

Model	Size	Valid	Test
<i>Existing results</i>			
Unregularzed LSTM	7M	120.7	114.5
NR-dropout	66M	82.2	78.4
Zoneout	66M	-	77.4
Variational LSTM	19M	-	73.4
CharCNN	21M	72.4	78.9
Pointer Sentinel-LSTM	51M	-	70.9
LSTM + continuous cache pointer	-	-	72.1
Variational LSTM + augmented loss	51M	71.1	68.5
Variational RHN	23M	67.9	65.4
NAS Cell	54M	-	62.4
4-layer skip connection LSTM	24M	60.9	58.3
AWD-LSTM w/o finetune	24M	60.7	58.8
AWD-LSTM (Baseline)	24M	60.0	57.3
<i>Our system</i>			
Sharpened Sigmoid AWD-LSTM w/o finetune	24M	61.6	59.4
Sharpened Sigmoid AWD-LSTM	24M	59.9	57.5
G^2 -LSTM w/o finetune	24M	60.4	58.2
G^2 -LSTM	24M	58.5	56.1
<i>+continuous cache pointer</i>			
AWD-LSTM + continuous cache pointer	24M	53.9	52.8
Sharpened Sigmoid AWD-LSTM + continuous cache pointer	24M	53.9	53.2
G^2 -LSTM + continuous cache pointer	24M	52.9	52.1