Large-Scale QA-SRL Parsing

Nicholas FitzGerald* **Julian Michael*** **Luheng He** **Luke Zettlemoyer***
Paul G. Allen School, University of Washington, Seattle, WA
{nfitz,julianjm,luheng,lsz}@cs.washington.edu

A Supplemental Material

A.1 Experimental Setup

Hyperparameters The parameters of our LSTMs are initialized with random orthonormal matrices as described by Saxe et al. (2014). Input tokens are lower-cased, and the word vectors are pre-initialized with the 100-dimensional Glove embeddings trained on 6B tokens (Pennington et al., 2014) and fine-tuned during training. Tokens which are not covered by the Glove embeddings are assigned to the UNK vector. The embedding of the binary predicate indicator feature is also 100 dimensions. The text-encoder BiLSTM consists of 4 layers, uses a hidden size of 300 and . The output prediction feed-forward neural network for each model consists of a single 100 dimensional hidden layer with the non-rectified linear unit nonlinearity. For the sequential question generation model, each timestep consists of 4 layers of LSTMCells with a hidden size of 200.

Training All models are trained using Adadelta (Zeiler, 2012) with $\epsilon = 1e^{-6}$ and $\rho = 0.95$ and a mini-batch size of 80. The span encoding BiLSTM uses a recurrent dropout rate of 0.1, and we clip gradients with a norm greater than 1. All models were trained until performance on the development set did not improve for 10 epochs[1]. Our models were implemented in PyTorch[2] using the AllenNLP toolkit (Gardner et al.).

*Much of this work was done while the indicated authors were at the Allen Institute for Artificial Intelligence.

[1] All models completed training within 40 epochs, which took less than 4 hours on a single Titan X Pascal GPU.

[2] http://pytorch.org/

References

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. Allennlp: A deep semantic natural language processing platform.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *EMNLP 2014*.

Andrew M Saxe, James L McClelland, and Surya Ganguli. 2014. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *ICLR 2014*.

Matthew D Zeiler. 2012. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.