

Description	Domain	Initial value	CSJ Generation 10
vocabulary size	$0 < x$	10000	25743
# of hidden layers	$0 < x$	2	1
# of layer units	$0 < x$	300	407
NNLM weight	$0 < x < 1$	0.500	0.546
acoustic weight	$0 < x$	14.000	14.525
minibatch size	$0 < x$	32	213
dropout ratio	$0 < x < 1$	0.500	0.302
initial learn rate	$0 < x < 1$	1.000	0.990
learn decay	$0 < x < 1$	0.500	0.398
learn decay epochs	$1 < x$	6	14
momentum	$0 < x < 1$	1.000E-10	7.268E-9
gradient clipping	$0 < x$	5.000	6.003
unit type	{LSTM, RNN, FF}	LSTM	LSTM
# of proj. layer units	$0 < x$	300	563
initial forget gate bias	$0 < x$	1.000	1.257
optimizer type	{SGD, ADAM, ADADELTA, rmsprop}	SGD	SGD
ADAM α	$0 < x$	1.000E-3	5.437E-4
ADAM ϵ	$0 < x$	1.000E-8	1.074E-8
ADAM λ	$x < 1$	$1 - 1.000E-8$	$1 - 6.000E-9$
rmsprop γ	$0 < x$	1.000E-4	1.405E-4
rmsprop learn rate	$0 < x < 1$	1.000E-4	7.940E-5
rmsprop δ	$x < 1$	0.950	0.941
rmsprop momentum	$0 < x < 1$	0.900	0.886
ADADELTA ϵ	$0 < x$	1.000E-3	7.490E-4
ADADELTA ρ	$x < 1$	0.950	0.917