

IDGAN hyper-parameter settings for high-ID corpus	
Parameter	Value
Word embedding size	256
Discriminator layer sizes	256, 128, 1
Number of data buckets	8
Dropout probability	0.5
Sampled softmax samples (reconstruction objective)	25
L2 regularization beta	1e-5
Gradient norm upper bound	None
Optimizer	ADAM
Discriminator label smoothing	True
Label flipping probability (GAN   WGAN-GP)	0.1   0.3
Generator learning rate (GAN   WGAN-GP)	0.0001   0.00005
Discriminator learning rate (GAN   WGAN-GP)	0.0001   0.00005
Learning rate annealing steps	5
Learning rate annealing factor	0.9
Scheduled sampling function	Inverse sigmoid
Scheduled sampling constant	10.6
Batch size	32
Warm-up epochs	5
Stagnant epochs before early stopping	20
Max. training epochs	100

Table 1: IDGAN hyper-parameter choices during the adversarial training phase.