Test TCN

Herdiantri Sufriyana

I used Rstudio to run either python by R via reticulate package, or python directly. To run py file of TCN demo, I used bash in R studio; thus, I can show the process for reproducibility purpose.

I usually uses pytorch 1.2.0, torchvision 0.4.0, cudatoolkit 10.0, numpy 1.18.5, and PyQt5.

Below we just used torch for this demo, but I think the py file would use other as called by the file.

```
import torch
```

I cloned the TCN files from https://github.com/locuslab/TCN (https://github.com/locuslab/TCN).

Below is the content.

```
ls TCN/TCN/
```

```
## __init__.py
## __pycache__
## adding_problem
## char_cnn
## copy_memory
## lambada_language
## mnist_pixel
## poly_music
## tcn.py
## word_cnn
```

We would only use one of the example below.

```
ls TCN/TCN/word_cnn
```

```
## README.md
## __pycache__
## data
## model.pt
## model.py
## utils.py
## word_cnn_test.py
```

To print, the code would be re-run. Since it took about 7-8 hours to run, I turned off this code and just show you the code as an output below.

```
##
## cd TCN
## cd TCN
## cd word_cnn
## python word_cnn_test.py
##
```

After running the code,

we could see the results below.

```
##
## Namespace(batch_size=16, clip=0.35, corpus=False, cuda=True, data='./data/penn', dropout=0.45, emb_dropout=0.25, em
size=600, epochs=100, ksize=3, levels=4, log_interval=100, lr=4, nhid=600, optim='SGD', seed=1111, seq_len=80, tied=Tr
ue, validseqlen=40)
## Weight tied
## | epoch   1 |   100/ 1452 batches | lr 4.00000 | ms/batch 213.31373 | loss  7.54 | ppl  1880.00
## | epoch   1 |   200/ 1452 batches | lr 4.00000 | ms/batch 185.89694 | loss  6.81 | ppl   908.39
## | epoch   1 |   300/ 1452 batches | lr 4.00000 | ms/batch 187.20691 | loss  6.57 | ppl   711.84
## | epoch   1 |   400/ 1452 batches | lr 4.00000 | ms/batch 193.41650 | loss  6.35 | ppl   571.59
## | epoch   1 |   500/ 1452 batches | lr 4.00000 | ms/batch 194.45185 | loss  6.20 | ppl   493.92
## | epoch   1 |   600/ 1452 batches | lr 4.00000 | ms/batch 201.14858 | loss  6.19 | ppl   489.85
## | epoch   1 |   700/ 1452 batches | lr 4.00000 | ms/batch 201.56429 | loss  6.10 | ppl   445.51
## | epoch   1 |   800/ 1452 batches | lr 4.00000 | ms/batch 199.86704 | loss  5.99 | ppl   399.70
## | epoch   1 |   900/ 1452 batches | lr 4.00000 | ms/batch 202.62938 | loss  5.96 | ppl   388.77
## | epoch   1 |  1000/ 1452 batches | lr 4.00000 | ms/batch 201.62596 | loss  5.90 | ppl   366.33
## | epoch   1 |  1100/ 1452 batches | lr 4.00000 | ms/batch 205.04670 | loss  5.87 | ppl   354.88
## | epoch   1 |  1200/ 1452 batches | lr 4.00000 | ms/batch 206.55519 | loss  5.86 | ppl   352.43
## | epoch   1 |  1300/ 1452 batches | lr 4.00000 | ms/batch 207.36968 | loss  5.74 | ppl   310.85
## | epoch   1 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.18492 | loss  5.73 | ppl   306.89
## -----------------------------------------------------------------------------------------
## | end of epoch   1 | time: 310.02s | valid loss  5.64 | valid ppl   281.32
## | end of epoch   1 | time: 310.02s | test loss  5.61 | test ppl   271.97
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch   2 |   100/ 1452 batches | lr 4.00000 | ms/batch 207.76594 | loss  5.79 | ppl   327.40
## | epoch   2 |   200/ 1452 batches | lr 4.00000 | ms/batch 210.23370 | loss  5.67 | ppl   291.22
## | epoch   2 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.12593 | loss  5.64 | ppl   282.45
## | epoch   2 |   400/ 1452 batches | lr 4.00000 | ms/batch 213.22784 | loss  5.53 | ppl   251.53
## | epoch   2 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.70381 | loss  5.50 | ppl   245.64
## | epoch   2 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.55480 | loss  5.56 | ppl   259.67
## | epoch   2 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.38481 | loss  5.53 | ppl   251.57
## | epoch   2 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.42319 | loss  5.48 | ppl   240.35
## | epoch   2 |   900/ 1452 batches | lr 4.00000 | ms/batch 209.18406 | loss  5.48 | ppl   239.94
## | epoch   2 |  1000/ 1452 batches | lr 4.00000 | ms/batch 216.88229 | loss  5.46 | ppl   235.08
## | epoch   2 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.68234 | loss  5.48 | ppl   239.75
## | epoch   2 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.54750 | loss  5.49 | ppl   241.81
## | epoch   2 |  1300/ 1452 batches | lr 4.00000 | ms/batch 206.93573 | loss  5.36 | ppl   211.72
## | epoch   2 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.54007 | loss  5.39 | ppl   220.17
## -----------------------------------------------------------------------------------------
## | end of epoch   2 | time: 322.26s | valid loss  5.35 | valid ppl   211.36
## | end of epoch   2 | time: 322.26s | test loss  5.31 | test ppl   202.18
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch   3 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.69461 | loss  5.49 | ppl   241.29
## | epoch   3 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.64407 | loss  5.39 | ppl   220.07
## | epoch   3 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.65640 | loss  5.38 | ppl   217.21
## | epoch   3 |   400/ 1452 batches | lr 4.00000 | ms/batch 210.45569 | loss  5.26 | ppl   192.11
## | epoch   3 |   500/ 1452 batches | lr 4.00000 | ms/batch 211.68313 | loss  5.25 | ppl   190.90
## | epoch   3 |   600/ 1452 batches | lr 4.00000 | ms/batch 209.61010 | loss  5.32 | ppl   203.79
## | epoch   3 |   700/ 1452 batches | lr 4.00000 | ms/batch 209.69868 | loss  5.30 | ppl   199.83
## | epoch   3 |   800/ 1452 batches | lr 4.00000 | ms/batch 210.06772 | loss  5.27 | ppl   194.21
## | epoch   3 |   900/ 1452 batches | lr 4.00000 | ms/batch 209.59599 | loss  5.27 | ppl   193.74
## | epoch   3 |  1000/ 1452 batches | lr 4.00000 | ms/batch 210.64381 | loss  5.25 | ppl   190.78
## | epoch   3 |  1100/ 1452 batches | lr 4.00000 | ms/batch 211.45980 | loss  5.29 | ppl   198.44
## | epoch   3 |  1200/ 1452 batches | lr 4.00000 | ms/batch 211.16803 | loss  5.30 | ppl   199.64
## | epoch   3 |  1300/ 1452 batches | lr 4.00000 | ms/batch 210.12538 | loss  5.16 | ppl   174.07
## | epoch   3 |  1400/ 1452 batches | lr 4.00000 | ms/batch 210.72939 | loss  5.22 | ppl   185.46
## -----------------------------------------------------------------------------------------
## | end of epoch   3 | time: 324.30s | valid loss  5.20 | valid ppl   181.95
## | end of epoch   3 | time: 324.30s | test loss  5.16 | test ppl   174.06
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch   4 |   100/ 1452 batches | lr 4.00000 | ms/batch 211.13621 | loss  5.32 | ppl   204.35
## | epoch   4 |   200/ 1452 batches | lr 4.00000 | ms/batch 217.09299 | loss  5.23 | ppl   187.45
## | epoch   4 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.35256 | loss  5.23 | ppl   186.81
## | epoch   4 |   400/ 1452 batches | lr 4.00000 | ms/batch 209.98574 | loss  5.10 | ppl   164.41
## | epoch   4 |   500/ 1452 batches | lr 4.00000 | ms/batch 213.11988 | loss  5.10 | ppl   164.25
## | epoch   4 |   600/ 1452 batches | lr 4.00000 | ms/batch 210.56105 | loss  5.17 | ppl   175.58
## | epoch   4 |   700/ 1452 batches | lr 4.00000 | ms/batch 210.48958 | loss  5.16 | ppl   174.71
```

```
## | epoch   4 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.39828 | loss  5.12 | ppl   168.15
## | epoch   4 |   900/ 1452 batches | lr 4.00000 | ms/batch 215.68474 | loss  5.13 | ppl   169.14
## | epoch   4 |  1000/ 1452 batches | lr 4.00000 | ms/batch 205.74007 | loss  5.11 | ppl   165.46
## | epoch   4 |  1100/ 1452 batches | lr 4.00000 | ms/batch 206.78384 | loss  5.17 | ppl   175.86
## | epoch   4 |  1200/ 1452 batches | lr 4.00000 | ms/batch 209.35117 | loss  5.16 | ppl   174.55
## | epoch   4 |  1300/ 1452 batches | lr 4.00000 | ms/batch 206.65563 | loss  5.01 | ppl   150.31
## | epoch   4 |  1400/ 1452 batches | lr 4.00000 | ms/batch 210.99112 | loss  5.10 | ppl   164.01
## -----------------------------------------------------------------------------------------
## | end of epoch   4 | time: 324.81s | valid loss  5.10 | valid ppl   163.73
## | end of epoch   4 | time: 324.81s | test loss  5.05 | test ppl   156.80
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch   5 |   100/ 1452 batches | lr 4.00000 | ms/batch 215.17044 | loss  5.20 | ppl   180.40
## | epoch   5 |   200/ 1452 batches | lr 4.00000 | ms/batch 210.40326 | loss  5.12 | ppl   167.76
## | epoch   5 |   300/ 1452 batches | lr 4.00000 | ms/batch 215.68785 | loss  5.12 | ppl   166.66
## | epoch   5 |   400/ 1452 batches | lr 4.00000 | ms/batch 211.18521 | loss  4.98 | ppl   145.53
## | epoch   5 |   500/ 1452 batches | lr 4.00000 | ms/batch 211.11784 | loss  4.99 | ppl   146.95
## | epoch   5 |   600/ 1452 batches | lr 4.00000 | ms/batch 215.19255 | loss  5.05 | ppl   156.60
## | epoch   5 |   700/ 1452 batches | lr 4.00000 | ms/batch 211.37289 | loss  5.05 | ppl   156.10
## | epoch   5 |   800/ 1452 batches | lr 4.00000 | ms/batch 212.21538 | loss  5.01 | ppl   150.38
## | epoch   5 |   900/ 1452 batches | lr 4.00000 | ms/batch 217.82060 | loss  5.02 | ppl   151.67
## | epoch   5 |  1000/ 1452 batches | lr 4.00000 | ms/batch 211.14945 | loss  5.00 | ppl   148.59
## | epoch   5 |  1100/ 1452 batches | lr 4.00000 | ms/batch 213.39612 | loss  5.07 | ppl   159.76
## | epoch   5 |  1200/ 1452 batches | lr 4.00000 | ms/batch 211.42601 | loss  5.07 | ppl   158.42
## | epoch   5 |  1300/ 1452 batches | lr 4.00000 | ms/batch 211.07869 | loss  4.91 | ppl   135.29
## | epoch   5 |  1400/ 1452 batches | lr 4.00000 | ms/batch 212.93417 | loss  5.01 | ppl   149.98
## -----------------------------------------------------------------------------------------
## | end of epoch   5 | time: 328.01s | valid loss  5.03 | valid ppl   153.01
## | end of epoch   5 | time: 328.01s | test loss  4.99 | test ppl   146.40
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch   6 |   100/ 1452 batches | lr 4.00000 | ms/batch 212.65275 | loss  5.10 | ppl   163.83
## | epoch   6 |   200/ 1452 batches | lr 4.00000 | ms/batch 220.03440 | loss  5.03 | ppl   152.74
## | epoch   6 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.45225 | loss  5.04 | ppl   153.92
## | epoch   6 |   400/ 1452 batches | lr 4.00000 | ms/batch 216.58135 | loss  4.89 | ppl   132.32
## | epoch   6 |   500/ 1452 batches | lr 4.00000 | ms/batch 210.33716 | loss  4.90 | ppl   134.32
## | epoch   6 |   600/ 1452 batches | lr 4.00000 | ms/batch 211.37767 | loss  4.96 | ppl   143.30
## | epoch   6 |   700/ 1452 batches | lr 4.00000 | ms/batch 211.65438 | loss  4.96 | ppl   142.98
## | epoch   6 |   800/ 1452 batches | lr 4.00000 | ms/batch 213.01646 | loss  4.93 | ppl   138.13
## | epoch   6 |   900/ 1452 batches | lr 4.00000 | ms/batch 214.75693 | loss  4.93 | ppl   139.00
## | epoch   6 |  1000/ 1452 batches | lr 4.00000 | ms/batch 211.87914 | loss  4.91 | ppl   135.72
## | epoch   6 |  1100/ 1452 batches | lr 4.00000 | ms/batch 211.57452 | loss  5.00 | ppl   148.76
## | epoch   6 |  1200/ 1452 batches | lr 4.00000 | ms/batch 213.16926 | loss  4.98 | ppl   146.17
## | epoch   6 |  1300/ 1452 batches | lr 4.00000 | ms/batch 223.06965 | loss  4.82 | ppl   123.40
## | epoch   6 |  1400/ 1452 batches | lr 4.00000 | ms/batch 209.68592 | loss  4.94 | ppl   139.85
## -----------------------------------------------------------------------------------------
## | end of epoch   6 | time: 328.79s | valid loss  4.96 | valid ppl   143.14
## | end of epoch   6 | time: 328.79s | test loss  4.92 | test ppl   136.89
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch   7 |   100/ 1452 batches | lr 4.00000 | ms/batch 212.58477 | loss  5.02 | ppl   151.21
## | epoch   7 |   200/ 1452 batches | lr 4.00000 | ms/batch 212.03506 | loss  4.96 | ppl   142.22
## | epoch   7 |   300/ 1452 batches | lr 4.00000 | ms/batch 213.88812 | loss  4.96 | ppl   142.65
## | epoch   7 |   400/ 1452 batches | lr 4.00000 | ms/batch 213.57654 | loss  4.81 | ppl   122.43
## | epoch   7 |   500/ 1452 batches | lr 4.00000 | ms/batch 210.71936 | loss  4.82 | ppl   124.38
## | epoch   7 |   600/ 1452 batches | lr 4.00000 | ms/batch 212.56867 | loss  4.89 | ppl   133.42
## | epoch   7 |   700/ 1452 batches | lr 4.00000 | ms/batch 210.66742 | loss  4.89 | ppl   133.14
## | epoch   7 |   800/ 1452 batches | lr 4.00000 | ms/batch 211.27162 | loss  4.86 | ppl   128.73
## | epoch   7 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.63145 | loss  4.86 | ppl   129.31
## | epoch   7 |  1000/ 1452 batches | lr 4.00000 | ms/batch 236.13865 | loss  4.84 | ppl   126.00
## | epoch   7 |  1100/ 1452 batches | lr 4.00000 | ms/batch 211.52704 | loss  4.93 | ppl   138.91
## | epoch   7 |  1200/ 1452 batches | lr 4.00000 | ms/batch 212.91221 | loss  4.92 | ppl   136.74
## | epoch   7 |  1300/ 1452 batches | lr 4.00000 | ms/batch 210.75504 | loss  4.74 | ppl   114.50
## | epoch   7 |  1400/ 1452 batches | lr 4.00000 | ms/batch 214.39593 | loss  4.87 | ppl   130.40
## -----------------------------------------------------------------------------------------
## | end of epoch   7 | time: 329.47s | valid loss  4.91 | valid ppl   136.23
## | end of epoch   7 | time: 329.47s | test loss  4.87 | test ppl   130.78
## -----------------------------------------------------------------------------------------
## Save model!
```

```
##
## | epoch   8 |   100/ 1452 batches | lr 4.00000 | ms/batch 212.64537 | loss  4.95 | ppl   141.58
## | epoch   8 |   200/ 1452 batches | lr 4.00000 | ms/batch 213.14901 | loss  4.88 | ppl   132.08
## | epoch   8 |   300/ 1452 batches | lr 4.00000 | ms/batch 213.50942 | loss  4.89 | ppl   132.95
## | epoch   8 |   400/ 1452 batches | lr 4.00000 | ms/batch 211.24087 | loss  4.75 | ppl   115.08
## | epoch   8 |   500/ 1452 batches | lr 4.00000 | ms/batch 216.52609 | loss  4.77 | ppl   117.42
## | epoch   8 |   600/ 1452 batches | lr 4.00000 | ms/batch 211.71264 | loss  4.83 | ppl   125.14
## | epoch   8 |   700/ 1452 batches | lr 4.00000 | ms/batch 211.77630 | loss  4.83 | ppl   125.05
## | epoch   8 |   800/ 1452 batches | lr 4.00000 | ms/batch 211.77871 | loss  4.79 | ppl   120.46
## | epoch   8 |   900/ 1452 batches | lr 4.00000 | ms/batch 212.35066 | loss  4.80 | ppl   121.52
## | epoch   8 |  1000/ 1452 batches | lr 4.00000 | ms/batch 212.10021 | loss  4.78 | ppl   118.68
## | epoch   8 |  1100/ 1452 batches | lr 4.00000 | ms/batch 212.47463 | loss  4.88 | ppl   131.28
## | epoch   8 |  1200/ 1452 batches | lr 4.00000 | ms/batch 220.87684 | loss  4.86 | ppl   128.93
## | epoch   8 |  1300/ 1452 batches | lr 4.00000 | ms/batch 210.58892 | loss  4.68 | ppl   108.04
## | epoch   8 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.59220 | loss  4.81 | ppl   123.04
## -----------------------------------------------------------------------------------------
## | end of epoch   8 | time: 326.97s | valid loss  4.89 | valid ppl   133.05
## | end of epoch   8 | time: 326.97s | test loss  4.84 | test ppl   126.86
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch   9 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.82642 | loss  4.90 | ppl   134.08
## | epoch   9 |   200/ 1452 batches | lr 4.00000 | ms/batch 206.86843 | loss  4.83 | ppl   125.69
## | epoch   9 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.39298 | loss  4.83 | ppl   125.56
## | epoch   9 |   400/ 1452 batches | lr 4.00000 | ms/batch 214.39891 | loss  4.69 | ppl   108.74
## | epoch   9 |   500/ 1452 batches | lr 4.00000 | ms/batch 212.13166 | loss  4.70 | ppl   110.10
## | epoch   9 |   600/ 1452 batches | lr 4.00000 | ms/batch 217.61060 | loss  4.77 | ppl   117.63
## | epoch   9 |   700/ 1452 batches | lr 4.00000 | ms/batch 206.99830 | loss  4.78 | ppl   118.76
## | epoch   9 |   800/ 1452 batches | lr 4.00000 | ms/batch 210.13510 | loss  4.74 | ppl   114.36
## | epoch   9 |   900/ 1452 batches | lr 4.00000 | ms/batch 206.78366 | loss  4.75 | ppl   115.26
## | epoch   9 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.56088 | loss  4.71 | ppl   111.55
## | epoch   9 |  1100/ 1452 batches | lr 4.00000 | ms/batch 216.22261 | loss  4.83 | ppl   124.65
## | epoch   9 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.90635 | loss  4.80 | ppl   121.12
## | epoch   9 |  1300/ 1452 batches | lr 4.00000 | ms/batch 206.73856 | loss  4.62 | ppl   101.46
## | epoch   9 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.45782 | loss  4.76 | ppl   116.66
## -----------------------------------------------------------------------------------------
## | end of epoch   9 | time: 322.94s | valid loss  4.84 | valid ppl   127.10
## | end of epoch   9 | time: 322.94s | test loss  4.80 | test ppl   121.46
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  10 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.74550 | loss  4.84 | ppl   127.00
## | epoch  10 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.65660 | loss  4.78 | ppl   119.41
## | epoch  10 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.23805 | loss  4.79 | ppl   120.06
## | epoch  10 |   400/ 1452 batches | lr 4.00000 | ms/batch 207.18810 | loss  4.63 | ppl   102.94
## | epoch  10 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.05722 | loss  4.65 | ppl   104.94
## | epoch  10 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.01829 | loss  4.72 | ppl   112.27
## | epoch  10 |   700/ 1452 batches | lr 4.00000 | ms/batch 210.36484 | loss  4.72 | ppl   111.92
## | epoch  10 |   800/ 1452 batches | lr 4.00000 | ms/batch 209.65066 | loss  4.69 | ppl   108.96
## | epoch  10 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.54467 | loss  4.69 | ppl   109.01
## | epoch  10 |  1000/ 1452 batches | lr 4.00000 | ms/batch 216.45369 | loss  4.67 | ppl   106.24
## | epoch  10 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.78155 | loss  4.78 | ppl   119.10
## | epoch  10 |  1200/ 1452 batches | lr 4.00000 | ms/batch 209.49575 | loss  4.76 | ppl   116.17
## | epoch  10 |  1300/ 1452 batches | lr 4.00000 | ms/batch 208.23702 | loss  4.57 | ppl    96.46
## | epoch  10 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.35204 | loss  4.72 | ppl   112.00
## -----------------------------------------------------------------------------------------
## | end of epoch  10 | time: 322.16s | valid loss  4.81 | valid ppl   123.29
## | end of epoch  10 | time: 322.16s | test loss  4.77 | test ppl   118.11
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  11 |   100/ 1452 batches | lr 4.00000 | ms/batch 207.57771 | loss  4.80 | ppl   121.13
## | epoch  11 |   200/ 1452 batches | lr 4.00000 | ms/batch 212.97724 | loss  4.73 | ppl   113.48
## | epoch  11 |   300/ 1452 batches | lr 4.00000 | ms/batch 206.64866 | loss  4.73 | ppl   113.84
## | epoch  11 |   400/ 1452 batches | lr 4.00000 | ms/batch 211.27393 | loss  4.59 | ppl    98.61
## | epoch  11 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.78749 | loss  4.61 | ppl   100.49
## | epoch  11 |   600/ 1452 batches | lr 4.00000 | ms/batch 208.01726 | loss  4.68 | ppl   107.60
## | epoch  11 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.55673 | loss  4.68 | ppl   107.75
## | epoch  11 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.70652 | loss  4.65 | ppl   104.38
## | epoch  11 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.06330 | loss  4.64 | ppl   104.01
## | epoch  11 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.71654 | loss  4.62 | ppl   101.32
## | epoch  11 |  1100/ 1452 batches | lr 4.00000 | ms/batch 214.21094 | loss  4.74 | ppl   114.00
```

```
## | epoch  11 |  1200/ 1452 batches | lr 4.00000 | ms/batch 206.33896 | loss  4.71 | ppl   111.22
## | epoch  11 |  1300/ 1452 batches | lr 4.00000 | ms/batch 208.23703 | loss  4.52 | ppl    91.91
## | epoch  11 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.05722 | loss  4.68 | ppl   107.34
## -----------------------------------------------------------------------------------------
## | end of epoch  11 | time: 321.52s | valid loss  4.79 | valid ppl   120.80
## | end of epoch  11 | time: 321.52s | test loss  4.75 | test ppl   115.45
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  12 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.68554 | loss  4.75 | ppl   115.94
## | epoch  12 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.76650 | loss  4.70 | ppl   109.49
## | epoch  12 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.97627 | loss  4.70 | ppl   109.44
## | epoch  12 |   400/ 1452 batches | lr 4.00000 | ms/batch 211.74345 | loss  4.55 | ppl    94.48
## | epoch  12 |   500/ 1452 batches | lr 4.00000 | ms/batch 209.60564 | loss  4.57 | ppl    96.57
## | epoch  12 |   600/ 1452 batches | lr 4.00000 | ms/batch 225.52935 | loss  4.63 | ppl   102.62
## | epoch  12 |   700/ 1452 batches | lr 4.00000 | ms/batch 205.80951 | loss  4.64 | ppl   103.13
## | epoch  12 |   800/ 1452 batches | lr 4.00000 | ms/batch 206.48392 | loss  4.61 | ppl   100.23
## | epoch  12 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.14712 | loss  4.60 | ppl    99.81
## | epoch  12 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.50296 | loss  4.58 | ppl    97.59
## | epoch  12 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.02741 | loss  4.70 | ppl   109.59
## | epoch  12 |  1200/ 1452 batches | lr 4.00000 | ms/batch 209.90043 | loss  4.66 | ppl   106.14
## | epoch  12 |  1300/ 1452 batches | lr 4.00000 | ms/batch 207.83743 | loss  4.48 | ppl    88.42
## | epoch  12 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.25207 | loss  4.64 | ppl   103.05
## -----------------------------------------------------------------------------------------
## | end of epoch  12 | time: 322.91s | valid loss  4.77 | valid ppl   118.02
## | end of epoch  12 | time: 322.91s | test loss  4.72 | test ppl   112.44
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  13 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.78544 | loss  4.71 | ppl   111.20
## | epoch  13 |   200/ 1452 batches | lr 4.00000 | ms/batch 211.13408 | loss  4.65 | ppl   105.02
## | epoch  13 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.94733 | loss  4.66 | ppl   105.15
## | epoch  13 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.41193 | loss  4.51 | ppl    90.55
## | epoch  13 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.47195 | loss  4.53 | ppl    92.58
## | epoch  13 |   600/ 1452 batches | lr 4.00000 | ms/batch 208.10718 | loss  4.60 | ppl    99.35
## | epoch  13 |   700/ 1452 batches | lr 4.00000 | ms/batch 209.23599 | loss  4.60 | ppl    99.14
## | epoch  13 |   800/ 1452 batches | lr 4.00000 | ms/batch 210.11511 | loss  4.57 | ppl    96.33
## | epoch  13 |   900/ 1452 batches | lr 4.00000 | ms/batch 209.02129 | loss  4.57 | ppl    96.07
## | epoch  13 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.89753 | loss  4.54 | ppl    93.92
## | epoch  13 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.92734 | loss  4.66 | ppl   105.87
## | epoch  13 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.51675 | loss  4.63 | ppl   102.80
## | epoch  13 |  1300/ 1452 batches | lr 4.00000 | ms/batch 208.76649 | loss  4.44 | ppl    84.93
## | epoch  13 |  1400/ 1452 batches | lr 4.00000 | ms/batch 214.47578 | loss  4.60 | ppl    99.45
## -----------------------------------------------------------------------------------------
## | end of epoch  13 | time: 322.36s | valid loss  4.76 | valid ppl   116.38
## | end of epoch  13 | time: 322.36s | test loss  4.71 | test ppl   110.63
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  14 |   100/ 1452 batches | lr 4.00000 | ms/batch 212.65271 | loss  4.68 | ppl   107.65
## | epoch  14 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.62765 | loss  4.62 | ppl   101.39
## | epoch  14 |   300/ 1452 batches | lr 4.00000 | ms/batch 210.57477 | loss  4.62 | ppl   101.60
## | epoch  14 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.52675 | loss  4.47 | ppl    87.37
## | epoch  14 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.47678 | loss  4.50 | ppl    89.85
## | epoch  14 |   600/ 1452 batches | lr 4.00000 | ms/batch 209.57076 | loss  4.56 | ppl    95.87
## | epoch  14 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.26701 | loss  4.56 | ppl    95.25
## | epoch  14 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.50185 | loss  4.53 | ppl    93.08
## | epoch  14 |   900/ 1452 batches | lr 4.00000 | ms/batch 214.36076 | loss  4.53 | ppl    92.52
## | epoch  14 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.12714 | loss  4.51 | ppl    90.52
## | epoch  14 |  1100/ 1452 batches | lr 4.00000 | ms/batch 211.65354 | loss  4.63 | ppl   102.84
## | epoch  14 |  1200/ 1452 batches | lr 4.00000 | ms/batch 210.71450 | loss  4.60 | ppl    99.22
## | epoch  14 |  1300/ 1452 batches | lr 4.00000 | ms/batch 210.80440 | loss  4.41 | ppl    82.33
## | epoch  14 |  1400/ 1452 batches | lr 4.00000 | ms/batch 206.24906 | loss  4.57 | ppl    96.34
## -----------------------------------------------------------------------------------------
## | end of epoch  14 | time: 322.79s | valid loss  4.73 | valid ppl   113.81
## | end of epoch  14 | time: 322.79s | test loss  4.69 | test ppl   108.32
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  15 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.45087 | loss  4.64 | ppl   103.19
## | epoch  15 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.80256 | loss  4.59 | ppl    98.43
## | epoch  15 |   300/ 1452 batches | lr 4.00000 | ms/batch 210.08515 | loss  4.59 | ppl    98.32
```

```
## | epoch  15 |   400/ 1452 batches | lr 4.00000 | ms/batch 217.65741 | loss  4.44 | ppl    84.69
## | epoch  15 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.12325 | loss  4.46 | ppl    86.58
## | epoch  15 |   600/ 1452 batches | lr 4.00000 | ms/batch 214.36077 | loss  4.53 | ppl    92.31
## | epoch  15 |   700/ 1452 batches | lr 4.00000 | ms/batch 206.24907 | loss  4.53 | ppl    92.71
## | epoch  15 |   800/ 1452 batches | lr 4.00000 | ms/batch 207.83744 | loss  4.50 | ppl    90.30
## | epoch  15 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.26210 | loss  4.50 | ppl    89.60
## | epoch  15 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.04738 | loss  4.48 | ppl    87.85
## | epoch  15 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.05722 | loss  4.60 | ppl    99.14
## | epoch  15 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.17230 | loss  4.56 | ppl    95.66
## | epoch  15 |  1300/ 1452 batches | lr 4.00000 | ms/batch 207.86743 | loss  4.37 | ppl    79.06
## | epoch  15 |  1400/ 1452 batches | lr 4.00000 | ms/batch 209.59564 | loss  4.53 | ppl    93.02
## -----------------------------------------------------------------------------------------
## | end of epoch  15 | time: 322.33s | valid loss  4.72 | valid ppl   111.91
## | end of epoch  15 | time: 322.33s | test loss  4.67 | test ppl   106.34
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  16 |   100/ 1452 batches | lr 4.00000 | ms/batch 210.46474 | loss  4.61 | ppl   100.70
## | epoch  16 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.54673 | loss  4.56 | ppl    96.05
## | epoch  16 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.51184 | loss  4.56 | ppl    95.42
## | epoch  16 |   400/ 1452 batches | lr 4.00000 | ms/batch 206.99828 | loss  4.42 | ppl    82.76
## | epoch  16 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.84644 | loss  4.43 | ppl    83.80
## | epoch  16 |   600/ 1452 batches | lr 4.00000 | ms/batch 210.22498 | loss  4.50 | ppl    90.24
## | epoch  16 |   700/ 1452 batches | lr 4.00000 | ms/batch 207.06329 | loss  4.50 | ppl    89.68
## | epoch  16 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.81644 | loss  4.48 | ppl    87.93
## | epoch  16 |   900/ 1452 batches | lr 4.00000 | ms/batch 211.37109 | loss  4.47 | ppl    87.03
## | epoch  16 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.38301 | loss  4.44 | ppl    84.87
## | epoch  16 |  1100/ 1452 batches | lr 4.00000 | ms/batch 209.66556 | loss  4.57 | ppl    96.82
## | epoch  16 |  1200/ 1452 batches | lr 4.00000 | ms/batch 211.08429 | loss  4.54 | ppl    93.25
## | epoch  16 |  1300/ 1452 batches | lr 4.00000 | ms/batch 208.59666 | loss  4.34 | ppl    76.85
## | epoch  16 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.29696 | loss  4.51 | ppl    90.49
## -----------------------------------------------------------------------------------------
## | end of epoch  16 | time: 321.96s | valid loss  4.71 | valid ppl   111.39
## | end of epoch  16 | time: 321.96s | test loss  4.66 | test ppl   105.82
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  17 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.65662 | loss  4.59 | ppl    98.02
## | epoch  17 |   200/ 1452 batches | lr 4.00000 | ms/batch 213.76136 | loss  4.52 | ppl    92.04
## | epoch  17 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.85047 | loss  4.53 | ppl    92.41
## | epoch  17 |   400/ 1452 batches | lr 4.00000 | ms/batch 207.15329 | loss  4.38 | ppl    80.23
## | epoch  17 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.93734 | loss  4.40 | ppl    81.69
## | epoch  17 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.65761 | loss  4.47 | ppl    86.99
## | epoch  17 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.48192 | loss  4.47 | ppl    87.49
## | epoch  17 |   800/ 1452 batches | lr 4.00000 | ms/batch 209.98384 | loss  4.44 | ppl    84.91
## | epoch  17 |   900/ 1452 batches | lr 4.00000 | ms/batch 209.26118 | loss  4.44 | ppl    84.48
## | epoch  17 |  1000/ 1452 batches | lr 4.00000 | ms/batch 210.77955 | loss  4.42 | ppl    83.10
## | epoch  17 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.96627 | loss  4.55 | ppl    94.25
## | epoch  17 |  1200/ 1452 batches | lr 4.00000 | ms/batch 207.52776 | loss  4.51 | ppl    90.61
## | epoch  17 |  1300/ 1452 batches | lr 4.00000 | ms/batch 212.92223 | loss  4.31 | ppl    74.71
## | epoch  17 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.31796 | loss  4.47 | ppl    87.70
## -----------------------------------------------------------------------------------------
## | end of epoch  17 | time: 322.26s | valid loss  4.70 | valid ppl   110.47
## | end of epoch  17 | time: 322.26s | test loss  4.66 | test ppl   105.27
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  18 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.44088 | loss  4.55 | ppl    95.05
## | epoch  18 |   200/ 1452 batches | lr 4.00000 | ms/batch 210.10512 | loss  4.50 | ppl    90.01
## | epoch  18 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.04824 | loss  4.50 | ppl    89.97
## | epoch  18 |   400/ 1452 batches | lr 4.00000 | ms/batch 209.14123 | loss  4.36 | ppl    78.07
## | epoch  18 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.17710 | loss  4.37 | ppl    79.44
## | epoch  18 |   600/ 1452 batches | lr 4.00000 | ms/batch 209.81542 | loss  4.45 | ppl    85.24
## | epoch  18 |   700/ 1452 batches | lr 4.00000 | ms/batch 210.12510 | loss  4.44 | ppl    84.80
## | epoch  18 |   800/ 1452 batches | lr 4.00000 | ms/batch 211.81337 | loss  4.41 | ppl    82.44
## | epoch  18 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.82646 | loss  4.41 | ppl    82.15
## | epoch  18 |  1000/ 1452 batches | lr 4.00000 | ms/batch 209.17603 | loss  4.38 | ppl    80.11
## | epoch  18 |  1100/ 1452 batches | lr 4.00000 | ms/batch 206.62002 | loss  4.52 | ppl    91.90
## | epoch  18 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.52181 | loss  4.47 | ppl    87.78
## | epoch  18 |  1300/ 1452 batches | lr 4.00000 | ms/batch 214.68044 | loss  4.29 | ppl    72.86
## | epoch  18 |  1400/ 1452 batches | lr 4.00000 | ms/batch 205.71962 | loss  4.45 | ppl    85.40
## -----------------------------------------------------------------------------------------
```

```
## | end of epoch  18 | time: 322.20s | valid loss  4.69 | valid ppl   109.28
## | end of epoch  18 | time: 322.20s | test loss  4.65 | test ppl   104.16
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  19 |   100/ 1452 batches | lr 4.00000 | ms/batch 211.41379 | loss  4.53 | ppl    92.61
## | epoch  19 |   200/ 1452 batches | lr 4.00000 | ms/batch 212.20296 | loss  4.47 | ppl    87.36
## | epoch  19 |   300/ 1452 batches | lr 4.00000 | ms/batch 210.27131 | loss  4.47 | ppl    87.77
## | epoch  19 |   400/ 1452 batches | lr 4.00000 | ms/batch 210.26505 | loss  4.33 | ppl    75.89
## | epoch  19 |   500/ 1452 batches | lr 4.00000 | ms/batch 216.62846 | loss  4.35 | ppl    77.55
## | epoch  19 |   600/ 1452 batches | lr 4.00000 | ms/batch 205.15020 | loss  4.42 | ppl    82.91
## | epoch  19 |   700/ 1452 batches | lr 4.00000 | ms/batch 206.56874 | loss  4.41 | ppl    82.56
## | epoch  19 |   800/ 1452 batches | lr 4.00000 | ms/batch 207.06823 | loss  4.39 | ppl    80.43
## | epoch  19 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.19706 | loss  4.38 | ppl    79.98
## | epoch  19 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.60768 | loss  4.36 | ppl    78.37
## | epoch  19 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.32693 | loss  4.49 | ppl    89.16
## | epoch  19 |  1200/ 1452 batches | lr 4.00000 | ms/batch 211.62378 | loss  4.46 | ppl    86.23
## | epoch  19 |  1300/ 1452 batches | lr 4.00000 | ms/batch 211.20888 | loss  4.26 | ppl    70.98
## | epoch  19 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.99727 | loss  4.43 | ppl    84.14
## -----------------------------------------------------------------------------------------
## | end of epoch  19 | time: 322.68s | valid loss  4.68 | valid ppl   107.47
## | end of epoch  19 | time: 322.68s | test loss  4.62 | test ppl   101.98
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  20 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.11633 | loss  4.50 | ppl    90.20
## | epoch  20 |   200/ 1452 batches | lr 4.00000 | ms/batch 210.60992 | loss  4.45 | ppl    85.99
## | epoch  20 |   300/ 1452 batches | lr 4.00000 | ms/batch 205.48494 | loss  4.45 | ppl    85.65
## | epoch  20 |   400/ 1452 batches | lr 4.00000 | ms/batch 206.92368 | loss  4.30 | ppl    74.01
## | epoch  20 |   500/ 1452 batches | lr 4.00000 | ms/batch 206.37426 | loss  4.33 | ppl    76.08
## | epoch  20 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.55792 | loss  4.39 | ppl    80.58
## | epoch  20 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.25740 | loss  4.39 | ppl    80.80
## | epoch  20 |   800/ 1452 batches | lr 4.00000 | ms/batch 205.97466 | loss  4.37 | ppl    78.67
## | epoch  20 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.63425 | loss  4.36 | ppl    78.15
## | epoch  20 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.53723 | loss  4.34 | ppl    76.73
## | epoch  20 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.04355 | loss  4.47 | ppl    87.51
## | epoch  20 |  1200/ 1452 batches | lr 4.00000 | ms/batch 210.42536 | loss  4.43 | ppl    84.19
## | epoch  20 |  1300/ 1452 batches | lr 4.00000 | ms/batch 208.58327 | loss  4.24 | ppl    69.42
## | epoch  20 |  1400/ 1452 batches | lr 4.00000 | ms/batch 206.97372 | loss  4.40 | ppl    81.76
## -----------------------------------------------------------------------------------------
## | end of epoch  20 | time: 320.49s | valid loss  4.67 | valid ppl   106.46
## | end of epoch  20 | time: 320.49s | test loss  4.62 | test ppl   101.16
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  21 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.03642 | loss  4.48 | ppl    88.24
## | epoch  21 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.10840 | loss  4.43 | ppl    84.04
## | epoch  21 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.53309 | loss  4.42 | ppl    83.49
## | epoch  21 |   400/ 1452 batches | lr 4.00000 | ms/batch 209.89070 | loss  4.29 | ppl    72.70
## | epoch  21 |   500/ 1452 batches | lr 4.00000 | ms/batch 206.30946 | loss  4.31 | ppl    74.26
## | epoch  21 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.82768 | loss  4.36 | ppl    78.60
## | epoch  21 |   700/ 1452 batches | lr 4.00000 | ms/batch 209.65600 | loss  4.37 | ppl    78.70
## | epoch  21 |   800/ 1452 batches | lr 4.00000 | ms/batch 207.30829 | loss  4.34 | ppl    76.77
## | epoch  21 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.54507 | loss  4.34 | ppl    76.59
## | epoch  21 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.37326 | loss  4.31 | ppl    74.81
## | epoch  21 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.71694 | loss  4.45 | ppl    85.51
## | epoch  21 |  1200/ 1452 batches | lr 4.00000 | ms/batch 211.69931 | loss  4.41 | ppl    81.91
## | epoch  21 |  1300/ 1452 batches | lr 4.00000 | ms/batch 213.14338 | loss  4.22 | ppl    67.97
## | epoch  21 |  1400/ 1452 batches | lr 4.00000 | ms/batch 206.97875 | loss  4.39 | ppl    80.40
## -----------------------------------------------------------------------------------------
## | end of epoch  21 | time: 321.46s | valid loss  4.66 | valid ppl   105.48
## | end of epoch  21 | time: 321.46s | test loss  4.61 | test ppl   100.16
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  22 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.49187 | loss  4.47 | ppl    87.06
## | epoch  22 |   200/ 1452 batches | lr 4.00000 | ms/batch 206.90371 | loss  4.41 | ppl    81.99
## | epoch  22 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.05843 | loss  4.41 | ppl    81.92
## | epoch  22 |   400/ 1452 batches | lr 4.00000 | ms/batch 206.95854 | loss  4.26 | ppl    71.02
## | epoch  22 |   500/ 1452 batches | lr 4.00000 | ms/batch 209.82541 | loss  4.28 | ppl    72.27
## | epoch  22 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.15833 | loss  4.35 | ppl    77.09
## | epoch  22 |   700/ 1452 batches | lr 4.00000 | ms/batch 210.95974 | loss  4.35 | ppl    77.10
```

```
## | epoch  22 |   800/ 1452 batches | lr 4.00000 | ms/batch 205.42010 | loss  4.32 | ppl    75.37
## | epoch  22 |   900/ 1452 batches | lr 4.00000 | ms/batch 206.59893 | loss  4.32 | ppl    75.31
## | epoch  22 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.41194 | loss  4.30 | ppl    73.52
## | epoch  22 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.84284 | loss  4.43 | ppl    84.05
## | epoch  22 |  1200/ 1452 batches | lr 4.00000 | ms/batch 212.74776 | loss  4.39 | ppl    80.39
## | epoch  22 |  1300/ 1452 batches | lr 4.00000 | ms/batch 208.17266 | loss  4.19 | ppl    66.12
## | epoch  22 |  1400/ 1452 batches | lr 4.00000 | ms/batch 206.93836 | loss  4.36 | ppl    78.38
## -----------------------------------------------------------------------------------------
## | end of epoch  22 | time: 320.55s | valid loss  4.65 | valid ppl   105.04
## | end of epoch  22 | time: 320.55s | test loss  4.60 | test ppl    99.73
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  23 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.82179 | loss  4.44 | ppl    84.50
## | epoch  23 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.21806 | loss  4.39 | ppl    80.41
## | epoch  23 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.84121 | loss  4.39 | ppl    80.42
## | epoch  23 |   400/ 1452 batches | lr 4.00000 | ms/batch 209.47595 | loss  4.24 | ppl    69.51
## | epoch  23 |   500/ 1452 batches | lr 4.00000 | ms/batch 206.66913 | loss  4.27 | ppl    71.24
## | epoch  23 |   600/ 1452 batches | lr 4.00000 | ms/batch 214.80545 | loss  4.32 | ppl    75.35
## | epoch  23 |   700/ 1452 batches | lr 4.00000 | ms/batch 206.12916 | loss  4.33 | ppl    75.88
## | epoch  23 |   800/ 1452 batches | lr 4.00000 | ms/batch 211.33898 | loss  4.30 | ppl    73.49
## | epoch  23 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.96542 | loss  4.29 | ppl    73.17
## | epoch  23 |  1000/ 1452 batches | lr 4.00000 | ms/batch 206.34407 | loss  4.27 | ppl    71.33
## | epoch  23 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.01848 | loss  4.41 | ppl    82.56
## | epoch  23 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.88665 | loss  4.37 | ppl    78.76
## | epoch  23 |  1300/ 1452 batches | lr 4.00000 | ms/batch 217.12310 | loss  4.17 | ppl    64.98
## | epoch  23 |  1400/ 1452 batches | lr 4.00000 | ms/batch 205.90963 | loss  4.34 | ppl    76.93
## -----------------------------------------------------------------------------------------
## | end of epoch  23 | time: 321.77s | valid loss  4.65 | valid ppl   104.94
## | end of epoch  23 | time: 321.77s | test loss  4.60 | test ppl    99.69
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  24 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.33104 | loss  4.42 | ppl    82.99
## | epoch  24 |   200/ 1452 batches | lr 4.00000 | ms/batch 209.97120 | loss  4.37 | ppl    78.68
## | epoch  24 |   300/ 1452 batches | lr 4.00000 | ms/batch 206.71902 | loss  4.37 | ppl    79.06
## | epoch  24 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.60703 | loss  4.22 | ppl    68.19
## | epoch  24 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.59303 | loss  4.24 | ppl    69.70
## | epoch  24 |   600/ 1452 batches | lr 4.00000 | ms/batch 209.94548 | loss  4.30 | ppl    73.61
## | epoch  24 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.01243 | loss  4.31 | ppl    74.30
## | epoch  24 |   800/ 1452 batches | lr 4.00000 | ms/batch 206.77852 | loss  4.28 | ppl    72.24
## | epoch  24 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.02726 | loss  4.28 | ppl    72.12
## | epoch  24 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.62273 | loss  4.25 | ppl    70.38
## | epoch  24 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.86741 | loss  4.39 | ppl    80.40
## | epoch  24 |  1200/ 1452 batches | lr 4.00000 | ms/batch 207.59768 | loss  4.34 | ppl    76.95
## | epoch  24 |  1300/ 1452 batches | lr 4.00000 | ms/batch 206.98388 | loss  4.16 | ppl    63.84
## | epoch  24 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.36303 | loss  4.31 | ppl    74.81
## -----------------------------------------------------------------------------------------
## | end of epoch  24 | time: 320.44s | valid loss  4.66 | valid ppl   105.87
## | end of epoch  24 | time: 320.44s | test loss  4.61 | test ppl   100.35
## -----------------------------------------------------------------------------------------
## | epoch  25 |   100/ 1452 batches | lr 4.00000 | ms/batch 213.25702 | loss  4.40 | ppl    81.43
## | epoch  25 |   200/ 1452 batches | lr 4.00000 | ms/batch 206.83357 | loss  4.35 | ppl    77.80
## | epoch  25 |   300/ 1452 batches | lr 4.00000 | ms/batch 213.66150 | loss  4.35 | ppl    77.42
## | epoch  25 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.99541 | loss  4.20 | ppl    66.91
## | epoch  25 |   500/ 1452 batches | lr 4.00000 | ms/batch 213.42682 | loss  4.23 | ppl    68.88
## | epoch  25 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.25337 | loss  4.28 | ppl    72.34
## | epoch  25 |   700/ 1452 batches | lr 4.00000 | ms/batch 207.47318 | loss  4.29 | ppl    72.96
## | epoch  25 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.09264 | loss  4.26 | ppl    70.90
## | epoch  25 |   900/ 1452 batches | lr 4.00000 | ms/batch 211.33897 | loss  4.26 | ppl    70.56
## | epoch  25 |  1000/ 1452 batches | lr 4.00000 | ms/batch 212.67247 | loss  4.24 | ppl    69.12
## | epoch  25 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.49803 | loss  4.36 | ppl    78.61
## | epoch  25 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.63204 | loss  4.33 | ppl    75.79
## | epoch  25 |  1300/ 1452 batches | lr 4.00000 | ms/batch 207.99765 | loss  4.13 | ppl    62.26
## | epoch  25 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.84793 | loss  4.30 | ppl    73.64
## -----------------------------------------------------------------------------------------
## | end of epoch  25 | time: 322.70s | valid loss  4.65 | valid ppl   104.42
## | end of epoch  25 | time: 322.70s | test loss  4.60 | test ppl    99.36
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  26 |   100/ 1452 batches | lr 4.00000 | ms/batch 210.26006 | loss  4.38 | ppl    79.90
```

```
## | epoch  26 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.01241 | loss  4.34 | ppl    76.48
## | epoch  26 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.81077 | loss  4.33 | ppl    75.85
## | epoch  26 |   400/ 1452 batches | lr 4.00000 | ms/batch 210.19523 | loss  4.19 | ppl    65.96
## | epoch  26 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.31313 | loss  4.21 | ppl    67.39
## | epoch  26 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.32817 | loss  4.27 | ppl    71.47
## | epoch  26 |   700/ 1452 batches | lr 4.00000 | ms/batch 209.12148 | loss  4.27 | ppl    71.45
## | epoch  26 |   800/ 1452 batches | lr 4.00000 | ms/batch 210.82968 | loss  4.25 | ppl    69.95
## | epoch  26 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.32023 | loss  4.24 | ppl    69.42
## | epoch  26 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.30229 | loss  4.22 | ppl    67.83
## | epoch  26 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.20724 | loss  4.35 | ppl    77.82
## | epoch  26 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.43222 | loss  4.31 | ppl    74.08
## | epoch  26 |  1300/ 1452 batches | lr 4.00000 | ms/batch 210.62997 | loss  4.12 | ppl    61.36
## | epoch  26 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.23807 | loss  4.29 | ppl    73.10
## -----------------------------------------------------------------------------------------
## | end of epoch  26 | time: 321.80s | valid loss  4.65 | valid ppl   104.29
## | end of epoch  26 | time: 321.80s | test loss  4.60 | test ppl    99.28
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  27 |   100/ 1452 batches | lr 4.00000 | ms/batch 213.98627 | loss  4.37 | ppl    78.90
## | epoch  27 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.97240 | loss  4.32 | ppl    74.87
## | epoch  27 |   300/ 1452 batches | lr 4.00000 | ms/batch 212.10350 | loss  4.31 | ppl    74.47
## | epoch  27 |   400/ 1452 batches | lr 4.00000 | ms/batch 206.28900 | loss  4.17 | ppl    64.52
## | epoch  27 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.51188 | loss  4.19 | ppl    66.33
## | epoch  27 |   600/ 1452 batches | lr 4.00000 | ms/batch 211.67680 | loss  4.25 | ppl    69.93
## | epoch  27 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.85704 | loss  4.25 | ppl    70.41
## | epoch  27 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.73676 | loss  4.22 | ppl    68.18
## | epoch  27 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.24342 | loss  4.22 | ppl    67.94
## | epoch  27 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.54712 | loss  4.20 | ppl    66.82
## | epoch  27 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.49290 | loss  4.33 | ppl    75.77
## | epoch  27 |  1200/ 1452 batches | lr 4.00000 | ms/batch 210.72412 | loss  4.29 | ppl    72.85
## | epoch  27 |  1300/ 1452 batches | lr 4.00000 | ms/batch 208.98144 | loss  4.10 | ppl    60.35
## | epoch  27 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.34306 | loss  4.27 | ppl    71.39
## -----------------------------------------------------------------------------------------
## | end of epoch  27 | time: 322.04s | valid loss  4.63 | valid ppl   102.71
## | end of epoch  27 | time: 322.04s | test loss  4.58 | test ppl    97.63
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  28 |   100/ 1452 batches | lr 4.00000 | ms/batch 211.06437 | loss  4.35 | ppl    77.79
## | epoch  28 |   200/ 1452 batches | lr 4.00000 | ms/batch 209.04620 | loss  4.29 | ppl    73.33
## | epoch  28 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.54775 | loss  4.29 | ppl    73.25
## | epoch  28 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.39712 | loss  4.15 | ppl    63.62
## | epoch  28 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.44702 | loss  4.17 | ppl    64.91
## | epoch  28 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.51145 | loss  4.23 | ppl    68.97
## | epoch  28 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.70168 | loss  4.24 | ppl    69.30
## | epoch  28 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.48209 | loss  4.21 | ppl    67.20
## | epoch  28 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.19760 | loss  4.20 | ppl    66.78
## | epoch  28 |  1000/ 1452 batches | lr 4.00000 | ms/batch 213.92878 | loss  4.18 | ppl    65.63
## | epoch  28 |  1100/ 1452 batches | lr 4.00000 | ms/batch 206.92859 | loss  4.32 | ppl    75.23
## | epoch  28 |  1200/ 1452 batches | lr 4.00000 | ms/batch 207.31331 | loss  4.27 | ppl    71.57
## | epoch  28 |  1300/ 1452 batches | lr 4.00000 | ms/batch 206.94857 | loss  4.09 | ppl    59.51
## | epoch  28 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.96261 | loss  4.25 | ppl    70.12
## -----------------------------------------------------------------------------------------
## | end of epoch  28 | time: 321.12s | valid loss  4.64 | valid ppl   103.41
## | end of epoch  28 | time: 321.12s | test loss  4.59 | test ppl    98.47
## -----------------------------------------------------------------------------------------
## | epoch  29 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.56078 | loss  4.33 | ppl    75.99
## | epoch  29 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.69267 | loss  4.28 | ppl    72.60
## | epoch  29 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.31358 | loss  4.28 | ppl    72.24
## | epoch  29 |   400/ 1452 batches | lr 4.00000 | ms/batch 207.77750 | loss  4.14 | ppl    62.56
## | epoch  29 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.04334 | loss  4.17 | ppl    64.40
## | epoch  29 |   600/ 1452 batches | lr 4.00000 | ms/batch 223.30163 | loss  4.21 | ppl    67.45
## | epoch  29 |   700/ 1452 batches | lr 4.00000 | ms/batch 204.60105 | loss  4.22 | ppl    68.08
## | epoch  29 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.93159 | loss  4.19 | ppl    65.81
## | epoch  29 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.09230 | loss  4.18 | ppl    65.60
## | epoch  29 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.98162 | loss  4.16 | ppl    64.20
## | epoch  29 |  1100/ 1452 batches | lr 4.00000 | ms/batch 209.02655 | loss  4.29 | ppl    73.25
## | epoch  29 |  1200/ 1452 batches | lr 4.00000 | ms/batch 209.72601 | loss  4.26 | ppl    70.70
## | epoch  29 |  1300/ 1452 batches | lr 4.00000 | ms/batch 207.67781 | loss  4.08 | ppl    58.86
## | epoch  29 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.06249 | loss  4.24 | ppl    69.38
## -----------------------------------------------------------------------------------------
```

```
## | end of epoch  29 | time: 322.16s | valid loss  4.63 | valid ppl   102.28
## | end of epoch  29 | time: 322.16s | test loss  4.57 | test ppl    97.02
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  30 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.76677 | loss  4.31 | ppl    74.70
## | epoch  30 |   200/ 1452 batches | lr 4.00000 | ms/batch 210.08034 | loss  4.27 | ppl    71.28
## | epoch  30 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.42094 | loss  4.26 | ppl    70.92
## | epoch  30 |   400/ 1452 batches | lr 4.00000 | ms/batch 214.07616 | loss  4.12 | ppl    61.82
## | epoch  30 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.17810 | loss  4.15 | ppl    63.33
## | epoch  30 |   600/ 1452 batches | lr 4.00000 | ms/batch 208.90656 | loss  4.19 | ppl    66.32
## | epoch  30 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.84660 | loss  4.21 | ppl    67.12
## | epoch  30 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.14737 | loss  4.17 | ppl    64.98
## | epoch  30 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.65761 | loss  4.17 | ppl    64.96
## | epoch  30 |  1000/ 1452 batches | lr 4.00000 | ms/batch 206.71369 | loss  4.15 | ppl    63.45
## | epoch  30 |  1100/ 1452 batches | lr 4.00000 | ms/batch 212.32289 | loss  4.28 | ppl    72.50
## | epoch  30 |  1200/ 1452 batches | lr 4.00000 | ms/batch 209.17822 | loss  4.24 | ppl    69.52
## | epoch  30 |  1300/ 1452 batches | lr 4.00000 | ms/batch 209.45112 | loss  4.05 | ppl    57.65
## | epoch  30 |  1400/ 1452 batches | lr 4.00000 | ms/batch 213.88146 | loss  4.22 | ppl    68.11
## -----------------------------------------------------------------------------------------
## | end of epoch  30 | time: 322.83s | valid loss  4.63 | valid ppl   102.24
## | end of epoch  30 | time: 322.83s | test loss  4.58 | test ppl    97.18
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  31 |   100/ 1452 batches | lr 4.00000 | ms/batch 206.92863 | loss  4.30 | ppl    73.39
## | epoch  31 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.25314 | loss  4.24 | ppl    69.71
## | epoch  31 |   300/ 1452 batches | lr 4.00000 | ms/batch 214.80032 | loss  4.25 | ppl    70.13
## | epoch  31 |   400/ 1452 batches | lr 4.00000 | ms/batch 215.04026 | loss  4.10 | ppl    60.46
## | epoch  31 |   500/ 1452 batches | lr 4.00000 | ms/batch 210.86939 | loss  4.13 | ppl    62.40
## | epoch  31 |   600/ 1452 batches | lr 4.00000 | ms/batch 206.61420 | loss  4.19 | ppl    66.14
## | epoch  31 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.06257 | loss  4.19 | ppl    65.92
## | epoch  31 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.48724 | loss  4.16 | ppl    64.07
## | epoch  31 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.25029 | loss  4.16 | ppl    64.12
## | epoch  31 |  1000/ 1452 batches | lr 4.00000 | ms/batch 209.30975 | loss  4.14 | ppl    62.51
## | epoch  31 |  1100/ 1452 batches | lr 4.00000 | ms/batch 209.72551 | loss  4.27 | ppl    71.76
## | epoch  31 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.54179 | loss  4.23 | ppl    68.47
## | epoch  31 |  1300/ 1452 batches | lr 4.00000 | ms/batch 218.91120 | loss  4.05 | ppl    57.14
## | epoch  31 |  1400/ 1452 batches | lr 4.00000 | ms/batch 216.62378 | loss  4.21 | ppl    67.44
## -----------------------------------------------------------------------------------------
## | end of epoch  31 | time: 324.37s | valid loss  4.63 | valid ppl   102.02
## | end of epoch  31 | time: 324.37s | test loss  4.57 | test ppl    96.96
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  32 |   100/ 1452 batches | lr 4.00000 | ms/batch 213.05229 | loss  4.29 | ppl    72.69
## | epoch  32 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.36816 | loss  4.23 | ppl    68.88
## | epoch  32 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.86247 | loss  4.24 | ppl    69.07
## | epoch  32 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.70166 | loss  4.10 | ppl    60.22
## | epoch  32 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.70755 | loss  4.11 | ppl    61.25
## | epoch  32 |   600/ 1452 batches | lr 4.00000 | ms/batch 210.27028 | loss  4.17 | ppl    64.74
## | epoch  32 |   700/ 1452 batches | lr 4.00000 | ms/batch 209.09638 | loss  4.17 | ppl    64.96
## | epoch  32 |   800/ 1452 batches | lr 4.00000 | ms/batch 209.76563 | loss  4.14 | ppl    62.80
## | epoch  32 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.62295 | loss  4.14 | ppl    62.61
## | epoch  32 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.12817 | loss  4.11 | ppl    61.21
## | epoch  32 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.96733 | loss  4.26 | ppl    71.04
## | epoch  32 |  1200/ 1452 batches | lr 4.00000 | ms/batch 207.85758 | loss  4.22 | ppl    67.96
## | epoch  32 |  1300/ 1452 batches | lr 4.00000 | ms/batch 220.03884 | loss  4.02 | ppl    55.83
## | epoch  32 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.25314 | loss  4.19 | ppl    66.31
## -----------------------------------------------------------------------------------------
## | end of epoch  32 | time: 322.23s | valid loss  4.61 | valid ppl   100.87
## | end of epoch  32 | time: 322.23s | test loss  4.56 | test ppl    95.75
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  33 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.50697 | loss  4.27 | ppl    71.39
## | epoch  33 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.78764 | loss  4.22 | ppl    67.89
## | epoch  33 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.73269 | loss  4.22 | ppl    68.03
## | epoch  33 |   400/ 1452 batches | lr 4.00000 | ms/batch 212.76749 | loss  4.08 | ppl    59.03
## | epoch  33 |   500/ 1452 batches | lr 4.00000 | ms/batch 209.88041 | loss  4.11 | ppl    60.89
## | epoch  33 |   600/ 1452 batches | lr 4.00000 | ms/batch 206.97850 | loss  4.15 | ppl    63.68
## | epoch  33 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.83154 | loss  4.16 | ppl    64.04
```

```
## | epoch  33 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.22219 | loss  4.13 | ppl    62.26
## | epoch  33 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.33713 | loss  4.13 | ppl    62.17
## | epoch  33 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.14328 | loss  4.11 | ppl    61.05
## | epoch  33 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.99649 | loss  4.25 | ppl    69.85
## | epoch  33 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.57419 | loss  4.20 | ppl    66.67
## | epoch  33 |  1300/ 1452 batches | lr 4.00000 | ms/batch 209.19116 | loss  4.01 | ppl    55.40
## | epoch  33 |  1400/ 1452 batches | lr 4.00000 | ms/batch 217.20354 | loss  4.18 | ppl    65.35
## -----------------------------------------------------------------------------------------
## | end of epoch  33 | time: 322.20s | valid loss  4.61 | valid ppl   100.71
## | end of epoch  33 | time: 322.20s | test loss  4.56 | test ppl    96.05
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  34 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.33619 | loss  4.25 | ppl    70.42
## | epoch  34 |   200/ 1452 batches | lr 4.00000 | ms/batch 210.52512 | loss  4.20 | ppl    66.75
## | epoch  34 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.52711 | loss  4.20 | ppl    66.85
## | epoch  34 |   400/ 1452 batches | lr 4.00000 | ms/batch 210.47493 | loss  4.06 | ppl    58.15
## | epoch  34 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.60194 | loss  4.09 | ppl    59.87
## | epoch  34 |   600/ 1452 batches | lr 4.00000 | ms/batch 210.02570 | loss  4.14 | ppl    62.59
## | epoch  34 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.80694 | loss  4.15 | ppl    63.20
## | epoch  34 |   800/ 1452 batches | lr 4.00000 | ms/batch 209.50104 | loss  4.12 | ppl    61.38
## | epoch  34 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.30208 | loss  4.11 | ppl    61.11
## | epoch  34 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.06231 | loss  4.10 | ppl    60.52
## | epoch  34 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.52203 | loss  4.23 | ppl    68.61
## | epoch  34 |  1200/ 1452 batches | lr 4.00000 | ms/batch 210.17529 | loss  4.19 | ppl    65.89
## | epoch  34 |  1300/ 1452 batches | lr 4.00000 | ms/batch 209.03155 | loss  4.00 | ppl    54.46
## | epoch  34 |  1400/ 1452 batches | lr 4.00000 | ms/batch 210.25517 | loss  4.17 | ppl    64.80
## -----------------------------------------------------------------------------------------
## | end of epoch  34 | time: 322.76s | valid loss  4.61 | valid ppl   100.59
## | end of epoch  34 | time: 322.76s | test loss  4.56 | test ppl    95.58
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  35 |   100/ 1452 batches | lr 4.00000 | ms/batch 207.99258 | loss  4.24 | ppl    69.58
## | epoch  35 |   200/ 1452 batches | lr 4.00000 | ms/batch 209.06634 | loss  4.19 | ppl    66.21
## | epoch  35 |   300/ 1452 batches | lr 4.00000 | ms/batch 215.92557 | loss  4.18 | ppl    65.57
## | epoch  35 |   400/ 1452 batches | lr 4.00000 | ms/batch 206.53795 | loss  4.06 | ppl    57.86
## | epoch  35 |   500/ 1452 batches | lr 4.00000 | ms/batch 206.37919 | loss  4.08 | ppl    58.98
## | epoch  35 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.23850 | loss  4.13 | ppl    62.05
## | epoch  35 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.25252 | loss  4.13 | ppl    62.45
## | epoch  35 |   800/ 1452 batches | lr 4.00000 | ms/batch 207.31837 | loss  4.11 | ppl    61.16
## | epoch  35 |   900/ 1452 batches | lr 4.00000 | ms/batch 212.30819 | loss  4.10 | ppl    60.63
## | epoch  35 |  1000/ 1452 batches | lr 4.00000 | ms/batch 204.18137 | loss  4.09 | ppl    59.57
## | epoch  35 |  1100/ 1452 batches | lr 4.00000 | ms/batch 206.57407 | loss  4.22 | ppl    67.89
## | epoch  35 |  1200/ 1452 batches | lr 4.00000 | ms/batch 207.27820 | loss  4.16 | ppl    64.39
## | epoch  35 |  1300/ 1452 batches | lr 4.00000 | ms/batch 207.00360 | loss  3.99 | ppl    53.86
## | epoch  35 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.51308 | loss  4.15 | ppl    63.49
## -----------------------------------------------------------------------------------------
## | end of epoch  35 | time: 320.92s | valid loss  4.62 | valid ppl   101.72
## | end of epoch  35 | time: 320.92s | test loss  4.57 | test ppl    96.74
## -----------------------------------------------------------------------------------------
## | epoch  36 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.37111 | loss  4.23 | ppl    68.62
## | epoch  36 |   200/ 1452 batches | lr 4.00000 | ms/batch 209.05157 | loss  4.18 | ppl    65.48
## | epoch  36 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.02653 | loss  4.18 | ppl    65.48
## | epoch  36 |   400/ 1452 batches | lr 4.00000 | ms/batch 207.39826 | loss  4.05 | ppl    57.39
## | epoch  36 |   500/ 1452 batches | lr 4.00000 | ms/batch 212.18349 | loss  4.06 | ppl    57.99
## | epoch  36 |   600/ 1452 batches | lr 4.00000 | ms/batch 211.50923 | loss  4.12 | ppl    61.32
## | epoch  36 |   700/ 1452 batches | lr 4.00000 | ms/batch 207.14859 | loss  4.12 | ppl    61.86
## | epoch  36 |   800/ 1452 batches | lr 4.00000 | ms/batch 206.82878 | loss  4.10 | ppl    60.17
## | epoch  36 |   900/ 1452 batches | lr 4.00000 | ms/batch 209.96549 | loss  4.08 | ppl    59.43
## | epoch  36 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.77793 | loss  4.07 | ppl    58.48
## | epoch  36 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.72792 | loss  4.21 | ppl    67.15
## | epoch  36 |  1200/ 1452 batches | lr 4.00000 | ms/batch 207.20357 | loss  4.16 | ppl    64.17
## | epoch  36 |  1300/ 1452 batches | lr 4.00000 | ms/batch 213.07253 | loss  3.97 | ppl    53.14
## | epoch  36 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.55644 | loss  4.14 | ppl    62.61
## -----------------------------------------------------------------------------------------
## | end of epoch  36 | time: 321.78s | valid loss  4.61 | valid ppl   100.37
## | end of epoch  36 | time: 321.78s | test loss  4.56 | test ppl    95.29
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  37 |   100/ 1452 batches | lr 4.00000 | ms/batch 210.38023 | loss  4.22 | ppl    67.79
```

```
## | epoch  37 |   200/ 1452 batches | lr 4.00000 | ms/batch 212.14838 | loss  4.17 | ppl    64.52
## | epoch  37 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.35919 | loss  4.16 | ppl    64.38
## | epoch  37 |   400/ 1452 batches | lr 4.00000 | ms/batch 207.74794 | loss  4.03 | ppl    56.27
## | epoch  37 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.36221 | loss  4.05 | ppl    57.54
## | epoch  37 |   600/ 1452 batches | lr 4.00000 | ms/batch 208.26744 | loss  4.10 | ppl    60.40
## | epoch  37 |   700/ 1452 batches | lr 4.00000 | ms/batch 206.95869 | loss  4.11 | ppl    61.11
## | epoch  37 |   800/ 1452 batches | lr 4.00000 | ms/batch 211.94324 | loss  4.08 | ppl    59.21
## | epoch  37 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.11734 | loss  4.08 | ppl    59.16
## | epoch  37 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.66175 | loss  4.06 | ppl    57.83
## | epoch  37 |  1100/ 1452 batches | lr 4.00000 | ms/batch 216.25391 | loss  4.19 | ppl    66.25
## | epoch  37 |  1200/ 1452 batches | lr 4.00000 | ms/batch 206.06925 | loss  4.14 | ppl    62.95
## | epoch  37 |  1300/ 1452 batches | lr 4.00000 | ms/batch 214.57572 | loss  3.95 | ppl    52.11
## | epoch  37 |  1400/ 1452 batches | lr 4.00000 | ms/batch 206.25438 | loss  4.13 | ppl    61.99
## -----------------------------------------------------------------------------------------
## | end of epoch  37 | time: 322.30s | valid loss  4.61 | valid ppl   100.52
## | end of epoch  37 | time: 322.30s | test loss  4.56 | test ppl    95.61
## -----------------------------------------------------------------------------------------
## | epoch  38 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.44698 | loss  4.21 | ppl    67.15
## | epoch  38 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.86149 | loss  4.16 | ppl    64.07
## | epoch  38 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.01247 | loss  4.16 | ppl    64.39
## | epoch  38 |   400/ 1452 batches | lr 4.00000 | ms/batch 216.65349 | loss  4.01 | ppl    55.34
## | epoch  38 |   500/ 1452 batches | lr 4.00000 | ms/batch 206.48388 | loss  4.04 | ppl    57.03
## | epoch  38 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.17835 | loss  4.09 | ppl    59.86
## | epoch  38 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.71021 | loss  4.09 | ppl    59.98
## | epoch  38 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.91144 | loss  4.07 | ppl    58.39
## | epoch  38 |   900/ 1452 batches | lr 4.00000 | ms/batch 210.74462 | loss  4.06 | ppl    58.14
## | epoch  38 |  1000/ 1452 batches | lr 4.00000 | ms/batch 206.87352 | loss  4.05 | ppl    57.18
## | epoch  38 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.77261 | loss  4.18 | ppl    65.50
## | epoch  38 |  1200/ 1452 batches | lr 4.00000 | ms/batch 208.01745 | loss  4.13 | ppl    62.11
## | epoch  38 |  1300/ 1452 batches | lr 4.00000 | ms/batch 214.71551 | loss  3.94 | ppl    51.61
## | epoch  38 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.76750 | loss  4.11 | ppl    60.76
## -----------------------------------------------------------------------------------------
## | end of epoch  38 | time: 322.28s | valid loss  4.61 | valid ppl   100.09
## | end of epoch  38 | time: 322.28s | test loss  4.55 | test ppl    94.95
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  39 |   100/ 1452 batches | lr 4.00000 | ms/batch 207.95749 | loss  4.19 | ppl    65.73
## | epoch  39 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.68184 | loss  4.14 | ppl    62.79
## | epoch  39 |   300/ 1452 batches | lr 4.00000 | ms/batch 206.70880 | loss  4.14 | ppl    62.82
## | epoch  39 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.06747 | loss  4.01 | ppl    54.88
## | epoch  39 |   500/ 1452 batches | lr 4.00000 | ms/batch 211.96321 | loss  4.03 | ppl    56.13
## | epoch  39 |   600/ 1452 batches | lr 4.00000 | ms/batch 206.25416 | loss  4.08 | ppl    58.98
## | epoch  39 |   700/ 1452 batches | lr 4.00000 | ms/batch 207.38793 | loss  4.08 | ppl    59.40
## | epoch  39 |   800/ 1452 batches | lr 4.00000 | ms/batch 212.42781 | loss  4.05 | ppl    57.39
## | epoch  39 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.48779 | loss  4.05 | ppl    57.53
## | epoch  39 |  1000/ 1452 batches | lr 4.00000 | ms/batch 209.62560 | loss  4.03 | ppl    56.47
## | epoch  39 |  1100/ 1452 batches | lr 4.00000 | ms/batch 207.77264 | loss  4.17 | ppl    64.52
## | epoch  39 |  1200/ 1452 batches | lr 4.00000 | ms/batch 206.50390 | loss  4.12 | ppl    61.28
## | epoch  39 |  1300/ 1452 batches | lr 4.00000 | ms/batch 212.52378 | loss  3.93 | ppl    50.98
## | epoch  39 |  1400/ 1452 batches | lr 4.00000 | ms/batch 205.70468 | loss  4.11 | ppl    60.93
## -----------------------------------------------------------------------------------------
## | end of epoch  39 | time: 321.15s | valid loss  4.61 | valid ppl   100.79
## | end of epoch  39 | time: 321.15s | test loss  4.56 | test ppl    95.79
## -----------------------------------------------------------------------------------------
## | epoch  40 |   100/ 1452 batches | lr 4.00000 | ms/batch 209.08663 | loss  4.18 | ppl    65.06
## | epoch  40 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.18246 | loss  4.13 | ppl    62.35
## | epoch  40 |   300/ 1452 batches | lr 4.00000 | ms/batch 210.87943 | loss  4.13 | ppl    62.12
## | epoch  40 |   400/ 1452 batches | lr 4.00000 | ms/batch 207.89248 | loss  4.00 | ppl    54.37
## | epoch  40 |   500/ 1452 batches | lr 4.00000 | ms/batch 210.71465 | loss  4.01 | ppl    55.34
## | epoch  40 |   600/ 1452 batches | lr 4.00000 | ms/batch 219.47084 | loss  4.07 | ppl    58.60
## | epoch  40 |   700/ 1452 batches | lr 4.00000 | ms/batch 214.93529 | loss  4.07 | ppl    58.77
## | epoch  40 |   800/ 1452 batches | lr 4.00000 | ms/batch 206.56893 | loss  4.05 | ppl    57.34
## | epoch  40 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.66197 | loss  4.04 | ppl    56.91
## | epoch  40 |  1000/ 1452 batches | lr 4.00000 | ms/batch 209.43112 | loss  4.02 | ppl    55.49
## | epoch  40 |  1100/ 1452 batches | lr 4.00000 | ms/batch 206.35429 | loss  4.17 | ppl    64.50
## | epoch  40 |  1200/ 1452 batches | lr 4.00000 | ms/batch 209.24639 | loss  4.10 | ppl    60.31
## | epoch  40 |  1300/ 1452 batches | lr 4.00000 | ms/batch 212.52262 | loss  3.93 | ppl    50.72
## | epoch  40 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.59214 | loss  4.09 | ppl    59.91
## -----------------------------------------------------------------------------------------
## | end of epoch  40 | time: 323.39s | valid loss  4.62 | valid ppl   101.15
## | end of epoch  40 | time: 323.39s | test loss  4.57 | test ppl    96.21
```

```
## -----------------------------------------------------------------------------------------
## | epoch  41 |   100/ 1452 batches | lr 4.00000 | ms/batch 214.55584 | loss  4.17 | ppl    64.50
## | epoch  41 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.59810 | loss  4.12 | ppl    61.61
## | epoch  41 |   300/ 1452 batches | lr 4.00000 | ms/batch 207.24336 | loss  4.12 | ppl    61.52
## | epoch  41 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.18263 | loss  3.99 | ppl    54.13
## | epoch  41 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.62783 | loss  4.00 | ppl    54.82
## | epoch  41 |   600/ 1452 batches | lr 4.00000 | ms/batch 209.66601 | loss  4.06 | ppl    57.94
## | epoch  41 |   700/ 1452 batches | lr 4.00000 | ms/batch 207.65804 | loss  4.06 | ppl    58.22
## | epoch  41 |   800/ 1452 batches | lr 4.00000 | ms/batch 208.51717 | loss  4.03 | ppl    56.42
## | epoch  41 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.49815 | loss  4.03 | ppl    56.20
## | epoch  41 |  1000/ 1452 batches | lr 4.00000 | ms/batch 206.98358 | loss  4.01 | ppl    55.09
## | epoch  41 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.76330 | loss  4.15 | ppl    63.43
## | epoch  41 |  1200/ 1452 batches | lr 4.00000 | ms/batch 209.04212 | loss  4.09 | ppl    59.66
## | epoch  41 |  1300/ 1452 batches | lr 4.00000 | ms/batch 209.95847 | loss  3.92 | ppl    50.38
## | epoch  41 |  1400/ 1452 batches | lr 4.00000 | ms/batch 211.03709 | loss  4.08 | ppl    59.03
## -----------------------------------------------------------------------------------------
## | end of epoch  41 | time: 322.29s | valid loss  4.61 | valid ppl   100.45
## | end of epoch  41 | time: 322.29s | test loss  4.56 | test ppl    95.47
## -----------------------------------------------------------------------------------------
## | epoch  42 |   100/ 1452 batches | lr 4.00000 | ms/batch 212.09435 | loss  4.16 | ppl    63.96
## | epoch  42 |   200/ 1452 batches | lr 4.00000 | ms/batch 210.90525 | loss  4.11 | ppl    61.10
## | epoch  42 |   300/ 1452 batches | lr 4.00000 | ms/batch 220.45823 | loss  4.11 | ppl    61.07
## | epoch  42 |   400/ 1452 batches | lr 4.00000 | ms/batch 207.89867 | loss  3.97 | ppl    52.91
## | epoch  42 |   500/ 1452 batches | lr 4.00000 | ms/batch 212.65358 | loss  4.00 | ppl    54.42
## | epoch  42 |   600/ 1452 batches | lr 4.00000 | ms/batch 209.50730 | loss  4.05 | ppl    57.59
## | epoch  42 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.09777 | loss  4.05 | ppl    57.55
## | epoch  42 |   800/ 1452 batches | lr 4.00000 | ms/batch 219.91530 | loss  4.02 | ppl    55.97
## | epoch  42 |   900/ 1452 batches | lr 4.00000 | ms/batch 206.05961 | loss  4.02 | ppl    55.62
## | epoch  42 |  1000/ 1452 batches | lr 4.00000 | ms/batch 207.74287 | loss  4.00 | ppl    54.58
## | epoch  42 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.16222 | loss  4.14 | ppl    62.74
## | epoch  42 |  1200/ 1452 batches | lr 4.00000 | ms/batch 212.08311 | loss  4.09 | ppl    59.51
## | epoch  42 |  1300/ 1452 batches | lr 4.00000 | ms/batch 211.56361 | loss  3.90 | ppl    49.57
## | epoch  42 |  1400/ 1452 batches | lr 4.00000 | ms/batch 206.17438 | loss  4.07 | ppl    58.57
## -----------------------------------------------------------------------------------------
## | end of epoch  42 | time: 324.42s | valid loss  4.61 | valid ppl   100.75
## | end of epoch  42 | time: 324.42s | test loss  4.56 | test ppl    95.31
## -----------------------------------------------------------------------------------------
## | epoch  43 |   100/ 1452 batches | lr 4.00000 | ms/batch 210.00027 | loss  4.15 | ppl    63.38
## | epoch  43 |   200/ 1452 batches | lr 4.00000 | ms/batch 209.05638 | loss  4.11 | ppl    60.75
## | epoch  43 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.50080 | loss  4.10 | ppl    60.09
## | epoch  43 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.15225 | loss  3.97 | ppl    52.72
## | epoch  43 |   500/ 1452 batches | lr 4.00000 | ms/batch 214.53099 | loss  3.98 | ppl    53.66
## | epoch  43 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.05840 | loss  4.03 | ppl    56.32
## | epoch  43 |   700/ 1452 batches | lr 4.00000 | ms/batch 208.39704 | loss  4.04 | ppl    56.88
## | epoch  43 |   800/ 1452 batches | lr 4.00000 | ms/batch 213.04235 | loss  4.01 | ppl    55.25
## | epoch  43 |   900/ 1452 batches | lr 4.00000 | ms/batch 216.54893 | loss  4.01 | ppl    55.26
## | epoch  43 |  1000/ 1452 batches | lr 4.00000 | ms/batch 209.53584 | loss  3.99 | ppl    54.30
## | epoch  43 |  1100/ 1452 batches | lr 4.00000 | ms/batch 208.72167 | loss  4.14 | ppl    62.49
## | epoch  43 |  1200/ 1452 batches | lr 4.00000 | ms/batch 213.04721 | loss  4.08 | ppl    58.98
## | epoch  43 |  1300/ 1452 batches | lr 4.00000 | ms/batch 207.31329 | loss  3.89 | ppl    48.92
## | epoch  43 |  1400/ 1452 batches | lr 4.00000 | ms/batch 208.54673 | loss  4.06 | ppl    57.69
## -----------------------------------------------------------------------------------------
## | end of epoch  43 | time: 323.48s | valid loss  4.61 | valid ppl   100.40
## | end of epoch  43 | time: 323.48s | test loss  4.56 | test ppl    95.73
## -----------------------------------------------------------------------------------------
## | epoch  44 |   100/ 1452 batches | lr 4.00000 | ms/batch 210.60483 | loss  4.13 | ppl    62.29
## | epoch  44 |   200/ 1452 batches | lr 4.00000 | ms/batch 208.84150 | loss  4.09 | ppl    59.64
## | epoch  44 |   300/ 1452 batches | lr 4.00000 | ms/batch 208.75161 | loss  4.08 | ppl    59.43
## | epoch  44 |   400/ 1452 batches | lr 4.00000 | ms/batch 206.95836 | loss  3.96 | ppl    52.45
## | epoch  44 |   500/ 1452 batches | lr 4.00000 | ms/batch 208.68658 | loss  3.97 | ppl    53.16
## | epoch  44 |   600/ 1452 batches | lr 4.00000 | ms/batch 207.21319 | loss  4.03 | ppl    56.44
## | epoch  44 |   700/ 1452 batches | lr 4.00000 | ms/batch 207.19797 | loss  4.03 | ppl    56.14
## | epoch  44 |   800/ 1452 batches | lr 4.00000 | ms/batch 212.72264 | loss  4.00 | ppl    54.52
## | epoch  44 |   900/ 1452 batches | lr 4.00000 | ms/batch 207.87773 | loss  4.01 | ppl    54.88
## | epoch  44 |  1000/ 1452 batches | lr 4.00000 | ms/batch 209.14132 | loss  3.98 | ppl    53.28
## | epoch  44 |  1100/ 1452 batches | lr 4.00000 | ms/batch 211.67881 | loss  4.13 | ppl    61.96
## | epoch  44 |  1200/ 1452 batches | lr 4.00000 | ms/batch 211.28919 | loss  4.07 | ppl    58.55
## | epoch  44 |  1300/ 1452 batches | lr 4.00000 | ms/batch 209.67587 | loss  3.88 | ppl    48.58
## | epoch  44 |  1400/ 1452 batches | lr 4.00000 | ms/batch 210.87480 | loss  4.05 | ppl    57.17
## -----------------------------------------------------------------------------------------
## | end of epoch  44 | time: 322.97s | valid loss  4.61 | valid ppl   100.42
## | end of epoch  44 | time: 322.97s | test loss  4.56 | test ppl    95.35
```

```
## -----------------------------------------------------------------------------------------
## | epoch  45 |   100/ 1452 batches | lr 4.00000 | ms/batch 208.27722 | loss  4.12 | ppl    61.79
## | epoch  45 |   200/ 1452 batches | lr 4.00000 | ms/batch 207.04862 | loss  4.07 | ppl    58.80
## | epoch  45 |   300/ 1452 batches | lr 4.00000 | ms/batch 209.25131 | loss  4.08 | ppl    59.08
## | epoch  45 |   400/ 1452 batches | lr 4.00000 | ms/batch 208.37237 | loss  3.95 | ppl    51.83
## | epoch  45 |   500/ 1452 batches | lr 4.00000 | ms/batch 207.42313 | loss  3.97 | ppl    52.98
## | epoch  45 |   600/ 1452 batches | lr 4.00000 | ms/batch 208.72695 | loss  4.01 | ppl    55.28
## | epoch  45 |   700/ 1452 batches | lr 4.00000 | ms/batch 209.52615 | loss  4.02 | ppl    55.97
## | epoch  45 |   800/ 1452 batches | lr 4.00000 | ms/batch 210.05054 | loss  3.99 | ppl    54.26
## | epoch  45 |   900/ 1452 batches | lr 4.00000 | ms/batch 208.29211 | loss  3.98 | ppl    53.56
## | epoch  45 |  1000/ 1452 batches | lr 4.00000 | ms/batch 208.68701 | loss  3.97 | ppl    53.01
## | epoch  45 |  1100/ 1452 batches | lr 4.00000 | ms/batch 211.93861 | loss  4.11 | ppl    60.98
## | epoch  45 |  1200/ 1452 batches | lr 4.00000 | ms/batch 207.52308 | loss  4.05 | ppl    57.45
## | epoch  45 |  1300/ 1452 batches | lr 4.00000 | ms/batch 209.56084 | loss  3.88 | ppl    48.37
## | epoch  45 |  1400/ 1452 batches | lr 4.00000 | ms/batch 207.55284 | loss  4.04 | ppl    56.95
## -----------------------------------------------------------------------------------------
## | end of epoch  45 | time: 322.01s | valid loss  4.62 | valid ppl   101.34
## | end of epoch  45 | time: 322.01s | test loss  4.57 | test ppl    96.24
## -----------------------------------------------------------------------------------------
## | epoch  46 |   100/ 1452 batches | lr 2.00000 | ms/batch 211.55392 | loss  4.09 | ppl    60.03
## | epoch  46 |   200/ 1452 batches | lr 2.00000 | ms/batch 210.88456 | loss  4.03 | ppl    56.36
## | epoch  46 |   300/ 1452 batches | lr 2.00000 | ms/batch 207.94752 | loss  4.02 | ppl    55.58
## | epoch  46 |   400/ 1452 batches | lr 2.00000 | ms/batch 212.37307 | loss  3.88 | ppl    48.42
## | epoch  46 |   500/ 1452 batches | lr 2.00000 | ms/batch 208.89167 | loss  3.89 | ppl    49.06
## | epoch  46 |   600/ 1452 batches | lr 2.00000 | ms/batch 211.08140 | loss  3.94 | ppl    51.52
## | epoch  46 |   700/ 1452 batches | lr 2.00000 | ms/batch 210.86453 | loss  3.94 | ppl    51.53
## | epoch  46 |   800/ 1452 batches | lr 2.00000 | ms/batch 207.28322 | loss  3.90 | ppl    49.42
## | epoch  46 |   900/ 1452 batches | lr 2.00000 | ms/batch 213.04341 | loss  3.89 | ppl    49.11
## | epoch  46 |  1000/ 1452 batches | lr 2.00000 | ms/batch 215.76510 | loss  3.88 | ppl    48.23
## | epoch  46 |  1100/ 1452 batches | lr 2.00000 | ms/batch 211.13379 | loss  4.00 | ppl    54.78
## | epoch  46 |  1200/ 1452 batches | lr 2.00000 | ms/batch 207.68336 | loss  3.95 | ppl    52.10
## | epoch  46 |  1300/ 1452 batches | lr 2.00000 | ms/batch 211.08435 | loss  3.77 | ppl    43.30
## | epoch  46 |  1400/ 1452 batches | lr 2.00000 | ms/batch 206.16446 | loss  3.93 | ppl    50.68
## -----------------------------------------------------------------------------------------
## | end of epoch  46 | time: 323.66s | valid loss  4.58 | valid ppl    97.16
## | end of epoch  46 | time: 323.66s | test loss  4.52 | test ppl    91.88
## -----------------------------------------------------------------------------------------
## Save model!
##
```

```
##
## | epoch  47 |   100/ 1452 batches | lr 2.00000 | ms/batch 208.03744 | loss  4.05 | ppl    57.20
## | epoch  47 |   200/ 1452 batches | lr 2.00000 | ms/batch 214.32629 | loss  3.99 | ppl    54.13
## | epoch  47 |   300/ 1452 batches | lr 2.00000 | ms/batch 204.66580 | loss  3.98 | ppl    53.52
## | epoch  47 |   400/ 1452 batches | lr 2.00000 | ms/batch 205.40035 | loss  3.86 | ppl    47.31
## | epoch  47 |   500/ 1452 batches | lr 2.00000 | ms/batch 216.62355 | loss  3.87 | ppl    47.80
## | epoch  47 |   600/ 1452 batches | lr 2.00000 | ms/batch 204.35610 | loss  3.91 | ppl    50.01
## | epoch  47 |   700/ 1452 batches | lr 2.00000 | ms/batch 205.63480 | loss  3.92 | ppl    50.60
## | epoch  47 |   800/ 1452 batches | lr 2.00000 | ms/batch 212.01842 | loss  3.88 | ppl    48.55
## | epoch  47 |   900/ 1452 batches | lr 2.00000 | ms/batch 206.11449 | loss  3.88 | ppl    48.36
## | epoch  47 |  1000/ 1452 batches | lr 2.00000 | ms/batch 208.28718 | loss  3.86 | ppl    47.49
## | epoch  47 |  1100/ 1452 batches | lr 2.00000 | ms/batch 207.42316 | loss  4.00 | ppl    54.56
## | epoch  47 |  1200/ 1452 batches | lr 2.00000 | ms/batch 211.81374 | loss  3.93 | ppl    51.11
## | epoch  47 |  1300/ 1452 batches | lr 2.00000 | ms/batch 207.56309 | loss  3.76 | ppl    43.12
## | epoch  47 |  1400/ 1452 batches | lr 2.00000 | ms/batch 208.96158 | loss  3.92 | ppl    50.42
## -----------------------------------------------------------------------------------------
## | end of epoch  47 | time: 321.40s | valid loss  4.57 | valid ppl    97.00
## | end of epoch  47 | time: 321.40s | test loss  4.52 | test ppl    91.94
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  48 |   100/ 1452 batches | lr 2.00000 | ms/batch 208.19726 | loss  4.03 | ppl    56.36
## | epoch  48 |   200/ 1452 batches | lr 2.00000 | ms/batch 207.79794 | loss  3.98 | ppl    53.48
## | epoch  48 |   300/ 1452 batches | lr 2.00000 | ms/batch 208.75103 | loss  3.97 | ppl    52.91
## | epoch  48 |   400/ 1452 batches | lr 2.00000 | ms/batch 207.71282 | loss  3.84 | ppl    46.57
## | epoch  48 |   500/ 1452 batches | lr 2.00000 | ms/batch 208.49691 | loss  3.86 | ppl    47.66
## | epoch  48 |   600/ 1452 batches | lr 2.00000 | ms/batch 208.47210 | loss  3.91 | ppl    49.95
## | epoch  48 |   700/ 1452 batches | lr 2.00000 | ms/batch 207.04823 | loss  3.91 | ppl    49.70
## | epoch  48 |   800/ 1452 batches | lr 2.00000 | ms/batch 209.29549 | loss  3.88 | ppl    48.52
## | epoch  48 |   900/ 1452 batches | lr 2.00000 | ms/batch 208.85657 | loss  3.87 | ppl    47.82
## | epoch  48 |  1000/ 1452 batches | lr 2.00000 | ms/batch 208.60682 | loss  3.85 | ppl    46.98
## | epoch  48 |  1100/ 1452 batches | lr 2.00000 | ms/batch 209.20639 | loss  3.98 | ppl    53.73
## | epoch  48 |  1200/ 1452 batches | lr 2.00000 | ms/batch 208.34710 | loss  3.93 | ppl    50.67
## | epoch  48 |  1300/ 1452 batches | lr 2.00000 | ms/batch 210.23524 | loss  3.75 | ppl    42.38
## | epoch  48 |  1400/ 1452 batches | lr 2.00000 | ms/batch 207.60821 | loss  3.91 | ppl    50.02
## -----------------------------------------------------------------------------------------
## | end of epoch  48 | time: 320.84s | valid loss  4.57 | valid ppl    96.92
## | end of epoch  48 | time: 320.84s | test loss  4.52 | test ppl    91.75
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  49 |   100/ 1452 batches | lr 2.00000 | ms/batch 208.12227 | loss  4.01 | ppl    54.93
## | epoch  49 |   200/ 1452 batches | lr 2.00000 | ms/batch 208.15737 | loss  3.96 | ppl    52.65
## | epoch  49 |   300/ 1452 batches | lr 2.00000 | ms/batch 212.52284 | loss  3.96 | ppl    52.72
## | epoch  49 |   400/ 1452 batches | lr 2.00000 | ms/batch 210.98458 | loss  3.83 | ppl    46.00
## | epoch  49 |   500/ 1452 batches | lr 2.00000 | ms/batch 208.38212 | loss  3.84 | ppl    46.74
## | epoch  49 |   600/ 1452 batches | lr 2.00000 | ms/batch 207.95752 | loss  3.90 | ppl    49.24
## | epoch  49 |   700/ 1452 batches | lr 2.00000 | ms/batch 207.90503 | loss  3.90 | ppl    49.20
## | epoch  49 |   800/ 1452 batches | lr 2.00000 | ms/batch 210.52489 | loss  3.86 | ppl    47.51
## | epoch  49 |   900/ 1452 batches | lr 2.00000 | ms/batch 208.25226 | loss  3.86 | ppl    47.55
## | epoch  49 |  1000/ 1452 batches | lr 2.00000 | ms/batch 207.61298 | loss  3.84 | ppl    46.42
## | epoch  49 |  1100/ 1452 batches | lr 2.00000 | ms/batch 211.17952 | loss  3.97 | ppl    53.21
## | epoch  49 |  1200/ 1452 batches | lr 2.00000 | ms/batch 217.97240 | loss  3.91 | ppl    49.97
## | epoch  49 |  1300/ 1452 batches | lr 2.00000 | ms/batch 210.97463 | loss  3.74 | ppl    42.06
## | epoch  49 |  1400/ 1452 batches | lr 2.00000 | ms/batch 215.82455 | loss  3.91 | ppl    49.70
## -----------------------------------------------------------------------------------------
## | end of epoch  49 | time: 323.75s | valid loss  4.57 | valid ppl    96.97
## | end of epoch  49 | time: 323.75s | test loss  4.52 | test ppl    91.76
## -----------------------------------------------------------------------------------------
## | epoch  50 |   100/ 1452 batches | lr 2.00000 | ms/batch 212.74790 | loss  4.00 | ppl    54.79
## | epoch  50 |   200/ 1452 batches | lr 2.00000 | ms/batch 216.27414 | loss  3.95 | ppl    52.04
## | epoch  50 |   300/ 1452 batches | lr 2.00000 | ms/batch 205.79987 | loss  3.95 | ppl    51.94
## | epoch  50 |   400/ 1452 batches | lr 2.00000 | ms/batch 207.89263 | loss  3.82 | ppl    45.58
## | epoch  50 |   500/ 1452 batches | lr 2.00000 | ms/batch 212.14343 | loss  3.84 | ppl    46.38
## | epoch  50 |   600/ 1452 batches | lr 2.00000 | ms/batch 213.92654 | loss  3.89 | ppl    48.75
## | epoch  50 |   700/ 1452 batches | lr 2.00000 | ms/batch 207.74756 | loss  3.88 | ppl    48.47
## | epoch  50 |   800/ 1452 batches | lr 2.00000 | ms/batch 207.50313 | loss  3.86 | ppl    47.41
## | epoch  50 |   900/ 1452 batches | lr 2.00000 | ms/batch 207.18821 | loss  3.85 | ppl    47.21
## | epoch  50 |  1000/ 1452 batches | lr 2.00000 | ms/batch 213.56178 | loss  3.83 | ppl    46.20
## | epoch  50 |  1100/ 1452 batches | lr 2.00000 | ms/batch 205.99931 | loss  3.97 | ppl    52.94
## | epoch  50 |  1200/ 1452 batches | lr 2.00000 | ms/batch 208.86663 | loss  3.91 | ppl    49.98
## | epoch  50 |  1300/ 1452 batches | lr 2.00000 | ms/batch 209.32104 | loss  3.73 | ppl    41.84
```

```
## | epoch  50 |  1400/ 1452 batches | lr 2.00000 | ms/batch 211.05435 | loss  3.90 | ppl    49.17
## -----------------------------------------------------------------------------------------
## | end of epoch  50 | time: 323.13s | valid loss  4.57 | valid ppl    96.88
## | end of epoch  50 | time: 323.13s | test loss  4.52 | test ppl    91.57
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  51 |   100/ 1452 batches | lr 2.00000 | ms/batch 208.31712 | loss  4.00 | ppl    54.52
## | epoch  51 |   200/ 1452 batches | lr 2.00000 | ms/batch 208.43727 | loss  3.95 | ppl    51.92
## | epoch  51 |   300/ 1452 batches | lr 2.00000 | ms/batch 216.09925 | loss  3.94 | ppl    51.27
## | epoch  51 |   400/ 1452 batches | lr 2.00000 | ms/batch 208.80153 | loss  3.81 | ppl    45.19
## | epoch  51 |   500/ 1452 batches | lr 2.00000 | ms/batch 206.11449 | loss  3.83 | ppl    45.99
## | epoch  51 |   600/ 1452 batches | lr 2.00000 | ms/batch 208.57707 | loss  3.88 | ppl    48.41
## | epoch  51 |   700/ 1452 batches | lr 2.00000 | ms/batch 208.72181 | loss  3.87 | ppl    48.18
## | epoch  51 |   800/ 1452 batches | lr 2.00000 | ms/batch 207.32801 | loss  3.84 | ppl    46.63
## | epoch  51 |   900/ 1452 batches | lr 2.00000 | ms/batch 210.53997 | loss  3.85 | ppl    46.85
## | epoch  51 |  1000/ 1452 batches | lr 2.00000 | ms/batch 212.64281 | loss  3.82 | ppl    45.44
## | epoch  51 |  1100/ 1452 batches | lr 2.00000 | ms/batch 208.58707 | loss  3.96 | ppl    52.67
## | epoch  51 |  1200/ 1452 batches | lr 2.00000 | ms/batch 208.76173 | loss  3.91 | ppl    49.69
## | epoch  51 |  1300/ 1452 batches | lr 2.00000 | ms/batch 207.84263 | loss  3.73 | ppl    41.72
## | epoch  51 |  1400/ 1452 batches | lr 2.00000 | ms/batch 209.71592 | loss  3.88 | ppl    48.66
## -----------------------------------------------------------------------------------------
## | end of epoch  51 | time: 322.14s | valid loss  4.57 | valid ppl    97.03
## | end of epoch  51 | time: 322.14s | test loss  4.52 | test ppl    91.74
## -----------------------------------------------------------------------------------------
## | epoch  52 |   100/ 1452 batches | lr 2.00000 | ms/batch 209.25630 | loss  3.99 | ppl    53.84
## | epoch  52 |   200/ 1452 batches | lr 2.00000 | ms/batch 207.41318 | loss  3.94 | ppl    51.21
## | epoch  52 |   300/ 1452 batches | lr 2.00000 | ms/batch 207.21355 | loss  3.93 | ppl    50.75
## | epoch  52 |   400/ 1452 batches | lr 2.00000 | ms/batch 208.14749 | loss  3.80 | ppl    44.67
## | epoch  52 |   500/ 1452 batches | lr 2.00000 | ms/batch 213.39711 | loss  3.82 | ppl    45.69
## | epoch  52 |   600/ 1452 batches | lr 2.00000 | ms/batch 206.30439 | loss  3.87 | ppl    48.07
## | epoch  52 |   700/ 1452 batches | lr 2.00000 | ms/batch 209.07617 | loss  3.87 | ppl    47.95
## | epoch  52 |   800/ 1452 batches | lr 2.00000 | ms/batch 209.42604 | loss  3.84 | ppl    46.66
## | epoch  52 |   900/ 1452 batches | lr 2.00000 | ms/batch 208.39690 | loss  3.84 | ppl    46.30
## | epoch  52 |  1000/ 1452 batches | lr 2.00000 | ms/batch 205.57485 | loss  3.82 | ppl    45.63
## | epoch  52 |  1100/ 1452 batches | lr 2.00000 | ms/batch 207.90884 | loss  3.96 | ppl    52.57
## | epoch  52 |  1200/ 1452 batches | lr 2.00000 | ms/batch 212.88741 | loss  3.90 | ppl    49.22
## | epoch  52 |  1300/ 1452 batches | lr 2.00000 | ms/batch 211.63886 | loss  3.72 | ppl    41.21
## | epoch  52 |  1400/ 1452 batches | lr 2.00000 | ms/batch 208.00260 | loss  3.88 | ppl    48.57
## -----------------------------------------------------------------------------------------
## | end of epoch  52 | time: 321.63s | valid loss  4.58 | valid ppl    97.20
## | end of epoch  52 | time: 321.63s | test loss  4.52 | test ppl    91.71
## -----------------------------------------------------------------------------------------
## | epoch  53 |   100/ 1452 batches | lr 1.00000 | ms/batch 210.76489 | loss  3.98 | ppl    53.76
## | epoch  53 |   200/ 1452 batches | lr 1.00000 | ms/batch 212.86561 | loss  3.93 | ppl    50.93
## | epoch  53 |   300/ 1452 batches | lr 1.00000 | ms/batch 209.27846 | loss  3.92 | ppl    50.39
## | epoch  53 |   400/ 1452 batches | lr 1.00000 | ms/batch 206.62405 | loss  3.79 | ppl    44.14
## | epoch  53 |   500/ 1452 batches | lr 1.00000 | ms/batch 206.75895 | loss  3.81 | ppl    44.94
## | epoch  53 |   600/ 1452 batches | lr 1.00000 | ms/batch 207.00870 | loss  3.85 | ppl    46.77
## | epoch  53 |   700/ 1452 batches | lr 1.00000 | ms/batch 206.92390 | loss  3.85 | ppl    46.79
## | epoch  53 |   800/ 1452 batches | lr 1.00000 | ms/batch 207.03358 | loss  3.82 | ppl    45.56
## | epoch  53 |   900/ 1452 batches | lr 1.00000 | ms/batch 210.65495 | loss  3.81 | ppl    45.04
## | epoch  53 |  1000/ 1452 batches | lr 1.00000 | ms/batch 205.09554 | loss  3.78 | ppl    43.99
## | epoch  53 |  1100/ 1452 batches | lr 1.00000 | ms/batch 206.92859 | loss  3.92 | ppl    50.35
## | epoch  53 |  1200/ 1452 batches | lr 1.00000 | ms/batch 207.02358 | loss  3.86 | ppl    47.24
## | epoch  53 |  1300/ 1452 batches | lr 1.00000 | ms/batch 206.22541 | loss  3.68 | ppl    39.49
## | epoch  53 |  1400/ 1452 batches | lr 1.00000 | ms/batch 209.97573 | loss  3.83 | ppl    46.14
## -----------------------------------------------------------------------------------------
## | end of epoch  53 | time: 320.33s | valid loss  4.56 | valid ppl    95.75
## | end of epoch  53 | time: 320.33s | test loss  4.50 | test ppl    90.31
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  54 |   100/ 1452 batches | lr 1.00000 | ms/batch 212.33307 | loss  3.95 | ppl    52.04
## | epoch  54 |   200/ 1452 batches | lr 1.00000 | ms/batch 205.84481 | loss  3.91 | ppl    49.91
## | epoch  54 |   300/ 1452 batches | lr 1.00000 | ms/batch 210.21549 | loss  3.90 | ppl    49.46
## | epoch  54 |   400/ 1452 batches | lr 1.00000 | ms/batch 206.45419 | loss  3.77 | ppl    43.20
## | epoch  54 |   500/ 1452 batches | lr 1.00000 | ms/batch 207.20846 | loss  3.79 | ppl    44.19
## | epoch  54 |   600/ 1452 batches | lr 1.00000 | ms/batch 207.93270 | loss  3.83 | ppl    46.26
## | epoch  54 |   700/ 1452 batches | lr 1.00000 | ms/batch 206.10792 | loss  3.83 | ppl    46.11
## | epoch  54 |   800/ 1452 batches | lr 1.00000 | ms/batch 206.46423 | loss  3.80 | ppl    44.83
## | epoch  54 |   900/ 1452 batches | lr 1.00000 | ms/batch 206.38408 | loss  3.81 | ppl    45.10
```

```
## | epoch  54 |  1000/ 1452 batches | lr 1.00000 | ms/batch 207.28332 | loss  3.78 | ppl    43.72
## | epoch  54 |  1100/ 1452 batches | lr 1.00000 | ms/batch 213.50847 | loss  3.91 | ppl    50.05
## | epoch  54 |  1200/ 1452 batches | lr 1.00000 | ms/batch 205.63699 | loss  3.85 | ppl    47.09
## | epoch  54 |  1300/ 1452 batches | lr 1.00000 | ms/batch 213.39708 | loss  3.66 | ppl    38.97
## | epoch  54 |  1400/ 1452 batches | lr 1.00000 | ms/batch 206.87399 | loss  3.83 | ppl    45.86
## -----------------------------------------------------------------------------------------
## | end of epoch  54 | time: 321.84s | valid loss  4.56 | valid ppl    95.43
## | end of epoch  54 | time: 321.84s | test loss  4.50 | test ppl    89.99
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  55 |   100/ 1452 batches | lr 1.00000 | ms/batch 206.25929 | loss  3.95 | ppl    51.83
## | epoch  55 |   200/ 1452 batches | lr 1.00000 | ms/batch 213.35204 | loss  3.90 | ppl    49.61
## | epoch  55 |   300/ 1452 batches | lr 1.00000 | ms/batch 206.53390 | loss  3.89 | ppl    48.72
## | epoch  55 |   400/ 1452 batches | lr 1.00000 | ms/batch 206.67879 | loss  3.76 | ppl    43.01
## | epoch  55 |   500/ 1452 batches | lr 1.00000 | ms/batch 207.48780 | loss  3.79 | ppl    44.22
## | epoch  55 |   600/ 1452 batches | lr 1.00000 | ms/batch 208.13762 | loss  3.82 | ppl    45.66
## | epoch  55 |   700/ 1452 batches | lr 1.00000 | ms/batch 207.12815 | loss  3.83 | ppl    46.09
## | epoch  55 |   800/ 1452 batches | lr 1.00000 | ms/batch 211.05927 | loss  3.80 | ppl    44.76
## | epoch  55 |   900/ 1452 batches | lr 1.00000 | ms/batch 208.51676 | loss  3.80 | ppl    44.55
## | epoch  55 |  1000/ 1452 batches | lr 1.00000 | ms/batch 206.50392 | loss  3.77 | ppl    43.36
## | epoch  55 |  1100/ 1452 batches | lr 1.00000 | ms/batch 209.96036 | loss  3.91 | ppl    49.76
## | epoch  55 |  1200/ 1452 batches | lr 1.00000 | ms/batch 206.52416 | loss  3.84 | ppl    46.51
## | epoch  55 |  1300/ 1452 batches | lr 1.00000 | ms/batch 208.35226 | loss  3.67 | ppl    39.19
## | epoch  55 |  1400/ 1452 batches | lr 1.00000 | ms/batch 210.58487 | loss  3.82 | ppl    45.66
## -----------------------------------------------------------------------------------------
## | end of epoch  55 | time: 320.74s | valid loss  4.56 | valid ppl    95.62
## | end of epoch  55 | time: 320.74s | test loss  4.50 | test ppl    90.08
## -----------------------------------------------------------------------------------------
## | epoch  56 |   100/ 1452 batches | lr 1.00000 | ms/batch 209.83774 | loss  3.94 | ppl    51.67
## | epoch  56 |   200/ 1452 batches | lr 1.00000 | ms/batch 206.74371 | loss  3.90 | ppl    49.21
## | epoch  56 |   300/ 1452 batches | lr 1.00000 | ms/batch 207.58790 | loss  3.89 | ppl    48.86
## | epoch  56 |   400/ 1452 batches | lr 1.00000 | ms/batch 206.91884 | loss  3.75 | ppl    42.65
## | epoch  56 |   500/ 1452 batches | lr 1.00000 | ms/batch 207.93756 | loss  3.78 | ppl    43.75
## | epoch  56 |   600/ 1452 batches | lr 1.00000 | ms/batch 207.34841 | loss  3.82 | ppl    45.44
## | epoch  56 |   700/ 1452 batches | lr 1.00000 | ms/batch 208.57708 | loss  3.82 | ppl    45.67
## | epoch  56 |   800/ 1452 batches | lr 1.00000 | ms/batch 207.10359 | loss  3.79 | ppl    44.22
## | epoch  56 |   900/ 1452 batches | lr 1.00000 | ms/batch 208.32205 | loss  3.78 | ppl    44.03
## | epoch  56 |  1000/ 1452 batches | lr 1.00000 | ms/batch 207.66970 | loss  3.76 | ppl    42.93
## | epoch  56 |  1100/ 1452 batches | lr 1.00000 | ms/batch 208.94671 | loss  3.90 | ppl    49.28
## | epoch  56 |  1200/ 1452 batches | lr 1.00000 | ms/batch 207.54289 | loss  3.84 | ppl    46.64
## | epoch  56 |  1300/ 1452 batches | lr 1.00000 | ms/batch 208.42192 | loss  3.66 | ppl    39.00
## | epoch  56 |  1400/ 1452 batches | lr 1.00000 | ms/batch 210.80440 | loss  3.82 | ppl    45.51
## -----------------------------------------------------------------------------------------
## | end of epoch  56 | time: 320.46s | valid loss  4.56 | valid ppl    95.52
## | end of epoch  56 | time: 320.46s | test loss  4.50 | test ppl    89.98
## -----------------------------------------------------------------------------------------
## | epoch  57 |   100/ 1452 batches | lr 1.00000 | ms/batch 214.18289 | loss  3.93 | ppl    51.15
## | epoch  57 |   200/ 1452 batches | lr 1.00000 | ms/batch 206.66399 | loss  3.89 | ppl    48.76
## | epoch  57 |   300/ 1452 batches | lr 1.00000 | ms/batch 206.68395 | loss  3.88 | ppl    48.34
## | epoch  57 |   400/ 1452 batches | lr 1.00000 | ms/batch 207.00829 | loss  3.75 | ppl    42.52
## | epoch  57 |   500/ 1452 batches | lr 1.00000 | ms/batch 207.55793 | loss  3.76 | ppl    43.11
## | epoch  57 |   600/ 1452 batches | lr 1.00000 | ms/batch 206.77878 | loss  3.81 | ppl    45.25
## | epoch  57 |   700/ 1452 batches | lr 1.00000 | ms/batch 206.41930 | loss  3.82 | ppl    45.47
## | epoch  57 |   800/ 1452 batches | lr 1.00000 | ms/batch 206.98342 | loss  3.78 | ppl    43.90
## | epoch  57 |   900/ 1452 batches | lr 1.00000 | ms/batch 206.34917 | loss  3.78 | ppl    43.75
## | epoch  57 |  1000/ 1452 batches | lr 1.00000 | ms/batch 206.71859 | loss  3.76 | ppl    42.85
## | epoch  57 |  1100/ 1452 batches | lr 1.00000 | ms/batch 208.21704 | loss  3.89 | ppl    49.13
## | epoch  57 |  1200/ 1452 batches | lr 1.00000 | ms/batch 206.38407 | loss  3.84 | ppl    46.61
## | epoch  57 |  1300/ 1452 batches | lr 1.00000 | ms/batch 217.82721 | loss  3.65 | ppl    38.64
## | epoch  57 |  1400/ 1452 batches | lr 1.00000 | ms/batch 208.28699 | loss  3.83 | ppl    45.87
## -----------------------------------------------------------------------------------------
## | end of epoch  57 | time: 320.66s | valid loss  4.56 | valid ppl    95.86
## | end of epoch  57 | time: 320.66s | test loss  4.50 | test ppl    90.19
## -----------------------------------------------------------------------------------------
## | epoch  58 |   100/ 1452 batches | lr 1.00000 | ms/batch 210.38506 | loss  3.93 | ppl    50.68
## | epoch  58 |   200/ 1452 batches | lr 1.00000 | ms/batch 210.12558 | loss  3.88 | ppl    48.44
## | epoch  58 |   300/ 1452 batches | lr 1.00000 | ms/batch 208.17221 | loss  3.87 | ppl    47.89
## | epoch  58 |   400/ 1452 batches | lr 1.00000 | ms/batch 207.98729 | loss  3.74 | ppl    42.16
## | epoch  58 |   500/ 1452 batches | lr 1.00000 | ms/batch 209.82080 | loss  3.76 | ppl    42.86
## | epoch  58 |   600/ 1452 batches | lr 1.00000 | ms/batch 207.60787 | loss  3.81 | ppl    45.01
## | epoch  58 |   700/ 1452 batches | lr 1.00000 | ms/batch 209.36099 | loss  3.81 | ppl    45.01
```

```
## | epoch  58 |   800/ 1452 batches | lr 1.00000 | ms/batch 207.30023 | loss  3.78 | ppl    43.80
## | epoch  58 |   900/ 1452 batches | lr 1.00000 | ms/batch 219.13100 | loss  3.78 | ppl    43.84
## | epoch  58 |  1000/ 1452 batches | lr 1.00000 | ms/batch 205.58483 | loss  3.75 | ppl    42.53
## | epoch  58 |  1100/ 1452 batches | lr 1.00000 | ms/batch 207.06333 | loss  3.89 | ppl    49.05
## | epoch  58 |  1200/ 1452 batches | lr 1.00000 | ms/batch 207.93246 | loss  3.83 | ppl    46.26
## | epoch  58 |  1300/ 1452 batches | lr 1.00000 | ms/batch 206.31900 | loss  3.66 | ppl    38.71
## | epoch  58 |  1400/ 1452 batches | lr 1.00000 | ms/batch 209.67577 | loss  3.81 | ppl    45.01
## -----------------------------------------------------------------------------------------
## | end of epoch  58 | time: 321.90s | valid loss  4.56 | valid ppl    95.49
## | end of epoch  58 | time: 321.90s | test loss  4.50 | test ppl    89.96
## -----------------------------------------------------------------------------------------
## | epoch  59 |   100/ 1452 batches | lr 1.00000 | ms/batch 210.02057 | loss  3.93 | ppl    50.79
## | epoch  59 |   200/ 1452 batches | lr 1.00000 | ms/batch 209.04668 | loss  3.87 | ppl    48.16
## | epoch  59 |   300/ 1452 batches | lr 1.00000 | ms/batch 208.21137 | loss  3.87 | ppl    47.75
## | epoch  59 |   400/ 1452 batches | lr 1.00000 | ms/batch 208.44702 | loss  3.74 | ppl    42.22
## | epoch  59 |   500/ 1452 batches | lr 1.00000 | ms/batch 214.35127 | loss  3.76 | ppl    42.77
## | epoch  59 |   600/ 1452 batches | lr 1.00000 | ms/batch 207.47832 | loss  3.80 | ppl    44.79
## | epoch  59 |   700/ 1452 batches | lr 1.00000 | ms/batch 207.19852 | loss  3.80 | ppl    44.89
## | epoch  59 |   800/ 1452 batches | lr 1.00000 | ms/batch 210.22520 | loss  3.78 | ppl    43.94
## | epoch  59 |   900/ 1452 batches | lr 1.00000 | ms/batch 208.09251 | loss  3.78 | ppl    43.81
## | epoch  59 |  1000/ 1452 batches | lr 1.00000 | ms/batch 209.02644 | loss  3.76 | ppl    42.75
## | epoch  59 |  1100/ 1452 batches | lr 1.00000 | ms/batch 210.40502 | loss  3.89 | ppl    48.78
## | epoch  59 |  1200/ 1452 batches | lr 1.00000 | ms/batch 207.28856 | loss  3.82 | ppl    45.78
## | epoch  59 |  1300/ 1452 batches | lr 1.00000 | ms/batch 211.18452 | loss  3.65 | ppl    38.40
## | epoch  59 |  1400/ 1452 batches | lr 1.00000 | ms/batch 207.72773 | loss  3.81 | ppl    45.01
## -----------------------------------------------------------------------------------------
## | end of epoch  59 | time: 322.10s | valid loss  4.56 | valid ppl    96.05
## | end of epoch  59 | time: 322.10s | test loss  4.50 | test ppl    90.39
## -----------------------------------------------------------------------------------------
## | epoch  60 |   100/ 1452 batches | lr 0.50000 | ms/batch 211.34891 | loss  3.93 | ppl    50.71
## | epoch  60 |   200/ 1452 batches | lr 0.50000 | ms/batch 208.08721 | loss  3.88 | ppl    48.19
## | epoch  60 |   300/ 1452 batches | lr 0.50000 | ms/batch 217.18322 | loss  3.87 | ppl    47.90
## | epoch  60 |   400/ 1452 batches | lr 0.50000 | ms/batch 206.96365 | loss  3.73 | ppl    41.83
## | epoch  60 |   500/ 1452 batches | lr 0.50000 | ms/batch 207.61295 | loss  3.76 | ppl    43.04
## | epoch  60 |   600/ 1452 batches | lr 0.50000 | ms/batch 210.23993 | loss  3.80 | ppl    44.59
## | epoch  60 |   700/ 1452 batches | lr 0.50000 | ms/batch 207.62329 | loss  3.80 | ppl    44.77
## | epoch  60 |   800/ 1452 batches | lr 0.50000 | ms/batch 208.54210 | loss  3.77 | ppl    43.42
## | epoch  60 |   900/ 1452 batches | lr 0.50000 | ms/batch 207.33796 | loss  3.77 | ppl    43.49
## | epoch  60 |  1000/ 1452 batches | lr 0.50000 | ms/batch 208.53189 | loss  3.74 | ppl    42.11
## | epoch  60 |  1100/ 1452 batches | lr 0.50000 | ms/batch 210.22027 | loss  3.87 | ppl    47.87
## | epoch  60 |  1200/ 1452 batches | lr 0.50000 | ms/batch 207.65274 | loss  3.82 | ppl    45.67
## | epoch  60 |  1300/ 1452 batches | lr 0.50000 | ms/batch 207.10818 | loss  3.63 | ppl    37.81
## | epoch  60 |  1400/ 1452 batches | lr 0.50000 | ms/batch 207.92247 | loss  3.79 | ppl    44.08
## -----------------------------------------------------------------------------------------
## | end of epoch  60 | time: 322.00s | valid loss  4.55 | valid ppl    94.63
## | end of epoch  60 | time: 322.00s | test loss  4.49 | test ppl    89.01
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  61 |   100/ 1452 batches | lr 0.50000 | ms/batch 217.11326 | loss  3.91 | ppl    50.09
## | epoch  61 |   200/ 1452 batches | lr 0.50000 | ms/batch 207.52801 | loss  3.87 | ppl    47.71
## | epoch  61 |   300/ 1452 batches | lr 0.50000 | ms/batch 210.23037 | loss  3.86 | ppl    47.47
## | epoch  61 |   400/ 1452 batches | lr 0.50000 | ms/batch 208.76842 | loss  3.72 | ppl    41.26
## | epoch  61 |   500/ 1452 batches | lr 0.50000 | ms/batch 207.53822 | loss  3.75 | ppl    42.53
## | epoch  61 |   600/ 1452 batches | lr 0.50000 | ms/batch 212.69778 | loss  3.79 | ppl    44.31
## | epoch  61 |   700/ 1452 batches | lr 0.50000 | ms/batch 207.16835 | loss  3.80 | ppl    44.68
## | epoch  61 |   800/ 1452 batches | lr 0.50000 | ms/batch 208.57796 | loss  3.76 | ppl    42.90
## | epoch  61 |   900/ 1452 batches | lr 0.50000 | ms/batch 210.06564 | loss  3.77 | ppl    43.29
## | epoch  61 |  1000/ 1452 batches | lr 0.50000 | ms/batch 208.67689 | loss  3.74 | ppl    42.02
## | epoch  61 |  1100/ 1452 batches | lr 0.50000 | ms/batch 208.33717 | loss  3.86 | ppl    47.65
## | epoch  61 |  1200/ 1452 batches | lr 0.50000 | ms/batch 210.25520 | loss  3.81 | ppl    45.10
## | epoch  61 |  1300/ 1452 batches | lr 0.50000 | ms/batch 208.38688 | loss  3.63 | ppl    37.62
## | epoch  61 |  1400/ 1452 batches | lr 0.50000 | ms/batch 210.39503 | loss  3.79 | ppl    44.19
## -----------------------------------------------------------------------------------------
## | end of epoch  61 | time: 322.82s | valid loss  4.55 | valid ppl    94.72
## | end of epoch  61 | time: 322.82s | test loss  4.49 | test ppl    89.11
## -----------------------------------------------------------------------------------------
## | epoch  62 |   100/ 1452 batches | lr 0.50000 | ms/batch 211.28434 | loss  3.91 | ppl    50.02
## | epoch  62 |   200/ 1452 batches | lr 0.50000 | ms/batch 207.58282 | loss  3.86 | ppl    47.60
## | epoch  62 |   300/ 1452 batches | lr 0.50000 | ms/batch 208.15786 | loss  3.86 | ppl    47.33
## | epoch  62 |   400/ 1452 batches | lr 0.50000 | ms/batch 211.63377 | loss  3.72 | ppl    41.22
## | epoch  62 |   500/ 1452 batches | lr 0.50000 | ms/batch 207.75752 | loss  3.75 | ppl    42.39
```

```
## | epoch  62 |   600/ 1452 batches | lr 0.50000 | ms/batch 210.09535 | loss  3.78 | ppl    43.98
## | epoch  62 |   700/ 1452 batches | lr 0.50000 | ms/batch 208.80157 | loss  3.79 | ppl    44.45
## | epoch  62 |   800/ 1452 batches | lr 0.50000 | ms/batch 207.51308 | loss  3.76 | ppl    43.16
## | epoch  62 |   900/ 1452 batches | lr 0.50000 | ms/batch 209.96036 | loss  3.76 | ppl    43.07
## | epoch  62 |  1000/ 1452 batches | lr 0.50000 | ms/batch 217.30796 | loss  3.73 | ppl    41.84
## | epoch  62 |  1100/ 1452 batches | lr 0.50000 | ms/batch 207.88253 | loss  3.87 | ppl    48.02
## | epoch  62 |  1200/ 1452 batches | lr 0.50000 | ms/batch 206.79850 | loss  3.81 | ppl    45.04
## | epoch  62 |  1300/ 1452 batches | lr 0.50000 | ms/batch 206.20911 | loss  3.63 | ppl    37.88
## | epoch  62 |  1400/ 1452 batches | lr 0.50000 | ms/batch 207.13816 | loss  3.79 | ppl    44.24
## -----------------------------------------------------------------------------------------
## | end of epoch  62 | time: 322.09s | valid loss  4.55 | valid ppl    94.69
## | end of epoch  62 | time: 322.09s | test loss  4.49 | test ppl    89.14
## -----------------------------------------------------------------------------------------
## | epoch  63 |   100/ 1452 batches | lr 0.50000 | ms/batch 217.78253 | loss  3.91 | ppl    49.76
## | epoch  63 |   200/ 1452 batches | lr 0.50000 | ms/batch 206.73898 | loss  3.86 | ppl    47.63
## | epoch  63 |   300/ 1452 batches | lr 0.50000 | ms/batch 208.95654 | loss  3.85 | ppl    47.20
## | epoch  63 |   400/ 1452 batches | lr 0.50000 | ms/batch 208.75647 | loss  3.72 | ppl    41.16
## | epoch  63 |   500/ 1452 batches | lr 0.50000 | ms/batch 208.62690 | loss  3.74 | ppl    42.05
## | epoch  63 |   600/ 1452 batches | lr 0.50000 | ms/batch 208.43705 | loss  3.79 | ppl    44.09
## | epoch  63 |   700/ 1452 batches | lr 0.50000 | ms/batch 208.01574 | loss  3.78 | ppl    43.97
## | epoch  63 |   800/ 1452 batches | lr 0.50000 | ms/batch 214.77054 | loss  3.76 | ppl    43.08
## | epoch  63 |   900/ 1452 batches | lr 0.50000 | ms/batch 208.76692 | loss  3.77 | ppl    43.44
## | epoch  63 |  1000/ 1452 batches | lr 0.50000 | ms/batch 209.02669 | loss  3.73 | ppl    41.55
## | epoch  63 |  1100/ 1452 batches | lr 0.50000 | ms/batch 214.12100 | loss  3.86 | ppl    47.57
## | epoch  63 |  1200/ 1452 batches | lr 0.50000 | ms/batch 207.17811 | loss  3.81 | ppl    45.25
## | epoch  63 |  1300/ 1452 batches | lr 0.50000 | ms/batch 207.26802 | loss  3.62 | ppl    37.32
## | epoch  63 |  1400/ 1452 batches | lr 0.50000 | ms/batch 209.14611 | loss  3.79 | ppl    44.08
## -----------------------------------------------------------------------------------------
## | end of epoch  63 | time: 322.97s | valid loss  4.55 | valid ppl    94.65
## | end of epoch  63 | time: 322.97s | test loss  4.49 | test ppl    89.08
## -----------------------------------------------------------------------------------------
## | epoch  64 |   100/ 1452 batches | lr 0.50000 | ms/batch 212.67765 | loss  3.90 | ppl    49.47
## | epoch  64 |   200/ 1452 batches | lr 0.50000 | ms/batch 209.20603 | loss  3.86 | ppl    47.64
## | epoch  64 |   300/ 1452 batches | lr 0.50000 | ms/batch 212.01827 | loss  3.85 | ppl    47.10
## | epoch  64 |   400/ 1452 batches | lr 0.50000 | ms/batch 207.53309 | loss  3.72 | ppl    41.22
## | epoch  64 |   500/ 1452 batches | lr 0.50000 | ms/batch 208.41195 | loss  3.74 | ppl    42.01
## | epoch  64 |   600/ 1452 batches | lr 0.50000 | ms/batch 207.97758 | loss  3.78 | ppl    43.96
## | epoch  64 |   700/ 1452 batches | lr 0.50000 | ms/batch 208.83663 | loss  3.79 | ppl    44.09
## | epoch  64 |   800/ 1452 batches | lr 0.50000 | ms/batch 208.12713 | loss  3.76 | ppl    42.87
## | epoch  64 |   900/ 1452 batches | lr 0.50000 | ms/batch 208.26703 | loss  3.75 | ppl    42.68
## | epoch  64 |  1000/ 1452 batches | lr 0.50000 | ms/batch 208.93140 | loss  3.72 | ppl    41.43
## | epoch  64 |  1100/ 1452 batches | lr 0.50000 | ms/batch 209.71552 | loss  3.86 | ppl    47.45
## | epoch  64 |  1200/ 1452 batches | lr 0.50000 | ms/batch 207.75772 | loss  3.80 | ppl    44.91
## | epoch  64 |  1300/ 1452 batches | lr 0.50000 | ms/batch 208.03725 | loss  3.62 | ppl    37.16
## | epoch  64 |  1400/ 1452 batches | lr 0.50000 | ms/batch 209.51111 | loss  3.79 | ppl    44.15
## -----------------------------------------------------------------------------------------
## | end of epoch  64 | time: 322.00s | valid loss  4.55 | valid ppl    94.68
## | end of epoch  64 | time: 322.00s | test loss  4.49 | test ppl    89.17
## -----------------------------------------------------------------------------------------
## | epoch  65 |   100/ 1452 batches | lr 0.50000 | ms/batch 213.86128 | loss  3.89 | ppl    49.15
## | epoch  65 |   200/ 1452 batches | lr 0.50000 | ms/batch 208.36711 | loss  3.86 | ppl    47.27
## | epoch  65 |   300/ 1452 batches | lr 0.50000 | ms/batch 211.88860 | loss  3.84 | ppl    46.73
## | epoch  65 |   400/ 1452 batches | lr 0.50000 | ms/batch 208.54204 | loss  3.71 | ppl    40.88
## | epoch  65 |   500/ 1452 batches | lr 0.50000 | ms/batch 208.45729 | loss  3.74 | ppl    42.11
## | epoch  65 |   600/ 1452 batches | lr 0.50000 | ms/batch 207.91245 | loss  3.78 | ppl    43.64
## | epoch  65 |   700/ 1452 batches | lr 0.50000 | ms/batch 209.43108 | loss  3.78 | ppl    44.02
## | epoch  65 |   800/ 1452 batches | lr 0.50000 | ms/batch 208.71677 | loss  3.75 | ppl    42.52
## | epoch  65 |   900/ 1452 batches | lr 0.50000 | ms/batch 211.02969 | loss  3.76 | ppl    42.88
## | epoch  65 |  1000/ 1452 batches | lr 0.50000 | ms/batch 207.66298 | loss  3.73 | ppl    41.49
## | epoch  65 |  1100/ 1452 batches | lr 0.50000 | ms/batch 214.09141 | loss  3.86 | ppl    47.56
## | epoch  65 |  1200/ 1452 batches | lr 0.50000 | ms/batch 208.67682 | loss  3.81 | ppl    45.05
## | epoch  65 |  1300/ 1452 batches | lr 0.50000 | ms/batch 210.10549 | loss  3.61 | ppl    37.15
## | epoch  65 |  1400/ 1452 batches | lr 0.50000 | ms/batch 208.07252 | loss  3.78 | ppl    43.73
## -----------------------------------------------------------------------------------------
## | end of epoch  65 | time: 323.07s | valid loss  4.55 | valid ppl    94.83
## | end of epoch  65 | time: 323.07s | test loss  4.49 | test ppl    89.24
## -----------------------------------------------------------------------------------------
## | epoch  66 |   100/ 1452 batches | lr 0.25000 | ms/batch 211.21950 | loss  3.90 | ppl    49.58
## | epoch  66 |   200/ 1452 batches | lr 0.25000 | ms/batch 209.34105 | loss  3.86 | ppl    47.70
## | epoch  66 |   300/ 1452 batches | lr 0.25000 | ms/batch 209.24113 | loss  3.85 | ppl    46.99
## | epoch  66 |   400/ 1452 batches | lr 0.25000 | ms/batch 208.62203 | loss  3.72 | ppl    41.07
## | epoch  66 |   500/ 1452 batches | lr 0.25000 | ms/batch 210.28007 | loss  3.74 | ppl    42.10
```

```
## | epoch  66 |   600/ 1452 batches | lr 0.25000 | ms/batch 207.90737 | loss  3.79 | ppl    44.15
## | epoch  66 |   700/ 1452 batches | lr 0.25000 | ms/batch 208.83684 | loss  3.79 | ppl    44.16
## | epoch  66 |   800/ 1452 batches | lr 0.25000 | ms/batch 216.12922 | loss  3.75 | ppl    42.60
## | epoch  66 |   900/ 1452 batches | lr 0.25000 | ms/batch 207.59300 | loss  3.76 | ppl    42.86
## | epoch  66 |  1000/ 1452 batches | lr 0.25000 | ms/batch 208.93398 | loss  3.73 | ppl    41.60
## | epoch  66 |  1100/ 1452 batches | lr 0.25000 | ms/batch 208.41710 | loss  3.86 | ppl    47.56
## | epoch  66 |  1200/ 1452 batches | lr 0.25000 | ms/batch 208.56204 | loss  3.80 | ppl    44.53
## | epoch  66 |  1300/ 1452 batches | lr 0.25000 | ms/batch 215.21120 | loss  3.61 | ppl    37.02
## | epoch  66 |  1400/ 1452 batches | lr 0.25000 | ms/batch 209.43113 | loss  3.77 | ppl    43.57
## -----------------------------------------------------------------------------------------
## | end of epoch  66 | time: 323.18s | valid loss  4.54 | valid ppl    94.14
## | end of epoch  66 | time: 323.18s | test loss  4.48 | test ppl    88.55
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  67 |   100/ 1452 batches | lr 0.25000 | ms/batch 211.03994 | loss  3.90 | ppl    49.31
## | epoch  67 |   200/ 1452 batches | lr 0.25000 | ms/batch 209.52594 | loss  3.86 | ppl    47.32
## | epoch  67 |   300/ 1452 batches | lr 0.25000 | ms/batch 207.94251 | loss  3.85 | ppl    46.89
## | epoch  67 |   400/ 1452 batches | lr 0.25000 | ms/batch 208.41678 | loss  3.71 | ppl    40.69
## | epoch  67 |   500/ 1452 batches | lr 0.25000 | ms/batch 209.75568 | loss  3.74 | ppl    41.98
## | epoch  67 |   600/ 1452 batches | lr 0.25000 | ms/batch 208.65173 | loss  3.78 | ppl    43.83
## | epoch  67 |   700/ 1452 batches | lr 0.25000 | ms/batch 212.12096 | loss  3.79 | ppl    44.11
## | epoch  67 |   800/ 1452 batches | lr 0.25000 | ms/batch 208.98141 | loss  3.75 | ppl    42.43
## | epoch  67 |   900/ 1452 batches | lr 0.25000 | ms/batch 207.69780 | loss  3.75 | ppl    42.64
## | epoch  67 |  1000/ 1452 batches | lr 0.25000 | ms/batch 208.06231 | loss  3.72 | ppl    41.23
## | epoch  67 |  1100/ 1452 batches | lr 0.25000 | ms/batch 210.74473 | loss  3.85 | ppl    47.10
## | epoch  67 |  1200/ 1452 batches | lr 0.25000 | ms/batch 208.78666 | loss  3.80 | ppl    44.78
## | epoch  67 |  1300/ 1452 batches | lr 0.25000 | ms/batch 211.74366 | loss  3.61 | ppl    36.98
## | epoch  67 |  1400/ 1452 batches | lr 0.25000 | ms/batch 207.76287 | loss  3.77 | ppl    43.37
## -----------------------------------------------------------------------------------------
## | end of epoch  67 | time: 323.05s | valid loss  4.55 | valid ppl    94.20
## | end of epoch  67 | time: 323.05s | test loss  4.48 | test ppl    88.58
## -----------------------------------------------------------------------------------------
## | epoch  68 |   100/ 1452 batches | lr 0.25000 | ms/batch 209.46593 | loss  3.89 | ppl    49.09
## | epoch  68 |   200/ 1452 batches | lr 0.25000 | ms/batch 208.58175 | loss  3.85 | ppl    47.02
## | epoch  68 |   300/ 1452 batches | lr 0.25000 | ms/batch 207.92735 | loss  3.84 | ppl    46.74
## | epoch  68 |   400/ 1452 batches | lr 0.25000 | ms/batch 209.24601 | loss  3.71 | ppl    40.77
## | epoch  68 |   500/ 1452 batches | lr 0.25000 | ms/batch 208.03234 | loss  3.74 | ppl    42.12
## | epoch  68 |   600/ 1452 batches | lr 0.25000 | ms/batch 208.22748 | loss  3.78 | ppl    43.90
## | epoch  68 |   700/ 1452 batches | lr 0.25000 | ms/batch 210.19022 | loss  3.78 | ppl    43.68
## | epoch  68 |   800/ 1452 batches | lr 0.25000 | ms/batch 209.61600 | loss  3.74 | ppl    42.28
## | epoch  68 |   900/ 1452 batches | lr 0.25000 | ms/batch 208.19730 | loss  3.76 | ppl    42.94
## | epoch  68 |  1000/ 1452 batches | lr 0.25000 | ms/batch 210.04561 | loss  3.72 | ppl    41.29
## | epoch  68 |  1100/ 1452 batches | lr 0.25000 | ms/batch 210.88481 | loss  3.86 | ppl    47.30
## | epoch  68 |  1200/ 1452 batches | lr 0.25000 | ms/batch 210.97770 | loss  3.81 | ppl    45.14
## | epoch  68 |  1300/ 1452 batches | lr 0.25000 | ms/batch 209.64654 | loss  3.61 | ppl    36.82
## | epoch  68 |  1400/ 1452 batches | lr 0.25000 | ms/batch 209.75576 | loss  3.77 | ppl    43.58
## -----------------------------------------------------------------------------------------
## | end of epoch  68 | time: 322.53s | valid loss  4.54 | valid ppl    94.11
## | end of epoch  68 | time: 322.53s | test loss  4.48 | test ppl    88.43
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  69 |   100/ 1452 batches | lr 0.25000 | ms/batch 209.12138 | loss  3.89 | ppl    49.05
## | epoch  69 |   200/ 1452 batches | lr 0.25000 | ms/batch 211.05946 | loss  3.86 | ppl    47.41
## | epoch  69 |   300/ 1452 batches | lr 0.25000 | ms/batch 210.71959 | loss  3.84 | ppl    46.62
## | epoch  69 |   400/ 1452 batches | lr 0.25000 | ms/batch 206.57895 | loss  3.70 | ppl    40.36
## | epoch  69 |   500/ 1452 batches | lr 0.25000 | ms/batch 207.71268 | loss  3.73 | ppl    41.57
## | epoch  69 |   600/ 1452 batches | lr 0.25000 | ms/batch 217.79235 | loss  3.77 | ppl    43.38
## | epoch  69 |   700/ 1452 batches | lr 0.25000 | ms/batch 204.83070 | loss  3.78 | ppl    43.80
## | epoch  69 |   800/ 1452 batches | lr 0.25000 | ms/batch 208.21726 | loss  3.74 | ppl    42.28
## | epoch  69 |   900/ 1452 batches | lr 0.25000 | ms/batch 216.32875 | loss  3.76 | ppl    42.89
## | epoch  69 |  1000/ 1452 batches | lr 0.25000 | ms/batch 212.29290 | loss  3.71 | ppl    40.95
## | epoch  69 |  1100/ 1452 batches | lr 0.25000 | ms/batch 206.13916 | loss  3.86 | ppl    47.26
## | epoch  69 |  1200/ 1452 batches | lr 0.25000 | ms/batch 206.21909 | loss  3.81 | ppl    44.94
## | epoch  69 |  1300/ 1452 batches | lr 0.25000 | ms/batch 207.50828 | loss  3.61 | ppl    36.80
## | epoch  69 |  1400/ 1452 batches | lr 0.25000 | ms/batch 214.22600 | loss  3.78 | ppl    43.77
## -----------------------------------------------------------------------------------------
## | end of epoch  69 | time: 322.89s | valid loss  4.54 | valid ppl    94.02
## | end of epoch  69 | time: 322.89s | test loss  4.48 | test ppl    88.39
## -----------------------------------------------------------------------------------------
## Save model!
```

```
##
## | epoch  70 |   100/ 1452 batches | lr 0.25000 | ms/batch 210.38015 | loss  3.90 | ppl    49.28
## | epoch  70 |   200/ 1452 batches | lr 0.25000 | ms/batch 209.78056 | loss  3.85 | ppl    46.86
## | epoch  70 |   300/ 1452 batches | lr 0.25000 | ms/batch 208.97655 | loss  3.83 | ppl    46.24
## | epoch  70 |   400/ 1452 batches | lr 0.25000 | ms/batch 208.56180 | loss  3.71 | ppl    40.79
## | epoch  70 |   500/ 1452 batches | lr 0.25000 | ms/batch 210.77977 | loss  3.73 | ppl    41.75
## | epoch  70 |   600/ 1452 batches | lr 0.25000 | ms/batch 208.92677 | loss  3.77 | ppl    43.31
## | epoch  70 |   700/ 1452 batches | lr 0.25000 | ms/batch 208.77183 | loss  3.77 | ppl    43.59
## | epoch  70 |   800/ 1452 batches | lr 0.25000 | ms/batch 209.59565 | loss  3.74 | ppl    42.04
## | epoch  70 |   900/ 1452 batches | lr 0.25000 | ms/batch 209.54102 | loss  3.75 | ppl    42.32
## | epoch  70 |  1000/ 1452 batches | lr 0.25000 | ms/batch 218.29725 | loss  3.71 | ppl    40.94
## | epoch  70 |  1100/ 1452 batches | lr 0.25000 | ms/batch 207.80284 | loss  3.85 | ppl    47.05
## | epoch  70 |  1200/ 1452 batches | lr 0.25000 | ms/batch 208.00263 | loss  3.80 | ppl    44.85
## | epoch  70 |  1300/ 1452 batches | lr 0.25000 | ms/batch 210.10031 | loss  3.61 | ppl    36.91
## | epoch  70 |  1400/ 1452 batches | lr 0.25000 | ms/batch 209.81545 | loss  3.77 | ppl    43.44
## -----------------------------------------------------------------------------------------
## | end of epoch  70 | time: 322.96s | valid loss  4.54 | valid ppl    94.13
## | end of epoch  70 | time: 322.96s | test loss  4.48 | test ppl    88.48
## -----------------------------------------------------------------------------------------
## | epoch  71 |   100/ 1452 batches | lr 0.25000 | ms/batch 210.23500 | loss  3.89 | ppl    48.80
## | epoch  71 |   200/ 1452 batches | lr 0.25000 | ms/batch 207.55774 | loss  3.85 | ppl    47.11
## | epoch  71 |   300/ 1452 batches | lr 0.25000 | ms/batch 214.95017 | loss  3.84 | ppl    46.74
## | epoch  71 |   400/ 1452 batches | lr 0.25000 | ms/batch 208.58690 | loss  3.70 | ppl    40.26
## | epoch  71 |   500/ 1452 batches | lr 0.25000 | ms/batch 207.95777 | loss  3.72 | ppl    41.43
## | epoch  71 |   600/ 1452 batches | lr 0.25000 | ms/batch 208.45220 | loss  3.77 | ppl    43.27
## | epoch  71 |   700/ 1452 batches | lr 0.25000 | ms/batch 208.38688 | loss  3.77 | ppl    43.51
## | epoch  71 |   800/ 1452 batches | lr 0.25000 | ms/batch 209.01623 | loss  3.74 | ppl    42.10
## | epoch  71 |   900/ 1452 batches | lr 0.25000 | ms/batch 207.65763 | loss  3.74 | ppl    42.19
## | epoch  71 |  1000/ 1452 batches | lr 0.25000 | ms/batch 207.89250 | loss  3.72 | ppl    41.06
## | epoch  71 |  1100/ 1452 batches | lr 0.25000 | ms/batch 207.55289 | loss  3.85 | ppl    47.18
## | epoch  71 |  1200/ 1452 batches | lr 0.25000 | ms/batch 212.00386 | loss  3.79 | ppl    44.43
## | epoch  71 |  1300/ 1452 batches | lr 0.25000 | ms/batch 208.61225 | loss  3.61 | ppl    36.80
## | epoch  71 |  1400/ 1452 batches | lr 0.25000 | ms/batch 209.59670 | loss  3.77 | ppl    43.32
## -----------------------------------------------------------------------------------------
## | end of epoch  71 | time: 322.06s | valid loss  4.55 | valid ppl    94.18
## | end of epoch  71 | time: 322.06s | test loss  4.48 | test ppl    88.49
## -----------------------------------------------------------------------------------------
## | epoch  72 |   100/ 1452 batches | lr 0.25000 | ms/batch 211.21450 | loss  3.88 | ppl    48.41
## | epoch  72 |   200/ 1452 batches | lr 0.25000 | ms/batch 209.17139 | loss  3.84 | ppl    46.68
## | epoch  72 |   300/ 1452 batches | lr 0.25000 | ms/batch 214.61823 | loss  3.83 | ppl    45.95
## | epoch  72 |   400/ 1452 batches | lr 0.25000 | ms/batch 207.56822 | loss  3.70 | ppl    40.30
## | epoch  72 |   500/ 1452 batches | lr 0.25000 | ms/batch 213.31724 | loss  3.73 | ppl    41.65
## | epoch  72 |   600/ 1452 batches | lr 0.25000 | ms/batch 206.99360 | loss  3.76 | ppl    42.97
## | epoch  72 |   700/ 1452 batches | lr 0.25000 | ms/batch 218.39706 | loss  3.78 | ppl    43.80
## | epoch  72 |   800/ 1452 batches | lr 0.25000 | ms/batch 206.50419 | loss  3.74 | ppl    42.20
## | epoch  72 |   900/ 1452 batches | lr 0.25000 | ms/batch 207.77750 | loss  3.74 | ppl    42.29
## | epoch  72 |  1000/ 1452 batches | lr 0.25000 | ms/batch 208.54203 | loss  3.71 | ppl    40.75
## | epoch  72 |  1100/ 1452 batches | lr 0.25000 | ms/batch 211.61387 | loss  3.84 | ppl    46.59
## | epoch  72 |  1200/ 1452 batches | lr 0.25000 | ms/batch 208.00265 | loss  3.80 | ppl    44.61
## | epoch  72 |  1300/ 1452 batches | lr 0.25000 | ms/batch 215.64472 | loss  3.61 | ppl    36.83
## | epoch  72 |  1400/ 1452 batches | lr 0.25000 | ms/batch 214.03646 | loss  3.76 | ppl    43.16
## -----------------------------------------------------------------------------------------
## | end of epoch  72 | time: 324.63s | valid loss  4.54 | valid ppl    94.15
## | end of epoch  72 | time: 324.63s | test loss  4.48 | test ppl    88.45
## -----------------------------------------------------------------------------------------
## | epoch  73 |   100/ 1452 batches | lr 0.25000 | ms/batch 211.26922 | loss  3.88 | ppl    48.20
## | epoch  73 |   200/ 1452 batches | lr 0.25000 | ms/batch 208.10790 | loss  3.84 | ppl    46.53
## | epoch  73 |   300/ 1452 batches | lr 0.25000 | ms/batch 209.27134 | loss  3.83 | ppl    46.13
## | epoch  73 |   400/ 1452 batches | lr 0.25000 | ms/batch 209.42112 | loss  3.70 | ppl    40.58
## | epoch  73 |   500/ 1452 batches | lr 0.25000 | ms/batch 209.20146 | loss  3.72 | ppl    41.21
## | epoch  73 |   600/ 1452 batches | lr 0.25000 | ms/batch 216.47907 | loss  3.76 | ppl    42.88
## | epoch  73 |   700/ 1452 batches | lr 0.25000 | ms/batch 208.20394 | loss  3.77 | ppl    43.50
## | epoch  73 |   800/ 1452 batches | lr 0.25000 | ms/batch 208.40240 | loss  3.74 | ppl    41.95
## | epoch  73 |   900/ 1452 batches | lr 0.25000 | ms/batch 208.86174 | loss  3.75 | ppl    42.53
## | epoch  73 |  1000/ 1452 batches | lr 0.25000 | ms/batch 208.95145 | loss  3.71 | ppl    41.01
## | epoch  73 |  1100/ 1452 batches | lr 0.25000 | ms/batch 212.25600 | loss  3.85 | ppl    47.22
## | epoch  73 |  1200/ 1452 batches | lr 0.25000 | ms/batch 208.01241 | loss  3.80 | ppl    44.75
## | epoch  73 |  1300/ 1452 batches | lr 0.25000 | ms/batch 209.06618 | loss  3.60 | ppl    36.70
## | epoch  73 |  1400/ 1452 batches | lr 0.25000 | ms/batch 206.04435 | loss  3.76 | ppl    42.96
## -----------------------------------------------------------------------------------------
## | end of epoch  73 | time: 322.67s | valid loss  4.54 | valid ppl    94.12
## | end of epoch  73 | time: 322.67s | test loss  4.48 | test ppl    88.45
```

```
## -----------------------------------------------------------------------------------------
## | epoch  74 |   100/ 1452 batches | lr 0.25000 | ms/batch 213.28187 | loss  3.89 | ppl    48.68
## | epoch  74 |   200/ 1452 batches | lr 0.25000 | ms/batch 206.41400 | loss  3.84 | ppl    46.61
## | epoch  74 |   300/ 1452 batches | lr 0.25000 | ms/batch 207.42785 | loss  3.82 | ppl    45.72
## | epoch  74 |   400/ 1452 batches | lr 0.25000 | ms/batch 207.68759 | loss  3.69 | ppl    40.14
## | epoch  74 |   500/ 1452 batches | lr 0.25000 | ms/batch 207.95732 | loss  3.73 | ppl    41.49
## | epoch  74 |   600/ 1452 batches | lr 0.25000 | ms/batch 210.22501 | loss  3.76 | ppl    43.10
## | epoch  74 |   700/ 1452 batches | lr 0.25000 | ms/batch 208.64661 | loss  3.78 | ppl    43.61
## | epoch  74 |   800/ 1452 batches | lr 0.25000 | ms/batch 210.97934 | loss  3.74 | ppl    42.05
## | epoch  74 |   900/ 1452 batches | lr 0.25000 | ms/batch 207.12816 | loss  3.74 | ppl    42.30
## | epoch  74 |  1000/ 1452 batches | lr 0.25000 | ms/batch 208.80173 | loss  3.71 | ppl    40.89
## | epoch  74 |  1100/ 1452 batches | lr 0.25000 | ms/batch 210.33489 | loss  3.84 | ppl    46.67
## | epoch  74 |  1200/ 1452 batches | lr 0.25000 | ms/batch 209.77545 | loss  3.79 | ppl    44.36
## | epoch  74 |  1300/ 1452 batches | lr 0.25000 | ms/batch 211.20401 | loss  3.60 | ppl    36.69
## | epoch  74 |  1400/ 1452 batches | lr 0.25000 | ms/batch 208.52690 | loss  3.76 | ppl    42.91
## -----------------------------------------------------------------------------------------
## | end of epoch  74 | time: 322.07s | valid loss  4.55 | valid ppl    94.19
## | end of epoch  74 | time: 322.07s | test loss  4.48 | test ppl    88.46
## -----------------------------------------------------------------------------------------
## | epoch  75 |   100/ 1452 batches | lr 0.12500 | ms/batch 220.89917 | loss  3.89 | ppl    49.07
## | epoch  75 |   200/ 1452 batches | lr 0.12500 | ms/batch 212.53773 | loss  3.85 | ppl    47.01
## | epoch  75 |   300/ 1452 batches | lr 0.12500 | ms/batch 210.31004 | loss  3.83 | ppl    46.22
## | epoch  75 |   400/ 1452 batches | lr 0.12500 | ms/batch 209.43582 | loss  3.70 | ppl    40.39
## | epoch  75 |   500/ 1452 batches | lr 0.12500 | ms/batch 208.78647 | loss  3.73 | ppl    41.58
## | epoch  75 |   600/ 1452 batches | lr 0.12500 | ms/batch 212.19315 | loss  3.76 | ppl    42.99
## | epoch  75 |   700/ 1452 batches | lr 0.12500 | ms/batch 210.03519 | loss  3.78 | ppl    43.82
## | epoch  75 |   800/ 1452 batches | lr 0.12500 | ms/batch 215.34488 | loss  3.74 | ppl    42.01
## | epoch  75 |   900/ 1452 batches | lr 0.12500 | ms/batch 217.88224 | loss  3.74 | ppl    42.27
## | epoch  75 |  1000/ 1452 batches | lr 0.12500 | ms/batch 209.20605 | loss  3.71 | ppl    40.74
## | epoch  75 |  1100/ 1452 batches | lr 0.12500 | ms/batch 209.88043 | loss  3.85 | ppl    46.94
## | epoch  75 |  1200/ 1452 batches | lr 0.12500 | ms/batch 209.45578 | loss  3.79 | ppl    44.44
## | epoch  75 |  1300/ 1452 batches | lr 0.12500 | ms/batch 209.49080 | loss  3.60 | ppl    36.68
## | epoch  75 |  1400/ 1452 batches | lr 0.12500 | ms/batch 210.67966 | loss  3.76 | ppl    42.81
## -----------------------------------------------------------------------------------------
## | end of epoch  75 | time: 325.85s | valid loss  4.54 | valid ppl    93.75
## | end of epoch  75 | time: 325.85s | test loss  4.48 | test ppl    88.10
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  76 |   100/ 1452 batches | lr 0.12500 | ms/batch 210.84943 | loss  3.88 | ppl    48.56
## | epoch  76 |   200/ 1452 batches | lr 0.12500 | ms/batch 216.41376 | loss  3.85 | ppl    46.79
## | epoch  76 |   300/ 1452 batches | lr 0.12500 | ms/batch 217.57251 | loss  3.83 | ppl    46.14
## | epoch  76 |   400/ 1452 batches | lr 0.12500 | ms/batch 208.55670 | loss  3.69 | ppl    40.06
## | epoch  76 |   500/ 1452 batches | lr 0.12500 | ms/batch 209.27619 | loss  3.73 | ppl    41.74
## | epoch  76 |   600/ 1452 batches | lr 0.12500 | ms/batch 210.08512 | loss  3.77 | ppl    43.21
## | epoch  76 |   700/ 1452 batches | lr 0.12500 | ms/batch 211.75344 | loss  3.77 | ppl    43.51
## | epoch  76 |   800/ 1452 batches | lr 0.12500 | ms/batch 217.01804 | loss  3.73 | ppl    41.89
## | epoch  76 |   900/ 1452 batches | lr 0.12500 | ms/batch 209.23108 | loss  3.75 | ppl    42.34
## | epoch  76 |  1000/ 1452 batches | lr 0.12500 | ms/batch 209.06617 | loss  3.71 | ppl    40.67
## | epoch  76 |  1100/ 1452 batches | lr 0.12500 | ms/batch 208.79647 | loss  3.84 | ppl    46.76
## | epoch  76 |  1200/ 1452 batches | lr 0.12500 | ms/batch 209.32591 | loss  3.79 | ppl    44.30
## | epoch  76 |  1300/ 1452 batches | lr 0.12500 | ms/batch 212.31302 | loss  3.60 | ppl    36.56
## | epoch  76 |  1400/ 1452 batches | lr 0.12500 | ms/batch 212.19804 | loss  3.75 | ppl    42.60
## -----------------------------------------------------------------------------------------
## | end of epoch  76 | time: 325.40s | valid loss  4.54 | valid ppl    93.74
## | end of epoch  76 | time: 325.40s | test loss  4.48 | test ppl    88.12
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  77 |   100/ 1452 batches | lr 0.12500 | ms/batch 210.84436 | loss  3.89 | ppl    48.78
## | epoch  77 |   200/ 1452 batches | lr 0.12500 | ms/batch 210.95426 | loss  3.84 | ppl    46.69
## | epoch  77 |   300/ 1452 batches | lr 0.12500 | ms/batch 210.23009 | loss  3.83 | ppl    46.12
## | epoch  77 |   400/ 1452 batches | lr 0.12500 | ms/batch 212.54778 | loss  3.69 | ppl    39.94
## | epoch  77 |   500/ 1452 batches | lr 0.12500 | ms/batch 215.11511 | loss  3.73 | ppl    41.55
## | epoch  77 |   600/ 1452 batches | lr 0.12500 | ms/batch 206.64396 | loss  3.76 | ppl    43.03
## | epoch  77 |   700/ 1452 batches | lr 0.12500 | ms/batch 217.15318 | loss  3.78 | ppl    43.63
## | epoch  77 |   800/ 1452 batches | lr 0.12500 | ms/batch 206.85863 | loss  3.73 | ppl    41.54
## | epoch  77 |   900/ 1452 batches | lr 0.12500 | ms/batch 208.64679 | loss  3.75 | ppl    42.35
## | epoch  77 |  1000/ 1452 batches | lr 0.12500 | ms/batch 209.33098 | loss  3.71 | ppl    40.71
## | epoch  77 |  1100/ 1452 batches | lr 0.12500 | ms/batch 210.84435 | loss  3.85 | ppl    46.84
## | epoch  77 |  1200/ 1452 batches | lr 0.12500 | ms/batch 210.75446 | loss  3.79 | ppl    44.28
## | epoch  77 |  1300/ 1452 batches | lr 0.12500 | ms/batch 211.18402 | loss  3.60 | ppl    36.62
```

```
## | epoch  77 |  1400/ 1452 batches | lr 0.12500 | ms/batch 209.78054 | loss  3.75 | ppl    42.60
## -----------------------------------------------------------------------------------------
## | end of epoch  77 | time: 325.01s | valid loss  4.54 | valid ppl    93.73
## | end of epoch  77 | time: 325.01s | test loss  4.48 | test ppl    88.08
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  78 |   100/ 1452 batches | lr 0.12500 | ms/batch 210.02029 | loss  3.87 | ppl    48.08
## | epoch  78 |   200/ 1452 batches | lr 0.12500 | ms/batch 216.64844 | loss  3.84 | ppl    46.48
## | epoch  78 |   300/ 1452 batches | lr 0.12500 | ms/batch 208.10715 | loss  3.82 | ppl    45.56
## | epoch  78 |   400/ 1452 batches | lr 0.12500 | ms/batch 224.30570 | loss  3.69 | ppl    40.10
## | epoch  78 |   500/ 1452 batches | lr 0.12500 | ms/batch 208.61170 | loss  3.72 | ppl    41.23
## | epoch  78 |   600/ 1452 batches | lr 0.12500 | ms/batch 215.60747 | loss  3.76 | ppl    43.12
## | epoch  78 |   700/ 1452 batches | lr 0.12500 | ms/batch 209.89534 | loss  3.77 | ppl    43.40
## | epoch  78 |   800/ 1452 batches | lr 0.12500 | ms/batch 210.89431 | loss  3.74 | ppl    41.99
## | epoch  78 |   900/ 1452 batches | lr 0.12500 | ms/batch 209.56567 | loss  3.74 | ppl    42.16
## | epoch  78 |  1000/ 1452 batches | lr 0.12500 | ms/batch 222.42253 | loss  3.71 | ppl    40.83
## | epoch  78 |  1100/ 1452 batches | lr 0.12500 | ms/batch 207.55772 | loss  3.85 | ppl    46.83
## | epoch  78 |  1200/ 1452 batches | lr 0.12500 | ms/batch 210.10512 | loss  3.79 | ppl    44.23
## | epoch  78 |  1300/ 1452 batches | lr 0.12500 | ms/batch 211.73880 | loss  3.60 | ppl    36.64
## | epoch  78 |  1400/ 1452 batches | lr 0.12500 | ms/batch 211.04415 | loss  3.76 | ppl    42.81
## -----------------------------------------------------------------------------------------
## | end of epoch  78 | time: 327.20s | valid loss  4.54 | valid ppl    93.71
## | end of epoch  78 | time: 327.20s | test loss  4.48 | test ppl    88.07
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  79 |   100/ 1452 batches | lr 0.12500 | ms/batch 211.20907 | loss  3.88 | ppl    48.24
## | epoch  79 |   200/ 1452 batches | lr 0.12500 | ms/batch 214.70548 | loss  3.84 | ppl    46.40
## | epoch  79 |   300/ 1452 batches | lr 0.12500 | ms/batch 209.47090 | loss  3.83 | ppl    45.91
## | epoch  79 |   400/ 1452 batches | lr 0.12500 | ms/batch 212.44781 | loss  3.70 | ppl    40.25
## | epoch  79 |   500/ 1452 batches | lr 0.12500 | ms/batch 211.25397 | loss  3.73 | ppl    41.50
## | epoch  79 |   600/ 1452 batches | lr 0.12500 | ms/batch 212.32796 | loss  3.76 | ppl    42.87
## | epoch  79 |   700/ 1452 batches | lr 0.12500 | ms/batch 209.89043 | loss  3.77 | ppl    43.48
## | epoch  79 |   800/ 1452 batches | lr 0.12500 | ms/batch 213.92121 | loss  3.73 | ppl    41.88
## | epoch  79 |   900/ 1452 batches | lr 0.12500 | ms/batch 207.61783 | loss  3.74 | ppl    42.06
## | epoch  79 |  1000/ 1452 batches | lr 0.12500 | ms/batch 215.13997 | loss  3.71 | ppl    40.67
## | epoch  79 |  1100/ 1452 batches | lr 0.12500 | ms/batch 210.02534 | loss  3.85 | ppl    46.80
## | epoch  79 |  1200/ 1452 batches | lr 0.12500 | ms/batch 217.61745 | loss  3.79 | ppl    44.16
## | epoch  79 |  1300/ 1452 batches | lr 0.12500 | ms/batch 207.07330 | loss  3.60 | ppl    36.51
## | epoch  79 |  1400/ 1452 batches | lr 0.12500 | ms/batch 209.47575 | loss  3.76 | ppl    43.07
## -----------------------------------------------------------------------------------------
## | end of epoch  79 | time: 325.67s | valid loss  4.54 | valid ppl    93.76
## | end of epoch  79 | time: 325.67s | test loss  4.48 | test ppl    88.11
## -----------------------------------------------------------------------------------------
## | epoch  80 |   100/ 1452 batches | lr 0.12500 | ms/batch 212.73277 | loss  3.88 | ppl    48.66
## | epoch  80 |   200/ 1452 batches | lr 0.12500 | ms/batch 209.10615 | loss  3.84 | ppl    46.56
## | epoch  80 |   300/ 1452 batches | lr 0.12500 | ms/batch 209.37112 | loss  3.83 | ppl    45.94
## | epoch  80 |   400/ 1452 batches | lr 0.12500 | ms/batch 209.01623 | loss  3.69 | ppl    39.97
## | epoch  80 |   500/ 1452 batches | lr 0.12500 | ms/batch 211.08412 | loss  3.72 | ppl    41.19
## | epoch  80 |   600/ 1452 batches | lr 0.12500 | ms/batch 209.86047 | loss  3.75 | ppl    42.63
## | epoch  80 |   700/ 1452 batches | lr 0.12500 | ms/batch 210.09022 | loss  3.77 | ppl    43.20
## | epoch  80 |   800/ 1452 batches | lr 0.12500 | ms/batch 211.24905 | loss  3.73 | ppl    41.85
## | epoch  80 |   900/ 1452 batches | lr 0.12500 | ms/batch 209.26598 | loss  3.74 | ppl    42.18
## | epoch  80 |  1000/ 1452 batches | lr 0.12500 | ms/batch 221.12893 | loss  3.70 | ppl    40.62
## | epoch  80 |  1100/ 1452 batches | lr 0.12500 | ms/batch 209.45089 | loss  3.85 | ppl    46.90
## | epoch  80 |  1200/ 1452 batches | lr 0.12500 | ms/batch 208.15711 | loss  3.79 | ppl    44.05
## | epoch  80 |  1300/ 1452 batches | lr 0.12500 | ms/batch 207.93735 | loss  3.60 | ppl    36.63
## | epoch  80 |  1400/ 1452 batches | lr 0.12500 | ms/batch 208.47678 | loss  3.76 | ppl    42.77
## -----------------------------------------------------------------------------------------
## | end of epoch  80 | time: 324.14s | valid loss  4.54 | valid ppl    93.79
## | end of epoch  80 | time: 324.14s | test loss  4.48 | test ppl    88.10
## -----------------------------------------------------------------------------------------
## | epoch  81 |   100/ 1452 batches | lr 0.06250 | ms/batch 211.29390 | loss  3.88 | ppl    48.46
## | epoch  81 |   200/ 1452 batches | lr 0.06250 | ms/batch 210.92938 | loss  3.84 | ppl    46.63
## | epoch  81 |   300/ 1452 batches | lr 0.06250 | ms/batch 207.66268 | loss  3.83 | ppl    46.05
## | epoch  81 |   400/ 1452 batches | lr 0.06250 | ms/batch 216.83825 | loss  3.69 | ppl    40.16
## | epoch  81 |   500/ 1452 batches | lr 0.06250 | ms/batch 209.68554 | loss  3.72 | ppl    41.36
## | epoch  81 |   600/ 1452 batches | lr 0.06250 | ms/batch 208.88636 | loss  3.76 | ppl    43.06
## | epoch  81 |   700/ 1452 batches | lr 0.06250 | ms/batch 209.69554 | loss  3.77 | ppl    43.42
## | epoch  81 |   800/ 1452 batches | lr 0.06250 | ms/batch 216.07919 | loss  3.74 | ppl    42.28
## | epoch  81 |   900/ 1452 batches | lr 0.06250 | ms/batch 207.95239 | loss  3.74 | ppl    42.21
```

```
## | epoch  81 |  1000/ 1452 batches | lr 0.06250 | ms/batch 209.54569 | loss  3.70 | ppl    40.49
## | epoch  81 |  1100/ 1452 batches | lr 0.06250 | ms/batch 215.90919 | loss  3.84 | ppl    46.42
## | epoch  81 |  1200/ 1452 batches | lr 0.06250 | ms/batch 209.10615 | loss  3.79 | ppl    44.19
## | epoch  81 |  1300/ 1452 batches | lr 0.06250 | ms/batch 211.64354 | loss  3.59 | ppl    36.18
## | epoch  81 |  1400/ 1452 batches | lr 0.06250 | ms/batch 208.65172 | loss  3.75 | ppl    42.52
## -----------------------------------------------------------------------------------------
## | end of epoch  81 | time: 324.77s | valid loss  4.54 | valid ppl    93.42
## | end of epoch  81 | time: 324.77s | test loss  4.48 | test ppl    87.81
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  82 |   100/ 1452 batches | lr 0.06250 | ms/batch 211.35385 | loss  3.87 | ppl    47.89
## | epoch  82 |   200/ 1452 batches | lr 0.06250 | ms/batch 209.68555 | loss  3.85 | ppl    46.84
## | epoch  82 |   300/ 1452 batches | lr 0.06250 | ms/batch 209.58075 | loss  3.82 | ppl    45.73
## | epoch  82 |   400/ 1452 batches | lr 0.06250 | ms/batch 216.61868 | loss  3.69 | ppl    39.97
## | epoch  82 |   500/ 1452 batches | lr 0.06250 | ms/batch 209.99028 | loss  3.72 | ppl    41.15
## | epoch  82 |   600/ 1452 batches | lr 0.06250 | ms/batch 209.42583 | loss  3.76 | ppl    43.03
## | epoch  82 |   700/ 1452 batches | lr 0.06250 | ms/batch 210.45476 | loss  3.77 | ppl    43.56
## | epoch  82 |   800/ 1452 batches | lr 0.06250 | ms/batch 208.36200 | loss  3.73 | ppl    41.69
## | epoch  82 |   900/ 1452 batches | lr 0.06250 | ms/batch 210.94937 | loss  3.74 | ppl    42.20
## | epoch  82 |  1000/ 1452 batches | lr 0.06250 | ms/batch 211.74852 | loss  3.70 | ppl    40.38
## | epoch  82 |  1100/ 1452 batches | lr 0.06250 | ms/batch 209.72549 | loss  3.85 | ppl    46.80
## | epoch  82 |  1200/ 1452 batches | lr 0.06250 | ms/batch 216.60871 | loss  3.79 | ppl    44.26
## | epoch  82 |  1300/ 1452 batches | lr 0.06250 | ms/batch 204.65068 | loss  3.60 | ppl    36.67
## | epoch  82 |  1400/ 1452 batches | lr 0.06250 | ms/batch 204.71065 | loss  3.76 | ppl    42.79
## -----------------------------------------------------------------------------------------
## | end of epoch  82 | time: 323.59s | valid loss  4.54 | valid ppl    93.40
## | end of epoch  82 | time: 323.59s | test loss  4.48 | test ppl    87.83
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  83 |   100/ 1452 batches | lr 0.06250 | ms/batch 210.74964 | loss  3.88 | ppl    48.30
## | epoch  83 |   200/ 1452 batches | lr 0.06250 | ms/batch 212.10305 | loss  3.83 | ppl    46.21
## | epoch  83 |   300/ 1452 batches | lr 0.06250 | ms/batch 217.14792 | loss  3.82 | ppl    45.73
## | epoch  83 |   400/ 1452 batches | lr 0.06250 | ms/batch 207.83746 | loss  3.69 | ppl    40.23
## | epoch  83 |   500/ 1452 batches | lr 0.06250 | ms/batch 209.49086 | loss  3.72 | ppl    41.39
## | epoch  83 |   600/ 1452 batches | lr 0.06250 | ms/batch 214.65066 | loss  3.76 | ppl    42.95
## | epoch  83 |   700/ 1452 batches | lr 0.06250 | ms/batch 209.48089 | loss  3.77 | ppl    43.35
## | epoch  83 |   800/ 1452 batches | lr 0.06250 | ms/batch 212.77744 | loss  3.74 | ppl    41.94
## | epoch  83 |   900/ 1452 batches | lr 0.06250 | ms/batch 208.96158 | loss  3.73 | ppl    41.78
## | epoch  83 |  1000/ 1452 batches | lr 0.06250 | ms/batch 209.89553 | loss  3.70 | ppl    40.46
## | epoch  83 |  1100/ 1452 batches | lr 0.06250 | ms/batch 210.36998 | loss  3.84 | ppl    46.40
## | epoch  83 |  1200/ 1452 batches | lr 0.06250 | ms/batch 209.05621 | loss  3.79 | ppl    44.13
## | epoch  83 |  1300/ 1452 batches | lr 0.06250 | ms/batch 217.15300 | loss  3.60 | ppl    36.49
## | epoch  83 |  1400/ 1452 batches | lr 0.06250 | ms/batch 208.52674 | loss  3.75 | ppl    42.72
## -----------------------------------------------------------------------------------------
## | end of epoch  83 | time: 325.14s | valid loss  4.54 | valid ppl    93.37
## | end of epoch  83 | time: 325.14s | test loss  4.47 | test ppl    87.79
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  84 |   100/ 1452 batches | lr 0.06250 | ms/batch 211.48879 | loss  3.87 | ppl    47.97
## | epoch  84 |   200/ 1452 batches | lr 0.06250 | ms/batch 208.74667 | loss  3.83 | ppl    46.15
## | epoch  84 |   300/ 1452 batches | lr 0.06250 | ms/batch 212.16807 | loss  3.82 | ppl    45.79
## | epoch  84 |   400/ 1452 batches | lr 0.06250 | ms/batch 208.99134 | loss  3.68 | ppl    39.81
## | epoch  84 |   500/ 1452 batches | lr 0.06250 | ms/batch 208.74164 | loss  3.72 | ppl    41.38
## | epoch  84 |   600/ 1452 batches | lr 0.06250 | ms/batch 209.87548 | loss  3.75 | ppl    42.73
## | epoch  84 |   700/ 1452 batches | lr 0.06250 | ms/batch 213.10206 | loss  3.77 | ppl    43.50
## | epoch  84 |   800/ 1452 batches | lr 0.06250 | ms/batch 209.68555 | loss  3.73 | ppl    41.88
## | epoch  84 |   900/ 1452 batches | lr 0.06250 | ms/batch 208.06720 | loss  3.74 | ppl    41.95
## | epoch  84 |  1000/ 1452 batches | lr 0.06250 | ms/batch 209.96525 | loss  3.71 | ppl    40.80
## | epoch  84 |  1100/ 1452 batches | lr 0.06250 | ms/batch 209.45597 | loss  3.84 | ppl    46.75
## | epoch  84 |  1200/ 1452 batches | lr 0.06250 | ms/batch 206.52878 | loss  3.79 | ppl    44.25
## | epoch  84 |  1300/ 1452 batches | lr 0.06250 | ms/batch 207.32796 | loss  3.60 | ppl    36.61
## | epoch  84 |  1400/ 1452 batches | lr 0.06250 | ms/batch 211.35383 | loss  3.76 | ppl    42.86
## -----------------------------------------------------------------------------------------
## | end of epoch  84 | time: 322.64s | valid loss  4.54 | valid ppl    93.44
## | end of epoch  84 | time: 322.64s | test loss  4.48 | test ppl    87.83
## -----------------------------------------------------------------------------------------
## | epoch  85 |   100/ 1452 batches | lr 0.06250 | ms/batch 211.41883 | loss  3.87 | ppl    48.06
## | epoch  85 |   200/ 1452 batches | lr 0.06250 | ms/batch 209.00135 | loss  3.84 | ppl    46.62
## | epoch  85 |   300/ 1452 batches | lr 0.06250 | ms/batch 208.40687 | loss  3.83 | ppl    46.01
```

```
## | epoch  85 |   400/ 1452 batches | lr 0.06250 | ms/batch 216.78336 | loss  3.68 | ppl    39.73
## | epoch  85 |   500/ 1452 batches | lr 0.06250 | ms/batch 210.53975 | loss  3.72 | ppl    41.14
## | epoch  85 |   600/ 1452 batches | lr 0.06250 | ms/batch 209.33603 | loss  3.76 | ppl    42.87
## | epoch  85 |   700/ 1452 batches | lr 0.06250 | ms/batch 208.44700 | loss  3.77 | ppl    43.27
## | epoch  85 |   800/ 1452 batches | lr 0.06250 | ms/batch 214.94018 | loss  3.73 | ppl    41.88
## | epoch  85 |   900/ 1452 batches | lr 0.06250 | ms/batch 210.65962 | loss  3.73 | ppl    41.49
## | epoch  85 |  1000/ 1452 batches | lr 0.06250 | ms/batch 218.25186 | loss  3.70 | ppl    40.55
## | epoch  85 |  1100/ 1452 batches | lr 0.06250 | ms/batch 209.82540 | loss  3.84 | ppl    46.63
## | epoch  85 |  1200/ 1452 batches | lr 0.06250 | ms/batch 207.87742 | loss  3.78 | ppl    44.00
## | epoch  85 |  1300/ 1452 batches | lr 0.06250 | ms/batch 207.74753 | loss  3.60 | ppl    36.57
## | epoch  85 |  1400/ 1452 batches | lr 0.06250 | ms/batch 209.72062 | loss  3.74 | ppl    42.24
## -----------------------------------------------------------------------------------------
## | end of epoch  85 | time: 324.63s | valid loss  4.54 | valid ppl    93.42
## | end of epoch  85 | time: 324.63s | test loss  4.48 | test ppl    87.81
## -----------------------------------------------------------------------------------------
## | epoch  86 |   100/ 1452 batches | lr 0.06250 | ms/batch 210.93959 | loss  3.87 | ppl    48.15
## | epoch  86 |   200/ 1452 batches | lr 0.06250 | ms/batch 210.13019 | loss  3.84 | ppl    46.38
## | epoch  86 |   300/ 1452 batches | lr 0.06250 | ms/batch 208.56670 | loss  3.82 | ppl    45.70
## | epoch  86 |   400/ 1452 batches | lr 0.06250 | ms/batch 207.41307 | loss  3.68 | ppl    39.77
## | epoch  86 |   500/ 1452 batches | lr 0.06250 | ms/batch 209.66558 | loss  3.72 | ppl    41.14
## | epoch  86 |   600/ 1452 batches | lr 0.06250 | ms/batch 219.67041 | loss  3.75 | ppl    42.50
## | epoch  86 |   700/ 1452 batches | lr 0.06250 | ms/batch 208.48699 | loss  3.77 | ppl    43.24
## | epoch  86 |   800/ 1452 batches | lr 0.06250 | ms/batch 208.75649 | loss  3.73 | ppl    41.71
## | epoch  86 |   900/ 1452 batches | lr 0.06250 | ms/batch 208.01746 | loss  3.74 | ppl    41.98
## | epoch  86 |  1000/ 1452 batches | lr 0.06250 | ms/batch 209.39584 | loss  3.71 | ppl    40.82
## | epoch  86 |  1100/ 1452 batches | lr 0.06250 | ms/batch 210.19012 | loss  3.84 | ppl    46.64
## | epoch  86 |  1200/ 1452 batches | lr 0.06250 | ms/batch 209.14119 | loss  3.78 | ppl    43.79
## | epoch  86 |  1300/ 1452 batches | lr 0.06250 | ms/batch 208.74669 | loss  3.60 | ppl    36.49
## | epoch  86 |  1400/ 1452 batches | lr 0.06250 | ms/batch 208.98626 | loss  3.75 | ppl    42.45
## -----------------------------------------------------------------------------------------
## | end of epoch  86 | time: 323.16s | valid loss  4.54 | valid ppl    93.40
## | end of epoch  86 | time: 323.16s | test loss  4.47 | test ppl    87.79
## -----------------------------------------------------------------------------------------
## | epoch  87 |   100/ 1452 batches | lr 0.06250 | ms/batch 212.25292 | loss  3.88 | ppl    48.46
## | epoch  87 |   200/ 1452 batches | lr 0.06250 | ms/batch 209.29595 | loss  3.83 | ppl    46.22
## | epoch  87 |   300/ 1452 batches | lr 0.06250 | ms/batch 212.14304 | loss  3.82 | ppl    45.46
## | epoch  87 |   400/ 1452 batches | lr 0.06250 | ms/batch 208.56670 | loss  3.68 | ppl    39.59
## | epoch  87 |   500/ 1452 batches | lr 0.06250 | ms/batch 208.73161 | loss  3.71 | ppl    41.04
## | epoch  87 |   600/ 1452 batches | lr 0.06250 | ms/batch 212.06312 | loss  3.75 | ppl    42.73
## | epoch  87 |   700/ 1452 batches | lr 0.06250 | ms/batch 215.49477 | loss  3.77 | ppl    43.33
## | epoch  87 |   800/ 1452 batches | lr 0.06250 | ms/batch 215.73443 | loss  3.73 | ppl    41.80
## | epoch  87 |   900/ 1452 batches | lr 0.06250 | ms/batch 208.08231 | loss  3.73 | ppl    41.79
## | epoch  87 |  1000/ 1452 batches | lr 0.06250 | ms/batch 215.89433 | loss  3.70 | ppl    40.57
## | epoch  87 |  1100/ 1452 batches | lr 0.06250 | ms/batch 206.88352 | loss  3.84 | ppl    46.58
## | epoch  87 |  1200/ 1452 batches | lr 0.06250 | ms/batch 211.48383 | loss  3.79 | ppl    44.05
## | epoch  87 |  1300/ 1452 batches | lr 0.06250 | ms/batch 207.26804 | loss  3.60 | ppl    36.54
## | epoch  87 |  1400/ 1452 batches | lr 0.06250 | ms/batch 214.59053 | loss  3.76 | ppl    42.95
## -----------------------------------------------------------------------------------------
## | end of epoch  87 | time: 325.16s | valid loss  4.54 | valid ppl    93.37
## | end of epoch  87 | time: 325.16s | test loss  4.47 | test ppl    87.76
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  88 |   100/ 1452 batches | lr 0.06250 | ms/batch 209.39586 | loss  3.88 | ppl    48.27
## | epoch  88 |   200/ 1452 batches | lr 0.06250 | ms/batch 208.81643 | loss  3.83 | ppl    46.04
## | epoch  88 |   300/ 1452 batches | lr 0.06250 | ms/batch 208.86151 | loss  3.82 | ppl    45.45
## | epoch  88 |   400/ 1452 batches | lr 0.06250 | ms/batch 210.61692 | loss  3.68 | ppl    39.73
## | epoch  88 |   500/ 1452 batches | lr 0.06250 | ms/batch 214.28492 | loss  3.71 | ppl    40.98
## | epoch  88 |   600/ 1452 batches | lr 0.06250 | ms/batch 213.96782 | loss  3.75 | ppl    42.63
## | epoch  88 |   700/ 1452 batches | lr 0.06250 | ms/batch 212.37468 | loss  3.76 | ppl    43.05
## | epoch  88 |   800/ 1452 batches | lr 0.06250 | ms/batch 212.71940 | loss  3.73 | ppl    41.62
## | epoch  88 |   900/ 1452 batches | lr 0.06250 | ms/batch 210.99473 | loss  3.74 | ppl    42.02
## | epoch  88 |  1000/ 1452 batches | lr 0.06250 | ms/batch 213.94924 | loss  3.70 | ppl    40.38
## | epoch  88 |  1100/ 1452 batches | lr 0.06250 | ms/batch 218.77737 | loss  3.84 | ppl    46.50
## | epoch  88 |  1200/ 1452 batches | lr 0.06250 | ms/batch 212.80096 | loss  3.78 | ppl    44.03
## | epoch  88 |  1300/ 1452 batches | lr 0.06250 | ms/batch 212.09353 | loss  3.60 | ppl    36.60
## | epoch  88 |  1400/ 1452 batches | lr 0.06250 | ms/batch 210.43155 | loss  3.75 | ppl    42.68
## -----------------------------------------------------------------------------------------
## | end of epoch  88 | time: 327.17s | valid loss  4.54 | valid ppl    93.38
## | end of epoch  88 | time: 327.17s | test loss  4.48 | test ppl    87.80
## -----------------------------------------------------------------------------------------
## | epoch  89 |   100/ 1452 batches | lr 0.06250 | ms/batch 217.04732 | loss  3.87 | ppl    48.16
```

```
## | epoch  89 |   200/ 1452 batches | lr 0.06250 | ms/batch 217.80686 | loss  3.83 | ppl    46.24
## | epoch  89 |   300/ 1452 batches | lr 0.06250 | ms/batch 210.85811 | loss  3.82 | ppl    45.50
## | epoch  89 |   400/ 1452 batches | lr 0.06250 | ms/batch 211.60734 | loss  3.69 | ppl    39.85
## | epoch  89 |   500/ 1452 batches | lr 0.06250 | ms/batch 215.40104 | loss  3.72 | ppl    41.31
## | epoch  89 |   600/ 1452 batches | lr 0.06250 | ms/batch 214.08292 | loss  3.75 | ppl    42.61
## | epoch  89 |   700/ 1452 batches | lr 0.06250 | ms/batch 211.54234 | loss  3.76 | ppl    43.16
## | epoch  89 |   800/ 1452 batches | lr 0.06250 | ms/batch 216.60958 | loss  3.73 | ppl    41.75
## | epoch  89 |   900/ 1452 batches | lr 0.06250 | ms/batch 211.52163 | loss  3.73 | ppl    41.82
## | epoch  89 |  1000/ 1452 batches | lr 0.06250 | ms/batch 211.30060 | loss  3.71 | ppl    40.66
## | epoch  89 |  1100/ 1452 batches | lr 0.06250 | ms/batch 212.88247 | loss  3.84 | ppl    46.40
## | epoch  89 |  1200/ 1452 batches | lr 0.06250 | ms/batch 211.32766 | loss  3.79 | ppl    44.14
## | epoch  89 |  1300/ 1452 batches | lr 0.06250 | ms/batch 215.86939 | loss  3.60 | ppl    36.42
## | epoch  89 |  1400/ 1452 batches | lr 0.06250 | ms/batch 210.56768 | loss  3.75 | ppl    42.53
## -----------------------------------------------------------------------------------------
## | end of epoch  89 | time: 328.75s | valid loss  4.54 | valid ppl    93.38
## | end of epoch  89 | time: 328.75s | test loss  4.48 | test ppl    87.79
## -----------------------------------------------------------------------------------------
## | epoch  90 |   100/ 1452 batches | lr 0.06250 | ms/batch 213.62976 | loss  3.87 | ppl    47.94
## | epoch  90 |   200/ 1452 batches | lr 0.06250 | ms/batch 212.41892 | loss  3.83 | ppl    46.11
## | epoch  90 |   300/ 1452 batches | lr 0.06250 | ms/batch 210.96205 | loss  3.82 | ppl    45.51
## | epoch  90 |   400/ 1452 batches | lr 0.06250 | ms/batch 207.93735 | loss  3.69 | ppl    39.94
## | epoch  90 |   500/ 1452 batches | lr 0.06250 | ms/batch 208.04722 | loss  3.71 | ppl    40.71
## | epoch  90 |   600/ 1452 batches | lr 0.06250 | ms/batch 207.53775 | loss  3.76 | ppl    42.77
## | epoch  90 |   700/ 1452 batches | lr 0.06250 | ms/batch 207.88884 | loss  3.76 | ppl    43.12
## | epoch  90 |   800/ 1452 batches | lr 0.06250 | ms/batch 209.35588 | loss  3.72 | ppl    41.45
## | epoch  90 |   900/ 1452 batches | lr 0.06250 | ms/batch 207.61765 | loss  3.73 | ppl    41.84
## | epoch  90 |  1000/ 1452 batches | lr 0.06250 | ms/batch 208.91634 | loss  3.70 | ppl    40.31
## | epoch  90 |  1100/ 1452 batches | lr 0.06250 | ms/batch 207.74754 | loss  3.84 | ppl    46.38
## | epoch  90 |  1200/ 1452 batches | lr 0.06250 | ms/batch 207.66762 | loss  3.78 | ppl    43.77
## | epoch  90 |  1300/ 1452 batches | lr 0.06250 | ms/batch 208.25716 | loss  3.60 | ppl    36.63
## | epoch  90 |  1400/ 1452 batches | lr 0.06250 | ms/batch 207.89253 | loss  3.75 | ppl    42.55
## -----------------------------------------------------------------------------------------
## | end of epoch  90 | time: 321.69s | valid loss  4.54 | valid ppl    93.41
## | end of epoch  90 | time: 321.69s | test loss  4.48 | test ppl    87.81
## -----------------------------------------------------------------------------------------
## | epoch  91 |   100/ 1452 batches | lr 0.06250 | ms/batch 212.64251 | loss  3.87 | ppl    48.00
## | epoch  91 |   200/ 1452 batches | lr 0.06250 | ms/batch 210.19506 | loss  3.83 | ppl    46.00
## | epoch  91 |   300/ 1452 batches | lr 0.06250 | ms/batch 208.96136 | loss  3.82 | ppl    45.78
## | epoch  91 |   400/ 1452 batches | lr 0.06250 | ms/batch 209.11122 | loss  3.68 | ppl    39.82
## | epoch  91 |   500/ 1452 batches | lr 0.06250 | ms/batch 209.25600 | loss  3.71 | ppl    40.81
## | epoch  91 |   600/ 1452 batches | lr 0.06250 | ms/batch 220.77420 | loss  3.75 | ppl    42.73
## | epoch  91 |   700/ 1452 batches | lr 0.06250 | ms/batch 206.74855 | loss  3.76 | ppl    43.08
## | epoch  91 |   800/ 1452 batches | lr 0.06250 | ms/batch 208.88755 | loss  3.73 | ppl    41.72
## | epoch  91 |   900/ 1452 batches | lr 0.06250 | ms/batch 209.66556 | loss  3.73 | ppl    41.88
## | epoch  91 |  1000/ 1452 batches | lr 0.06250 | ms/batch 218.37667 | loss  3.70 | ppl    40.30
## | epoch  91 |  1100/ 1452 batches | lr 0.06250 | ms/batch 209.26597 | loss  3.84 | ppl    46.35
## | epoch  91 |  1200/ 1452 batches | lr 0.06250 | ms/batch 208.46679 | loss  3.79 | ppl    44.04
## | epoch  91 |  1300/ 1452 batches | lr 0.06250 | ms/batch 209.70072 | loss  3.60 | ppl    36.51
## | epoch  91 |  1400/ 1452 batches | lr 0.06250 | ms/batch 208.74157 | loss  3.75 | ppl    42.60
## -----------------------------------------------------------------------------------------
## | end of epoch  91 | time: 324.24s | valid loss  4.54 | valid ppl    93.41
## | end of epoch  91 | time: 324.24s | test loss  4.48 | test ppl    87.80
## -----------------------------------------------------------------------------------------
## | epoch  92 |   100/ 1452 batches | lr 0.03125 | ms/batch 211.95832 | loss  3.87 | ppl    47.79
## | epoch  92 |   200/ 1452 batches | lr 0.03125 | ms/batch 209.15124 | loss  3.84 | ppl    46.39
## | epoch  92 |   300/ 1452 batches | lr 0.03125 | ms/batch 216.11895 | loss  3.82 | ppl    45.76
## | epoch  92 |   400/ 1452 batches | lr 0.03125 | ms/batch 208.34710 | loss  3.69 | ppl    39.86
## | epoch  92 |   500/ 1452 batches | lr 0.03125 | ms/batch 210.67476 | loss  3.72 | ppl    41.10
## | epoch  92 |   600/ 1452 batches | lr 0.03125 | ms/batch 209.20603 | loss  3.75 | ppl    42.62
## | epoch  92 |   700/ 1452 batches | lr 0.03125 | ms/batch 209.92533 | loss  3.76 | ppl    43.14
## | epoch  92 |   800/ 1452 batches | lr 0.03125 | ms/batch 209.68554 | loss  3.72 | ppl    41.41
## | epoch  92 |   900/ 1452 batches | lr 0.03125 | ms/batch 211.98320 | loss  3.73 | ppl    41.79
## | epoch  92 |  1000/ 1452 batches | lr 0.03125 | ms/batch 207.58769 | loss  3.70 | ppl    40.28
## | epoch  92 |  1100/ 1452 batches | lr 0.03125 | ms/batch 209.48576 | loss  3.84 | ppl    46.50
## | epoch  92 |  1200/ 1452 batches | lr 0.03125 | ms/batch 209.42582 | loss  3.78 | ppl    43.81
## | epoch  92 |  1300/ 1452 batches | lr 0.03125 | ms/batch 208.93631 | loss  3.59 | ppl    36.22
## | epoch  92 |  1400/ 1452 batches | lr 0.03125 | ms/batch 212.40278 | loss  3.75 | ppl    42.57
## -----------------------------------------------------------------------------------------
## | end of epoch  92 | time: 323.67s | valid loss  4.54 | valid ppl    93.31
## | end of epoch  92 | time: 323.67s | test loss  4.47 | test ppl    87.72
## -----------------------------------------------------------------------------------------
## Save model!
```

```
##
## | epoch  93 |   100/ 1452 batches | lr 0.03125 | ms/batch 209.20604 | loss  3.87 | ppl    47.94
## | epoch  93 |   200/ 1452 batches | lr 0.03125 | ms/batch 210.59479 | loss  3.84 | ppl    46.52
## | epoch  93 |   300/ 1452 batches | lr 0.03125 | ms/batch 209.47576 | loss  3.82 | ppl    45.74
## | epoch  93 |   400/ 1452 batches | lr 0.03125 | ms/batch 210.67454 | loss  3.68 | ppl    39.62
## | epoch  93 |   500/ 1452 batches | lr 0.03125 | ms/batch 207.97729 | loss  3.72 | ppl    41.10
## | epoch  93 |   600/ 1452 batches | lr 0.03125 | ms/batch 209.06619 | loss  3.75 | ppl    42.47
## | epoch  93 |   700/ 1452 batches | lr 0.03125 | ms/batch 210.83946 | loss  3.76 | ppl    43.03
## | epoch  93 |   800/ 1452 batches | lr 0.03125 | ms/batch 207.83255 | loss  3.73 | ppl    41.56
## | epoch  93 |   900/ 1452 batches | lr 0.03125 | ms/batch 217.46269 | loss  3.73 | ppl    41.80
## | epoch  93 |  1000/ 1452 batches | lr 0.03125 | ms/batch 207.95238 | loss  3.70 | ppl    40.47
## | epoch  93 |  1100/ 1452 batches | lr 0.03125 | ms/batch 209.45098 | loss  3.84 | ppl    46.53
## | epoch  93 |  1200/ 1452 batches | lr 0.03125 | ms/batch 209.08127 | loss  3.79 | ppl    44.23
## | epoch  93 |  1300/ 1452 batches | lr 0.03125 | ms/batch 212.71757 | loss  3.59 | ppl    36.40
## | epoch  93 |  1400/ 1452 batches | lr 0.03125 | ms/batch 208.46679 | loss  3.75 | ppl    42.35
## -----------------------------------------------------------------------------------------
## | end of epoch  93 | time: 324.06s | valid loss  4.54 | valid ppl    93.29
## | end of epoch  93 | time: 324.06s | test loss  4.47 | test ppl    87.72
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch  94 |   100/ 1452 batches | lr 0.03125 | ms/batch 209.72549 | loss  3.87 | ppl    47.78
## | epoch  94 |   200/ 1452 batches | lr 0.03125 | ms/batch 217.70735 | loss  3.83 | ppl    46.27
## | epoch  94 |   300/ 1452 batches | lr 0.03125 | ms/batch 211.73868 | loss  3.81 | ppl    45.34
## | epoch  94 |   400/ 1452 batches | lr 0.03125 | ms/batch 208.09718 | loss  3.68 | ppl    39.63
## | epoch  94 |   500/ 1452 batches | lr 0.03125 | ms/batch 215.11998 | loss  3.72 | ppl    41.20
## | epoch  94 |   600/ 1452 batches | lr 0.03125 | ms/batch 211.90841 | loss  3.74 | ppl    42.21
## | epoch  94 |   700/ 1452 batches | lr 0.03125 | ms/batch 212.80235 | loss  3.77 | ppl    43.35
## | epoch  94 |   800/ 1452 batches | lr 0.03125 | ms/batch 207.63276 | loss  3.73 | ppl    41.59
## | epoch  94 |   900/ 1452 batches | lr 0.03125 | ms/batch 209.57076 | loss  3.73 | ppl    41.63
## | epoch  94 |  1000/ 1452 batches | lr 0.03125 | ms/batch 207.91735 | loss  3.70 | ppl    40.45
## | epoch  94 |  1100/ 1452 batches | lr 0.03125 | ms/batch 209.64559 | loss  3.82 | ppl    45.77
## | epoch  94 |  1200/ 1452 batches | lr 0.03125 | ms/batch 208.74158 | loss  3.78 | ppl    43.91
## | epoch  94 |  1300/ 1452 batches | lr 0.03125 | ms/batch 209.88047 | loss  3.59 | ppl    36.33
## | epoch  94 |  1400/ 1452 batches | lr 0.03125 | ms/batch 209.01132 | loss  3.75 | ppl    42.51
## -----------------------------------------------------------------------------------------
## | end of epoch  94 | time: 324.20s | valid loss  4.54 | valid ppl    93.29
## | end of epoch  94 | time: 324.20s | test loss  4.47 | test ppl    87.71
## -----------------------------------------------------------------------------------------
## | epoch  95 |   100/ 1452 batches | lr 0.03125 | ms/batch 212.48269 | loss  3.87 | ppl    47.94
## | epoch  95 |   200/ 1452 batches | lr 0.03125 | ms/batch 209.32102 | loss  3.83 | ppl    46.05
## | epoch  95 |   300/ 1452 batches | lr 0.03125 | ms/batch 208.78645 | loss  3.81 | ppl    45.29
## | epoch  95 |   400/ 1452 batches | lr 0.03125 | ms/batch 209.74063 | loss  3.68 | ppl    39.76
## | epoch  95 |   500/ 1452 batches | lr 0.03125 | ms/batch 214.11102 | loss  3.72 | ppl    41.32
## | epoch  95 |   600/ 1452 batches | lr 0.03125 | ms/batch 207.33796 | loss  3.74 | ppl    42.13
## | epoch  95 |   700/ 1452 batches | lr 0.03125 | ms/batch 208.96135 | loss  3.76 | ppl    43.00
## | epoch  95 |   800/ 1452 batches | lr 0.03125 | ms/batch 216.21887 | loss  3.74 | ppl    41.90
## | epoch  95 |   900/ 1452 batches | lr 0.03125 | ms/batch 212.43783 | loss  3.73 | ppl    41.77
## | epoch  95 |  1000/ 1452 batches | lr 0.03125 | ms/batch 207.83272 | loss  3.70 | ppl    40.35
## | epoch  95 |  1100/ 1452 batches | lr 0.03125 | ms/batch 209.04620 | loss  3.83 | ppl    46.18
## | epoch  95 |  1200/ 1452 batches | lr 0.03125 | ms/batch 208.90142 | loss  3.78 | ppl    43.96
## | epoch  95 |  1300/ 1452 batches | lr 0.03125 | ms/batch 209.39599 | loss  3.60 | ppl    36.45
## | epoch  95 |  1400/ 1452 batches | lr 0.03125 | ms/batch 217.09819 | loss  3.76 | ppl    42.73
## -----------------------------------------------------------------------------------------
## | end of epoch  95 | time: 324.41s | valid loss  4.54 | valid ppl    93.33
## | end of epoch  95 | time: 324.41s | test loss  4.47 | test ppl    87.73
## -----------------------------------------------------------------------------------------
## | epoch  96 |   100/ 1452 batches | lr 0.03125 | ms/batch 211.42396 | loss  3.87 | ppl    47.92
## | epoch  96 |   200/ 1452 batches | lr 0.03125 | ms/batch 210.06025 | loss  3.83 | ppl    45.97
## | epoch  96 |   300/ 1452 batches | lr 0.03125 | ms/batch 207.68759 | loss  3.82 | ppl    45.61
## | epoch  96 |   400/ 1452 batches | lr 0.03125 | ms/batch 212.91224 | loss  3.68 | ppl    39.72
## | epoch  96 |   500/ 1452 batches | lr 0.03125 | ms/batch 208.60666 | loss  3.72 | ppl    41.17
## | epoch  96 |   600/ 1452 batches | lr 0.03125 | ms/batch 209.15117 | loss  3.75 | ppl    42.67
## | epoch  96 |   700/ 1452 batches | lr 0.03125 | ms/batch 208.62677 | loss  3.77 | ppl    43.17
## | epoch  96 |   800/ 1452 batches | lr 0.03125 | ms/batch 209.29103 | loss  3.73 | ppl    41.82
## | epoch  96 |   900/ 1452 batches | lr 0.03125 | ms/batch 208.89144 | loss  3.73 | ppl    41.56
## | epoch  96 |  1000/ 1452 batches | lr 0.03125 | ms/batch 209.72069 | loss  3.70 | ppl    40.38
## | epoch  96 |  1100/ 1452 batches | lr 0.03125 | ms/batch 210.27003 | loss  3.83 | ppl    46.16
## | epoch  96 |  1200/ 1452 batches | lr 0.03125 | ms/batch 208.65170 | loss  3.78 | ppl    43.84
## | epoch  96 |  1300/ 1452 batches | lr 0.03125 | ms/batch 221.51853 | loss  3.59 | ppl    36.41
## | epoch  96 |  1400/ 1452 batches | lr 0.03125 | ms/batch 206.09431 | loss  3.74 | ppl    42.28
## -----------------------------------------------------------------------------------------
```

```
## | end of epoch  96 | time: 323.59s | valid loss  4.54 | valid ppl    93.33
## | end of epoch  96 | time: 323.59s | test loss  4.47 | test ppl    87.72
## -----------------------------------------------------------------------------------------
## | epoch  97 |   100/ 1452 batches | lr 0.03125 | ms/batch 211.76343 | loss  3.87 | ppl    48.01
## | epoch  97 |   200/ 1452 batches | lr 0.03125 | ms/batch 208.37215 | loss  3.83 | ppl    46.09
## | epoch  97 |   300/ 1452 batches | lr 0.03125 | ms/batch 208.71671 | loss  3.82 | ppl    45.53
## | epoch  97 |   400/ 1452 batches | lr 0.03125 | ms/batch 208.00741 | loss  3.69 | ppl    39.90
## | epoch  97 |   500/ 1452 batches | lr 0.03125 | ms/batch 207.59769 | loss  3.71 | ppl    41.05
## | epoch  97 |   600/ 1452 batches | lr 0.03125 | ms/batch 208.84168 | loss  3.75 | ppl    42.53
## | epoch  97 |   700/ 1452 batches | lr 0.03125 | ms/batch 209.55077 | loss  3.76 | ppl    43.03
## | epoch  97 |   800/ 1452 batches | lr 0.03125 | ms/batch 208.04246 | loss  3.73 | ppl    41.47
## | epoch  97 |   900/ 1452 batches | lr 0.03125 | ms/batch 208.47696 | loss  3.73 | ppl    41.81
## | epoch  97 |  1000/ 1452 batches | lr 0.03125 | ms/batch 211.83336 | loss  3.69 | ppl    40.08
## | epoch  97 |  1100/ 1452 batches | lr 0.03125 | ms/batch 207.86742 | loss  3.84 | ppl    46.39
## | epoch  97 |  1200/ 1452 batches | lr 0.03125 | ms/batch 208.01725 | loss  3.77 | ppl    43.53
## | epoch  97 |  1300/ 1452 batches | lr 0.03125 | ms/batch 209.62071 | loss  3.59 | ppl    36.21
## | epoch  97 |  1400/ 1452 batches | lr 0.03125 | ms/batch 209.37099 | loss  3.75 | ppl    42.50
## -----------------------------------------------------------------------------------------
## | end of epoch  97 | time: 321.77s | valid loss  4.54 | valid ppl    93.31
## | end of epoch  97 | time: 321.81s | test loss  4.47 | test ppl    87.72
## -----------------------------------------------------------------------------------------
## | epoch  98 |   100/ 1452 batches | lr 0.03125 | ms/batch 215.86923 | loss  3.87 | ppl    47.97
## | epoch  98 |   200/ 1452 batches | lr 0.03125 | ms/batch 207.49288 | loss  3.83 | ppl    46.10
## | epoch  98 |   300/ 1452 batches | lr 0.03125 | ms/batch 209.01132 | loss  3.82 | ppl    45.41
## | epoch  98 |   400/ 1452 batches | lr 0.03125 | ms/batch 210.17012 | loss  3.69 | ppl    39.93
## | epoch  98 |   500/ 1452 batches | lr 0.03125 | ms/batch 211.63356 | loss  3.72 | ppl    41.12
## | epoch  98 |   600/ 1452 batches | lr 0.03125 | ms/batch 208.72652 | loss  3.75 | ppl    42.66
## | epoch  98 |   700/ 1452 batches | lr 0.03125 | ms/batch 209.03134 | loss  3.75 | ppl    42.53
## | epoch  98 |   800/ 1452 batches | lr 0.03125 | ms/batch 208.01725 | loss  3.73 | ppl    41.54
## | epoch  98 |   900/ 1452 batches | lr 0.03125 | ms/batch 209.40584 | loss  3.74 | ppl    41.97
## | epoch  98 |  1000/ 1452 batches | lr 0.03125 | ms/batch 208.46422 | loss  3.70 | ppl    40.38
## | epoch  98 |  1100/ 1452 batches | lr 0.03125 | ms/batch 211.76341 | loss  3.84 | ppl    46.56
## | epoch  98 |  1200/ 1452 batches | lr 0.03125 | ms/batch 206.92346 | loss  3.78 | ppl    43.84
## | epoch  98 |  1300/ 1452 batches | lr 0.03125 | ms/batch 209.06637 | loss  3.59 | ppl    36.28
## | epoch  98 |  1400/ 1452 batches | lr 0.03125 | ms/batch 208.26235 | loss  3.75 | ppl    42.52
## -----------------------------------------------------------------------------------------
## | end of epoch  98 | time: 322.53s | valid loss  4.54 | valid ppl    93.31
## | end of epoch  98 | time: 322.53s | test loss  4.47 | test ppl    87.71
## -----------------------------------------------------------------------------------------
## | epoch  99 |   100/ 1452 batches | lr 0.03125 | ms/batch 210.59969 | loss  3.87 | ppl    48.01
## | epoch  99 |   200/ 1452 batches | lr 0.03125 | ms/batch 209.14131 | loss  3.83 | ppl    46.08
## | epoch  99 |   300/ 1452 batches | lr 0.03125 | ms/batch 211.95323 | loss  3.82 | ppl    45.44
## | epoch  99 |   400/ 1452 batches | lr 0.03125 | ms/batch 209.56567 | loss  3.68 | ppl    39.65
## | epoch  99 |   500/ 1452 batches | lr 0.03125 | ms/batch 211.56362 | loss  3.71 | ppl    40.76
## | epoch  99 |   600/ 1452 batches | lr 0.03125 | ms/batch 208.77159 | loss  3.75 | ppl    42.54
## | epoch  99 |   700/ 1452 batches | lr 0.03125 | ms/batch 218.47172 | loss  3.77 | ppl    43.19
## | epoch  99 |   800/ 1452 batches | lr 0.03125 | ms/batch 212.52264 | loss  3.73 | ppl    41.66
## | epoch  99 |   900/ 1452 batches | lr 0.03125 | ms/batch 207.79257 | loss  3.74 | ppl    41.91
## | epoch  99 |  1000/ 1452 batches | lr 0.03125 | ms/batch 210.15012 | loss  3.70 | ppl    40.45
## | epoch  99 |  1100/ 1452 batches | lr 0.03125 | ms/batch 212.26291 | loss  3.83 | ppl    46.21
## | epoch  99 |  1200/ 1452 batches | lr 0.03125 | ms/batch 206.27904 | loss  3.78 | ppl    43.83
## | epoch  99 |  1300/ 1452 batches | lr 0.03125 | ms/batch 208.79664 | loss  3.60 | ppl    36.45
## | epoch  99 |  1400/ 1452 batches | lr 0.03125 | ms/batch 208.02724 | loss  3.75 | ppl    42.67
## -----------------------------------------------------------------------------------------
## | end of epoch  99 | time: 323.91s | valid loss  4.54 | valid ppl    93.29
## | end of epoch  99 | time: 323.91s | test loss  4.47 | test ppl    87.70
## -----------------------------------------------------------------------------------------
## Save model!
##
## | epoch 100 |   100/ 1452 batches | lr 0.03125 | ms/batch 209.37826 | loss  3.87 | ppl    48.08
## | epoch 100 |   200/ 1452 batches | lr 0.03125 | ms/batch 208.93408 | loss  3.82 | ppl    45.77
## | epoch 100 |   300/ 1452 batches | lr 0.03125 | ms/batch 209.22605 | loss  3.82 | ppl    45.65
## | epoch 100 |   400/ 1452 batches | lr 0.03125 | ms/batch 208.74650 | loss  3.68 | ppl    39.62
## | epoch 100 |   500/ 1452 batches | lr 0.03125 | ms/batch 208.88148 | loss  3.71 | ppl    40.88
## | epoch 100 |   600/ 1452 batches | lr 0.03125 | ms/batch 208.92147 | loss  3.75 | ppl    42.46
## | epoch 100 |   700/ 1452 batches | lr 0.03125 | ms/batch 207.35304 | loss  3.77 | ppl    43.17
## | epoch 100 |   800/ 1452 batches | lr 0.03125 | ms/batch 217.04313 | loss  3.73 | ppl    41.66
## | epoch 100 |   900/ 1452 batches | lr 0.03125 | ms/batch 209.99030 | loss  3.74 | ppl    41.92
## | epoch 100 |  1000/ 1452 batches | lr 0.03125 | ms/batch 206.29429 | loss  3.69 | ppl    40.16
## | epoch 100 |  1100/ 1452 batches | lr 0.03125 | ms/batch 208.13713 | loss  3.83 | ppl    46.24
## | epoch 100 |  1200/ 1452 batches | lr 0.03125 | ms/batch 208.77153 | loss  3.78 | ppl    43.88
## | epoch 100 |  1300/ 1452 batches | lr 0.03125 | ms/batch 210.99423 | loss  3.59 | ppl    36.27
```

```
## | epoch 100 |  1400/ 1452 batches | lr 0.03125 | ms/batch 208.53672 | loss  3.75 | ppl    42.61
## -----------------------------------------------------------------------------------------
## | end of epoch 100 | time: 322.26s | valid loss  4.54 | valid ppl    93.32
## | end of epoch 100 | time: 322.26s | test loss  4.47 | test ppl    87.73
## -----------------------------------------------------------------------------------------
## =========================================================================================
## | End of training | test loss  4.47 | test ppl    87.70
## =========================================================================================
##
```