hate as offensive

alpha=0.05

input: last layer

1 run...

Downloading: 100%

411/411 [00:00<00:00, 14.5kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 23.3MB/s]

training 1 epoch...

49.58000731468201 seconds used......

1 training finished! average train loss: 0.654838579004271

total clf loss: 114.85966019332409 total adv loss: 57.6452586427331

evaluating...

Confusion matrix, without normalization

[[ 257 25 12]

[1536 2221 64]

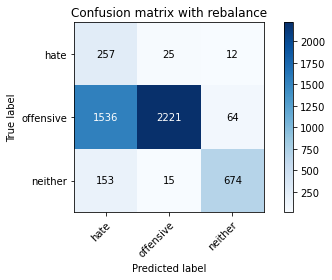
[ 153 15 674]]

Normalized confusion matrix

[[0.87414966 0.08503401 0.04081633]

[0.40198901 0.58126145 0.01674954]

[0.18171021 0.01781473 0.80047506]]





precision recall f1-score support

hate 0.13 0.87 0.23 294

offensive 0.98 0.58 0.73 3821

neither 0.90 0.80 0.85 842

accuracy 0.64 4957

macro avg 0.67 0.75 0.60 4957

weighted avg 0.92 0.64 0.72 4957

average eval\_loss: 0.7911994327460566, accuracy: 0.6358684688319548

training 2 epoch...

49.327197790145874 seconds used......

2 training finished! average train loss: 0.389756515622139

total clf loss: 67.97876465320587 total adv loss: 26.608010333031416

evaluating...

Confusion matrix, without normalization

[[ 229 44 21]

[ 750 2933 138]

[ 46 15 781]]

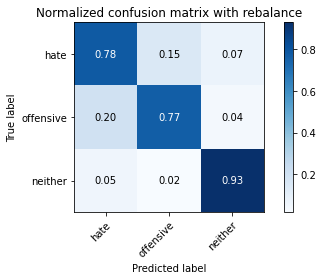
Normalized confusion matrix

[[0.77891156 0.14965986 0.07142857]

[0.1962837 0.7676001 0.0361162 ]

[0.05463183 0.01781473 0.92755344]]





precision recall f1-score support

hate 0.22 0.78 0.35 294

offensive 0.98 0.77 0.86 3821

neither 0.83 0.93 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.80 0.83 4957

average eval\_loss: 0.4827515036318331, accuracy: 0.7954407908008876

training 3 epoch...

49.38491344451904 seconds used......

3 training finished! average train loss: 0.26793055048985787

total clf loss: 46.54096095263958 total adv loss: 14.49673150468152

evaluating...

Confusion matrix, without normalization

[[ 214 54 26]

[ 622 3028 171]

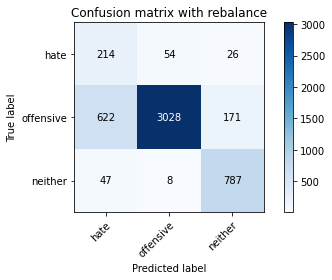
[ 47 8 787]]

Normalized confusion matrix

[[0.72789116 0.18367347 0.08843537]

[0.16278461 0.79246271 0.04475268]

[0.05581948 0.00950119 0.93467933]]





precision recall f1-score support

hate 0.24 0.73 0.36 294

offensive 0.98 0.79 0.88 3821

neither 0.80 0.93 0.86 842

accuracy 0.81 4957

macro avg 0.67 0.82 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.5138891376554966, accuracy: 0.8127899939479524

training 4 epoch...

49.52403521537781 seconds used......

4 training finished! average train loss: 0.17385147230928405

total clf loss: 30.096534913405776 total adv loss: 7.3586638532578945

evaluating...

Confusion matrix, without normalization

[[ 199 80 15]

[ 524 3186 111]

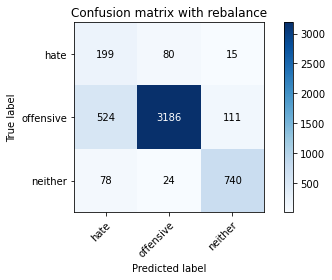
[ 78 24 740]]

Normalized confusion matrix

[[0.67687075 0.27210884 0.05102041]

[0.13713688 0.83381314 0.02904999]

[0.09263658 0.02850356 0.87885986]]





precision recall f1-score support

hate 0.25 0.68 0.36 294

offensive 0.97 0.83 0.90 3821

neither 0.85 0.88 0.87 842

accuracy 0.83 4957

macro avg 0.69 0.80 0.71 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.474959924103572, accuracy: 0.8321565462981642

training 5 epoch...

49.50388956069946 seconds used......

5 training finished! average train loss: 0.09543537587058126

total clf loss: 16.498415769077837 total adv loss: 3.579329052969115

evaluating...

Confusion matrix, without normalization

[[ 194 73 27]

[ 556 3120 145]

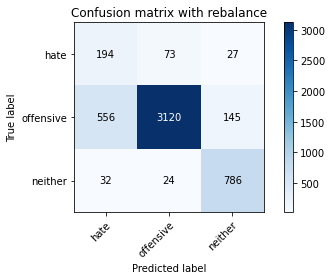
[ 32 24 786]]

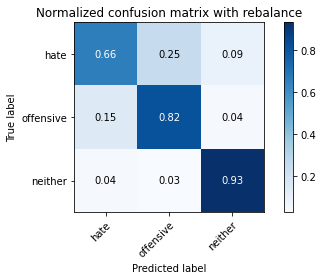
Normalized confusion matrix

[[0.65986395 0.24829932 0.09183673]

[0.14551165 0.81654017 0.03794818]

[0.03800475 0.02850356 0.93349169]]





precision recall f1-score support

hate 0.25 0.66 0.36 294

offensive 0.97 0.82 0.89 3821

neither 0.82 0.93 0.87 842

accuracy 0.83 4957

macro avg 0.68 0.80 0.71 4957

weighted avg 0.90 0.83 0.85 4957

average eval\_loss: 0.5807200958877201, accuracy: 0.8271131732902965

training 6 epoch...

49.375930309295654 seconds used......

6 training finished! average train loss: 0.0449423863408183

total clf loss: 7.792692891205661 total adv loss: 2.150896782928612

evaluating...

Confusion matrix, without normalization

[[ 216 60 18]

[ 643 3051 127]

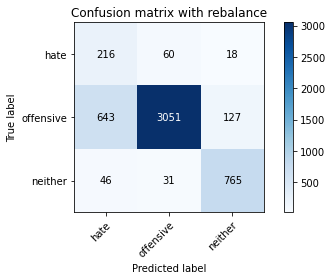
[ 46 31 765]]

Normalized confusion matrix

[[0.73469388 0.20408163 0.06122449]

[0.16828055 0.79848207 0.03323737]

[0.05463183 0.0368171 0.90855107]]





precision recall f1-score support

hate 0.24 0.73 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.84 0.91 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.81 0.85 4957

average eval\_loss: 0.751746162404156, accuracy: 0.8133951987088965

training 7 epoch...

49.428297996520996 seconds used......

7 training finished! average train loss: 0.03040703973182809

total clf loss: 5.289755240781233 total adv loss: 1.8030288748486782

evaluating...

Confusion matrix, without normalization

[[ 218 51 25]

[ 801 2872 148]

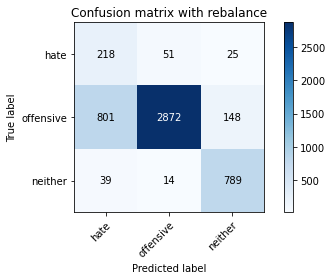
[ 39 14 789]]

Normalized confusion matrix

[[0.7414966 0.17346939 0.08503401]

[0.20963099 0.7516357 0.03873332]

[0.04631829 0.01662708 0.93705463]]





precision recall f1-score support

hate 0.21 0.74 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.82 0.94 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.9790758894185626, accuracy: 0.7825297559007465

training 8 epoch...

49.38671827316284 seconds used......

8 training finished! average train loss: 0.02199925715461161

total clf loss: 3.836230487882858 total adv loss: 1.487150501954602

evaluating...

Confusion matrix, without normalization

[[ 218 60 16]

[ 734 2962 125]

[ 47 27 768]]

Normalized confusion matrix

[[0.7414966 0.20408163 0.05442177]

[0.19209631 0.77518974 0.03271395]

[0.05581948 0.03206651 0.91211401]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.84 0.91 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.80 0.83 4957

average eval\_loss: 0.9508754986060423, accuracy: 0.7964494654024612

training 9 epoch...

49.3772075176239 seconds used......

9 training finished! average train loss: 0.013443030279445631

total clf loss: 2.314671358995838 total adv loss: 0.3182637675017759

evaluating...

Confusion matrix, without normalization

[[ 212 60 22]

[ 711 2962 148]

[ 38 20 784]]

Normalized confusion matrix

[[0.72108844 0.20408163 0.07482993]

[0.18607694 0.77518974 0.03873332]

[0.04513064 0.02375297 0.93111639]]





precision recall f1-score support

hate 0.22 0.72 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.82 0.93 0.87 842

accuracy 0.80 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 1.0511982386787198, accuracy: 0.7984668146056082

training 10 epoch...

49.58288526535034 seconds used......

10 training finished! average train loss: 0.01865227343698375

total clf loss: 3.1974314939870965 total adv loss: 0.1578539557322074

evaluating...

Confusion matrix, without normalization

[[ 217 59 18]

[ 737 2965 119]

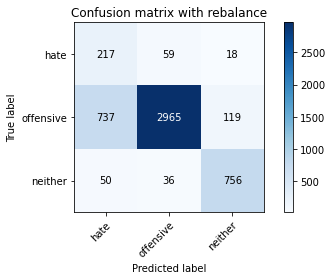
[ 50 36 756]]

Normalized confusion matrix

[[0.73809524 0.20068027 0.06122449]

[0.19288144 0.77597488 0.03114368]

[0.05938242 0.04275534 0.89786223]]





precision recall f1-score support

hate 0.22 0.74 0.33 294

offensive 0.97 0.78 0.86 3821

neither 0.85 0.90 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 1.039998809987728, accuracy: 0.7944321161993141

2 run...

training 1 epoch...

49.45879364013672 seconds used......

1 training finished! average train loss: 0.657312250799603

total clf loss: 115.22110691666603 total adv loss: 56.41424855962396

evaluating...

Confusion matrix, without normalization

[[ 230 38 26]

[ 844 2847 130]

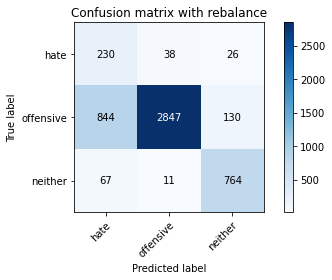
[ 67 11 764]]

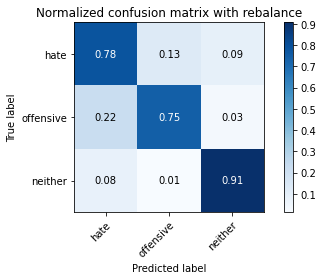
Normalized confusion matrix

[[0.78231293 0.1292517 0.08843537]

[0.22088459 0.74509291 0.03402251]

[0.07957245 0.01306413 0.90736342]]





precision recall f1-score support

hate 0.20 0.78 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.83 0.91 0.87 842

accuracy 0.77 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 0.5460158578690982, accuracy: 0.7748638289287876

training 2 epoch...

49.40034818649292 seconds used......

2 training finished! average train loss: 0.3918449045186154

total clf loss: 68.3989368379116 total adv loss: 27.86916401796043

evaluating...

Confusion matrix, without normalization

[[ 215 51 28]

[ 632 3044 145]

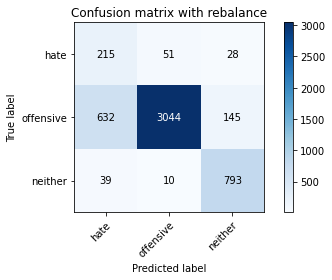
[ 39 10 793]]

Normalized confusion matrix

[[0.73129252 0.17346939 0.0952381 ]

[0.16540173 0.79665009 0.03794818]

[0.04631829 0.01187648 0.94180523]]





precision recall f1-score support

hate 0.24 0.73 0.36 294

offensive 0.98 0.80 0.88 3821

neither 0.82 0.94 0.88 842

accuracy 0.82 4957

macro avg 0.68 0.82 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.451831110753119, accuracy: 0.8174298971151907

training 3 epoch...

49.32570219039917 seconds used......

3 training finished! average train loss: 0.268162022491819

total clf loss: 46.66051021590829 total adv loss: 16.096085745142773

evaluating...

Confusion matrix, without normalization

[[ 202 62 30]

[ 556 3072 193]

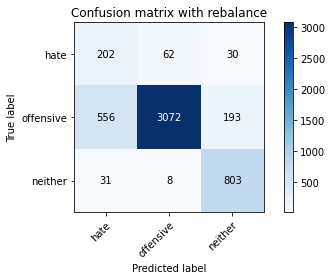
[ 31 8 803]]

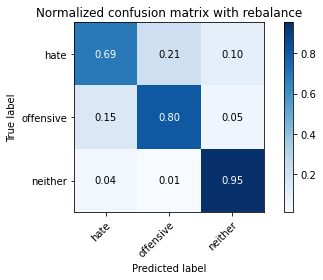
Normalized confusion matrix

[[0.68707483 0.21088435 0.10204082]

[0.14551165 0.80397802 0.05051034]

[0.0368171 0.00950119 0.95368171]]





precision recall f1-score support

hate 0.26 0.69 0.37 294

offensive 0.98 0.80 0.88 3821

neither 0.78 0.95 0.86 842

accuracy 0.82 4957

macro avg 0.67 0.81 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.46522411109218675, accuracy: 0.8224732701230583

training 4 epoch...

49.361449241638184 seconds used......

4 training finished! average train loss: 0.15214965443469977

total clf loss: 26.360317127779126 total adv loss: 6.8545245920540765

evaluating...

Confusion matrix, without normalization

[[ 225 54 15]

[ 817 2906 98]

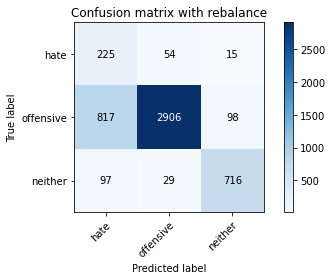
[ 97 29 716]]

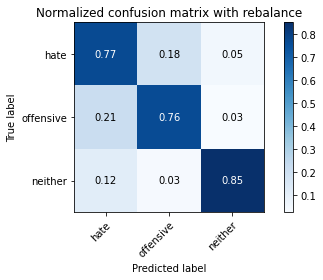
Normalized confusion matrix

[[0.76530612 0.18367347 0.05102041]

[0.21381837 0.76053389 0.02564774]

[0.1152019 0.03444181 0.85035629]]





precision recall f1-score support

hate 0.20 0.77 0.31 294

offensive 0.97 0.76 0.85 3821

neither 0.86 0.85 0.86 842

accuracy 0.78 4957

macro avg 0.68 0.79 0.67 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.6325248306647183, accuracy: 0.7760742384506758

training 5 epoch...

49.38177561759949 seconds used......

5 training finished! average train loss: 0.08642421828784397

total clf loss: 14.920516277197748 total adv loss: 2.83949845895404

evaluating...

Confusion matrix, without normalization

[[ 208 72 14]

[ 558 3172 91]

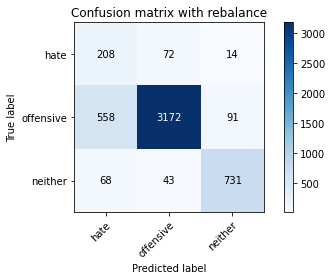
[ 68 43 731]]

Normalized confusion matrix

[[0.70748299 0.24489796 0.04761905]

[0.14603507 0.83014918 0.02381576]

[0.0807601 0.05106888 0.86817102]]





precision recall f1-score support

hate 0.25 0.71 0.37 294

offensive 0.97 0.83 0.89 3821

neither 0.87 0.87 0.87 842

accuracy 0.83 4957

macro avg 0.70 0.80 0.71 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.6054572149991028, accuracy: 0.8293322574137583

training 6 epoch...

49.320380449295044 seconds used......

6 training finished! average train loss: 0.05160060200883805

total clf loss: 8.986114224418998 total adv loss: 3.2482264040445443

evaluating...

Confusion matrix, without normalization

[[ 221 54 19]

[ 784 2946 91]

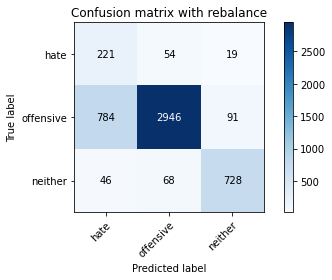
[ 46 68 728]]

Normalized confusion matrix

[[0.75170068 0.18367347 0.06462585]

[0.20518189 0.77100236 0.02381576]

[0.05463183 0.0807601 0.86460808]]





precision recall f1-score support

hate 0.21 0.75 0.33 294

offensive 0.96 0.77 0.86 3821

neither 0.87 0.86 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.8026752079641747, accuracy: 0.7857575146257817

training 7 epoch...

49.3191032409668 seconds used......

7 training finished! average train loss: 0.022740948340060376

total clf loss: 3.9614411715301685 total adv loss: 1.4547803718305659

evaluating...

Confusion matrix, without normalization

[[ 226 52 16]

[ 812 2899 110]

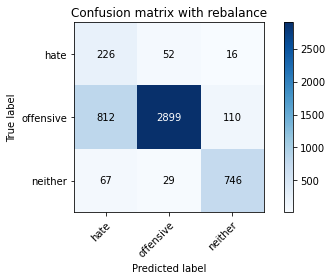
[ 67 29 746]]

Normalized confusion matrix

[[0.76870748 0.17687075 0.05442177]

[0.21250981 0.75870191 0.02878828]

[0.07957245 0.03444181 0.88598575]]





precision recall f1-score support

hate 0.20 0.77 0.32 294

offensive 0.97 0.76 0.85 3821

neither 0.86 0.89 0.87 842

accuracy 0.78 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 1.0258174459900586, accuracy: 0.7809158765382288

training 8 epoch...

49.36000919342041 seconds used......

8 training finished! average train loss: 0.020915960017196856

total clf loss: 3.6202737657004036 total adv loss: 0.8728923179805861

evaluating...

Confusion matrix, without normalization

[[ 224 53 17]

[ 672 3028 121]

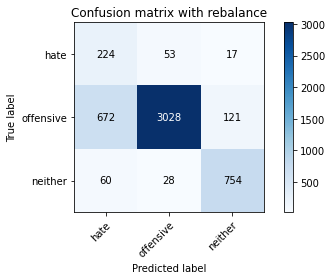
[ 60 28 754]]

Normalized confusion matrix

[[0.76190476 0.18027211 0.05782313]

[0.17587019 0.79246271 0.0316671 ]

[0.07125891 0.03325416 0.89548694]]





precision recall f1-score support

hate 0.23 0.76 0.36 294

offensive 0.97 0.79 0.87 3821

neither 0.85 0.90 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.9190786312138962, accuracy: 0.8081500907807141

training 9 epoch...

49.46071434020996 seconds used......

9 training finished! average train loss: 0.017085381978932137

total clf loss: 3.0328086482477374 total adv loss: 2.2241662931992323

evaluating...

Confusion matrix, without normalization

[[ 213 63 18]

[ 600 3123 98]

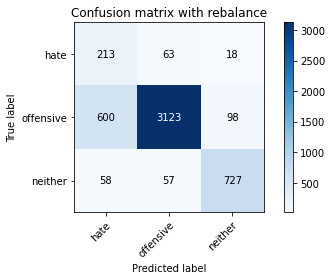
[ 58 57 727]]

Normalized confusion matrix

[[0.7244898 0.21428571 0.06122449]

[0.15702696 0.81732531 0.02564774]

[0.06888361 0.06769596 0.86342043]]





precision recall f1-score support

hate 0.24 0.72 0.37 294

offensive 0.96 0.82 0.88 3821

neither 0.86 0.86 0.86 842

accuracy 0.82 4957

macro avg 0.69 0.80 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.8872432185284372, accuracy: 0.8196489812386524

training 10 epoch...

49.34267258644104 seconds used......

10 training finished! average train loss: 0.014163884755846233

total clf loss: 2.4543182500783587 total adv loss: 0.645878883144178

evaluating...

Confusion matrix, without normalization

[[ 202 70 22]

[ 630 3090 101]

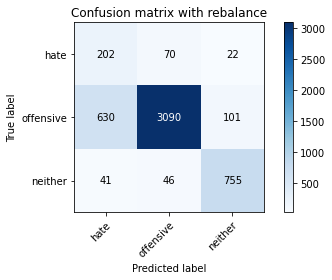
[ 41 46 755]]

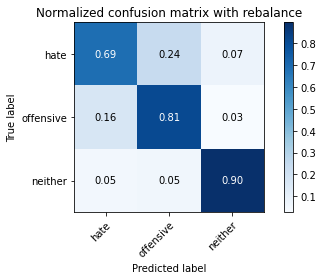
Normalized confusion matrix

[[0.68707483 0.23809524 0.07482993]

[0.1648783 0.80868882 0.02643287]

[0.04869359 0.05463183 0.89667458]]





precision recall f1-score support

hate 0.23 0.69 0.35 294

offensive 0.96 0.81 0.88 3821

neither 0.86 0.90 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.80 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.9072781273036173, accuracy: 0.8164212225136172

3 run...

training 1 epoch...

49.452704429626465 seconds used......

1 training finished! average train loss: 0.6763842827395389

total clf loss: 118.56076012551785 total adv loss: 57.9809526540339

evaluating...

Confusion matrix, without normalization

[[ 231 37 26]

[ 892 2820 109]

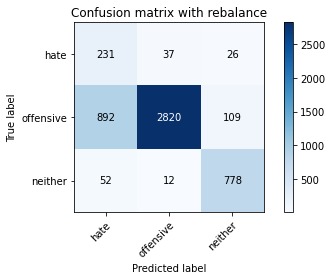
[ 52 12 778]]

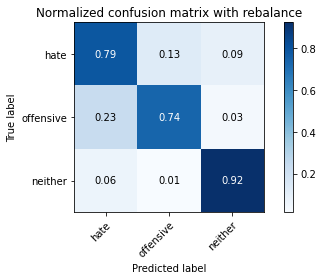
Normalized confusion matrix

[[0.78571429 0.12585034 0.08843537]

[0.23344674 0.73802669 0.02852656]

[0.06175772 0.01425178 0.9239905 ]]





precision recall f1-score support

hate 0.20 0.79 0.31 294

offensive 0.98 0.74 0.84 3821

neither 0.85 0.92 0.89 842

accuracy 0.77 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 0.5741890408098698, accuracy: 0.7724430098850111

training 2 epoch...

49.38859963417053 seconds used......

2 training finished! average train loss: 0.38943725689287073

total clf loss: 67.95193448662758 total adv loss: 27.163270908407867

evaluating...

Confusion matrix, without normalization

[[ 233 43 18]

[ 752 2964 105]

[ 44 18 780]]

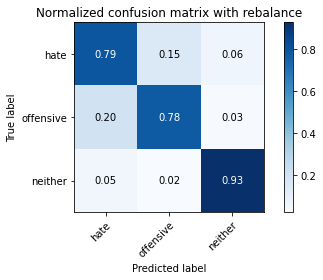
Normalized confusion matrix

[[0.79251701 0.1462585 0.06122449]

[0.19680712 0.77571316 0.02747972]

[0.05225653 0.02137767 0.9263658 ]]





precision recall f1-score support

hate 0.23 0.79 0.35 294

offensive 0.98 0.78 0.87 3821

neither 0.86 0.93 0.89 842

accuracy 0.80 4957

macro avg 0.69 0.83 0.70 4957

weighted avg 0.92 0.80 0.84 4957

average eval\_loss: 0.4885691492367656, accuracy: 0.8022997780915877

training 3 epoch...

49.36817526817322 seconds used......

3 training finished! average train loss: 0.2670600215593974

total clf loss: 46.32143725827336 total adv loss: 13.083474647370167

evaluating...

Confusion matrix, without normalization

[[ 251 30 13]

[1105 2638 78]

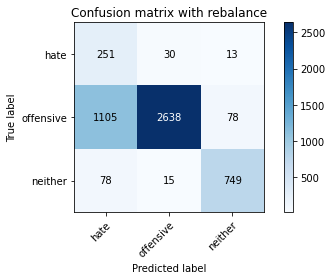
[ 78 15 749]]

Normalized confusion matrix

[[0.8537415 0.10204082 0.04421769]

[0.28919131 0.69039518 0.0204135 ]

[0.09263658 0.01781473 0.88954869]]





precision recall f1-score support

hate 0.18 0.85 0.29 294

offensive 0.98 0.69 0.81 3821

neither 0.89 0.89 0.89 842

accuracy 0.73 4957

macro avg 0.68 0.81 0.66 4957

weighted avg 0.92 0.73 0.79 4957

average eval\_loss: 0.7014281617838048, accuracy: 0.7339116401049022

training 4 epoch...

49.38482642173767 seconds used......

4 training finished! average train loss: 0.15680508146726946

total clf loss: 27.138143429532647 total adv loss: 6.489491540240124

evaluating...

Confusion matrix, without normalization

[[ 201 62 31]

[ 530 3138 153]

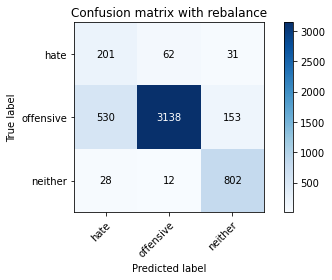
[ 28 12 802]]

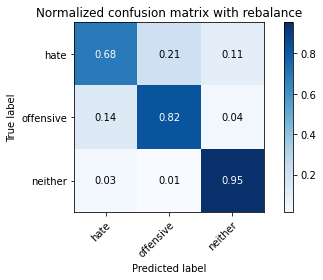
Normalized confusion matrix

[[0.68367347 0.21088435 0.10544218]

[0.13870714 0.82125098 0.04004187]

[0.03325416 0.01425178 0.95249406]]





precision recall f1-score support

hate 0.26 0.68 0.38 294

offensive 0.98 0.82 0.89 3821

neither 0.81 0.95 0.88 842

accuracy 0.84 4957

macro avg 0.69 0.82 0.72 4957

weighted avg 0.91 0.84 0.86 4957

average eval\_loss: 0.5377456161809424, accuracy: 0.8353843050231995

training 5 epoch...

49.419774770736694 seconds used......

5 training finished! average train loss: 0.07780598478707654

total clf loss: 13.454064008779824 total adv loss: 2.9848111591127235

evaluating...

Confusion matrix, without normalization

[[ 244 39 11]

[1127 2595 99]

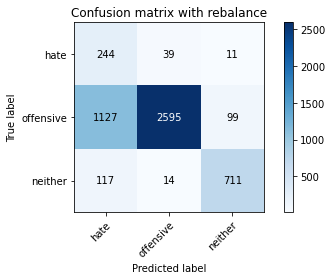
[ 117 14 711]]

Normalized confusion matrix

[[0.82993197 0.13265306 0.03741497]

[0.29494897 0.67914159 0.02590945]

[0.13895487 0.01662708 0.84441805]]





precision recall f1-score support

hate 0.16 0.83 0.27 294

offensive 0.98 0.68 0.80 3821

neither 0.87 0.84 0.86 842

accuracy 0.72 4957

macro avg 0.67 0.78 0.64 4957

weighted avg 0.91 0.72 0.78 4957

average eval\_loss: 1.07314474865674, accuracy: 0.716158967117208

training 6 epoch...

49.415913581848145 seconds used......

6 training finished! average train loss: 0.04431861204677942

total clf loss: 7.7427794493269175 total adv loss: 3.2859367345809005

evaluating...

Confusion matrix, without normalization

[[ 226 48 20]

[ 882 2828 111]

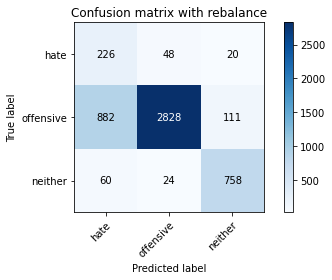
[ 60 24 758]]

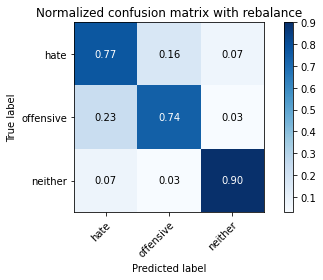
Normalized confusion matrix

[[0.76870748 0.16326531 0.06802721]

[0.23082963 0.74012039 0.02904999]

[0.07125891 0.02850356 0.90023753]]





precision recall f1-score support

hate 0.19 0.77 0.31 294

offensive 0.98 0.74 0.84 3821

neither 0.85 0.90 0.88 842

accuracy 0.77 4957

macro avg 0.67 0.80 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 1.0168306442191675, accuracy: 0.7690135162396611

training 7 epoch...

49.32111072540283 seconds used......

7 training finished! average train loss: 0.02868709848051324

total clf loss: 4.974206372804474 total adv loss: 1.3742501533706672

evaluating...

Confusion matrix, without normalization

[[ 217 60 17]

[ 762 2958 101]

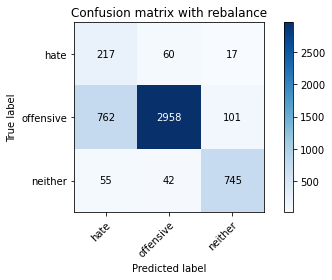
[ 55 42 745]]

Normalized confusion matrix

[[0.73809524 0.20408163 0.05782313]

[0.19942423 0.77414289 0.02643287]

[0.06532067 0.04988124 0.8847981 ]]





precision recall f1-score support

hate 0.21 0.74 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.86 0.88 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.9420604362347794, accuracy: 0.7908008876336494

training 8 epoch...

49.3179292678833 seconds used......

8 training finished! average train loss: 0.032953108775590094

total clf loss: 5.677420039195567 total adv loss: 0.848768961361202

evaluating...

Confusion matrix, without normalization

[[ 222 61 11]

[ 742 2985 94]

[ 66 39 737]]

Normalized confusion matrix

[[0.75510204 0.20748299 0.03741497]

[0.19419 0.78120911 0.02460089]

[0.0783848 0.04631829 0.87529691]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.88 0.88 0.88 842

accuracy 0.80 4957

macro avg 0.69 0.80 0.69 4957

weighted avg 0.91 0.80 0.83 4957

average eval\_loss: 0.8090411524829124, accuracy: 0.7956425257212023

training 9 epoch...

49.46766543388367 seconds used......

9 training finished! average train loss: 0.010803642361227434

total clf loss: 1.8927580203162506 total adv loss: 0.9067036471096799

evaluating...

Confusion matrix, without normalization

[[ 206 64 24]

[ 625 3071 125]

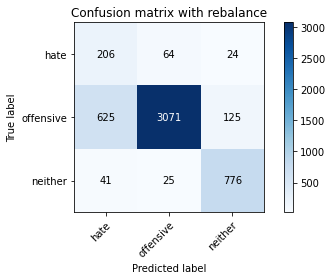
[ 41 25 776]]

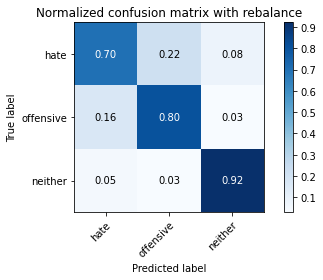
Normalized confusion matrix

[[0.70068027 0.21768707 0.08163265]

[0.16356975 0.8037163 0.03271395]

[0.04869359 0.02969121 0.9216152 ]]





precision recall f1-score support

hate 0.24 0.70 0.35 294

offensive 0.97 0.80 0.88 3821

neither 0.84 0.92 0.88 842

accuracy 0.82 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.9223410525117933, accuracy: 0.8176316320355054

training 10 epoch...

49.526840686798096 seconds used......

10 training finished! average train loss: 0.007392435065880472

total clf loss: 1.2733903017360717 total adv loss: 0.1856781349197263

evaluating...

Confusion matrix, without normalization

[[ 197 68 29]

[ 481 3206 134]

[ 24 41 777]]

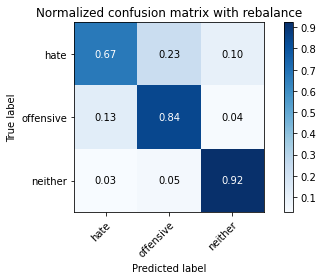
Normalized confusion matrix

[[0.67006803 0.23129252 0.09863946]

[0.12588328 0.83904737 0.03506935]

[0.02850356 0.04869359 0.92280285]]





precision recall f1-score support

hate 0.28 0.67 0.40 294

offensive 0.97 0.84 0.90 3821

neither 0.83 0.92 0.87 842

accuracy 0.84 4957

macro avg 0.69 0.81 0.72 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.8237747852452876, accuracy: 0.8432519669154731

4 run...

training 1 epoch...

49.35277986526489 seconds used......

1 training finished! average train loss: 0.6503857222565433

total clf loss: 113.99430339038372 total adv loss: 55.56690185330808

evaluating...

Confusion matrix, without normalization

[[ 235 36 23]

[ 871 2802 148]

[ 47 10 785]]

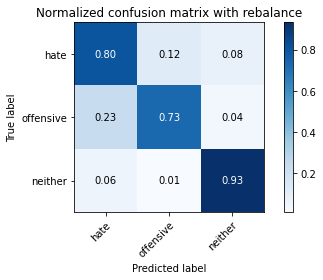
Normalized confusion matrix

[[0.79931973 0.12244898 0.07823129]

[0.2279508 0.73331589 0.03873332]

[0.05581948 0.01187648 0.93230404]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.73 0.84 3821

neither 0.82 0.93 0.87 842

accuracy 0.77 4957

macro avg 0.67 0.82 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 0.5701835586298858, accuracy: 0.7710308654428082

training 2 epoch...

49.29232382774353 seconds used......

2 training finished! average train loss: 0.391009986923452

total clf loss: 68.2115368321538 total adv loss: 26.976582024712116

evaluating...

Confusion matrix, without normalization

[[ 247 36 11]

[ 910 2820 91]

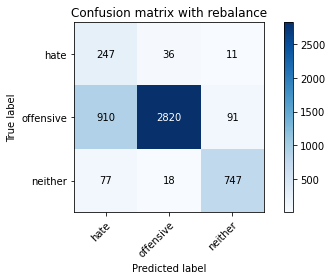
[ 77 18 747]]

Normalized confusion matrix

[[0.84013605 0.12244898 0.03741497]

[0.23815755 0.73802669 0.02381576]

[0.09144893 0.02137767 0.8871734 ]]





precision recall f1-score support

hate 0.20 0.84 0.32 294

offensive 0.98 0.74 0.84 3821

neither 0.88 0.89 0.88 842

accuracy 0.77 4957

macro avg 0.69 0.82 0.68 4957

weighted avg 0.92 0.77 0.82 4957

average eval\_loss: 0.5499561582842181, accuracy: 0.7694169860802905

training 3 epoch...

49.31545639038086 seconds used......

3 training finished! average train loss: 0.2593005023548129

total clf loss: 45.033746641129255 total adv loss: 13.86721410881728

evaluating...

Confusion matrix, without normalization

[[ 247 33 14]

[ 978 2694 149]

[ 52 11 779]]

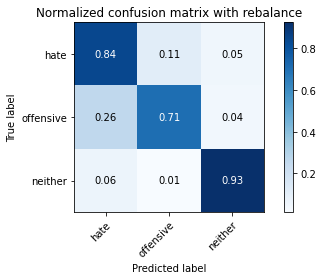
Normalized confusion matrix

[[0.84013605 0.1122449 0.04761905]

[0.25595394 0.70505103 0.03899503]

[0.06175772 0.01306413 0.92517815]]





precision recall f1-score support

hate 0.19 0.84 0.31 294

offensive 0.98 0.71 0.82 3821

neither 0.83 0.93 0.87 842

accuracy 0.75 4957

macro avg 0.67 0.82 0.67 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 0.62697653736799, accuracy: 0.7504539035707081

training 4 epoch...

49.31212854385376 seconds used......

4 training finished! average train loss: 0.15478775235797054

total clf loss: 26.818007856607437 total adv loss: 6.98604304890614

evaluating...

Confusion matrix, without normalization

[[ 235 37 22]

[1007 2629 185]

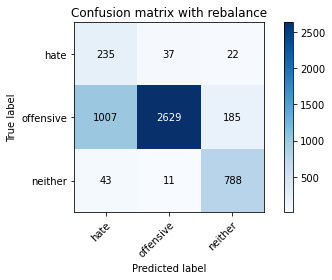
[ 43 11 788]]

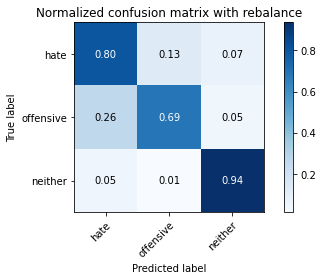
Normalized confusion matrix

[[0.79931973 0.12585034 0.07482993]

[0.26354357 0.68803978 0.04841664]

[0.05106888 0.01306413 0.93586698]]





precision recall f1-score support

hate 0.18 0.80 0.30 294

offensive 0.98 0.69 0.81 3821

neither 0.79 0.94 0.86 842

accuracy 0.74 4957

macro avg 0.65 0.81 0.65 4957

weighted avg 0.90 0.74 0.79 4957

average eval\_loss: 0.7613389603192767, accuracy: 0.7367359289893081

training 5 epoch...

49.284369230270386 seconds used......

5 training finished! average train loss: 0.07647997009627826

total clf loss: 13.185048330575228 total adv loss: 2.1394692114699865

evaluating...

Confusion matrix, without normalization

[[ 217 51 26]

[ 652 3015 154]

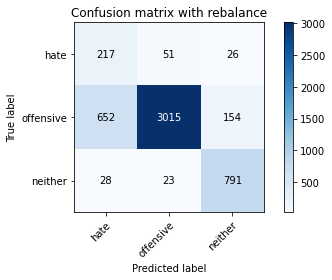
[ 28 23 791]]

Normalized confusion matrix

[[0.73809524 0.17346939 0.08843537]

[0.17063596 0.78906046 0.04030359]

[0.03325416 0.02731591 0.93942993]]





precision recall f1-score support

hate 0.24 0.74 0.36 294

offensive 0.98 0.79 0.87 3821

neither 0.81 0.94 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.6648657635838214, accuracy: 0.8115795844260641

training 6 epoch...

49.25497102737427 seconds used......

6 training finished! average train loss: 0.053307667310616515

total clf loss: 9.280209878226742 total adv loss: 3.291976683656685

evaluating...

Confusion matrix, without normalization

[[ 227 56 11]

[ 783 2938 100]

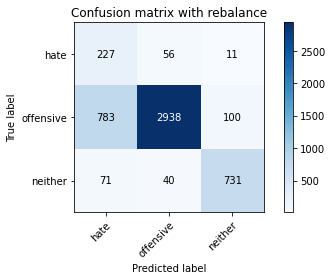
[ 71 40 731]]

Normalized confusion matrix

[[0.77210884 0.19047619 0.03741497]

[0.20492018 0.76890866 0.02617116]

[0.08432304 0.04750594 0.86817102]]





precision recall f1-score support

hate 0.21 0.77 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.87 0.87 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.8841471221793683, accuracy: 0.7859592495460964

training 7 epoch...

49.316011905670166 seconds used......

7 training finished! average train loss: 0.02978175061915549

total clf loss: 5.161136779468507 total adv loss: 1.3691485760500655

evaluating...

Confusion matrix, without normalization

[[ 219 51 24]

[ 801 2885 135]

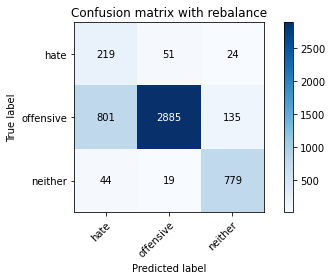
[ 44 19 779]]

Normalized confusion matrix

[[0.74489796 0.17346939 0.08163265]

[0.20963099 0.75503795 0.03533107]

[0.05225653 0.02256532 0.92517815]]





precision recall f1-score support

hate 0.21 0.74 0.32 294

offensive 0.98 0.76 0.85 3821

neither 0.83 0.93 0.88 842

accuracy 0.78 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 1.0525382429064671, accuracy: 0.7833366955820052

training 8 epoch...

49.229655265808105 seconds used......

8 training finished! average train loss: 0.011725110765369555

total clf loss: 2.0213617271510884 total adv loss: 0.3273554753541248

evaluating...

Confusion matrix, without normalization

[[ 238 42 14]

[ 927 2780 114]

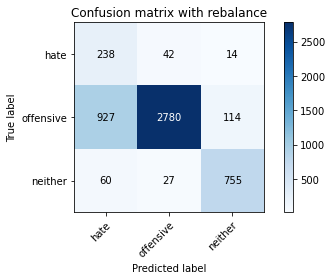
[ 60 27 755]]

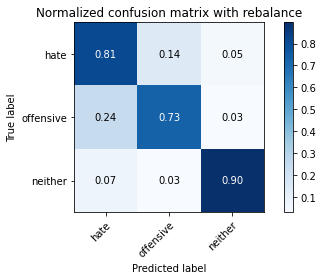
Normalized confusion matrix

[[0.80952381 0.14285714 0.04761905]

[0.24260665 0.72755823 0.02983512]

[0.07125891 0.03206651 0.89667458]]





precision recall f1-score support

hate 0.19 0.81 0.31 294

offensive 0.98 0.73 0.83 3821

neither 0.86 0.90 0.88 842

accuracy 0.76 4957

macro avg 0.68 0.81 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.2074502668673954, accuracy: 0.7611458543473876

training 9 epoch...

49.27281951904297 seconds used......

9 training finished! average train loss: 0.005661530140208691

total clf loss: 0.9794496392132714 total adv loss: 0.22655969935294706

evaluating...

Confusion matrix, without normalization

[[ 207 60 27]

[ 632 3050 139]

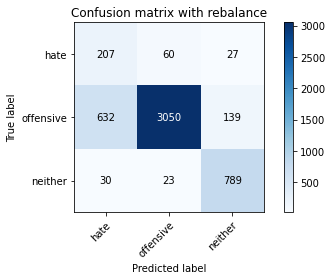
[ 30 23 789]]

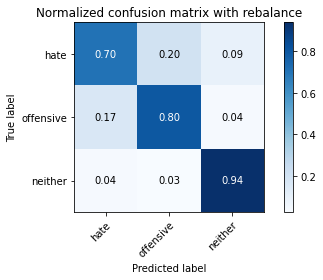
Normalized confusion matrix

[[0.70408163 0.20408163 0.09183673]

[0.16540173 0.79822036 0.03637791]

[0.03562945 0.02731591 0.93705463]]





precision recall f1-score support

hate 0.24 0.70 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.83 0.94 0.88 842

accuracy 0.82 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 1.0675350366901482, accuracy: 0.8162194875933024

training 10 epoch...

49.271275758743286 seconds used......

10 training finished! average train loss: 0.009122831023850326

total clf loss: 1.5807170501939254 total adv loss: 0.41425869925660663

evaluating...

Confusion matrix, without normalization

[[ 222 57 15]

[ 789 2931 101]

[ 67 42 733]]

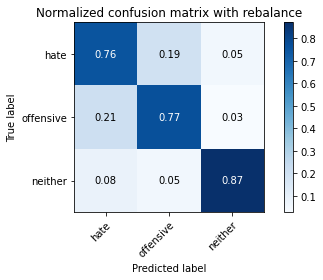
Normalized confusion matrix

[[0.75510204 0.19387755 0.05102041]

[0.20649045 0.76707668 0.02643287]

[0.07957245 0.04988124 0.87054632]]





precision recall f1-score support

hate 0.21 0.76 0.32 294

offensive 0.97 0.77 0.86 3821

neither 0.86 0.87 0.87 842

accuracy 0.78 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.90 0.78 0.83 4957

average eval\_loss: 1.1434454863305175, accuracy: 0.7839419003429494

5 run...

training 1 epoch...

49.31991791725159 seconds used......

1 training finished! average train loss: 0.6589183525035256

total clf loss: 115.66087305545807 total adv loss: 59.71668849140406

evaluating...

Confusion matrix, without normalization

[[ 218 43 33]

[ 743 2893 185]

[ 35 11 796]]

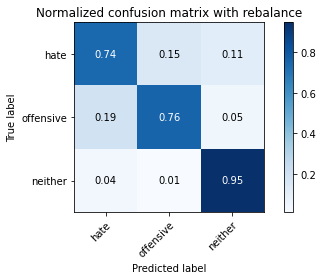
Normalized confusion matrix

[[0.7414966 0.1462585 0.1122449 ]

[0.19445171 0.75713164 0.04841664]

[0.0415677 0.01306413 0.94536817]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.98 0.76 0.85 3821

neither 0.79 0.95 0.86 842

accuracy 0.79 4957

macro avg 0.66 0.81 0.68 4957

weighted avg 0.90 0.79 0.82 4957

average eval\_loss: 0.474362586474707, accuracy: 0.7881783336695583

training 2 epoch...

49.33419132232666 seconds used......

2 training finished! average train loss: 0.383036938414239

total clf loss: 66.84178426861763 total adv loss: 26.849356526508927

evaluating...

Confusion matrix, without normalization

[[ 244 32 18]

[ 982 2746 93]

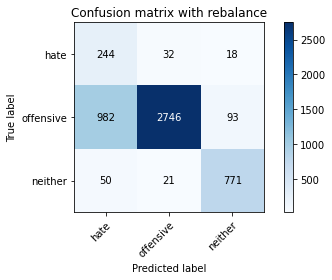
[ 50 21 771]]

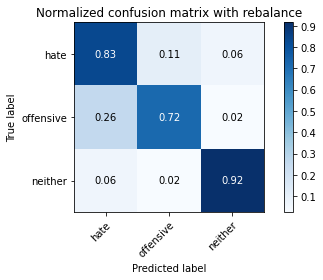
Normalized confusion matrix

[[0.82993197 0.10884354 0.06122449]

[0.25700079 0.71866004 0.02433918]

[0.05938242 0.02494062 0.91567696]]





precision recall f1-score support

hate 0.19 0.83 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.87 0.92 0.89 842

accuracy 0.76 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.92 0.76 0.81 4957

average eval\_loss: 0.5240807833690797, accuracy: 0.758725035303611

training 3 epoch...

49.35660529136658 seconds used......

3 training finished! average train loss: 0.25503418766702823

total clf loss: 44.30069216713309 total adv loss: 13.796918857377023

evaluating...

Confusion matrix, without normalization

[[ 214 63 17]

[ 547 3151 123]

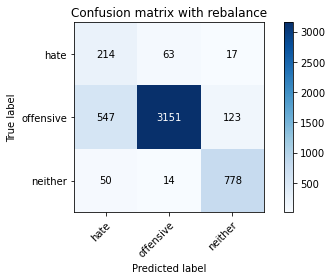
[ 50 14 778]]

Normalized confusion matrix

[[0.72789116 0.21428571 0.05782313]

[0.14315624 0.82465323 0.03219053]

[0.05938242 0.01662708 0.9239905 ]]





precision recall f1-score support

hate 0.26 0.73 0.39 294

offensive 0.98 0.82 0.89 3821

neither 0.85 0.92 0.88 842

accuracy 0.84 4957

macro avg 0.70 0.83 0.72 4957

weighted avg 0.91 0.84 0.86 4957

average eval\_loss: 0.4266079334662326, accuracy: 0.8357877748638289

training 4 epoch...

49.36906051635742 seconds used......

4 training finished! average train loss: 0.1628950636354629

total clf loss: 28.187359208241105 total adv loss: 6.646065367967822

evaluating...

Confusion matrix, without normalization

[[ 241 35 18]

[1031 2672 118]

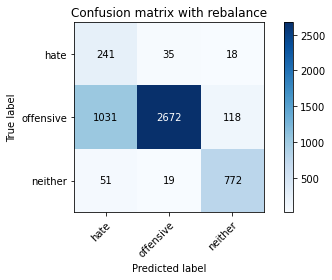
[ 51 19 772]]

Normalized confusion matrix

[[0.81972789 0.11904762 0.06122449]

[0.26982465 0.69929338 0.03088197]

[0.06057007 0.02256532 0.91686461]]





precision recall f1-score support

hate 0.18 0.82 0.30 294

offensive 0.98 0.70 0.82 3821

neither 0.85 0.92 0.88 842

accuracy 0.74 4957

macro avg 0.67 0.81 0.67 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 0.6972027557992166, accuracy: 0.7433931813596933

training 5 epoch...

49.301265478134155 seconds used......

5 training finished! average train loss: 0.08285169215731279

total clf loss: 14.30082057369873 total adv loss: 2.6636242564418353

evaluating...

Confusion matrix, without normalization

[[ 205 70 19]

[ 537 3164 120]

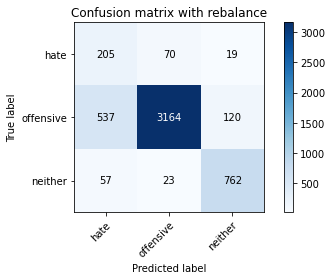
[ 57 23 762]]

Normalized confusion matrix

[[0.69727891 0.23809524 0.06462585]

[0.14053913 0.82805548 0.03140539]

[0.06769596 0.02731591 0.90498812]]





precision recall f1-score support

hate 0.26 0.70 0.38 294

offensive 0.97 0.83 0.89 3821

neither 0.85 0.90 0.87 842

accuracy 0.83 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.5937950251654031, accuracy: 0.8333669558200525

training 6 epoch...

49.28462219238281 seconds used......

6 training finished! average train loss: 0.046459048401635165

total clf loss: 8.02577655587811 total adv loss: 1.6255859548109584

evaluating...

Confusion matrix, without normalization

[[ 234 42 18]

[ 897 2816 108]

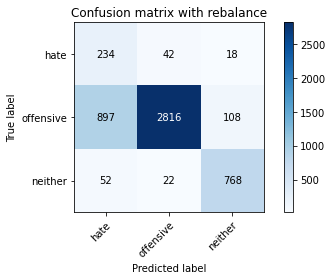
[ 52 22 768]]

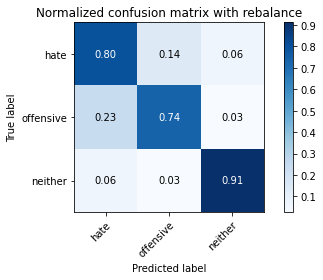
Normalized confusion matrix

[[0.79591837 0.14285714 0.06122449]

[0.2347553 0.73697985 0.02826485]

[0.06175772 0.02612827 0.91211401]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.74 0.84 3821

neither 0.86 0.91 0.88 842

accuracy 0.77 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 1.0219857898829956, accuracy: 0.7702239257615493

training 7 epoch...

49.39573907852173 seconds used......

7 training finished! average train loss: 0.021589580502589804

total clf loss: 3.7517181003349833 total adv loss: 1.1979965406935662

evaluating...

Confusion matrix, without normalization

[[ 192 89 13]

[ 452 3282 87]

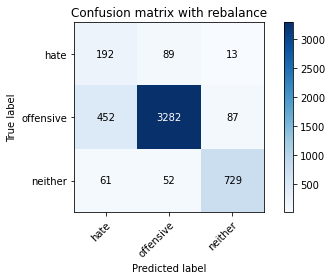
[ 61 52 729]]

Normalized confusion matrix

[[0.65306122 0.30272109 0.04421769]

[0.11829364 0.85893745 0.02276891]

[0.07244656 0.06175772 0.86579572]]





precision recall f1-score support

hate 0.27 0.65 0.38 294

offensive 0.96 0.86 0.91 3821

neither 0.88 0.87 0.87 842

accuracy 0.85 4957

macro avg 0.70 0.79 0.72 4957

weighted avg 0.90 0.85 0.87 4957

average eval\_loss: 0.646365995280869, accuracy: 0.8478918700827113

training 8 epoch...

49.40034627914429 seconds used......

8 training finished! average train loss: 0.021933481376856422

total clf loss: 3.8255709366349038 total adv loss: 1.4989115679854876

evaluating...

Confusion matrix, without normalization

[[ 225 55 14]

[ 711 3006 104]

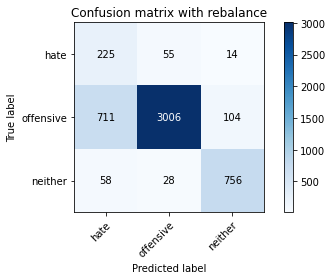
[ 58 28 756]]

Normalized confusion matrix

[[0.76530612 0.18707483 0.04761905]

[0.18607694 0.78670505 0.02721801]

[0.06888361 0.03325416 0.89786223]]





precision recall f1-score support

hate 0.23 0.77 0.35 294

offensive 0.97 0.79 0.87 3821

neither 0.86 0.90 0.88 842

accuracy 0.80 4957

macro avg 0.69 0.82 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.9248786530008675, accuracy: 0.8043171272947347

training 9 epoch...

49.2836012840271 seconds used......

9 training finished! average train loss: 0.01311870086527966

total clf loss: 2.2800016222754493 total adv loss: 0.7340754565120733

evaluating...

Confusion matrix, without normalization

[[ 189 82 23]

[ 517 3178 126]

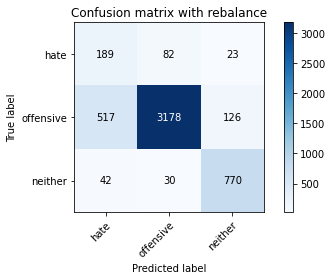
[ 42 30 770]]

Normalized confusion matrix

[[0.64285714 0.27891156 0.07823129]

[0.13530489 0.83171945 0.03297566]

[0.04988124 0.03562945 0.91448931]]





precision recall f1-score support

hate 0.25 0.64 0.36 294

offensive 0.97 0.83 0.89 3821

neither 0.84 0.91 0.87 842

accuracy 0.83 4957

macro avg 0.69 0.80 0.71 4957

weighted avg 0.90 0.83 0.86 4957

average eval\_loss: 0.815724046829517, accuracy: 0.8345773653419407

training 10 epoch...

49.39024829864502 seconds used......

10 training finished! average train loss: 0.012168435558655615

total clf loss: 2.0928971982648363 total adv loss: 0.2418947635233053

evaluating...

Confusion matrix, without normalization

[[ 205 64 25]

[ 596 3103 122]

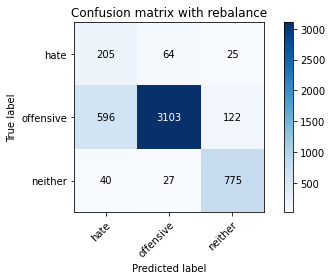
[ 40 27 775]]

Normalized confusion matrix

[[0.69727891 0.21768707 0.08503401]

[0.15598011 0.81209108 0.03192881]

[0.04750594 0.03206651 0.92042755]]





precision recall f1-score support

hate 0.24 0.70 0.36 294

offensive 0.97 0.81 0.88 3821

neither 0.84 0.92 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.9193765851002442, accuracy: 0.8236836796449465