From Weighted Conditionals of Multilayer Perceptrons to a Gradual Argumentation Semantics

Laura Giordano[1]

[1]DISIT - Università del Piemonte Orientale, Italy

Abstract
A fuzzy multipreference semantics has been recently proposed for weighted conditional knowledge bases, and used to develop a logical semantics for Multilayer Perceptrons, by regarding a deep neural network (after training) as a weighted conditional knowledge base. Based on some different variants of this semantics, we propose some new gradual argumentation semantics, and relate them to the family of the gradual semantics. The relationships between weighted conditional knowledge bases and MLPs extend to the proposed gradual semantics to capture the stationary states of MPs, in agreement with previous results on the relationship between argumentation frameworks and neural networks.

1. Introduction

Argumentation is a reasoning approach which, in its different formulations and semantics, has been used in different contexts in the multi-agent setting, from social networks [1] to classification [2], and it is very relevant for decision making and for explanation [3]. The argumentation semantics are strongly related to other non-monotonic reasoning formalisms and semantics [4, 5].

Our starting point in this paper is a preferential semantics for commonsense reasoning which has been proposed for a description logic with typicality. Preferential description logics have been studied in the last fifteen years to deal with inheritance with exceptions in ontologies, based on the idea of extending the language of Description Logics (DLs), by allowing for non-strict forms of inclusions, called *typicality or defeasible inclusions*, of the form $\mathbf{T}(C) \sqsubseteq D$ (meaning "the typical C-elements are D-elements" or "normally C's are D's"), with different preferential semantics [6, 7] and closure constructions [8, 9, 10, 11, 12]. Such defeasible inclusions correspond to KLM conditionals $C \mid\sim D$ [13, 14], and defeasible DLs inherit and extend some of the preferential semantics and the closure constructions developed within preferential and conditional approaches to commonsense reasoning [13, 15, 14, 16, 17].

In previous work [18], a concept-wise multipreference semantics for weighted conditional knowledge bases (KBs) has been proposed to account for preferences with respect to different concepts, by allowing a set of typicality inclusions of the form $\mathbf{T}(C) \sqsubseteq D$ with positive or negative weights, for distinguished concepts C. The concept-wise multipreference semantics has been first introduced as a semantics for ranked DL knowledge bases [19] (where conditionals

5th Workshop on Advances In Argumentation In Artificial Intelligence (AIxIA 2021)
laura.giordano@uniupo.it (L. Giordano)
© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3086/paper8.pdf

are associated a positive integer rank), and later extended to weighted conditional KBs (in the two-valued and in the fuzzy case), based on a different semantic closure construction, still in the spirit of Lehmann's lexicographic closure [14] and Kern-Isberner's c-representations [20, 21], but exploiting multiple preferences with respect to concepts.

The concept-wise multipreference semantics has been proven to have some desired properties from the knowledge representation point of view in the two-valued case [19, 22]: it satisfies the KLM properties of a preferential consequence relation [13, 14], it allows to deal with specificity and irrelevance and avoids inheritance blocking or the "drowning problem" [15, 17], and deals with "ambiguity preservation" [16]. The plausibility of the concept-wise multipreference semantics has also been supported [23, 24] by showing that it is able to provide a logical interpretation to Kohonen' Self-Organising Maps [25], which are psychologically and biologically plausible neural network models. In the fuzzy case, the KLM properties of non-monotomic entailment have been studied in [26], showing that *most* KLM postulates are satisfied, depending on their reformulation and on the choice of fuzzy combination functions. It has been shown [18] that both in the two-valued and in the fuzzy case, the multi-preferential semantics allows to describe the input-output behavior of Multilayer Perceptrons (MLPs), after training, in terms of a preferential interpretation which, in the fuzzy case, can be proved to be a model (in a logical sense) of the weighted KB which is associated to the neural network.

The relationships between preferential and conditional approaches to non-monotonic reasoning and argumentation semantics are strong. Let us also mention, the work by Geffner and Pearl on Conditional Entailment, whose proof theory is defined in terms of "arguments" [16]. In this paper we aim at investigating the relationships between the fuzzy multipreference semantics for weighted conditionals and gradual argumentation semantics [27, 28, 29, 30, 31, 32, 33, 34]. To this purpose, in addition to the notions of coherent and faithful fuzzy multipreference semantics [18, 26], in Section 4, we introduce a notion of φ-coherent (fuzzy) multipreference semantics. In Section 5, we propose three new gradual semantics for a weighted argumentation graph (namely, a coherent, a faithful and a φ-coherent semantics) inspired by the fuzzy preferential semantics of weighted conditionals and, in Section 6, we investigate the relationship of φ-coherent semantics with the family of gradual semantics studied by Amgoud and Doder. The relationships between weighted conditional knowledge bases and MLPs easily extend to the proposed gradual semantics, which captures the stationary states of MLPs. This is in agreement with the previous results on the relationships between argumentation frameworks and neural networks by Garces, Gabbay and Lamb [35] and by Potyca [36]. Section 7 concludes the paper by suggesting a possible approach for defeasible reasoning building on a gradual semantics, as considered in an extended version of this paper [37].

2. The description logic $\mathcal{LC}$ and fuzzy $\mathcal{LC}$

In this section we recall the syntax and semantics of the description logic $\mathcal{ALC}$ [38] and of its fuzzy extension [39]. For sake of simplicity, we only focus on $\mathcal{LC}$, the boolean fragment of $\mathcal{ALC}$, which does not allow for roles. Let N_C be a set of concept names, and N_I a set of individual names. The set of $\mathcal{LC}$ *concepts* (or, simply, concepts) can be defined inductively:
- $A \in N_C$, $\top$ and $\bot$ are concepts;

- if C and D are concepts, and $r \in N_R$, then $C \sqcap D$, $C \sqcup D$, $\neg C$ are concepts.
An $\mathcal{LC}$ knowledge base (KB) K is a pair $(\mathcal{T}_K, \mathcal{A}_K)$, where $\mathcal{T}_K$ is a TBox and $\mathcal{A}_K$ is an ABox. The TBox $\mathcal{T}_K$ is a set of concept inclusions (or subsumptions) $C \sqsubseteq D$, where C, D are concepts. The ABox $\mathcal{A}_K$ is a set of assertions of the form $C(a)$, where C is a concept and a an individual name in N_I.

An $\mathcal{LC}$ *interpretation* is defined as a pair $I = \langle \Delta, \cdot^I \rangle$ where: Δ is a domain—a set whose elements are denoted by $x, y, z, \ldots$—and $\cdot^I$ is an extension function that maps each concept name $C \in N_C$ to a set $C^I \subseteq \Delta$, and each individual name $a \in N_I$ to an element $a^I \in \Delta$. It is extended to complex concepts as follows:

$$\top^I = \Delta \qquad \bot^I = \emptyset \qquad (\neg C)^I = \Delta \backslash C^I$$
$$(C \sqcap D)^I = C^I \cap D^I \qquad (C \sqcup D)^I = C^I \cup D^I$$

The notion of satisfiability of a KB in an interpretation and the notion of entailment are defined as follows:

Definition 1 (Satisfiability and entailment). *Given an $\mathcal{LC}$ interpretation $I = \langle \Delta, \cdot^I \rangle$:*
- I satisfies an inclusion $C \sqsubseteq D$ if $C^I \subseteq D^I$;
- I satisfies an assertion $C(a)$ if $a^I \in C^I$.
Given a KB $K = (\mathcal{T}_K, \mathcal{A}_K)$, an interpretation I satisfies $\mathcal{T}_K$ (resp. $\mathcal{A}_K$) if I satisfies all inclusions in $\mathcal{T}_K$ (resp. all assertions in $\mathcal{A}_K$); I is a model *of K if I satisfies $\mathcal{T}_K$ and $\mathcal{A}_K$.*

A subsumption $F = C \sqsubseteq D$ (resp., an assertion $C(a)$), is entailed by K, written $K \models F$, if for all models $I = \langle \Delta, \cdot^I \rangle$ of K, I satisfies F.

Given a knowledge base K, the *subsumption* problem is the problem of deciding whether an inclusion $C \sqsubseteq D$ is entailed by K.

Fuzzy description logics have been widely studied in the literature for representing vagueness in DLs [40, 41, 39, 42, 43], based on the idea that concepts and roles can be interpreted as fuzzy sets. Formulas in Mathematical Fuzzy Logic [44] have a degree of truth in an interpretation rather than being true or false; similarly, axioms in a fuzzy DL have a degree of truth, usually in the interval $[0, 1]$. In the following we shortly recall the semantics of a fuzzy extension of $\mathcal{ALC}$ for the fragment $\mathcal{LC}$, referring to the survey by Lukasiewicz and Straccia [39]. We limit our consideration to a few features of a fuzzy DL, without considering roles, datatypes, and restricting to the language of $\mathcal{LC}$.

A *fuzzy interpretation* for $\mathcal{LC}$ is a pair $I = \langle \Delta, \cdot^I \rangle$ where: Δ is a non-empty domain and $\cdot^I$ is *fuzzy interpretation function* that assigns to each concept name $A \in N_C$ a function $A^I : \Delta \rightarrow [0, 1]$, and to each individual name $a \in N_I$ an element $a^I \in \Delta$. A domain element $x \in \Delta$ belongs to the extension of A to some degree in $[0, 1]$, i.e., A^I is a fuzzy set.

The interpretation function $\cdot^I$ is extended to complex concepts as follows:
$$\top^I(x) = 1, \qquad \bot^I(x) = 0, \qquad (\neg C)^I(x) = \ominus C^I(x),$$
$$(C \sqcap D)^I(x) = C^I(x) \otimes D^I(x), \qquad (C \sqcup D)^I(x) = C^I(x) \oplus D^I(x).$$
where $x \in \Delta$ and $\otimes$, $\oplus$, $\rhd$ and $\ominus$ are arbitrary but fixed t-norm, s-norm, implication function, and negation function, chosen among the combination functions of various fuzzy logics (we refer to [39] for details). For instance, in Zadeh logic $a \otimes b = min\{a, b\}$, $a \oplus b = max\{a, b\}$, $a \rhd b = max\{1 - a, b\}$ and $\ominus a = 1 - a$.

The interpretation function $\cdot^I$ is also extended to non-fuzzy axioms (i.e., to strict inclusions and assertions of an $\mathcal{LC}$ knowledge base) as follows:
$(C \sqsubseteq D)^I = inf_{x \in \Delta} C^I(x) \rhd D^I(x)$, $\quad$ $(C(a))^I = C^I(a^I)$.

A *fuzzy $\mathcal{LC}$ knowledge base* K is a pair $(\mathcal{T}_f, \mathcal{A}_f)$ where $\mathcal{T}_f$ is a fuzzy TBox and $\mathcal{A}_f$ a fuzzy ABox. A fuzzy TBox is a set of *fuzzy concept inclusions* of the form $C \sqsubseteq D\ \theta\ n$, where $C \sqsubseteq D$ is an $\mathcal{LC}$ concept inclusion axiom, $\theta \in \{\geq, \leq, >, <\}$ and $n \in [0, 1]$. A fuzzy ABox $\mathcal{A}_f$ is a set of *fuzzy assertions* of the form $C(a)\theta n$, where C is an $\mathcal{LC}$ concept, $a \in N_I$, $\theta \in \{\geq, \leq, >, <\}$ and $n \in [0, 1]$. Following Bobillo and Straccia [43], we assume that fuzzy interpretations are *witnessed*, i.e., the *sup* and *inf* are attained at some point of the involved domain. The notions of satisfiability of a KB in a fuzzy interpretation and of entailment are defined in the natural way.

Definition 2 (Satisfiability and entailment for fuzzy KBs). *A fuzzy interpretation I satisfies a fuzzy $\mathcal{LC}$ axiom E (denoted $I \models E$), as follows, for $\theta \in \{\geq, \leq, >, <\}$:*
- I satisfies a fuzzy $\mathcal{LC}$ inclusion axiom $C \sqsubseteq D\ \theta\ n$ if $(C \sqsubseteq D)^I \theta\ n$;
- I satisfies a fuzzy $\mathcal{LC}$ assertion $C(a)\ \theta\ n$ if $C^I(a^I)\theta\ n$;
Given a fuzzy $\mathcal{LC}$ KB $K = (\mathcal{T}_f, \mathcal{A}_f)$, a fuzzy interpretation I satisfies $\mathcal{T}_f$ (resp. $\mathcal{A}_f$) if I satisfies all fuzzy inclusions in $\mathcal{T}_f$ (resp. all fuzzy assertions in $\mathcal{A}_f$). A fuzzy interpretation I is a model *of K if I satisfies $\mathcal{T}_f$ and $\mathcal{A}_f$. A fuzzy axiom E is entailed by a fuzzy knowledge base K (i.e., $K \models E$) if for all models $I = \langle \Delta, \cdot^I \rangle$ of K, I satisfies E.*

3. Fuzzy $\mathcal{LC}$ with typicality: $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$

In this section, we describe an extension of fuzzy $\mathcal{LC}$ with typicality following [18, 26]. Typicality concepts of the form $\mathbf{T}(C)$ are added, where C is a concept in fuzzy $\mathcal{LC}$. The idea is similar to the extension of $\mathcal{ALC}$ with typicality under the two-valued semantics [6] but transposed to the fuzzy case. The extension allows for the definition of *fuzzy typicality inclusions* of the form $\mathbf{T}(C) \sqsubseteq D\ \theta\ n$, meaning that typical C-elements are D-elements with a degree greater than n. A typicality inclusion $\mathbf{T}(C) \sqsubseteq D$, as in the two-valued case, stands for a KLM conditional implication $C \mid\!\sim D$ [13, 14], but now it has an associated degree. We call $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ the extension of fuzzy $\mathcal{LC}$ with typicality. As in the two-valued case, and in the propositional typicality logic, PTL, [45] the nesting of the typicality operator is not allowed.

Observe that, in a fuzzy $\mathcal{LC}$ interpretation $I = \langle \Delta, \cdot^I \rangle$, the degree of membership $C^I(x)$ of the domain elements x in a concept C, induces a preference relation $<_C$ on Δ, as follows:

$$x <_C y \text{ iff } C^I(x) > C^I(y) \tag{1}$$

Each $<_C$ has the properties of preference relations in KLM-style ranked interpretations [14], that is, $<_C$ is a modular and well-founded strict partial order. Let us recall that, $<_C$ is *well-founded* if there is no infinite descending chain $x_1 <_C x_0$, $x_2 <_C x_1$, $x_3 <_C x_2, \ldots$ of domain elements; $<_C$ is *modular* if, for all $x, y, z \in \Delta$, $x <_C y$ implies ($x <_C z$ or $z <_C y$). Well-foundedness holds for the induced preference $<_C$ defined by condition (1) under the assumption that fuzzy interpretations are witnessed [43] (see Section 2) or that Δ is finite. For simplicity, we will assume Δ to be finite.

Each preference relation $<_C$ has the properties of a preference relation in KLM rational interpretations [14] (also called ranked interpretations), but here there are multiple preferences and, therefore, fuzzy interpretations can be regarded as *multipreferential* interpretations, which have been also studied in the two-valued case [19, 46, 47]. Preference relation $<_C$ captures the relative typicality of domain elements wrt concept C and may then be used to identify the *typical C-elements*. We will regard typical C-elements as the domain elements x that are preferred with respect to relation $<_C$ among those such that $C^I(x) \neq 0$. Let $C^I_{>0}$ be the crisp set containing all domain elements x such that $C^I(x) > 0$, that is, $C^I_{>0} = \{x \in \Delta \mid C^I(x) > 0\}$. One can provide a (two-valued) interpretation of typicality concepts $\mathbf{T}(C)$ in a fuzzy interpretation I, by letting:

$$(\mathbf{T}(C))^I(x) = \begin{cases} 1 & \text{if } x \in min_{<_C}(C^I_{>0}) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $min_<(S) = \{u : u \in S \text{ and } \nexists z \in S \text{ s.t. } z < u\}$. When $(\mathbf{T}(C))^I(x) = 1$, we say that x is a typical C-element in I.

Note that, if $C^I(x) > 0$ for some $x \in \Delta$, $min_{<_C}(C^I_{>0})$ is non-empty.

Definition 3 ($\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ interpretation). *An $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ interpretation $I = \langle \Delta, \cdot^I \rangle$ is a fuzzy $\mathcal{LC}$ interpretation, extended by interpreting typicality concepts as in (2).*

The fuzzy interpretation $I = \langle \Delta, \cdot^I \rangle$ implicitly defines a multipreference interpretation, where any concept C is associated to a preference relation $<_C$. This is different from the two-valued multipreference semantics in [19], where only a subset of distinguished concepts have an associated preference, and a notion of global preference $<$ is introduced to define the interpretation of the typicality concept $\mathbf{T}(C)$, for an arbitrary C. Here, we do not need to introduce a notion of global preference. The interpretation of any $\mathcal{LC}$ concept C is defined compositionally from the interpretation of atomic concepts, and the preference relation $<_C$ associated to C is defined from C^I. The notions of *satisfiability* in $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$, *model* of an $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ knowledge base, and $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ *entailment* can be defined in a similar way as in fuzzy $\mathcal{LC}$ (see Section 2).

3.1. Strengthening $\mathcal{LC}^{\mathbf{F}}\mathrm{T}$: a closure construction

To overcome the weakness of preferential entailment, the rational closure [14] and the lexicographic closure of a conditional knowledge base [48] have been introduced, to allow for further inferences. In this section, we recall a closure construction introduced to strengthen $\mathcal{ALC}^{\mathbf{F}}\mathbf{T}$ entailment for weighted conditional knowledge bases, where typicality inclusions are associated real-valued weights. In the two-valued case, the construction is related to the definition of Kern-Isberner's c-representations [20, 21], which include penalty points for falsified conditionals. In the fuzzy case, the construction also relates to the fuzzy extension of rational closure by Casini and Straccia [49].

A *weighted $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ knowledge base K*, over a set $\mathcal{C} = \{C_1, \ldots, C_k\}$ of distinguished $\mathcal{LC}$ concepts, is a tuple $\langle \mathcal{T}_f, \mathcal{T}_{C_1}, \ldots, \mathcal{T}_{C_k}, \mathcal{A}_f \rangle$, where $\mathcal{T}_f$ is a set of fuzzy $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ inclusion axiom, $\mathcal{A}_f$ is a set of fuzzy $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ assertions and $\mathcal{T}_{C_i} = \{(d^i_h, w^i_h)\}$ is a set of all weighted typicality inclusions $d^i_h = \mathbf{T}(C_i) \sqsubseteq D_{i,h}$ for C_i, indexed by h, where each inclusion d^i_h has weight w^i_h, a real number. As in [18], the typicality operator is assumed to occur only on the left hand side of a weighted typicality inclusion, and we call *distinguished concepts* those concepts C_i occurring on the l.h.s. of some typicality inclusion $\mathbf{T}(C_i) \sqsubseteq D$. Arbitrary $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ inclusions and assertions may belong to $\mathcal{T}_f$ and $\mathcal{A}_f$. Let us consider the following example from [26].

Example 1. *Consider the weighted knowledge base* $K = \langle \mathcal{T}_f, \mathcal{T}_{Bird}, \mathcal{T}_{Penguin}, \mathcal{T}_{Canary}, \mathcal{A}_f \rangle$*, over the set of distinguished concepts* $\mathcal{C} = \{Bird, Penguin, Canary\}$*, with empty* $\mathcal{A}_f$ *and* $\mathcal{T}_f$ *containing, for instance, the inclusions:*

$Yellow \sqcap Black \sqsubseteq \bot \geq 1$ $\quad$ $Yellow \sqcap Red \sqsubseteq \bot \geq 1$ $\quad$ $Black \sqcap Red \sqsubseteq \bot \geq 1$

The weighted TBox $\mathcal{T}_{Bird}$ *contains the following weighted defeasible inclusions:*

(d_1) $\mathbf{T}(Bird) \sqsubseteq Fly$, *+20* $\quad$ (d_2) $\mathbf{T}(Bird) \sqsubseteq Has\_Wings$, *+50*

(d_3) $\mathbf{T}(Bird) \sqsubseteq Has\_Feather$, *+50;*

$\mathcal{T}_{Penguin}$ *and* $\mathcal{T}_{Canary}$ *contain, respectively, the following defeasible inclusions:*

(d_4) $\mathbf{T}(Penguin) \sqsubseteq Bird$, *+100* $\quad$ (d_7) $\mathbf{T}(Canary) \sqsubseteq Bird$, *+100*

(d_5) $\mathbf{T}(Penguin) \sqsubseteq Fly$, *- 70* $\quad$ (d_8) $\mathbf{T}(Canary) \sqsubseteq Yellow$, *+30*

(d_6) $\mathbf{T}(Penguin) \sqsubseteq Black$, *+50;* $\quad$ (d_9) $\mathbf{T}(Canary) \sqsubseteq Red$, *+20*

The meaning is that a bird normally has wings, has feathers and flies, but having wings and feather (both with weight 50) for a bird is more plausible than flying (weight 20), although flying is regarded as being plausible. For a penguin, flying is not plausible (inclusion (d_5) *has negative weight -70), while being a bird and being black are plausible properties of prototypical penguins, and* (d_4) *and* (d_6) *have positive weights (100 and 50, respectively). Similar considerations can be done for concept* $Canary$*. Given an Abox in which Reddy is red, has wings, has feather and flies (all with degree 1) and Opus has wings and feather (with degree 1), is black with degree 0.8 and does not fly (*$Fly^I(opus) = 0$*), considering the weights of defeasible inclusions, we may expect Reddy to be more typical than Opus as a bird, but less typical than Opus as a penguin.*

The semantics of a weighted knowledge base is defined in [18] trough a *semantic closure construction*, similar in spirit to Lehmann's lexicographic closure [48], but strictly related to c-representations and, additionally, based on multiple preferences. The construction allows a subset of the $\mathcal{ALC}^{\mathbf{F}}\mathbf{T}$ interpretations to be selected, the interpretations whose induced preference relations $<_{C_i}$, for the distinguished concepts C_i, *faithfully* represent the defeasible part of the knowledge base K.

Let $\mathcal{T}_{C_i} = \{(d^i_h, w^i_h)\}$ be the set of weighted typicality inclusions $d^i_h = \mathbf{T}(C_i) \sqsubseteq D_{i,h}$ associated to the distinguished concept C_i, and let $I = \langle \Delta, \cdot^I \rangle$ be a fuzzy $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ interpretation. In the two-valued case, we would associate to each domain element $x \in \Delta$ and each distinguished concept C_i, a weight $W_i(x)$ of x wrt C_i in I, by *summing the weights* of the defeasible inclusions satisfied by x. However, as I is a fuzzy interpretation, we do not only distinguish between the typicality inclusions satisfied or falsified by x; we also need to consider, for all inclusions $\mathbf{T}(C_i) \sqsubseteq D_{i,h} \in \mathcal{T}_{C_i}$, the degree of membership of x in $D_{i,h}$. Furthermore, in comparing the weight of domain elements with respect to $<_{C_i}$, we give higher preference to the domain elements belonging to C_i (with a degree greater than 0), with respect to those not belonging to C_i (having membership degree 0).

For each domain element $x \in \Delta$ and distinguished concept C_i, *the weight* $W_i(x)$ *of* x *wrt* C_i in the $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ interpretation $I = \langle \Delta, \cdot^I \rangle$ is defined as follows:

$$W_i(x) = \begin{cases} \sum_h w^i_h \; D^I_{i,h}(x) & \text{if } C^I_i(x) > 0 \\ -\infty & \text{otherwise} \end{cases} \tag{3}$$

where $-\infty$ is added at the bottom of all real values.

The value of $W_i(x)$ is $-\infty$ when x is not a C-element (i.e., $C_i^I(x) = 0$). Otherwise, $C_i^I(x) > 0$ and the higher is the sum $W_i(x)$, the more typical is the element x relative to the defeasible properties of C_i. How much x satisfies a typicality property $\mathbf{T}(C_i) \sqsubseteq D_{i,h}$ depends on the value of $D_{i,h}^I(x) \in [0, 1]$, which is weighted by w_h^i in the sum. In the *two-valued case*, $D_{i,h}^I(x) \in \{0, 1\}$, and $W_i(x)$ is the sum of the weights of the typicality inclusions for C satisfied by x, if x is a C-element, and is $-\infty$, otherwise.

Example 2. *Let us consider again Example 1. Let I be an $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ interpretation such that $Fly^I(reddy) = (Has\_Wings)^I(reddy) = (Has\_Feather)^I(reddy) = 1$ and $Red^I(red\text{-}dy) = 1$, i.e., Reddy flies, has wings and feather and is red (and $Black^I(reddy) = 0$). Suppose further that $Fly^I(opus) = 0$ and $(Has\_Wings)^I(opus) = (Has\_\ Feather)^I\ (opus) = 1$ and $Black^I(opus) = 0.8$, i.e., Opus does not fly, has wings and feather, and is black with degree 0.8. Considering the weights of typicality inclusions for Bird, $W_{Bird}(reddy) = 20 + 50 + 50 = 120$ and $W_{Bird}(opus) = 0 + 50 + 50 = 100$. This suggests that reddy should be more typical as a bird than opus. On the other hand, if we suppose further that $Bird^I(reddy) = 1$ and $Bird^I(opus) = 0.8$, then $W_{Penguin}\ (reddy) = 100 - 70 + 0 = 30$ and $W_{Penguin}(opus) = 0.8 \times 100 - 0 + 0.8 \times 50 = 120$, and Reddy should be less typical as a penguin than Opus.*

In [18] a notion of *coherence* is introduced, to force an agreement between the preference relations $<_{C_i}$ induced by a fuzzy interpretation I, for each distinguished concept C_i, and the weights $W_i(x)$ computed, for each $x \in \Delta$, from the conditional knowledge base K, given the interpretation I. This leads to the following definition of a coherent fuzzy multipreference model of a weighted a $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ knowledge base.

Definition 4 (Coherent (fuzzy) multipreference model of K [18]). *Let $K = \langle \mathcal{T}_f, \mathcal{T}_{C_1}, \ldots, \mathcal{T}_{C_k}, \mathcal{A}_f \rangle$ be a weighted $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ knowledge base over $\mathcal{C}$. A* coherent (fuzzy) multipreference model *(cf^m-model) of K is a fuzzy $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ interpretation $I = \langle \Delta, \cdot^I \rangle$ s.t.:*

- *I satisfies the fuzzy inclusions in $\mathcal{T}_f$ and the fuzzy assertions in $\mathcal{A}_f$;*
- *for all $C_i \in \mathcal{C}$, the preference $<_{C_i}$* is coherent to *$\mathcal{T}_{C_i}$, that is, for all $x, y \in \Delta$,*

$$x <_{C_i} y \iff W_i(x) > W_i(y) \tag{4}$$

In a similar way, one can define a *faithful (fuzzy) multipreference model (fm-model) of K* by replacing the *coherence* condition (4) with the following *faithfulness* condition (called weak coherence in [50]): for all $x, y \in \Delta$,

$$x <_{C_i} y \Rightarrow W_i(x) > W_i(y). \tag{5}$$

The weaker notion of faithfulness allows to define a larger class of fuzzy multipreference models of a weighted knowledge base, compared to the class of coherent models. This allows a larger class of monotone non-decreasing activation functions in neural network models to be captured, whose activation function is monotonically non-decreasing (we refer to [50], Sec. 7).

Example 3. *Referring to Example 2 above, where $Bird^I(reddy) = 1$, $Bird^I(opus) = 0.8$, let us further assume that $Penguin^I(reddy) = 0.2$ and $Penguin^I(opus) = 0.8$. Clearly, reddy $<_{Bird}$*

opus and opus $<_{Penguin}$ *reddy. For the interpretation* I *to be faithful, it is necessary that the conditions* $W_{Bird}(reddy) > W_{Bird}(opus)$ *and* $W_{Penguin}\ (opus) > W_{Penguin}(reddy)$ *hold; which is true. On the contrary, if it were* $Penguin^I\ (reddy) = 0.9$*, the interpretation* I *would not be faithful. For* $Penguin^I(reddy) = 0.8$*, the interpretation* I *would be faithful, but not coherent, as* $W_{Penguin}(opus) > W_{Penguin}\ (reddy)$*, but* $Penguin^I(opus) = Penguin^I(reddy)$*.*

It has been shown [18, 50] that the proposed semantics allows the input-output behavior of a deep network (considered after training) to be captured by a fuzzy multipreference interpretation built over a set of input stimuli, through a simple construction which exploits the activity level of neurons for the stimuli. Each unit h of $\mathcal{N}$ can be associated to a concept name C_h and, for a given domain Δ of input stimuli, the activation value of unit h for a stimulus x is interpreted as the degree of membership of x in concept C_h. The resulting preferential interpretation can be used for verifying properties of the network by model checking.

For MLPs, the deep network itself can be regarded as a conditional knowledge base, by mapping synaptic connections to weighted conditionals, so that the input-output model of the network can be regarded as a coherent-model of the associated conditional knowledge base [18].

4. φ-coherent models

In this section we consider a new notion of coherence of a fuzzy interpretation I wrt a KB, that we call *φ-coherence*, where φ is a function from $\mathbb{R}$ to the interval $[0, 1]$, i.e., $\varphi : \mathbb{R} \rightarrow [0, 1]$. We also establish it relationships with coherent and faithful models.

Definition 5 (φ-coherence). *Let* $K = \langle \mathcal{T}_f, \mathcal{T}_{C_1}, \ldots, \mathcal{T}_{C_k}, \mathcal{A}_f \rangle$ *be a weighted* $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ *knowledge base, and* $\varphi : \mathbb{R} \rightarrow [0, 1]$*. A fuzzy* $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ *interpretation* $I = \langle \Delta, \cdot^I \rangle$ *is* φ-coherent *if, for all concepts* $C_i \in \mathcal{C}$ *and* $x \in \Delta$*,*

$$C_i^I(x) = \varphi(\sum_h w_h^i\ D_{i,h}^I(x)) \tag{6}$$

where $\mathcal{T}_{C_i} = \{(\mathbf{T}(C_i) \sqsubseteq D_{i,h}, w_h^i)\}$ *is the set of weighted conditionals for* C_i*.*

To define *φ-coherent multipreference model* of a knowledge base K, we can replace the *coherence* condition (4) in Definition 4 with the notion of *φ-coherence* of an interpretation I wrt the knowledge base K.

Observe that, for all x such that $C_i(x) > 0$, condition (6) above corresponds to condition $C_i^I(x) = \varphi(W_i(x))$. While the notions of *coherence* and of weight $W_i(x)$ (of an element x wrt a concept C_i) consider, as a special case, the case when $C_i(x) = 0$, in condition (6) we impose the same constraint to all domain elements x (including those with $C_i(x) = 0$).

For Multilayer Perceptrons, let us associate a concept name C_i to each unit i in a deep network $\mathcal{N}$, and let us interpret, as in [18], a synaptic connection between neuron h and neuron i with weight w_{ih} as the conditional $\mathbf{T}(C_i) \sqsubseteq C_j$ with weight $w_h^i = w_{ih}$. If we assume that φ is the *activation function* of *all units* in the network $\mathcal{N}$, then condition (6) characterizes the stationary states of MLPs, where $C_i^I(x)$ corresponds to the activation of neuron i for some input stimulus

x and $\sum_h w_h^i \; D_{i,h}^I(x)$ corresponds to the *induced local field* of neuron i, which is obtained by summing the input signals to the neuron, $x_1, \ldots, x_n$, weighted by the respective synaptic weights: $\sum_{h=1}^n w_{ih} x_h$ [51]. Here, each $D_{i,h}^I(x)$ corresponds to the input signal x_h, for input stimulus x. Of course, φ*-coherence* could be easily extended to deal with different activation functions φ_i, one for each concept C_i (i.e., for each unit i).

Proposition 1. *Let K be a weighted conditional knowledge base and $\varphi : \mathbb{R} \rightarrow [0,1]$. (1) if φ is a* monotonically non-decreasing *function, a φ-coherent fuzzy multipreference model I of K is also an fm-model of K; (2) if φ is a* monotonically increasing *function, a φ-coherent fuzzy multipreference model I of K is also an cf^m-model of K.*

All proofs can be found in the technical report [37]. Item 2 can be regarded as the analog of Proposition 1 in [18, 50], where the fuzzy multi-preferential interpretation $\mathcal{M}_{\mathcal{N}}^{f,\Delta}$ of a deep neural network $\mathcal{N}$, built over the domain of input stimuli Δ, is proven to be a coherent model of the knowledge base $K^{\mathcal{N}}$ associated to $\mathcal{N}$, under the specified conditions on the activation function φ, and the assumption that each stimulus in Δ corresponds to a stationary state in the neural network. Item 1 in Proposition 1 is as well the analog of Proposition 2 in [50] stating that $\mathcal{M}_{\mathcal{N}}^{f,\Delta}$ is a faithful (or weakly-coerent) model of $K^{\mathcal{N}}$.

A notion of *coherent/faithful/φ-coherent multipreference entailment* from a weighted $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ knowledge base K can be defined in the obvious way (see [18, 26] for the definitions of coherent and faithful (fuzzy) multipreference entailment). The properties of faithful entailment have been studied in [26]. Faithful entailment is reasonably well-behaved: it deals with specificity and irrelevance; it is not subject to inheritance blocking; it satisfies most KLM properties [13, 14], depending on their fuzzy reformulation and on the chosen combination functions.

As MLPs are usually represented as a weighted graphs [51], whose nodes are units and whose edges are the synaptic connections between units with their weight, it is very tempting to extend the different semantics of weighted knowledge bases considered above, to weighted argumentation graphs.

5. Coherent, faithful and φ-coherent semantics for weighted argumentation graphs

There is much work in the literature concerning extension of Dung's argumentation framework [4] with weights attached to arguments and/or to the attacks between arguments. Many different proposals have been investigated and compared in the literature. Let us just mention [27, 28, 29, 30, 31, 33] for the moment, which also include extensive comparisons. In the following, we will propose some semantics for weighted argumentation with the purpose of establishing some links with the semantics of conditional knowledge bases considered in the previous section.

In the following, we will consider a notion of *weighted argumentation graph* as a triple $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$, where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ and $\pi : \mathcal{R} \rightarrow \mathbb{R}$. This definition of weighted argumentation graph corresponds to the definition of *weighted argument system* in [29], but here we admit both positive and negative weights, while [29] only allows for positive weights representing the strength of attacks. In our notion of weighted graph, a pair $(A, B) \in \mathcal{R}$

can be regarded as a *support* relation when the weight is positive and an *attack* relation when the weight is negative, and it leads to bipolar argumentation [52]. The argumentation semantics we will consider in the following, as in the case of weighted conditionals, deals with both the positive and the negative weights in a uniform way. For the moment we do not include in G a function determining the *basic strength* of arguments [31].

Given a weighted argumentation graph $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$, we define a *labelling of the graph G* as a function $\sigma : \mathcal{A} \rightarrow [0, 1]$ which assigns to each argument and *acceptability degree*, i.e., a value in the interval $[0, 1]$. Let $R^-(A) = \{B \mid (B, A) \in \mathcal{R}\}$. When $R^-(A) = \emptyset$, argument A has neither supports nor attacks.

For a weighted graph $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$ and a labelling σ, we introduce a *weight* W^G_σ *on* $\mathcal{A}$, as a partial function $W^G_\sigma : \mathcal{A} \rightarrow \mathbb{R}$, assigning a positive or negative support to the arguments $A_i \in \mathcal{A}$ such that $R^-(A_i) \neq \emptyset$, as follows:

$$W^G_\sigma(A_i) = \sum_{(A_j, A_i) \in \mathcal{R}} \pi(A_j, A_i)\, \sigma(A_j) \tag{7}$$

When $R^-(A_i) = \emptyset$, $W^G_\sigma(A_i)$ is let undefined.

We can now exploit this notion of weight of an argument to define different argumentation semantics for a graph G as follows.

Definition 6. *Given a weighted graph $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$ and a labelling σ:*

- *σ is a* coherent labelling of G *if, for all arguments $A, B \in \mathcal{A}$ s.t. $R^-(A) \neq \emptyset$ and $R^-(B) \neq \emptyset$,*
$$\sigma(A) < \sigma(B) \iff W^G_\sigma(A) < W^G_\sigma(B);$$
- *σ is a* faithfull labelling of G *if, for all arguments $A, B \in \mathcal{A}$ s.t. $R^-(A) \neq \emptyset$ and $R^-(B) \neq \emptyset$,*
$$\sigma(A) < \sigma(B) \Rightarrow W^G_\sigma(A) < W^G_\sigma(B);$$
- *for a function $\varphi : \mathbb{R} \rightarrow [0, 1]$, σ is a* φ-coherent labelling of G *if, for all arguments $A \in \mathcal{A}$ s.t. $R^-(A) \neq \emptyset$, $\sigma(A) = \varphi(W^G_\sigma(A))$.*

These definitions do not put any constraint on the labelling of arguments which do not have incoming edges in G: their labelling is arbitrary, provided the constraints on the labelings of all other arguments can be satisfied, depending on the semantics considered.

The definition of φ-coherent labelling of G is defined through a set of equations, as in Gabbay's equational approach to argumentation networks [53]. Here, we use equations for defining the weights of arguments starting from the weights of attacks/supports.

A φ-coherent labelling of a weigthed graph G can be proven to be as well a coherent labelling or a faithful labelling, under some conditions on the function φ.

Proposition 2. *Given a weighted graph $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$: (1) A coherent labelling of G is a faithful labelling of G; (2) if φ is a* monotonically non-decreasing *function, a φ-coherent labelling σ of G is a faithful labelling of G; (3) if φ is a* monotonically increasing *function, a φ-coherent labelling σ of G is a coherent labelling of G.*

The proof is similar to the one of Proposition 1, and can be found in [37]. It exploits the property of a φ-labelling that $\sigma(A) = \varphi(W^G_\sigma(A))$, for all arguments A with $R^-(A) \neq \emptyset$, as well as the properties of φ.

6. φ-coherent labellings and the gradual semantics

The notion of φ-coherent labelling relates to the framework of gradual semantics studied by Amgoud and Doder [33] where, for the sake of simplicity, the weights of arguments and attacks are in the interval $[0, 1]$. Here, as we have seen, positive and negative weights are admitted to represent the strength of attacks and supports. To define an evaluation method for φ-coherent labellings, we need to consider a slightly extended definition of an evaluation method for a graph G in [33]. Following [33] we include a function $\sigma_0 : \mathcal{A} \rightarrow [0, 1]$ in the definition of a weighted graph, where σ_0 assigns to each argument $A \in \mathcal{A}$ its basic strength. Hence a graph G becomes a quadruple $G = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$.

An *evaluation method* for a graph $G = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$ is a triple $M = \langle h, g, f \rangle$, where[1]:

$$\begin{array}{l} h : \mathbb{R} \times [0,1] \rightarrow \mathbb{R} \\ g : \bigcup_{n=0}^{+\infty} \mathbb{R}^n \rightarrow \mathbb{R} \\ f : [0,1] \times Range(g) \rightarrow [0,1] \end{array}$$

Function h is intended to calculate the strength of an attack/support by aggregating the weight on the edge between two arguments with the strength of the attacker/supporter. Function g aggregates the strength of all attacks and supports to a given argument, and function f returns a value for an argument, given the strength of the argument and aggregated weight of its attacks and supports.

As in [33], a gradual semantics S is a function assigning to any graph $G = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$ a weighting Deg^S_G on $\mathcal{A}$, i.e., $Deg^S_G : \mathcal{A} \rightarrow [0, 1]$, where $Deg^S_G(A)$ represents the strength of an argument A (or its acceptability degree).

A *gradual semantics* S is *based on an evaluation method* M iff, $\forall\ G = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$, $\forall A \in \mathcal{A}$,

$$Deg^S_G(A) = f(\sigma_0(A), g(h(\pi(B_1, A), Deg^S_G(B_1)), \ldots, h(\pi(B_n, A), Deg^S_G(B_n))) \quad (8)$$

where $B_1, \ldots, B_n$ are all arguments attacking or supporting A (i.e., $R^-(A) = \{B_1, \ldots, B_n\}$).

Let us consider the evaluation method $M^\varphi = \langle h_{prod}, g_{sum}, f_\varphi \rangle$, where the functions h_{prod} and g_{sum} are defined as in [33], i.e., $h_{prod}(x, y) = x \cdot y$ and $g_{sum}(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$, but we let $g_{sum}()$ to be *undefined*. We let $f_\varphi(x, y) = x$ when y is undefined, and $f_\varphi(x, y) = \varphi(y)$ otherwise. The function f_φ returns a value which is independent from the first argument, when the second argument is not undefined (i.e., there is some support/attack for the argument). When A has neither attacks nor supports ($R^-(A) = \emptyset$), f_φ returns the basic strength of A, $\sigma_0(A)$.

The evaluation method $M^\varphi = \langle h_{prod}, g_{sum}, f_\varphi \rangle$ provides a characterization of the φ-coherent labelling for an argumentation graph, in the following sense.

Proposition 3. *Let $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$ be a weighted argumentation graph. If, for some $\sigma_0 : \mathcal{A} \rightarrow [0, 1]$, S is a gradual semantics of graph $G' = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$ based on the evaluation method $M^\varphi = \langle h_{prod}, g_{sum}, f_\varphi \rangle$, then $Deg^S_{G'}$ is a φ-coherent labelling for G.*

Vice-versa, if σ is a φ-coherent labelling for G, then there are a function σ_0 and a gradual semantics S based on the evaluation method $M^\varphi = \langle h_{prod}, g_{sum}, f_\varphi \rangle$, such that, for the graph $G' = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$, $Deg^S_{G'} \equiv \sigma$.

[1] This definition is the same as in [33], but for the fact that in the domain/range of functions h and g interval $[0, 1]$ is sometimes replaced by $\mathbb{R}$.

The proof can be found in [37].

Amgoud and Doder [33] study a large family of *determinative* and *well-behaved* evaluation models for weighted graphs in which attacks have positive weights in the interval $[0,1]$. For weighted graph G with positive and negative weights, the evaluation method M^{φ} cannot be guaranteed to be determinative, even under the conditions that φ is monotonically increasing and continuous. In general, there is not a unique semantics S based on M^{φ}, and there is not a unique φ-coherent labelling for a weighted graph G, given a basic strength σ_0. This is not surprising, considering that φ-coherent labelings of a graph correspond to stationary states (or equilibrium states) [51] in a deep neural network.

A deep neural network can than be seen as a weighted argumentation graph, with positive and negative weights, where each unit in the network is associated to an argument, and the activation value of the unit can be regarded as the weight (in the interval $[0,1]$) of the corresponding argument. Synaptic positive and negative weights correspond to the strength of supports (when positive) and attacks (when negative). In this view, φ-coherent labelings, assigning to each argument a weight in the interval $[0,1]$, correspond to stationary states of the network, the solutions of a set of equations. This is in agreement with previous results on the relationship between weighted argumentation graphs and MLPs established by Garcez, Gabbay and Lamb [35] and, more recently, by Potyca [36]. We refer to [37] for comparisons.

Unless the network is feedforward (and the corresponding graph is acyclic), stationary states cannot be uniquely determined by an iterative process from the values of input units (that is, from an initial labelling σ_0). On the other hand, a semantics S based on M^{φ} satisfies some of the properties considered in [33], including *anonymity, independence, directionality, equivalence* and *maximality*, provided the last two properties are properly reformulated to deal with both positive and negative weights (i.e., by replacing $R^{-}(x)$ to $Att(x)$, for each argument x in the formulation in [33]). However, a semantics S based on M^{φ} cannot be expected to satisfy the properties of *neutrality, weakening, proportionality* and *resilience*. In fact, function f_{φ} completely disregard the initial valuation σ_0 in graph $G = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$, for those arguments having some incoming edge (even if their weight is 0). So, for instance, it is not the same, for an argument to have a support with weight 0 or no support or attack at all: *neutrality* does not hold.

7. Conclusions

In this paper, drawing inspiration from a fuzzy preferential semantics for weighted conditionals, which has been introduced for modeling the behavior of Multilayer Perceptrons [18], we develop some semantics for weighted argumentation graphs, where positive and negative weights can be associated to pairs of arguments. In particular, we introduce the notions of coherent/faithful/φ-coherent labellings, and establish some relationships among them. While in [18] a deep neural network is mapped to a weighted conditional knowledge base, a deep neural network can as well be seen as a weighted argumentation graph, with positive and negative weights, under the proposed semantics. In this view, φ-coherent labellings correspond to stationary states in the network (where each unit in the network is associated to an argument and the activation value of the unit can be regarded as the weight of the corresponding argument). This is in agreement with previous work on the relationship between argumentation frameworks and neural networks first investigated by Garcez, et al. [35] and recently by Potyca [36]. See in [37] for comparisons.

The proposed approach suggests interesting directions for future work. On the one hand, the generality of the fuzzy conditional logic, where in $\mathbf{T}(C) \sqsubseteq D$, C and D are boolean concepts, suggests a simple approach to deal with attacks/supports by boolean combination of arguments, based on the fuzzy semantics of weighted conditionals [37]. On the other hand, it has been shown in [37] that, under suitable conditions on φ, a multipreference model can be constructed over a (finite) set of φ-labelling Σ. This allows (fuzzy) conditional formulas over arguments to be validated by model checking over a preferential model. For instance, the property: "does normally argument A_2 follows from argument A_1 with a degree greater than 0.7?" can be formalized by the fuzzy inclusion $\mathbf{T}(A_1) \sqsubseteq A_2 > 0.7$. Whether this approach can be extended to the other gradual semantics, and under which conditions on the evaluation method, is subject of future work.

References

[1] J. Leite, J. G. Martins, Social abstract argumentation, in: IJCAI 2011, Barcelona, Spain, July 16-22, 2011, IJCAI/AAAI, 2011, pp. 2287–2292.

[2] L. Amgoud, M. Serrurier, Agents that argue and explain classifications, Auton. Agents Multi Agent Syst. 16 (2008) 187–209.

[3] Q. Zhong, X. Fan, X. Luo, F. Toni, An explainable multi-attribute decision model based on argumentation, Expert Syst. Appl. 117 (2019) 42–61.

[4] P. M. Dung, On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (1995) 321–357.

[5] G. Alfano, S. Greco, F. Parisi, I. Trubitsyna, On the semantics of abstract argumentation frameworks: A logic programming approach, TPLP 20 (2020) 703–718.

[6] L. Giordano, V. Gliozzi, N. Olivetti, G. L. Pozzato, Preferential Description Logics, in: LPAR 2007, volume 4790 of *LNAI*, Springer, Yerevan, Armenia, 2007, pp. 257–272.

[7] K. Britz, J. Heidema, T. Meyer, Semantic preferential subsumption, in: G. Brewka, J. Lang (Eds.), KR 2008, AAAI Press, Sidney, Australia, 2008, pp. 476–484.

[8] G. Casini, U. Straccia, Rational Closure for Defeasible Description Logics, in: T. Janhunen, I. Niemelä (Eds.), JELIA 2010, volume 6341 of *LNCS*, Springer, Helsinki, 2010, pp. 77–90.

[9] G. Casini, T. Meyer, I. J. Varzinczak, , K. Moodley, Nonmonotonic Reasoning in Description Logics: Rational Closure for the ABox, in: 26th International Workshop on Description Logics (DL 2013), volume 1014 of *CEUR Workshop Proceedings*, 2013, pp. 600–615.

[10] L. Giordano, V. Gliozzi, N. Olivetti, G. L. Pozzato, Semantic characterization of rational closure: From propositional logic to description logics, Artif. Intell. 226 (2015) 1–33.

[11] P. A. Bonatti, L. Sauro, On the logical properties of the nonmonotonic description logic DL^{N}, Artif. Intell. 248 (2017) 85–111.

[12] G. Casini, U. Straccia, T. Meyer, A polynomial time subsumption algorithm for nominal safe elo$\perp$ under rational closure, Inf. Sci. 501 (2019) 588–620.

[13] S. Kraus, D. Lehmann, M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, Artificial Intelligence 44 (1990) 167–207.

[14] D. Lehmann, M. Magidor, What does a conditional knowledge base entail?, Artificial Intelligence 55 (1992) 1–60.

[15] J. Pearl, System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning, in: TARK'90, Pacific Grove, CA, USA, 1990, pp. 121–135.

[16] H. Geffner, J. Pearl, Conditional entailment: Bridging two approaches to default reasoning, Artif. Intell. 53 (1992) 209–244.

[17] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, H. Prade, Inconsistency management and prioritized syntax-based entailment, in: Proc. IJCAI'93, Chambéry,, 1993, pp. 640–647.

[18] L. Giordano, D. T. Dupré, Weighted defeasible knowledge bases and a multipreference semantics for a deep neural network model, in: Proc. JELIA 2021, May 17-20, volume 12678 of *LNCS*, Springer, 2021, pp. 225–242.

[19] L. Giordano, D. T. Dupré, An ASP approach for reasoning in a concept-aware multipreferential lightweight DL, Theory Pract. Log. Program. 20 (2020) 751–766.

[20] G. Kern-Isberner, Conditionals in Nonmonotonic Reasoning and Belief Revision - Considering Conditionals as Agents, volume 2087 of *LNCS*, Springer, 2001.

[21] G. Kern-Isberner, C. Eichhorn, Structural inference from conditional knowledge bases, Stud Logica 102 (2014) 751–769.

[22] L. Giordano, D. T. Dupré, A framework for a modular multi-concept lexicographic closure semantics, in: Proc. 18th Int. Workshop on Non-Monotonic Reasoning, NMR2020, September 12th - 14th, 2020.

[23] L. Giordano, V. Gliozzi, D. T. Dupré, On a plausible concept-wise multipreference semantics and its relations with self-organising maps, in: F. Calimeri, S. Perri, E. Zumpano (Eds.), CILC 2020, Rende, Italy, October 13-15, 2020, volume 2710 of *CEUR*, 2020, pp. 127–140.

[24] L. Giordano, V. Gliozzi, D. T. Dupré, A conditional, a fuzzy and a probabilistic interpretation of self-organising maps, CoRR abs/2103.06854 (2021). URL: https://arxiv.org/abs/2103.06854.

[25] T. Kohonen, M. Schroeder, T. Huang (Eds.), Self-Organizing Maps, Third Edition, Springer Series in Information Sciences, Springer, 2001.

[26] L. Giordano, On the KLM properties of a fuzzy DL with Typicality, in: Proc. ECSQARU 2021, volume 12897 of *LNCS*, Springer, 2021, pp. 557–571.

[27] C. Cayrol, M. Lagasquie-Schiex, Graduality in argumentation, J. Artif. Intell. Res. 23 (2005) 245–297. doi:10.1613/jair.1411.

[28] J. Janssen, M. D. Cock, D. Vermeir, Fuzzy argumentation frameworks., in: IPMU 2008, 2008, pp. 513–520.

[29] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, M. J. Wooldridge, Weighted argument systems: Basic definitions, algorithms, and complexity results, Artif. Intell. 175 (2011) 457–486.

[30] S. Egilmez, J. G. Martins, J. Leite, Extending social abstract argumentation with votes on attacks, in: TAFA 2013, Beijing, China, Aug. 3-5, LNCS 8306, Springer, 2013, pp. 16–31.

[31] L. Amgoud, J. Ben-Naim, D. Doder, S. Vesic, Acceptability semantics for weighted argumentation frameworks, in: IJCAI 2017, Melbourne, Australia, 2017, pp. 56–62.

[32] P. Baroni, A. Rago, F. Toni, How many properties do we need for gradual argumentation?, in: Proc. AAAI 2018, New Orleans, Louisiana, USA, February 2-7, 2018, pp. 1736–1743.

[33] L. Amgoud, D. Doder, Gradual semantics accounting for varied-strength attacks, in: Proceedings AAMAS '19, Montreal, QC, Canada, May 13-17, 2019, 2019, pp. 1270–1278.

[34] B. Yun, S. Vesic, Gradual semantics for weighted bipolar setafs, in: Proc. ECSQARU 2021,

Prague, Czech Republic, Sept. 21-24, 2021, volume 12897 of *LNCS*, 2021, pp. 201–214.
[35] A. S. d'Avila Garcez, D. M. Gabbay, L. C. Lamb, Value-based argumentation frameworks as neural-symbolic learning systems, J. Log. Comput. 15 (2005) 1041–1058.
[36] N. Potyka, Interpreting neural networks as quantitative argumentation frameworks, in: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, February 2-9, 2021, AAAI Press, 2021, pp. 6463–6470.
[37] L. Giordano, From weighted conditionals of multilayer perceptrons to gradual argumentation and back, CoRR abs/2110.03643 (2021). URL: https://arxiv.org/abs/2110.03643.
[38] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider, The Description Logic Handbook - Theory, Implementation, and Applications, Cambridge, 2007.
[39] T. Lukasiewicz, U. Straccia, Description logic programs under probabilistic uncertainty and fuzzy vagueness, Int. J. Approx. Reason. 50 (2009) 837–853.
[40] U. Straccia, Towards a fuzzy description logic for the semantic web (preliminary report), in: ESWC 2005, Heraklion, Crete, Greece, May 29 - June 1, 2005, Proceedings, volume 3532 of *LNCS*, Springer, 2005, pp. 167–181.
[41] G. Stoilos, G. B. Stamou, V. Tzouvaras, J. Z. Pan, I. Horrocks, Fuzzy OWL: uncertainty and the semantic web, in: Proceedings of the OWLED*05 Workshop on OWL: Experiences and Directions, Galway, Ireland, Nov 11-12, 2005, volume 188 of *CEUR Workshop Proc.*, 2005.
[42] S. Borgwardt, F. Distel, R. Peñaloza, The limits of decidability in fuzzy description logics with general concept inclusions, Artif. Intell. 218 (2015) 23–55.
[43] F. Bobillo, U. Straccia, Reasoning within fuzzy OWL 2 EL revisited, Fuzzy Sets Syst. 351 (2018) 1–40.
[44] P. Cintula, P. Hájek, C. Noguera (Eds.), Handbook of Mathematical Fuzzy Logic, volume 37-38, College Publications, 2011.
[45] R. Booth, G. Casini, T. Meyer, I. Varzinczak, On rational entailment for propositional typicality logic, Artif. Intell. 277 (2019).
[46] J. Delgrande, C. Rantsoudis, A preference-based approach for representing defaults in first-order logic, in: Proc. 18th Int. Workshop on Non-Monotonic Reasoning, NMR, 2020.
[47] L. Giordano, V. Gliozzi, A reconstruction of multipreference closure, Artif. Intell. 290 (2021).
[48] D. J. Lehmann, Another perspective on default reasoning, Ann. Math. Artif. Intell. 15 (1995) 61–82.
[49] G. Casini, U. Straccia, Towards rational closure for fuzzy logic: The case of propositional gödel logic, in: Proc. LPAR-19, Stellenbosch, South Africa, December 14-19, 2013, volume 8312 of *LNCS*, Springer, 2013, pp. 213–227.
[50] L. Giordano, D. T. Dupré, Weighted defeasible knowledge bases and a multipreference semantics for a deep neural network model, CoRR abs/2012.13421 (2020). URL: https://arxiv.org/abs/2012.13421v2.
[51] S. Haykin, Neural Networks - A Comprehensive Foundation, Pearson, 1999.
[52] L. Amgoud, C. Cayrol, M. Lagasquie-Schiex, On the bipolarity in argumentation frameworks, in: 10th International Workshop on Non-Monotonic Reasoning (NMR 2004), Whistler, Canada, June 6-8, 2004, Proceedings, 2004, pp. 1–9.
[53] D. M. Gabbay, Equational approach to argumentation networks, Argument Comput. 3 (2012) 87–142.