Improving Out-of-Distribution Detection with Markov Logic Networks

Konstantin Kirchheim, Frank Ortmeier
firstname.lastname@ovgu.de
University of Magdeburg, Department of Computer Science

ICML International Conference On Machine Learning

Out-of-distribution (OOD) detection is essential for ensuring the reliability of deep learning models operating in open-world scenarios. Current OOD detectors mainly rely on statistical models to identify unusual patterns in the latent representations of a deep neural network. This work proposes to augment existing OOD detectors with probabilistic reasoning, utilizing Markov logic networks (MLNs). MLNs connect first-order logic with probabilistic reasoning to assign probabilities to inputs based on weighted logical constraints defined over human-understandable concepts, which offers improved explainability. Through extensive experiments on multiple datasets, we demonstrate that MLNs can significantly enhance the performance of a wide range of existing OOD detectors while maintaining computational efficiency. Furthermore, we introduce a simple algorithm for learning logical constraints for OOD detection from a dataset and showcase its effectiveness.

Background: OOD Detection with Logic

Idea

- Train DNNs to predict some concepts, formulate constraints on plausible concepts
- Inputs that violate a constraint are marked as OOD

Problem

- Strict logic too rigid for real-world applications where statistical associations dominate
- Rather, we want constraints that are violated often only have slight impact, etc.

Background: Markov Logic Network (MLN)

- Probabilistic Generalization of First-order Logic (FOL)
- Can be seen as Neuro-Symbolic or templates for large Markov Networks
- Each FOL formula φ_i is associated with a weight w_i
- For some input z, a MLN can predict:

$$P_{\mathcal{M}}(z) = \frac{1}{Z} \exp\left(\sum_i w_i \varphi_i(z)\right)$$

Standalone Markov Logic Network

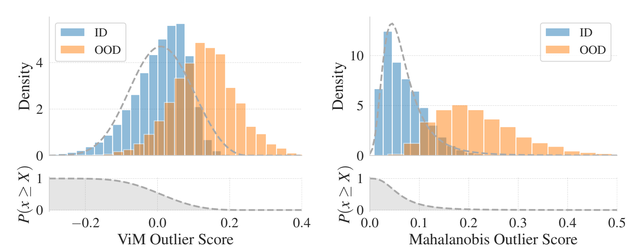
- Train DNNs to approximate interpretation of FOL predicates $\{\mathcal{P}_n\}_{n=1}^{N}$
- Create constraint set $\{\varphi_i\}_{i=1}^{N}$ with predicates
- Train MLN weights w_i by maximizing likelihood on ID training set
- Inference time outlier score:

$$D_{\mathcal{M}}(\mathbf{x}) = -\sum_i w_i \varphi_i(\mathbf{x})$$

- **Explainability**: we know exactly by what amount a violated rule changed the outlier score

Combination with other Detectors

- Normalizing outlier scores is necessary
- For detector $D : \mathcal{X} \to \mathbb{R}$, fit some distribution to outlier scores for ID data
- Then, use survival function p_D, which normalizes scores into the $[0, 1]$ range
- Outlier score $= p_D(\mathbf{x}) \times -\sum_i w_i \varphi_i(\mathbf{x})$



Constraint Search

- For some datasets, no prior knowledge available
- Idea: take dataset with ID and OOD examples and optimize set of rules

$$\max_{\varphi \in \mathscr{P}(\mathcal{T})} \underbrace{\mathbb{E}_{(\mathbf{x}_{\text{ID}}, \mathbf{x}_{\text{OOD}})} \left[J(\varphi, \mathbf{x}_{\text{ID}}, \mathbf{x}_{\text{OOD}}) \right]}_{\text{Performance}} - \lambda \underbrace{C(\varphi)}_{\text{Complexity}}$$

Input: Training set $\mathcal{D}_{\text{train}}$, validation set $\mathcal{D}_{\text{val}}$, baseline J_0, rule set $\mathcal{T}$
Output: Selected constraints φ
Initialize $\varphi \leftarrow \emptyset$
Initialize $J \leftarrow J_0$
for all $\varphi_i \in \mathcal{T}$ **do**
    $\varphi' \leftarrow \varphi \cup \{\varphi_i\}$
    Train detector with φ' on $\mathcal{D}_{\text{train}}$
    $J' \leftarrow$ Evaluate on $\mathcal{D}_{\text{val}}$
    if $J' > J + \delta_{\min}$ **then**
        $J \leftarrow J'$
        $\varphi \leftarrow \varphi'$
    end if
end for
return φ

Traffic Sign Classification (GTSRB)

- We have 43 constraints over the predicates: class, shape and color
- Statistically significant performance gains, reduces FPR95 by 37% (relative)
- Combination with MLN improves performance of all tested detectors

Face Attribute Prediction (CelebA)

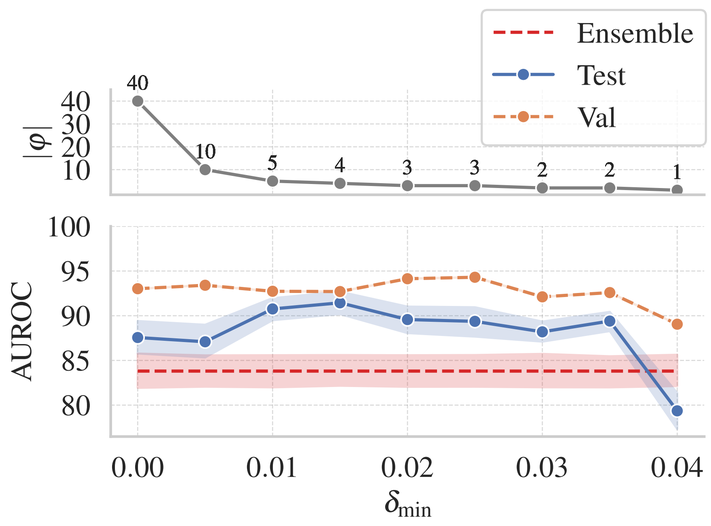
Constraint search on CelebA yields the following result:

$$\forall \mathbf{x} \quad \text{young}(\mathbf{x}) \tag{1}$$
$$\forall \mathbf{x} \quad \text{heavy\_makeup}(\mathbf{x}) \to \text{gray\_hair}(\mathbf{x}) \tag{2}$$
$$\forall \mathbf{x} \quad \text{wearing\_lipstick}(\mathbf{x}) \to \text{gray\_hair}(\mathbf{x}) \tag{3}$$
$$\forall \mathbf{x} \quad \text{wearing\_lipstick}(\mathbf{x}) \to \text{no\_beard}(\mathbf{x}) \tag{4}$$
$$\forall \mathbf{x} \quad \neg\text{male}(\mathbf{x}) \to \text{no\_beard}(\mathbf{x}) \tag{5}$$

- Since constraints are human-understandable, we can manually curate them
- Combination with MLN improves performance of all tested detectors
- e.g.: for MLN+Ensemble, FPR95 is reduced by 20% (relative)

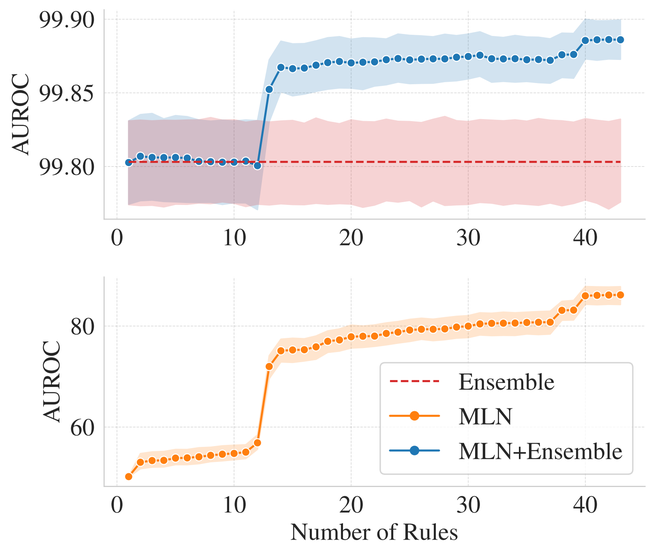
Constraint Search Regularization

- Regularizing constraint optimization improves results



Omiting Rules (GTSRB)

- Some rules are more important then others



Constraint Search Dataset

- Rules depend on dataset used for optimization
- Sufficient variability seems beneficial