

SAE hyper-parameter settings for ID-specific corpora	
Parameter	Value
Word embedding size	256
Encoder hidden layer size	256
Encoder number of layers	2, unidirectional
Decoder hidden layer size	256
Decoder number of layers	2, unidirectional
Decoder attention size	512
Number of data buckets	8
Dropout probability	0.5
Sampled softmax samples	25
L2 regularization beta	1e-5
Gradient norm upper bound	None
Optimizer	ADAM
Learning rate	0.0001
Learning rate annealing steps	5
Learning rate annealing factor	0.9
Scheduled sampling function	Inverse sigmoid
Scheduled sampling constant	17.48
Batch size	32
Warm-up epochs	10
Stagnant epochs before early stopping	30
Max. training epochs	200

Table 1: SAE hyper-parameter choices during pre-training phase.