Preferential Reasoning with Typicality and Neural Network Models (Extended Abstract)

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

[1]DISIT - Università del Piemonte Orientale, Italy
[2]Università degli Studi di Torino

Abstract

In this extended abstract we report some results concerning the relationships between a multipreferential semantics for defeasible reasoning in knowledge representation and some neural network models, namely Self-Organising Maps and Multilayer Perceptrons.

Keywords

Common Sense Reasoning, Preferential semantics, Weighted Conditionals, Neural Networks

1. Introduction

We report some results concerning the relationships between a multipreference semantics for defeasible reasoning in knowledge representation and some neural network models, namely, Self-Organising Maps (SOMs) and Multilayer Perceptrons (MLPs). In particular, weighted knowledge bases for description logics are considered under a "concept-wise" multipreference semantics and, in the fuzzy case, provide a preferential interpretation of MLPs.

Preferential approaches have been used to provide axiomatic foundations of non-monotonic and common sense reasoning [1, 2, 3, 4], and, more recently, they have been extended to description logics (DLs) to deal with inheritance with exceptions in ontologies, by allowing for non-strict forms of inclusions, called *typicality or defeasible inclusions*, with different preferential semantics [5, 6] and closure constructions [7, 8, 9, 10, 11]. In this abstract, we consider a concept-wise multipreference semantics as a semantics for weighted knowledge bases, i.e. knowledge bases in which defeasible or typicality inclusions of the form $\mathbf{T}(C) \sqsubseteq D$ (meaning "the typical C's are D's" or "normally C's are D's") are given a positive or negative weight. A multipreference semantics taking into account preferences with respect to different concepts, was first introduced by the authors as a semantics for ranked DL knowledge bases [12]. For weighted knowledge bases, a different semantic closure construction is developed, still in the spirit of other semantic constructions in the literature, and is further extended to the fuzzy case.

The concept-wise multipreference semantics has been used to develop a semantic interpretations for some neural network models. Both an unsupervised model, Self-organising maps

OVERLAY 2021: 3rd Workshop on Artificial Intelligence and Formal Verification, Logic, Automata, and Synthesis, September 22, 2021, Padova, Italy
laura.giordano@uniupo.it (L. Giordano); valentina.gliozzi@unito.it (V. Gliozzi); dtd@uniupo.it (D. Theseider Dupré)

(SOMs)[13], which is regarded as psychologically and biologically plausible neural network models, and a supervised one, Multilayer Perceptrons (MLPs) [14] have been considered. 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 associated to concepts. For SOMs, the learned categories are regarded as concepts $C_1, \ldots, C_n$ so that a preference relation (over the domain of input stimuli) is associated to each category [15, 16]. In case of MLPs, each neuron in the deep network (including hidden neurons) is associated to a concept, so that preference relations are associated to neurons [17]. For MLPs, the relationship between these logics of commonsense reasoning and deep neural networks is even stronger, as a deep neural network can be regarded as a conditional knowledge base, i.e., a set weighted conditionals. This has been achieved by developing a concept-wise fuzzy multipreference semantics for a DL with weighted defeasible inclusions. In the following we shortly recall these results and discuss some challenges from the standpoint of explainable AI [18, 19].

2. A Multi-preferential interpretation for SOMs and MPLs

The multipreference semantics (cw^m-semantics) has been first developed as a semantics for strengthening rational closure [20] and then it has been made *concept-wise* to provide a semantics for ranked $\mathcal{EL}$ knowledge bases [21] by capturing, with different preference relationships, the preferences among domain elements with respect to different concepts.

In weighted knowledge bases [17], besides standard inclusions (called *strict* inclusions), defeasible inclusions of the form $\mathbf{T}(C) \sqsubseteq D$ are allowed with a weight w, whose meaning is that "typical Cs are Ds" (or "normally Cs are Ds") with weight w. Such inclusions correspond to conditionals $C \mathrel{|\!\sim} D$ in Kraus, Lehmann and Magidor (KLM) preferential logics [3], while the positive or negative weights of defeasible inclusions represent their plausibility (implausibility).

A *cw^m-interpretation* is defined by adding to a standard DL interpretation, a set of preference relations $<_{C_1}, \ldots, <_{C_n}$ each one associated with a distinguished concept C_i. A DL interpretation is a pair $\langle \Delta, \cdot^I \rangle$, where Δ is a domain, and $\cdot^I$ an interpretation function, and each preference $<_{C_i}$ captures the relative typicality of domain individuals with respect to C_i. Preferences with respect to different concepts do not need to agree, as 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, by Pareto combination of the preference relations $<_{C_1}, \ldots, <_{C_n}$, but a more sophisticated notion of preference combination has also been considered [21], which takes into account the specificity relation among concepts. A typicality concept $\mathbf{T}(C)$ is then interpreted as the set of all minimal C elements with respect to $<$. It has been proven [21] that cw^m-entailment satisfies the KLM postulates of a preferential consequence relation [3].

Self-organising maps are psychologically and biologically plausible neural network models [13] that can learn after limited exposure to positive category examples, without need of contrastive information. They have been proposed as possible candidates to explain the psychological mechanisms underlying category generalisation. Multilayer Perceptrons [14] are deep networks. Learning algorithms in the two cases are quite different but our approach aims to capture, through a semantic interpretation, the behavior of the network obtained after training and not to model

learning. We have seen that this can be accomplished in both cases in a similar way, based on the multi-preferential semantics above and its fuzzy extension.

The result of the training phase is represented very differently in the two models: for SOMs it is given by a set of units spatially organized in a grid (where each unit u in the map is associated with a weight vector w_u of the same dimensionality as the input vectors); for MLPs, as a result of training, the weights of the synaptic connections have been learned. 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 associated to concepts. For SOMs, the learned categories are regarded as concepts $C_1, \ldots, C_n$ so that a preference relation (over the domain of input stimuli) is associated to each category by a notion of *relative distance* of a stimulus from its *Best Matching Unit* [15, 16]. For MLPs, units in the deep network (including hidden units), or a subset thereof, can be associated to concepts, each related to a preference (a well-founded modular partial order). In both cases, a multipreference interpretation can be constructed from the network after training, describing the input-output behavior of the network on the input stimuli considered. Such preferential interpretation can be used for checking properties like: are the instances of a category C_1 also instances of category C_2? are typical instances of a category C_1 also instances of category C_2? The verification can be done by *model-checking* on the multipreference interpretation describing the input-output behavior of the network [16, 17].

This kind of construction establishes strong relationships between the logics of commonsense reasoning and the neural network models, as the first ones are able to reason about the properties of the second ones. These relationships can be made even stronger for MLPs, as the neural network itself can be regarded as a conditional knowledge base.

Under a fuzzy extension of the multipreference semantics, it has been proven [17] that MLPs can be regarded as weighted conditional knowledge bases. The multipreference interpretation constructed over the set of input stimuli to describe the input-output behavior of the deep network, in the fuzzy case, exploits the activation value of each unit h for a stimulus x, which can be interpreted as the degree of membership of x in concept C_h. The fuzzy interpretation also induces a preference on the domain for each concept C_h. Such an interpretation can be proven to be a fuzzy multipreference model of the knowledge base extracted from the network.

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 $j_1, \ldots, j_m$, whose output signals are the input signals for unit k, with synaptic weights $w_{k,j_1}, \ldots, w_{k,j_m}$. One can define for each unit k a set $\mathcal{T}_{C_k}$ of typicality inclusions, with their associated weights, as follows: $\mathbf{T}(C_k) \sqsubseteq C_{j_1}$ with w_{k,j_1}, $\ldots$, $\mathbf{T}(C_k) \sqsubseteq C_{j_m}$ with w_{k,j_m}. The collection of the defeasible inclusions $\mathcal{T}_{C_k}$ for all concepts (units) defines the weighted conditional KB associated to the network.

The definition of the cw^m-interpretation for a weighted conditional knowledge base exploits a closure construction in the same spirit of the one considered by Lehmann [22] to define the lexicographic closure, but more similar to Kern-Isberner's c-representations [23, 24]. As a difference, our construction in [17] is concept-wise, thus considering the modular structure of the knowledge base (and of the network). In the fuzzy case, to guarantee that the weights computed from the KB are coherent with the fuzzy interpretation of concepts, a notion of *coherent (fuzzy) multipreference interpretation* is introduced. We refer to [25] for a study of its KLM properties.

3. Conclusions

In [15, 17, 16] we have studied the relationships between multi-preferential (and fuzzy) logics 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 a 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 expect that a similar approach might be extended to other neural network models and learning approaches.

Much work has been devoted, in recent years, to the combination of neural networks and symbolic reasoning [26, 27, 28], leading to the definition of new computational models [29, 30, 31, 32] and to extensions of logic programming languages with neural predicates [33, 34]. Among the earliest systems combining logical reasoning and neural learning are the KBANN [35] and the CLIP [36] systems and Penalty Logic [37]. The relationships between normal logic programs and connectionist network have been investigated by Garcez and Gabbay [36, 26] and by Hitzler et al. [38]. The correspondence between neural network models and fuzzy systems has been first investigated by Kosko in his seminal work [39]. A fuzzy extension of preferential logics has been studied by Casini and Straccia [40] based on a Rational Closure construction.

For Multilayer Perceptrons, it has been proven [17] that a deep network can itself be regarded as a weighted conditional knowledge base. 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; combining 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 [17]. In the fuzzy case, whether the notion of coherent fuzzy multipreference entailment is decidable is an open problem even for the small fragment of $\mathcal{EL}^{\perp}$ without roles. Undecidability results for fuzzy description logics with general inclusion axioms [41, 42] motivate the investigation of decidable approximations of fuzzy-multipreference entailment.

An 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 [29]. 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. Indeed, interpreting concepts as fuzzy sets suggests a probabilistic account based on Zadeh's probability of fuzzy events [43], an approach exploited for SOMs [16]. We refer to [44] for a preliminary account for MLPs.

Acknowledgments We thank the anonymous referees for their helpful comments.

References

[1] J. Delgrande, A first-order conditional logic for prototypical properties, Artificial Intelligence 33 (1987) 105–130.

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

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

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

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

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

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

[15] L. Giordano, V. Gliozzi, D. Theseider Dupré, On a plausible concept-wise multipreference semantics and its relations with self-organising maps, in: CILC 2020, Rende, Italy, October 13-15, 2020, volume 2710 of *CEUR*, 2020, pp. 127–140.

[16] 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.

[17] 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. URL: https://arxiv.org/abs/2012.13421, extended version.

[18] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160.

[19] 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.

[20] V. Gliozzi, Reasoning about multiple aspects in rational closure for DLs, in: Proc. AI*IA 2016 - XVth International Conference of the Italian Association for Artificial Intelligence, Genova, Italy, November 29 - December 1, 2016, 2016, pp. 392–405.
[21] 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.
[22] D. J. Lehmann, Another perspective on default reasoning, Ann. Math. Artif. Intell. 15 (1995) 61–82.
[23] G. Kern-Isberner, Conditionals in Nonmonotonic Reasoning and Belief Revision - Considering Conditionals as Agents, volume 2087 of *LNCS*, Springer, 2001.
[24] G. Kern-Isberner, C. Eichhorn, Structural inference from conditional knowledge bases, Stud Logica 102 (2014) 751–769.
[25] L. Giordano, On the KLM properties of a fuzzy DL with Typicality, 2021. arXiv:2106.00390, to appear in ECSQARU 2021.
[26] 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.
[27] A. S. d'Avila Garcez, L. C. Lamb, D. M. Gabbay, Neural-Symbolic Cognitive Reasoning, Cognitive Technologies, Springer, 2009.
[28] 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.
[29] 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.), IJCAI 2020, ijcai.org, 2020, pp. 4877–4884.
[30] 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, 2016, volume 10037 of *LNCS*, Springer, 2016, pp. 334–348.
[31] P. Hohenecker, T. Lukasiewicz, Ontology reasoning with deep neural networks, J. Artif. Intell. Res. 68 (2020) 503–540.
[32] 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.
[33] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, L. D. Raedt, Deepproblog: Neural probabilistic logic programming, in: NeurIPS 2018, 3-8 December 2018, Montréal, Canada, 2018, pp. 3753–3763.
[34] Z. Yang, A. Ishay, J. Lee, Neurasp: Embracing neural networks into answer set programming, in: C. Bessiere (Ed.), IJCAI 2020, ijcai.org, 2020, pp. 1755–1762.
[35] G. G. Towell, J. W. Shavlik, Knowledge-based artificial neural networks, Artif. Intell. 70 (1994) 119–165.
[36] A. S. d'Avila Garcez, G. Zaverucha, The connectionist inductive learning and logic programming system, Appl. Intell. 11 (1999) 59–77.
[37] G. Pinkas, Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge, Artif. Intell. 77 (1995) 203–247.
[38] P. Hitzler, S. Hölldobler, A. K. Seda, Logic programs and connectionist networks, J. Appl. Log. 2 (2004) 245–272.

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

[40] 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 Int. Conf., LPAR-19, Stellenbosch, South Africa, December 14-19, 2013. Proceedings, volume 8312 of *LNCS*, Springer, 2013, pp. 213–227.

[41] F. Baader, R. Peñaloza, Are fuzzy description logics with general concept inclusion axioms decidable?, in: FUZZ-IEEE 2011, Taipei, 27-30 June, 2011, IEEE, 2011, pp. 1735–1742.

[42] 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.

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

[44] 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.