hate as offensive

alpha=0.05

1 run...

Downloading: 100%

411/411 [00:00<00:00, 14.7kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 22.9MB/s]

training 1 epoch...

104.27458906173706 seconds used......

1 training finished! average train loss: 0.6844569076920113

total clf loss: 120.05704921483994 total adv loss: 60.29835822060704

evaluating...

Confusion matrix, without normalization

[[ 201 77 16]

[ 410 3331 80]

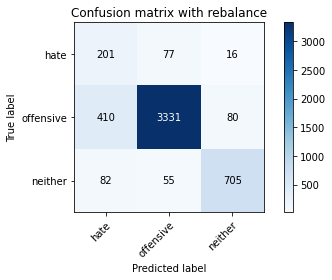
[ 82 55 705]]

Normalized confusion matrix

[[0.68367347 0.26190476 0.05442177]

[0.10730175 0.87176132 0.02093693]

[0.09738717 0.06532067 0.83729216]]





precision recall f1-score support

hate 0.29 0.68 0.41 294

offensive 0.96 0.87 0.91 3821

neither 0.88 0.84 0.86 842

accuracy 0.85 4957

macro avg 0.71 0.80 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.3918466524130875, accuracy: 0.8547508573734114

training 2 epoch...

103.684321641922 seconds used......

2 training finished! average train loss: 0.3878088651431931

total clf loss: 67.5725331902504 total adv loss: 25.144349020905793

evaluating...

Confusion matrix, without normalization

[[ 226 35 33]

[1006 2673 142]

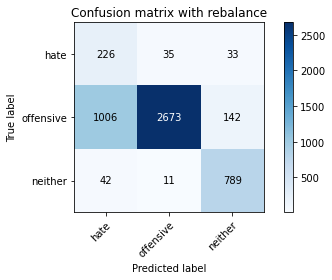
[ 42 11 789]]

Normalized confusion matrix

[[0.76870748 0.11904762 0.1122449 ]

[0.26328186 0.69955509 0.03716305]

[0.04988124 0.01306413 0.93705463]]





precision recall f1-score support

hate 0.18 0.77 0.29 294

offensive 0.98 0.70 0.82 3821

neither 0.82 0.94 0.87 842

accuracy 0.74 4957

macro avg 0.66 0.80 0.66 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 0.5879411535758164, accuracy: 0.7439983861206375

training 3 epoch...

103.66180109977722 seconds used......

3 training finished! average train loss: 0.27420620626786296

total clf loss: 47.51442924514413 total adv loss: 12.50335866806563

evaluating...

Confusion matrix, without normalization

[[ 228 48 18]

[ 790 2935 96]

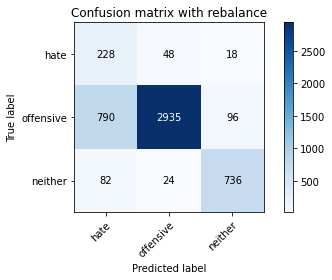
[ 82 24 736]]

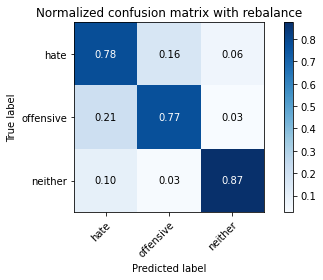
Normalized confusion matrix

[[0.7755102 0.16326531 0.06122449]

[0.20675216 0.76812353 0.02512431]

[0.09738717 0.02850356 0.87410926]]





precision recall f1-score support

hate 0.21 0.78 0.33 294

offensive 0.98 0.77 0.86 3821

neither 0.87 0.87 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.5216961504681216, accuracy: 0.7865644543070406

training 4 epoch...

103.65726780891418 seconds used......

4 training finished! average train loss: 0.16718430203870374

total clf loss: 28.841583028435707 total adv loss: 5.061347450937319

evaluating...

Confusion matrix, without normalization

[[ 214 58 22]

[ 583 3141 97]

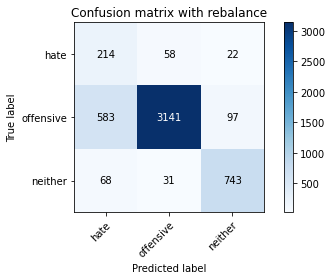
[ 68 31 743]]

Normalized confusion matrix

[[0.72789116 0.19727891 0.07482993]

[0.15257786 0.82203612 0.02538602]

[0.0807601 0.0368171 0.8824228 ]]





precision recall f1-score support

hate 0.25 0.73 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.86 0.88 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: 0.5502652401712933, accuracy: 0.8267097034496671

training 5 epoch...

103.69307518005371 seconds used......

5 training finished! average train loss: 0.10035125691748677

total clf loss: 17.31856274791062 total adv loss: 3.1699554593215

evaluating...

Confusion matrix, without normalization

[[ 205 69 20]

[ 627 3104 90]

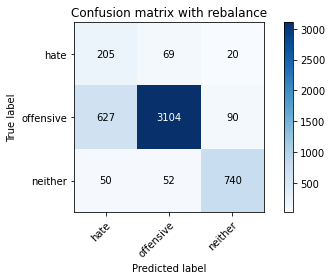
[ 50 52 740]]

Normalized confusion matrix

[[0.69727891 0.23469388 0.06802721]

[0.16409317 0.81235279 0.02355404]

[0.05938242 0.06175772 0.87885986]]





precision recall f1-score support

hate 0.23 0.70 0.35 294

offensive 0.96 0.81 0.88 3821

neither 0.87 0.88 0.87 842

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: 0.5677428264152108, accuracy: 0.8168246923542465

training 6 epoch...

103.6640272140503 seconds used......

6 training finished! average train loss: 0.04688712648845861

total clf loss: 8.065520328469574 total adv loss: 0.9564346594829658

evaluating...

Confusion matrix, without normalization

[[ 219 55 20]

[ 701 3030 90]

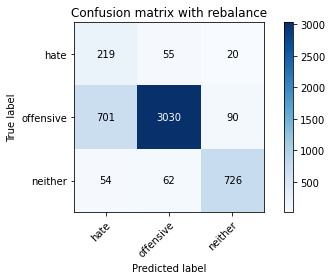
[ 54 62 726]]

Normalized confusion matrix

[[0.74489796 0.18707483 0.06802721]

[0.18345983 0.79298613 0.02355404]

[0.06413302 0.0736342 0.86223278]]





precision recall f1-score support

hate 0.22 0.74 0.35 294

offensive 0.96 0.79 0.87 3821

neither 0.87 0.86 0.87 842

accuracy 0.80 4957

macro avg 0.69 0.80 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.7605715051984354, accuracy: 0.8018963082509583

training 7 epoch...

103.64301180839539 seconds used......

7 training finished! average train loss: 0.027753650824190797

total clf loss: 4.851699155755341 total adv loss: 2.1164971712128136

evaluating...

Confusion matrix, without normalization

[[ 236 42 16]

[1100 2611 110]

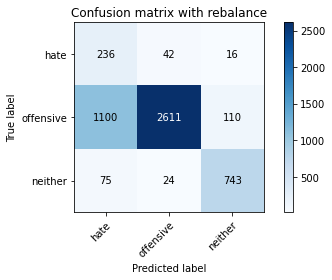
[ 75 24 743]]

Normalized confusion matrix

[[0.80272109 0.14285714 0.05442177]

[0.28788275 0.68332897 0.02878828]

[0.08907363 0.02850356 0.8824228 ]]





precision recall f1-score support

hate 0.17 0.80 0.28 294

offensive 0.98 0.68 0.80 3821

neither 0.86 0.88 0.87 842

accuracy 0.72 4957

macro avg 0.67 0.79 0.65 4957

weighted avg 0.91 0.72 0.78 4957

average eval\_loss: 1.3215596664757017, accuracy: 0.7242283639297963

training 8 epoch...

103.6696662902832 seconds used......

8 training finished! average train loss: 0.02315102457337773

total clf loss: 4.023577623418532 total adv loss: 1.2950482537708012

evaluating...

Confusion matrix, without normalization

[[ 202 72 20]

[ 486 3243 92]

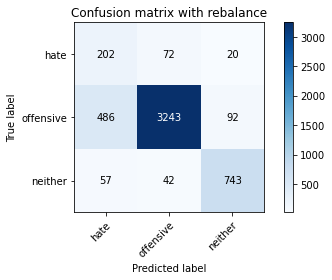
[ 57 42 743]]

Normalized confusion matrix

[[0.68707483 0.24489796 0.06802721]

[0.12719183 0.8487307 0.02407747]

[0.06769596 0.04988124 0.8824228 ]]





precision recall f1-score support

hate 0.27 0.69 0.39 294

offensive 0.97 0.85 0.90 3821

neither 0.87 0.88 0.88 842

accuracy 0.84 4957

macro avg 0.70 0.81 0.72 4957

weighted avg 0.91 0.84 0.87 4957

average eval\_loss: 0.750139553366734, accuracy: 0.8448658462779908

training 9 epoch...

103.60516285896301 seconds used......

9 training finished! average train loss: 0.01397495261939482

total clf loss: 2.428485456475755 total adv loss: 0.7753713110123499

evaluating...

Confusion matrix, without normalization

[[ 225 42 27]

[ 820 2872 129]

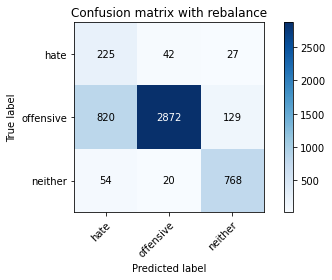
[ 54 20 768]]

Normalized confusion matrix

[[0.76530612 0.14285714 0.09183673]

[0.21460351 0.7516357 0.0337608 ]

[0.06413302 0.02375297 0.91211401]]





precision recall f1-score support

hate 0.20 0.77 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.83 0.91 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: 1.200200519663462, accuracy: 0.7797054670163406

training 10 epoch...

103.66526675224304 seconds used......

10 training finished! average train loss: 0.01119292868071165

total clf loss: 1.948591654887423 total adv loss: 0.6920171739875514

evaluating...

Confusion matrix, without normalization

[[ 223 52 19]

[ 847 2883 91]

[ 77 40 725]]

Normalized confusion matrix

[[0.7585034 0.17687075 0.06462585]

[0.22166972 0.75451452 0.02381576]

[0.09144893 0.04750594 0.86104513]]





precision recall f1-score support

hate 0.19 0.76 0.31 294

offensive 0.97 0.75 0.85 3821

neither 0.87 0.86 0.86 842

accuracy 0.77 4957

macro avg 0.68 0.79 0.67 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 1.198279129925199, accuracy: 0.7728464797256405

2 run...

training 1 epoch...

103.71934175491333 seconds used......

1 training finished! average train loss: 0.6368288375132265

total clf loss: 111.5516742169857 total adv loss: 53.07886011712253

evaluating...

Confusion matrix, without normalization

[[ 190 64 40]

[ 429 3226 166]

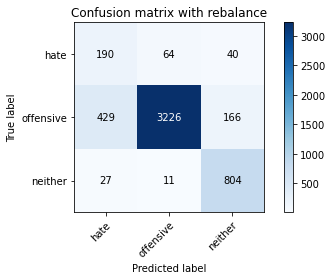
[ 27 11 804]]

Normalized confusion matrix

[[0.6462585 0.21768707 0.13605442]

[0.11227427 0.8442816 0.04344412]

[0.03206651 0.01306413 0.95486936]]





precision recall f1-score support

hate 0.29 0.65 0.40 294

offensive 0.98 0.84 0.91 3821

neither 0.80 0.95 0.87 842

accuracy 0.85 4957

macro avg 0.69 0.82 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.42180826887488365, accuracy: 0.8513213637280613

training 2 epoch...

103.64938569068909 seconds used......

2 training finished! average train loss: 0.38514608044547644

total clf loss: 67.19283699244261 total adv loss: 26.657142265466973

evaluating...

Confusion matrix, without normalization

[[ 219 59 16]

[ 599 3124 98]

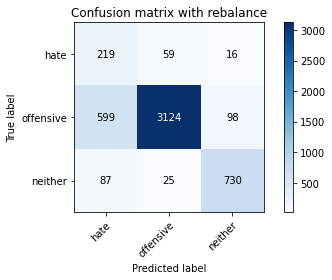
[ 87 25 730]]

Normalized confusion matrix

[[0.74489796 0.20068027 0.05442177]

[0.15676524 0.81758702 0.02564774]

[0.10332542 0.02969121 0.86698337]]





precision recall f1-score support

hate 0.24 0.74 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.86 0.87 0.87 842

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: 0.4323575993879668, accuracy: 0.8216663304417995

training 3 epoch...

103.60568308830261 seconds used......

3 training finished! average train loss: 0.25839657353901724

total clf loss: 44.82703058794141 total adv loss: 12.824332285526907

evaluating...

Confusion matrix, without normalization

[[ 222 44 28]

[ 788 2881 152]

[ 43 12 787]]

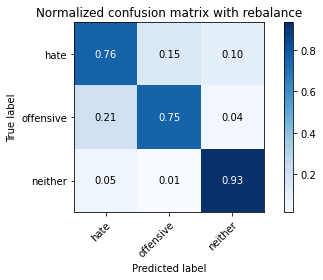
Normalized confusion matrix

[[0.75510204 0.14965986 0.0952381 ]

[0.20622874 0.7539911 0.03978016]

[0.05106888 0.01425178 0.93467933]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.98 0.75 0.85 3821

neither 0.81 0.93 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: 0.5371254216428967, accuracy: 0.7847488400242082

training 4 epoch...

103.78992867469788 seconds used......

4 training finished! average train loss: 0.16046835641153376

total clf loss: 27.69056124612689 total adv loss: 5.009444475770579

evaluating...

Confusion matrix, without normalization

[[ 201 72 21]

[ 485 3181 155]

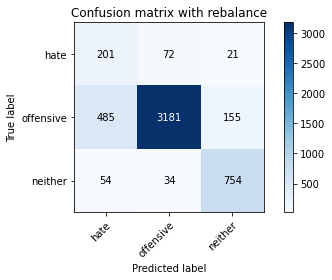
[ 54 34 754]]

Normalized confusion matrix

[[0.68367347 0.24489796 0.07142857]

[0.12693012 0.83250458 0.0405653 ]

[0.06413302 0.04038005 0.89548694]]





precision recall f1-score support

hate 0.27 0.68 0.39 294

offensive 0.97 0.83 0.90 3821

neither 0.81 0.90 0.85 842

accuracy 0.83 4957

macro avg 0.68 0.80 0.71 4957

weighted avg 0.90 0.83 0.86 4957

average eval\_loss: 0.5037219350812056, accuracy: 0.834375630421626

training 5 epoch...

103.63298153877258 seconds used......

5 training finished! average train loss: 0.08154016374256352

total clf loss: 14.071595068555325 total adv loss: 2.564540129234956

evaluating...

Confusion matrix, without normalization

[[ 260 22 12]

[1712 2028 81]

[ 146 26 670]]

Normalized confusion matrix

[[0.88435374 0.07482993 0.04081633]

[0.44805025 0.53075111 0.02119864]

[0.17339667 0.03087886 0.79572447]]





precision recall f1-score support

hate 0.12 0.88 0.22 294

offensive 0.98 0.53 0.69 3821

neither 0.88 0.80 0.83 842

accuracy 0.60 4957

macro avg 0.66 0.74 0.58 4957

weighted avg 0.91 0.60 0.68 4957

average eval\_loss: 1.5923498083627032, accuracy: 0.5967318942909018

training 6 epoch...

103.70072865486145 seconds used......

6 training finished! average train loss: 0.049641201353621024

total clf loss: 8.651437763939612 total adv loss: 3.255846619254953

evaluating...

Confusion matrix, without normalization

[[ 223 53 18]

[ 760 2970 91]

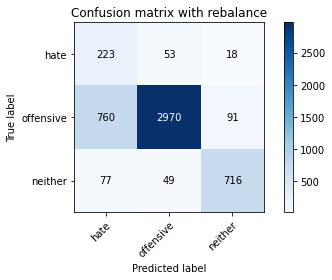
[ 77 49 716]]

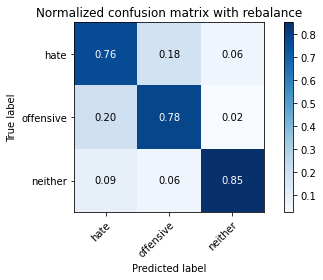
Normalized confusion matrix

[[0.7585034 0.18027211 0.06122449]

[0.19890081 0.77728343 0.02381576]

[0.09144893 0.05819477 0.85035629]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.97 0.78 0.86 3821

neither 0.87 0.85 0.86 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.8897625267986329, accuracy: 0.7885818035101876

training 7 epoch...

103.69493913650513 seconds used......

7 training finished! average train loss: 0.018669944012140016

total clf loss: 3.2179650561884046 total adv loss: 0.5080926873126828

evaluating...

Confusion matrix, without normalization

[[ 239 42 13]

[ 931 2781 109]

[ 111 20 711]]

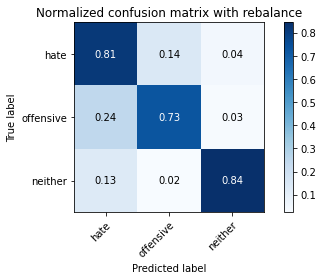
Normalized confusion matrix

[[0.81292517 0.14285714 0.04421769]

[0.24365349 0.72781994 0.02852656]

[0.13182898 0.02375297 0.84441805]]





precision recall f1-score support

hate 0.19 0.81 0.30 294

offensive 0.98 0.73 0.83 3821

neither 0.85 0.84 0.85 842

accuracy 0.75 4957

macro avg 0.67 0.80 0.66 4957

weighted avg 0.91 0.75 0.81 4957

average eval\_loss: 1.2788505471281466, accuracy: 0.7526729876941699

training 8 epoch...

103.69373083114624 seconds used......

8 training finished! average train loss: 0.01969745872121533

total clf loss: 3.4146130627195816 total adv loss: 0.926952178426518

evaluating...

Confusion matrix, without normalization

[[ 208 65 21]

[ 675 3043 103]

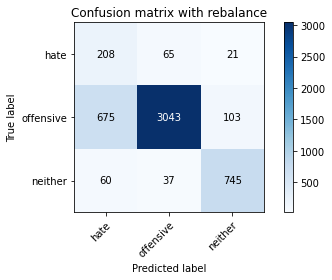
[ 60 37 745]]

Normalized confusion matrix

[[0.70748299 0.22108844 0.07142857]

[0.17665533 0.79638838 0.02695629]

[0.07125891 0.04394299 0.8847981 ]]





precision recall f1-score support

hate 0.22 0.71 0.34 294

offensive 0.97 0.80 0.87 3821

neither 0.86 0.88 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.8949729990169045, accuracy: 0.8061327415775671

training 9 epoch...

103.68613028526306 seconds used......

9 training finished! average train loss: 0.019956544771441938

total clf loss: 3.47082848395803 total adv loss: 1.1651869868454465

evaluating...

Confusion matrix, without normalization

[[ 210 63 21]

[ 659 3054 108]

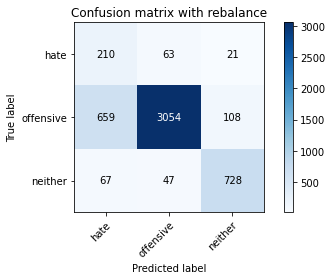
[ 67 47 728]]

Normalized confusion matrix

[[0.71428571 0.21428571 0.07142857]

[0.17246794 0.79926721 0.02826485]

[0.07957245 0.05581948 0.86460808]]





precision recall f1-score support

hate 0.22 0.71 0.34 294

offensive 0.97 0.80 0.87 3821

neither 0.85 0.86 0.86 842

accuracy 0.81 4957

macro avg 0.68 0.79 0.69 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 1.0061989261828843, accuracy: 0.8053258018963082

training 10 epoch...

103.71328163146973 seconds used......

10 training finished! average train loss: 0.033487084267214376

total clf loss: 5.854512112797238 total adv loss: 2.564414410739573

evaluating...

Confusion matrix, without normalization

[[ 208 65 21]

[ 642 3073 106]

[ 36 49 757]]

Normalized confusion matrix

[[0.70748299 0.22108844 0.07142857]

[0.16801884 0.80423973 0.02774143]

[0.04275534 0.05819477 0.89904988]]





precision recall f1-score support

hate 0.23 0.71 0.35 294

offensive 0.96 0.80 0.88 3821

neither 0.86 0.90 0.88 842

accuracy 0.81 4957

macro avg 0.69 0.80 0.70 4957

weighted avg 0.90 0.81 0.85 4957

average eval\_loss: 0.877612539394308, accuracy: 0.8146056082307848

3 run...

training 1 epoch...

103.70312333106995 seconds used......

1 training finished! average train loss: 0.6777155476006848

total clf loss: 118.84809389710426 total adv loss: 59.17470365203917

evaluating...

Confusion matrix, without normalization

[[ 217 36 41]

[ 866 2795 160]

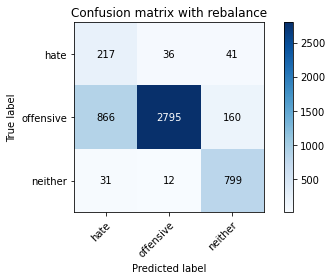
[ 31 12 799]]

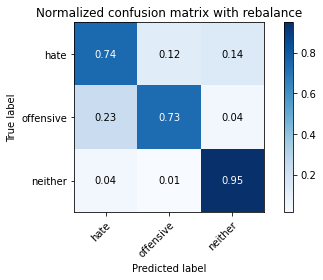
Normalized confusion matrix

[[0.73809524 0.12244898 0.13945578]

[0.22664224 0.7314839 0.04187386]

[0.0368171 0.01425178 0.94893112]]





precision recall f1-score support

hate 0.19 0.74 0.31 294

offensive 0.98 0.73 0.84 3821

neither 0.80 0.95 0.87 842

accuracy 0.77 4957

macro avg 0.66 0.81 0.67 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 0.5889325775806943, accuracy: 0.7688117813193464

training 2 epoch...

103.69455146789551 seconds used......

2 training finished! average train loss: 0.3880470647797947

total clf loss: 67.62347564101219 total adv loss: 25.34855239163153

evaluating...

Confusion matrix, without normalization

[[ 213 45 36]

[ 754 2917 150]

[ 30 12 800]]

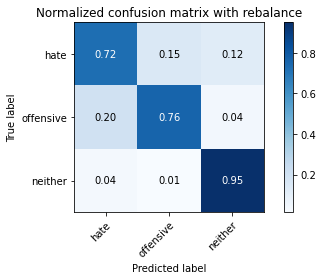
Normalized confusion matrix

[[0.7244898 0.15306122 0.12244898]

[0.19733054 0.76341272 0.03925674]

[0.03562945 0.01425178 0.95011876]]





precision recall f1-score support

hate 0.21 0.72 0.33 294

offensive 0.98 0.76 0.86 3821

neither 0.81 0.95 0.88 842

accuracy 0.79 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.4645954527020935, accuracy: 0.7928182368367964

training 3 epoch...

103.73793506622314 seconds used......

3 training finished! average train loss: 0.2724475003234301

total clf loss: 47.29296046681702 total adv loss: 14.088759783655405

evaluating...

Confusion matrix, without normalization

[[ 189 72 33]

[ 391 3268 162]

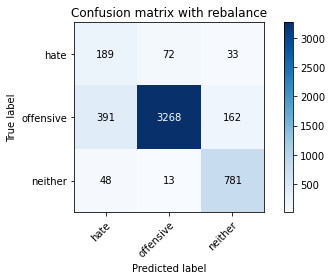
[ 48 13 781]]

Normalized confusion matrix

[[0.64285714 0.24489796 0.1122449 ]

[0.10232923 0.85527349 0.04239728]

[0.05700713 0.01543943 0.92755344]]





precision recall f1-score support

hate 0.30 0.64 0.41 294

offensive 0.97 0.86 0.91 3821

neither 0.80 0.93 0.86 842

accuracy 0.85 4957

macro avg 0.69 0.81 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.4284143947936114, accuracy: 0.854952592293726

training 4 epoch...

103.68037629127502 seconds used......

4 training finished! average train loss: 0.17477363418809502

total clf loss: 30.191583277657628 total adv loss: 6.105838309653336

evaluating...

Confusion matrix, without normalization

[[ 217 51 26]

[ 643 3041 137]

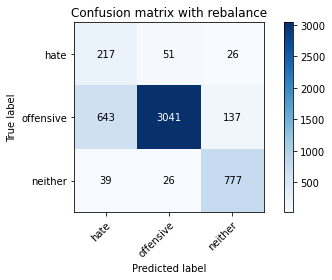
[ 39 26 777]]

Normalized confusion matrix

[[0.73809524 0.17346939 0.08843537]

[0.16828055 0.79586496 0.03585449]

[0.04631829 0.03087886 0.92280285]]





precision recall f1-score support

hate 0.24 0.74 0.36 294

offensive 0.98 0.80 0.88 3821

neither 0.83 0.92 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.91 0.81 0.85 4957

average eval\_loss: 0.5361361259354218, accuracy: 0.8140004034698406

training 5 epoch...

103.72400426864624 seconds used......

5 training finished! average train loss: 0.10477363095994581

total clf loss: 18.061541688628495 total adv loss: 2.9050160921542556

evaluating...

Confusion matrix, without normalization

[[ 201 74 19]

[ 442 3263 116]

[ 48 36 758]]

Normalized confusion matrix

[[0.68367347 0.25170068 0.06462585]

[0.11567652 0.85396493 0.03035854]

[0.05700713 0.04275534 0.90023753]]





precision recall f1-score support

hate 0.29 0.68 0.41 294

offensive 0.97 0.85 0.91 3821

neither 0.85 0.90 0.87 842

accuracy 0.85 4957

macro avg 0.70 0.81 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.493588256739801, accuracy: 0.8517248335686908

training 6 epoch...

103.88526272773743 seconds used......

6 training finished! average train loss: 0.050495304998516906

total clf loss: 8.697990742744878 total adv loss: 1.26587138816285

evaluating...

Confusion matrix, without normalization

[[ 217 64 13]

[ 610 3118 93]

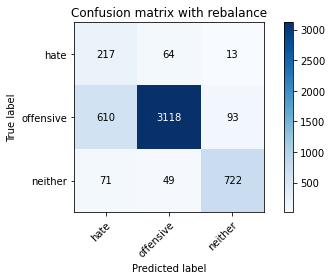
[ 71 49 722]]

Normalized confusion matrix

[[0.73809524 0.21768707 0.04421769]

[0.15964407 0.81601675 0.02433918]

[0.08432304 0.05819477 0.85748219]]





precision recall f1-score support

hate 0.24 0.74 0.36 294

offensive 0.97 0.82 0.88 3821

neither 0.87 0.86 0.86 842

accuracy 0.82 4957

macro avg 0.69 0.80 0.70 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.7168351608486245, accuracy: 0.8184385717167642

training 7 epoch...

103.73753070831299 seconds used......

7 training finished! average train loss: 0.03510966367661485

total clf loss: 6.092650398466503 total adv loss: 1.7779579577963887

evaluating...

Confusion matrix, without normalization

[[ 209 60 25]

[ 638 3062 121]

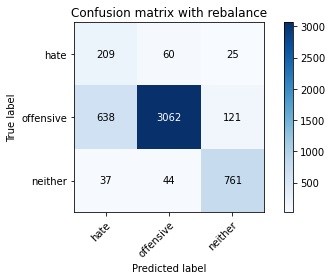
[ 37 44 761]]

Normalized confusion matrix

[[0.71088435 0.20408163 0.08503401]

[0.166972 0.8013609 0.0316671 ]

[0.04394299 0.05225653 0.90380048]]





precision recall f1-score support

hate 0.24 0.71 0.35 294

offensive 0.97 0.80 0.88 3821

neither 0.84 0.90 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.7918505280798737, accuracy: 0.8133951987088965

training 8 epoch...

103.75524687767029 seconds used......

8 training finished! average train loss: 0.029986660777079215

total clf loss: 5.211077349085826 total adv loss: 1.667166449634351

evaluating...

Confusion matrix, without normalization

[[ 213 53 28]

[ 852 2823 146]

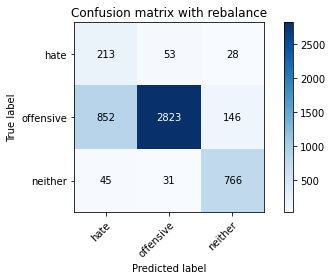
[ 45 31 766]]

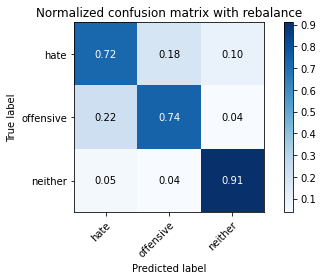
Normalized confusion matrix

[[0.7244898 0.18027211 0.0952381 ]

[0.22297828 0.73881183 0.03820989]

[0.05344418 0.0368171 0.90973872]]





precision recall f1-score support

hate 0.19 0.72 0.30 294

offensive 0.97 0.74 0.84 3821

neither 0.81 0.91 0.86 842

accuracy 0.77 4957

macro avg 0.66 0.79 0.67 4957

weighted avg 0.90 0.77 0.81 4957

average eval\_loss: 1.1504127165873446, accuracy: 0.7669961670365141

training 9 epoch...

103.69775414466858 seconds used......

9 training finished! average train loss: 0.03225239068821171

total clf loss: 5.532641621073708 total adv loss: 0.34965639871779786

evaluating...

Confusion matrix, without normalization

[[ 209 63 22]

[ 648 3036 137]

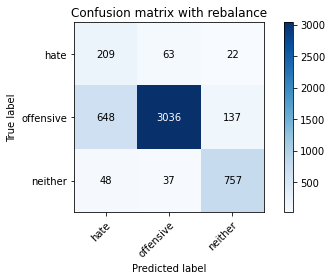
[ 48 37 757]]

Normalized confusion matrix

[[0.71088435 0.21428571 0.07482993]

[0.16958911 0.7945564 0.03585449]

[0.05700713 0.04394299 0.89904988]]





precision recall f1-score support

hate 0.23 0.71 0.35 294

offensive 0.97 0.79 0.87 3821

neither 0.83 0.90 0.86 842

accuracy 0.81 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.8501476356534348, accuracy: 0.8073431510994553

training 10 epoch...

103.6158881187439 seconds used......

10 training finished! average train loss: 0.011140943273400877

total clf loss: 1.914213470154209 total adv loss: 0.18224327027951404

evaluating...

Confusion matrix, without normalization

[[ 210 67 17]

[ 678 3026 117]

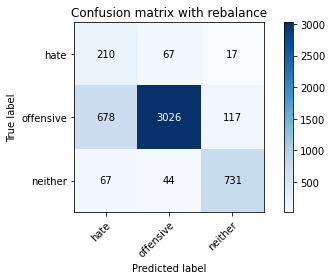
[ 67 44 731]]

Normalized confusion matrix

[[0.71428571 0.22789116 0.05782313]

[0.17744046 0.79193928 0.03062026]

[0.07957245 0.05225653 0.86817102]]





precision recall f1-score support

hate 0.22 0.71 0.34 294

offensive 0.96 0.79 0.87 3821

neither 0.85 0.87 0.86 842

accuracy 0.80 4957

macro avg 0.68 0.79 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.9650626998590005, accuracy: 0.8002824288884406

4 run...

training 1 epoch...

103.67081475257874 seconds used......

1 training finished! average train loss: 0.6286267165900671

total clf loss: 110.07919825613499 total adv loss: 51.68059182912111

evaluating...

Confusion matrix, without normalization

[[ 226 35 33]

[1159 2537 125]

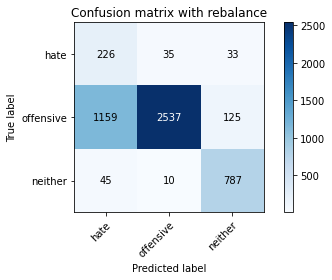
[ 45 10 787]]

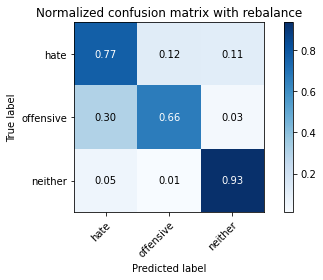
Normalized confusion matrix

[[0.76870748 0.11904762 0.1122449 ]

[0.30332374 0.66396231 0.03271395]

[0.05344418 0.01187648 0.93467933]]





precision recall f1-score support

hate 0.16 0.77 0.26 294

offensive 0.98 0.66 0.79 3821

neither 0.83 0.93 0.88 842

accuracy 0.72 4957

macro avg 0.66 0.79 0.65 4957

weighted avg 0.91 0.72 0.78 4957

average eval\_loss: 0.5864162663778951, accuracy: 0.716158967117208

training 2 epoch...

103.70569849014282 seconds used......

2 training finished! average train loss: 0.38822311829579503

total clf loss: 67.67335653305054 total adv loss: 25.744068494066596

evaluating...

Confusion matrix, without normalization

[[ 246 36 12]

[ 965 2777 79]

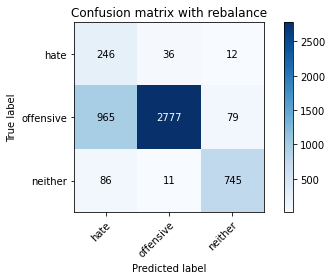
[ 86 11 745]]

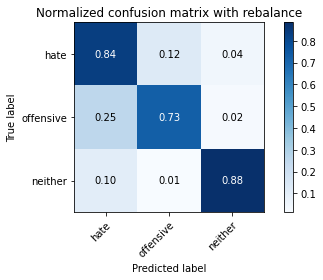
Normalized confusion matrix

[[0.83673469 0.12244898 0.04081633]

[0.25255169 0.7267731 0.02067522]

[0.10213777 0.01306413 0.8847981 ]]





precision recall f1-score support

hate 0.19 0.84 0.31 294

offensive 0.98 0.73 0.84 3821

neither 0.89 0.88 0.89 842

accuracy 0.76 4957

macro avg 0.69 0.82 0.68 4957

weighted avg 0.92 0.76 0.81 4957

average eval\_loss: 0.5173016200743376, accuracy: 0.760137179745814

training 3 epoch...

103.6911358833313 seconds used......

3 training finished! average train loss: 0.26785468349331304

total clf loss: 46.52026155218482 total adv loss: 14.34221309935674

evaluating...

Confusion matrix, without normalization

[[ 239 38 17]

[1167 2543 111]

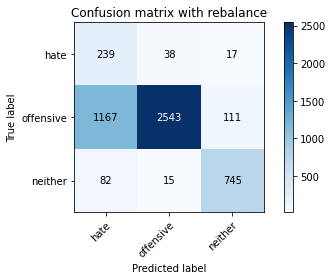
[ 82 15 745]]

Normalized confusion matrix

[[0.81292517 0.1292517 0.05782313]

[0.30541743 0.66553258 0.02904999]

[0.09738717 0.01781473 0.8847981 ]]





precision recall f1-score support

hate 0.16 0.81 0.27 294

offensive 0.98 0.67 0.79 3821

neither 0.85 0.88 0.87 842

accuracy 0.71 4957

macro avg 0.66 0.79 0.64 4957

weighted avg 0.91 0.71 0.77 4957

average eval\_loss: 0.613865188443132, accuracy: 0.7115190639499698

training 4 epoch...

103.68575024604797 seconds used......

4 training finished! average train loss: 0.17097370197985604

total clf loss: 29.557428122498095 total adv loss: 6.418502860375156

evaluating...

Confusion matrix, without normalization

[[ 224 52 18]

[ 769 2945 107]

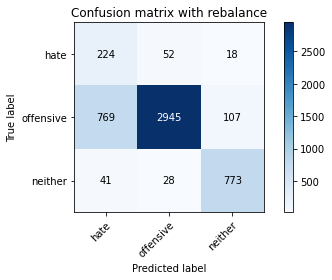
[ 41 28 773]]

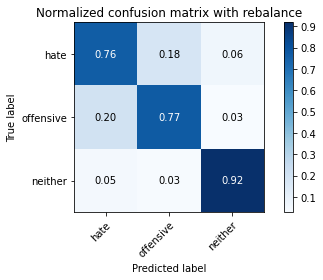
Normalized confusion matrix

[[0.76190476 0.17687075 0.06122449]

[0.20125622 0.77074064 0.02800314]

[0.04869359 0.03325416 0.91805226]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.77 0.86 3821

neither 0.86 0.92 0.89 842

accuracy 0.80 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.91 0.80 0.83 4957

average eval\_loss: 0.5588795601750814, accuracy: 0.795239055880573

training 5 epoch...

103.63429260253906 seconds used......

5 training finished! average train loss: 0.09750504454608723

total clf loss: 16.763419203925878 total adv loss: 1.8011317223517835

evaluating...

Confusion matrix, without normalization

[[ 219 52 23]

[ 717 2971 133]

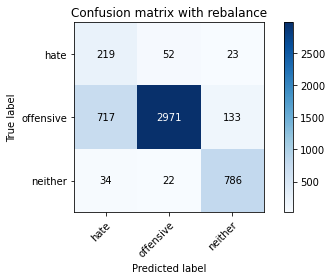
[ 34 22 786]]

Normalized confusion matrix

[[0.74489796 0.17687075 0.07823129]

[0.18764721 0.77754515 0.03480764]

[0.04038005 0.02612827 0.93349169]]





precision recall f1-score support

hate 0.23 0.74 0.35 294

offensive 0.98 0.78 0.87 3821

neither 0.83 0.93 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.6126530118842399, accuracy: 0.802098043171273

training 6 epoch...

103.6587028503418 seconds used......

6 training finished! average train loss: 0.04944631497631645

total clf loss: 8.526706682983786 total adv loss: 1.4277367557269827

evaluating...

Confusion matrix, without normalization

[[ 208 71 15]

[ 565 3170 86]

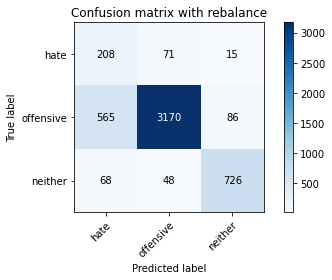
[ 68 48 726]]

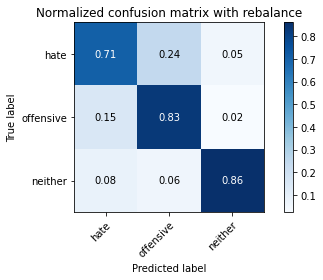
Normalized confusion matrix

[[0.70748299 0.2414966 0.05102041]

[0.14786705 0.82962575 0.0225072 ]

[0.0807601 0.05700713 0.86223278]]





precision recall f1-score support

hate 0.25 0.71 0.37 294

offensive 0.96 0.83 0.89 3821

neither 0.88 0.86 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: 0.6818721769602909, accuracy: 0.8279201129715553

training 7 epoch...

103.6878674030304 seconds used......

7 training finished! average train loss: 0.024815748308169825

total clf loss: 4.32898313854821 total adv loss: 1.7098035634369353

evaluating...

Confusion matrix, without normalization

[[ 202 75 17]

[ 496 3206 119]

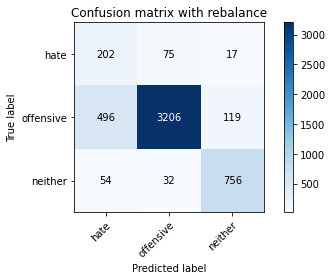
[ 54 32 756]]

Normalized confusion matrix

[[0.68707483 0.25510204 0.05782313]

[0.12980895 0.83904737 0.03114368]

[0.06413302 0.03800475 0.89786223]]





precision recall f1-score support

hate 0.27 0.69 0.39 294

offensive 0.97 0.84 0.90 3821

neither 0.85 0.90 0.87 842

accuracy 0.84 4957

macro avg 0.69 0.81 0.72 4957

weighted avg 0.91 0.84 0.86 4957

average eval\_loss: 0.6920074575032378, accuracy: 0.8400242081904378

training 8 epoch...

103.6503336429596 seconds used......

8 training finished! average train loss: 0.023284460388992678

total clf loss: 4.072949591529323 total adv loss: 1.8261370893480944

evaluating...

Confusion matrix, without normalization

[[ 223 55 16]

[ 795 2908 118]

[ 66 26 750]]

Normalized confusion matrix

[[0.7585034 0.18707483 0.05442177]

[0.20806072 0.76105731 0.03088197]

[0.0783848 0.03087886 0.89073634]]





precision recall f1-score support

hate 0.21 0.76 0.32 294

offensive 0.97 0.76 0.85 3821

neither 0.85 0.89 0.87 842

accuracy 0.78 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 1.0243688080189448, accuracy: 0.7829332257413758

training 9 epoch...

103.68855094909668 seconds used......

9 training finished! average train loss: 0.00839241172308892

total clf loss: 1.4529680747946259 total adv loss: 0.3573134748235134

evaluating...

Confusion matrix, without normalization

[[ 223 53 18]

[ 734 2976 111]

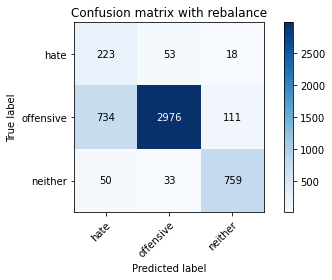
[ 50 33 759]]

Normalized confusion matrix

[[0.7585034 0.18027211 0.06122449]

[0.19209631 0.7788537 0.02904999]

[0.05938242 0.0391924 0.90142518]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.85 0.90 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 1.0757639510195605, accuracy: 0.7984668146056082

training 10 epoch...

103.73037457466125 seconds used......

10 training finished! average train loss: 0.014911668646259625

total clf loss: 2.652082533983048 total adv loss: 2.0437436349359928

evaluating...

Confusion matrix, without normalization

[[ 215 61 18]

[ 632 3089 100]

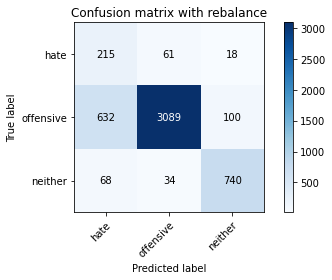
[ 68 34 740]]

Normalized confusion matrix

[[0.73129252 0.20748299 0.06122449]

[0.16540173 0.80842711 0.02617116]

[0.0807601 0.04038005 0.87885986]]





precision recall f1-score support

hate 0.23 0.73 0.36 294

offensive 0.97 0.81 0.88 3821

neither 0.86 0.88 0.87 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.70 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 1.0188636911041085, accuracy: 0.815816017752673