Cautious Rule-Based Collective Inference

Martin Svatoš

Department of Computer Science, Czech Technical University in Prague

svatoma1@fel.cvut.cz

Abstract

Collective inference is a popular approach for solving tasks as knowledge graph completion within the statistical relational learning field. There are many existing solutions for this task, however, each of them is subjected to some limitation, either by restriction to only some learning settings, lacking interpretability of the model or theoretical test error bounds. We propose an approach based on cautious inference process which uses first-order rules and provides PAC-style bounds.

1 Introduction

We are interested in the task of learning a model on a relational structure $\mathcal{A}$ capable of predicting missing facts on a relational structure $\mathcal{E}$. The relational structure may represent relations among people within a social network, interactions of proteins in a biological system, or relations between n-tuples and properties of single entities.

Let us first show a well-known example of Statistical Relational Learning (SRL) considering a network of people corresponding to their friendships and information whether they smoke or not. Using the imperfect rule that friend of a smoker is also a smoker can turn all the people in the network to be smokers or non-smokers. For example, the former may occur when using classical logic entailment, because one incorrect fact, i.e. a fact that is predicted as true but is false in $\mathcal{E}$, can lead to error propagation by deriving other incorrect facts based on this one. However, we can use bounded reasoning, e.g. in the form of k-entailment [Kuzelka *et al.*, 2018], which derives facts given size-k subsets of entities in a relational structure. In this way, potentially long proofs of false positively derived facts are broken. Besides that, bounded reasoning provides a PAC-style error bound for a set of rules on the test set $\mathcal{E}$ given its error on the training set $\mathcal{A}$ and number of entities in both the training and the test set.

To define our task, given a training data $\mathcal{A}$ we learn a theory composed of definite first-order rules Φ, e.g. $smoker(X) \wedge friends(X,Y) \implies smoker(Y)$, and a set of constraints Γ which hold in the data; these are expressed as universally quantified disjunctions, e.g. $\neg human(X) \vee \neg animal(X)$ meaning that no one can be a human and an animal at the same time. Using cautious inference mechanism and a theory $\Phi \cup \Gamma$ we want to predict missing facts in $\mathcal{E}$. The cautious inference mechanism is a part of our research. There are two main requirements within our research interest: i) the inference mechanism must provide some PAC-style bound, and ii) our learned model should be interpretable.

Our aim is to provide an SRL method consisting of both the theory learning part and a cautious inference mechanism. We have already published a method for the former, while the latter is under development.

2 Related Work

As discussed in the previous section, a straightforward application of learned imperfect first-order rules can lead to error propagation during the inference process. Bounding the derivation process to only one-step forward rule pass, as done in [Galárraga *et al.*, 2013], is too cautious and does not allow to use predicted facts for further predictions. Possibilistic logic uses ordered rules and removes them until it is consistent with the provided evidence [Lang *et al.*, 1991]. Both of these methods share interpretability, because there is a proof for each of the predicted facts, but lack error bounds. Markov Logic Networks (MLNs) [Richardson and Domingos, 2006] use weighted first-order rules to encode statistical regularities observed in training data. A popular inference method for MLNs is MAP inference, i.e. finding the most probable assignment to all variables given evidence. However, this inference possesses worse test-error bound than k-entailment [Kuzelka *et al.*, 2018]. From the computational point of view, MAP inference is NP-hard while k-entailment is polynomial in the number of entities and evidence size. Interpretability of MLNs is often hard because of complex transformations of rules weights.

Beside logic-based approaches discussed earlier, there are embeddings, e.g. ComplEx [Trouillon *et al.*, 2016], and neural network-based methods. From the latter group, Neural Theorem Provers [Rocktäschel and Riedel, 2017] learn to derive facts based on sub-symbolic representations. The vast majority of these methods are applicable only in the transductive setting since they cannot work on unseen entities. On the other hand, Deep Collective Inference [Moore and Neville, 2017] is not restricted to this setting, because it predicts labels of a graph's node based on the node's neighborhood. However, methods from this group lack interpretability because of underlying sub-symbolic vector representations. They are

also limited to at most binary relations, which makes them applicable only to graphs, i.e. predicting only properties of either one entity or relation between two of them. In contrast, logic-based approaches are not subjected to this limitation and allow n-tuple relations.

3 Progress

Our approach can be split into two parts: i) learning a theory (a set of rules), ii) designing suitable inference mechanism.

For learning first-order theories we introduced heuristically driven beam search with novel candidate rule scoring based on minimizing log-likelihood of the train data. We have also invented a pruning method [Svatoš *et al.*, 2017], using domain theory, which can exponentially prune the hypothesis space. However, this pruning method is not subjected to be used for our task, but, in fact, can be used in other rule-mining algorithms in relational learning, e.g. [Galárraga *et al.*, 2013].

Considering the inference mechanism part of our approach, we have followed the theoretical framework [Kuzelka *et al.*, 2018] which provides PAC-style bounds for a learned model. As the first step, we have developed an efficient algorithm for computing k-entailment, extended it into a stratified version and experimentally evaluated that both of them behave according to the theory, while they provide both the PAC-style bounds and interpretability of the models. However, the theory is suitable for a learning setting in which the training set $\mathcal{A}$ and the test set $\mathcal{E}$ do not share any entity. Therefore, our next goal is to invent a new inference mechanism for transductive setting where $\mathcal{A}$ actually equals $\mathcal{E}$. This, in turn, will allow us a direct comparison with embedding and neural network-based methods of which the majority is subjected to this setting.

4 Future Work

Our future work is focused on both directions towards theory learning and cautious inference. Generally, the main bottleneck of learning first-order rules is the size of the hypothesis space, which standard rule learning methods narrow by using language bias provided by an expert. Nonetheless, we do not have an expert for each dataset, therefore we need a refinement operator capable of learning which refinement operations are useless and should be avoided in order to lower computational time. Relational datasets often contain regularities like symmetries and mutexes. While the pruning method can handle symmetries quite well, it is not the case for mutexes. Naturally, our next step in this part is to develop a refinement operator which learns from the data and can handle these kinds of situations well. The final issue of this part is sorting learned rules, because of the order matters. So, finding the best possible order is another topic that we are interested in.

Considering the inference part, we want to exploit the high-level idea of bounded reasoning in transductive settings. The k-entailment, however, cannot be directly used in this setting, mainly because of the fast growth of the error bound w.r.t. the number of entities. Therefore our goal is to propose a new inference mechanism which suits this setting well. Under the assumptions of missing completely at random, it naturally stems to base the mechanism on literals instead of entities, which is in direct contrast to k-entailment. Thus, the new cautious inference should be stable w.r.t. flipping the truth value of a literal in $\mathcal{E}$. The most straightforward way is to give each literal one vote which, in turn, will be distributed among predicted facts. Finally, a predicted literal is as confident as the sum of partial votes gathered from evidence's literals. The key mechanism, which is yet to be investigated in our future work, is distributing of votes from literals in $\mathcal{E}$ to the predicted ones.

5 Contribution

Our main contribution is to present a new method capable of collective inference while preserving both PAC-style bounds and interpretability of learned models.

Acknowledgements

References

[Galárraga *et al.*, 2013] Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In *The 22nd International Conference on World Wide Web*, pages 413–422, 2013.

[Kuzelka *et al.*, 2018] Ondrej Kuzelka, Yuyi Wang, Jesse Davis, and Steven Schockaert. Pac-reasoning in relational domains. In *The Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 927–936, 2018.

[Lang *et al.*, 1991] Jérôme Lang, Didier Dubois, and Henri Prade. A logic of graded possibility and certainty coping with partial inconsistency. In *The Seventh conference on Uncertainty in Artificial Intelligence*, pages 188–196, 1991.

[Moore and Neville, 2017] John Moore and Jeninfer Neville. Deep collective inference. In *Thirty-First AAAI Conference on Artificial Intelligence*, pages 2364–2372, 2017.

[Richardson and Domingos, 2006] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine learning*, 62(1-2):107–136, 2006.

[Rocktäschel and Riedel, 2017] Tim Rocktäschel and Sebastian Riedel. End-to-end differentiable proving. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, pages 3791–3803, 2017.

[Svatoš *et al.*, 2017] Martin Svatoš, Gustav Šourek, Filip Železný, Steven Schockaert, and Ondřej Kuželka. Pruning hypothesis spaces using learned domain theories. In *International Conference on Inductive Logic Programming*, pages 152–168. Springer, 2017.

[Trouillon *et al.*, 2016] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080, 2016.