Core Semantic First: A Top-down Approach for AMR Parsing (Supplementary Material)

Deng Cai
The Chinese University of Hong Kong
thisisjcykcd@gmail.com

Wai Lam
The Chinese University of Hong Kong
wlam@se.cuhk.edu.hk

model component	hyper-parameter	value
char-level CNN	number of filters	256
	width of filters	3
	char embedding size	32
	final hidden size	128
Transformer	number of heads	8
	hidden state size	512
	feed-forward hidden size	1024
Sentence Encoder	Transformer layers	4
	lemma embedding size	200
	POS tag embedding size	32
	NER tag embedding size	16
Graph Encoder	Transformer layers	1
	concept embedding size	300
Focus Selection	attention layers	3
Relation Identification	number of heads	8
Relation Classification	hidden state size	100

Table 1: Hyper-parameters settings.

A Implementation Details

In all experiments, we use the same char-level CNN settings in the sentence encoder and the graph encoder. In addition, all Transformer (Vaswani et al., 2017) layers in our model share the same hyper-parameter settings. For computation efficiency, we only allow each concept to attend to its previously generated concepts in the graph encoder.[1] Table 1 summarizes the chosen hyper-parameters after we tuned on the development set. To mitigate overfitting, we also apply dropout (Srivastava et al., 2014) with the drop rate 0.2 between different layers. We use a special UNK token to replace the input lemmas, POS tags, and NER tags with a rate of 0.33. Parameter optimization is performed with the Adam optimizer (Kingma and Ba, 2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The same learning rate schedule of (Vaswani et al., 2017) is adopted in our experiments. We use early stopping on the development set for choosing the best model.

[1] Otherwise, we will need to re-compute the hidden states for all existing nodes at each parsing step.

Following Lyu and Titov (2018), for word sense disambiguation, we simply use the most frequent sense in the training set, or `-01` if not presented. For wikification, we look-up in the training set for the most frequent one and default to "-".

References

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Chunchuan Lyu and Ivan Titov. 2018. AMR parsing as graph prediction with latent alignment. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 397–407.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.