A fuzzy institution for neural conceptors

Till Mossakowski[1], Razvan Diaconescu[2], and Martin Glauer[1]

[1] Otto-von-Guericke-Universität Magdeburg, Germany
till@iws.cs.ovgu.de
[2] Simion Stoilow Institute of Mathematics of the Romanian Academy, Bucharest, Romania
razvansdiaconescu@gmail.com

Neural networks have been successfully used for learning tasks [8], but they exhibit the problem that the way they compute their output generally cannot be interpreted or explained at a higher conceptual level [9]. The field of neuro-symbolic integration [1] addresses this problem by combining neural networks with logical methods. However, most approaches in the field (like e.g. logic tensor networks [3]) are localist, that is, predicates or other symbolic items are represented in small sub-networks. This contrasts with the distributed representation of knowledge in (deep learning) neural networks, which seems to be much more flexible and powerful.

Jaeger's conceptors [6, 7] provide such a distributed representation while simultaneously providing logical operators and concept hierarchies that foster explainability. The basic idea is to to take a recurrent neural network and not use it for learning through back-propagation, but rather feed it with input signals, leading to a state space that can be captured as a certain ellipsoid using a conceptor matrix. Conceptor matrices are positive semidefinite matrices with singular values (which represent the lengths of the ellipsoid axes) ranging in [0,1].

In [7], Jaeger introduces and studies an algebra of conceptors, providing the logical operations "and", "or" and "not" (which however satisfy only part of the laws of Boolean algebra and do not even form a lattice) as well as a scaling operation called aperture adaption, and an interpolation operation. A crucial advantage of conceptors over ordinary neural networks is that using the algebra of conceptors, training examples can easily be added to conceptors, without the need of re-training with the whole sample. Moreover, the Löwner ordering on conceptor matrices expresses a concept hierarchy. For reasoning about conceptors, two logics are introduced, an extrinsic and an intrinsic one. Both logics are based on the conceptor algebra operations. The extrinsic logic provides a first-order logic with atomic formulas based on the Löwner ordering. This leads to two levels of Boolean operations: one within conceptor terms, and one within the first-order logic. The intrinsic operation avoids this duplication by only working on conceptor terms and comparing them with a fixed conceptor. In [7], Jaeger formalises both logics as institutions [4], which are an abstract formalisation of the notion of logical system. Moreover, he states the open problem of developing a proof calculus and theorem proving support for these logics.

We here argue that both of these logics are not completely adequate for reasoning about conceptors, because they both can ultimately speak only about the Löwner ordering, i.e. crisp statements that can be either true or false. We propose that a more promising approach is to view conceptors as a kind of fuzzy sets. Indeed, their Boolean operators satisfy the (appropriate generalisation of) T-norm and T-conorm laws, and form a (generalised) De Morgan Triplet [10, 5]. This is remarkable, because conceptors have not been introduced as a neuro-fuzzy approach (and note that neuro-fuzzy approaches generally are localist in the above sense, while conceptors provide a global distributed representation of knowledge).

We argue that an appropriate conceptor logic should not have crisp but fuzzy statements as its atomic constituents:

- classification of an N-dimensional signal vector z by a $N \times N$ conceptor matrix C, yielding the fuzzy truth value $z^T Cz/N$ (which can be seen as fuzzy set membership),
- a "fuzzy subset" relation $C_1 \leq C_2$ between conceptors.

Atomic formulas use conceptor terms formed with the same operations as in Jaeger's logics. On this basis, we develop a many-valued institution in the sense of [2] for conceptors. A (fuzzy) first-order logic on top of that can be obtained using general methods of defining fuzzy connectives and quantifiers.

References

1. Tarek R. Besold, Artur S. d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro M. Domingos, Pascal Hitzler, Kai-Uwe Kühnberger, Luís C. Lamb, Daniel Lowd, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, Hoifung Poon, and Gerson Zaverucha. Neural-symbolic learning and reasoning: A survey and interpretation. *CoRR*, abs/1711.03902, 2017.
2. Razvan Diaconescu. Institutional semantics for many-valued logics. *Fuzzy Sets and Systems*, 218:32–52, 2013.
3. Ivan Donadello, Luciano Serafini, and Artur S. d'Avila Garcez. Logic tensor networks for semantic image interpretation. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 1596–1602. ijcai.org, 2017.
4. J. A. Goguen and R. M. Burstall. Institutions: Abstract model theory for specification and programming. *Journal of the Association for Computing Machinery*, 39:95–146, 1992. Predecessor in: LNCS 164, 221–256, 1984.
5. M.M. Gupta and J. Qi. Theory of T-norms and fuzzy inference methods. *Fuzzy Sets and Systems*, 40:431–450, 1991.
6. Herbert Jaeger. Conceptors: an easy introduction. *CoRR*, abs/1406.2671, 2014.
7. Herbert Jaeger. Controlling recurrent neural networks by conceptors. *CoRR*, abs/1403.3369, 2014.
8. Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
9. E. Smith and S. Kosslyn. *Cognitive psychology: Mind and brain.* Upper Saddle River, NJ: Prentice-Hall Inc., 2006.
10. Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.