Inadequacies of Logic to Address Big Data

Marcelo Finger*
Universidade de São Paulo, São Paulo, Brazil

Keywords: Big Data, First-order Logic, Propositional Logic, SAT, Łukasiewicz Logic

Logic and logicians have not been able to address the recent challenges proposed by the Big Data phenomenon. Initially, this might have been discarded as a passing fashion, but the consequential achievements of this field of investigation cannot be discarded.

This work, which may be seen as a position paper, intends to address the inadequacies of Logic**S**, in their several forms, to deal with several problems arising from the need to deal with vast quantities of data. Our exposition covers: efficiency, modelling capacity and explicability. But, we argue, there are unexplored avenues that may be left for logic researchers to explore some important aspects of Big Data modelling.

Efficiency. Simply put, the tools of logic are too complex. First-order logic is undecidable, and something as simple as propositional logic is intractable. There are known tractable fragments of propositional logics [Schaefer, 1978], but modelling real world applications, even when they can be transformed into propositional logics, rarely fits into one of those fragments.

Modelling. Formal logic was developed as a tentative to formalize mathematics. After a bumpy start, first-order logic was consolidated as a tool In which the foundations of mathematics could be formalized, and so set theories could be axiomatized and compared, and now formal proofs may even be automated. But then computer scientist try to apply first order predicates as a model of real world entities and relationships. And for a while it succeeded, by enforcing that the Universe of this course was all contained in a finite database. But this initial View does not transfer to Big Data, which represents fuzzy, contradictory, and mostly inexplicable observations of human behavior. The difference between mathematical concepts and human concepts are that the former is designed to be precise, timeless, immutable, while the latter are anything but. So there is an unbridgeable impedance mismatch between first order predicates and human concepts, a chasm that has been researched and described by recent research in neuroscience [Barrett, 2017].

Explainability. Explainability is a big problem for modern machine learning methods based on neural networks, and logic has been volunteered as the ideal vehicle in which explainable AI should be developed [Darwiche, 2023]. However, the sheer size of models developed with big data makes this approach impractical. A large formula that fits into a 200 MB file, no matter what logic it encodes, is no more satisfying as an explanation as any neural network.

Is there a way out for logic in the context of Big Data? I suggest that yes, logic should be used for what logic is good at: proofs. The idea is to model neural networks, or more generally, any continuous function, using a specific logic. Any continuous function can be densely approximated by piecewise linear functions, and recent work has shown how to represent any such function as a pair of formulas in Łukasiewicz infinitely valued logic [Preto and Finger, 2022b]. The idea is to represent or approximate a neural network as a piecewise linear function, automatically translated

*mfinger@ime.usp.br

into a formula and, no matter how big the formula, use logic's proof methods to demonstrate the existence of refutability of some properties of the original neural network [Preto and Finger, 2022a]. For efficiency, approximations of the target logic are being developed [Finger and Preto, 2023] For further technical details referred to [Preto and Finger, 2023, Preto et al., 2023]

References

L. F. Barrett. *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt, Boston, 2017.

A. Darwiche. Logic for explainable ai, 2023. URL https://arxiv.org/abs/2305.05172.

M. Finger and S. Preto. Polyhedral semantics and the tractable approximation of Łukasiewicz infinitely-valued logic. *Journal of Logic and Computation*, page exad059, 10 2023. ISSN 0955-792X. doi: 10.1093/logcom/exad059. URL https://doi.org/10.1093/logcom/exad059.

S. Preto and M. Finger. Proving properties of binary classification neural networks via Łukasiewicz logic. *Logic Journal of the IGPL*, 2022a. ISSN 1367-0751. doi: 10.1093/jigpal/jzac050. URL https://doi.org/10.1093/jigpal/jzac050.

S. Preto and M. Finger. Efficient representation of piecewise linear functions into Łukasiewicz logic modulo satisfiability. page 1?26, 2022b. doi: 10.1017/S096012952200010X.

S. Preto and M. Finger. Effective reasoning over neural networks using łukasiewicz logic. In *Compendium of Neurosymbolic Artificial Intelligence*, pages 609–630. IOS Press, 2023.

S. Preto, F. Manyà, and M. Finger. Benchmarking łukasiewicz logic solvers with properties of neural networks. In *2023 IEEE 53rd International Symposium on Multiple-Valued Logic (ISMVL)*, pages 158–163, 2023. doi: 10.1109/ISMVL57333.2023.00039.

T. J. Schaefer. The complexity of satisfiability problems. In *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing*, STOC '78, page 216–226, New York, NY, USA, 1978. Association for Computing Machinery. ISBN 9781450374378. doi: 10.1145/800133.804350. URL https://doi.org/10.1145/800133.804350.