training adversary 1 epoch...

891.5740084648132 seconds used......

1 training finished! train loss: 498.20323710143566

evaluating...

Confusion matrix, without normalization

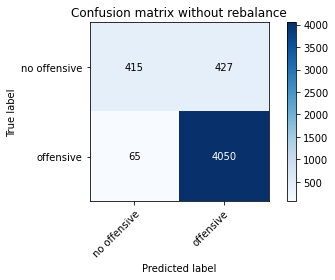
[[ 415 427]

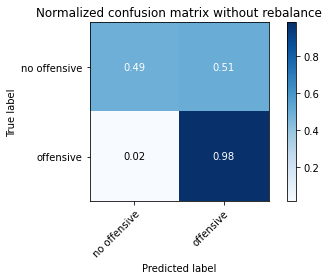
[ 65 4050]]

Normalized confusion matrix

[[0.49287411 0.50712589]

[0.01579587 0.98420413]]





precision recall f1-score support

no offensive 0.86 0.49 0.63 842

offensive 0.90 0.98 0.94 4115

accuracy 0.90 4957

macro avg 0.88 0.74 0.79 4957

weighted avg 0.90 0.90 0.89 4957

training adversary 2 epoch...

893.2636218070984 seconds used......

2 training finished! train loss: 406.94186740298755

evaluating...

Confusion matrix, without normalization

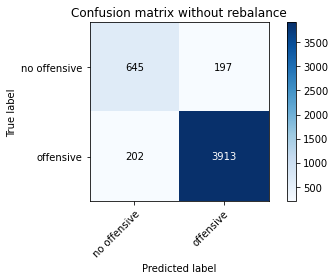
[[ 645 197]

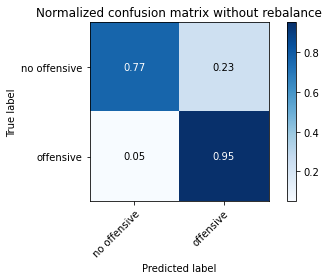
[ 202 3913]]

Normalized confusion matrix

[[0.76603325 0.23396675]

[0.0490887 0.9509113 ]]





precision recall f1-score support

no offensive 0.76 0.77 0.76 842

offensive 0.95 0.95 0.95 4115

accuracy 0.92 4957

macro avg 0.86 0.86 0.86 4957

weighted avg 0.92 0.92 0.92 4957

alpha = 0.05

training 1 epoch...

949.3248662948608 seconds used......

1 training finished! average train loss: -0.3896813025558109

total clf loss: 593.0612462237477 total adv loss: 27315.984932668507

evaluating...

Confusion matrix, without normalization

[[ 117 138 39]

[ 116 3601 104]

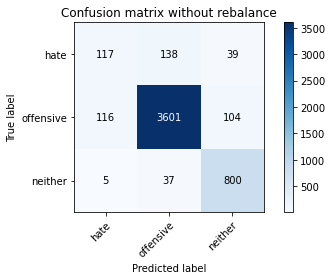
[ 5 37 800]]

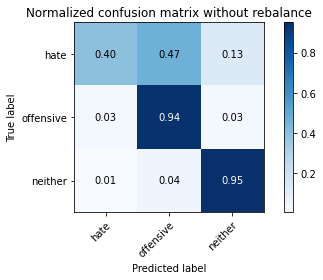
Normalized confusion matrix

[[0.39795918 0.46938776 0.13265306]

[0.03035854 0.94242345 0.02721801]

[0.00593824 0.04394299 0.95011876]]





precision recall f1-score support

hate 0.49 0.40 0.44 294

offensive 0.95 0.94 0.95 3821

neither 0.85 0.95 0.90 842

accuracy 0.91 4957

macro avg 0.76 0.76 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.25757695604566366, accuracy: 0.9114383699818439

training 2 epoch...

948.4784214496613 seconds used......

2 training finished! average train loss: -0.5383754046168883

total clf loss: 447.6900619259104 total adv loss: 30305.769266605377

evaluating...

Confusion matrix, without normalization

[[ 72 202 20]

[ 44 3704 73]

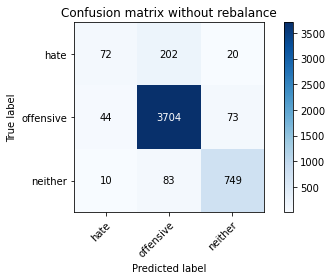
[ 10 83 749]]

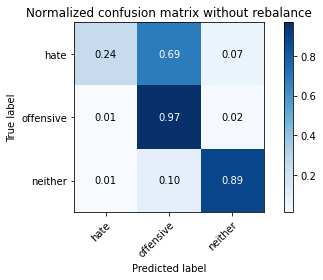
Normalized confusion matrix

[[0.24489796 0.68707483 0.06802721]

[0.01151531 0.96937974 0.01910495]

[0.01187648 0.09857482 0.88954869]]





precision recall f1-score support

hate 0.57 0.24 0.34 294

offensive 0.93 0.97 0.95 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.80 0.70 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.23994861985932314, accuracy: 0.9128505144240469

training 3 epoch...

948.828705072403 seconds used......

3 training finished! average train loss: -0.6394804770665161

total clf loss: 338.966584214475 total adv loss: 32141.12693119049

evaluating...

Confusion matrix, without normalization

[[ 29 255 10]

[ 10 3764 47]

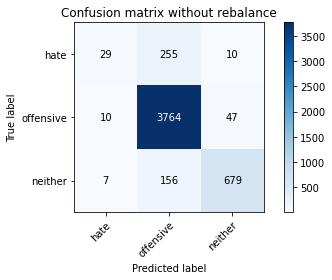
[ 7 156 679]]

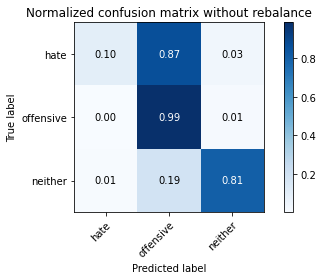
Normalized confusion matrix

[[0.09863946 0.86734694 0.03401361]

[0.00261712 0.98508244 0.01230044]

[0.00831354 0.18527316 0.8064133 ]]





precision recall f1-score support

hate 0.63 0.10 0.17 294

offensive 0.90 0.99 0.94 3821

neither 0.92 0.81 0.86 842

accuracy 0.90 4957

macro avg 0.82 0.63 0.66 4957

weighted avg 0.89 0.90 0.88 4957

average eval\_loss: 0.28026105210770885, accuracy: 0.9021585636473674

training 4 epoch...

949.7152018547058 seconds used......

4 training finished! average train loss: -2.9072837145474

total clf loss: 329.61297197197564 total adv loss: 121895.12926483154

evaluating...

Confusion matrix, without normalization

[[ 89 174 31]

[ 72 3641 108]

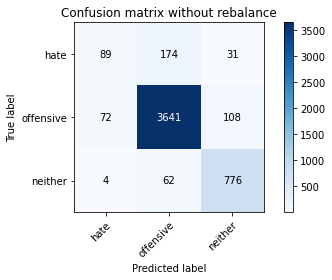
[ 4 62 776]]

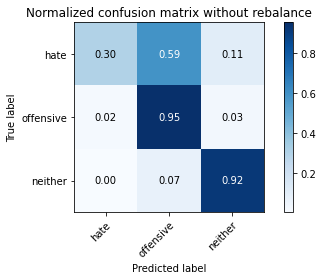
Normalized confusion matrix

[[0.30272109 0.59183673 0.10544218]

[0.01884323 0.95289191 0.02826485]

[0.00475059 0.0736342 0.9216152 ]]





precision recall f1-score support

hate 0.54 0.30 0.39 294

offensive 0.94 0.95 0.95 3821

neither 0.85 0.92 0.88 842

accuracy 0.91 4957

macro avg 0.78 0.73 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.27281670417306164, accuracy: 0.9090175509380674

training 5 epoch...

948.8770034313202 seconds used......

5 training finished! average train loss: -3.9506635414309654

total clf loss: 250.71967641753145 total adv loss: 161697.70657730103

evaluating...

Confusion matrix, without normalization

[[ 90 173 31]

[ 68 3645 108]

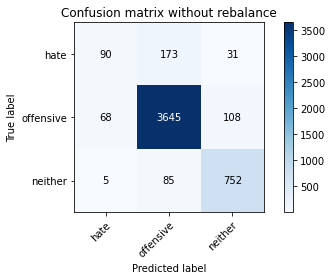
[ 5 85 752]]

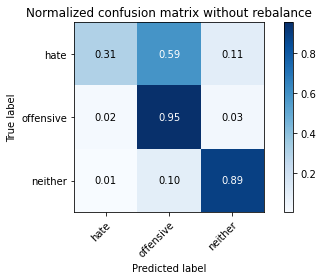
Normalized confusion matrix

[[0.30612245 0.58843537 0.10544218]

[0.01779639 0.95393876 0.02826485]

[0.00593824 0.10095012 0.89311164]]





precision recall f1-score support

hate 0.55 0.31 0.39 294

offensive 0.93 0.95 0.94 3821

neither 0.84 0.89 0.87 842

accuracy 0.91 4957

macro avg 0.78 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.31455428750984626, accuracy: 0.905184587452088