A Unified Linear-Time Framework for Sentence-Level Discourse Parsing (Supplementary)

Xiang Lin*¶, Shafiq Joty*¶§, Prathyusha Jwalapuram¶ and M Saiful Bari¶
¶Nanyang Technological University, Singapore
§Salesforce Research Asia, Singapore
{linx0057@e., srjoty@, jwal0001@e., bari0001@e.}ntu.edu.sg

A Supplementary Material

A.1 Segmentation Model Hyper-parameters

The optimal hyper-parameters for our segmenter are shown in table 1.

Hyper-parameters	Segmenter
Minibatch size	80
Embedding size	1024
Encoder hidden size	64
Decoder hidden size	64
ELMo dropout rate	0.5
Encoder dropout rate	0.2
Decoder dropout rate	0.2
Initial Learning rate	0.01
Adam β_1	0.9
Adam β_2	0.999
L_2 Regularization strength	0.0001

Table 1: Optimal hyper-parameter settings for segmenter

A.2 Parsing Model Hyper-parameters

The optimal hyper-parameters for our parser are shown in table 2.

A.3 Joint Training Model Hyper-parameters

The optimal hyper-parameters for our joint training model are shown in table 3.

Hyper-parameters	Parser
Minibatch size	64
Embedding size	1024
Encoder hidden size	64
Decoder hidden size	64
ELMo dropout rate	0.5
Encoder dropout rate	0.33
Decoder dropout rate	0.5
Classifier dropout rate	0.5
Initial Learning rate	0.001
Adam β_1 and β_2	0.9
L_2 Regularization strength	0.0005

Table 2: Optimal hyper-parameter settings for parser

Hyper-parameters	Joint Model
Minibatch size	80
Embedding size	1024
Encoder hidden size	64
Decoder hidden size	64
ELMo dropout rate	0.5
Encoder dropout rate	0.33
Seg. decoder dropout rate	0.5
Par. decoder dropout rate	0.5
Classifier dropout rate	0.5
Initial Learning rate	0.001
Adam β_1 and β_2	0.9
L_2 Regularization strength	0.0005

Table 3: Optimal hyper-parameter settings for joint training model

*equal contribution