A Multipreference Semantics from Common Sense Reasoning to Neural Network Models: an Overview

Laura Giordano[1], Valentina Gliozzi[2] and Daniele Theseider Dupré[1]

[1]DISIT - Università del Piemonte Orientale, Viale Michel 11, I-15121, Alessandria, Italy
[2]Dipartimento di Informatica, Università degli Studi di Torino, Corso Svizzera 185, I-10149,Torino, Italy

Abstract

In this short paper we report about a "concept-wise" multipreference semantics for weighted conditionals and its use to provide a logical interpretation to some neural network models, Self-Organising Maps (SOMs) and Multilayer Perceptrons (MLPs). For MLPs, a deep network can be regarded as a conditional knowledge base, in which the synaptic connections correspond to weighted conditionals.

Keywords

Common Sense Reasoning, Preferential semantics, Typicality in Description Logics, Neural Network models

1. Introduction

Preferential approaches to common sense reasoning [1, 2, 3, 4, 5, 6, 7] have their roots in conditional logics [8, 9], and have been recently extended to Description Logics (DLs), to deal with inheritance with exceptions in ontologies, by allowing non-strict form of inclusions, called *defeasible* or *typicality* inclusions.

Different preferential semantics [10, 11] and closure constructions [12, 13, 14, 15, 16, 17, 18] have been proposed for such defeasible DLs and, in this paper, we report about a concept-wise multipreference semantics [19], which has been recently introduced as a semantics of ranked knowledge bases in a lightweight DL to account for preferences with respect to different concepts, and has been proposed as a semantics for some neural network models.

We have considered both an unsupervised model, Self-organising maps (SOMs)[20], which is considered as a psychologically and biologically plausible neural network model, and a supervised one, Multilayer Perceptrons (MLPs) [21]. Learning algorithms in the two cases are quite different but our aim is to capture, through a semantic interpretation, the behavior of the network resulting after training and not to deal with the learning process. We will see that this can be accomplished in both cases in a similar way, based on the multi-preferential semantics.

In both cases, considering the domain of all input stimuli presented to the network during training (or in the generalization phase), one can build a semantic interpretation describing the *input-output behavior* of the network as a multi-preference interpretation, where preferences are

CAOS 2021: 5th Workshop on Cognition And OntologieS, held at JOWO 2021: Episode VII The Bolzano Summer of Knowledge, September 11-18, 2021, Bolzano, Italy

laura.giordano@uniupo.it (L. Giordano); valentina.gliozzi@unito.it (V. Gliozzi); dtd@uniupo.it (D. Theseider Dupré)

associated to concepts. For SOMs, the learned categories $C_1, \dots, C_n$ are regarded as concepts so that a preference relation (over the domain of input stimuli) is associated to each category [22, 23]. For MLPs, each neuron in the deep network (including hidden neurons) can be associated with a concept and with a preference relation on the domain [24].

The idea is that, given two input stimuli x and y, and two categories/concepts, e.g., $Horse$ and $Zebra$, the neural model can assign to x a degree of membership in the category $Horse$ which is higher than the degree of membership of y, so that x can be regarded as a being more typical than y as a horse ($x <_{Horse} y$), while vice-versa x can be regarded as a being less typical than y as a zebra ($y <_{Zebra} x$). A preferential interpretation can be built over the domain of input stimuli and can be used for checking properties such as: are the instances of category C_1 also instances of category C_2? Are typical instances of category C_1 also instances of category C_2? This verification can be done by *model-checking* given a multipreference interpretation describing the input-output behavior of the network [23].

For MLPs, the relationship between our logic of commonsense reasoning and deep neural networks is even stronger, as a deep neural network can itself be regarded as a conditional knowledge base, i.e., as a set weighted conditionals. This has been achieved by developing a concept-wise fuzzy multipreference semantics for a DLs with weighted defeasible inclusions.

The strong relationship between neural networks and conditional logics of commonsense reasoning raises several issues from the standpoint of knowledge representation, from the standpoint of neuro-symbolic integration, as well as from the standpoint of explainable AI [25, 26, 27]. We will hint at some of these issues in the extended abstract after shortly describing the approach.

2. The Concept-Wise Multipreference Semantics

The concept-wise multipreference semantics (cw^m-semantics) has been introduced as a semantics for ranked $\mathcal{EL}$ knowledge bases [19], and later extended to weighted knowledge bases [24]. In both cases the knowledge base contains (besides standard inclusions, called *strict*) defeasible or typicality inclusions of the form $\mathbf{T}(C) \sqsubseteq D$ (meaning “the typical Cs are Ds" or “normally Cs are Ds") with a rank (resp. a weight). They correspond to KLM conditionals $C \mathrel{|\!\sim} D$ [4]. Ranks (weights) of defeasible inclusions represent their strength (plausibility/implausibility). The preferential semantics of ranked and weighted knowledge bases are defined in terms of concept-wise multipreference interpretations, based on different constructions.

Concept-wise multipreference interpretations (cw^m-interpretations) are defined by adding to standard DL interpretations, which are pairs $\langle \Delta, \cdot^I \rangle$, where Δ is a domain, and $\cdot^I$ an interpretation function, the preference relations $<_{C_1}, \dots, <_{C_n}$ associated with a set of distinguished concepts $C_1, \dots, C_n$, representing the relative typicality of domain individuals with respect to these concepts. Each preference relation $<_{C_i}$ is a modular and well-founded strict partial order on Δ. Preferences with respect to different concepts do not need to agree; as we have seen, a domain element x may be more typical than y as a horse, but less typical as a zebra. A global preference relation $<$ can be defined starting from the $<_{C_i}$'s, and concept $\mathbf{T}(C)$ is interpreted as the set of all $<$-minimal C elements. A simple notion of global preference $<$ exploits Pareto combination of the preference relations $<_{C_i}$, but a more sophisticated notion of preference combination has been considered in [19], by exploiting a modified Pareto condition which takes into account the specificity relation among concepts (e.g., that concept $Penguin$ is more specific than concept $Bird$). It has been proven [19] that global preference in a cw^m-interpretation determines a KLM-style preferential interpretation, and cw^m-entailment satisfies the KLM postulates of a preferential consequence relation [4].

3. A Preferential Interpretation of Self-Organising Maps

Once the SOM has learned to categorize, one can look at the result of the categorization as a concept-wise multipreference interpretation over a domain of input stimuli, in which a preference relation is associated to each concept (each learned category), and the combination of the preferences into a global one (following the approach described above) defines a KLM-style preferential model of the SOM. More precisely, once the SOM has learned to categorize, to assess category generalization, Gliozzi and Plunkett [28] define the map's disposition to consider a new stimulus y as a member of a known category C as a function of the *distance* of y from the *map's representation* of C. The relative distance $rd(x, C_i)$ of a stimulus x from a category C_i can be used to build a binary preference relation $<_{C_i}$ among the stimuli in Δ with respect to category C_i [22, 29], by letting $x <_{C_i} y$ if and only if $rd(x, C_i) > rd(y, C_i)$ (x is more typical than y with respect to category C_i if its relative distance from category C_i is lower than the relative distance of y).

This preferential model can be exploited to learn or validate conditional knowledge from empirical data, by verifying conditional formulas over the preferential interpretation constructed from the SOM. Both a two-valued and a fuzzy semantics have been considered [23]. In both cases, model checking can be used for the verification of inclusions (either defeasible inclusions or fuzzy inclusion axioms) over the respective models of the SOM (for instance, do the most typical penguins belong to the category Bird with at least a degree of membership 0.8?). Starting from the fuzzy interpretation of the SOM, a probabilistic account can also be given based on Zadeh's probability of fuzzy events [30].

4. A Preferential Interpretation of Multilayer Perceptrons

For MLPs, a deep network is considered after the training phase, when the synaptic weights have been learned. The input-output behaviour of the network can be captured in a similar way as for SOMs by constructing a preferential interpretation over the domain Δ of the input stimuli considered during training (or generalization) [24]. Each neuron k of interest can be associated to a concept C_k and, for each distinguished concept C_j, a preference relation $<_{C_j}$ is defined over the domain Δ based on the activity values, $y_j(v)$, of neuron j for each input $v \in \Delta$. In a similar way, a fuzzy interpretation of the network can be constructed over the domain Δ, as well as a fuzzy-multipreference semantics.

All the three semantics allow the input-output behavior of the network to be captured by interpretations built over a set of input stimuli through simple constructions, which exploits the activity level of neurons for the stimuli. In particular, for the fuzzy-multipreference interpretations, the idea [24] is to extend a fuzzy DL interpretation with a set of induced preferences.

In a fuzzy DL interpretation I, the interpretation of a concept C_h is a mapping $C_i^I : \Delta \rightarrow [0, 1]$, associating to each $x \in \Delta$ the degree of membership of x in C_h. The activation value of unit h for a stimulus x in the network (assumed to be in the interval $[0, 1]$) is taken as the degree of membership of x in concept C_h. The fuzzy interpretation also induces an ordering $<_{C_h}$ on the domain Δ, for each C_h, to be regarded as the preference relation associated to concept C_h. This allows a notion of typicality to be defined in a fuzzy interpretation. Let us call $\mathcal{M}_{\mathcal{N}}^{f,\Delta}$ the fuzzy multipreference interpretation built from the network $\mathcal{N}$ over a domain Δ of input stimuli.

As for SOMs, logical properties of the neural network (both typicality properties and fuzzy axioms) can then be verified by model checking over such an interpretation. Evaluating properties involving hidden units might be of interest, although their meaning is usually unknown. In the well known Hinton's family example [31], one may want to verify whether, normally, given an old Person 1 and relationship Husband, Person 2 would also be old, i.e., whether $\mathbf{T}(Old_1 \sqcap Husband) \sqsubseteq Old_2$ is satisfied. Here, concept Old_1 (resp., Old_2) is associated to a (known, in this case) hidden unit for Person 1 (and Person 2), while Husband is associated to an input unit. If the properties of interest involve some specific units, only the concepts associated to those units may be considered in the language to build the interpretation.

All the three kinds of interpretations considered above for MLPs describe the input-output behavior of the network. However, the fuzzy multipreference interpretation $\mathcal{M}_{\mathcal{N}}^{f,\Delta}$ described above can be also proven to be a model of the neural network $\mathcal{N}$ in a logical sense, by mapping the multilayer network into a weighted conditional knowledge base.

4.1. Weighted $\mathcal{ALC}$ Knowledge Bases

In this section, we shortly recall the definition of weighted conditional knowledge bases through an example, and give some hints about the two-valued and fuzzy multipreference semantics, referring to [24] for a detailed description for $\mathcal{EL}$.

A weighted $\mathcal{ALC}$ knowledge base K over a set $\mathcal{C} = \{C_1, \ldots, C_k\}$ of distinguished $\mathcal{ALC}$ concepts is a tuple $\langle \mathcal{T}, \mathcal{T}_{C_1}, \ldots, \mathcal{T}_{C_k}, \mathcal{A} \rangle$, where the TBOX $\mathcal{T}$ is a set of $\mathcal{ALC}$ inclusion axiom, the ABox $\mathcal{A}$ is a set of $\mathcal{ALC}$ assertions and, for each distinguished concept $C_i \in \mathcal{C}$, $\mathcal{T}_{C_i}$ is a set of weighted typicality inclusions of the form $\mathbf{T}(C_i) \sqsubseteq D$, with a positive or negative weight (a real number). In the fuzzy case, $\mathcal{T}$ and $\mathcal{A}$ contain fuzzy axioms.

Consider the weighted knowledge base $K = \langle \mathcal{T}, \mathcal{T}_{Bird}, \mathcal{T}_{Penguin}, \mathcal{A} \rangle$, over the set of distinguished concepts $\mathcal{C} = \{Bird, Penguin\}$, with empty ABox and with $\mathcal{T}$ containing the inclusions $Penguin \sqsubseteq Bird$ and $Black \sqcap Grey \sqsubseteq \bot$.

The weighted TBox $\mathcal{T}_{Bird}$ contains the following weighted defeasible inclusions:
(d_1) $\mathbf{T}(Bird) \sqsubseteq Fly$, +20
(d_2) $\mathbf{T}(Bird) \sqsubseteq \exists has\_Wings.\top$, +50
(d_3) $\mathbf{T}(Bird) \sqsubseteq \exists has\_Feathers.\top$, +50;
$\mathcal{T}_{Penguin}$ contains the defeasible inclusions:
(d_4) $\mathbf{T}(Penguin) \sqsubseteq Fly$, - 70
(d_5) $\mathbf{T}(Penguin) \sqsubseteq Black$, +50;
(d_6) $\mathbf{T}(Penguin) \sqsubseteq Grey$, +10;

The meaning is that a bird normally has wings, has feathers and flies, but having wings and feathers (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_4 has a negative weight -70), while being black or being grey are plausible properties of prototypical penguins, in fact, d_5 and d_6 have positive weights, resp. 50 and 10, so that being black is more plausible than being grey.

A two-valued semantics for weighted $\mathcal{ALC}$ knowledge bases can be defined by developing a semantic closure construction in the same spirit as Lehmann's lexicographic closure [32], but more similar to Kern-Isberner's semantics of c-representations [7, 33], in which the world ranks are generated as a sum of impacts of falsified conditionals. Here, the (positive or negative) weights of the satisfied defaults are summed, but in a concept-wise manner, so to determine the plausibility of a domain elements with respect to certain concepts. In this way, the modular structure of the knowledge base can be considered. More precisely, for a domain element x in Δ, and a distinguished concept C_i, the weight $W_i(x)$ of x wrt C_i is defined as the sum of the weights w_h^i of the typicality inclusions $\mathbf{T}(C_i) \sqsubseteq D_{i,h}$ in $\mathcal{T}_{C_i}$ verified by x (and is $-\infty$ when x is not an instance of C_i). From the weights $W_i(x)$ the *preference relation* $\leq_{C_i}$ can be defined by letting: for $x, y \in \Delta$, $x \leq_{C_i} y$ iff $W_i(x) \geq W_i(y)$. The higher the weight of x wrt C_i the higher its typicality relative to C_i. This closure construction defines preferences $<_{C_i}$ (strict modular partial orders) and allows for the definition of *concept-wise multipreference interpretations* as in Section 2.

In the fuzzy case, the fuzzy logic combination functions are used for complex concepts to compute the $W_i(x)$'s and to determine the associated preference relations. To guarantee that the preferences determined from the knowledge base are coherent with the fuzzy interpretation of concepts, a notions of *coherent (fuzzy) multipreference interpretation* (cf^m-interpretation) is also introduced [24].

4.2. MLPs as Conditional Knowledge Bases

Let us describe how the multilayer network $\mathcal{N}$ can be mapped to a weighted conditional knowledge base $K^{\mathcal{N}}$, i.e., to a set of weighted typicality inclusions. The idea is to consider, for each unit k, all the units $j_1, \ldots, j_m$, whose output signals are the input signals of unit k, with synaptic weights $w_{k,j_1}, \ldots, w_{k,j_m}$. Let C_k be the concept name associated to unit k and $C_{j_1}, \ldots, C_{j_m}$ be the concept names associated to units $j_1, \ldots, j_m$. One can define, for unit k, a set $\mathcal{T}_{C_k}$ of m typicality inclusions, with their associated weights, as follows: $\mathbf{T}(C_k) \sqsubseteq C_{j_1}$ with w_{k,j_1}, ... , $\mathbf{T}(C_k) \sqsubseteq C_{j_m}$ with w_{k,j_m}. The network $\mathcal{N}$ can than be mapped to a conditional knowledge base $K^{\mathcal{N}}$ containing, for each neuron k, a set of typicality inclusions $\mathcal{T}_{C_k}$ as defined above.

Let us consider the fuzzy multipreference interpretation $\mathcal{M}_{\mathcal{N}}^{f,\Delta}$ built from $\mathcal{N}$ over a domain Δ of input stimuli, as described above. Let us further assume that, in the construction, all units are considered and a concept C_k is introduced in the language for each unit k. It has been proven [24] that the interpretation $\mathcal{M}_{\mathcal{N}}^{f,\Delta}$ is a cf^m-model of the knowledge base $K^{\mathcal{N}}$, under some condition on the activation functions in $\mathcal{N}$. In particular, the properties that are entailed from $K^{\mathcal{N}}$ are properties satisfied by $\mathcal{M}_{\mathcal{N}}^{f,\Delta}$, for any choice of the input stimuli in the domain Δ.

5. Discussion and Conclusions

In [22, 23, 24] we have studied the relationships between a preferential logic of common sense reasoning and two different neural network models, Self-Organising Maps and Multilayer Perceptrons, showing that a multi-preferential semantics can be used to provide a logical model of the neural network behavior after training. Such a model can be used to learn or to validate conditional knowledge from the empirical data used for training and generalization, by model checking of logical properties. A two-valued KLM-style preferential interpretation with multiple preferences and a fuzzy semantics have been considered, based on the idea of associating preference relations to categories (in the case of SOMs) or to neurons (for Multilayer Perceptrons). Due to the diversity of the two models we would expect that a similar approach might be extended to other neural network models and learning approaches. The plausibility of concept-wise multipreference semantics is supported by the fact that self-organising maps are considered as psychologically and biologically plausible neural network models. This multipreference semantics has been shown to satisfy the KLM properties in the two-valued case [19], and most of the KLM properties in the fuzzy case, depending on their reformulation and on the fuzzy combination functions considered [34].

Much work has been devoted, in recent years, to the combination of neural networks and symbolic reasoning [35, 36, 37], leading to the definition of new computational models [38, 39, 40, 41] and to extensions of logic programming languages with neural predicates [42, 43]. Among the earliest systems combining logical reasoning and neural learning are the Knowledge-Based Artificial Neural Network (KBANN) [44] and the Connectionist Inductive Learning and Logic Programming (CILP) [45] systems and Penalty Logic [46], a non-monotonic reasoning formalism used to establish a correspondence with symmetric connectionist networks. The relationships between normal logic programs under the stable model semantics [47] and neural networks have been investigated by Garcez and Gabbay [45, 35] and by Hitzler et al. [48].

The correspondence between neural network models and fuzzy systems has been first investigated by Kosko in his seminal work [49]. We have adopted the usual way of viewing concepts in fuzzy DLs [50, 51, 52], and we have used fuzzy concepts within a multipreference semantics, based on a semantic closure construction in the line of Lehmann's semantics for lexicographic closure [32] and strictly related to Kern-Isberner's c-representations [7, 33]. Furthermore, we have adopted a preferential semantics with multiple preferences, in order to make it concept-wise: each distinguished concept C_i has its own set $\mathcal{T}_{C_i}$ of (weighted) typicality inclusions, and an associated preference relation $<_{C_i}$. This allows a preference relation to be associated to each category (e.g., in the preferential interpretation of SOMs) or to neurons (in a deep network). A combination of fuzzy logic with the preferential semantics of conditional knowledge bases has been first studied by Casini and Straccia [53], who have developed a rational closure construction for propositional Gödel logic.

For Multilayer Perceptrons, it has been proven [24] that a deep network can itself be regarded as a weighted conditional knowledge base (under some conditions on the activation function). This opens to the possibility of adopting a conditional logics as a basis for neuro-symbolic integration. While a neural network, once trained, is able and fast in classifying the new stimuli (that is, it is able to do instance checking), all other reasoning services such as satisfiability, entailment and model-checking are missing. These capabilities would be needed for dealing with tasks combining empirical and symbolic knowledge, such as, for instance: proving whether the network satisfies some (strict or conditional) properties; learning the weights of a conditional knowledge base from empirical data, and combine the defeasible inclusions extracted from a neural network with other defeasible or strict inclusions for inference.

To make these tasks possible, the development of proof methods for such logics is a preliminary step. In the two-valued case multipreference entailment is decidable for weighted $\mathcal{EL}^{\perp}$ knowledge bases, and proof methods for reasoning with weighted conditional knowledge bases in $\mathcal{EL}^{\perp}$ can, for instance, exploit Answer Set Programming (ASP) encodings of the concept-wise multipreference semantics [54], using *asprin* [55] to achieve defeasible reasoning, an approach already considered for ranked $\mathcal{EL}^{+}_{\perp}$ knowledge bases [19]. In the fuzzy case, an open problem is whether the notion of fuzzy-multipreference entailment is decidable (even for the small fragment of $\mathcal{EL}$ without roles), and under which choice of fuzzy logic combination functions. Undecidability results for fuzzy description logics with general inclusion axioms [56, 57, 58] motivate the investigation of decidable approximations of fuzzy-multipreference entailment.

An interesting issue is whether the mapping of deep neural networks to weighted conditional knowledge bases can be extended to more complex neural network models, such as Graph neural networks [38], or whether different logical formalisms and semantics would be needed. Another issue is whether the fuzzy-preferential interpretation of neural networks can be related with the probabilistic interpretation of neural networks based on statistical AI. This is an interesting issue, as the fuzzy DL interpretations we have considered in [24], where concepts are regarded as fuzzy sets, also suggests a probabilistic account based on Zadeh's probability of fuzzy events [30]. We refer to [23] for some results concerning a probabilistic interpretation of SOMs and to [59] for a preliminary account for MLPs.

Acknowledgments

We thank the anonymous referees for their helpful comments and suggestions. This research has been partially supported by INDAM-GNCS Projects 2019 and 2020.

References

[1] D. Gabbay, Theoretical foundations for non-monotonic reasoning in expert systems, in: A. K.R. (Ed.), Logics and Models of Concurrent Systems, volume 13 of *NATO ASI Series (Series F: Computer and Systems Sciences)*, Springer, 1985.

[2] D. Makinson, General theory of cumulative inference, in: Non-Monotonic Reasoning, 2nd International Workshop, Grassau, FRG, June 13-15, 1988, Proceedings, 1988, pp. 1–18.

[3] J. Pearl, Probabilistic Reasoning in Intelligent Systems Networks of Plausible Inference, Morgan Kaufmann, 1988.

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

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

[6] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, H. Prade, Inconsistency management and prioritized syntax-based entailment, in: Proc. IJCAI'93, Chambéry, France, August 28 - September 3, Morgan Kaufmann, 1993, pp. 640–647.

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

[8] D. Lewis, Counterfactuals, Basil Blackwell Ltd, 1973.

[9] D. Nute, Topics in conditional logic, Reidel, Dordrecht (1980).

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

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

[16] 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.

[17] K. Britz, G. Casini, T. Meyer, K. Moodley, U. Sattler, I. Varzinczak, Principles of KLM-style defeasible description logics, ACM Trans. Comput. Log. 22 (2021) 1:1–1:46.

[18] L. Giordano, V. Gliozzi, A reconstruction of multipreference closure, Artif. Intell. 290 (2021).

[19] L. Giordano, D. Theseider Dupré, An ASP approach for reasoning in a concept-aware multipreferential lightweight DL, Theory and Practice of Logic programming, TPLP 10(5) (2020) 751–766.

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

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

[22] L. Giordano, V. Gliozzi, D. Theseider 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.

[23] L. Giordano, V. Gliozzi, D. Theseider 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.

[24] L. Giordano, D. Theseider Dupré, Weighted defeasible knowledge bases and a multipreference semantics for a deep neural network model, in: Proc17th European Conf. on Logics in AI, JELIA 2021, May 17-20, volume 12678 of *LNCS*, Springer, 2021, pp. 225–242.

[25] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial

intelligence (XAI), IEEE Access 6 (2018) 52138–52160.

[26] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (2019) 93:1–93:42.

[27] A. B. Arrieta, N. D. Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.

[28] V. Gliozzi, K. Plunkett, Grounding bayesian accounts of numerosity and variability effects in a similarity-based framework: the case of self-organising maps, Journal of Cognitive Psychology 31 (2019).

[29] L. Giordano, V. Gliozzi, D. Theseider Dupré, Towards a conditional interpretation of self organising maps, in: Italian Workshop on Explainable Artificial Intelligence, XAI.it, November 25-26, 2020, volume 2742 of *CEUR*, 2020, pp. 127–134.

[30] L. Zadeh, Probability measures of fuzzy events, J.Math.Anal.Appl 23 (1968) 421–427.

[31] G. Hinton, Learning distributed representation of concepts, in: Proceedings 8th Annual Conference of the Cognitive Science Society. Erlbaum, Hillsdale, NJ, 1986.

[32] D. J. Lehmann, Another perspective on default reasoning, Ann. Math. Artif. Intell. 15 (1995) 61–82.

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

[34] L. Giordano, On the KLM properties of a fuzzy DL with Typicality, 2021. arXiv:2106.00390, submitted.

[35] A. S. d'Avila Garcez, K. Broda, D. M. Gabbay, Symbolic knowledge extraction from trained neural networks: A sound approach, Artif. Intell. 125 (2001) 155–207.

[36] A. S. d'Avila Garcez, L. C. Lamb, D. M. Gabbay, Neural-Symbolic Cognitive Reasoning, Cognitive Technologies, Springer, 2009.

[37] A. S. d'Avila Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, S. N. Tran, Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning, FLAP 6 (2019) 611–632.

[38] L. C. Lamb, A. S. d'Avila Garcez, M. Gori, M. O. R. Prates, P. H. C. Avelar, M. Y. Vardi, Graph neural networks meet neural-symbolic computing: A survey and perspective, in: C. Bessiere (Ed.), Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, ijcai.org, 2020, pp. 4877–4884.

[39] L. Serafini, A. S. d'Avila Garcez, Learning and reasoning with logic tensor networks, in: Proc. AI*IA 2016, Genova, Italy, November 29 - December 1, volume 10037 of *LNCS*, Springer, 2016, pp. 334–348.

[40] P. Hohenecker, T. Lukasiewicz, Ontology reasoning with deep neural networks, J. Artif. Intell. Res. 68 (2020) 503–540.

[41] D. Le-Phuoc, T. Eiter, A. Le-Tuan, A scalable reasoning and learning approach for neural-symbolic stream fusion, in: AAAI 2021, February 2-9, AAAI Press, 2021, pp. 4996–5005.

[42] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, L. D. Raedt, Deepproblog: Neural probabilistic logic programming, in: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada, 2018, pp. 3753–3763.

[43] Z. Yang, A. Ishay, J. Lee, Neurasp: Embracing neural networks into answer set programming, in: C. Bessiere (Ed.), Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, ijcai.org, 2020, pp. 1755–1762.

[44] G. G. Towell, J. W. Shavlik, Knowledge-based artificial neural networks, Artif. Intell. 70 (1994) 119–165.

[45] A. S. d'Avila Garcez, G. Zaverucha, The connectionist inductive learning and logic programming system, Appl. Intell. 11 (1999) 59–77.

[46] G. Pinkas, Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge, Artif. Intell. 77 (1995) 203–247.

[47] M. Gelfond, V. Lifschitz, The stable model semantics for logic programming, in: Logic Programming, Proc. of the 5th Int. Conf. and Symposium, 1988, pp. 1070–1080.

[48] P. Hitzler, S. Hölldobler, A. K. Seda, Logic programs and connectionist networks, J. Appl. Log. 2 (2004) 245–272.

[49] B. Kosko, Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence, Prentice Hall, 1992.

[50] U. Straccia, Towards a fuzzy description logic for the semantic web (preliminary report), in: The Semantic Web: Research and Applications, Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29 - June 1, 2005, Proceedings, volume 3532 of *Lecture Notes in Computer Science*, Springer, 2005, pp. 167–181.

[51] T. Lukasiewicz, U. Straccia, Managing uncertainty and vagueness in description logics for the semantic web, J. Web Semant. 6 (2008) 291–308.

[52] F. Bobillo, U. Straccia, The fuzzy ontology reasoner fuzzydl, Knowl. Based Syst. 95 (2016) 12–34.

[53] G. Casini, U. Straccia, Towards rational closure for fuzzy logic: The case of propositional gödel logic, in: Logic for Programming, Artificial Intelligence, and Reasoning - 19th International Conference, LPAR-19, Stellenbosch, South Africa, December 14-19, 2013. Proceedings, volume 8312 of *LNCS*, Springer, 2013, pp. 213–227. URL: https://doi.org/10.1007/978-3-642-45221-5_16.

[54] L. Giordano, D. Theseider Dupré, Weighted conditional $\mathcal{EL}$ knowledge bases with integer weights: an ASP approach, in: Int. Conf. on logic Programming, ICLP 2021, 2021. To appear.

[55] G. Brewka, J. P. Delgrande, J. Romero, T. Schaub, asprin: Customizing answer set preferences without a headache, in: Proc. AAAI 2015, 2015, pp. 1467–1474.

[56] F. Baader, R. Peñaloza, Are fuzzy description logics with general concept inclusion axioms decidable?, in: FUZZ-IEEE 2011, IEEE International Conference on Fuzzy Systems, Taipei, Taiwan, 27-30 June, 2011, Proceedings, IEEE, 2011, pp. 1735–1742.

[57] M. Cerami, U. Straccia, On the undecidability of fuzzy description logics with gcis with lukasiewicz t-norm, CoRR abs/1107.4212 (2011). URL: http://arxiv.org/abs/1107.4212.

[58] S. Borgwardt, R. Peñaloza, Undecidability of fuzzy description logics, in: G. Brewka, T. Eiter, S. A. McIlraith (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012, AAAI Press, 2012, pp. 232–242.

[59] L. Giordano, D. Theseider 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/2012.13421. arXiv:2012.13421.