \text{HasTusks} \sqcap \text{IsBig} \sqcap \text{IsGrey}) \sqcup \\ & (\text{HasTrunk} \sqcap \text{IsBig} \sqcap \neg \text{HasAntlers}) \sqcup \\ & (\text{HasTrunk} \sqcap \text{IsGrey} \sqcap \neg \text{HasAntlers}). \end{aligned}$$

More in general, as mentioned also in [15], any tooth operator is always reducible to some Boolean combination of its components (this follows directly from the fact that all Boolean functions may be represented using disjunctions and negations).

We now consider the other type of tooth operator  $\mathbb{W}_{\mathcal{K}, \bar{w}}^t(C_1, \dots, C_m)$ , or  $\mathbb{W}_{\mathcal{K}}(C_1 : w_1, \dots, C_m : w_m)$ , which we will call the *Knowledge-Dependent Tooth Operator*. In this type of tooth operator, as we will now see, the sub-expressions are evaluated not with respect to the current interpretation  $I$  but with respect to the knowledge base  $\mathcal{K}$ :

**Definition 3** (Values of Individual Names in Knowledge-Dependent Tooth Operators).

Let  $\mathbb{C}_{\mathcal{K}} = \mathbb{W}_{\mathcal{K}}^t(C_1 : w_1 \dots C_n : w_n)$  be a knowledge-dependent tooth operator and let  $a$  be any individual name. Then we define the value  $\omega_{\mathbb{C}_{\mathcal{K}}}(a)$  of  $a$  according to  $\mathbb{C}_{\mathcal{K}}$  as

$$\omega_{\mathbb{C}_{\mathcal{K}}}(a) = \sum \{w_i : \mathcal{K} \models C_i(a)\}$$

where, as usual,  $\mathcal{K} \models C_i(a)$  if and only if, for all interpretations  $J$ ,  $J \models \mathcal{K} \Rightarrow J \models C_i(a)$ .

**Definition 4** (Semantics of Knowledge-Dependent Tooth Operators).

Let  $\mathbb{C}_{\mathcal{K}} = \mathbb{W}_{\mathcal{K}}^t(C_1 : w_1 \dots C_n : w_n)$  be a knowledge-dependent tooth operator and let  $I$  be an interpretation. Then

$$(\mathbb{C}_{\mathcal{K}})^I = \{a^I \mid a \in N_{\text{Ind}}, a^I \text{ is defined, } \omega_{\mathbb{C}_{\mathcal{K}}}(a) \geq t\}$$

where, in the above expression,  $a$  ranges over the set  $N_{\text{Ind}}$  of all individual names.

This notion of knowledge-dependent tooth operator is somewhat different from the knowledge-independent notion introduced before. Most importantly, it follows immediately from our definitions that  $(\mathbb{W}_{\mathcal{K}}^t(C_1 : w_1 \dots C_n : w_n))^I$  is always of the form  $\{a^I, b^I, \dots\}$  for some set of individual names  $\{a, b, \dots\}$ ; and that, more importantly, this set is the same no matter the interpretation  $I$  with respect to which the operator is interpreted. Nonetheless, we believe it useful to define the semantics of this operator in the usual way, i.e. in terms of satisfaction with respect to interpretations, both for uniformity and for ease of combination with other connectives. Additionally, we point out here that both our tooth operators are well-defined, even in the case of nested expressions. This is verified easily by remarking that the interpretation of a tooth expression is defined in terms of the interpretation of its *subexpressions*, in the knowledge-independent case, or in terms of their entailment given a knowledge base (itself a sub-expression) in the knowledge-dependent one.

**Example 2.** *Continuing from Example 1, let  $\mathcal{K}$  contain the axioms*

- $\text{HasTrunk}(\text{Surus})$ ;<sup>3</sup>
- $\top \sqsubseteq \text{IsBig} \sqcup \text{IsGrey}$ ;
- $\text{HasAntlers} \sqsubseteq \text{HasTusks} \sqcap \text{IsBig} \sqcap \text{IsGrey}$ .

Then  $\mathcal{K} \models \text{Elephant}(\text{Surus})$ , where *Elephant* is defined as in Example 1 as

$$\mathbb{W}^{1.4}(\text{HasTrunk} : 1.0, \text{HasAntlers} : -0.5, \text{HasTusks} : 0.3, \text{IsBig} : 0.4, \text{IsGrey} : 0.4).$$

Indeed, let  $I$  be any interpretation such that  $I \models \mathcal{K}$ . Then clearly  $\text{Surus}^I \in \text{HasTrunk}^I$ . If moreover  $\text{Surus}^I \in \text{HasAntlers}^I$ , we have that  $\text{Surus}^I \in \text{HasTusks}^I \sqcap \text{IsBig}^I \sqcap \text{IsGrey}^I$  and therefore that

$$v_{\text{Elephant}}^I(\text{Surus}) \geq 1.0 - 0.5 + 0.3 + 0.4 + 0.4 \geq 1.4.$$

If instead  $\text{Surus}^I \notin \text{HasAntlers}^I$ , consider that by the second axiom  $\text{Surus}^I \in \text{IsBig}^I$  or  $\text{Surus}^I \in \text{IsGrey}^I$ ; and in either case,

$$v_{\text{Elephant}}^I(\text{Surus}) \geq 1.0 + 0.4 = 1.4,$$

since the only concept that has a negative weight in the tooth expression does not apply to *Surus*. So, for every interpretation  $I$  such that  $I \models \mathcal{K}$  we have that  $I \models \text{Elephant}(\text{Surus})$ , and therefore  $\mathcal{K} \models \text{Elephant}(\text{Surus})$ .

However, if we write  $\text{Elephant}_{\mathcal{K}}$  for the knowledge-dependent tooth expression

$$\mathbb{W}_{\mathcal{K}}^{1.4}(\text{HasTrunk} : 1.0, \text{HasAntlers} : -0.5, \text{HasTusks} : 0.3, \text{IsBig} : 0.4, \text{IsGrey} : 0.4)$$

then it is never the case that  $I \models \text{Elephant}_{\mathcal{K}}(\text{Surus})$ . Indeed, of all subconcepts of our tooth operators, it is trivial to see that  $\mathcal{K} \models \text{HasTrunk}(\text{Surus})$  but  $\mathcal{K} \not\models \text{HasTusks}(\text{Surus})$ ,  $\mathcal{K} \not\models \text{IsBig}(\text{Surus})$ ,  $\mathcal{K} \not\models \text{HasAntlers}(\text{Surus})$  and  $\mathcal{K} \not\models \text{IsGrey}(\text{Surus})$ .

Hence  $\omega_{\text{Elephant}_{\mathcal{K}}}(\text{Surus}) = 1.0 < 1.4$  and it is never the case that  $\text{Surus}^I \in (\text{Elephant}_{\mathcal{K}})^I$ .

Thus, the knowledge-independent and knowledge-dependent variants of our tooth operator do not in general agree. These two types of tooth operator can also be used together, for example for *aggregating* the results of multiple knowledge bases:

<sup>3</sup>According to Pliny the Elder, the elephant that Hannibal personally rode during his invasion of Italy was named Surus.

**Example 3.** Let us suppose that we want to decide whether Tantor is an elephant according to the knowledge-dependent tooth operator  $\text{Elephant}_{\mathcal{K}}$  discussed in Example 2; but that, instead of a single knowledge base  $\mathcal{K}$ , we have  $n$  distinct knowledge bases  $\mathcal{K}_1 \dots \mathcal{K}_n$ . A straightforward way to answer this question is then to ask whether Tantor is an elephant according to at least half of the knowledge bases, or – more in general – according to a certain percentage  $r \in [0, 100]$  of the knowledge bases.

This can be represented easily as  $\mathbb{W}^{n/2}(\text{Elephant}_{\mathcal{K}_1} : 1, \dots, \text{Elephant}_{\mathcal{K}_n} : 1)$  or, more in general, as  $\mathbb{W}^{nr/100}(\text{Elephant}_{\mathcal{K}_1} : 1, \dots, \text{Elephant}_{\mathcal{K}_n} : 1)$ .

In the next two sections, we will discuss some basic properties of these two variants of the tooth operator.

### 3 Knowledge-Independent Teeth

In this section, we begin a brief exploration of the problem of evaluating if a knowledge-independent tooth expression  $\mathbb{W}_{\bar{w}}^t(C_1, \dots, C_m)$  is *entailed* by a knowledge base  $\mathcal{K}$ .

As per the usual notion of entailment, we have that  $\mathcal{K} \models \mathbb{W}_{\bar{w}}^t(C_1, \dots, C_m)(a)$  if and only if  $I \models \mathbb{W}_{\bar{w}}^t(C_1, \dots, C_m)(a)$  for all interpretations  $I$  such that  $I \models \mathcal{K}$ ; but can we find a notion of value of a (knowledge-independent) tooth operator with respect to a knowledge base that leads directly to this same notion of entailment? The answer is positive:

**Definition 5** (Evaluating Knowledge-Independent Tooth Operators over Knowledge Bases). *Let  $\mathcal{K}$  be a knowledge base, let  $a$  be an individual name, and let  $\mathbf{C} = \mathbb{W}^t((C_i : w_i)_{i=1\dots n})$  be a knowledge-independent tooth expression. Then we define the value of  $a$  in  $\mathbf{C}$  with respect to  $\mathcal{K}$  as*

$$\mu_{\mathbf{C}}^{\mathcal{K}}(a) = \min\{v_{\mathbf{C}}^I(a^I) : I \models \mathcal{K}\}.$$

Note that the minimum introduced in the above definition always exists, because any possible value of  $v_{\mathbf{C}}^I(a^I)$  must belong in the (finite) set of values  $S = \{\sum(X) : X \subseteq \{w_1, \dots, w_n\}\}$ . Furthermore, we have that

**Proposition 1.** *Let  $\mathcal{K}$  be any knowledge base, let  $a$  be an individual name, and let  $\mathbf{C} = \mathbb{W}^t((C_i : w_i)_{i=1\dots n})$  be any knowledge-independent tooth expression. Then*

$$\mathcal{K} \models \mathbf{C}(a) \Leftrightarrow \mu_{\mathbf{C}}^{\mathcal{K}}(a) \geq t.$$

*Proof.* By definition,

$$\begin{aligned} \mathcal{K} \models \mathbf{C}(a) &\text{ iff } \forall I, (I \models \mathcal{K} \Rightarrow I \models \mathbf{C}(a)) \\ &\text{ iff } \forall I, (I \models \mathcal{K} \Rightarrow v_{\mathbf{C}}^I(a^I) \geq t) \\ &\text{ iff } \min\{v_{\mathbf{C}}^I(a^I) : I \models \mathcal{K}\} \geq t \text{ iff } \mu_{\mathbf{C}}^{\mathcal{K}}(a) \geq t. \end{aligned}$$

□

It is worth remarking that the knowledge-independent value  $\mu_{\mathbf{C}}^{\mathcal{K}}$  of a tooth expression  $\mathbf{C}$  with respect to a knowledge base  $\mathcal{K}$  just defined is *not* necessarily the same as the knowledge-dependent value  $\omega_{\mathbf{C}\mathcal{K}}$  of the same tooth operator with respect to the same knowledge base. For instance, let  $\mathcal{K} = \emptyset$  be the empty theory, let  $a$  be some individual name, let  $A$  be an atomic concept, and consider the tooth operator  $\mathbf{C} = \mathbb{W}^1(A : 1, \neg A : 1)$ . Then  $\mu_{\mathbf{C}}^{\emptyset}(a) = 1$ , since any interpretation  $I$  which assigns some individual to  $a$  will be such that  $a^I \in A^I$  or such that  $a^I \notin A^I$  (and, therefore,  $v_{\mathbf{C}}^I = 1$  in either case). However,  $\omega_{\mathbf{C}\emptyset}(a) = 0$ , because  $\emptyset \not\models A(a)$  and  $\emptyset \not\models \neg A(a)$ . For this example we also have that  $\emptyset \models \mathbf{C}(a)$  even

though  $\emptyset \not\models C_\emptyset(a)$ . This illustrates the difference between entailment in the knowledge-independent and knowledge-dependent cases: very briefly, in the above case we have that  $\mathcal{K} \models C(a)$  if and only if  $\mathcal{K} \models (A \sqcup \neg A)(a)$ , but  $\mathcal{K} \models C_{\mathcal{K}}(a)$  if and only if  $\mathcal{K} \models A(a)$  or  $\mathcal{K} \models \neg A(a)$ .

The negation of a knowledge-independent tooth expression is always equivalent to some other knowledge-independent tooth expression: for example, the negation of Equation (3) is equivalent to

$$\mathbb{W}^{-1.3}((\text{HasTrunk} : -1.0), (\text{HasAntlers} : 0.5), (\text{HasTusks} : -0.3), (\text{IsBig} : -0.4), (\text{IsGrey} : -0.4)).$$

More in general, we have that

**Proposition 2.** *Let  $\mathbf{C} = \mathbb{W}^t(C_1 : w_1, \dots, C_n : w_n)$ , and let  $\mathbf{D} = \mathbb{W}^{-t+\epsilon}(C_1 : -w_1, \dots, C_n : -w_n)$ , where  $\epsilon = \min\{t - \sum(A) : A \subseteq (w_1 \dots w_n), t > \sum(A)\}$ . Then for all interpretations  $I$  and individuals  $d \in \Delta^I$ ,*

$$d \in \mathbf{C}^I \Leftrightarrow d \notin \mathbf{D}^I.$$

*Proof.* Consider an interpretation  $I$  and an individual  $d \in \Delta^I$ , and suppose that  $d \in \mathbf{C}^I$ . By definition, this means that  $\sum\{w_i : i \in 1 \dots n, d \in C_i^I\} = q \geq t$ . But then  $\sum\{-w_i : i \in 1 \dots n, d \in C_i^I\} = -q \leq -t < -t + \epsilon$ , and hence  $d \notin \mathbf{D}^I$ .

Suppose instead that  $d \notin \mathbf{C}^I$ . Then  $\sum\{w_i : i \in 1 \dots n, d \in C_i^I\} = q < t$ , and therefore – by the definition of  $\epsilon$  – we have that  $q \leq t - \epsilon$ . But this implies that  $\sum\{-w_i : i \in 1 \dots n, d \in C_i^I\} = -q \geq -t + \epsilon$ , and therefore that  $d \in \mathbf{D}^I$ .  $\square$

Another interesting and easy result is that negative weights can be disposed of in knowledge-independent tooth operators as long as concept complementation is in the language: for example, Equation (3) can be seen to be equivalent to

$$\mathbb{W}^{1.9}((\text{HasTrunk} : 1.0), (\neg\text{HasAntlers} : 0.5), (\text{HasTusks} : 0.3), (\text{IsBig} : 0.4), (\text{IsGrey} : 0.4)).$$

More in general, we have that

**Proposition 3.** *Let  $\mathbf{C}$  be of the form  $\mathbb{W}^t((C_1 : w_1), \dots, (C_n, w_n))$ , let  $k \in 1 \dots n$ , and let  $\mathbf{D}$  be instead  $\mathbb{W}^{t-w_k}((C_1 : w_1) \dots (\neg C_k : -w_k) \dots (C_n : w_n))$  (that is, let  $\mathbf{D}$  be obtained from  $\mathbf{C}$  by negating the value of  $w_k$ , replacing  $C_k$  with  $\neg C_k$ , and subtracting  $w_k$  from the threshold  $t$ ).*

*Then  $\mathbf{C}$  and  $\mathbf{D}$  are equivalent, in the sense that  $d \in \mathbf{C}^I \Leftrightarrow d \in \mathbf{D}^I$  for all interpretations  $I$  and all individuals  $d \in \Delta^I$ .*

*Proof.* Observe that, for any interpretation  $I$  and individual  $d \in \Delta^I$ ,  $v_{\mathbf{D}}^I(d) = v_{\mathbf{C}}^I(d) - w_k$ .

Indeed, suppose that  $d \in C_k^I$ : then  $d \notin (\neg C_k)^I$  and hence

$$v_{\mathbf{D}}^I(d) = \sum\{w_i : i \in 1 \dots n, i \neq k, d \in C_i^I\}$$

and

$$v_{\mathbf{C}}^I(d) = \sum\{w_i : i \in 1 \dots n, d \in C_i^I\} = w_k + \sum\{w_i : i \in 1 \dots n, i \neq k, d \in C_i^I\} = w_k + v_{\mathbf{D}}^I(d).$$

If instead  $d \notin C_k^I$ , we have at once that  $d \in (\neg C_k)^I$  and hence that

$$v_{\mathbf{D}}^I(d) = -w_k + \sum\{w_i : i \in 1 \dots n, i \neq k, d \in C_i^I\}$$

and

$$v_{\mathbf{C}}^I(d) = \sum\{w_i : i \in 1 \dots n, d \in C_i^I\} = \sum\{w_i : i \in 1 \dots n, i \neq k, d \in C_i^I\} = w_k + v_{\mathbf{D}}^I(d).$$

Thus, in either case,  $v_C^I(d) - w_k = v_D^I(d)$ . It implies at once that

$$d \in \mathbf{C}^I \Leftrightarrow v_C^I(d) \geq t \Leftrightarrow v_C^I(d) - w_k \geq t - w_k \Leftrightarrow v_D^I(d) \geq t - w_k \Leftrightarrow I \models d \in \mathbf{D}^I$$

as required. □

As a consequence of this result, it follows at once that:

**Corollary 1.** *Every knowledge-independent tooth expression*

$$\mathbf{C} := \mathbb{W}^t((C_1 : w_1) \dots (C_n, w_n))$$

*is equivalent to some knowledge-independent tooth expression*

$$\mathbf{D} := \mathbb{W}^{t'}((D_1 : w'_1) \dots (D_n, w'_n))$$

*such that all weights  $w'_i$  are positive and every  $D_i$  is equivalent to  $C_i$  or to  $\neg C_i$ .*

*Proof.* Apply repeatedly Proposition 3 to all negative weights  $w_i$ , until none is left. □

This is a particularly interesting result in the light of the results of [4] showing that, if a threshold operator (whose semantics corresponds precisely to that of our tooth operators) has only positive weights, it may be simulated by a Boolean circuit that is monotone (that is, is composed only of AND and OR gates) and has size which is polynomial in the number of inputs of the threshold. The paper mentions that the problem whether there exists a polynomial-size monotone *formula* (rather than circuit) that simulates such a threshold operator and uses only conjunctions and disjunctions is open, and to our current knowledge it is still so; but in any case, adding new concepts  $D_1 \dots D_s$  to our knowledge base to represent the intermediate nodes of the circuits, it follows at once that every knowledge base containing a knowledge-independent tooth operator may be reduced to some other knowledge base *not* containing it by adding to our TBox a polynomial number of polynomial-size definitions  $D_i \equiv \dots$  (where the right size is some Boolean combination of concepts, including possibly  $D_1 \dots D_{i-1}$ ) and the tooth operator itself is replaced by some (also polynomial-size) Boolean combination of the  $D_i$ .

On the other hand, it is straightforward to see that negation and disjunction can be defined in terms of knowledge-independent tooth operators: indeed,  $\neg C_1$  is equivalent to  $\mathbb{W}^0((C_1, -1))$  and  $C_1 \sqcup C_2$  is equivalent to  $\mathbb{W}^1((C_1, 1), (C_2, 1))$ .

Additionally, it is not difficult to express the notion that a knowledge-independent tooth operator takes values *between* two thresholds:

**Proposition 4.** *Let  $C_1 \dots C_n$  be concepts, let  $w_1 \dots w_n \in \mathbb{R}$  be weights, and let  $t, t' \in \mathbb{R}$  be such that  $t \leq t'$ . Then, for all interpretations  $I$ ,*

$$d \in (\mathbb{W}^t(C_1 : w_1 \dots C_n : w_n) \sqcap \mathbb{W}^{-t'}(C_1 : -w_1 \dots C_n : -w_n))^I$$

*if and only if  $t \leq v_C^I(d) \leq t'$ , where  $\mathbf{C}$  represents the knowledge-independent tooth expression  $\mathbb{W}^t(C_1 : w_1 \dots C_n : w_n)$ .*

As we will now see, even in the case in which all subconcepts of knowledge-independent tooth operators are atomic and all the weights are integer, the problem of evaluating whether a tooth operator subsumes another with respect to the empty knowledge base is co-NP-hard. This is proved easily by reducing the *Knapsack Problem* to it:



**Definition 6** (Knapsack Decision Problem). *Let  $\vec{o} = o_1 \dots o_n$  be a finite number of distinct objects, each one of which has a weight  $w_i \in \mathbb{Z}$  and a value  $v_i \in \mathbb{Z}$ . Furthermore, let  $W \in \mathbb{R}$  be a maximum weight capacity and let  $V \in \mathbb{Z}$  be the target value.*

*Then the Knapsack Decision Problem  $\mathbf{KN}(\vec{w}, \vec{v}, W, V)$  asks whether there is a possible selection of objects having total value  $\geq V$  and total weight  $\leq W$ , that is, whether there exists some  $\vec{x} = x_1 \dots x_n \in \{0, 1\}^n$  such that  $\sum_{i=1}^n v_i x_i \geq V$  and  $\sum_{i=1}^n w_i x_i \leq W$ .*

It is a standard result, found in most computational complexity textbooks (see e.g. [7]), that the Knapsack Decision Problem is NP-complete with respect to the binary length of its input.<sup>4</sup> Furthermore, it is not difficult to see how the Knapsack Problem may be encoded in two knowledge-independent tooth expressions with only atomic subexpressions:

**Proposition 5.** *Let  $\mathbf{KN}(\vec{w}, \vec{v}, W, V)$  be an instance of the Knapsack Decision Problem, where  $\vec{w} = w_1 \dots w_n$  and  $\vec{v} = v_1 \dots v_n$ . Then this instance has a solution if and only if, given  $n$  atomic concepts  $A_1 \dots A_n$ , the conjunction of the concepts*

$$\mathbb{W}^{-W}(A_1 : -w_1, \dots, A_n : -w_n) \quad (4)$$

and

$$\mathbb{W}^V(A_1 : v_1, \dots, A_n : v_n) \quad (5)$$

is satisfiable.

*Proof.* Suppose that  $\mathbf{KN}(\vec{w}, \vec{v}, W, V)$  has a solution  $\vec{x} = x_1 \dots x_n \in \{0, 1\}^n$ . Then let  $I$  be an interpretation such that, for all concepts  $A_1 \dots A_n$ ,  $a^I \in A_i^I$  if and only if  $x_i = 1$ : then since  $\sum_i w_i x_i \leq W$  we have that  $\sum_i (-w_i) x_i \geq -W$  and  $I \models \mathbb{W}^{-W}(A_1 : -w_1 \dots A_n : -w_n)(a)$ , as required; and similarly, since  $\sum_i v_i x_i \geq V$  we have that  $I \models \mathbb{W}^V(A_1 : v_1 \dots A_n : v_n)(a)$ .

Conversely, suppose that there exists an interpretation  $I$  such that  $I \models \mathbb{W}^{-W}(A_1 : -w_1 \dots A_n : -w_n)(a)$  and  $I \models \mathbb{W}^V(A_1 : v_1 \dots A_n : v_n)(a)$ . Then if we let, for all  $i = 1 \dots n$ ,  $x_i$  be 1 if  $a^I \in A_i^I$  and 0 otherwise, we have at once that  $\sum_i w_i x_i \leq W$  and  $\sum_i v_i x_i \geq V$ , as required.  $\square$

**Corollary 2.** *Let  $\mathbf{C}$  and  $\mathbf{D}$  be two knowledge-independent tooth expressions with integer weights and thresholds and only atomic subexpressions. Then the problem of deciding whether  $\models \mathbf{C} \sqsubseteq \mathbf{D}$  is co-NP-Hard with respect to the binary encodings of the weights and thresholds of  $\mathbf{C}$  and  $\mathbf{D}$ .*

*Proof.* By Proposition 5, an instance  $\mathbf{KN}(\vec{w}, \vec{v}, W, V)$  of the Knapsack Decision Problem has no solution if and only if the conjunction (4) and (5) is not satisfiable, that is, if and only if  $\models \mathbb{W}^{-W}(A_1 : -w_1, \dots, A_n : -w_n) \sqsubseteq \neg \mathbb{W}^V(A_1 : v_1, \dots, A_n : v_n)$ .

Thus, by Proposition 2, this is the case if and only if

$$\models \mathbb{W}^{-W}(A_1 : -w_1, \dots, A_n : -w_n) \sqsubseteq \mathbb{W}^{-V+\epsilon}(A_1 : -v_1, \dots, A_n : -v_n). \quad (6)$$

for  $\epsilon = \min\{V - \sum(S) : S \subseteq (v_1 \dots v_n), V \geq \sum(S)\}$ .

Observing that the length of the binary encoding of  $-V + \epsilon$  is of the same order of magnitude as the length of the binary encoding of  $V$ , we reach our conclusion.  $\square$

Thus, even evaluating subsumptions between simple tooth expressions with respect to the empty knowledge base is already computationally nontrivial. This illustrates the power of this type of operator.

<sup>4</sup>However, this problem is *not* NP-hard if we assume that the input is instead encoded in unary.



## 4 Knowledge-Dependent Teeth

In this section, we will in turn analyze our knowledge-dependent tooth operators, the semantics of which we introduced in Def. 4 above.

Let us begin by observing that, as mentioned already in [14], such operators are *non-monotonic* with respect to the knowledge base  $\mathcal{K}$ :

**Example 4.** Consider the usual definition of  $\text{Elephant}_{\mathcal{K}}$  from Example 2 as

$$\mathbb{W}_{\mathcal{K}}^{1.4}(\text{HasTrunk} : 1.0, \text{HasAntlers} : -0.5, \text{HasTusks} : 0.3, \text{IsBig} : 0.4, \text{IsGrey} : 0.4) \quad (7)$$

where now  $\mathcal{K}$  contains the axioms  $\text{HasTrunk}(\text{Hanno})$ ,<sup>5</sup> and  $\text{IsBig}(\text{Hanno})$ .

Then  $\omega_{\text{Elephant}_{\mathcal{K}}}(\text{Hanno}) = 1.0 + 0.4 \geq 1.4$ , and hence  $I \models \text{Elephant}_{\mathcal{K}}(\text{Hanno})$  for any interpretation  $I$ . However, if we updated our knowledge base<sup>6</sup>  $\mathcal{K}$  by adding to it the axiom  $\text{HasAntlers}(\text{Hanno})$  we would instead have that  $\omega_{\text{Elephant}_{\mathcal{K}}}(\text{Hanno}) = 1.0 - 0.5 + 0.4 < 1.4$ , and hence  $I \not\models \text{Elephant}_{\mathcal{K}}(\text{Hanno})$ .

Furthermore, as already mentioned, whether some individual  $d$  is or is not accepted by a knowledge-dependent tooth operator  $\mathbb{C}_{\mathcal{K}}$  depends exclusively on which individual names correspond to it according to the current interpretation  $I$ , and not on any other feature of  $I$ .

**Proposition 6.** Let  $I$  and  $I'$  be any two interpretations, let  $d \in \Delta^I$  and  $d' \in \Delta^{I'}$  be two individuals such that  $a^I = d \Leftrightarrow a^{I'} = d'$  for all individual names  $a \in N_{\text{Ind}}$ . Then, for all knowledge-dependent tooth expressions  $\mathbb{C}_{\mathcal{K}} = \mathbb{W}_{\mathcal{K}}^t((C_i : w_i)_{i=1\dots n})$ ,  $d \in (\mathbb{C}_{\mathcal{K}})^I \Leftrightarrow d' \in (\mathbb{C}_{\mathcal{K}})^{I'}$ .

Differently from the case of knowledge-independent tooth operators, the negation of a knowledge-dependent tooth operator is not in general equivalent to any other knowledge-dependent tooth operator. The reason is that, given any knowledge-dependent tooth operator  $\mathbb{C}_{\mathcal{K}} = \mathbb{W}_{\mathcal{K}}^t((C_i : w_i)_{i=1\dots n})$ , any interpretation  $I$  and any individual  $d \in \Delta^I$ , by definition  $d \in (\mathbb{C}_{\mathcal{K}})^I$  if and only if there is *some* individual name  $a \in N_{\text{Ind}}$  such that  $a^I = d$  and  $\omega_{\mathbb{C}_{\mathcal{K}}}(a) \geq t$ . Thus, the following result is straightforward:

**Proposition 7.** Let  $\mathbb{C}_{\mathcal{K}} = \mathbb{W}_{\mathcal{K}}^t((C_i : w_i)_{i=1\dots n})$  be any knowledge-dependent tooth operator, let  $I$  be any interpretation and let  $d \in \Delta^I$  be an individual such that  $d \in (\mathbb{C}_{\mathcal{K}})^I$ . Furthermore, let  $a$  be any individual name such that  $a^I \neq d$ , and let  $I'$  be an interpretation that differs from  $I$  only in that  $a^{I'} = d$ . Then  $d \in (\mathbb{C}_{\mathcal{K}})^{I'}$  as well.

On the other hand  $d \notin (\mathbb{C}_{\mathcal{K}})^{I'}$  if and only if for all  $a \in N_{\text{Ind}}$  for which  $a^{I'}$  is defined it is the case that  $\omega_{\mathbb{C}_{\mathcal{K}}}(a) < t$ . Thus, the above property does not hold for the negation of a knowledge-dependent tooth operator, as shown in more detail in the following simple example:

**Example 5.** Let  $\mathcal{K} = \{A(a)\}$  and let  $I$  be an interpretation with  $\Delta^I = \{0, 1\}$ ,  $a^I = 0$  and  $b^I = 1$ . Then it is easy to see that  $1 \notin (\mathbb{W}_{\mathcal{K}}^1(A : 1))^I$ , and hence  $I \models \neg \mathbb{W}_{\mathcal{K}}^1(A : 1)(b)$ .

However, let  $I'$  differ from  $I$  only in that  $a^{I'} = b^{I'} = 1$ : then  $1 \in (\mathbb{W}_{\mathcal{K}}^1(A : 1))^{I'}$ , since  $1 = a^{I'}$  and  $\omega_{\mathbb{W}_{\mathcal{K}}^1(A:1)}(a) \geq 1$ , and therefore  $I' \not\models \neg \mathbb{W}_{\mathcal{K}}^1(A : 1)(b)$ .

Therefore, by Proposition 7, there is no knowledge-dependent tooth expression  $\mathbb{C}_{\mathcal{K}}$  that is equivalent to the complement of  $\mathbb{W}_{\mathcal{K}}^1(A : 1)$ .

<sup>5</sup>“Hanno” was the name of the white elephant gifted from King Manuel I of Portugal to Pope Leo X, who apparently appreciated it so much that he commissioned Raphael to make a portrait of it (now lost).

<sup>6</sup>For instance, because Raphael’s portrait of Hanno was discovered, and it turns out that Hanno *did* have antlers.

Thus, it is not in general true that the negation of a knowledge-dependent tooth operator is equivalent to some other knowledge-dependent tooth operator.

Similarly, and again differently from the knowledge-independent case, it is not in general possible – given a knowledge-dependent tooth operator  $C_{\mathcal{K}}$  – to find an equivalent knowledge-dependent tooth operator  $D_{\mathcal{K}}$  in which all weights are positive. This is a consequence of the fact that knowledge-dependent tooth operators in which all weights are positive are *monotone* with respect to the knowledge base  $\mathcal{K}$ :

**Proposition 8.** *Let  $C_{\mathcal{K}} = \mathbb{W}_{\mathcal{K}}^t((C_i : w_i)_{i=1\dots n})$  be any knowledge-dependent tooth operator such that all weights  $w_i$  are positive, let  $\mathcal{K}' \supseteq \mathcal{K}$  be any knowledge base that contains  $\mathcal{K}$ , and let us write  $C_{\mathcal{K}'}$  for  $\mathbb{W}_{\mathcal{K}'}^t((C_i : w_i)_{i=1\dots n})$ . Then for all interpretations  $I$ ,  $I \models C_{\mathcal{K}} \sqsubseteq C_{\mathcal{K}'}$ .*

*Proof.* Let  $d \in \Delta^I$  be such that  $d \in (C_{\mathcal{K}})^I$ . Then, by definition,  $d = a^I$  for some individual name  $a$  such that  $\sum\{w_i : \mathcal{K} \models C_i(a)\} = q \geq t$ . But for any  $i$ , if  $\mathcal{K} \models C_i(a)$  then  $\mathcal{K}' \models C_i(a)$ . Therefore, since all  $w_i$  are positive,  $\sum\{w_i : \mathcal{K}' \models C_i(a)\} \geq q \geq t$  and  $d \in (C_{\mathcal{K}'})^I$ .  $\square$

Since, as we saw in Example 4, there exist knowledge-dependent tooth expressions  $C_{\mathcal{K}}$  with negative weights that are not monotone with respect to  $\mathcal{K}$ , it follows at once that these expressions are not equivalent to any other tooth expression with only positive weights.

On the other hand, it is still possible to express the notion that a knowledge-dependent tooth operator takes values between two thresholds:

**Proposition 9.** *Let  $C_1 \dots C_n$  be concepts, let  $w_1 \dots w_n \in \mathbb{R}$  be weights, and let  $t, t' \in \mathbb{R}$  be such that  $t \leq t'$ , and let  $\mathcal{K}$  be a knowledge base. Then, for all interpretations  $I$ ,*

$$d \in (\mathbb{W}_{\mathcal{K}}^t(C_1 : w_1 \dots C_n : w_n) \sqcap \mathbb{W}_{\mathcal{K}}^{-t'}(C_1 : -w_1 \dots C_n : -w_n))^I$$

*if and only if  $t \leq \omega_{C_{\mathcal{K}}}(d) \leq t'$ , where  $C$  represents the knowledge-dependent tooth expression  $\mathbb{W}_{\mathcal{K}}^t(C_1 : w_1 \dots C_n : w_n)$ .*

## 5 Learning Tooth Expressions: A Toy Example

We will now briefly discuss a toy example of *learning teeth expressions from data*.

The scenario that we will examine is the following:

- We have a knowledge base  $\mathcal{K}$ ;
- We have a *target concept*  $C_T$ , i.e., the concept we wish to be able to approximate via a tooth expression;
- We have a finite number of *concept features*  $C_1 \dots C_n$ , representing the various concepts that may appear in the tooth expression simulating  $C_T$ .

The output of our procedure will be a tooth expression  $\mathbb{W}^t(C_1 : w_1 \dots C_n : w_n)$  which will, in some sense that will depend on the algorithm used, approximate  $C_T$  on the basis of  $C_1 \dots C_n$ .

Let  $a$  be any named individual occurring in the knowledge base  $\mathcal{K}$ . Then  $a$  induces a pair  $(\vec{x}(a), y(a))$ , where  $\vec{x}(a) \in \{0, 1\}^n$  is such that its  $i$ -th element  $\vec{x}(a)_i$  is 1 if  $\mathcal{K} \models C_i(a)$  and it is 0 otherwise; and likewise,  $y(a)$  is 1 if  $\mathcal{K} \models C_T(a)$  and  $y(a) = 0$  otherwise.

Then we can define the dataset  $\mathbf{D}(\mathcal{K}) = \{(\vec{x}(a), y(a)) : a \text{ named individual in } \mathcal{K}\}$ ; and our task will be to learn from it a *linear classification model* of the sort

$$y(a) = \begin{cases} 1 & \text{if } \sum_i x(a)_i w_i + c \geq 0; \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

according to one of the many learning algorithms that are available for such linear models. Once the weights  $w_1 \dots w_n$  and the constant term  $c$  are learned, it will be trivial to transform this linear model in the equivalent tooth expression  $\mathbb{W}_{\mathcal{K}}^{-c}(C_1 : w_1 \dots C_n : w_n)$ .

For the purpose of this toy example, we chose to learn the weights and the constant using *logistic regression*. Very briefly, the logistic regression model computes, for every input instance  $\vec{x} \in \{0, 1\}^n$ , the probability that  $y = 1$  as

$$\Pr(y = 1) = \frac{1}{1 + e^{-(\sum_i w_i x_i + c)}}.$$

It is straightforward then to verify that  $\Pr(y = 1) \geq 0.5$  if and only if  $e^{-(\sum_i w_i x_i + c)} \leq 1$ , that is,  $\sum_i w_i x_i + c \geq 0$ ; thus, if for any possible input  $\vec{x}$  we predict  $y = 1$  if  $\Pr(y = 1) \geq 0.5$  and  $y = 0$  otherwise, we obtain precisely Equation (8). We can then learn the weights and the threshold, given the training dataset  $\mathbf{D}(\mathcal{K})$ , simply by minimizing (for example by gradient descent) the *loss function*

$$-\sum \{\log \Pr(y(a) = 1) : y(a) = 1\} - \sum \{\log \Pr(y(a) = 0) : y(a) = 0\} + \gamma \cdot \sum_i w_i^2$$

where  $\gamma$  is a *ridge regularization parameter* that penalizes high values for the  $w_i$ .

In order to test this approach we generated a synthetic knowledge base  $\mathcal{K}$  containing the TBox axioms  $\text{IsBig} \sqcap \text{IsSmall} \sqsubseteq \perp$  and  $\text{IsGrey} \sqcap \text{IsPink} \sqsubseteq \perp$  and the following individuals and ABox axioms:

- 200 individuals Elephant0...Elephant199, every one of which satisfies – of course –  $\text{IsElephant}$  and each one of which satisfies  $\text{HasTrunk}$  with probability 0.9,  $\text{HasAntlers}$  with probability 0.1,  $\text{HasTusks}$  with probability 0.7,  $\text{IsBig}$  with probability 0.7 and otherwise  $\text{IsSmall}$  with probability 0.5,  $\text{IsGrey}$  with probability 0.8 and otherwise  $\text{IsPink}$  with probability 0.5;
- 200 individuals Deer0...Deer199, none of which satisfies  $\text{IsElephant}$  and each one of which satisfies  $\text{HasTrunk}$  with probability 0.1,  $\text{HasAntlers}$  with probability 0.5,  $\text{HasTusks}$  with probability 0.05,  $\text{IsBig}$  with probability 0.3 and otherwise  $\text{IsSmall}$  with probability 0.3,  $\text{IsGrey}$  with probability 0.3 and otherwise  $\text{IsPink}$  with probability 0.1;
- 200 individuals Mouse0...Mouse199, none of which satisfies  $\text{IsElephant}$  and each one of which satisfies  $\text{HasTrunk}$  with probability 0.1,  $\text{HasAntlers}$  with probability 0.05,  $\text{HasTusks}$  with probability 0.05,  $\text{IsSmall}$  with probability 0.9,  $\text{IsGrey}$  with probability 0.8 and otherwise  $\text{IsPink}$  with probability 0.8.

Then we used the above sketched approach to attempt to learn a tooth operator that would approximate  $\text{IsElephant}$  on the basis of  $\text{HasTrunk}$ ,  $\text{HasAntlers}$ ,  $\text{HasTusks}$ ,  $\text{IsBig}$  and  $\text{IsPink}$ . The ridge parameter  $\gamma$  was set (fairly arbitrarily) to the small value of  $10^{-8}$ .

The result was the tooth expression  $\text{LearnedElephant}_{\mathcal{K}}$

$$\mathbb{W}_{\mathcal{K}}^{5.26}(\text{HasTrunk} : 5.11, \text{IsBig} : 2.95, \text{HasTusks} : 3.75, \text{HasAntlers} : -1.93, \text{IsPink} : 0.07)$$

which may be seen as an approximation of the property  $\text{IsElephant}$  in terms of the feature concepts, learned on the basis of the data of our synthetic knowledge base. Thus, for instance, we see that the property of having a trunk affects positively our estimation that the individual in exam might be an elephant; that the one of having antlers affects it negatively instead; and that – given this particular dataset, that has what some might describe as a unusually high frequency of pink elephants – the property of being pink has essentially no effect on it. Furthermore, given another knowledge base  $\mathcal{K}'$  which contains the atomic concepts  $\text{HasTrunk}$ ,  $\text{HasAntlers}$ ,  $\text{HasTusks}$ ,  $\text{IsBig}$  and  $\text{IsPink}$  but not the concept  $\text{IsElephant}$ , we may use this  $\text{LearnedElephant}_{\mathcal{K}'}$  as a “replacement” for the missing  $\text{IsElephant}$  concept.

Of course, there is nothing really new in this logistic regression model; and, if this was anything more than a toy example, we should perform a thorough cross-validation and report on the behaviour of the logistic model and the corresponding tooth operator on an alternate dataset. We report this here merely as a simple illustration of how the tooth operator may be used as a “bridge” of sorts between statistical learning and knowledge representation.

## 6 Conclusions and Further Work

In this work we continued the exploration, began in [15] and [16], of a new family of concept-forming operators which take a list of weighted concepts and return a new concept that satisfies ‘enough’ of the input concepts. While [15] studied basic logical properties of these operators and [16] explored applications to cognitive modelling, we here focus on analysing in greater precision different notions of the semantics of tooth expressions, clarifying in particular the distinction between “knowledge-dependent” and “knowledge-independent” interpretations of tooth expressions and their semantics. Here, the “knowledge-dependent” version of a tooth considers only the weights of features entailed by a knowledge base, and is a non-monotonic operator, whilst the “knowledge-independent” version evaluates any given model, and behaves monotonically.

We explored some basic properties of these two types of tooth expressions, studied some of their interdependencies, and finally presented a simple example of an application bridging the gap between statistical learning and knowledge representation. The obvious next steps, aside from further exploring the formal properties of these operators, consist in studying the behaviour and the possibilities of these operators over real datasets and in studying in more detail *nested* tooth expressions and their connection to non-linear classification models (and, in particular, to multi-layer perceptrons), and in exploring the connections and similarities between our approach and other approaches to learning over knowledge bases such as [17], [5], [12], [13] and [11]. We also plan to study the relationship between our approach and other extensions of description logics with graded membership values and thresholds such as [2, 1].<sup>7</sup>

We hope that a deeper exploration of the correspondence theory between tooth logic, statistical learning and classification models, and concept learning in DL, will not only contribute to these fields individually, but will allow for hybrid frameworks where e.g. statistically learned concepts can be naturally represented in DL knowledge bases, thereby contributing also to neural-symbolic integration and to the AI explainability challenge [6].

## References

- [1] Franz Baader, Gerhard Brewka, and Oliver Fernández Gil. Adding threshold concepts to the description logic  $\mathcal{EL}$ . In *International Symposium on Frontiers of Combining Systems*, pages 33–48. Springer, 2015.
- [2] Franz Baader and Andreas Ecke. Reasoning with prototypes in the description logic  $\mathcal{ALC}$  using weighted tree automata. In Adrian-Horia Dediu, Jan Janoušek, Carlos Martín-Vide, and Bianca Truthe, editors, *Language and Automata Theory and Applications*, pages 63–75. Cham, 2016. Springer International Publishing.
- [3] Franz Baader, Ian Horrocks, Carsten Lutz, and Uli Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.
- [4] Amos Beimel and Enav Weinreb. Monotone circuits for monotone weighted threshold functions. *Information Processing Letters*, 97(1):12 – 18, 2006.
- [5] Nicola Fanizzi, Claudia d’Amato, and Floriana Esposito. DL-FOIL concept learning in description logics. In *International Conference on Inductive Logic Programming*, pages 107–121. Springer, 2008.

<sup>7</sup>Our approach is substantially different from the ones cited in that it does not require us to change the nature of interpretations or the semantics of the logic.

- [6] Artur d'Avila Garcez, Marco Gori, Luis C Lamb, Luciano Serafini, Michael Spranger, and Son N Tran. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *arXiv preprint arXiv:1905.06088*, 2019.
- [7] Michael R Garey and David S Johnson. *Computers and intractability*, volume 29. wh freeman New York, 2002.
- [8] Mikael Goldmann, Johan Håstad, and Alexander Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2(4):277–300, Dec 1992.
- [9] Mikael Goldmann and Marek Karpinski. Simulating threshold circuits by majority circuits. *SIAM Journal on Computing*, 27(1):230–246, 1998.
- [10] András Hajnal, Wolfgang Maass, Pavel Pudlák, Máriaó Szegedy, and György Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46(2):129 – 154, 1993.
- [11] Patrick Hohenecker and Thomas Lukasiewicz. Deep learning for ontology reasoning. *arXiv preprint arXiv:1705.10342*, 2017.
- [12] Jens Lehmann. DL-Learner: Learning Concepts in Description Logics. *Journal of Machine Learning Research*, 10(Nov):2639–2642, 2009.
- [13] Jens Lehmann and Pascal Hitzler. Concept learning in description logics using refinement operators. *Machine Learning*, 78(1-2):203, 2010.
- [14] Claudio Masolo and Daniele Porello. Representing concepts by weighted formulas. In *Formal Ontology in Information Systems - Proceedings of the 10th International Conference, FOIS 2018, Cape Town, South Africa, 19-21 September 2018*, pages 55–68, 2018.
- [15] Daniele Porello, Oliver Kutz, Guendalina Righetti, Nicolas Troquard, Pietro Galliani, and Claudio Masolo. A toothful of concepts: Towards a theory of weighted concept combination. In *Proceedings of the 32nd International Workshop on Description Logics*, volume 2373. CEUR-WS, 2019.
- [16] Guendalina Righetti, Daniele Porello, Oliver Kutz, Nicolas Troquard, and Claudio Masolo. Pink panthers and toothless tigers: Three problems in classification. In *Proc. of the 5th International Workshop on Artificial Intelligence and Cognition*, Manchester, September 10–11, 2019.
- [17] Volker Tresp, Markus Bundschuh, Achim Rettinger, and Yi Huang. Towards machine learning on the semantic web. In *Uncertainty reasoning for the Semantic Web I*, pages 282–314. Springer, 2006.