alpha=0.05

1 run...

<tqdm.auto.tqdm object at 0x7f7b4e757a10>

<tqdm.auto.tqdm object at 0x7f7b453fbc90>

training 1 epoch...

391.8352394104004 seconds used......

1 training finished! average train loss: 0.5940741152580382

total clf loss: 270.26212649047375 total adv loss: 141.7458663880825

evaluating...

Confusion matrix, without normalization

[[ 628 210 184]

[ 709 4556 142]

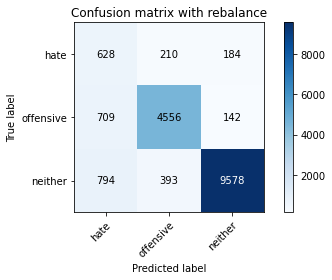
[ 794 393 9578]]

Normalized confusion matrix

[[0.61448141 0.20547945 0.18003914]

[0.13112632 0.84261143 0.02626225]

[0.07375755 0.0365072 0.88973525]]





precision recall f1-score support

hate 0.29 0.61 0.40 1022

offensive 0.88 0.84 0.86 5407

neither 0.97 0.89 0.93 10765

accuracy 0.86 17194

macro avg 0.71 0.78 0.73 17194

weighted avg 0.90 0.86 0.88 17194

average eval\_loss: 0.4010233576907668, accuracy: 0.8585553099918576

training 2 epoch...

390.74973130226135 seconds used......

2 training finished! average train loss: 0.4303062265869188

total clf loss: 196.69807105511427 total adv loss: 121.44825713336468

evaluating...

Confusion matrix, without normalization

[[ 671 189 162]

[ 758 4486 163]

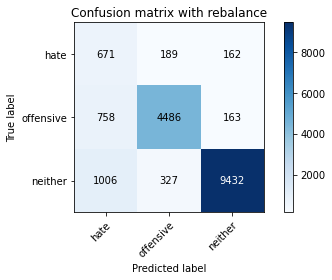
[1006 327 9432]]

Normalized confusion matrix

[[0.65655577 0.18493151 0.15851272]

[0.14018864 0.82966525 0.03014611]

[0.093451 0.03037622 0.87617278]]





precision recall f1-score support

hate 0.28 0.66 0.39 1022

offensive 0.90 0.83 0.86 5407

neither 0.97 0.88 0.92 10765

accuracy 0.85 17194

macro avg 0.71 0.79 0.72 17194

weighted avg 0.90 0.85 0.87 17194

average eval\_loss: 0.42881042193188224, accuracy: 0.8484936605792719

training 3 epoch...

390.7471115589142 seconds used......

3 training finished! average train loss: 0.3168165859667496

total clf loss: 146.54802548140287 total adv loss: 123.96556268632412

evaluating...

Confusion matrix, without normalization

[[ 651 190 181]

[ 819 4401 187]

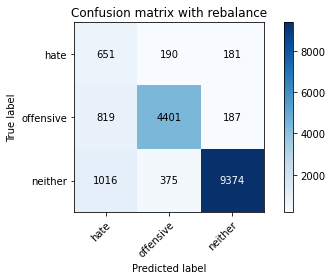
[1016 375 9374]]

Normalized confusion matrix

[[0.6369863 0.18590998 0.17710372]

[0.15147032 0.81394489 0.0345848 ]

[0.09437993 0.03483511 0.87078495]]





precision recall f1-score support

hate 0.26 0.64 0.37 1022

offensive 0.89 0.81 0.85 5407

neither 0.96 0.87 0.91 10765

accuracy 0.84 17194

macro avg 0.70 0.77 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.47394989245939395, accuracy: 0.8390136093986275

training 4 epoch...

390.87255477905273 seconds used......

4 training finished! average train loss: 0.20577785407422156

total clf loss: 97.56899807229638 total adv loss: 128.18816926702857

evaluating...

Confusion matrix, without normalization

[[ 632 196 194]

[ 856 4346 205]

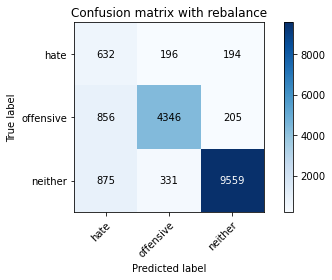
[ 875 331 9559]]

Normalized confusion matrix

[[0.6183953 0.19178082 0.18982387]

[0.1583133 0.80377289 0.03791382]

[0.08128193 0.03074779 0.88797027]]





precision recall f1-score support

hate 0.27 0.62 0.37 1022

offensive 0.89 0.80 0.85 5407

neither 0.96 0.89 0.92 10765

accuracy 0.85 17194

macro avg 0.71 0.77 0.71 17194

weighted avg 0.90 0.85 0.87 17194

average eval\_loss: 0.5047043225273143, accuracy: 0.8454693497731767

training 5 epoch...

390.93914771080017 seconds used......

5 training finished! average train loss: 0.13787660613639224

total clf loss: 67.74473801720887 total adv loss: 133.30802722647786

evaluating...

Confusion matrix, without normalization

[[ 624 211 187]

[ 851 4389 167]

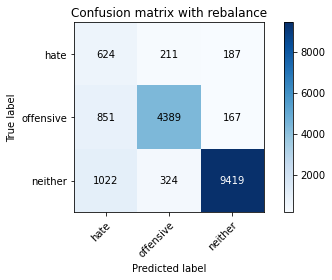
[1022 324 9419]]

Normalized confusion matrix

[[0.61056751 0.20645793 0.18297456]

[0.15738857 0.81172554 0.03088589]

[0.0949373 0.03009754 0.87496516]]





precision recall f1-score support

hate 0.25 0.61 0.35 1022

offensive 0.89 0.81 0.85 5407

neither 0.96 0.87 0.92 10765

accuracy 0.84 17194

macro avg 0.70 0.77 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.5969034542589513, accuracy: 0.8393625683377922

2 run...

training 1 epoch...

391.27211570739746 seconds used......

1 training finished! average train loss: 0.588535609619343

total clf loss: 267.7625135332346 total adv loss: 140.82477332651615

evaluating...

Confusion matrix, without normalization

[[ 655 207 160]

[ 686 4594 127]

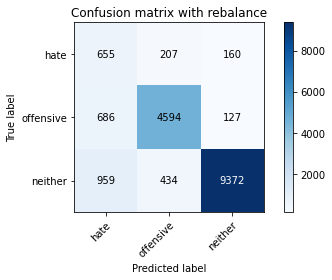
[ 959 434 9372]]

Normalized confusion matrix

[[0.6409002 0.20254403 0.15655577]

[0.12687257 0.84963936 0.02348807]

[0.089085 0.04031584 0.87059916]]





precision recall f1-score support

hate 0.28 0.64 0.39 1022

offensive 0.88 0.85 0.86 5407

neither 0.97 0.87 0.92 10765

accuracy 0.85 17194

macro avg 0.71 0.79 0.73 17194

weighted avg 0.90 0.85 0.87 17194

average eval\_loss: 0.4321718946619089, accuracy: 0.8503547749214843

training 2 epoch...

391.33130049705505 seconds used......

2 training finished! average train loss: 0.4263417563205648

total clf loss: 194.93843717873096 total adv loss: 121.3807807341218

evaluating...

Confusion matrix, without normalization

[[ 691 189 142]

[ 759 4506 142]

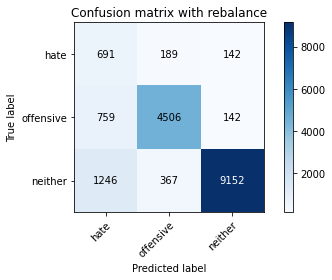
[1246 367 9152]]

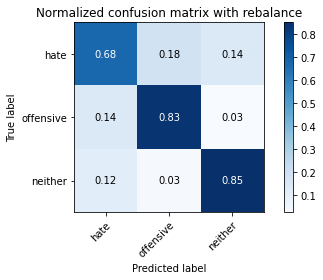
Normalized confusion matrix

[[0.67612524 0.18493151 0.13894325]

[0.14037359 0.83336416 0.02626225]

[0.11574547 0.03409196 0.85016256]]





precision recall f1-score support

hate 0.26 0.68 0.37 1022

offensive 0.89 0.83 0.86 5407

neither 0.97 0.85 0.91 10765

accuracy 0.83 17194

macro avg 0.71 0.79 0.71 17194

weighted avg 0.90 0.83 0.86 17194

average eval\_loss: 0.46482590198690116, accuracy: 0.8345353030126789

training 3 epoch...

391.0501477718353 seconds used......

3 training finished! average train loss: 0.31103769348383487

total clf loss: 144.15584121644497 total adv loss: 127.32285994291306

evaluating...

Confusion matrix, without normalization

[[ 561 284 177]

[ 462 4775 170]

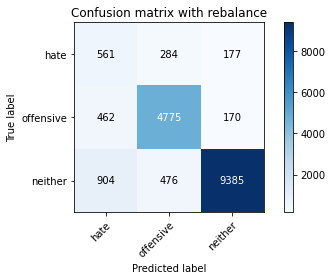
[ 904 476 9385]]

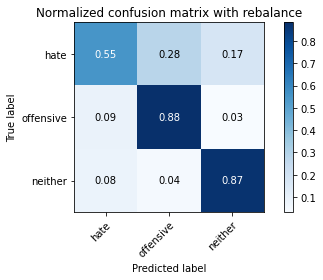
Normalized confusion matrix

[[0.54892368 0.2778865 0.17318982]

[0.08544479 0.88311448 0.03144072]

[0.08397585 0.04421737 0.87180678]]





precision recall f1-score support

hate 0.29 0.55 0.38 1022

offensive 0.86 0.88 0.87 5407

neither 0.96 0.87 0.92 10765

accuracy 0.86 17194

macro avg 0.71 0.77 0.72 17194

weighted avg 0.89 0.86 0.87 17194

average eval\_loss: 0.47486559919969634, accuracy: 0.856170757240898

training 4 epoch...

390.87001967430115 seconds used......

4 training finished! average train loss: 0.20007548490476124

total clf loss: 95.16817565076053 total adv loss: 130.69471663981676

evaluating...

Confusion matrix, without normalization

[[ 698 178 146]

[1131 4128 148]

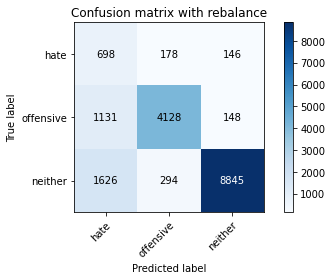
[1626 294 8845]]

Normalized confusion matrix

[[0.68297456 0.1741683 0.14285714]

[0.20917329 0.76345478 0.02737193]

[0.15104505 0.02731073 0.82164422]]





precision recall f1-score support

hate 0.20 0.68 0.31 1022

offensive 0.90 0.76 0.83 5407

neither 0.97 0.82 0.89 10765

accuracy 0.80 17194

macro avg 0.69 0.76 0.68 17194

weighted avg 0.90 0.80 0.83 17194

average eval\_loss: 0.6645661615614974, accuracy: 0.7951029428870536

training 5 epoch...

391.0373351573944 seconds used......

5 training finished! average train loss: 0.13028801501671833

total clf loss: 64.42983462847769 total adv loss: 134.24487794190645

evaluating...

Confusion matrix, without normalization

[[ 578 270 174]

[ 597 4636 174]

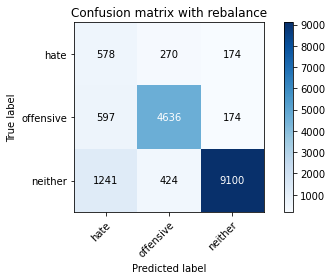
[1241 424 9100]]

Normalized confusion matrix

[[0.56555773 0.26418787 0.1702544 ]

[0.11041243 0.85740706 0.03218051]

[0.115281 0.0393869 0.84533209]]





precision recall f1-score support

hate 0.24 0.57 0.34 1022

offensive 0.87 0.86 0.86 5407

neither 0.96 0.85 0.90 10765

accuracy 0.83 17194

macro avg 0.69 0.76 0.70 17194

weighted avg 0.89 0.83 0.86 17194

average eval\_loss: 0.6710190067500915, accuracy: 0.832499709200884

3 run...

training 1 epoch...

391.0976207256317 seconds used......

1 training finished! average train loss: 0.5857853594133062

total clf loss: 266.53900504112244 total adv loss: 140.72182459011674

evaluating...

Confusion matrix, without normalization

[[ 710 180 132]

[ 829 4462 116]

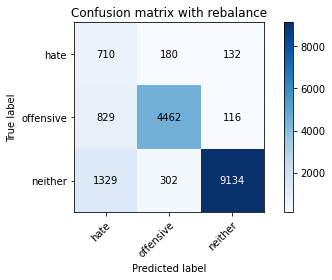
[1329 302 9134]]

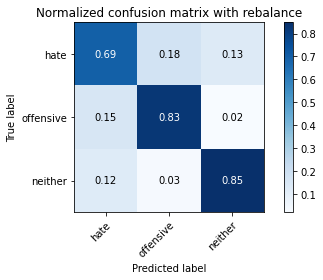
Normalized confusion matrix

[[0.69471624 0.17612524 0.12915851]

[0.15331977 0.82522656 0.02145367]

[0.12345564 0.02805388 0.84849048]]





precision recall f1-score support

hate 0.25 0.69 0.37 1022

offensive 0.90 0.83 0.86 5407

neither 0.97 0.85 0.91 10765

accuracy 0.83 17194

macro avg 0.71 0.79 0.71 17194

weighted avg 0.91 0.83 0.86 17194

average eval\_loss: 0.45830610475914424, accuracy: 0.8320344306153309

training 2 epoch...

391.4947922229767 seconds used......

2 training finished! average train loss: 0.4262119742030484

total clf loss: 194.89034324884415 total adv loss: 121.56877487897873

evaluating...

Confusion matrix, without normalization

[[ 731 163 128]

[ 969 4292 146]

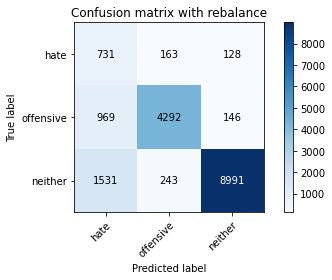
[1531 243 8991]]

Normalized confusion matrix

[[0.71526419 0.15949119 0.12524462]

[0.17921213 0.79378583 0.02700203]

[0.14222016 0.02257315 0.83520669]]





precision recall f1-score support

hate 0.23 0.72 0.34 1022

offensive 0.91 0.79 0.85 5407

neither 0.97 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.78 0.70 17194

weighted avg 0.91 0.82 0.85 17194

average eval\_loss: 0.5114824831832286, accuracy: 0.8150517622426428

training 3 epoch...

391.9814829826355 seconds used......

3 training finished! average train loss: 0.31308003921171074

total clf loss: 145.02736653387547 total adv loss: 126.65819097310305

evaluating...

Confusion matrix, without normalization

[[ 711 143 168]

[1223 3978 206]

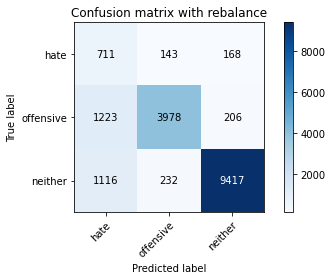
[1116 232 9417]]

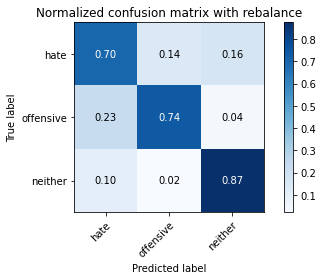
Normalized confusion matrix

[[0.69569472 0.13992172 0.16438356]

[0.22618827 0.73571296 0.03809876]

[0.1036693 0.02155132 0.87477938]]





precision recall f1-score support

hate 0.23 0.70 0.35 1022

offensive 0.91 0.74 0.82 5407

neither 0.96 0.87 0.92 10765

accuracy 0.82 17194

macro avg 0.70 0.77 0.69 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.5054221170846113, accuracy: 0.8204024659765035

training 4 epoch...

391.8848147392273 seconds used......

4 training finished! average train loss: 0.19971214720672395

total clf loss: 95.1698844190687 total adv loss: 133.9480661228299

evaluating...

Confusion matrix, without normalization

[[ 673 226 123]

[ 865 4410 132]

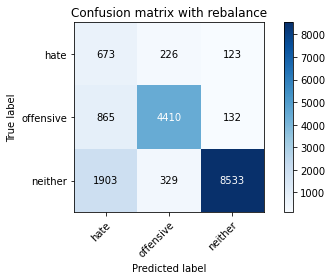
[1903 329 8533]]

Normalized confusion matrix

[[0.65851272 0.22113503 0.12035225]

[0.15997781 0.8156094 0.0244128 ]

[0.17677659 0.03056201 0.7926614 ]]





precision recall f1-score support

hate 0.20 0.66 0.30 1022

offensive 0.89 0.82 0.85 5407

neither 0.97 0.79 0.87 10765

accuracy 0.79 17194

macro avg 0.68 0.76 0.67 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 0.7508447578913251, accuracy: 0.791904152611376

training 5 epoch...

391.3831379413605 seconds used......

5 training finished! average train loss: 0.1263538682370806

total clf loss: 62.79221768025309 total adv loss: 136.34908234328032

evaluating...

Confusion matrix, without normalization

[[ 583 237 202]

[ 757 4412 238]

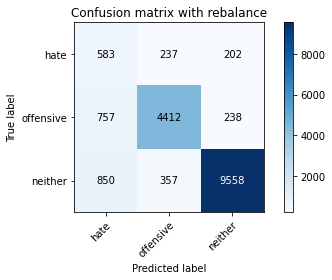
[ 850 357 9558]]

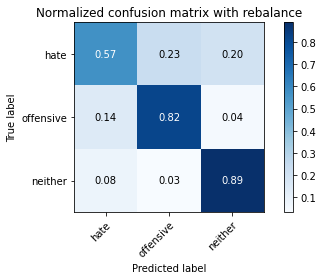
Normalized confusion matrix

[[0.5704501 0.23189824 0.19765166]

[0.1400037 0.81597929 0.04401701]

[0.07895959 0.03316303 0.88787738]]





precision recall f1-score support

hate 0.27 0.57 0.36 1022

offensive 0.88 0.82 0.85 5407

neither 0.96 0.89 0.92 10765

accuracy 0.85 17194

macro avg 0.70 0.76 0.71 17194

weighted avg 0.89 0.85 0.86 17194

average eval\_loss: 0.6000364317450413, accuracy: 0.8463999069442829

4 run...

training 1 epoch...

391.39071106910706 seconds used......

1 training finished! average train loss: 0.5925867687364195

total clf loss: 269.59577637910843 total adv loss: 141.59674472734332

evaluating...

Confusion matrix, without normalization

[[ 713 155 154]

[ 935 4323 149]

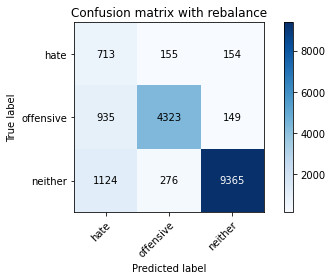
[1124 276 9365]]

Normalized confusion matrix

[[0.69765166 0.15166341 0.15068493]

[0.17292399 0.79951914 0.02755687]

[0.10441245 0.02563864 0.86994891]]





precision recall f1-score support

hate 0.26 0.70 0.38 1022

offensive 0.91 0.80 0.85 5407

neither 0.97 0.87 0.92 10765

accuracy 0.84 17194

macro avg 0.71 0.79 0.71 17194

weighted avg 0.91 0.84 0.86 17194

average eval\_loss: 0.4399004669643419, accuracy: 0.837559613818774

training 2 epoch...

391.37899947166443 seconds used......

2 training finished! average train loss: 0.43094922151441767

total clf loss: 196.98315598070621 total adv loss: 121.45302010327578

evaluating...

Confusion matrix, without normalization

[[ 720 157 145]

[1005 4264 138]

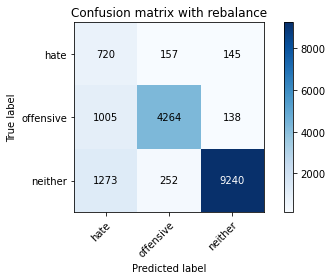
[1273 252 9240]]

Normalized confusion matrix

[[0.70450098 0.15362035 0.14187867]

[0.18587017 0.78860736 0.02552247]

[0.1182536 0.0234092 0.8583372 ]]





precision recall f1-score support

hate 0.24 0.70 0.36 1022

offensive 0.91 0.79 0.85 5407

neither 0.97 0.86 0.91 10765

accuracy 0.83 17194

macro avg 0.71 0.78 0.71 17194

weighted avg 0.91 0.83 0.86 17194

average eval\_loss: 0.4745067310887714, accuracy: 0.8272653251134117

training 3 epoch...

391.4031901359558 seconds used......

3 training finished! average train loss: 0.3117296984061834

total clf loss: 144.39963893964887 total adv loss: 126.06764525175095

evaluating...

Confusion matrix, without normalization

[[ 713 154 155]

[1087 4172 148]

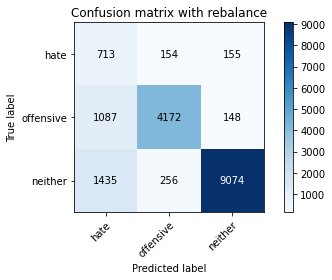
[1435 256 9074]]

Normalized confusion matrix

[[0.69765166 0.15068493 0.15166341]

[0.20103569 0.77159238 0.02737193]

[0.13330237 0.02378077 0.84291686]]





precision recall f1-score support

hate 0.22 0.70 0.33 1022

offensive 0.91 0.77 0.84 5407

neither 0.97 0.84 0.90 10765

accuracy 0.81 17194

macro avg 0.70 0.77 0.69 17194

weighted avg 0.91 0.81 0.85 17194

average eval\_loss: 0.5468817230270699, accuracy: 0.8118529719669653

training 4 epoch...

391.4453754425049 seconds used......

4 training finished! average train loss: 0.20387977527761433

total clf loss: 96.9225309966132 total adv loss: 132.07580871134996

evaluating...

Confusion matrix, without normalization

[[ 668 245 109]

[ 685 4623 99]

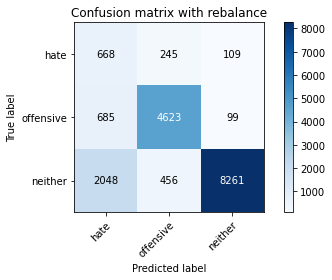
[2048 456 8261]]

Normalized confusion matrix

[[0.65362035 0.23972603 0.10665362]

[0.12668763 0.85500277 0.0183096 ]

[0.19024617 0.0423595 0.76739433]]





precision recall f1-score support

hate 0.20 0.65 0.30 1022

offensive 0.87 0.86 0.86 5407

neither 0.98 0.77 0.86 10765

accuracy 0.79 17194

macro avg 0.68 0.76 0.67 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 0.7272766569845899, accuracy: 0.7881819239269513

training 5 epoch...

391.4394872188568 seconds used......

5 training finished! average train loss: 0.13271901904527544

total clf loss: 65.52105619898066 total adv loss: 134.5306094288826

evaluating...

Confusion matrix, without normalization

[[ 660 229 133]

[ 834 4468 105]

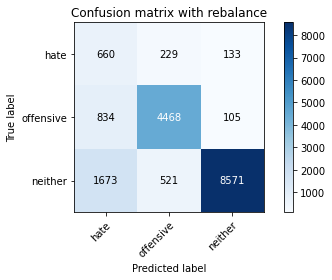
[1673 521 8571]]

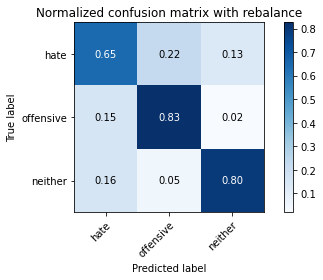
Normalized confusion matrix

[[0.64579256 0.22407045 0.13013699]

[0.1542445 0.82633623 0.01941927]

[0.15541105 0.04839758 0.79619136]]





precision recall f1-score support

hate 0.21 0.65 0.32 1022

offensive 0.86 0.83 0.84 5407

neither 0.97 0.80 0.88 10765

accuracy 0.80 17194

macro avg 0.68 0.76 0.68 17194

weighted avg 0.89 0.80 0.83 17194

average eval\_loss: 0.8091883992256467, accuracy: 0.7967314179364895

5 run...

training 1 epoch...

391.24678659439087 seconds used......

1 training finished! average train loss: 0.5970398259283995

total clf loss: 271.65123049914837 total adv loss: 143.25175654515624

evaluating...

Confusion matrix, without normalization

[[ 633 186 203]

[ 718 4515 174]

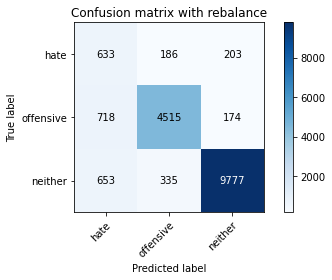
[ 653 335 9777]]

Normalized confusion matrix

[[0.61937378 0.18199609 0.19863014]

[0.13279083 0.83502867 0.03218051]

[0.06065954 0.03111937 0.90822109]]





precision recall f1-score support

hate 0.32 0.62 0.42 1022

offensive 0.90 0.84 0.86 5407

neither 0.96 0.91 0.93 10765

accuracy 0.87 17194

macro avg 0.73 0.79 0.74 17194

weighted avg 0.90 0.87 0.88 17194

average eval\_loss: 0.39168164280438145, accuracy: 0.868035361172502

training 2 epoch...

391.00310134887695 seconds used......

2 training finished! average train loss: 0.43338809095104447

total clf loss: 198.0312587916851 total adv loss: 120.8066960722208

evaluating...

Confusion matrix, without normalization

[[ 747 133 142]

[1133 4152 122]

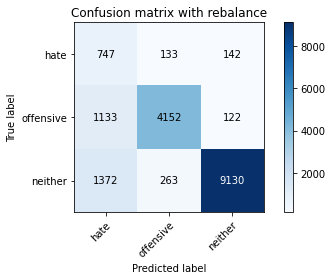
[1372 263 9130]]

Normalized confusion matrix

[[0.73091977 0.13013699 0.13894325]

[0.20954318 0.76789347 0.02256334]

[0.12745007 0.02443103 0.8481189 ]]





precision recall f1-score support

hate 0.23 0.73 0.35 1022

offensive 0.91 0.77 0.83 5407

neither 0.97 0.85 0.91 10765

accuracy 0.82 17194

macro avg 0.70 0.78 0.70 17194

weighted avg 0.91 0.82 0.85 17194

average eval\_loss: 0.49365768509722036, accuracy: 0.8159241595905549

training 3 epoch...

391.0835962295532 seconds used......

3 training finished! average train loss: 0.32330102200197314

total clf loss: 149.47029858082533 total adv loss: 124.95891361869872

evaluating...

Confusion matrix, without normalization

[[ 685 170 167]

[ 982 4263 162]

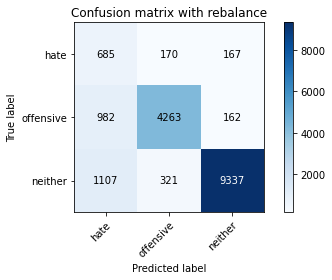
[1107 321 9337]]

Normalized confusion matrix

[[0.6702544 0.16634051 0.16340509]

[0.18161642 0.78842242 0.02996116]

[0.10283326 0.02981886 0.86734789]]





precision recall f1-score support

hate 0.25 0.67 0.36 1022

offensive 0.90 0.79 0.84 5407

neither 0.97 0.87 0.91 10765

accuracy 0.83 17194

macro avg 0.70 0.78 0.70 17194

weighted avg 0.90 0.83 0.86 17194

average eval\_loss: 0.5091678264572523, accuracy: 0.830813074328254

training 4 epoch...

391.13429522514343 seconds used......

4 training finished! average train loss: 0.20241770901113884

total clf loss: 96.23940055817366 total adv loss: 131.3671068660915

evaluating...

Confusion matrix, without normalization

[[ 677 199 146]

[ 927 4321 159]

[1451 296 9018]]

Normalized confusion matrix

[[0.66242661 0.19471624 0.14285714]

[0.17144442 0.79914925 0.02940633]

[0.13478867 0.02749652 0.83771482]]





precision recall f1-score support

hate 0.22 0.66 0.33 1022

offensive 0.90 0.80 0.85 5407

neither 0.97 0.84 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.77 0.69 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.6136576624933717, accuracy: 0.815168081889031

training 5 epoch...

390.9264466762543 seconds used......

5 training finished! average train loss: 0.13844050538072333

total clf loss: 68.01328322198242 total adv loss: 133.68278492614627

evaluating...

Confusion matrix, without normalization

[[ 640 187 195]

[ 937 4240 230]

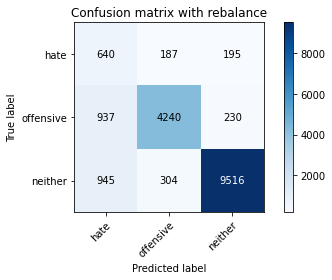
[ 945 304 9516]]

Normalized confusion matrix

[[0.62622309 0.18297456 0.19080235]

[0.17329388 0.78416867 0.04253745]

[0.08778449 0.02823967 0.88397585]]





precision recall f1-score support

hate 0.25 0.63 0.36 1022

offensive 0.90 0.78 0.84 5407

neither 0.96 0.88 0.92 10765

accuracy 0.84 17194

macro avg 0.70 0.76 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.5933425515402801, accuracy: 0.8372688147028033