

Supplement to: A Fast Unified Model for Parsing and Sentence Understanding

Anonymous ACL submission

1 Hyperparameters

We use random search to tune the hyperparameters of the model, setting the ranges for search for each hyperparameter heuristically (and validating the reasonableness of the ranges on the development set), and then launching eight copies of each experiment each with newly sampled hyperparameters from those ranges. Table 1 (on the following page) shows the hyperparameters used in the best run of each model.

Param.	Range	Strategy	RNN	SPINN-PI-NT	SPINN-PI	SPINN
Initial LR	2e-4–2e-2	LOG	5e-3	3e-4	7e-3	2e-3
L2 regularization λ	8e-7–3e-5	LOG	4e-6	3e-6	2e-5	3e-5
Transition cost α	0.5–4.0	LIN	—	—	—	3.9
Embedding transformation dropout keep rate	80–95%	LIN	—	83%	92%	86%
Classifier MLP dropout keep rate	80–95%	LIN	94%	94%	93%	94%
Tracking LSTM size D_{tracking}	24–128	LOG	—	—	61	79
Classifier MLP layers	1–3	LIN	2	2	2	1

Table 1: Hyperparameter ranges and values. *Range* shows the hyperparameter ranges explored during random search. *Strategy* indicates whether sampling from the range was uniform, or log–uniform.