



Jay Yoon Lee

Injecting constraints into neural NLP models

Friday, November 9, 2018 – 12:00 p.m. – GHC 6501

The goal of the thesis 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. The thesis proposes 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 models.

The thesis focuses on the structural knowledge of the output space and injects knowledge of correct or preferred structures as an objective to the model without modification to the model structure in a model-agnostic way. The first benefit in focusing on the knowledge of output space is that it is intuitive as we can directly enforce outputs to satisfy logical/linguistic constraints. Another advantage of structural knowledge is that it often does not require labeled dataset.

Focusing on deterministic constraints on the output values, the thesis first applies output constraints on inference time via proposed gradient-based inference (GBI) method. 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.

The thesis extends the inference-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, the thesis presents (1) structural constraint loss, (2) joint objective of structural loss and supervised loss on the training set and lastly (3) joint objective on a semi-supervised setting. All the loss functions show improvements and the (3) semi-supervised approach shows the largest improvement, particularly effective on the low-resource setting, among them. The analysis shows that the efforts on training time and on inference time are complementary rather than exclusive: the performance is best when efforts on train-time and inference-time methods are combined.

Lastly, the thesis proposes to extend the completed work to generalized span- based models and to domain adaptation where the target domain is unlabeled. Moreover, the thesis promises to explore additional methodology that might bring bigger gains through constraint injection compared to the currently proposed approaches.

Thesis Committee:
Jaime Carbonell, Chair
William Cohen
Graham Neubig
Dan Roth, University of Pennsylvania