

Description	Domain	Initial value	AMI Generation 12
vocabulary size	$0 < x$	10000	5587
# of hidden layers	$0 < x$	2	1
# of layer units	$0 < x$	300	476
NNLM weight	$0 < x < 1$	0.500	0.554
acoustic weight	$0 < x$	14.000	11.723
minibatch size	$0 < x$	32	52
dropout ratio	$0 < x < 1$	0.500	0.504
initial learn rate	$0 < x < 1$	1.000	0.981
learn decay	$0 < x < 1$	0.500	0.526
learn decay epochs	$1 < x$	6	13
momentum	$0 < x < 1$	1.000E-10	1.386E-10
gradient clipping	$0 < x$	5.000	11.441
unit type	{LSTM, RNN, FF}	LSTM	LSTM
# of proj. layer units	$0 < x$	300	399
initial forget gate bias	$0 < x$	1.000	0.450
optimizer type	{SGD, ADAM, ADADELTA, rmsprop}	SGD	SGD
ADAM α	$0 < x$	1.000E-3	1.364E-3
ADAM ϵ	$0 < x$	1.000E-8	1.246E-8
ADAM λ	$x < 1$	$1 - 1.000E-8$	$1 - 1.400E-8$
rmsprop γ	$0 < x$	1.000E-4	1.761E-4
rmsprop learn rate	$0 < x < 1$	1.000E-4	9.141E-5
rmsprop β	$x < 1$	0.950	0.950
rmsprop momentum	$0 < x < 1$	0.900	0.908
ADADELTA ϵ	$0 < x$	1.000E-3	9.730E-4
ADADELTA ρ	$x < 1$	0.950	0.958