train hate offensive none

7291 37070 51065 95426

7.6% 38.8% 53.5%

test 2540 10687 16947 30174

8.4% 35.4% 56.2%

hate as offensive alpha=0.05

1 run......

Downloading: 100%

466/466 [00:00<00:00, 16.7kB/s]

Downloading: 100%

542M/542M [00:14<00:00, 43.5MB/s]

training 1 epoch...

4240.714061498642 seconds used......

1 training finished! average train loss: 0.38868322511447934

total clf loss: 1919.70574991405 total adv loss: 1298.1879784241319

evaluating...

Confusion matrix, without normalization

[[ 623 923 994]

[ 164 9640 883]

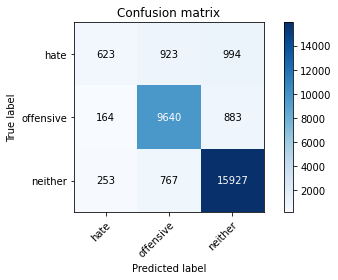
[ 253 767 15927]]

Normalized confusion matrix

[[0.24527559 0.36338583 0.39133858]

[0.01534575 0.9020305 0.08262375]

[0.0149289 0.04525875 0.93981236]]





precision recall f1-score support

hate 0.60 0.25 0.35 2540

offensive 0.85 0.90 0.88 10687

neither 0.89 0.94 0.92 16947

accuracy 0.87 30174

macro avg 0.78 0.70 0.71 30174

weighted avg 0.85 0.87 0.85 30174

average eval\_loss: 0.37247962566960363, accuracy: 0.8679657983694572

training 2 epoch...

4240.064320087433 seconds used......

2 training finished! average train loss: 0.30387746759597445

total clf loss: 1498.2815652117133 total adv loss: 963.5658016717061

evaluating...

Confusion matrix, without normalization

[[ 737 884 919]

[ 258 9664 765]

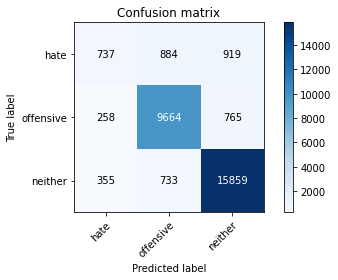
[ 355 733 15859]]

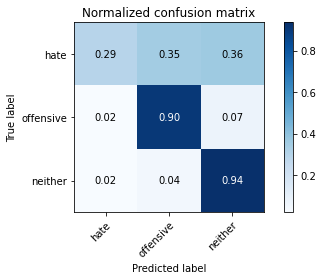
Normalized confusion matrix

[[0.29015748 0.3480315 0.36181102]

[0.02414148 0.90427622 0.0715823 ]

[0.02094766 0.04325249 0.93579985]]





precision recall f1-score support

hate 0.55 0.29 0.38 2540

offensive 0.86 0.90 0.88 10687

neither 0.90 0.94 0.92 16947

accuracy 0.87 30174

macro avg 0.77 0.71 0.73 30174

weighted avg 0.86 0.87 0.86 30174

average eval\_loss: 0.37068973719392334, accuracy: 0.8702856764101544

training 3 epoch...

4242.456351041794 seconds used......

3 training finished! average train loss: 0.23346409569525653

total clf loss: 1149.1962361913174 total adv loss: 702.11142817582

evaluating...

Confusion matrix, without normalization

[[ 813 959 768]

[ 281 9700 706]

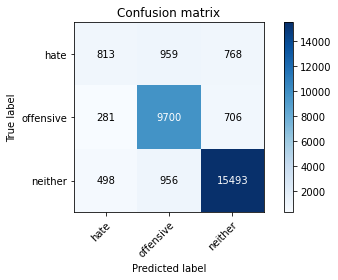
[ 498 956 15493]]

Normalized confusion matrix

[[0.32007874 0.37755906 0.3023622 ]

[0.02629363 0.9076448 0.06606157]

[0.02938573 0.05641116 0.9142031 ]]





precision recall f1-score support

hate 0.51 0.32 0.39 2540

offensive 0.84 0.91 0.87 10687

neither 0.91 0.91 0.91 16947

accuracy 0.86 30174

macro avg 0.75 0.71 0.73 30174

weighted avg 0.85 0.86 0.85 30174

average eval\_loss: 0.4178608756258126, accuracy: 0.86186783323391

training 4 epoch...

4242.636656045914 seconds used......

4 training finished! average train loss: 0.1660850574154124

total clf loss: 816.387321786955 total adv loss: 476.5885511174565

evaluating...

Confusion matrix, without normalization

[[ 805 763 972]

[ 391 9382 914]

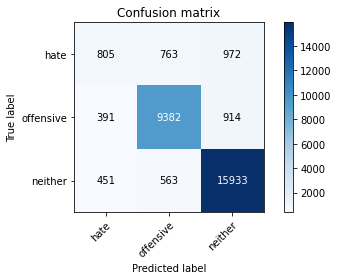
[ 451 563 15933]]

Normalized confusion matrix

[[0.31692913 0.3003937 0.38267717]

[0.03658651 0.87788902 0.08552447]

[0.02661238 0.03322122 0.9401664 ]]





precision recall f1-score support

hate 0.49 0.32 0.38 2540

offensive 0.88 0.88 0.88 10687

neither 0.89 0.94 0.92 16947

accuracy 0.87 30174

macro avg 0.75 0.71 0.73 30174

weighted avg 0.85 0.87 0.86 30174

average eval\_loss: 0.5269041202016264, accuracy: 0.8656459203287599