Scenarios Interpretation with Prior Knowledge

Alessandro Daniele[1,2] *Advisor:* Luciano Serafini[1]

[1] Fondazione Bruno Kessler, Trento, Italy
{daniele,serafini}@fbk.eu
[2] University of Florence, Florence, Italy

Abstract. Statistical Relational Learning (SRL) deals with relational domains, where the samples are neither independent nor uniformly distributed. Moreover, central to SRL is the integration of logical knowledge in the learning framework. The main tasks in SRL are *Collective Classification*, *Entity Resolution*, *Link Prediction* and *Knowledge Graph Completion*. In this extended abstract we propose a new supervised learning task called *Scenarios Interpretation* (SI) where a sample is a *Scenario*, i.e. a set of (typically few) objects where each object and pair of objects have its own features. The goal is to classify objects and relationships. We propose NIoS (Neural Interpeter of Scenarios), a method for solving SI that is able to inject Prior Knowledge expressed in First Order Logic (FOL) into a neural network model. We implemented a first version and tested it on Visual Relationship Detection task (VRD) showing that NIoS outperformed state of the art systems.

1 Introduction

SRL focus on exploiting relationships between different entities and it is characterized by the presence of given constraints that are often expressed as a logical knowledge base. Generally in SRL tasks a graph or a subset of it is given and the focal point is finding a classification for the graph itself or for nodes or edges.

We define *Scenarios Interpretation* (SI), a new SRL task where a Scenario is given (i.e. a set of features for objects and pairs) and the aim is to find an Interpretation, i.e. a labeled directed graph with objects as nodes where labels represent classes and relations. On the best of our knowledge this is the first SRL task where the prediction is an entire graph. For instance, in *Collective Classification* [6] the aim is to find a classification for the nodes (with relations given) while in *Link Prediction* [4] the focus is on finding missing relations. SI can be seen as a generalization of both tasks, considering that we aim at finding both classes and relations at the same time.

SI could have many applications in different fields, like Image Processing, NLP, Bioinformatics. In this extended abstract we will focus on VRD (Visual Relationship Detection) [3, 5, 13] where an image can be seen as a Scenario with bounding boxes as objects[3]. We are interested in finding triplets of the form

[3] A similar task is the Scene Graph Generation [11], that can also be seen as a specific instance of SI

$\langle subj, rel, obj\rangle$. Together with visual data we have some prior knowledge expressed as logic formulas (e.g. $Wear(x, y) \rightarrow Person(x)$).

Among the SRL approaches that exploit logical knowledge, there are *Logic Tensor Network* (LTN) [8], *Semantic Based Regularization* (SBR) [2] and *Markov Logic Network* (MLN) [7]. We propose NIoS, a method for injecting FOL clauses inside a neural network model that can deal with the graph structure of scenarios. The main difference with its major competitors is on the way logic formulas are used: in NIoS they become part of the predictors instead of being used during training. In particular, methods like LTN or SBR force the constraint satisfaction during training making the assumption that the knowledge is in general correct. Instead, we assume there is a relationship between clauses and correct results, but this relationship is not known. The logical constraint are seen as a Prior Belief rather than Prior Knowledge. More in details, NIoS has internal learnable parameters associated to the logic formulas. In this extent, the most similar approaches to ours are probably (Hybrid) Markov Logic Networks [10, 7] and *Probabilistic Soft Logic* (PSL) [1] where the formulas weights are learned.

We tested NIoS for the Predicate Detection subtask of VRD on the *Visual Relationship Dataset* [5] (VRD Dataset) where we outperformed state of the art results, in particular on the *Zero Shot Learning* evaluation.

2 Scenarios Interpretation task

A scenario $\mathcal{S} \in \mathbb{S}$ is a triple composed of a set of objects $\mathcal{O}$ and two functions $\mathbf{u} : \mathcal{O} \rightarrow \mathbb{R}^k$ and $\mathbf{b} : \mathcal{O} \times \mathcal{O} \rightarrow \mathbb{R}^l$. An interpretation of a scenario $\mathcal{S}$ is a pair $\mathcal{I} = \langle l_o, l_r\rangle$

$$l_o : \mathcal{O} \times \mathcal{C} \rightarrow [0, 1] \qquad l_r : \mathcal{O} \times \mathcal{R} \times \mathcal{O} \rightarrow [0, 1]$$

where $\mathcal{C}$ and $\mathcal{R}$ are two disjoint sets of symbols for classes and relations respectively. The set of all interpretation is $\mathbb{I}$, the set of interpretations of a particular scenario $\mathcal{S}$ is $\mathbb{I}_{\mathcal{S}}$. A *constraint* is a clause in First Order Logic where unary predicates are in $\mathcal{C}$ and binary predicates are in $\mathcal{R}$.

Let $\mathcal{I}^* : \mathbb{S} \rightarrow \mathbb{I}$ be a function that returns a correct interpretation of a scenario ($\mathcal{I}^*(\mathcal{S}) \in \mathbb{I}_{\mathcal{S}}$). Given a training set composed of scenarios and corresponding correct interpretation $\left(\mathcal{S}^{(i)}, \mathcal{I}^*(\mathcal{S}^{(i)})\right)_{i=1}^{n}$ and a tuple of clauses $\mathcal{K} = \langle c_1, \ldots, c_m\rangle$ representing the Prior Knowledge, the SI task is the problem of finding a function that predicts correct interpretations of unseen scenarios. In particular, we want to find a function $\tilde{\mathcal{I}}_{\mathcal{K}}$ parametrized by weights of clauses in $\mathcal{K}$, that given a Scenario returns an interpretation that minimize the error in the training set:

$$\tilde{\mathcal{I}}_{\mathcal{K}} = \operatorname*{argmin}_{\mathcal{I}_{\mathcal{K}}} \sum_{i=1}^{n} \mathcal{L}(\mathcal{I}_{\mathcal{K}}(\mathcal{S}^{(i)}), \mathcal{I}^*(\mathcal{S}^{(i)})) \tag{1}$$

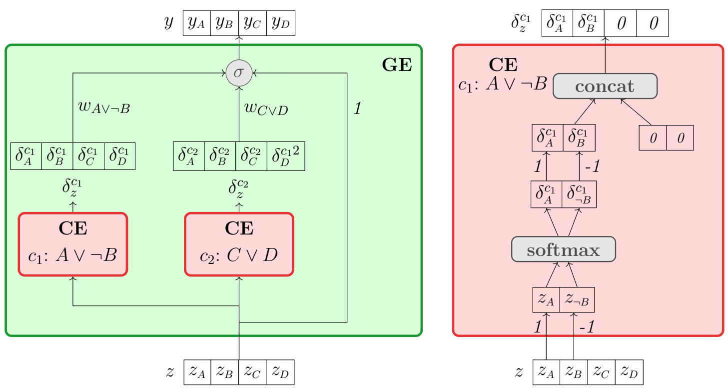
where $\mathcal{L} : \mathbb{I}_{\mathcal{S}} \times \mathbb{I}_{\mathcal{S}} \rightarrow \mathbb{R}^+$ is a function that returns a similarity score of two Interpretations. In our first implementation we used the L_2 loss function:

$$\mathcal{L}(\mathcal{I}^p, \mathcal{I}^t) = \sum_{\substack{o \in \mathcal{O} \\ c \in \mathcal{C}}} (l_o^p(o) - l_o^t(o))^2 + \sum_{\substack{(o_1, o_2) \in \mathcal{O}^2 \\ r \in \mathcal{R}}} (l_r^p(o_1, r, o_2) - l_r^t(o_1, r, o_2))^2$$

where $\mathcal{I}^t$ and $\mathcal{I}^p$ are the true and predicted interpretations.

3 NIoS: overview of the model

NIoS (Neural Interpreter of Scenarios) is a method for injecting logical knowledge into a Neural Network (NN). The original NN takes a Scenario as input and returns an initial Interpretation that is changed by a function, called *Global Enhancer* (GE), that modifies the initial predictions by enforcing the satisfaction of the logical constraints. The function must be differentiable and it can be seen as a new final layer for the original neural network. The entire network is still differentiable end-to-end, making it possible to train the model with back-propagation algorithm. GE contains additional parameters that can be learned as well. In particular *clause weights* determine the strength of each clause.



(a) (b)

Fig. 1: NIoS model: an example with four predicates and two clauses. Figure (a): *Global Enhancer*; Figure (b): *Clause Enhancer*.

Fig. 1(a) show GE implementation: it takes as input the preactivations z of the original neural network and produces the final activations y.

The GE contains a *Clause Enhancer* (CE) for each clause $c \in \mathcal{K}$ which returns δ_z^c, an adjustment to apply on preactivations to enforce c satisfaction. The CEs outputs are then combined linearly using *clauses weights* and summed to the initial preactivations. Lastly, the final predictions are calculated by applying the logistic function:

$$y = \sigma\Big(z + \sum_{c \in \mathcal{K}} w_c \cdot \delta_z^c\Big)$$

where w_c is the weight associated to clause c. Notice that setting w_c to zero make clause c irrelevant for the final predictions.

Fig. 1(b) shows current implementation of the CE. It is composed of a pre-elaboration step that select literals which appear in the clause (i.e. remove the absent predicates and change the sign of the negated ones). The next step applies the softmax function to the literals values. Intuitively, the idea is that, in order to satisfy a clause, at least one of its literal must be true. The softmax function act as a selector for the most promising true literal, that is the one with higher supporting evidences (biggest preactivation). Finally there is a post-elaboration step that works in reverse of the pre-elaboration (it sets the absent predicates adjustments to zero and change the sign of the negated ones).

4 Experimental evaluation

Visual Relationship Detection (VRD) is the task of finding objects in an image and capture their interactions [3, 5, 13]. It is composed of three subtask: Relationship Detection, Phrase Detection and Predicate Detection [5]. The VRD Dataset contains 100 classes for objects and 70 types of relations. It is composed of 4000 images for training and 1000 for testing with a total of 6672 triplets types. Among them 1877 can be find only in the Test Set and predicting them is the goal of the Zero Shot Learning variant of the task. For evaluating the results we used the *Recall*@n ($n \in \{50, 100\}$) metric proposed by Lu et al. [5] that is the percentage of times a correct relationship is found on the n predictions with highest score.

We evaluated NIoS on the Predicate Detection task using the knowledge base of [3]. We implemented NIoS using TensorFlow. As original NN we used a neural network with zero hidden layers and trained the entire network (original NN + GE) end-to-end using RMSProp [9]. Results are shown in Table 1.

	Standard L.		Zero Shot L.	
	R@50	R@100	R@50	R@100
Lu et al.[5]	47.87	47.87	8.45	8.45
LTN[3]	78.63	91.88	46.28	70.15
Yu et al.[12]	85.64	**94.65**	54.20	74.65
NIoS	**86.02**	91.91	**68.95**	**83.83**

Table 1: Results on VRD Predicate Detection task

NIoS outperformed other methods on all the metrics except for Recall@100 where it is surpassed by Yu et Al. [12]. The best results can be seen on the *Zero Shot Learning* task, where the difference between NIoS and the second best system is more than 10%. In *Zero Shot Learning* the aim is to predict previously unseen triplets, therefore it is rather difficult to learn to predict them from the Training Set. This confirms the ability of NIoS to use the Knowledge Base.

Another interesting result is the value obtained by NIoS compared to LTN[3]. In particular considering that the two works used the same Prior Knowledge. A possible explanation is given by the ability of NIoS to learn *clause weights*. Indeed, many weights results to be zero after learning. An example of a zero weighted clause is: $\neg Ride(x, y) \vee On(x, y)$.

Although the rule seems correct it is not in general satisfied on training and test set. This is because labels have been added manually, therefore there are plenty of missing relations. The hypothesis is that people have a tendency to add the most informative labels making some of the clauses unsatisfied.

5 Conclusions

We proposed SI, a new SRL task where the goal is to predict an entire graph, and we developed NIoS, a method for solving SI that can deal with learning in presence of a FOL Prior Knowledge. We reframed the VRD task as a SI instance and evaluated NIoS on it. With its results on VRD, NIoS showed to be competitive against other approaches, in particular tanks to its ability to effectively learn *clauses weights*.

References

1. Bach, S.H., Broecheler, M., Huang, B., Getoor, L.: Hinge-loss Markov random fields and probabilistic soft logic. Journal of Machine Learning Research **18**(109), 1–67 (2017)
2. Diligenti, M., Gori, M., Saccà, C.: Semantic-based regularization for learning and inference. Artif. Intell. **244**, 143–165 (2017)
3. Donadello, I.: Semantic Image Interpretation - Integration of Numerical Data and Logical Knowledge for Cognitive Vision. Ph.D. thesis, Trento Univ., Italy (2018)
4. Hasan, M.A., Chaoji, V., Salem, S., Zaki, M.: Link prediction using supervised learning. In: In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security (2006)
5. Lu, C., Krishna, R., Bernstein, M.S., Li, F.: Visual relationship detection with language priors. In: ECCV (1). Lecture Notes in Computer Science, vol. 9905, pp. 852–869. Springer (2016)
6. Pham, T., Tran, T., Phung, D.Q., Venkatesh, S.: Column networks for collective classification. In: AAAI. pp. 2485–2491. AAAI Press (2017)
7. Richardson, M., Domingos, P.: Markov logic networks. Mach. Learn. **62**(1-2), 107–136 (Feb 2006)
8. Serafini, L., d'Avila Garcez, A.S.: Logic tensor networks: Deep learning and logical reasoning from data and knowledge. CoRR **abs/1606.04422** (2016)
9. Tijmen, T., Hinton, G.: Lecture 6.5 − rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning (2012)
10. Wang, J., Domingos, P.: Hybrid markov logic networks. In: Proceedings of the 23rd National Conference on Artificial Intelligence. AAAI'08, vol. 2 (2008)
11. Xu, D., Zhu, Y., Choy, C.B., Fei-Fei, L.: Scene graph generation by iterative message passing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. vol. 2 (2017)
12. Yu, R., Li, A., Morariu, V.I., Davis, L.S.: Visual relationship detection with internal and external linguistic knowledge distillation. CoRR **abs/1707.09423** (2017)
13. Zhang, H., Kyaw, Z., Chang, S.F., Chua, T.S.: Visual translation embedding network for visual relation detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 3107–3115 (2017)