Comparing Differentiable Logics for Learning with Logical Constraints

Thomas Flinkow, Barak A. Pearlmutter, Rosemary Monahan

Department of Computer Science, Maynooth University

https://www.cs.nuim.ie/research/pop/maivv.html

thomas.flinkow@mu.ie

Maynooth University National University of Ireland Maynooth

Taighde Éireann Research Ireland

1. Background & Motivation: Property-driven ML

Standard ML: Given data $\boldsymbol{x}$, target $\boldsymbol{y}$, and loss $\mathcal{L}$,

$$\text{minimise} \underset{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}{\mathbb{E}} \mathcal{L}(\boldsymbol{x},\boldsymbol{y}).$$

Adversarial training and DL2 [1]: Learn to satisfy constraints ϕ of the form $\forall \boldsymbol{x}.\, P(\boldsymbol{x}) \rightarrow Q(\boldsymbol{x})$ by:

- finding a counterexample $\boldsymbol{x}^*$ that does *not satisfy* Q in the input space $\mathcal{S}$ induced by P (*outside* train set) using PGD:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}'\in\mathcal{S}}{\arg\max}\ \mathcal{L}_\phi(\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})$$

- and using this counterexample in training:

$$\text{minimise} \underset{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}{\mathbb{E}} [\lambda \underbrace{\mathcal{L}(\boldsymbol{x},\boldsymbol{y})}_{\text{prediction loss}} + (1-\lambda) \underbrace{\mathcal{L}_\phi(\boldsymbol{x},\boldsymbol{x}^*,\boldsymbol{y})}_{\text{logical constraint loss}}].$$

Differentiable Logics: Choice of many logics (e.g. DL2 [1], STL [2], fuzzy logics [3, 4], …) to translate logical constraints into logical loss, which differ in their domain and operators.

Research Question: How do they compare in terms of: (1) learning behaviour, (2) logical consistency, and (3) in practice?

2. Investigating Learning Behaviour (Derivatives)

- **Conjunction.** *Shadow-lifting* [2] requires the truth value of a conjunction to increase when the truth value of a conjunct does:

$$\left.\frac{\partial[\![x_1 \wedge x_2]\!]_L}{\partial x_i}\right|_{x_1=x_2=\rho} > 0 \quad \text{for all } i \in \{1,2\}.$$

- **Implication.** Derivatives of implication allow Modus tollens and Modus ponens reasoning [3]; two important inference rules.

Findings: DL2 and the Reichenbach fuzzy logic have shadow-lifting conjunctions. Only the Reichenbach implication closely follows Modus tollens and Modus ponens reasoning.

3. Investigating Logical Consistency

Idea [5]: A tautology τ should be true for all possible truth values:

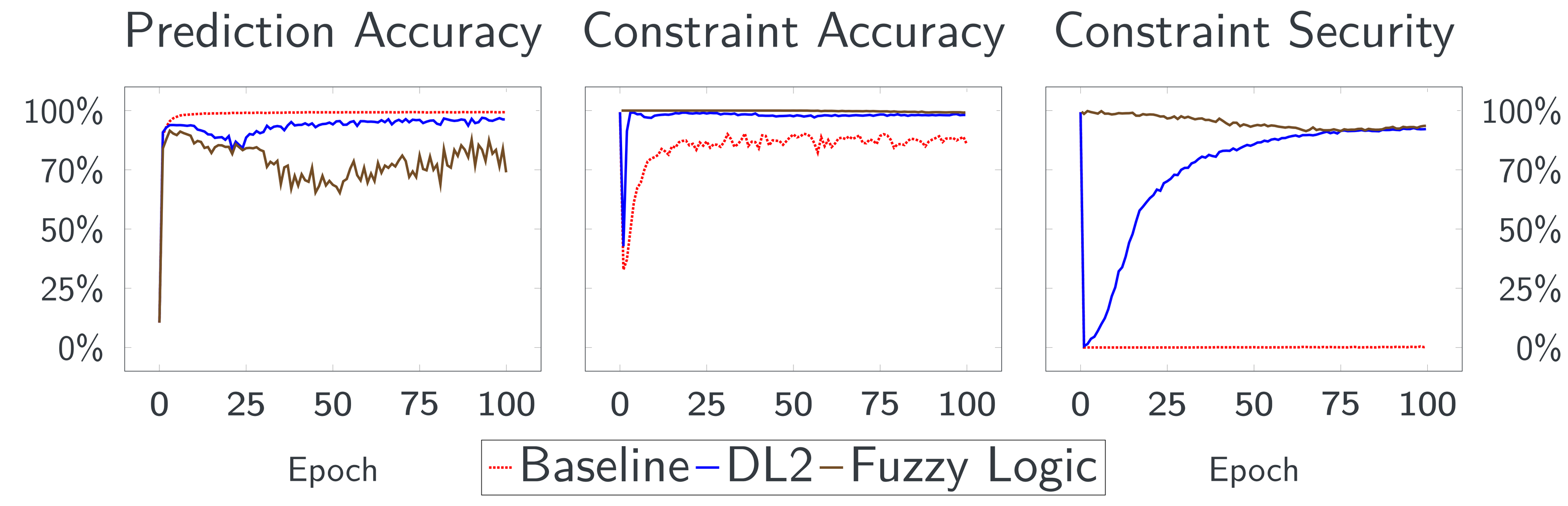
$$\int \cdots \int_{[0,1]} [\![\tau(x_1,\dots,x_n)]\!]_L \,\mathrm{d}x_n \cdots \mathrm{d}x_1$$

Tautology	Gödel	Łukasiewicz	Reichenbach
Primitive propositions			
$(P \vee P) \rightarrow P$	0.50	0.75	0.75
$Q \rightarrow (P \vee Q)$	0.83	1	0.92
$(P \vee Q) \rightarrow (Q \vee P)$	0.67	1	0.86
Law of excluded middle			
$P \vee \neg P$	0.75	1	0.83
Law of contradiction			
$\neg(P \wedge \neg P)$	0.75	1	0.83
Law of double negation			
$P \leftrightarrow \neg(\neg P)$	0.50	1	0.70
Laws of tautology			
$P \leftrightarrow (P \wedge P)$	0.50	0.75	0.69
$P \leftrightarrow (P \vee P)$	0.50	0.75	0.69
De Morgan's laws			
$\neg(P \wedge Q) \leftrightarrow (\neg P \vee \neg Q)$	0.67	1	0.75
$\neg(P \vee Q) \leftrightarrow (\neg P \wedge \neg Q)$	0.33	1	0.75
Average Consistency	**0.60**	**0.93**	**0.78**

Findings: R-implications (Łukasiewicz and Goguen)—except Gödel—are generally more consistent than S,N-implications (Reichenbach and Kleene-Dienes).

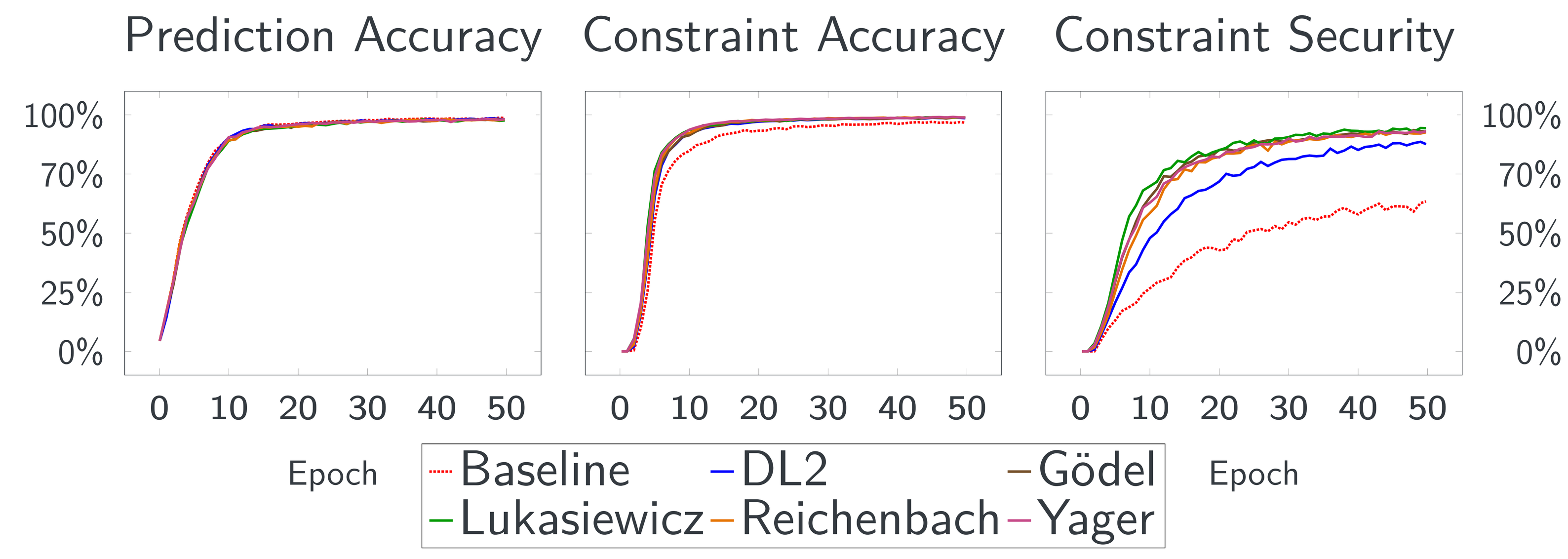
4. Training Experiments

Constraint: $\mathrm{SR}(\boldsymbol{x},\epsilon)\colon \forall \boldsymbol{x}' \in \mathbb{B}(\boldsymbol{x};\epsilon).\, \|f(\boldsymbol{x}') - f(\boldsymbol{x})\|_\infty \leq \delta.$



Constraint: The sum of probabilities of groups of related signs must be either very high or very low.

(a) unique signs (b) speed limit signs



Findings: Property-driven training with *any* differentiable logic generally leads to significantly improved constraint satisfaction.

5. Verification Experiment on MNIST

Constraint: $\mathrm{SCR}(\boldsymbol{x},\epsilon)\colon \forall \boldsymbol{x}' \in \mathbb{B}(\boldsymbol{x};\epsilon).\, f(\boldsymbol{x}')_y \geq \delta.$

Using Marabou [6] to determine verified constraint satisfaction on 500 randomly chosen images on networks trained for $\epsilon = 0.4$.

Logic	Prediction Accuracy	Constraint Security	Verified Satisfaction $\epsilon=0.2$	Verified Satisfaction $\epsilon=0.3$	Verified Satisfaction $\epsilon=0.4$
Baseline	96.50 %	79.68 %	0.68 % ($^3/_{444}$)	0 % ($^0/_{500}$)	**0 %** ($^0/_{500}$)
DL2	93.07 %	100 %	92.98 % ($^{384}/_{413}$)	55.29 % ($^{183}/_{331}$)	**20.51 %** ($^{73}/_{356}$)
Fuzzy Logic	94.87 %	100 %	92.70 % ($^{368}/_{397}$)	52.16 % ($^{157}/_{301}$)	**9.22 %** ($^{27}/_{293}$)

Marabou was run with a per-image timeout of 30s.

Findings: Property-driven training yields some formal guarantees but fails to establish strong ones.

6. Future Work: Formal Guarantees & Expressiveness

- Expressive specifications for ML & temporal differentiable logics.
- Adopt certified training to establish formal guarantees.

References

1. Fischer, M. *et al.* *DL2: Training and Querying Neural Networks with Logic.* in *Proceedings of the 36th International Conference on Machine Learning* International Conference on Machine Learning (PMLR, 24th May 2019), 1931–1941.
2. Varnai, P. & Dimarogonas, D. V. *On Robustness Metrics for Learning STL Tasks.* in *2020 American Control Conference (ACC)* (July 2020), 5394–5399.
3. Van Krieken, E., Acar, E. & van Harmelen, F. Analyzing Differentiable Fuzzy Logic Operators. *Artificial Intelligence* **302**, 103602. ISSN: 0004-3702 (1st Jan. 2022).
4. Ślusarz, N., Komendantskaya, E., Daggitt, M., Stewart, R. & Stark, K. *Logic of Differentiable Logics: Towards a Uniform Semantics of DL.* in *EPiC Series in Computing* Proceedings of 24th International Conference on Logic for Programming, Artificial Intelligence and Reasoning. **94** (EasyChair, 3rd June 2023), 473–493.
5. Grespan, M. M., Gupta, A. & Srikumar, V. *Evaluating Relaxations of Logic for Neural Networks: A Comprehensive Study.* arXiv: 2107.13646 [cs]. Pre-published.
6. Katz, G. *et al.* *The Marabou Framework for Verification and Analysis of Deep Neural Networks.* in *Computer Aided Verification* (eds Dillig, I. & Tasiran, S.) (Springer International Publishing, Cham, 2019), 443–452. ISBN: 978-3-030-25540-4.

T. Flinkow, B. A. Pearlmutter, R. Monahan: **'Comparing differentiable logics for learning with logical constraints'**, In *Science of Computer Programming*, **244**, 103280 (Sep 2025). DOI: 10.1016/j.scico.2025.103280

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under grant number 20/FFP-P/8853.