**Results of Char-aware NLM (Highway)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | English (PTB) | Czech | German | Spanish | French | Russian |
| Paper | 92.3 | 401 | 260 | 182 | 189 | 278 |
| Mine | 91.7 | 398 | 252 | 182 | 185 | 281 |

**Base parameter explicitly referred in paper**  
**\* configuration**  
(lr=1.0, seq\_len=35, char\_dim=15, highway\_layer=1, lstm\_layer=2, lstm\_dim=300, batch\_size=20, dropout=0.5, max\_norm=5.0, kernel\_width=[1, 2, 3, 4, 5, 6], kernel\_num = width\*25)

+ batch normalization after convolution layer (before highway layer)

+ embedding\_norm (max\_norm=5)

+ input data shuffling (freeze the random seed either in torch and numpy.)

+ max\_epoch 50 (but, if it reaching the result of paper (goal), early stop the training.)

(+ train with valid in other language except English)