alpha=0.05

1 run...

training 1 epoch...

393.912353515625 seconds used......

1 training finished! average train loss: 0.5910826010238506

total clf loss: 269.03492975234985 total adv loss: 143.70675076916814

evaluating...

Confusion matrix, without normalization

[[ 698 187 137]

[ 728 4560 119]

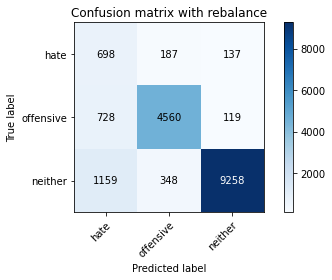
[1159 348 9258]]

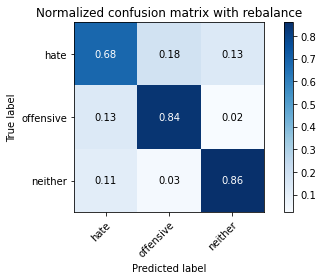
Normalized confusion matrix

[[0.68297456 0.18297456 0.13405088]

[0.13464028 0.84335121 0.02200851]

[0.10766373 0.03232699 0.86000929]]





precision recall f1-score support

hate 0.27 0.68 0.39 1022

offensive 0.89 0.84 0.87 5407

neither 0.97 0.86 0.91 10765

accuracy 0.84 17194

macro avg 0.71 0.80 0.72 17194

weighted avg 0.91 0.84 0.87 17194

average eval\_loss: 0.4290560420342656, accuracy: 0.8442479934860998

training 2 epoch...

393.08515882492065 seconds used......

2 training finished! average train loss: 0.4290536398375007

total clf loss: 196.2549402192235 total adv loss: 123.68356176093221

evaluating...

Confusion matrix, without normalization

[[ 740 168 114]

[ 933 4351 123]

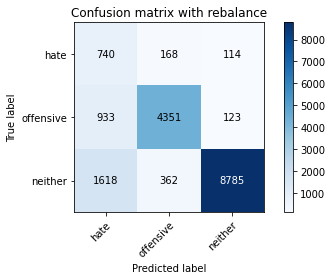
[1618 362 8785]]

Normalized confusion matrix

[[0.72407045 0.16438356 0.11154599]

[0.1725541 0.80469761 0.02274829]

[0.1503019 0.0336275 0.8160706 ]]





precision recall f1-score support

hate 0.22 0.72 0.34 1022

offensive 0.89 0.80 0.85 5407

neither 0.97 0.82 0.89 10765

accuracy 0.81 17194

macro avg 0.70 0.78 0.69 17194

weighted avg 0.90 0.81 0.84 17194

average eval\_loss: 0.5059188132455875, accuracy: 0.8070257066418518

training 3 epoch...

393.4071891307831 seconds used......

3 training finished! average train loss: 0.3092829887219127

total clf loss: 143.44358102977276 total adv loss: 128.6243401169777

evaluating...

Confusion matrix, without normalization

[[ 701 191 130]

[ 830 4449 128]

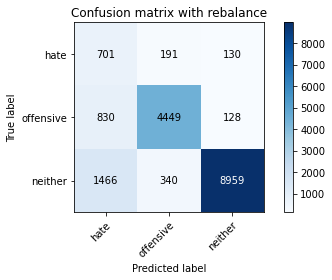
[1466 340 8959]]

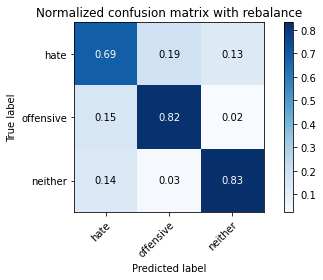
Normalized confusion matrix

[[0.68590998 0.18688845 0.12720157]

[0.15350472 0.82282227 0.02367302]

[0.13618207 0.03158384 0.83223409]]





precision recall f1-score support

hate 0.23 0.69 0.35 1022

offensive 0.89 0.82 0.86 5407

neither 0.97 0.83 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.78 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.5451706462907929, accuracy: 0.8205769454460858

training 4 epoch...

393.5327582359314 seconds used......

4 training finished! average train loss: 0.19757152318904012

total clf loss: 94.1883844230324 total adv loss: 133.28399340063334

evaluating...

Confusion matrix, without normalization

[[ 740 133 149]

[1365 3874 168]

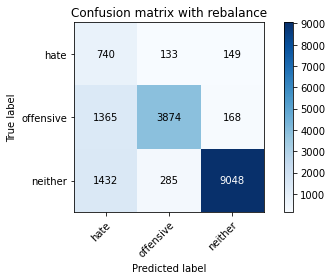
[1432 285 9048]]

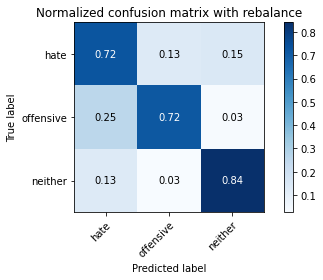
Normalized confusion matrix

[[0.72407045 0.13013699 0.14579256]

[0.25245053 0.71647864 0.03107083]

[0.13302369 0.02647469 0.84050163]]





precision recall f1-score support

hate 0.21 0.72 0.32 1022

offensive 0.90 0.72 0.80 5407

neither 0.97 0.84 0.90 10765

accuracy 0.79 17194

macro avg 0.69 0.76 0.67 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 0.6099787835031748, accuracy: 0.7945795044783064

training 5 epoch...

393.0964925289154 seconds used......

5 training finished! average train loss: 0.13736614214075607

total clf loss: 67.74684559181333 total adv loss: 137.87289094924927

evaluating...

Confusion matrix, without normalization

[[ 684 195 143]

[ 927 4313 167]

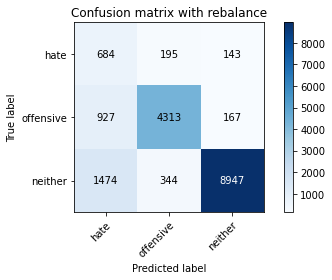
[1474 344 8947]]

Normalized confusion matrix

[[0.66927593 0.19080235 0.13992172]

[0.17144442 0.79766969 0.03088589]

[0.13692522 0.03195541 0.83111937]]





precision recall f1-score support

hate 0.22 0.67 0.33 1022

offensive 0.89 0.80 0.84 5407

neither 0.97 0.83 0.89 10765

accuracy 0.81 17194

macro avg 0.69 0.77 0.69 17194

weighted avg 0.90 0.81 0.84 17194

average eval\_loss: 0.6971800252816878, accuracy: 0.8109805746190532

2 run...

training 1 epoch...

393.0080142021179 seconds used......

1 training finished! average train loss: 0.5985383722704097

total clf loss: 272.1893193721771 total adv loss: 140.73641556501389

evaluating...

Confusion matrix, without normalization

[[ 699 195 128]

[ 739 4564 104]

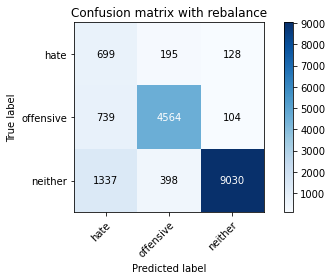
[1337 398 9030]]

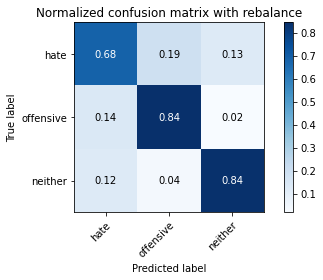
Normalized confusion matrix

[[0.68395303 0.19080235 0.12524462]

[0.13667468 0.84409099 0.01923433]

[0.12419879 0.03697167 0.83882954]]





precision recall f1-score support

hate 0.25 0.68 0.37 1022

offensive 0.89 0.84 0.86 5407

neither 0.97 0.84 0.90 10765

accuracy 0.83 17194

macro avg 0.70 0.79 0.71 17194

weighted avg 0.90 0.83 0.86 17194

average eval\_loss: 0.4667460239557333, accuracy: 0.8312783529138071

training 2 epoch...

392.9639046192169 seconds used......

2 training finished! average train loss: 0.4286168097406962

total clf loss: 195.89840939641 total adv loss: 120.42325323075056

evaluating...

Confusion matrix, without normalization

[[ 709 182 131]

[ 811 4485 111]

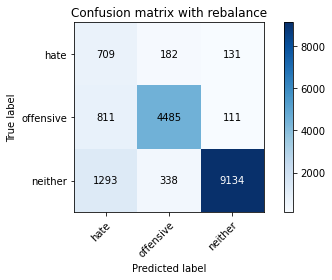
[1293 338 9134]]

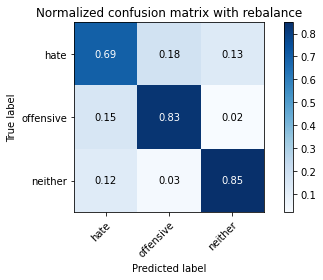
Normalized confusion matrix

[[0.69373777 0.17808219 0.12818004]

[0.14999075 0.8294803 0.02052894]

[0.12011147 0.03139805 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: 0.46966363256580607, accuracy: 0.833313946725602

training 3 epoch...

393.19957184791565 seconds used......

3 training finished! average train loss: 0.31885124610118887

total clf loss: 147.52844938635826 total adv loss: 125.5469424687326

evaluating...

Confusion matrix, without normalization

[[ 708 164 150]

[ 921 4307 179]

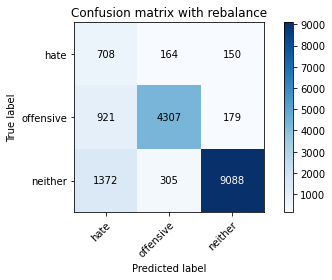
[1372 305 9088]]

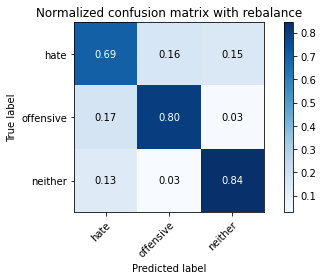
Normalized confusion matrix

[[0.6927593 0.16046967 0.14677104]

[0.17033475 0.79656001 0.03310523]

[0.12745007 0.02833256 0.84421737]]





precision recall f1-score support

hate 0.24 0.69 0.35 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 0.78 0.70 17194

weighted avg 0.90 0.82 0.85 17194

average eval\_loss: 0.5174369187549103, accuracy: 0.820227986506921

training 4 epoch...

393.11209893226624 seconds used......

4 training finished! average train loss: 0.2084827680586423

total clf loss: 98.96498181670904 total adv loss: 132.14230584353209

evaluating...

Confusion matrix, without normalization

[[ 692 221 109]

[ 722 4556 129]

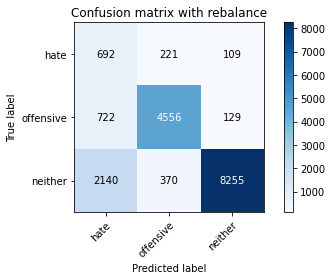
[2140 370 8255]]

Normalized confusion matrix

[[0.67710372 0.21624266 0.10665362]

[0.13353061 0.84261143 0.02385796]

[0.19879238 0.03437065 0.76683697]]





precision recall f1-score support

hate 0.19 0.68 0.30 1022

offensive 0.89 0.84 0.86 5407

neither 0.97 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: 0.7040436152628687, accuracy: 0.7853320925904386

training 5 epoch...

393.39275336265564 seconds used......

5 training finished! average train loss: 0.13121446398660777

total clf loss: 64.88557502999902 total adv loss: 135.15134997293353

evaluating...

Confusion matrix, without normalization

[[ 633 242 147]

[ 643 4611 153]

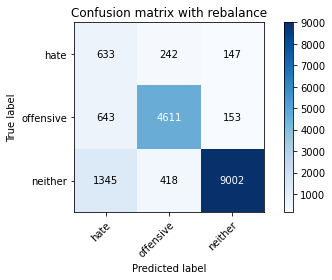
[1345 418 9002]]

Normalized confusion matrix

[[0.61937378 0.23679061 0.14383562]

[0.11891992 0.85278343 0.02829665]

[0.12494194 0.03882954 0.83622852]]





precision recall f1-score support

hate 0.24 0.62 0.35 1022

offensive 0.87 0.85 0.86 5407

neither 0.97 0.84 0.90 10765

accuracy 0.83 17194

macro avg 0.69 0.77 0.70 17194

weighted avg 0.90 0.83 0.85 17194

average eval\_loss: 0.6475423188598524, accuracy: 0.8285448412236827

3 run...

training 1 epoch...

393.2835810184479 seconds used......

1 training finished! average train loss: 0.5984191848448652

total clf loss: 272.31930112838745 total adv loss: 144.39203736186028

evaluating...

Confusion matrix, without normalization

[[ 738 169 115]

[ 907 4399 101]

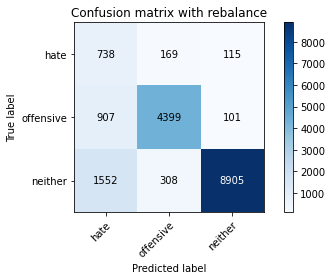
[1552 308 8905]]

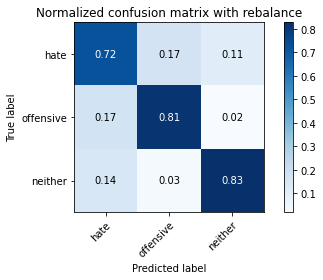
Normalized confusion matrix

[[0.7221135 0.16536204 0.11252446]

[0.16774552 0.813575 0.01867949]

[0.14417092 0.02861124 0.82721784]]





precision recall f1-score support

hate 0.23 0.72 0.35 1022

offensive 0.90 0.81 0.86 5407

neither 0.98 0.83 0.90 10765

accuracy 0.82 17194

macro avg 0.70 0.79 0.70 17194

weighted avg 0.91 0.82 0.85 17194

average eval\_loss: 0.5358114748846653, accuracy: 0.8166802372920786

training 2 epoch...

393.15869426727295 seconds used......

2 training finished! average train loss: 0.43041122030323836

total clf loss: 196.7489661499858 total adv loss: 121.5359159335494

evaluating...

Confusion matrix, without normalization

[[ 599 185 238]

[ 702 4458 247]

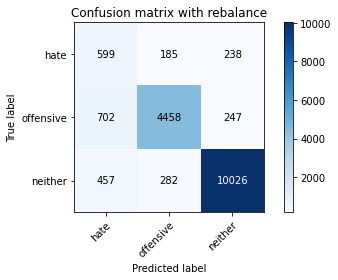
[ 457 282 10026]]

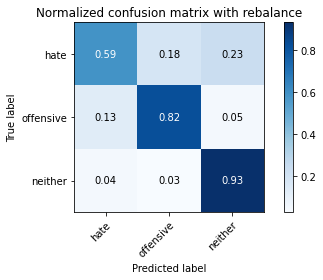
Normalized confusion matrix

[[0.58610568 0.18101761 0.23287671]

[0.1298317 0.82448678 0.04568152]

[0.04245239 0.02619601 0.9313516 ]]





precision recall f1-score support

hate 0.34 0.59 0.43 1022

offensive 0.91 0.82 0.86 5407

neither 0.95 0.93 0.94 10765

accuracy 0.88 17194

macro avg 0.73 0.78 0.75 17194

weighted avg 0.90 0.88 0.89 17194

average eval\_loss: 0.3502251865623822, accuracy: 0.8772246132371757

training 3 epoch...

392.8296980857849 seconds used......

3 training finished! average train loss: 0.3195549219838316

total clf loss: 147.92977569624782 total adv loss: 127.33890577405691

evaluating...

Confusion matrix, without normalization

[[ 724 206 92]

[ 763 4554 90]

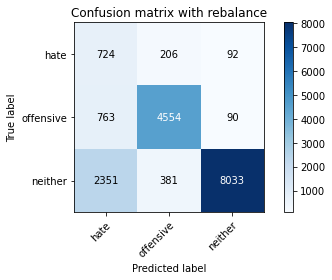
[2351 381 8033]]

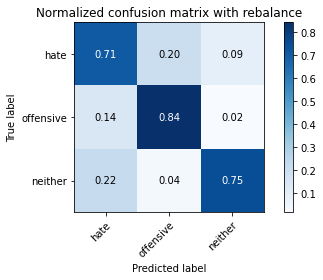
Normalized confusion matrix

[[0.70841487 0.20156556 0.09001957]

[0.14111337 0.84224154 0.01664509]

[0.21839294 0.03539248 0.74621458]]





precision recall f1-score support

hate 0.19 0.71 0.30 1022

offensive 0.89 0.84 0.86 5407

neither 0.98 0.75 0.85 10765

accuracy 0.77 17194

macro avg 0.68 0.77 0.67 17194

weighted avg 0.90 0.77 0.82 17194

average eval\_loss: 0.6995562240221472, accuracy: 0.7741654065371641

training 4 epoch...

393.23305463790894 seconds used......

4 training finished! average train loss: 0.20976533889812554

total clf loss: 99.61226699873805 total adv loss: 133.72443649172783

evaluating...

Confusion matrix, without normalization

[[ 745 148 129]

[1050 4217 140]

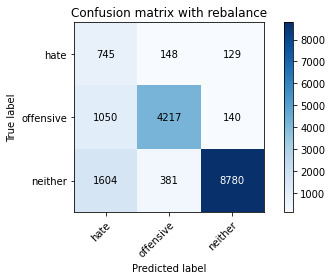
[1604 381 8780]]

Normalized confusion matrix

[[0.72896282 0.14481409 0.12622309]

[0.19419271 0.77991493 0.02589236]

[0.14900139 0.03539248 0.81560613]]





precision recall f1-score support

hate 0.22 0.73 0.34 1022

offensive 0.89 0.78 0.83 5407

neither 0.97 0.82 0.89 10765

accuracy 0.80 17194

macro avg 0.69 0.77 0.68 17194

weighted avg 0.90 0.80 0.84 17194

average eval\_loss: 0.6887806960273274, accuracy: 0.7992322903338374

training 5 epoch...

393.34270763397217 seconds used......

5 training finished! average train loss: 0.13366529070530464

total clf loss: 66.0879394877702 total adv loss: 137.48431219905615

evaluating...

Confusion matrix, without normalization

[[ 710 198 114]

[ 936 4344 127]

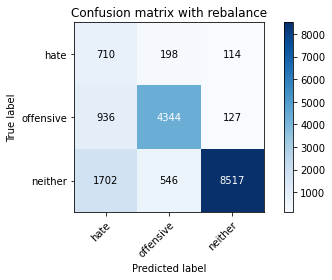
[1702 546 8517]]

Normalized confusion matrix

[[0.69471624 0.19373777 0.11154599]

[0.17310893 0.803403 0.02348807]

[0.15810497 0.05071993 0.7911751 ]]





precision recall f1-score support

hate 0.21 0.69 0.32 1022

offensive 0.85 0.80 0.83 5407

neither 0.97 0.79 0.87 10765

accuracy 0.79 17194

macro avg 0.68 0.76 0.68 17194

weighted avg 0.89 0.79 0.83 17194

average eval\_loss: 0.8774486211751219, accuracy: 0.7892869605676399

4 run...

training 1 epoch...

392.9338788986206 seconds used......

1 training finished! average train loss: 0.5887686193123897

total clf loss: 267.8497539758682 total adv loss: 140.50510859675705

evaluating...

Confusion matrix, without normalization

[[ 693 156 173]

[ 963 4302 142]

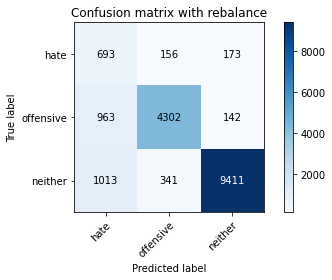
[1013 341 9411]]

Normalized confusion matrix

[[0.67808219 0.15264188 0.16927593]

[0.17810246 0.79563529 0.02626225]

[0.09410125 0.03167673 0.87422202]]





precision recall f1-score support

hate 0.26 0.68 0.38 1022

offensive 0.90 0.80 0.84 5407

neither 0.97 0.87 0.92 10765

accuracy 0.84 17194

macro avg 0.71 0.78 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.4382369611585556, accuracy: 0.8378504129347447

training 2 epoch...

393.00832319259644 seconds used......

2 training finished! average train loss: 0.42693001380017864

total clf loss: 195.23750314116478 total adv loss: 122.15014594048262

evaluating...

Confusion matrix, without normalization

[[ 613 241 168]

[ 577 4701 129]

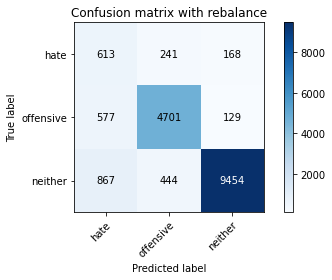
[ 867 444 9454]]

Normalized confusion matrix

[[0.59980431 0.23581213 0.16438356]

[0.10671352 0.86942852 0.02385796]

[0.08053878 0.04124477 0.87821644]]





precision recall f1-score support

hate 0.30 0.60 0.40 1022

offensive 0.87 0.87 0.87 5407

neither 0.97 0.88 0.92 10765

accuracy 0.86 17194

macro avg 0.71 0.78 0.73 17194

weighted avg 0.90 0.86 0.87 17194

average eval\_loss: 0.4118614681151717, accuracy: 0.8589042689310225

training 3 epoch...

393.07383584976196 seconds used......

3 training finished! average train loss: 0.31192151391479284

total clf loss: 144.42213498055935 total adv loss: 124.81807968765497

evaluating...

Confusion matrix, without normalization

[[ 751 163 108]

[1027 4278 102]

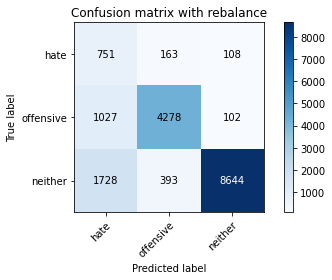
[1728 393 8644]]

Normalized confusion matrix

[[0.73483366 0.15949119 0.10567515]

[0.18993897 0.7911966 0.01886443]

[0.1605202 0.0365072 0.8029726 ]]





precision recall f1-score support

hate 0.21 0.73 0.33 1022

offensive 0.88 0.79 0.84 5407

neither 0.98 0.80 0.88 10765

accuracy 0.80 17194

macro avg 0.69 0.78 0.68 17194

weighted avg 0.90 0.80 0.83 17194

average eval\_loss: 0.5726185557454131, accuracy: 0.7952192625334419

training 4 epoch...

393.3122136592865 seconds used......

4 training finished! average train loss: 0.2088916242904709

total clf loss: 99.03605294041336 total adv loss: 129.94126466661692

evaluating...

Confusion matrix, without normalization

[[ 680 172 170]

[ 919 4268 220]

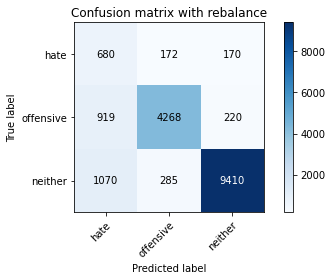
[1070 285 9410]]

Normalized confusion matrix

[[0.66536204 0.16829746 0.16634051]

[0.16996486 0.78934714 0.040688 ]

[0.09939619 0.02647469 0.87412912]]





precision recall f1-score support

hate 0.25 0.67 0.37 1022

offensive 0.90 0.79 0.84 5407

neither 0.96 0.87 0.92 10765

accuracy 0.84 17194

macro avg 0.71 0.78 0.71 17194

weighted avg 0.90 0.84 0.86 17194

average eval\_loss: 0.5409630639879249, accuracy: 0.8350587414214261

training 5 epoch...

393.30498337745667 seconds used......

5 training finished! average train loss: 0.12575560478329592

total clf loss: 62.50637637451291 total adv loss: 135.93286853283644

evaluating...

Confusion matrix, without normalization

[[ 668 183 171]

[ 999 4182 226]

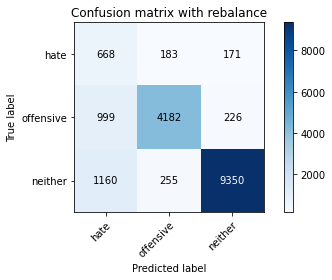
[1160 255 9350]]

Normalized confusion matrix

[[0.65362035 0.17906067 0.16731898]

[0.1847605 0.77344183 0.04179767]

[0.10775662 0.02368788 0.8685555 ]]





precision recall f1-score support

hate 0.24 0.65 0.35 1022

offensive 0.91 0.77 0.83 5407

neither 0.96 0.87 0.91 10765

accuracy 0.83 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.83 0.85 17194

average eval\_loss: 0.639871523693897, accuracy: 0.8258694893567523

5 run...

training 1 epoch...

393.5388352870941 seconds used......

1 training finished! average train loss: 0.590448200299024

total clf loss: 268.62189818918705 total adv loss: 141.06689766794443

evaluating...

Confusion matrix, without normalization

[[ 614 206 202]

[ 606 4645 156]

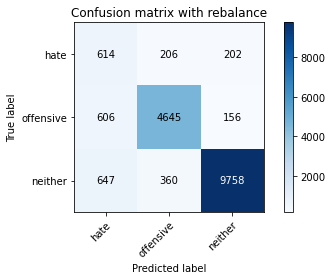
[ 647 360 9758]]

Normalized confusion matrix

[[0.60078278 0.20156556 0.19765166]

[0.11207694 0.85907157 0.02885149]

[0.06010218 0.03344171 0.90645611]]





precision recall f1-score support

hate 0.33 0.60 0.43 1022

offensive 0.89 0.86 0.87 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 0.89 17194

average eval\_loss: 0.39545522119936555, accuracy: 0.8733860649063627

training 2 epoch...

393.4244842529297 seconds used......

2 training finished! average train loss: 0.42368801607504386

total clf loss: 193.6865480169654 total adv loss: 119.85513857007027

evaluating...

Confusion matrix, without normalization

[[ 716 174 132]

[ 912 4377 118]

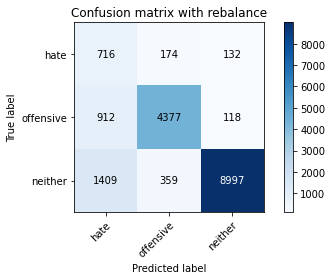
[1409 359 8997]]

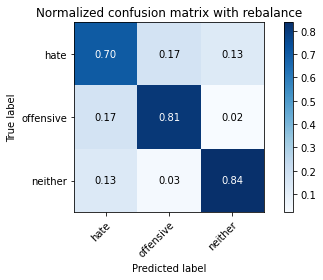
Normalized confusion matrix

[[0.70058708 0.1702544 0.12915851]

[0.16867024 0.8095062 0.02182356]

[0.13088713 0.03334882 0.83576405]]





precision recall f1-score support

hate 0.24 0.70 0.35 1022

offensive 0.89 0.81 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 0.90 0.82 0.85 17194

average eval\_loss: 0.48731640350333477, accuracy: 0.8194719088053972

training 3 epoch...

393.57328748703003 seconds used......

3 training finished! average train loss: 0.2994136274886858

total clf loss: 138.835799459368 total adv loss: 123.91124515235424

evaluating...

Confusion matrix, without normalization

[[ 664 192 166]

[ 801 4484 122]

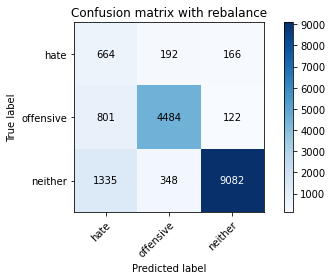
[1335 348 9082]]

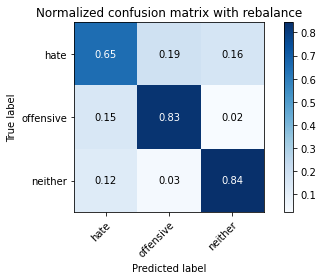
Normalized confusion matrix

[[0.64970646 0.18786693 0.16242661]

[0.1481413 0.82929536 0.02256334]

[0.12401301 0.03232699 0.84366001]]





precision recall f1-score support

hate 0.24 0.65 0.35 1022

offensive 0.89 0.83 0.86 5407

neither 0.97 0.84 0.90 10765

accuracy 0.83 17194

macro avg 0.70 0.77 0.70 17194

weighted avg 0.90 0.83 0.86 17194

average eval\_loss: 0.520413372466384, accuracy: 0.8276142840525765

training 4 epoch...

393.3239905834198 seconds used......

4 training finished! average train loss: 0.1887535327755369

total clf loss: 90.09981841780245 total adv loss: 129.6400657221675

evaluating...

Confusion matrix, without normalization

[[ 734 160 128]

[1042 4218 147]

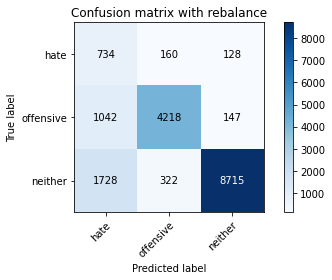
[1728 322 8715]]

Normalized confusion matrix

[[0.71819961 0.15655577 0.12524462]

[0.19271315 0.78009987 0.02718698]

[0.1605202 0.02991175 0.80956804]]





precision recall f1-score support

hate 0.21 0.72 0.32 1022

offensive 0.90 0.78 0.83 5407

neither 0.97 0.81 0.88 10765

accuracy 0.79 17194

macro avg 0.69 0.77 0.68 17194

weighted avg 0.90 0.79 0.83 17194

average eval\_loss: 0.71639371397147, accuracy: 0.7948703035942771

training 5 epoch...

393.3819453716278 seconds used......

5 training finished! average train loss: 0.12173347025016495

total clf loss: 60.65436325687915 total adv loss: 134.52871853485703

evaluating...

Confusion matrix, without normalization

[[ 610 246 166]

[ 680 4555 172]

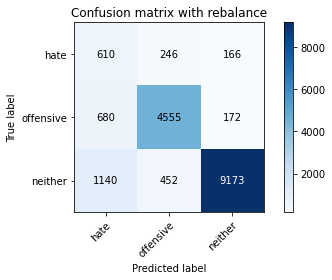
[1140 452 9173]]

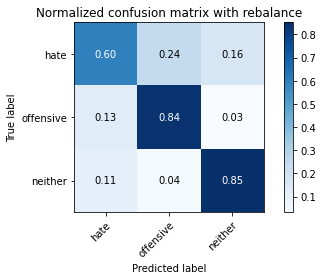
Normalized confusion matrix

[[0.59686888 0.2407045 0.16242661]

[0.1257629 0.84242648 0.03181062]

[0.10589875 0.04198792 0.85211333]]





precision recall f1-score support

hate 0.25 0.60 0.35 1022

offensive 0.87 0.84 0.85 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: 0.6389776974050112, accuracy: 0.8338955449575434