hate as no-off

alpha=2

training 1 epoch...

3052.8668484687805 seconds used......

1 training finished! average train loss: 0.6591092342715011

total clf loss: 992.1966930944473 total adv loss: 637.2399819176644

evaluating...

Confusion matrix, without normalization

[[ 127 201 162]

[ 38 2549 110]

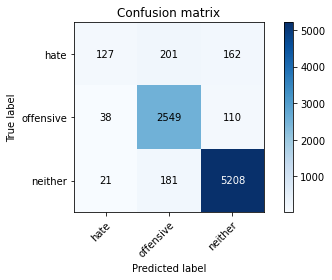
[ 21 181 5208]]

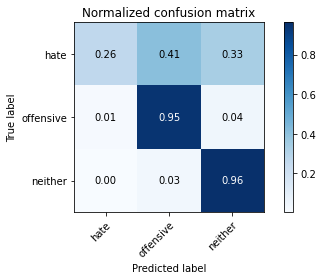
Normalized confusion matrix

[[0.25918367 0.41020408 0.33061224]

[0.01408973 0.94512421 0.04078606]

[0.0038817 0.03345656 0.96266174]]





precision recall f1-score support

hate 0.68 0.26 0.38 490

offensive 0.87 0.95 0.91 2697

neither 0.95 0.96 0.96 5410

accuracy 0.92 8597

macro avg 0.83 0.72 0.75 8597

weighted avg 0.91 0.92 0.91 8597

average eval\_loss: 0.2612306377828814, accuracy: 0.9170640921251599

training 2 epoch...

3053.484217643738 seconds used......

2 training finished! average train loss: 0.5401453669111191

total clf loss: 820.8959141038358 total adv loss: 518.3320018874947

evaluating...

Confusion matrix, without normalization

[[ 136 195 159]

[ 32 2554 111]

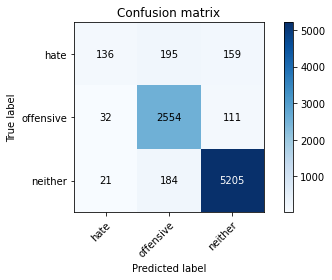
[ 21 184 5205]]

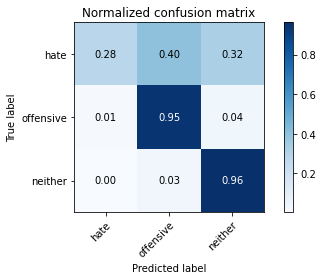
Normalized confusion matrix

[[0.27755102 0.39795918 0.3244898 ]

[0.01186504 0.94697812 0.04115684]

[0.0038817 0.03401109 0.96210721]]





precision recall f1-score support

hate 0.72 0.28 0.40 490

offensive 0.87 0.95 0.91 2697

neither 0.95 0.96 0.96 5410

accuracy 0.92 8597

macro avg 0.85 0.73 0.75 8597

weighted avg 0.91 0.92 0.91 8597

average eval\_loss: 0.2627821897654686, accuracy: 0.9183436082354309

training 3 epoch...

3052.7231590747833 seconds used......

3 training finished! average train loss: 0.438451979528949

total clf loss: 672.974058855325 total adv loss: 417.4311496524606

evaluating...

Confusion matrix, without normalization

[[ 184 161 145]

[ 79 2490 128]

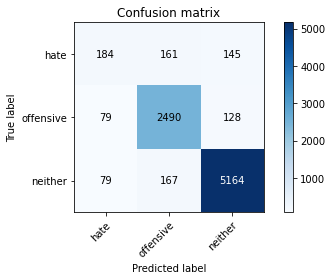
[ 79 167 5164]]

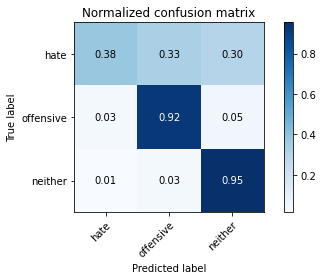
Normalized confusion matrix

[[0.3755102 0.32857143 0.29591837]

[0.02929181 0.92324805 0.04746014]

[0.01460259 0.03086876 0.95452865]]





precision recall f1-score support

hate 0.54 0.38 0.44 490

offensive 0.88 0.92 0.90 2697

neither 0.95 0.95 0.95 5410

accuracy 0.91 8597

macro avg 0.79 0.75 0.77 8597

weighted avg 0.91 0.91 0.91 8597

average eval\_loss: 0.290996188423488, accuracy: 0.9117133883912993

training 4 epoch...

3051.849383831024 seconds used......

4 training finished! average train loss: 0.3322301279970367

total clf loss: 513.7031463840976 total adv loss: 314.4181316415197

evaluating...

Confusion matrix, without normalization

[[ 183 163 144]

[ 111 2457 129]

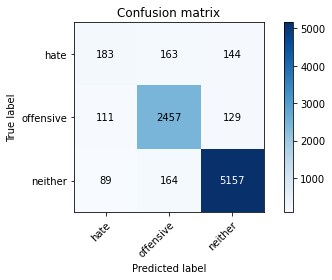
[ 89 164 5157]]

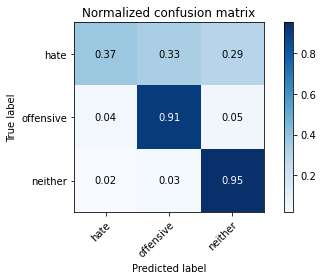
Normalized confusion matrix

[[0.37346939 0.33265306 0.29387755]

[0.04115684 0.91101224 0.04783092]

[0.01645102 0.03031423 0.95323475]]





precision recall f1-score support

hate 0.48 0.37 0.42 490

offensive 0.88 0.91 0.90 2697

neither 0.95 0.95 0.95 5410

accuracy 0.91 8597

macro avg 0.77 0.75 0.76 8597

weighted avg 0.90 0.91 0.90 8597

average eval\_loss: 0.33296782461341556, accuracy: 0.90694428288938

training 5 epoch...

3052.064789056778 seconds used......

5 training finished! average train loss: 0.25803807701631964

total clf loss: 394.7049998618895 total adv loss: 246.34397333697234

evaluating...

Confusion matrix, without normalization

[[ 161 171 158]

[ 97 2483 117]

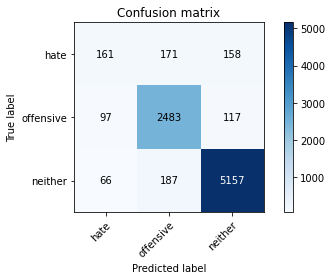
[ 66 187 5157]]

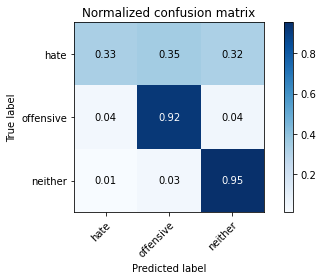
Normalized confusion matrix

[[0.32857143 0.34897959 0.32244898]

[0.03596589 0.92065258 0.04338154]

[0.01219963 0.03456562 0.95323475]]





precision recall f1-score support

hate 0.50 0.33 0.40 490

offensive 0.87 0.92 0.90 2697

neither 0.95 0.95 0.95 5410

accuracy 0.91 8597

macro avg 0.77 0.73 0.75 8597

weighted avg 0.90 0.91 0.90 8597

average eval\_loss: 0.40423222949273535, accuracy: 0.9074095614749331

model of 3 epoch

Confusion matrix, without normalization

[[ 182 192 158]

[ 68 2530 112]

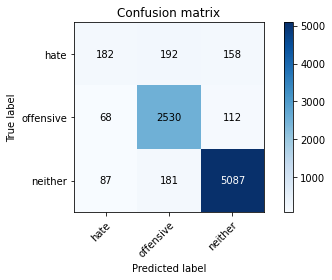
[ 87 181 5087]]

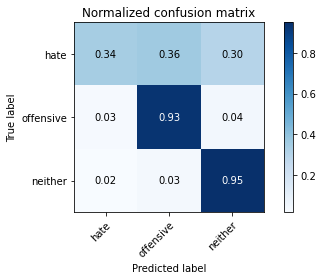
Normalized confusion matrix

[[0.34210526 0.36090226 0.29699248]

[0.02509225 0.93357934 0.04132841]

[0.0162465 0.03380019 0.94995331]]





precision recall f1-score support

hate 0.54 0.34 0.42 532

offensive 0.87 0.93 0.90 2710

neither 0.95 0.95 0.95 5355

accuracy 0.91 8597

macro avg 0.79 0.74 0.76 8597

weighted avg 0.90 0.91 0.90 8597

average eval\_loss: 0.3055757470407285, accuracy: 0.9071769221821566

[23]:

0.9071769221821566