Embedding Symbolic Knowledge into Deep Networks

Yaqi Xie*, Ziwei Xu*, Mohan S Kankanhalli, Kuldeep S. Meel, Harold Soh
{yaqixie, ziwei-xu, mohan, meel, harold}@comp.nus.edu.sg
* Equal contribution

NUS National University of Singapore

Overview

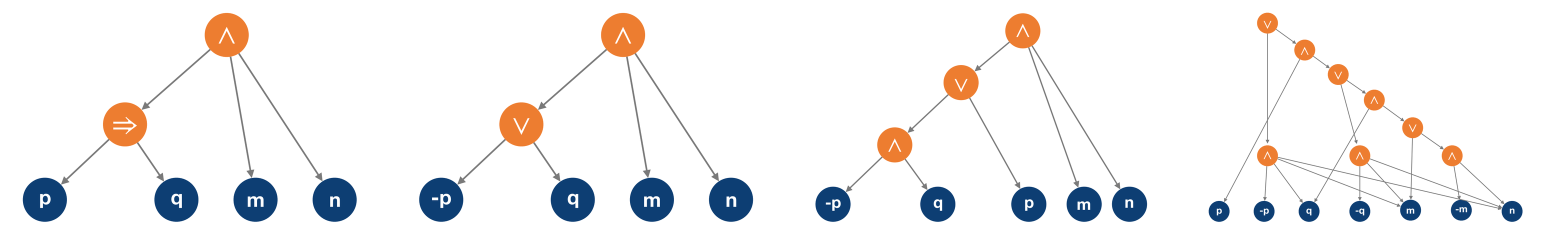
- We propose **L**ogic **E**mbedding **N**etwork with **S**emantic **R**egularization (**LENSR**), which embeds prior *symbolic knowledge* to enhance *deep models*.
- We compare the embedding quality of different logic representations: general form, CNF, and d-DNNF.
- Experiments show a connection between the tractability of a normal form and its amenability to embedding.
- LENSR is effective on an entailment task (synthetic dataset) and visual relationship prediction task (VRD Dataset).

Check out our code:
github.com/ZiweiXU/LENSR

Background

Notation

- **Propositions:** p (evaluation could be either True or False).
- **Formula:** F is a compound of propositions connected by logical connectives, $\neg, \wedge, \vee, \implies$.
- **Assignment:** τ is a function that maps propositions to their evaluations; There are satisfying assignments $\tau_{\mathsf{T}} \models$ F, and unsatisfying assignments s.t. $\tau_{\mathsf{F}} \models \neg$F.



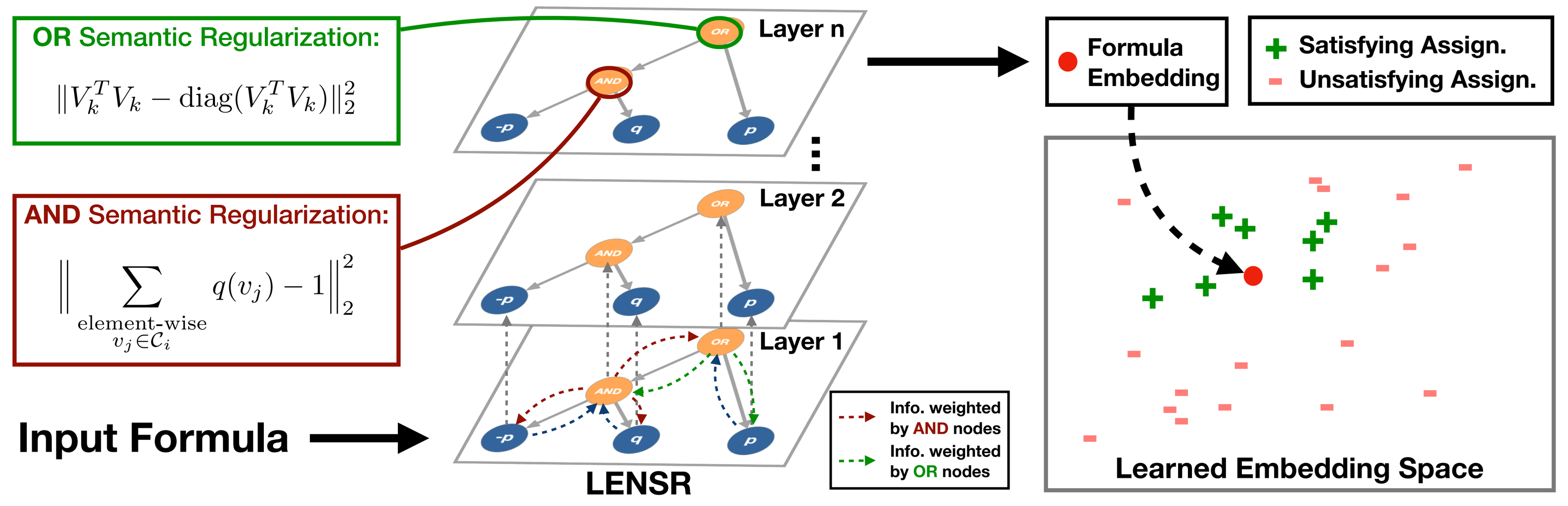
(a) General Form (b) CNF (c) d-DNNF (d) d-DNNF DAG

Logical forms

- **General form.**
- **Conjunctive Normal Form (CNF)**: a conjunction of clauses (a disjunction of literals).
- **Deterministic Decomposable Negation Normal Form (d-DNNF)**
  - Deterministic: operands of $\vee$ are mutually inconsistent.
  - Decomposable: operands of $\wedge$ are expressed on a mutually disjoint set of variables.

LENSR

Logic Formulae Embedder



- Augmented Graph Convolution Network (GCN):
  - **Heterogeneous Node:** type-dependent logical gate weights and attributes for each of the 5 node types (leaf, global, $\wedge$, $\vee$, $\Rightarrow$).
  - **Semantic Regularization:** regularizes the children embeddings of $\wedge$ gates to be orthogonal, the $\vee$ gate children embeddings to sum up to a unit vector:

$$L_r = \sum_{v_i \in \mathcal{N}_O} \left\| \sum_{\substack{\text{element-wise} \\ v_j \in \mathcal{C}_i}} q(v_j) - \mathbf{1} \right\|_2^2 + \sum_{v_k \in \mathcal{N}_A} \|V_k^T V_k - \mathrm{diag}(V_k^T V_k)\|_2^2$$

- **Train Embedder** $q(\mathsf{F}_x)$ with:

$$L_{\mathrm{emb}} = L_t + \lambda_r L_r$$

  - where L_t is triplet loss that encourages formulae embeddings to be close to satisfying assignments, and far from unsatisfying assignments:

$$L_t = \max\{d(q(\mathsf{F}), q(\tau_{\mathsf{F}})) - d(q(\mathsf{F}), q(\tau_{\mathsf{T}})) + m, 0\}$$

Target Task Training

- **Train Target Network** to minimize:

$$L = L_{\mathrm{c}} + \lambda L_{\mathrm{logic}}$$

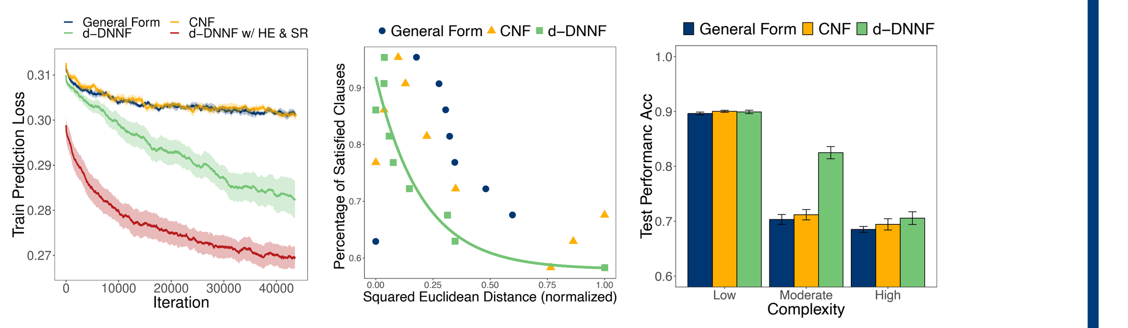
  - L_{c}: task specific loss (e.g., cross-entropy)
  - $L_{\mathrm{logic}} = \|q(\mathsf{F}_x) - q(\mathsf{G}_x)\|_2^2$
    - F_x: formula related to the input.
    - $\mathsf{G}_x = \wedge_i\, \mathsf{p}_i$: logic graph constructed from the predicted relationships p_i (in the predictive distribution $h(x)$).

Experiments

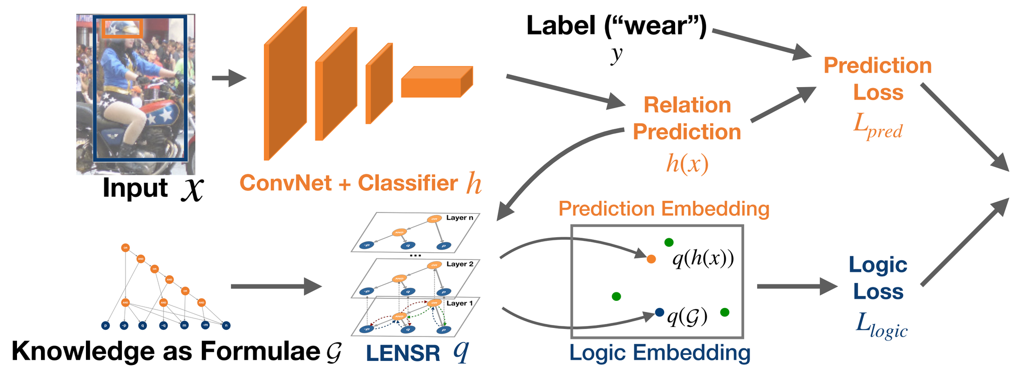
Synthetic

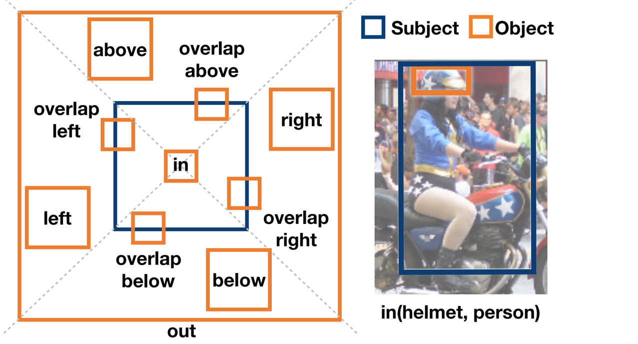
- **Complexity:** easy, medium, hard depending on num. variables and max depth

Formula Form	HE	SR	Acc.(%) Low	Acc.(%) Moderate	Acc.(%) High
General	-	-	89.63 (0.25)	70.32 (0.89)	68.51 (0.53)
CNF		-	90.02 (0.18)	71.19 (0.93)	69.42 (1.03)
	✓	-	90.25 (0.15)	73.92 (1.01)	68.79 (0.69)
d-DNNF			89.91 (0.31)	82.49 (1.11)	70.56 (1.16)
		✓	90.22 (0.23)	82.28 (1.40)	71.46 (1.17)
	✓		90.27 (0.55)	81.30 (1.29)	70.54 (0.62)
	✓	✓	**90.35** (0.32)	**83.04** (1.58)	**71.52** (0.54)



Visual Relation Prediction





- **Existence propositions**: e.g. proposition p=exist(person)
- **Visual relation proposition**: e.g. wear(person, glasses)
- **Spatial relation proposition**: e.g. in(glasses, person)
- **Existence constraint**: e.g. p(sub,obj) $\implies$ (exist(sub) $\wedge$ exist(obj))
- **Spatial constraint**: e.g. wear(person, glasses) $\implies$ in(glasses, person)

Model	Form	HE	SR	Top-5 Acc. (%)
without logic	-	-	-	84.30
with semantic loss [7]	-	-	-	84.76
with treeLSTM embedder [12]	CNF	-	-	85.76
	d-DNNF	-	-	82.99
LENSR	CNF		-	85.39
		✓	-	85.70
	d-DNNF			85.37
			✓	88.01
		✓		90.13
		✓	✓	**92.77**

Ground Truth:
<person - next to - truck>

Top-3 result of NN w/ LENSeR:
<person - next to - truck> ✓
<person - in the front of- truck>
<person - behind - truck>

Top-3 result of NN w/o LENSeR:
<person - skate on - truck>
<person - in - truck>
<person - next to - truck> ✓

Discussion

- Using symbolic knowledge via LENSR enhances deep networks performance in both synthetic and real world datasets.
- Concept of **embeddable-demanding** (Theorem 1; please see paper)
- Performance improvement due to embedding d-DNNF warrants:
  - a deeper examination of the relationship between tractability and embedding, and
  - the extension of semantic-aware embedding to alternative graph structures.

NeurIPS | 2019