From Weighted Conditionals with Typicality to a Gradual Argumentation Semantics and back*

Laura Giordano[1,*]

[1] *DISIT - Università del Piemonte Orientale, Alessandria, Italy*

Abstract

A fuzzy multipreference semantics has been recently proposed for weighted conditional knowledge bases with typicality, 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 different variants of this semantics, we propose some new gradual argumentation semantics, and relate them to the family of the gradual semantics. The paper also suggests an approach for defeasible reasoning over a weighted argumentation graph, building on the proposed semantics.

Keywords

Defeasible Reasoning, Gradual Argumentation, Fuzzy Description Logics

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 [54] to classification [5], and it is very relevant for decision making and for explanation [61]. The argumentation semantics are strongly related to other non-monotonic reasoning formalisms and semantics [29, 1].

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 [39, 18] and closure constructions, by Casini and Straccia [20, 21] and other researchers [40, 11, 23]. Such defeasible inclusions correspond to Kraus, Lehmann and Magidor (KLM) conditionals $C \mid\sim D$ [51, 52], and defeasible DLs inherit and extend some of the preferential semantics and closure constructions developed within preferential and conditional approaches to commonsense reasoning by Kraus, Lehmann and Magidor [51], Pearl [56], Lehmann [52], Geffner and Pearl [34], Benferhat et al.[9].

In previous work [42], 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 [41], where conditionals are given 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 [53] and Kern-Isberner's c-representations [47, 48], 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 [41]: it satisfies the KLM properties of a preferential consequence relation [51, 52], it allows to deal with specificity and irrelevance and avoids inheritance blocking or the "drowning problem" [56, 9], and deals with "ambiguity preservation" [34]. The plausibility of the concept-wise multipreference semantics has also been supported [38] by showing that it is able to provide a logical interpretation to Kohonen' Self-Organising Maps [49], which are psychologically and biologically plausible neural network models. In the fuzzy case, the KLM properties of non-monotomic entailment have been studied in [36], showing that most KLM postulates are satisfied, depending on their reformulation and on the choice of fuzzy combination functions. It has been shown [42] 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,

NMR 2022: 20th International Workshop on Non-Monotonic Reasoning, August 07–09, 2022, Haifa, Israel

* You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Corresponding author.
laura.giordano@uniupo.it (L. Giordano)
https://people.unipmn.it/laura.giordano/ (L. Giordano)

can be proven 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 just mention, the work by Geffner and Pearl on Conditional Entailment, whose proof theory is defined in terms of "arguments" [34]. In this paper we aim at investigating the relationships between the fuzzy multipreference semantics for weighted conditionals and gradual argumentation semantics [24, 46, 30, 31, 2, 7, 4, 60]. To this purpose, in addition to the notions of coherent [42] and faithful [36] fuzzy multipreference semantics, 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 behavior of MLPs. This is in agreement with the previous results on the relationships between argumentation frameworks and neural networks by Garces, Gabbay and Lamb [27] and by Potyca [57]. Section 7 suggests a possible approach for defeasible reasoning over an argumentation graph, building on the proposed gradual semantics.

A preliminary version of this work has been presented in [35]. For the proofs of the results we refer to https://arxiv.org/abs/2110.03643v2.

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

In this section we recall the syntax and semantics of a description logic and of its fuzzy extension [55]. For sake of simplicity, we only focus on $\mathcal{LC}$, the boolean fragment of $\mathcal{ALC}$ [6], which does not allow for roles. Let N_C be a set of concept names, and N_I a set of individual names. $\mathcal{LC}$ *concepts* (or, simply, concepts) can be defined inductively as follows:

- $A \in N_C$, $\top$ and $\bot$ are concepts;
- if C and D are concepts, then $C \sqcap D$, $C \sqcup D$, $\neg C$ are concepts.

An $\mathcal{LC}$ knowledge base 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 knowledge base $K = (\mathcal{T}_K, \mathcal{A}_K)$, 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 by Straccia [59], Stoilos [58], Lukasiewicz and Straccia [55], Borgwardt et al. [13], Bobillo and Straccia [10], based on the idea that concepts and roles can be interpreted as fuzzy sets. Formulas in Mathematical Fuzzy Logic [26] 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 [55]. We limit our consideration to a few features of a fuzzy DL, without considering datatypes, and restricting to constructs in $\mathcal{LC}$.

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 concept A with a membership degree $A^I(a^I)$ in $[0, 1]$, i.e., A^I is a fuzzy set.

The interpretation function $\cdot^I$ is extended to complex concepts as follows:

$$\top^I(x) = 1, \quad \bot^I(x) = 0,$$
$$(\neg C)^I(x) = \ominus C^I(x),$$
$$(C \sqcap D)^I(x) = C^I(x) \otimes D^I(x),$$
$$(C \sqcup D)^I(x) = C^I(x) \oplus D^I(x).$$

where $x \in \Delta$ and $\otimes$, $\oplus$, $\rhd$ and $\ominus$ are a t-norm, an s-norm, an implication function, and a negation function, chosen among the combination functions of fuzzy logics (we refer to [55] 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:
$(C \sqsubseteq D)^I = inf_{x \in \Delta} C^I(x) \rhd D^I(x)$,
$(C(a))^I = C^I(a^I)$.

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 [10], 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). *A fuzzy interpretation I satisfies a fuzzy $\mathcal{LC}$ axiom E (denoted $I \models E$), as follows:*

- *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 $C^I(a^I)\theta\ n$*

where for $\theta \in \{\geq, \leq, >, <\}$ and $n \in [0, 1]$.

Given a fuzzy $\mathcal{LC}$ knowledge base $K = (\mathcal{T}_f, \mathcal{A}_f)$, 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 [42, 36]. 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 [39] 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 m such that $m\theta n$ holds. In the two-valued case, a typicality inclusion $\mathbf{T}(C) \sqsubseteq D$ stands for a KLM conditional implication $C \mid\!\sim D$ [51, 52], 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, [12] 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:

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

Each $<_C$ has the properties of preference relations in KLM-style ranked interpretations [52], that is, $<_C$ is a modular and well-founded strict partial order, under the assumption that fuzzy interpretations are witnessed (see Section 2) or that Δ is finite. 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 $z <_C y$).

As there are multiple preferences, fuzzy interpretations can be regarded as *multipreferential* interpretations, which have been also studied in the two-valued case by Giordano and Theseider Dupré [41], by Delgrande and Rantsoudis [28], by Giordano and Gliozzi [37], by Casini et al. [19].

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:

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

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. Notice 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 [41], where only the 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 any 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 $\mathcal{LC}^{\mathbf{F}}\mathrm{T}$: some closure constructions

To overcome the weakness of preferential entailment, the rational closure [52] and the lexicographic closure of a conditional knowledge base [53] have been introduced. In this section, we recall a closure construction introduced by Giordano and Theseider Dupré [42] to strengthen $\mathcal{ALC}^{\mathbf{F}}\mathbf{T}$ entailment for weighted conditional knowledge bases, and then we consider some variants. In the two-valued case, the construction is related to the definition of Kern-Isberner's c-representations [47, 48], 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 [22].

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 [42], the typicality operator is assumed to occur only on the l.h.s. of a weighted typicality inclusion, and we call *distinguished concepts* those concepts C_i occurring in the l.h.s. of such inclusions. 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 of weighted $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ knowledge base adapted from [36]:

Example 1. *Consider the weighted knowledge base* $K = \langle \mathcal{T}_f, \mathcal{T}_{Bird}, \mathcal{T}_{Penguin}, \mathcal{A}_f \rangle$*, over the set of distinguished concepts* $\mathcal{C} = \{Bird, Penguin\}$*, with the strict TBox* $\mathcal{T}_f$ *containing the inclusion* $Black \sqcap Red \sqsubseteq \bot \geq 1$*; the weighted TBox* $\mathcal{T}_{Bird}$ *containing the weighted defeasible inclusions:*

(d_1) $\mathbf{T}(Bird) \sqsubseteq Fly$, *+20*
(d_2) $\mathbf{T}(Bird) \sqsubseteq Has\_Wings$, *+50*
(d_3) $\mathbf{T}(Bird) \sqsubseteq Has\_Feather$, *+50;*

$\mathcal{T}_{Penguin}$ *containing the weighted defeasible inclusions:*

(d_4) $\mathbf{T}(Penguin) \sqsubseteq Bird$, *+100*
(d_5) $\mathbf{T}(Penguin) \sqsubseteq Fly$, *- 70*
(d_6) $\mathbf{T}(Penguin) \sqsubseteq Black$, *+50.*

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. Given an ABox in which Reddy is red, has wings, has feather and flies (all with degree 1) and Opus has wings and feather, does not fly (with degree 1), and is black with degree 0.8, considering the weights of defeasible inclusions, we may expect Reddy to be more typical than Opus as a bird, but less typical as a penguin.*

The semantics of a weighted knowledge base is defined in [42] trough a *semantic closure construction*, which allows a subset of the $\mathcal{ALC}^{\mathbf{F}}\mathbf{T}$ interpretations to be selected, namely, the interpretations whose induced preference relations $<_{C_i}$, for the distinguished concepts C_i, *coherently* or *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 for C_i satisfied by x. However, as I is a fuzzy interpretation, 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 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.

In [42] 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). *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 $x, y \in \Delta$,*

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

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 [42], extended version): for all $x, y \in \Delta$,

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

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 [42], extended version, Sec. 7).

Example 2. *Referring to Example 1 above, let us further assume that $Bird^I(reddy) = 1$, $Bird^I(opus) = 0.8$, that $Penguin^I(reddy) = 0.2$ and $Penguin^I(opus) = 0.8$. Clearly, $reddy <_{Bird}$ opus and opus $<_{Penguin}$ $reddy$. The interpretation I to be faithful and coherent, as $W_{Bird}(reddy) > W_{Bird}(opus)$ and $W_{Penguin}$ $(opus) >$ $W_{Penguin}(reddy)$ hold. On the contrary, if we had $Penguin^I$ $(reddy) = 0.9$, 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 [42] 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 (e.g., $\mathbf{T}(Penguin) \sqsubseteq Has\_Wings \geq 0.7$, do typical penguins have wings with degree ≥ 0.7?).

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 [42].

4. Yet another closure construction: 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., $\varphi : \mathbb{R} \to [0, 1]$. 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 $\varphi : \mathbb{R} \to [0, 1]$. 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 $x \in \Delta$,*

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

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 in coherent and faithful models the notion of weight $W_i(x)$ considers, as a special case, the case $C_i(x) = 0$, condition (6) imposes the same constraint to all domain elements x.

To see the relation between this semantics and Multilayer Perceptrons, consider that a neuron k can be described by the following pair of equations: $u_k = \sum_{j=1}^n w_{kj} x_j$, and $y_k = \varphi(u_k + b_k)$, where $x_1, \ldots, x_n$ are the input signals and $w_{k1}, \ldots, w_{kn}$ are the weights of neuron k; b_k is the bias, φ the activation function, and y_k is the output signal of neuron k. By adding a new synapse with input $x_0 = +1$ and synaptic weight $w_{k0} = b_k$, one can write: $u_k = \sum_{j=0}^n w_{kj} x_j$, and $y_k = \varphi(u_k)$, where u_k is called the *induced local field* of the neuron. The neuron can be represented as a directed graph, where the input signals $x_1, \ldots, x_n$ and the output signal y_k of neuron k are nodes of the graph. An edge from x_j to y_k, labelled w_{kj}, means that x_j is an input signal of neuron k with synaptic weight w_{kj}. A neural network can then be seen as "a directed graph consisting of nodes with interconnecting synaptic and activation links" [44].

Let us associate a concept name C_i to each unit i in a deep neural network $\mathcal{N}$ (possibly allowing for feedback), and let us interpret, as in [42], 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 the network, 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, where each $D_{i,h}^I(x)$ represents 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). The following proposition establishes some relationships between φ-coherent, faithful and coherent fuzzy multipreference models of a weighted conditional knowledge base K.

Proposition 1. *Let K be a weighted conditional $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ 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 a faithful-model of K; (2) if φ is a* monotonically increasing *function, a φ-coherent fuzzy multipreference model I of K is also a coherent-model of K.*

Item 2 can be regarded as the analog of Proposition 1 in [42], 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 [42], extended version, 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 [42, 36] for the definitions of coherent and faithful (fuzzy) multipreference entailment). The properties of faithful entailment have been studied in [36]. Faithful entailment is reasonably well-behaved: it deals with specificity and irrelevance; it is not subject to inheritance blocking; it satisfies most KLM properties [51, 52], depending on their fuzzy reformulation and on the chosen combination functions.

As MLPs are usually represented as a weighted graphs [44], 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

There is much work in the literature concerning extension of Dung's argumentation framework [29] 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, for the moment, the work by Cayrol and Lagasquie-Schiex [24], Janssen and Cock [46], Dunne et al. [30], Egilmez et al. [31], Amgoud et al. [2], Amgoud and Doder [4], which also include extensive comparisons. In the following, we propose some semantics for weighted argumentation with the purpose of establishing some links with the semantics of conditional knowledge bases considered in the previous sections.

We 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 [30], but here we admit both positive and negative weights, while [30] 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 [3]. The argumentation semantics we 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 [2].

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 an *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_σ^G on $\mathcal{A}$*, as a partial function $W_\sigma^G : \mathcal{A} \rightarrow \mathbb{R}$, assigning a positive or negative support to the arguments $A_i \in \mathcal{A}$ such that $R^-(A_i) \neq \emptyset$, as follows:

$$W_\sigma^G(A_i) = \sum_{(A_j, A_i) \in \mathcal{R}} \pi(A_j, A_i) \, \sigma(A_j) \qquad (7)$$

When $R^-(A_i) = \emptyset$, $W_\sigma^G(A_i)$ 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 $R^-(B) \neq \emptyset$,*

$$\sigma(A) < \sigma(B) \iff W_\sigma^G(A) < W_\sigma^G(B);$$

- *σ is a* faithfull labelling of G *if, for all arguments $A, B \in \mathcal{A}$ s.t. $R^-(A) \neq \emptyset$ and $R^-(B) \neq \emptyset$,*

$$\sigma(A) < \sigma(B) \Rightarrow W_\sigma^G(A) < W_\sigma^G(B);$$

- *for a function $\varphi : \mathbb{R} \to [0,1]$, σ is a* φ-coherent labelling of G *if, for all arguments $A \in \mathcal{A}$ s.t. $R^-(A) \neq \emptyset$, $\sigma(A) = \varphi(W_\sigma^G(A))$.*

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 [32]. 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. It exploits the property of a φ-labelling that $\sigma(A) = \varphi(W_\sigma^G(A))$, 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 [4] 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 [4]. Following [4] we include a function $\sigma_0 : \mathcal{A} \to [0,1]$ 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]:

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

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

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

$$Deg_G^S(A) = f(\sigma_0(A), g(h(\pi(B_1, A), Deg_G^S(B_1)), \ldots, h(\pi(B_n, A), Deg_G^S(B_n)))$$

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 [4], 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 $\sigma_0 : \mathcal{A} \to [0,1]$, 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 $Deg_{G'}^S$ 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 $G' = \langle \mathcal{A}, \sigma_0, \mathcal{R}, \pi \rangle$, $Deg_{G'}^S \equiv \sigma$.

Amgoud and Doder [4] study a large family of *determinative* and *well-behaved* evaluation models for weighted

[1]This definition is the same as in [4], but for the fact that in the domain/range of functions h and g interval $[0,1]$ is sometimes replaced by $\mathbb{R}$.

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) in a deep neural network [44].

A deep neural network can indeed 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 [27] and, more recently, by Potyca [57]. We refer to the conclusions for some comparisons.

Unless the network is feedforward (and the corresponding graph is acyclic), stationary states cannot be uniquely determined by an iterative process from an initial labelling σ_0. On the other hand, a semantics S based on M^φ satisfies some of the properties considered in [4], 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 [4]). 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 incoming edges (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.

A detailed analysis of the properties of this argumentation semantics is left for an extended version of this work.

7. Back to conditional interpretations

An interesting question is whether, given a set of possible labelings $\Sigma = \{\sigma_1, \sigma_2, \ldots\}$ for a weighted argumentation graph G, where each labelling σ_i assigns to each argument a value in the interval $[0, 1]$ with respect to a given semantics, one can define a preferential structure starting from Σ to evaluate conditional properties of the argumentation graph. This would allow, for instance, to verify properties like: "does normally argument A_2 follows from argument A_1 with a degree greater than 0.7?" This query can be formalized as a fuzzy inclusion $\mathbf{T}(A_1) \sqsubseteq A_2 > 0.7$.

In particular, let Σ is a *finite set of* φ-coherent labelings $\sigma_1, \sigma_2, \ldots$ of a weighted graph $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$, for some function φ. One can define a fuzzy multipreference interpretation over Σ by adopting the construction used in [42] to build a fuzzy multipreference interpretation over the set of input stimuli of a neural network, where each input stimulus was associated to a fit vector [50] describing the activity levels of all units for that input. Here, each labelling σ_i plays the role of a fit vector and each argument A in $\mathcal{A}$ can be interpreted as a concept name of the language. Let $N_C = \mathcal{A}$ and $N_I = \{x_1, x_2, \ldots\}$. We assume that there is one individual name x_j in the language for each labelling $\sigma_j \in \Sigma$, and define a fuzzy multipreference interpretation $I^G_\Sigma = \langle \Sigma, \cdot^I \rangle$ as follows:

- for all $x_j \in N_I$, $x_j^I = \sigma_j$;
- for all $A \in N_C$, $A^I(x_j^I) = \sigma_j(A)$.

The fuzzy $\mathcal{LC}$ interpretation I^G_Σ induces a preference relation $<_{A_i}$ for each argument $A_i \in \mathcal{A}$. For all $\sigma_j, \sigma_k \in \Sigma$:

$$\begin{aligned} \sigma_j <_{A_i} \sigma_k \text{ iff } & A_i^I(x_j^I) > A_i^I(x_k^I) \\ \text{iff } & \sigma_j(A_i) > \sigma_j(A_i). \end{aligned}$$

Let K^G be the conditional knowledge base extracted from the weighted argumentation graph, as follows:

$$K^G = \{(\mathbf{T}(A_i) \sqsubseteq A_j, w_{j,i}) \mid (A_j, A_i) \in \mathcal{R} \text{ and } w((A_j, A_i)) = w_{ji}\}$$

It can be proven that:

Proposition 4. *Let Σ be a finite set of φ-coherent labelings of a weighted graph $G = \langle \mathcal{A}, \mathcal{R}, \pi \rangle$, for some function $\varphi : \mathbb{R} \rightarrow [0, 1]$. The following statements hold:*

(i) If φ is a monotonically increasing function and $\varphi : \mathbb{R} \rightarrow (0, 1]$, then I^Σ is a coherent (fuzzy) multipreference model of K^G.

(ii) If φ is a monotonically non-decreasing function, then I^Σ is a faithful (fuzzy) multipreference model of K^G.

The proof of item (i) is similar to the proof of Proposition 1 in [42] (extended version with proofs). The proof of item (ii) is similar to the proof of Proposition 2 therein. The restriction to a *finite set* Σ of φ-coherent labelings is needed to guarantee the well-foundedness of the resulting interpretation. In fact, in general, the set of all φ-coherent labelings of G might be infinite and, if Σ is the set of all φ-coherent labelings of G, there is no guarantee that the resulting fuzzy $\mathcal{LC}^{\mathbf{F}}\mathbf{T}$ interpretation is witnessed and the preference relations $<_{A_i}$ is well-founded.

While this allows (fuzzy) conditional formulas over arguments to be validated by model checking over a preferential model, whether this approach can be extended to the other gradual semantics, and under which conditions on the evaluation method, is subject of future work.

Observe also that, in the conditional semantics in Sections 3.1 and 4, in a typicality inclusion $\mathbf{T}(C) \sqsubseteq D$, concepts C and D are not required to be concept names, but they can be complex concepts. In particular, in the fragment $\mathcal{LC}$ of $\mathcal{ALC}$ considered in this paper, D can be any boolean combination of concept names. The correspondence between weighted conditionals $\mathbf{T}(A_i) \sqsubseteq A_j$ in K^G and weighted attacks/supports in the argumentation graph G, suggests a possible generalization of the structure of the weighted argumentation graph by allowing attacks/supports by a boolean combination of arguments. The labelling of arguments in the set $[0, 1]$ can indeed be extended to boolean combinations of arguments using the fuzzy combination functions, as for boolean concepts in the conditional semantics (e.g., by letting $\sigma(A_1 \wedge A_2) = min\{\sigma(A_1), \sigma(A_2)\}$, using the minimum t-norm as in Zadeh fuzzy logic). This also relates to the work considering "sets of attacking (resp. supporting) arguments"; i.e., several argument together attacking (or supporting) an argument. Indeed, for gradual semantics, the sets of attacking arguments framework (SETAF) has been studied by Yun and Vesic, by considering "the force of the set of attacking (resp. supporting) arguments to be the force of the weakest argument in the set" [60]. This would correspond to interpret the set of arguments as a conjunction, using minimum t-norm.

8. 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 [42], 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 labelings of a graph, and establish some relationships among them. While in [42] 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 labelings 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. [27] and recently by Potyca [57].

The work by Garcez, et al. [27] combines value-based argumentation frameworks [8] and neural-symbolic learning systems by providing a translation from argumentation networks to neural networks with 3 layers (input, output layer and one hidden layer). This enables the accrual of arguments through learning as well as the parallel computation of arguments. The work by Potyca [57] considers a quantitative bipolar argumentation frameworks (QBAFs) similar to [7] and exploits an *influence function* based on the logistic function to define an MLP-based semantics σ_{MLP} for a QBAF: for each argument $a \in \mathcal{A}$, $\sigma_{MLP}(a) = \lim_{k\to\infty} s_a^{(k)}$, when the limit exists, and is undefined otherwise; where $s_a^{(k)}$ is a value in the interval $[0, 1]$, and k represents the iteration. The paper studies convergence conditions both in the discrete and in the continuous case, as well as the semantic properties of MLP-based semantics, and proves that all properties for the QBAF semantics proposed in [2] are satisfied. As we have seen in Section 6, our semantics based on φ-coherent models fails to satisfy some of the properties in [2].

In this work we have investigated the relationships between φ-coherent labelings and the gradual semantics by Amgoud and Doder [4], by slightly extending their definitions to deal with positive and negative weights to capture the strength of supports and of attacks. A correspondence between the gradual semantics based on a specific evaluation method M^{φ} and φ-coherent labelings has been established. Differently from the Fuzzy Argumentation Frameworks by Jenssen et al. [46], where an attack relation is a fuzzy binary relation over the set of arguments, here we have considered real-valued weights associated to pairs of arguments. Our semantics also relates to the fuzzy extension of rational closure by Casini and Straccia [22].

The paper discusses an approach for defeasible reasoning over a weighted argumentation graph building on φ-coherent labelings. This allows a multipreference model to be constructed over a (finite) set of φ-labelling Σ and allows (fuzzy) conditional formulas over arguments to be validated over Σ by model checking over a preferential model. Whether this approach can be extended to the other gradual semantics, and under which conditions on the evaluation method, requires further investigation for future work. The paper also suggests an approach to deal with attack/supports by a boolean combination of arguments, by exploiting the fuzzy semantics of weighted conditionals.

The correspondence between Abstract Dialectical Frameworks [17] and Nonmonotonic Conditional Logics has been studied by Heyninck, Kern-Isberner and Thimm [45], with respect to the two-valued models, the stable, the preferred semantics and the grounded semantics of

ADFs. Whether the coherent/faithfull/φ-coherent semantics developed in the paper for weighted argumentation can be reformulated for a (weighted) Abstract Dialectical Frameworks, and which are the relationships with the work in [45], also requires investigation for future work.

Undecidability results for fuzzy description logics with general inclusion axioms (e.g., by Cerami and Straccia [25] and by Borgwardt and Peñaloza [14]) motivate restricting the logics to finitely valued semantics [15], and the investigation of decidable approximations of fuzzy multipreference entailment, under the different semantics. An ASP approach for reasoning under finitely multivalued fuzzy semantics for weighterd conditional knowledge bases has been proposed in [43], by exploiting ASP [33] and *asprin* [16] for defeasible reasoning with the concept-wise multipreference entailment under a φ-coherent semantics. through the computation of preferred answer sets. As a proof of concept, this approach has been experimented for checking properties of some trained Multilayer Perceptrons. A similar investigation is also of interest for the semantics of weighted argumentation graphs introduced in this paper, to study its extensions to the finitely many-valued case.

Acknowledgments

Thanks to the anonymous referees for their helpful comments and suggestions. This research is partially supported by INDAM-GNCS Project 2020. It was developed in the context of the European Cooperation in Science & Technology (COST) Action CA17124 Dig4ASP.

References

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

[2] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Acceptability semantics for weighted argumentation frameworks. In *IJCAI 2017, Melbourne, Australia*, pages 56–62, 2017.

[3] L. Amgoud, C. Cayrol, and 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*, pages 1–9, 2004.

[4] L. Amgoud and D. Doder. Gradual semantics accounting for varied-strength attacks. In *Proceedings AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*, pages 1270–1278, 2019.

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

[6] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider. *The Description Logic Handbook - Theory, Implementation, and Applications*. Cambridge, 2007.

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

[8] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. Log. Comput.*, 13(3):429–448, 2003.

[9] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, and H. Prade. Inconsistency management and prioritized syntax-based entailment. In *Proc. IJCAI'93, Chambéry,*, pages 640–647, 1993.

[10] F. Bobillo and U. Straccia. Reasoning within fuzzy OWL 2 EL revisited. *Fuzzy Sets Syst.*, 351:1–40, 2018.

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

[12] Richard Booth, Giovanni Casini, Thomas Meyer, and Ivan Varzinczak. On rational entailment for propositional typicality logic. *Artif. Intell.*, 277, 2019.

[13] S. Borgwardt, F. Distel, and R. Peñaloza. The limits of decidability in fuzzy description logics with general concept inclusions. *Artif. Intell.*, 218:23–55, 2015.

[14] S. Borgwardt and R. Peñaloza. Undecidability of fuzzy description logics. In Gerhard Brewka, Thomas Eiter, and Sheila A. McIlraith, editors, *Proc. KR 2012, Rome, Italy, June 10-14, 2012*. AAAI Press, 2012.

[15] S. Borgwardt and R. Peñaloza. The complexity of lattice-based fuzzy description logics. *J. Data Semant.*, 2(1):1–19, 2013.

[16] G. Brewka, J. P. Delgrande, J. Romero, and T. Schaub. asprin: Customizing answer set preferences without a headache. In *Proc. AAAI 2015*, pages 1467–1474, 2015.

[17] G Brewka, H. Strass, S. Ellmauthaler, J. P. Wallner, and S. Woltran. Abstract dialectical frameworks revisited. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pages 803–809. IJCAI/AAAI, 2013.

[18] K. Britz, J. Heidema, and T. Meyer. Semantic preferential subsumption. In G. Brewka and J. Lang, editors, *KR 2008*, pages 476–484, Sidney, Australia, September 2008. AAAI Press.

[19] G. Casini, T. A. Meyer, and I. Varzinczak. Contextual conditional reasoning. In *AAAI-21, Virtual Event, February 2-9, 2021*, pages 6254–6261. AAAI

Press, 2021.

[20] G. Casini and U. Straccia. Rational Closure for Defeasible Description Logics. In T. Janhunen and I. Niemelä, editors, *JELIA 2010*, volume 6341 of *LNCS*, pages 77–90, Helsinki, Sept. 2010. Springer.

[21] G. Casini and U. Straccia. Defeasible inheritance-based description logics. *Journal of Artificial Intelligence Research (JAIR)*, 48:415–473, 2013.

[22] G. Casini and 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*, pages 213–227. Springer, 2013.

[23] G. Casini, U. Straccia, and T. Meyer. A polynomial time subsumption algorithm for nominal safe elo$\bot$ under rational closure. *Inf. Sci.*, 501:588–620, 2019.

[24] C. Cayrol and M. Lagasquie-Schiex. Graduality in argumentation. *J. Artif. Intell. Res.*, 23:245–297, 2005.

[25] M. Cerami and U. Straccia. On the undecidability of fuzzy description logics with gcis with lukasiewicz t-norm. *CoRR*, abs/1107.4212, 2011.

[26] P. Cintula, P. Hájek, and C. Noguera, editors. *Handbook of Mathematical Fuzzy Logic*, volume 37-38. College Publications, 2011.

[27] A. S. d'Avila Garcez, D. M. Gabbay, and L. C. Lamb. Value-based argumentation frameworks as neural-symbolic learning systems. *J. Log. Comput.*, 15(6):1041–1058, 2005.

[28] J. Delgrande and C. Rantsoudis. A preference-based approach for representing defaults in first-order logic. In *Proc. 18th Int. Workshop on Non-Monotonic Reasoning, NMR*, 2020.

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

[30] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. J. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artif. Intell.*, 175(2):457–486, 2011.

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

[32] D. M. Gabbay. Equational approach to argumentation networks. *Argument Comput.*, 3(2-3):87–142, 2012.

[33] M. Gebser, R. Kaminski, B. Kaufmann, and T. Schaub. *Answer Set Solving in Practice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2012.

[34] Hector Geffner and Judea Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artif. Intell.*, 53(2-3):209–244, 1992.

[35] L. Giordano. From weighted conditionals of multilayer perceptrons to a gradual argumentation semantics. In *Proc. 5th Workshop on Advances in Argumentation in Artificial Intelligence 2021, Milan, Italy, Nov. 29*, volume 3086 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.

[36] L. Giordano. On the KLM properties of a fuzzy DL with Typicality. In *Proc. ECSQARU 2021, Prague, Sept. 21-24, 2021*, volume 12897 of *LNCS*, pages 557–571. Springer, 2021.

[37] L. Giordano and V. Gliozzi. A reconstruction of multipreference closure. *Artif. Intell.*, 290, 2021.

[38] L. Giordano, V. Gliozzi, and D. Theseider Dupré. A conditional, a fuzzy and a probabilistic interpretation of self-organizing maps. *J. Log. Comput.*, 32(2):178–205, 2022.

[39] L. Giordano, V. Gliozzi, N. Olivetti, and G. L. Pozzato. Preferential Description Logics. In *LPAR 2007*, volume 4790 of *LNAI*, pages 257–272, Yerevan, Armenia, October 2007. Springer.

[40] L. Giordano, V. Gliozzi, N. Olivetti, and G. L. Pozzato. Semantic characterization of rational closure: From propositional logic to description logics. *Art. Int.*, 226:1–33, 2015.

[41] L. Giordano and D. Theseider Dupré. An ASP approach for reasoning in a concept-aware multipreferential lightweight DL. *Theory Pract. Log. Program.*, 20(5):751–766, 2020.

[42] L. Giordano and D. Theseider 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*, pages 225–242. Springer, 2021.

[43] L. Giordano and D. Theseider Dupré. An ASP approach for reasoning on neural networks under a finitely many-valued semantics for weighted conditional knowledge bases. 2022. To appear in TPLP.

[44] S. Haykin. *Neural Networks - A Comprehensive Foundation*. Pearson, 1999.

[45] J. Heyninck, G. Kern-Isberner, and M. Thimm. On the correspondence between abstract dialectical frameworks and nonmonotonic conditional logics. In *Proceedings of the Thirty-Third International Florida Artificial Intelligence Research Society Conference, May 17-20, 2020*, pages 575–580. AAAI Press, 2020.

[46] J. Janssen, M. De Cock, and D. Vermeir. Fuzzy argumentation frameworks. In *IPMU 2008*, pages 513–520, 2008.

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

[48] G. Kern-Isberner and C. Eichhorn. Structural inference from conditional knowledge bases. *Stud*

Logica, 102(4):751–769, 2014.
[49] T. Kohonen, M.R. Schroeder, and T.S. Huang, editors. *Self-Organizing Maps, Third Edition*. Springer Series in Information Sciences. Springer, 2001.
[50] Bart Kosko. *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence*. Prentice Hall, 1992.
[51] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207, 1990.
[52] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.
[53] D. J. Lehmann. Another perspective on default reasoning. *Ann. Math. Artif. Intell.*, 15(1):61–82, 1995.
[54] J. Leite and J. G. Martins. Social abstract argumentation. In *IJCAI 2011, Barcelona, Spain, July 16-22, 2011*, pages 2287–2292. IJCAI/AAAI, 2011.
[55] T. Lukasiewicz and U. Straccia. Description logic programs under probabilistic uncertainty and fuzzy vagueness. *Int. J. Approx. Reason.*, 50(6):837–853, 2009.
[56] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *TARK'90, Pacific Grove, CA, USA*, pages 121–135, 1990.
[57] N. Potyka. Interpreting neural networks as quantitative argumentation frameworks. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, February 2-9, 2021*, pages 6463–6470. AAAI Press, 2021.
[58] G. Stoilos, G. B. Stamou, V. Tzouvaras, J. Z. Pan, and I. Horrocks. Fuzzy OWL: uncertainty and the semantic web. In *OWLED*05 Workshop on OWL Galway, Ireland, Nov 11-12, 2005*, volume 188 of *CEUR Workshop Proc.*, 2005.
[59] U. Straccia. Towards a fuzzy description logic for the semantic web (preliminary report). In *ESWC 2005, Heraklion, Crete, May 29 - June 1, 2005*, volume 3532 of *LNCS*, pages 167–181. Springer, 2005.
[60] B. Yun and S. Vesic. Gradual semantics for weighted bipolar setafs. In *Proc. ECSQARU 2021, Prague, Sept. 21-24, 2021*, volume 12897 of *LNCS*, pages 201–214, 2021.
[61] Q. Zhong, X. Fan, X. Luo, and F. Toni. An explainable multi-attribute decision model based on argumentation. *Expert Syst. Appl.*, 117:42–61, 2019.