Neural networks in recent years have achieved great breakthroughs in natural language processing. Though powerful, neural networks are statistically inefficient and require many labeled examples to train. One potential reason is that natural language has rich latent structure and a general purpose neural network module takes many examples to learn those complicated structures and generalize to new examples. In this thesis, we aim to improve the efficiency of neural networks by exploring structural properties of natural language in designing model architectures. We accomplish this by embedding prior knowledge of data structure into the model itself as a type of inductive bias.

In the first part, we find in a lot of supervised tasks on natural language—for example, visual question answering and document classification—the inputs have salient features that provide clues to the answers. The salient regions of inputs, however, is not directly annotated and cannot be directly leveraged for training. Moreover, the salient features must be reasoned and discovered according context in a step by step manner. By building a specific neural network module using iterative attention mechanism, we are able to localize the most important parts from inputs gradually and use them for prediction. The resulting systems not only achieved the state-of-the-art results, but also provided interpretations for their predictions.

In the second part, we find unsupervised learning models—including variational auto-encoders (VAEs) [54] and generative adversarial networks (GANs) [30] — designed for continuous inputs such as images do not perform well with natural languages as inputs. The main challenges lie in that the existing neural network modules in VAEs and GANs are not good at dealing with discrete and sequential inputs. To overcome the limitations, we designed network modules with input structure taken into consideration. Specifically, we proposed to use dilated CNNs as decoders for VAEs to control the contextual capacity. For GANs, we proposed to use more structured discriminators to replace the binary classifiers to provide better feedback to generators. The improved models achieved state-of-the-art results on text modeling and task of unsupervised text style transfer.

Thesis Committee:
Eric Xing, Co-chair
Taylor Berg-Kirkpatrick, Co-chair
Ruslan Salakhutdinov
Alexander Smola, Amazon
Li Deng, Citadel