training 1 epoch...

106.93076252937317 seconds used......

1 training finished! average train loss: 0.6410753614353862

total clf loss: 116.06140437722206 total adv loss: 64.64281207323074

evaluating...

Confusion matrix, without normalization

[[ 89 25 20]

[ 329 1556 52]

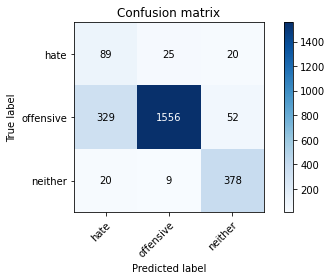
[ 20 9 378]]

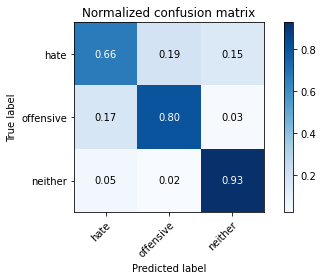
Normalized confusion matrix

[[0.6641791 0.18656716 0.14925373]

[0.16985028 0.80330408 0.02684564]

[0.04914005 0.02211302 0.92874693]]





precision recall f1-score support

hate 0.20 0.66 0.31 134

offensive 0.98 0.80 0.88 1937

neither 0.84 0.93 0.88 407

accuracy 0.82 2478

macro avg 0.67 0.80 0.69 2478

weighted avg 0.91 0.82 0.85 2478

average eval\_loss: 0.47400482683893175, accuracy: 0.8163841807909604

training 2 epoch...

106.92630577087402 seconds used......

2 training finished! average train loss: 0.36900606243447825

total clf loss: 67.30231335759163 total adv loss: 47.144924916327

evaluating...

Confusion matrix, without normalization

[[ 97 23 14]

[ 285 1608 44]

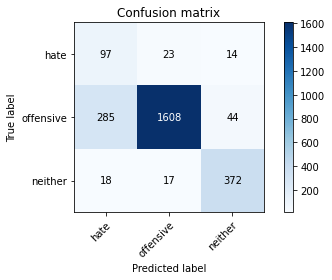
[ 18 17 372]]

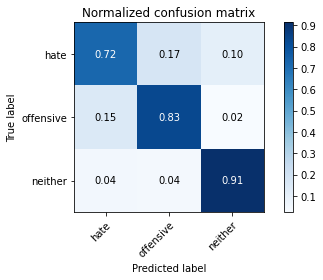
Normalized confusion matrix

[[0.7238806 0.17164179 0.10447761]

[0.14713474 0.83014972 0.02271554]

[0.04422604 0.04176904 0.91400491]]





precision recall f1-score support

hate 0.24 0.72 0.36 134

offensive 0.98 0.83 0.90 1937

neither 0.87 0.91 0.89 407

accuracy 0.84 2478

macro avg 0.69 0.82 0.72 2478

weighted avg 0.92 0.84 0.87 2478

average eval\_loss: 0.40852878152603106, accuracy: 0.8381759483454398

training 3 epoch...

106.96056342124939 seconds used......

3 training finished! average train loss: 0.24712451636812396

total clf loss: 45.9170378446579 total adv loss: 48.46246127039194

evaluating...

Confusion matrix, without normalization

[[ 94 24 16]

[ 290 1599 48]

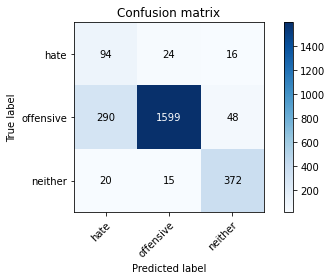
[ 20 15 372]]

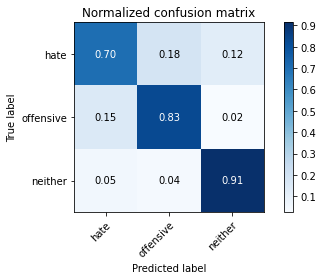
Normalized confusion matrix

[[0.70149254 0.17910448 0.11940299]

[0.14971606 0.82550336 0.02478059]

[0.04914005 0.03685504 0.91400491]]





precision recall f1-score support

hate 0.23 0.70 0.35 134

offensive 0.98 0.83 0.89 1937

neither 0.85 0.91 0.88 407

accuracy 0.83 2478

macro avg 0.69 0.81 0.71 2478

weighted avg 0.92 0.83 0.86 2478

average eval\_loss: 0.42091856608467715, accuracy: 0.8333333333333334

training 4 epoch...

106.88309407234192 seconds used......

4 training finished! average train loss: 0.13645912597844886

total clf loss: 26.5352702466771 total adv loss: 50.36928062886

evaluating...

Confusion matrix, without normalization

[[ 87 33 14]

[ 243 1655 39]

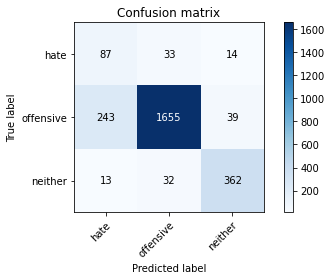
[ 13 32 362]]

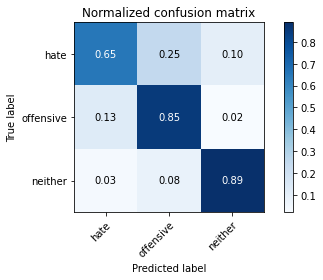
Normalized confusion matrix

[[0.64925373 0.24626866 0.10447761]

[0.12545173 0.85441404 0.02013423]

[0.03194103 0.07862408 0.88943489]]





precision recall f1-score support

hate 0.25 0.65 0.36 134

offensive 0.96 0.85 0.91 1937

neither 0.87 0.89 0.88 407

accuracy 0.85 2478

macro avg 0.70 0.80 0.72 2478

weighted avg 0.91 0.85 0.87 2478

average eval\_loss: 0.41369299990155045, accuracy: 0.8490718321226796

training 5 epoch...

106.82519006729126 seconds used......

5 training finished! average train loss: 0.07973556039823135

total clf loss: 16.580736194271594 total adv loss: 50.94555131345987

evaluating...

Confusion matrix, without normalization

[[ 72 40 22]

[ 155 1704 78]

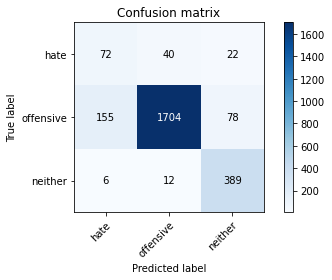
[ 6 12 389]]

Normalized confusion matrix

[[0.53731343 0.29850746 0.1641791 ]

[0.08002065 0.87971089 0.04026846]

[0.01474201 0.02948403 0.95577396]]





precision recall f1-score support

hate 0.31 0.54 0.39 134

offensive 0.97 0.88 0.92 1937

neither 0.80 0.96 0.87 407

accuracy 0.87 2478

macro avg 0.69 0.79 0.73 2478

weighted avg 0.91 0.87 0.89 2478

average eval\_loss: 0.4528681191826059, accuracy: 0.8736884584342212

add Codeadd Markdown

[33]:

best\_model **=** torch.load('./2Davidson\_re')

evaluate(best\_model, 0, test\_dataloader, 'test', '0')

Confusion matrix, without normalization

[[ 85 33 6]

[ 312 1579 51]

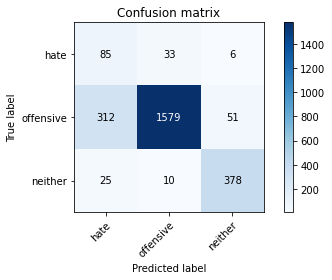
[ 25 10 378]]

Normalized confusion matrix

[[0.68548387 0.26612903 0.0483871 ]

[0.16065911 0.8130793 0.02626159]

[0.06053269 0.02421308 0.91525424]]





precision recall f1-score support

hate 0.20 0.69 0.31 124

offensive 0.97 0.81 0.89 1942

neither 0.87 0.92 0.89 413

accuracy 0.82 2479

macro avg 0.68 0.80 0.70 2479

weighted avg 0.92 0.82 0.86 2479

average eval\_loss: 0.45168009154017896, accuracy: 0.8237192416296893

[33]:

0.8237192416296893

2 run

training 1 epoch...

106.88987517356873 seconds used......

1 training finished! average train loss: 0.6473863983357494

total clf loss: 117.16383963823318 total adv loss: 64.4766666740179

evaluating...

Confusion matrix, without normalization

[[ 105 11 18]

[ 478 1395 64]

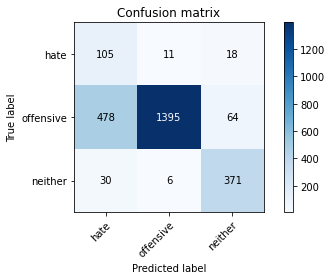
[ 30 6 371]]

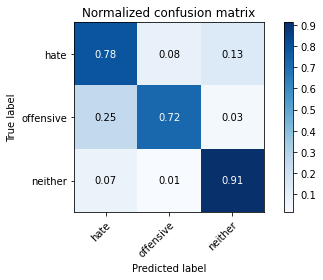
Normalized confusion matrix

[[0.78358209 0.08208955 0.13432836]

[0.24677336 0.72018585 0.03304078]

[0.07371007 0.01474201 0.91154791]]





precision recall f1-score support

hate 0.17 0.78 0.28 134

offensive 0.99 0.72 0.83 1937

neither 0.82 0.91 0.86 407

accuracy 0.76 2478

macro avg 0.66 0.81 0.66 2478

weighted avg 0.92 0.76 0.81 2478

average eval\_loss: 0.6767436696156379, accuracy: 0.755044390637611

training 2 epoch...

106.85667538642883 seconds used......

2 training finished! average train loss: 0.37787331852384587

total clf loss: 69.04987386614084 total adv loss: 50.883392594754696

evaluating...

Confusion matrix, without normalization

[[ 94 21 19]

[ 325 1549 63]

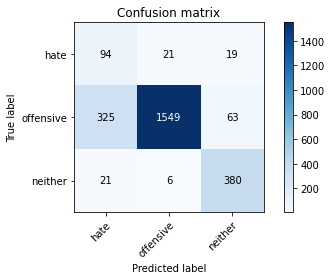
[ 21 6 380]]

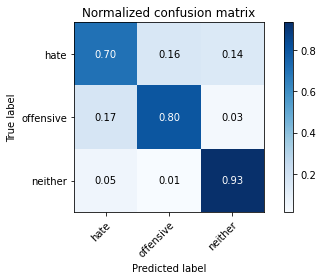
Normalized confusion matrix

[[0.70149254 0.15671642 0.14179104]

[0.16778523 0.79969024 0.03252452]

[0.05159705 0.01474201 0.93366093]]





precision recall f1-score support

hate 0.21 0.70 0.33 134

offensive 0.98 0.80 0.88 1937

neither 0.82 0.93 0.87 407

accuracy 0.82 2478

macro avg 0.67 0.81 0.69 2478

weighted avg 0.91 0.82 0.85 2478

average eval\_loss: 0.45851135530298753, accuracy: 0.8163841807909604

training 3 epoch...

106.94956398010254 seconds used......

3 training finished! average train loss: 0.25309913915539667

total clf loss: 47.09286404028535 total adv loss: 50.94830932468176

evaluating...

Confusion matrix, without normalization

[[ 86 29 19]

[ 217 1652 68]

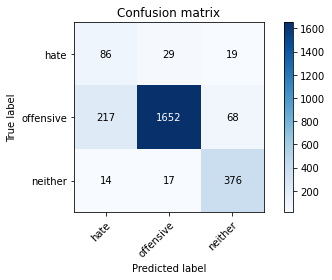
[ 14 17 376]]

Normalized confusion matrix

[[0.64179104 0.21641791 0.14179104]

[0.11202891 0.85286526 0.03510583]

[0.03439803 0.04176904 0.92383292]]





precision recall f1-score support

hate 0.27 0.64 0.38 134

offensive 0.97 0.85 0.91 1937

neither 0.81 0.92 0.86 407

accuracy 0.85 2478

macro avg 0.69 0.81 0.72 2478

weighted avg 0.91 0.85 0.87 2478

average eval\_loss: 0.4138399895520941, accuracy: 0.8531073446327684

training 4 epoch...

106.80899310112 seconds used......

4 training finished! average train loss: 0.1424362132357518

total clf loss: 27.555151605978608 total adv loss: 49.727562852203846

evaluating...

Confusion matrix, without normalization

[[ 86 30 18]

[ 210 1682 45]

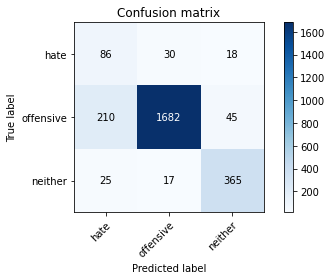
[ 25 17 365]]

Normalized confusion matrix

[[0.64179104 0.2238806 0.13432836]

[0.10841507 0.86835312 0.0232318 ]

[0.06142506 0.04176904 0.8968059 ]]





precision recall f1-score support

hate 0.27 0.64 0.38 134

offensive 0.97 0.87 0.92 1937

neither 0.85 0.90 0.87 407

accuracy 0.86 2478

macro avg 0.70 0.80 0.72 2478

weighted avg 0.91 0.86 0.88 2478

average eval\_loss: 0.4507750809162615, accuracy: 0.860774818401937

training 5 epoch...

106.90650391578674 seconds used......

5 training finished! average train loss: 0.06804601686731489

total clf loss: 14.558896010741591 total adv loss: 51.65593931823969

evaluating...

Confusion matrix, without normalization

[[ 92 30 12]

[ 324 1574 39]

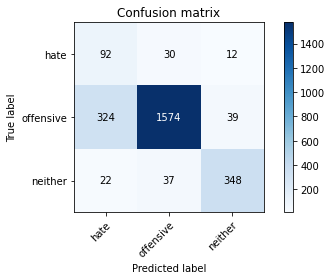
[ 22 37 348]]

Normalized confusion matrix

[[0.68656716 0.2238806 0.08955224]

[0.16726897 0.8125968 0.02013423]

[0.05405405 0.09090909 0.85503686]]





precision recall f1-score support

hate 0.21 0.69 0.32 134

offensive 0.96 0.81 0.88 1937

neither 0.87 0.86 0.86 407

accuracy 0.81 2478

macro avg 0.68 0.78 0.69 2478

weighted avg 0.90 0.81 0.85 2478

average eval\_loss: 0.6462068078258345, accuracy: 0.8127522195318806

Confusion matrix, without normalization

[[ 99 20 5]

[ 472 1399 71]

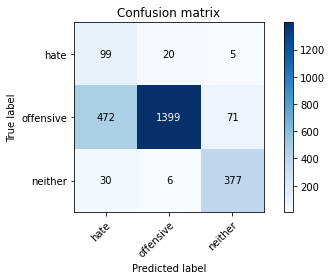
[ 30 6 377]]

Normalized confusion matrix

[[0.7983871 0.16129032 0.04032258]

[0.2430484 0.72039135 0.03656025]

[0.07263923 0.01452785 0.91283293]]





precision recall f1-score support

hate 0.16 0.80 0.27 124

offensive 0.98 0.72 0.83 1942

neither 0.83 0.91 0.87 413

accuracy 0.76 2479

macro avg 0.66 0.81 0.66 2479

weighted avg 0.92 0.76 0.81 2479

average eval\_loss: 0.6931507666745493, accuracy: 0.7563533682936668

[35]:

0.7563533682936668

3 run

training 1 epoch...

106.95307922363281 seconds used......

1 training finished! average train loss: 0.6143084647641941

total clf loss: 111.28153066337109 total adv loss: 63.26482602208853

evaluating...

Confusion matrix, without normalization

[[ 109 11 14]

[ 499 1385 53]

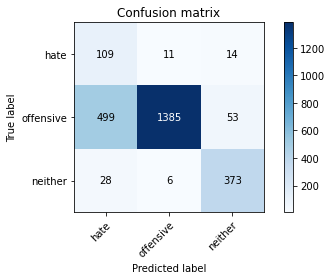
[ 28 6 373]]

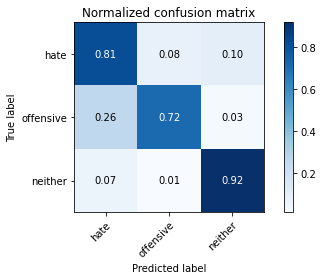
Normalized confusion matrix

[[0.81343284 0.08208955 0.10447761]

[0.25761487 0.71502323 0.0273619 ]

[0.06879607 0.01474201 0.91646192]]





precision recall f1-score support

hate 0.17 0.81 0.28 134

offensive 0.99 0.72 0.83 1937

neither 0.85 0.92 0.88 407

accuracy 0.75 2478

macro avg 0.67 0.81 0.66 2478

weighted avg 0.92 0.75 0.81 2478

average eval\_loss: 0.6207857540538234, accuracy: 0.7534301856335754

training 2 epoch...

106.86069321632385 seconds used......

2 training finished! average train loss: 0.37122893159870396

total clf loss: 67.7756290435791 total adv loss: 48.78674238175154

evaluating...

Confusion matrix, without normalization

[[ 86 32 16]

[ 228 1657 52]

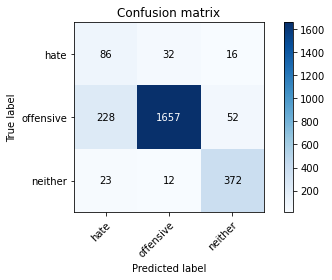
[ 23 12 372]]

Normalized confusion matrix

[[0.64179104 0.23880597 0.11940299]

[0.1177078 0.85544657 0.02684564]

[0.05651106 0.02948403 0.91400491]]





precision recall f1-score support

hate 0.26 0.64 0.37 134

offensive 0.97 0.86 0.91 1937

neither 0.85 0.91 0.88 407

accuracy 0.85 2478

macro avg 0.69 0.80 0.72 2478

weighted avg 0.91 0.85 0.88 2478

average eval\_loss: 0.417881355410622, accuracy: 0.8535108958837773

training 3 epoch...

106.89466214179993 seconds used......

3 training finished! average train loss: 0.24465730427552693

total clf loss: 45.45443149283528 total adv loss: 47.89491684362292

evaluating...

Confusion matrix, without normalization

[[ 101 19 14]

[ 381 1505 51]

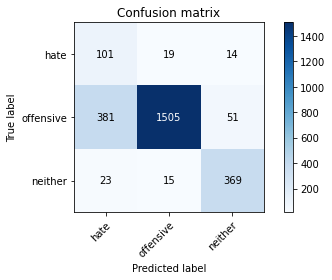
[ 23 15 369]]

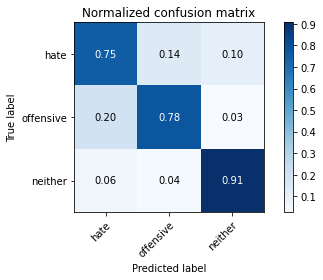
Normalized confusion matrix

[[0.75373134 0.14179104 0.10447761]

[0.19669592 0.7769747 0.02632938]

[0.05651106 0.03685504 0.90663391]]





precision recall f1-score support

hate 0.20 0.75 0.32 134

offensive 0.98 0.78 0.87 1937

neither 0.85 0.91 0.88 407

accuracy 0.80 2478

macro avg 0.68 0.81 0.69 2478

weighted avg 0.91 0.80 0.84 2478

average eval\_loss: 0.4813444235872838, accuracy: 0.7970137207425343

training 4 epoch...

106.89991164207458 seconds used......

4 training finished! average train loss: 0.1527483756018972

total clf loss: 29.242574896663427 total adv loss: 47.1772131100297

evaluating...

Confusion matrix, without normalization

[[ 75 43 16]

[ 143 1740 54]

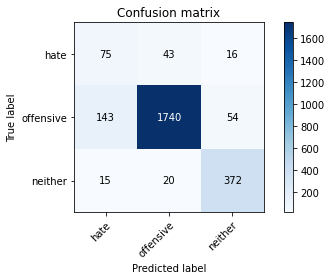
[ 15 20 372]]

Normalized confusion matrix

[[0.55970149 0.32089552 0.11940299]

[0.0738255 0.89829633 0.02787816]

[0.03685504 0.04914005 0.91400491]]





precision recall f1-score support

hate 0.32 0.56 0.41 134

offensive 0.97 0.90 0.93 1937

neither 0.84 0.91 0.88 407

accuracy 0.88 2478

macro avg 0.71 0.79 0.74 2478

weighted avg 0.91 0.88 0.89 2478

average eval\_loss: 0.37723102641382045, accuracy: 0.8825665859564165

training 5 epoch...

107.00659251213074 seconds used......

5 training finished! average train loss: 0.0764434445723468

total clf loss: 15.929378068540245 total adv loss: 49.50663647055626

evaluating...

Confusion matrix, without normalization

[[ 103 18 13]

[ 348 1530 59]

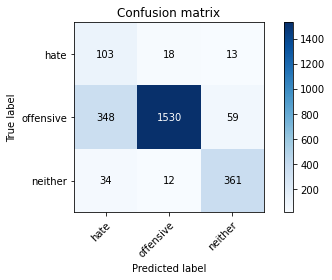
[ 34 12 361]]

Normalized confusion matrix

[[0.76865672 0.13432836 0.09701493]

[0.17965927 0.78988126 0.03045947]

[0.08353808 0.02948403 0.88697789]]





precision recall f1-score support

hate 0.21 0.77 0.33 134

offensive 0.98 0.79 0.88 1937

neither 0.83 0.89 0.86 407

accuracy 0.80 2478

macro avg 0.68 0.82 0.69 2478

weighted avg 0.92 0.80 0.84 2478

average eval\_loss: 0.6912833828959735, accuracy: 0.804681194511703

Confusion matrix, without normalization

[[ 98 19 7]

[ 488 1388 66]

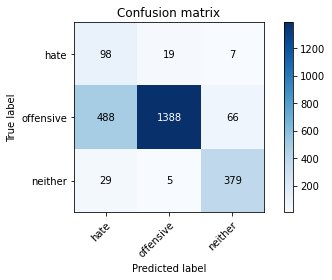
[ 29 5 379]]

Normalized confusion matrix

[[0.79032258 0.15322581 0.05645161]

[0.25128733 0.71472709 0.03398558]

[0.07021792 0.01210654 0.91767554]]





precision recall f1-score support

hate 0.16 0.79 0.27 124

offensive 0.98 0.71 0.83 1942

neither 0.84 0.92 0.88 413

accuracy 0.75 2479

macro avg 0.66 0.81 0.66 2479

weighted avg 0.92 0.75 0.81 2479

average eval\_loss: 0.6460328926482508, accuracy: 0.7523194836627672

[37]:

0.7523194836627672

4 run

training 1 epoch...

107.10649156570435 seconds used......

1 training finished! average train loss: 0.627587721683085

total clf loss: 113.63946677744389 total adv loss: 63.68055573105812

evaluating...

Confusion matrix, without normalization

[[ 90 25 19]

[ 341 1516 80]

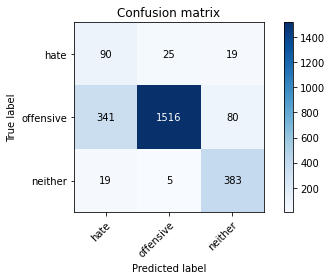
[ 19 5 383]]

Normalized confusion matrix

[[0.67164179 0.18656716 0.14179104]

[0.17604543 0.78265359 0.04130098]

[0.04668305 0.01228501 0.94103194]]





precision recall f1-score support

hate 0.20 0.67 0.31 134

offensive 0.98 0.78 0.87 1937

neither 0.79 0.94 0.86 407

accuracy 0.80 2478

macro avg 0.66 0.80 0.68 2478

weighted avg 0.91 0.80 0.84 2478

average eval\_loss: 0.5354033337245064, accuracy: 0.8026634382566586

training 2 epoch...

106.93660497665405 seconds used......

2 training finished! average train loss: 0.3725725298070095

total clf loss: 67.99587073922157 total adv loss: 48.46210526674986

evaluating...

Confusion matrix, without normalization

[[ 110 11 13]

[ 516 1389 32]

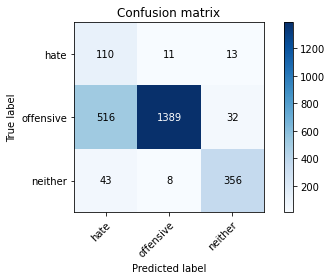
[ 43 8 356]]

Normalized confusion matrix

[[0.82089552 0.08208955 0.09701493]

[0.26639133 0.71708828 0.01652039]

[0.10565111 0.01965602 0.87469287]]





precision recall f1-score support

hate 0.16 0.82 0.27 134

offensive 0.99 0.72 0.83 1937

neither 0.89 0.87 0.88 407

accuracy 0.75 2478

macro avg 0.68 0.80 0.66 2478

weighted avg 0.93 0.75 0.81 2478

average eval\_loss: 0.5604934257365042, accuracy: 0.748587570621469

training 3 epoch...

106.8887550830841 seconds used......

3 training finished! average train loss: 0.25990121596789156

total clf loss: 48.18837746977806 total adv loss: 48.915271155536175

evaluating...

Confusion matrix, without normalization

[[ 102 16 16]

[ 376 1505 56]

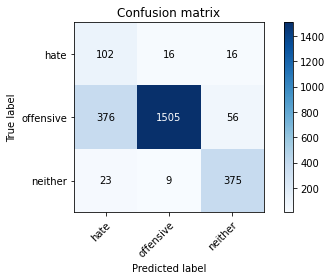
[ 23 9 375]]

Normalized confusion matrix

[[0.76119403 0.11940299 0.11940299]

[0.19411461 0.7769747 0.02891069]

[0.05651106 0.02211302 0.92137592]]





precision recall f1-score support

hate 0.20 0.76 0.32 134

offensive 0.98 0.78 0.87 1937

neither 0.84 0.92 0.88 407

accuracy 0.80 2478

macro avg 0.68 0.82 0.69 2478

weighted avg 0.92 0.80 0.84 2478

average eval\_loss: 0.5174631077435708, accuracy: 0.7998385794995965

training 4 epoch...

107.01079297065735 seconds used......

4 training finished! average train loss: 0.16249902344795622

total clf loss: 31.053617106750607 total adv loss: 49.075780130922794

evaluating...

Confusion matrix, without normalization

[[ 89 30 15]

[ 246 1635 56]

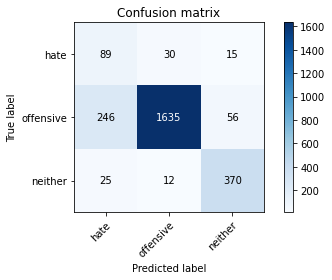
[ 25 12 370]]

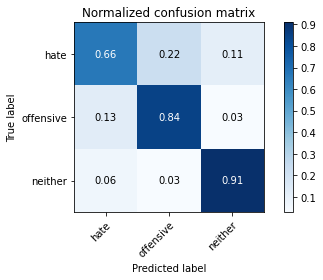
Normalized confusion matrix

[[0.6641791 0.2238806 0.1119403 ]

[0.12700052 0.8440888 0.02891069]

[0.06142506 0.02948403 0.90909091]]





precision recall f1-score support

hate 0.25 0.66 0.36 134

offensive 0.97 0.84 0.90 1937

neither 0.84 0.91 0.87 407

accuracy 0.85 2478

macro avg 0.69 0.81 0.71 2478

weighted avg 0.91 0.85 0.87 2478

average eval\_loss: 0.476361574364766, accuracy: 0.8450363196125908

training 5 epoch...

107.0664415359497 seconds used......

5 training finished! average train loss: 0.07827143292789432

total clf loss: 16.308078072033823 total adv loss: 50.646117605268955

evaluating...

Confusion matrix, without normalization

[[ 102 20 12]

[ 467 1441 29]

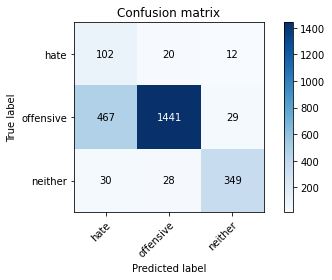
[ 30 28 349]]

Normalized confusion matrix

[[0.76119403 0.14925373 0.08955224]

[0.24109448 0.74393392 0.01497161]

[0.07371007 0.06879607 0.85749386]]





precision recall f1-score support

hate 0.17 0.76 0.28 134

offensive 0.97 0.74 0.84 1937

neither 0.89 0.86 0.88 407

accuracy 0.76 2478

macro avg 0.68 0.79 0.67 2478

weighted avg 0.91 0.76 0.82 2478

average eval\_loss: 0.7308836514430661, accuracy: 0.7635189669087974

Confusion matrix, without normalization

[[ 102 17 5]

[ 524 1379 39]

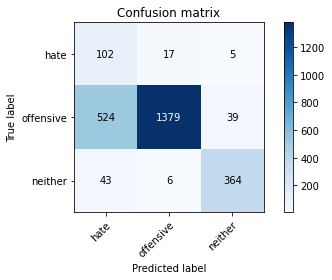
[ 43 6 364]]

Normalized confusion matrix

[[0.82258065 0.13709677 0.04032258]

[0.26982492 0.71009269 0.02008239]

[0.10411622 0.01452785 0.88135593]]





precision recall f1-score support

hate 0.15 0.82 0.26 124

offensive 0.98 0.71 0.82 1942

neither 0.89 0.88 0.89 413

accuracy 0.74 2479

macro avg 0.68 0.80 0.66 2479

weighted avg 0.93 0.74 0.81 2479

average eval\_loss: 0.591758681641471, accuracy: 0.7442517144009682

[39]:

0.7442517144009682