1. Introduction

The goal of my research is injecting knowledge/constraints into neural models, primarily for natural language processing (NLP) tasks. While neural models have set new state of the art performance in many tasks from vision to NLP, they often fail to learn simple rules necessary for well-formed structures unless there are an immense amount of training data. I propose that not all the aspects of the model have to be learned from the data itself and injecting simple knowledge/constraints into the neural models can help low-resource tasks as well as improving state-of-the-art (SOTA) models.

I focus on the structural constraints of the output space and inject knowledge of correct or preferred structures as an objective to the model without modification to the model structure in a model-agnostic way. [Lee et al., 2017, Lee et al.]. The first benefit in focusing on the constraint of output space is that it is *intuitive* as we can *directly enforce* outputs to satisfy logical/linguistic constraints. Another advantage of structural constraint is that it often *does not require* labeled dataset.

Focusing on deterministic constraints on the output values, I first applied output constraints on inference time via proposed gradient-based inference (GBI) method [Lee et al., 2017]. In the spirit of gradient-based training, GBI enforces constraints for each input at test-time by optimizing continuous model weights until the network's inference procedure generates an output that satisfies the constraints.

I extended the test time constraint injection to the training time: from instance-based optimization on inference time to generalization to multiple instances in training time. In training with structural constraints [Lee et al.], I have tried (1) structural constraint loss, (2) joint objective of structural loss and supervised loss on training set and lastly (3) joint objective on semi-supervised (SSL) setting. All the loss functions show improvements and the (3) SSL approach shows the largest improvement among them. The SSL improved the SOTA model by +0.5 F1 score and was particularly effective on low-resource (+1.58 F1). Additionally, the analysis shows that the efforts on training time and on inference time are complementary rather than mutually exclusive: the performance is best when efforts on train-time and inference-time methods are combined.

Lastly, I propose to extend the completed work to generalized span-based models and to domain adaptation where the target domain is unlabeled. Moreover, I aim to explore additional methodology that might bring bigger gains through constraint injection compared to current approaches.

2. Applications on Completed work

- *Semantic Role Labeling (SRL)* SRL jointly addresses two problems: chunking a sentence into spans and labeling each of the spans given the predicate. I use the output constraint that the spans of SRL outputs (set A) have to match with the spans of the parsing output (set B), i.e. $A \subset B$.

- *Syntactic constituency parsing* On constituency parsing, I focus on sequence-to-sequence models, where the model generates trees in linearized form with brackets. I use the simple constraint that the number of bracket opening and closing has to match in order to create a valid tree. This constraint resulted in significant improvements in structured prediction and domain adaptations in CoNLL2012 data.

3. Future work

- *Multi-task problems:* Many multi-task learning methods only share certain representations and simply apply separate (or summed) cross entropy as loss function. However, noting that there can exist a relation to the output space on a multi-task system. In such case, I propose to use agreements score as a regularizer similar to my completed work on single task with extrinsic knowledge.
- *Exploration of negative samples:* I want to know why even the SOTA networks fail in satisfying simple constraints as shown in my completed work. I hypothesize that the failure is due to the lack of introduction of negative samples, which refer to constraint-violating samples, in the training process. For the classification problem, it is well established that the training set should have balanced positive and negative examples in order to learn a good classifier from it. Can we sample constraint-violating instances systematically to inject constraints into the model?

References

*Jay Yoon Lee, *Sanket Vaibhav Mehta, and Jaime G. Carbonell. Towards semi-supervised learning for deep semantic role labeling. *EMNLP2018*. URL `http://arxiv.org/abs/1808.09543`.

Jay Yoon Lee, Michael Wick, Sanket Vaibhav Mehta, Jean-Baptiste Tristan, and Jaime Carbonell. Gradient-based inference for networks with output constraints. *arXiv preprint arXiv:1707.08608*, 2017.